



THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

La justice prédictive

Application au droit des marques

Vandenberghé, Johan

Award date:
2020

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

UNIVERSITÉ DE NAMUR
Faculté d'informatique
Année académique 2019–2020

**La justice prédictive
Application au droit des marques**

Johan Vandenberghe



Promoteur : _____ (Signature pour approbation du dépôt - REE art. 40)
Jean-Marie Jacquet

Mémoire présenté en vue de l'obtention du grade de
Master en Sciences Informatiques.

Remerciements

Nous aimerions remercier en premier lieu le Professeur Jean-Marie Jacquet pour son soutien et son aide dans la réalisation de ce travail. Les professeurs Zorana Rosic et Benoit Michaux pour avoir apporté l'idée de l'application de la justice prédictive au droit des marques, pour le bref résumé du droit des marques et pour la liste des arrêts de jurisprudence significatif dans le domaine des marques qui m'ont permis de débiter le travail d'analyse.

Nous tenons aussi à remercier ma famille pour son soutien pendant la période de réalisation du travail.

Résumé

Il existe beaucoup de domaines qui utilisent les prédictions dans la vie courante. On peut citer, par exemple, la météorologie, l'astronomie, l'automobile, la médecine,... . Mais qu'en est-il du domaine de la justice ? On la retrouve principalement dans les pays reconnus comme étant à la pointe de la robotisation comme l'Estonie ou dans les pays où le secteur public délègue facilement le travail à des sociétés privées pour diminuer les coûts liés au secteur de la justice. Mais en Belgique, la justice prédictive est principalement utilisée par des cabinets d'avocat ou des sociétés d'assurance pour prédire le verdict d'une éventuelle action judiciaire.

Notre travail sur la prédiction a pour but de prédire si un signe peut être enregistré comme une marque en respectant le droit européen des marques. Nous utiliserons en premier lieu la méthode de la distance cosinus et des modèles algorithmiques fournis par le logiciel Orange pour déterminer si, sur base des arrêts de jurisprudence, nous pouvons réaliser une prédiction sur l'acceptation d'une marque. Par la suite, nous utiliserons la programmation logique déductive pour réaliser un programme en langage Prolog avec l'aide du logiciel SWI-Prolog. En se basant sur les exemples fournis par la jurisprudence, ce programme déterminera si un signe a des chances d'être enregistré en tant que marque.

Table des matières

1	Introduction	6
2	La justice prédictive	9
2.1	La justice	9
2.2	L'intelligence artificielle	10
2.3	Les prédictions	11
2.4	L'intelligence artificielle et la justice	12
2.5	La justice prédictive	16
2.6	Conclusion	17
3	Le droit des marques	18
3.1	Une marque	18
3.2	Les conditions de validité et le caractère distinctif d'une marque	22
3.3	Conclusion	24
4	Analyse des textes	25
4.1	La Théorie	25
4.1.1	La distance cosinus et euclidienne	27
4.1.2	Les modèles algorithmes d'apprentissage	31
4.1.3	Les tests en détail	33
4.2	Analyse des résultats	41
4.2.1	Analyse des résultats sur le jeu de données "Conclusion" en utilisant le module de test et résultat ^{10 p34}	41
4.2.2	Analyse des résultats sur le jeu de données "Acceptation" en utilisant le module de test et résultat ^{10 p34}	44
4.2.3	Analyse des résultats sur le jeu de données "Conclusion" en utilisant le module de prédiction ^{12 p34}	47
4.2.4	Analyse des résultats sur le jeu de données "Acceptation" en utilisant le module de prédiction ^{12 p34}	49
4.2.5	Analyse des résultats obtenus en utilisant les données reprises dans le fichier <i>valeurTest.csv</i> en utilisant le module de prédiction ^{12 p34}	53
4.3	Conclusion	56
5	La programmation logique	58
5.1	Théorie sur la programmation logique	58
5.2	La programmation logique inductive	60
5.3	Mise en pratique de la théorie de la programmation logique	61
5.4	Conclusion	71

6	Conclusion	72
A	Article 4, paragraphe 1 de la directive 2015/2436	77
B	Article 7, paragraphe 1 du règlement EU 2017/1001	79
C	Résumé des petites applications "widgets" utilisé dans le programme Orange, version 3.24.1 (data mining) [Borondics et al., 2020]	81
D	Liste de mots à enlever lors de la phase de pré-traitement des textes	85
E	Liste de mots à comptabiliser lors de la phase de pré-traitement du texte	86
F	Liste des arrêts utilisés dans la base de données	87
G	Différentes Bases de données et leurs contenus	94
H	Résultats des analyses des textes sous forme de questions	96
I	Résultats des tests sur le schéma repris à la figure 4.9 de la page 36	99
J	Représentation graphique des résultats obtenus sur le schéma repris à la figure 4.12 de la page 40 avec les textes repris dans la base de données "Conclusion" et "Concl Test"(voir l'annexe G de la page 95	112
K	Représentation graphique des résultats obtenus sur le schéma repris à la figure 4.12 de la page 40 avec les textes repris dans la base de données "Accept Test" et "Accept Ex"(voir l'annexe G de la page 95)	114
L	Affichage des valeurs reprises dans le fichier <i>valeurTest.csv</i>	116

Chapitre 1

Introduction

De tous temps, l'homme a voulu connaître ou contrôler son avenir en allant voir des voyants, des devins, des astrologues, des sages, ... dans le seul but de faire les bons choix pour le futur. Ces hommes et ces femmes qui prédisaient l'avenir se basaient sur leurs vécus, sur des faits réels, sur des dictons ou ne donnaient que des généralités, comme "*Demain, il fera beau s'il ne pleut pas*".

Aujourd'hui, les météorologues font des prédictions sur la météo qu'il fera demain, la semaine prochaine ou le mois prochain. Ils utilisent des algorithmes qui calculent les mouvements des anticyclones et des dépressions futurs en se basant sur les mouvements déjà enregistrés auparavant, exemple la figure 1.1 ci-dessous.

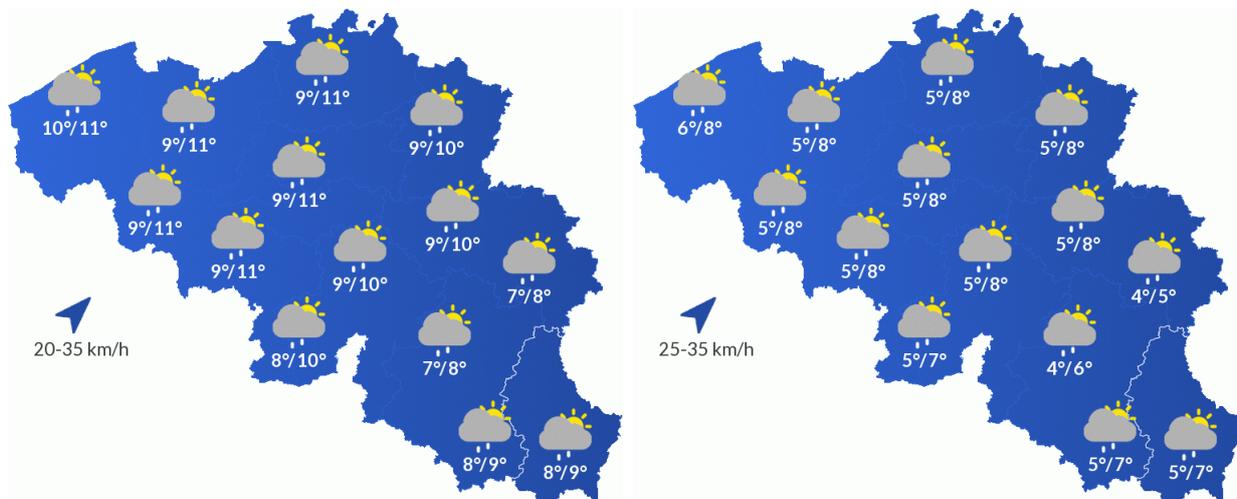


FIGURE 1.1 – Représentation graphique des prévisions météorologiques pour les 24 et 25 décembre 2019, consultée sur le site "<https://www.meteobelgique.be/previsions-meteo/belgique/mardi>", en date du 16 décembre 2019

D'autres domaines utilisent les algorithmes pour faciliter ou accélérer la recherche ou la prise de décision. Watson [Wikipédia, 2020o], le conseiller d'IBM, est un exemple de programme capable de comprendre le langage humain et est utilisé dans le domaine médical pour effectuer, en outre, des diagnostics.

D'une manière similaire la justice prédictive¹ évoque l'idée de pouvoir prédire, avec une certaine probabilité, une décision de justice sur base des décisions judiciaires précédentes. C'est un ensemble d'algorithmes qui donnera une réponse à une question qu'on se pose sur un fait qui pourrait arriver. Une prédiction, c'est donner un résultat probable d'un événement futur en se basant sur le passé.

Mais l'utilisation d'algorithmes n'est pas sans risque, car souvent un algorithme renforce les discriminations déjà existantes. Plusieurs sociétés en ont déjà fait les frais, comme Amazon qui après avoir lancé un programme d'aide à l'embauche de développeurs s'est vu contraint de le retirer car le programme discriminait tous les profils féminins. En effet, le programme avait été entraîné avec les profils des développeurs déjà présents dans la société (majoritairement masculin) [Signoret, 2018]. Un autre programme d'une grande société est soupçonné de sexisme dans l'octroi des plafonds pour des cartes de crédit [Dachez, 2019].

En matière de justice, les États-Unis utilisent des programmes pour, entre autre, aider les juges à déterminer le risque de récidives d'un criminel [van den Branden et al., 2019]. L'Estonie, un pays à la pointe de la numérisation des administrations, a mis en place une justice robotisée pour les litiges mineurs inférieurs à 7.000€ de dommages et intérêts [Balboni, 2019, Chabaud, 2020].

D'autres pays étudient la question de l'automatisation de la justice et des limites à y apporter. Comme pour la Belgique, où un certain nombre des membres de la justice comme l'avocat Adrien Van den Branden plaident pour une *justice 3.0* [van den Branden et al., 2019] pour aider la justice à résorber, entre autre, son retard. Pour passer d'une *justice 1.0*² [van den Branden et al., 2019] et arriver à cette digitalisation de la justice, il faudrait numériser toutes les décisions judiciaires et les rendre lisibles aux machines intelligentes.

Cependant l'application de l'automatisation des tâches ne peut pas se faire de manière uniforme. Il faut nuancer, certaines décisions de justice demandent un raisonnement et une sensibilité que la machine ne possède pas encore.

Ce travail, après avoir abordé l'automatisation dans la justice, se penche sur la prédictivité de l'acceptation d'un signe en tant que marque lors de la demande de protection d'un sigle dans le domaine du droit des marques. Les méthodes utilisées sont la distance cosinus et la programmation logique.

Le document se compose d'une introduction suivie du chapitre 2 qui donne un aperçu de la justice prédictive, les avantages et les risques d'utiliser des algorithmes pour prendre des décisions d'un point de vue de la justice.

Le chapitre 3 aborde les notions de droit des marques. Et le chapitre 4 explique les méthodes de la distance cosinus et euclidienne, et donne la définition des modèles algorithmiques utilisés dans le logiciel Orange [Borondics et al., 2020]. Ce chapitre se termine par l'analyse des résultats obtenus.

Dans le dernier chapitre, on abordera la théorie de la programmation logique pour finir par la mise en pratique de la théorie par la création d'un programme permettant de répondre à la question "*Un signe pourra-t-il être accepté en tant que marque en regard du droit des marques ?*".

1. en se basant sur la définition de la médecine prédictive du petit Larousse 2017

2. dans "*Les robots à l'assaut de la justice*" à la page 62

Le travail se clôture sur la conclusion qui nous rappelle les méthodes utilisées et nous donne des pistes pour améliorer le résultat dans un travail futur.

Chapitre 2

La justice prédictive

Ce chapitre aborde la problématique de l'intégration de l'intelligence artificielle dans le domaine de la justice. Mais pour bien comprendre l'ensemble, il faut comprendre chaque terme séparément. Les 3 premières sections définissent la justice, l'intelligence artificielle et les prédictions. La section qui suit abordera le rapport entre l'intelligence artificielle et la justice, leurs risques et leurs avantages. Et enfin, la dernière section aborde la justice prédictive qui est une des facettes de l'informatisation de la justice.

Commençons par définir la justice.

2.1 La justice

La justice est

- *"le principe moral qui exige le respect du et de l'équité."* [Lar, 2016].
- *"l'institution qui exerce un pouvoir juridictionnel"* [Lar, 2016].
- *"un principe philosophique, juridique et moral fondamental : suivant ce principe, les actions humaines doivent être approuvées ou rejetées en fonction de leur mérite au regard de la morale (le bien), du droit, de la vertu ou de tout autre norme de jugement des comportements"* [Wikipédia, 2020i].
- *"un ensemble d'institutions (police, tribunaux, prisons...) qui imposent le règne de la loi, sans lien nécessaire avec le principe philosophique. Elle est jugée fondamentale pour faire respecter les lois de l'autorité en place, légitime ou pas. La justice est censée punir quiconque ne respecte pas la loi avec une sanction ayant pour but de lui apprendre la loi et parfois de contribuer à la réparation des torts faits à autrui, au patrimoine privé ou commun ou à l'environnement"* [Wikipédia, 2020i].

On peut résumer l'histoire de la justice depuis le début de l'humanité à nos jours comme suit. Au commencement, la réparation du dommage était régi par la loi du plus fort (souvent avec l'aide du gourdin). Il fallut attendre l'Empire Romain pour obtenir *"la première justice-institutionnelle de l'histoire"* [Wikipédia, 2020i].

La justice peut être divisée en 3 phases de décision.

1. La décision d'aller en justice.
2. Le choix du jugement.
3. L'évaluation de la réparation.

Chacune des phases a un rapport différent avec l'intelligence artificielle

Mais avant d'aborder le rapport entre l'intelligence artificielle et la justice, il nous faut définir l'intelligence artificielle.

2.2 L'intelligence artificielle

L'intelligence artificielle est

- "*l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence*" [Wikipédia, 2020h].
- "*la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau, selon Marvin Lee Minsk*" [Wikipédia, 2020h].

Elle peut être divisée en 5 niveaux, selon Cédric de Serpos, CEO de Novasecur, dans l'article de Cecile Desjardins "Il y a cinq niveaux d'intelligence artificielle", [Desjardins, 2018].

- 1-2. La plupart des machines l'utilisent quotidiennement, la machine ne fait qu'exécuter du code qui lui a été fourni¹.
3. La machine effectue des prédictions sur base de données statistiques en sa possession.
4. Le "machine learning" et le "deep learning", la machine apprend et s'améliore en se corrigeant.
5. Le rêve de l'humanité mais aussi ce qu'il redoute le plus, l'autonomie complète de la machine (exemple "Terminator").

Remarque : ces niveaux sont discutables. Même au niveau 5, l'ordinateur ne fera qu'exécuter du code!

On situe les débuts de l'idée d'une intelligence artificielle vers 1950 avec le *test de turing* de Alan Turing [Wikipédia, 2020h]. Par la suite, il faudra attendre l'apparition de nouvelles technologies dans le domaine de l'informatique permettant d'augmenter la puissance des calculs pour voir apparaître l'apprentissage automatique dans les années 1980 qui aboutiront aux réseaux neuronaux [Wikipédia, 2020h]. Avec l'apparition du big data et de l'augmentation de la puissance de calcul des ordinateurs, on passe à l'apprentissage profond ou le "deep learning" [Wikipédia, 2020h].

Prenons quelques exemples d'intelligence artificielle de la vie courante ou de la fiction pour illustrer ce que l'homme a déjà réalisé ou a imaginé :

- Deep Blue qui a battu Garry Kasparov au jeu d'échecs en mai 1997 [Wikipédia, 2020h].
- AlphaGo qui a vaincu les meilleurs joueurs du monde au jeu de go entre octobre 2015 et mai 2017 [Wikipédia, 2020d].
- David, dans A.I. de Steven Spielberg, est un robot androïde qui devait remplacer un enfant malade qui a été cryogénisé en attente d'un remède efficace [Wikipédia, 2020b].
- Terminator, dans les "Terminator".

1. Cédric de Serpos regroupe les niveaux 1 et 2 en disant *Les niveaux 1 et 2 relèvent du traitement par des machines de scores, questionnaires, etc. en vue d'analyses de questionnaires ou d'informations statistiques.*"

- Sonny dans I, Robot de Alex Proyas.

Aujourd'hui, l'intelligence artificielle est partout. Elle a conquis tous les secteurs d'activités de notre quotidien. On parle de voitures autonomes, de drones, de chirurgies assistées, de nanorobots ou nanits, de feux de circulation dit intelligents, de robots ménagers, des chaînes de montage automatisées, ..., elle a pour but principal de nous aider dans nos tâches quotidiennes, de réduire les coûts de fabrication, d'aller là où l'être humain ne peut pas encore aller. En ce qui concerne le domaine de la justice, il existe des programmes de prédictions comme *Predictice* ou *Compas*, ou des sites d'arbitrages en ligne (eJust). Ces points seront vus plus en détails dans la section "L'intelligence artificielle et la justice".

Passons à la section suivante dans laquelle nous expliquerons ce qu'est une prédiction.

2.3 Les prédictions

Une prédiction est

- "*l'action de prédire ; divination*" [Lar, 2016].
- "*l'action d'annoncer à l'avance un événement par calcul, par raisonnement, par induction ; par métonymie, ce qui est ainsi annoncé*" [Wikipédia, 2019c].

Tout comme la justice, l'homme a, de tout temps, voulu prédire les événements à venir. Quelle serait l'issue d'une bataille ou quel temps fera-t-il demain, après demain, la semaine prochaine, ...? Les prédictions étaient réalisées par des devins, des oracles, des cartomanciennes, etc, ils utilisaient le raisonnement, l'observation de leur environnement ou l'histoire écrite ou orale pour annoncer les événements futurs d'une personne ou d'un peuple.

Prenons comme illustration, les dictons. Elles sont des prédictions basées sur l'observation du temps par nos ancêtres. Elles sont encore utilisées de nos jours.

Dans de nombreux domaines, on utilise l'informatique pour prédire ou prévenir des événements futurs. Prenons quelques modèles de la vie courante.

- Coyote a présenté, lors du salon de l'auto de Bruxelles de janvier 2020, son nouveau boîtier avec un système de sécurité prédictive [AdM avec A.Ca., 2020].
Le système Coyote utilise des algorithmes existant déjà dans les nouvelles voitures. Ces algorithmes informent les utilisateurs sur les zones dangereuses en se basant sur l'expérience cumulée de tous les utilisateurs du système Coyote.
- La Nasa utilise des modèles prédictifs pour déterminer si une planète découverte dans l'espace est potentiellement habitable (voir l'article "*La Nasa découvre une autre planète de la taille de la Terre dans une "zone habitable"*" publié sur le site de la rtbf le 7 janvier 2020 [Belga, 2020].

La section suivante aborde le rapport entre l'intelligence artificielle et la justice.

2.4 L'intelligence artificielle et la justice

Le rapport entre la justice et l'intelligence artificielle est difficile, car il peut être difficile de laisser à un robot l'autorité de juger. Accepterons-nous d'être jugés par un robot ? N'oublions pas que le "juge" est le représentant du 3^{ième} pouvoir, il fait respecter les lois prises au niveau législatif (le parlement) ou exécutif (le gouvernement). Et dès lors, on peut se poser la question "Peut-on laisser un "robot" contrôler le 3^{ième} pouvoir de l'État ?".

La robotisation de la justice ne veut pas dire "donner le contrôle absolu" comme dans le film *Judge Dredd* (1995) de Danny Cannon (adaptation de la bande dessinée "Judge Dredd") où les juges font respecter la loi en appliquant la loi au pied de la lettre, sans compromis comme des robots mais serait de tendre vers un "*fonctionnement plus efficient*" [van den Branden et al., 2019] c'est-à-dire, *des jugements d'une qualité acceptable, dans un délai raisonnable et à un prix abordable* [van den Branden et al., 2019]². Ceci pourrait être une solution à l'engorgement, à la lenteur et à l'accessibilité de la justice.

Alors que la plupart des domaines d'activités utilisent quotidiennement l'intelligence artificielle, seul la justice est restée dans l'expectative. L'avocat Adrien van den Branden³ compare la justice à une "*citadelle infranchissable à la technologie*" [van den Branden et al., 2019]. Elle est, selon lui, restée à l'air de la "*justice 1.0*" [van den Branden et al., 2019]. L'auteur fournit les 3 stades de l'informatisation de la justice pour étayer ses propos :

- * *justice 1.0 : informatisation de la justice*
- * *justice 2.0 : digitalisation de la justice*
- * *justice 3.0 : robotisation de la justice*

Analysons l'intégration de l'intelligence artificielle dans les différentes phases de la justice.

Pour la phase de décision d'aller en justice, l'intelligence artificielle est utilisée par les cabinets d'avocat. Ils utilisent, entre autre, des programmes qui consultent des bases de données privées pour prédire le résultat, comme *Predictice*, qui détermine les chances de résultat positif d'une action en justice [van den Branden et al., 2019]. On est, dans ce cas, dans de la justice prédictive⁴.

Pour la phase préparatoire des litiges, "*les robots d'analyse permettent de réduire le temps et le coût de recherche d'avant procès. Ces robots ne doivent pas être parfaits. Ils leur suffit d'être meilleur que l'humain*" [van den Branden et al., 2019]⁵. Mais le problème de la robotisation de la phase préparatoire ou de la phase de décision est l'accessibilité de la justice car ces programmes ne sont pas gratuit et dès lors pas disponible à tous les citoyens. Ce qui entraîne un fossé entre les gens qui ont accès à l'information et à une justice digitalisée et ceux qui ne l'ont pas.

Dans la phase du choix du jugement, les juges américains utilisent déjà des programmes d'aide comme *COMPAS* [van den Branden et al., 2019] qui déterminent le risque de récidive d'un condamné. Ou *ROSS*, développé par l'université de Toronto, qui est capable de rendre

2. selon Adrien van den Branden à la page 3 dans "*Les robots à l'assaut de la justice*"

3. dans "*Les robots à l'assaut de la justice*" à la page 2

4. Page 87 dans "*Les robots à l'assaut de la justice*", auteur Adrien van den Branden

5. selon Adrien van den Branden à la page 85 dans "*Les robots à l'assaut de la justice*"

un avis en se basant sur l'analyse de la jurisprudence [van den Branden et al., 2019]⁶. De même, l'Estonie qui est considérée par beaucoup comme à la pointe de l'intelligence artificielle a, depuis janvier 2020, automatisé les jugements pour les petits litiges inférieurs à 7.000€ [Chabaud, 2020].

Pour la phase d'évaluation de la réparation, l'intelligence artificielle est déjà bien implantée. Les assurances utilisent des algorithmes pour calculer les indemnités en se basant sur les indemnités déjà octroyées à leurs assurés. En Belgique, les amendes routières ont été automatisées depuis 2018, c'est le projet "Crossborder", ce qui a permis une augmentation de 21% des amendes envoyées [press@just.fgov.be, 2019]. Dans le film "*Le 5^{ème} élément*" de Luc Besson, sorti en 1997, l'acteur principal, Korb Dallas chauffeur de taxi, reçoit directement ses amendes dans le taxi au moment de l'infraction.

L'automatisation de la justice apporte des avantages non négligeables comme une diminution des coûts de fonctionnement, une diminution dans le temps de réponse d'un jugement, une impartialité théorique et une cohérence des jugements.

Cependant, il existe aussi des inconvénients à la robotisation de la justice. La mise en place aura un coût non négligeable au budget de la justice. Un algorithme ne prendra pas en compte l'environnement du procès ou la situation de l'accusé [van den Branden et al., 2019, Hubin et al., 2019]. Il ne fera qu'une application stricte de la loi, comme dans le film *Judge Dredd* (1995) de Danny Cannon, sans prendre en compte les circonstances atténuantes qui pourraient bénéficier à l'accusé.

Une robotisation n'est pas sans risque. Nous allons parcourir certains risques liés à l'introduction de l'intelligence artificielle dans le domaine de la justice.

Il existe le risque d'une privatisation de la justice comme le craignent les avocats et les juges suite à la mise en ligne du site "*justice.cool*" début 2020 [Chabaud, 2020]. Celui-ci propose, pour un forfait minimal, de régler en France les petits litiges ou bien encore l'arbitrage en ligne de la start-up suisse eJust [van den Branden et al., 2019]⁷.

Le risque d'une augmentation du nombre de personnes n'ayant plus les moyens d'accéder à la justice n'est pas négligeable à cause, entre autre, de la fracture numérique. En effet celle-ci existe déjà dans notre pays, on a pu le constater avec l'accès des enfants précarisés à l'information fournie par les écoles pendant la crise du "Corona virus" durant le printemps 2020.

Un autre exemple, "un assureur qui utilise la machine pour calculer les risques d'une *décision défavorable à son client contribue au désengorgement des tribunaux et à une marginalisation de l'assuré, il impose à son client la solution fournie par la machine sans en garantir la légitimité*"⁸ [Hubin et al., 2019].

Pouvons-nous toujours parler de procès équitables si l'accès à l'information n'est disponible qu'à une partie des justiciables car celui-ci n'est accessible qu'aux cabinets d'avocats rémunérant des sociétés donnant l'accès à l'information en "*open-data*"⁹.

6. A la page 86 et 87 dans "*Les robots à l'assaut de la justice*"

7. A la page 91 de "*Les robots à l'assaut de la justice*"

8. "*Le juge et l'algorithme : juges augmentés ou justice diminuée ?*" à la page 7 et 8

9. une information en "*open-data*" est une information accessible dans un format utilisable par des machines,

Dans les avantages, on parlait d'une "*impartialité théorique des jugements*", ce qui est vrai en théorie car si un robot dit que " $2 + 2 = 4$ " alors il ne dira jamais 5 sauf si on le programme pour le dire. En pratique, derrière le code source d'un programme, il y a toujours un être humain qui à la différence du robot est faillible. Il faut prendre en compte avec les opinions politiques, raciales, ... du développeur. Sans oublier que les programmes utilisent des bases de données pré-existantes, qui possèdent déjà des jugements pas toujours impartiaux. Il existe un risque d'augmenter les inégalités déjà existantes. Ils existent déjà beaucoup d'exemples d'intelligences artificielles qui ont fait scandale par leur discrimination envers les minorités car elles ne faisaient que reporter les discriminations pré-existantes. Citons, Amazon et son logiciel d'aide à l'embauche de développeurs et une grande société américaine et son algorithme d'octroi des plafonds de sa nouvelle carte de crédit.

Le programme d'Amazon [Signoret, 2018] devait sélectionner les profils qui se rapprochaient des profils des développeurs travaillant déjà dans l'entreprise. Mais le paramètre qu'Amazon n'avait pas pris en compte, c'est que les profils pour les postes de développeur travaillant déjà au sein de la société étaient en majorité masculin. L'algorithme rejetait dès lors tous les profils féminins car n'étant pas en adéquation avec le profil type et entraînait l'augmentation des profils masculins dans les postes de développeur ce qui confortait l'idée que le logiciel ne se trompait pas.

Le même problème se pose pour le logiciel d'octroi des crédits d'une grande société américaine [Dachez, 2019]. L'algorithme octroyant un plafond inférieur à la compagne par rapport à son conjoint, tout en renseignant les mêmes informations de revenus du couple.

Même si les erreurs des développeurs ne sont, en général, pas intentionnelles, on pourrait devoir faire face à des erreurs intentionnelles dictées par une poignée de grandes sociétés s'octroyant le droit de favoriser les jugements pour un petit nombre de personnes ayant les moyens car on pourrait légitimement accorder à ces sociétés le droit de cacher le code source des programmes qu'ils fournissent, au nom de la protection intellectuelle ou tout bêtement afin d'éviter qu'un groupe de criminel ne connaissent les failles des algorithmes qui devraient les juger.

Un autre problème qui se pose est le risque de lenteur dans l'optimisation des algorithmes [van den Branden et al., 2019] car pour optimiser un programme, il faut attendre le retour des résultats et identifier les erreurs possibles. Dans le cas de programme destiné à déterminer le risque de récidive d'un détenu. Il faudra attendre la fin de sa peine avant de constater si le résultat obtenu par le programme correspond à la réalité des faits, ce qui pourrait prendre quelques années avant d'obtenir suffisamment de résultats pour se faire une opinion des corrections à effectuer sur le programme.

Après avoir parcouru les principaux risques d'une robotisation de la justice, on peut se poser une question "quel est le niveau d'erreur qui pourrait être accepté dans le programme lié à la justice?". L'importance des erreurs est différente pour le citoyen et pour le condamné.

permettant des recherches et des comparaisons plus rapide que si un être humain devait effectué la recherche manuellement avec une information en "*open access*" car seulement accessible aux individus. Source page 65 de "*Les robots à l'assaut de la justice*" par Adrien van den Branden lisible par un algorithme ce qui permet, entre-autre, .

Le citoyen préférera un innocent en prison plutôt qu'un coupable en liberté, tant qu'il ne s'agit pas de lui-même, mais le condamné préférera être libre.

En règle général, les sociétés de développement comme Compas ont pris le parti qu'il vaut mieux un faux positif (une personne non coupable en prison) qu'un faux négatif (un coupable en liberté) car les conséquences d'une erreur "*sont supportées que par les récidivistes à qui l'entreprise n'a pas de comptes à rendre*"¹⁰ [van den Branden et al., 2019]. Les auteurs du livre "Le juge et l'algorithme : juges augmenté ou justice diminuée ?" à la page 29 nous disent la même chose "*les faux négatifs sont exclus, il est préférable un innocent en prison (faux-positifs) qu'un criminel en liberté (faux-négatifs) idée suivie par les sociétés qui fournissent les programmes à la justice*" [Hubin et al., 2019].

Pour terminer cette section, nous devons nous poser la question "quelle matière de la justice pourrions-nous automatiser?". Dans le livre "Les robots à l'assaut de la justice", Adrien van den Branden nous apporte une solution avec son spectre d'automatisation des litiges repris à la figure 2.1 à la page 16. On peut constater qu'il serait plus facile d'automatiser les litiges qui tombent dans le domaine de la justice réparatrice que dans le domaine de la justice distributive. La justice réparatrice tend à réparer un dommage qui a été occasionné. On peut lui appliquer les règles de manière stricte. En opposition avec la justice distributive pour laquelle le juge doit trancher pour déterminer la partie qui revient à chaque partie. Prenons un exemple, vous roulez à 100 km/h sur une route limitée à 90 km/h, l'amende sera de 56€ car c'est le montant habituellement pour ce type d'infraction. Mais pour déterminer la garde d'enfant, on ne peut pas se baser sur l'ensemble des arrêts déjà rendu pour déterminer le verdict car il devrait être personnalisé à l'affaire en cours.

Dans le livre "Le juge et l'algorithme : juges augmentés ou justice diminuée?", de la page 18 à 21, les auteurs ont une vision proche de l'avocat Adrien van den Branden. Les auteurs séparent les affaires en 2 catégories, les simples et les complexes. Les affaires simples sont des affaires où le juge applique des règles de droit objectif caractérisées par une grande précision [Hubin et al., 2019]. Et les affaires complexes sont des affaires dans lesquelles le juge est amené à appliquer des règles de droit objectif caractérisées par une certaine abstraction [Hubin et al., 2019] avec l'obligation de devoir motiver la réponse en l'absence de définition claire dans les textes légaux.

Pour finir ce chapitre, existe-t-il déjà une justice prédictive ?

10. A la page 11 de "*Les robots à l'assaut de la justice*"

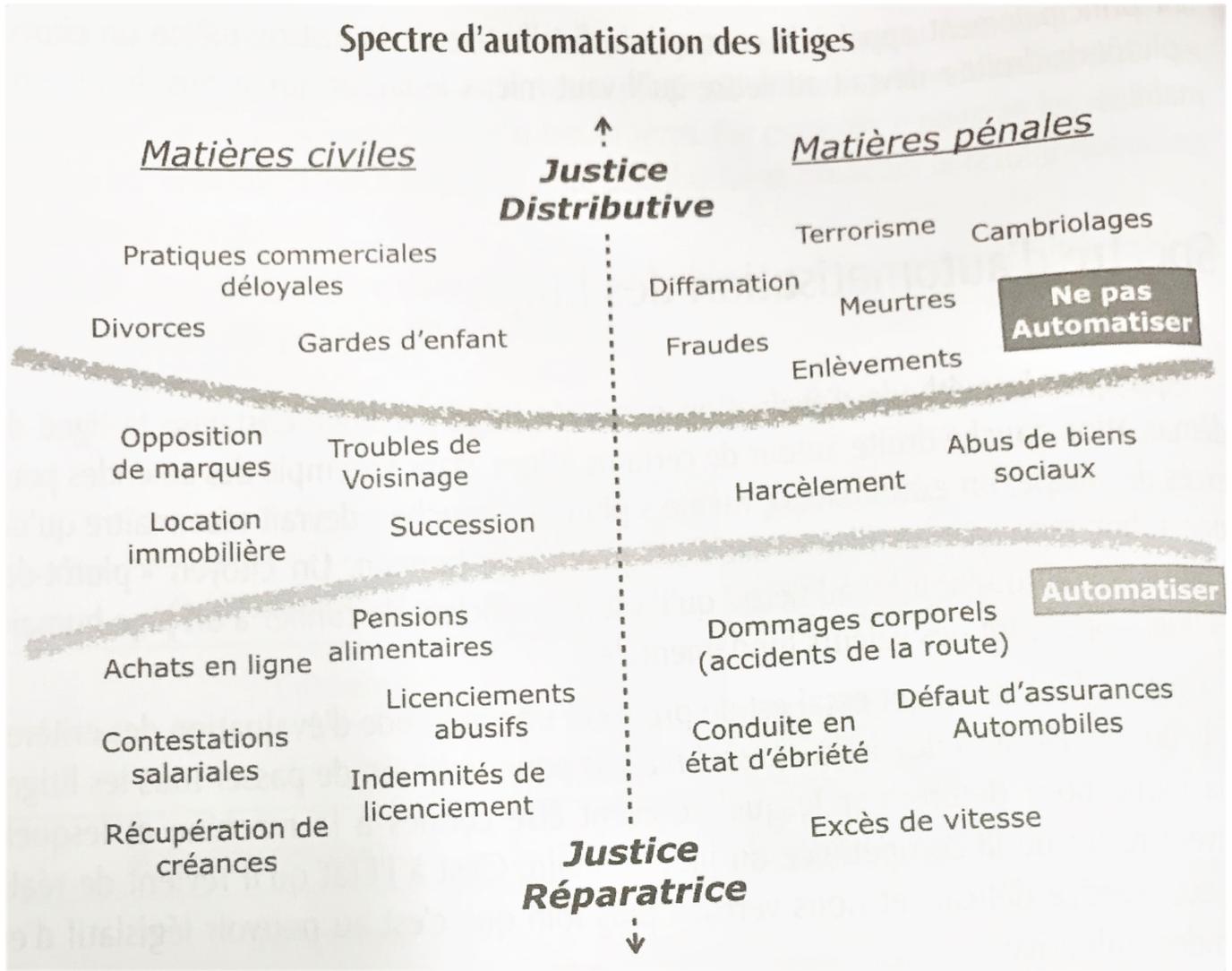


FIGURE 2.1 – Spectre d'automatisation des litiges
 Source : "Les robots à l'assaut de la justice" page 59

2.5 La justice prédictive

On peut dire, en se basant sur les définitions vues dans les sections précédentes, que la justice prédictive est l'annonce du résultat probable d'une décision judiciaire ou de réparation. Car on sait que l'homme a de tout temps voulu prédire des événements futurs et obtenir réparation pour des préjudices passés. Maintenant, il voudrait prédire la réparation d'un préjudice qui n'aura peut-être pas lieu si la prédiction n'est pas à son avantage ou prédire les infractions avant que celles-ci ne soient commises comme dans le film *"Minority Report"* de Steven Spielberg, sorti en 2002, où le simple fait de penser commettre un crime vous envoie directement en prison. Mais, en réalité, cela n'est pas possible actuellement, car les lois ne prévoient pas de condamnation sur le seul fait de penser commettre un crime cela reviendrait à nier la présomption d'innocence [Hubin et al., 2019] et n'oublions pas que "la charge de la preuve incombe à la partie poursuivante et que le doute profite à l'accusé, lequel a le droit d'adopter une attitude passive et bénéficie du droit au silence"¹¹ [Hubin et al., 2019] sauf en cas de terrorisme. Il reviendra quand même au procureur de prouver que le stade préparatoire de l'acte criminel est

11. à la page 66 du livre "Le juge et l'algorithme : juges augmentés ou justice diminuée"

déjà à un stade très avancé.

Au niveau de la police, la police prédictive est déjà une réalité. La police utilise des algorithmes pour déterminer les zones les plus sensibles afin d'adapter la réponse la plus appropriée à apporter. Car, comme l'écrit l'auteur Adrien van den Branden, dans "Les robots à l'assaut de la justice", à la page 89, "*mieux comprendre ce qui favorise l'émergence des crimes ne peut qu'aider les policiers dans leur travail*" [van den Branden et al., 2019].

Au niveau du fonctionnement de la justice, comme nous l'avons vu dans la section précédente, les avocats et les assurances utilisent déjà la prédiction pour déterminer les chances de réussite d'une affaire. Elles doivent pour cela utiliser les services de sociétés privées qui ont digitalisés l'information nécessaire à cette prédiction. Pour le grand public ou tout simplement pour éviter que l'accès à la justice soit réservée à l'élite en Belgique, la justice devrait évoluer vers une digitalisation des arrêts de justice ce qui rendrait possible un travail de prédiction.

Mais la justice prédictive n'est pas encore prête à rentrer dans tous les tribunaux car comme le dit Frédéric Rouvière, dans "*Le raisonnement par algorithmes : le fantasme du juge-robot*", lors d'un colloque organisé par la Cour de cassation de France le 12 février 2018, "les trois contributeurs viennent nourrir le sentiment général de défiance autour de la justice prédictive ... " [Rouvière, 2018]. Il existe encore beaucoup de préjuger au sein même de la magistrature.

La suite de notre travail se focalisera sur la justice prédictive appliquée au droit des marques. Nous pensons qu'il doit être possible de prédire si un signe peut-être enregistré comme marque en respectant les règles du droit européen des marques (comme nous le verrons dans le chapitre suivant). Seul l'interprétation de la loi peut engendrer des conflits sur la nullité de l'enregistrement d'un signe en tant que marque.

2.6 Conclusion

En conclusion, nous pouvons dire que la robotisation de la justice n'est pas vue d'un bon œil par tout le monde. Il faudra l'encadrer pour éviter les dérives et apaiser les réticences des plus anciens. Mais il pourrait aussi apporter une diminution des délais, une diminution des coûts ce dont la justice a grand besoin et ne devrait pas être appliquée à toutes les matières de la justice.

Pour ce qui concerne la justice prédictive, elle est déjà bien présente dans le domaine de la justice mais n'est pas accessible à tous le monde.

La réalisation de ce chapitre se base principalement sur deux livres, "Les robots à l'assaut de la justice : L'intelligence artificielle au service des justiciables" de Adrien van den Branden et "Le juge et l'algorithme : juges augmentés ou justice diminuée?" sous la coordination de Jean-Benoît Hubin, de Hervé Jacquemin et Benoît Michaux avec la contribution de Karim Benyekhlef, Valentin Callipel, Philippe Dambly, Jean de Codt, Loïck Gérard, Jean-Benoît Hubin, Hervé Jacquemin, Olivier Leroux, Bruno Mathis, Marie Messiaen, Dominique Mougenot, Yves Pouillet, Hugo Ruggieri et Jie Zhu. Complété par la lecture d'articles publiés sur le sujet et de mes connaissances acquises au cours de mes cursus antérieurs.

Dans le chapitre suivant, nous aborderons la théorie des marques et du droit qui l'encadre.

Chapitre 3

Le droit des marques

Nous aborderons, dans ce chapitre, la théorie sur le droit des marques. La première section définit la marque et fournit les représentations et les types de marques que l'on peut enregistrer. Dans la section suivante, nous verrons les conditions de validité pour l'enregistrement d'une marque et plus spécifiquement le caractère distinctif que doit posséder une marque pour être considérée comme valide.

3.1 Une marque

Une marque peut se définir comme

- l' *"Ensemble des produits fabriqués, vendus sous un label"* selon "Le petit Larousse illustré 2017" édité par Larousse.
- *"des signes utilisés dans le commerce pour identifier des produits. Votre marque est le symbole par lequel vos clients vous reconnaissent. Elle vous différencie de vos concurrents"* selon, le site "euipo.europa.eu" sur la page "<https://euipo.europa.eu/ohimportal/fr/trade-mark-definition>" en date du 13 octobre 2019.
- *"un signe qui est utilisé par une entreprise afin d'identifier ses produits et services. A la vue de ce signe, le public pourra les distinguer des produits et services mis sur le marché par une autre entreprise"* selon, le site économique du gouvernement fédéral de Belgique sur la page "<https://economie.fgov.be/fr/themes/propriete-intellectuelle/marque>" en date du 16 octobre 2019.

Mais attention de ne pas confondre une marque avec un signe. En effet une marque est un signe mais un signe peut ne pas être une marque. Imaginons la représentation d'une orange est un signe mais ne pourra pas être une marque si on désire qu'elle soit la représentation d'une marque de fruit car le signe de l'orange ne pourra plus être utilisé par les autres vendeurs d'orange et le public ne pourra pas distinguer la marque parmi tout les vendeurs de fruits comme appartenant à un vendeur spécifique. Mais le signe de l'orange pourrait être utilisé pour un vendeur d'IT. Comme nous vous l'expliquerons plus loin dans ce chapitre.

Il existe plusieurs représentations possible pour une marque [euipo.europa.eu, 2020b, Parlement Européen et Conseil de l'Union Européenne, 2015], on peut la représenter par :

- un mot, une suite de mots, des lettres, des chiffres, ou une combinaison de mots, de lettres ou de chiffres, c'est une marque verbale

Total, Peugeot, Ford

- un dessin, un logo,..., c'est une marque figurative



Logo de Total

- une combinaison d'éléments figuratifs et verbaux



- un produit qui par sa forme représente la marque, c'est une marque de forme



Marque de batterie, propriété de Duracell Batteries BVBA

- une combinaison d'éléments verbaux sur un produit, c'est une marque de forme comportant des éléments verbaux



- la position caractéristique d'une forme sur un produit, c'est une marque de position



Marque de vêtements propriété de JIMA Projects BV

- la répétition régulière d'éléments ou d'un motif, c'est une marque de motif



Marque de parfums, propriété de BURBERRY LIMITED

- une couleur, c'est une marque de couleur



Marque de friandise, propriété de Kraft Foods Schweiz Holding GmbH

- une combinaison de couleurs, c'est une marque de couleurs



Marque appartenant à BMW

- un son ou une combinaison de sons, c'est une marque sonore

<http://euipo.europa.eu/trademark/sound/EM500000017592031>

Marque appartenant à Lidl

- un mouvement d'un objet, d'une image ou d'un logo, c'est une marque de mouvement
<https://euipo.europa.eu/copla/image/UB2V7J5SX5HSOT2UA6C4K4UFFSCMH2SD4KY7WU4YJAAKRUTPKI63B4FTY7SAJGVMVSGC62BVJ73WU>
 Marque appartenant à Vodafone
- une combinaison d'images ou d'objet en mouvements et de sons, c'est une marque multi-média
<https://euipo.europa.eu/copla/image/UB2V7J5SX5HSOT2UA6C4K4UFFSBSWY2WJEO4Q43NOQJPBYJNH56IGMZ3EF0BAUXRUM3IKZVRQDAUG>
 Marque appartenant à "Fundació per a la Universitat Oberta de Catalunya"
- un hologramme, c'est une marque hologramme

On distingue aussi 3 types de marques [euipo.europa.eu, 2020a]

- les marques individuelles.
- les marques collectives¹ qui regroupent des produits ou des services ayant une même caractéristique, comme un signe régional garantissant la provenance du produits.
- les marques de certification² qui apportent la garantie que le produit ou le service respecte des critères de certification stricts émis par le propriétaire de la marque de certification, par exemple une marque de certification d'un produit bio.

Pour information, il y existe plus de 11 millions de marques³ déposées dans l'Union Européenne.

Il est donc nécessaire d'enregistrer une marque, car, contrairement au droit d'auteur⁴, la protection de la marque n'existe que si elle est enregistrée, que dans la zone géographique de son enregistrement et que pour une série de classe de produit ou de service précis et définis lors de l'enregistrement.

Pour enregistrer une marque, il existe 4 voix :

- la voie nationale.
- la voie régionale (exemple, l'Office Benelux de la Propriété Intellectuel (BOIP)).
- la voie européenne (via l'Office de l'Union Européenne pour la Propriété Intellectuel (EUIPO)).
- la voie internationale (via l'OMPI).

C'est l'autorité compétente qui vérifiera si les conditions d'enregistrement et si les règles régissant le droit des marques sont respectées.

Lors de l'enregistrement de la marque, il est important de choisir les classes de produits et de services associés à la marque. Car les conditions de validité de l'enregistrement de la marque dépendront aussi de ceux-ci.

1. article 74 à 82 du règlement EU 2017/1001

2. article 83 à 93 du règlement EU 2017/1001

3. chiffre fourni par l'EUIPO sur la page internet <https://euipo.europa.eu/ohimportal/fr/search-availability>

4. le droit d'auteur existe dès la naissance de la création artistique ou littéraire

Cheminement d'une demande d'enregistrement d'une propriété intellectuelle fourni par l'EUIPO.

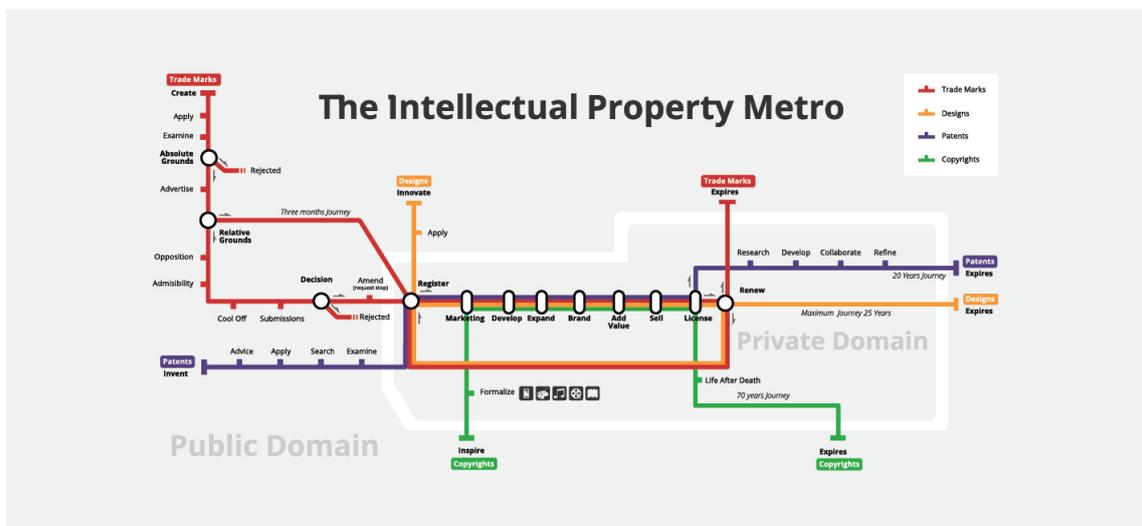


FIGURE 3.1 – Le métro de la propriété intellectuel
 source : <https://euipo.europa.eu/ohimportal/fr/trade-mark-definition>

La figure 3.1 représente la ligne de vie d'un brevet, d'une marque, d'un copyright ou d'un droit d'auteur de son enregistrement à l'expiration de la protection qui a été généré lors de l'enregistrement. Chaque ligne représente une des 4 propriétés intellectuelles et chaque station du métro indique les étapes pour chaque demande.

Prenons le cas d'une marque, la station de départ est la création de la marque suivi de la demande d'enregistrement. Il s'en suit l'examen de la demande et quelques stations plus loin l'enregistrement de la marque. Par la suite, il existe différentes stations reprenant la vie de la marque jusqu'à la station de renouvellement de la marque ou la fin de l'enregistrement de la marque. On constate que seul la marque possède une boucle permettant de renouveler l'enregistrement.

La section suivante aborde les conditions de validité de l'enregistrement d'une marque et plus spécifiquement le caractère distinctif de celle-ci.

3.2 Les conditions de validité et le caractère distinctif d'une marque

Pour considérer qu'une marque a été valablement déposée à l'autorité compétente⁵, elle doit respecter 3 conditions

1. avoir un caractère distinctif
2. être licite dans la zone géographique de la demande
3. être disponible dans
la classe de produit et de service souhaité
et la zone géographique de l'enregistrement

Nous commencerons par détailler les 2 derniers points. Une marque doit être licite dans la zone géographique de la demande. Elle pourrait donc être totalement illicite dans une autre zone de la demande et être valable pour la zone de la demande. Un signe est illicite si [SPF Economie, 2019]

- elle est *contraire à l'ordre public ou aux bonnes mœurs*.
- elle représente *les armoiries, drapeaux et autres emblèmes d'Etat, les emblèmes et les dénominations des organismes internationaux, des signes officiels de garantie et de contrôle*.
- elle est *susceptible de tromper le public, par exemple sur la nature, la qualité ou la provenance géographique des produits ou services*.
- elle *contient une indication géographique qui ne correspond pas à l'origine réelle des produits* cette règle s'applique pour les vins et spiritueux.

Une marque doit être disponible dans la classe de produits et de service souhaités et dans la zone géographique de l'enregistrement. *Disponible* signifie qu'aucune autre entité n'a déposé une marque semblable dans une des zones géographique et dans une des classes de produit et de service de la demande. Les 2 critères doivent être pris conjointement, car une marque pourrait être identique à une autre marque déjà enregistrée dans la zone géographique de la demande mais pour des classes de produit et de service différents et inversement si une autre marque est enregistrée dans la classe de produit et de service mais dans une autre zone territoriale. Ceci est vrai dans presque tout les cas sauf pour les marques ayant une renommée, mais ce cas sera vu à la page 23.

L'élément important dans l'étude du droit des marques est le caractère distinctif d'une marque. C'est ce qui distingue les produits ou les services d'une entreprise des autres produits ou services. Ce caractère ne peut pas être descriptif, car il ne doit pas décrire la marque mais la distinguer des produits ou services des autres marques.

Pour déterminer si une marque sera refusée sur base du caractère distinctif lors de l'enregistrement au sein de l'EUIPO, on prend en compte les articles :

- 4, paragraphe 1 de la directive 2015/2436
[Parlement Européen et Conseil de l'Union Européenne, 2015]
Cet article liste les différents motifs de refus ou de nullité lors de l'enregistrement d'une marque comme les marques qui n'ont pas de caractère distinctif ou les marques "*qui sont contraires à l'ordre public*" [Parlement Européen et Conseil de l'Union Européenne, 2015]

5. article 7 et 8 du règlement EU 2017/1001

ou les marques "*qui sont de nature à tromper le public*"

[Parlement Européen et Conseil de l'Union Européenne, 2015](voir annexe A).

- 7, paragraphe 1 du règlement EU 2017/1001

[Parlement Européen et Conseil de l'Union Européenne, 2017]

L'article nous indique les motifs entraînant le refus de l'enregistrement d'une marque comme les marques n'ayant pas de caractère distinctif ou les marques utilisant des signes "*devenus usuels dans le langage courant*"

[Parlement Européen et Conseil de l'Union Européenne, 2017] ou les marques constitués de signes qui pourrait être utilisés pour désigner "*l'espèce, la qualité, la quantité,...*"

[Parlement Européen et Conseil de l'Union Européenne, 2017] du produit ou du service ou "*les signes constitués exclusivement par la forme imposée par la nature même du produit*" [Parlement Européen et Conseil de l'Union Européenne, 2017] (voir annexe B)

Ces 2 articles donnent l'ensemble des cas où une demande d'enregistrement serait refusée.

Le caractère distinctif d'une marque s'apprécie en fonction ⁶

- du territoire dans lequel on effectue la demande d'enregistrement de la marque (nationale, régionale, européen).
- des classes de produits et services choisis au moment de la demande
- du consommateurs du produit ou du service (le "*public pertinent*")
- de l'intérêt général
- du moment du dépôt de la demande d'enregistrement de la marque

En cas d'acceptation du signe par l'autorité de l'enregistrement de la marque, celle-ci est dès lors protégée dans le territoire de la demande pour les classes de produits et de services sélectionnées(et les classes de produits et de services similaires) ⁷.

Comme nous l'avons vu plus haut, il existe des marques renommées. Une marque ayant une renommée est une marque qui est reconnue par un large public et qui ne pourra pas être utilisée par une autre entreprise même si la demande d'enregistrement à lieu dans une autre classe de produits ou de services. En effet, le public pourrait faire un lien avec le produit reconnu apportant un avantage au produit portant la même marque ou à l'inverse ayant un effet négatif sur la renommée du produit de la marque d'origine.

Expliquons cela en parlant du litige entre le fabricant de voiture "Jaguar" et le fabricant de vélo "Jaguar" [Maitre Frédéric Lejeune, 2019].



FIGURE 3.2 – Logo des marques en litige

Le fabricant de vélo et son distributeur ont été reconnus coupables de contrefaçon par le tribunal car il bénéficiait de la renommée de la marque Jaguar.

6. source : "Liste d'arrêts CJUE – caractère distinctif (DM) Bref résumé de la matière pour l'ILP à travers la jurisprudence de la Cour de justice de l'Union européenne (CJUE)" du 4 octobre 2019, auteur Zorana Rosic

7. article 9 et 10 du règlement EU 2017/1001

3.3 Conclusion

En conclusion, on peut dire que l'enregistrement d'une marque est régi par des règles européennes n'autorisant pas les arrangements entre copains. Elle doit respecter les règles strictes du droit des marques et les décisions de justice intervenue dans les arrêts de jurisprudence.

Ce Chapitre sur la théorie du droit des marques s'appuie principalement sur le résumé de la théorie du droit des marques fourni par la professeur Zorana Rosic intitulé "Liste d'arrêts CJUE – caractère distinctif (DM) Bref résumé de la matière pour l'ILP à travers la jurisprudence de la Cour de justice de l'Union européenne (CJUE)", sur les articles 4 et 7 repris aux annexes A et B et sur les sites internet de l'Office de l'Union européenne pour la propriété intellectuelle (EUIPO depuis le 23 mars 2016) et de la propriété intellectuelle du gouvernement fédéral belge.

La suite du travail portera sur l'analyse de texte de jurisprudence du droit des marques et par la création d'un programme sous forme de questions réponses qui nous permettra de déterminer si un signe respecte les règles pour l'enregistrement en tant que marque.

Chapitre 4

Analyse des textes

Après avoir vu la théorie sur les marques, voyons quelles méthodes nous pouvons utiliser pour comparer des textes. Par la suite, nous appliquerons ces méthodes aux arrêts de jurisprudence du droit des marques. Afin de déterminer s'il est possible de prédire l'acceptation d'un signe en tant que marque.



FIGURE 4.1 – Logo du logiciel Orange
source : <https://orange.biolab.si/>

Orange est un logiciel libre d'apprentissage machine ("machine learning") et d'analyse de données. En annexe C, page 84, nous avons mis les widgets utilisés pour les tests avec le logiciel Orange ainsi qu'un résumé de leurs fonctionnements.

Ce chapitre est séparé en 2 parties, la partie théorique et la partie d'analyse de résultats. Dans la première partie, nous détaillerons les méthodes de calcul de distance entre 2 textes afin de déterminer s'ils sont proches. Ensuite nous parcourrons les outils algorithmiques fournis par le logiciel Orange [Borondics et al., 2020]. Et pour finir, nous parcourrons en détail les différents tests qui seront réalisés.

La seconde partie, analysera en détail les résultats de prédiction obtenus avec le programme Orange [Borondics et al., 2020].

4.1 La Théorie

Pour analyser des textes, il faut pouvoir les comparer. Prenons comme illustration les 3 textes A , B et C repris ci-dessous.

A : *Le ciel est gris, il fera beau demain*

B : *Le ciel est bleu, il fera gris demain*

C : *Le ciel est vert, il faut que je change de lunettes demain*

On pourrait dire, à première vue que A et B sont "presque identiques", car ils ont le même nombre de mots et que 7 mots sur 8 sont identiques. Mais que vaut "presque identique"? Ce n'est pas mesurable. On peut aussi se poser une autre question, qui de A ou de B est le plus

proche de C ? C'est difficile à dire. On peut dès lors en conclure qu'il est difficile de comparer plusieurs textes dans leurs états d'origine. Nous devons les transformer en données numériques. Nous allons transformer un texte en une liste de nombres représentant les poids de chaque mots du texte en fonction de la totalité des mots des documents à comparer.

Ici, nous ne reprendrons pas dans la liste les verbes et les déterminants. Maintenant pour chaque phrase, on a un tableau reprenant l'ensemble des mots des 3 phrases et le poids de chaque mot dans chaque phrase :

A		B		C	
ciel	1	ciel	1	ciel	1
gris	1	gris	1	gris	0
bleu	0	bleu	1	bleu	0
beau	1	beau	0	beau	0
demain	1	demain	1	demain	1
vert	0	vert	0	vert	1
lunettes	0	lunettes	0	lunettes	1

FIGURE 4.2 – Tableau représentant les mots et leurs poids dans chaque phrase

On représente chaque phrase par une suite de chiffres représentant la fréquence de chaque mots dans les phrases :

$$A = \{1, 1, 0, 1, 1, 0, 0\}$$

$$B = \{1, 1, 1, 0, 1, 0, 0\}$$

$$C = \{1, 0, 0, 0, 1, 1, 1\}$$

Il existe différentes manières de calculer la fréquence ou le poids de chaque mot dans la phrase. La première méthode est la fréquence comme vu ci-dessus. Ensuite, on peut rapporter le nombre de mots dans la phrase sur le nombre total de mots différents dans les phrases, ce qui donnerait ceci [Negre, 2013]

$$A = \left\{ \frac{1}{7}, \frac{1}{7}, 0, \frac{1}{7}, \frac{1}{7}, 0, 0 \right\}$$

$$B = \left\{ \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, 0, \frac{1}{7}, 0, 0 \right\}$$

$$C = \left\{ \frac{1}{7}, 0, 0, 0, \frac{1}{7}, \frac{1}{7}, \frac{1}{7} \right\}$$

où 7 représente le nombre total de mots différents dans les 3 phrases.

Ou on peut aussi répondre à la question "*Le mot X est-il dans la phrase ?*" où

1 signifie que le mot existe dans la phrase

0 signifie que le mot n'existe pas dans la phrase

sans tenir compte de la fréquence des mots dans la phrase, il s'agit du poids booléen.

Tous les résultats obtenus ressemblent à des coordonnées de points dans un graphique à 7 dimensions.

Et si on possède les coordonnées de 2 points, on peut calculer la distance qui les sépare. Quand on connaît la distance entre 2 points, on peut les comparer aux autres points. Et si un

point est un texte alors on peut comparer plusieurs textes.

Nous parlerons dans la partie suivante la distance cosinus et la distance euclidienne qui sont les plus connues.

4.1.1 La distance cosinus et euclidienne

En analyse de texte, on utilise souvent la distance cosinus (ou similarité cosinus) ou la distance euclidienne pour comparer la similarité des documents [Wikipédia, 2020n, Negre, 2013].

La distance cosinus est une méthode qui *"permet de calculer la similarité entre deux vecteurs à n dimensions en déterminant le cosinus de l'angle entre eux"* [Wikipédia, 2020n].

La distance euclidienne *"calcule la similarité entre deux documents d_1 et d_2 comme la distance entre leurs représentations vectorielles ramenées à un seul point"* [Negre, 2013].

Le problème qui se pose est de calculer la distance entre 2 points ou la distance AB ou encore la distance qui sépare la phrase A de la phrase B . Ensuite, nous pourrions dès lors, déterminer la distance la plus courte entre les phrases et ainsi déduire le niveau de similarité entre les phrases.

Pour rappel, pour la distance euclidienne, plus la distance est petite plus il y a similarité entre 2 textes. Et pour la distance cosinus, il y a similitude entre 2 textes quand le résultat se rapproche de 1.

Prenons le cas où nous avons 3 points A , B et C ayant comme coordonnées (x_a, y_a) , (x_b, y_b) et (x_c, y_c) . Si nous les plaçons sur un graphe, on obtient pour $A = \{1, 1\}$, $B = \{2, 3\}$ et $C = \{1, 3\}$:

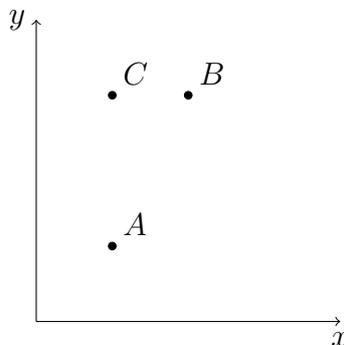


FIGURE 4.3 – Graphe représentant 3 points A , B et C sur un plan à 2 dimensions

Pour calculer la distance entre 2 points, on peut appliquer la formule de la distance euclidienne [Wikipédia, 2018, Wikipédia, 2020g]

$$AB = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}$$

On obtient

$$BC < AC < AB$$

car $BC = 1$, $AC = 2$ et $AB \approx 2,24$

Nous pouvons dire que C est plus proche de B que de A . Ce qui est facilement déductible en

regardant le graphe à la figure 4.3.

Mais que se passe-t-il si les points ne sont plus dans un système à 2 dimensions mais à n dimensions? Reprenons les points précédents A , B et C ayant comme nouvelles coordonnées (a_1, \dots, a_n) , (b_1, \dots, b_n) et (c_1, \dots, c_n) . Nous plaçons les points sur un graphe, nous obtenons une des représentations ci-dessous

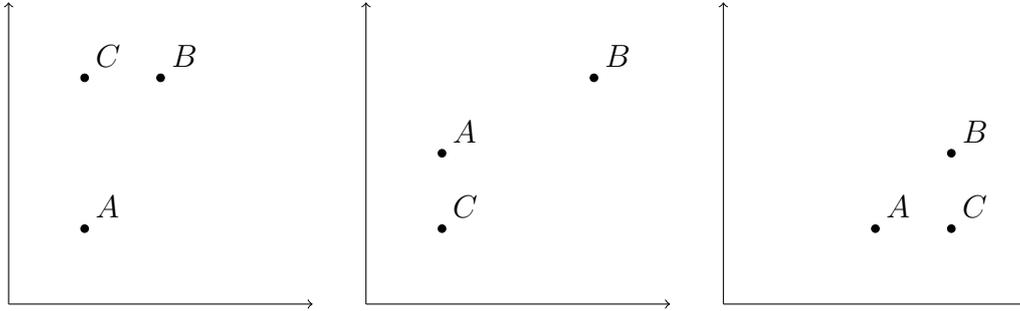


FIGURE 4.4 – Différentes vues d'un graphes à n dimensions représentant 3 points A , B et C

Maintenant, il est plus difficile, visuellement, de trouver les points les plus proches. Mais nous pouvons encore utiliser la formule de la distance euclidienne [Wikipédia, 2020g]

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

où n est le nombre de mot différent dans le document, $A = \{x_1, \dots, x_i, \dots, x_n\}$ et $B = \{y_1, \dots, y_i, \dots, y_n\}$

Appliquons la théorie à nos phrases

$$AB = \sqrt{((1 - 1)^2 + (1 - 1)^2 + (0 - 1)^2 + (1 - 0)^2 + (1 - 1)^2 + (0 - 0)^2 + (0 - 0)^2)}$$

$$BC = \sqrt{((1 - 1)^2 + (1 - 0)^2 + (1 - 0)^2 + (0 - 0)^2 + (1 - 1)^2 + (0 - 1)^2 + (0 - 1)^2)}$$

$$CA = \sqrt{((1 - 1)^2 + (0 - 1)^2 + (0 - 0)^2 + (0 - 1)^2 + (1 - 1)^2 + (1 - 0)^2 + (1 - 0)^2)}$$

on obtient

$$AB = 1,41$$

$$BC = 2$$

$$CA = 2$$

Nous pouvons en conclure que la phrase A est plus similaire à la phrase B qu'à la phrase C et que la phrase C a une similarité identique vis à vis de la phrase A que de la phrase B .

Utilisons maintenant la théorie de la distance cosinus face au problème. Reprenons le graphe de la figure 4.3, sur lequel on rajoute les vecteurs OA , OB et OC

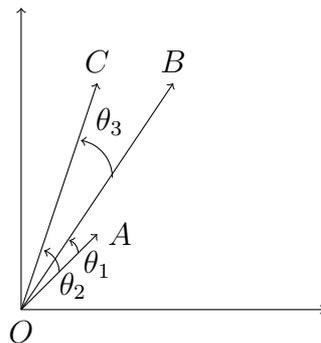


FIGURE 4.5 – Graphe représentant 3 vecteur OA , OB et OC sur un plan à n dimensions

Comparons les angles θ entre les vecteurs OA et OB , OB et OC et OA et OC .

La formule de la similarité cosinus [Wikipédia, 2020n, Negre, 2013] est

$$\cos \theta_1 = \frac{A.B}{\|A\| \|B\|}$$

où n est le nombre de mot différents dans les documents, $A = \{a_1, \dots, a_i, \dots, a_n\}$ et $B = \{b_1, \dots, b_i, \dots, b_n\}$

avec

$$A.B = \sum_{i=1}^n a_i * b_i$$

et

$$\|A\| = \sqrt{\sum_{i=1}^n a_i^2}$$

ce qui donne

$$\cos \theta = \frac{\sum_{i=1}^n a_i * b_i}{\sqrt{\sum_{i=1}^n a_i^2} * \sqrt{\sum_{i=1}^n b_i^2}}$$

Faisons le calcul pour le graphe de la figure A , B et C en 2 dimensions

$$\cos \theta_1 = \frac{1 * 2 + 1 * 3}{\sqrt{1^2 + 1^2} * \sqrt{2^2 + 3^2}} = \frac{5}{\sqrt{26}} \approx 0.98$$

$$\cos \theta_2 = \frac{1 * 1 + 1 * 3}{\sqrt{1^2 + 1^2} * \sqrt{1^2 + 3^2}} = \frac{4}{\sqrt{20}} \approx 0.89$$

$$\cos \theta_3 = \frac{1 * 2 + 3 * 3}{\sqrt{2^2 + 3^2} * \sqrt{1^2 + 3^2}} = \frac{11}{\sqrt{130}} \approx 0.96$$

Les résultats de la similarité cosinus nous disent que A est plus proche de B que B de C ce qui est différent des résultats de la distance euclidienne. Car avec la similarité cosinus, on analyse l'angle entre 2 vecteurs et plus la distance entre 2 points.

Appliquons la similarité cosinus à nos 3 phrases

$$\cos \theta_1 = \frac{(1 + 1 + 0 + 0 + 1 + 0 + 0)}{\sqrt{1 + 1 + 0 + 1 + 1 + 0 + 0} * \sqrt{1 + 1 + 1 + 0 + 1 + 0 + 0}} = \frac{3}{4} = 0.75$$

$$\cos \theta_2 = \frac{(1 + 0 + 0 + 0 + 1 + 0 + 0)}{\sqrt{1 + 1 + 0 + 1 + 1 + 0 + 0} * \sqrt{1 + 0 + 0 + 0 + 1 + 1 + 1}} = \frac{2}{4} = 0.50$$

$$\cos \theta_3 = \frac{(1 + 0 + 0 + 0 + 1 + 0 + 0)}{\sqrt{1 + 1 + 1 + 0 + 1 + 0 + 0} * \sqrt{1 + 0 + 0 + 0 + 1 + 1 + 1}} = \frac{2}{4} = 0.50$$

avec

- θ_1 est l'angle entre les vecteurs OA et OB
- θ_2 est l'angle entre les vecteurs OA et OC
- θ_3 est l'angle entre les vecteurs OB et OC

Donc c'est A et B qui sont les plus similaires.

Car les valeurs de $\cos \theta$ sont compris entre 0 et 1. Si la valeur de $\cos \theta$ vaut 0 alors les documents sont orthogonaux ou totalement différents. Et si la valeur de $\cos \theta$ vaut 1 alors les documents sont similaires. Donc, plus on se rapproche de 1, plus il y a similarité des documents.

Nous arrivons à la même conclusion pour les 2 méthodes de calcul. Mais peut-il y avoir des différences entre les résultats de la distance euclidienne et de la similarité cosinus ?

Prenons le cas extrême où une des phrases a les mêmes mots qu'une autre phrase mais avec une autre fréquence.

A : *Le ciel est gris, il fera beau demain*

B : *Le ciel est bleu, il fera gris demain*

C : *Le ciel est gris, il fera beau demain, le ciel est gris, il fera beau demain*

Ici la phrase C est une répétition de la phrase A . Résultat pour la distance euclidienne,

$$AB = 1,41$$

$$AC = 2$$

$$BC = 2,44$$

donc A est plus proche de B que de C .

Résultat pour la similarité cosinus,

$$\cos \theta_1 = 0,75$$

$$\cos \theta_2 = 1$$

$$\cos \theta_3 = 0,75$$

avec

- θ_1 est l'angle entre les vecteurs OA et OB
- θ_2 est l'angle entre les vecteurs OA et OC
- θ_3 est l'angle entre les vecteurs OB et OC

donc A est semblable à C , car le point C se trouve dans le prolongement du vecteur OA .

Sur le graphe de la figure 4.6 ci-dessous, on voit bien que la droite OA passe par les 2 points A et C . Mais on constate que C est plus proche de B que de A en calculant la distance euclidienne.

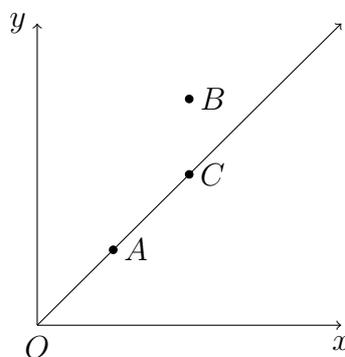


FIGURE 4.6 – Graphe représentant 3 points A , B et C sur un plan à 2 dimensions, avec C qui est sur le prolongement de la droite OA .

Nous pouvons donc conclure que les 2 méthodes ne donneront pas toujours le même résultat. La distance cosinus considérera que deux textes sont similaires si l'un des deux est la répétition de l'autre texte. Mais les textes seront considérés comme différents avec la méthode de la distance euclidienne.

Pour l'analyse des textes de jurisprudence lié au droit des marques, nous avons décidé de faire le choix de la distance cosinus pour effectuer les calculs des distances entre les textes. Car, sur base de ce que nous avons vu ci-dessus, elle nous paraît être intuitivement la mieux adaptée à l'analyse de texte.

Mais pour effectuer une prédiction sur les textes, il ne suffit pas de calculer la distance cosinus entre les textes, il faut aussi leur appliquer des algorithmes. Ces algorithmes vont tenter de prédire la réponse des textes soumis au test avant de les comparer à la réponse réel. Nous détaillerons à la sous-section suivante chaque algorithme utilisé dans les tests réalisés avec le programme Orange [Borondics et al., 2020].

4.1.2 Les modèles algorithmes d'apprentissage



Le premier modèle que nous voyons est le boosting adaptatif¹ ou adaboost de Yoav Freund et Robert Schapire [Wikipédia, 2020a]. Il s'agit d'un "*méta-algorithme probabiliste*" [Wikipédia, 2020a]. En résumé, l'algorithme applique différentes règles sur la base de donnée. Il récupère la règle ayant le taux d'erreur le plus faible et modifie le poids de chaque cas de la base de données en donnant un poids plus pour les cas qui étaient en erreur. Il recommence l'opération avec les nouveaux poids jusqu'à obtenir un algorithme fort. L'algorithme est une collection de règles simples, auquel il attribue une pondération. La prédiction se fait en combinant les règles en fonction de leur pondération respective [McCaffrey, 2016].



L'algorithme des k plus proches voisins² [Wikipédia, 2020j] ou kNN est un algorithme qui va déterminer la valeur d'un des points en sélectionnant la valeur ayant la majorité parmi les valeurs des points les plus proches. [Wikipédia, 2020j].



L'algorithme de régression logistique³ ou modèle de Logit [Wikipédia, 2019e, Gillet et al., 2011] est un algorithme qui construit une fonction logistique pour chaque caractéristique avec les exemples et qui fournit un coefficient de probabilité que la réponse tend vers une réponse plutôt qu'une autre. L'ensemble des coefficients fournissent la prédiction [Wikipédia, 2019e, Gillet et al., 2011].



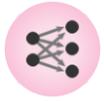
Le modèle bayésien naïf⁴ est un algorithme basé sur le théorème de Bayes (qui est fondé sur les probabilités conditionnelles [Benzaki, 2017]). En résumé, l'algorithme calcule la probabilité qu'un événement arrive sur plusieurs variables en prenant chaque variable indépendamment les uns des autres et en appliquant la probabilité

-
1. traduction de "Adaptive Boosting"
 2. traduction de "k-nearest neighbors algorithm"
 3. traduction de "Logistic Regression"
 4. traduction de "Naive Bayes"

conditionnelle pour chacun d'eux [Benzaki, 2017].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Ce qui signifie que la probabilité que l'événement B se réalise si l'événement A s'est réalisé est égale à la probabilité que l'événement A se réalise si l'événement B s'est réalisé multiplié par la probabilité l'événement A se réalise divisé par la probabilité que l'événement B se réalise.



Le réseau neuronale⁵ est une représentation simplifiée du fonctionnement des neurones de notre cerveau. L'algorithme va construire, au départ d'exemple, un réseau de neurone. Avec les données qu'il reçoit, il va construire les neurones suivants qui auront comme valeur l'addition des valeurs des neurones précédents multipliés par un poids auquel on rajoute une valeur de décalage de la fonction d'activation. On termine en multipliant le tout par une fonction d'activation

$$a^{(1)} = \sigma(Wa^{(0)} + b)$$

où a représente la valeur du neurone, σ est la fonction sigmoïde ou fonction d'activation et b est la valeur du biais.

Les valeurs de sortie sont alors comprise entre 0 et 1 ou 1 signifie que le neurone est activé. Le but de l'entraînement du réseau est l'activation d'un seul neurone de sortie. Pour cela nous devons corriger les poids ayant servi à calculer la valeur des neurones et remonter jusqu'aux valeurs d'entrée. Les valeurs de test ne devraient activer qu'un seul neurone de sortie donnant la solution au système [Bastien, 2019, Sanderson, 2017].

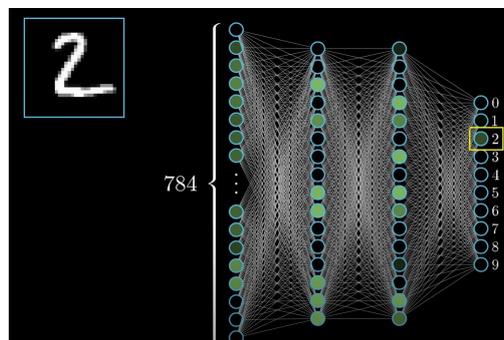


FIGURE 4.7 – Exemple de réseau neuronale pour la detection du chiffre 2 [Sanderson, 2017].



La méthode des forêts aléatoires⁶ ou les forêts d'arbres décisionnels fonctionnent comme un ensemble d'algorithmes d'arbre de décision (voir ci-dessous) où chaque arbre de décision correspond à une réponse potentielle du problème. Et le résultat final correspond au résultat de la majorité des arbre de décision [Wikipédia, 2020f].

5. traduction de "Neural network"

6. traduction de "Random Forest"



L'algorithme de descente stochastique du gradient ⁷ ou l'algorithme du gradient de descente stochastique [Wikipédia, 2020c]. La méthode du gradient de descente stochastique consiste à appliquer l'algorithme de gradient de descente sur une petite partie des données et de répéter l'opération sur les autres données tout en corrigeant la fonction à chaque itération sur un autre groupe de donnée. L'algorithme de gradient de descente recherche la valeur minimum des erreurs de prédictions en utilisant les fonction de perte [Sanderson, 2017].



L'algorithme des arbres de décision ⁸ [Wikipédia, 2020e]. Dans l'algorithme des arbres chaque nœud correspond à un choix auquel est confronté le programme et les feuilles correspondant à la réponse finale [Wikipédia, 2020e].

Après avoir vu les modèles algorithmiques qui seront utilisés dans les tests, nous passons maintenant à l'analyse des tests effectués dans le programme Orange [Borondics et al., 2020].

4.1.3 Les tests en détail

Les tests se dérouleront sur le logiciel Orange [Borondics et al., 2020], il utilise des modules qui, mis bout à bout, nous donnent un schéma représentant le chemin que parcourent les données des tests vers le module de prédiction. Ce même module nous fournira les résultats et le niveau d'efficacité de chaque modèle algorithmique de prédiction. Cependant ces modules de prédiction ont besoin d'un jeu de données d'entrainements et d'un jeu de données tests et comme la première partie des tests se réalisera avec un seul jeu de données, Orange [Borondics et al., 2020] va avec le module d'échantillonneur de données séparer de manière aléatoire le jeu de données en 2 jeux distincts.

Dans la suite de la section, nous allons détailler les jeux de données et le cheminement de l'information ainsi que les modules qui seront utilisés.

Les tests sont réalisés sur 2 jeux de données repris à la figure 4.8.

Liste des arrêts	Base de donnée				
	Conclusion	Concl Test	Acceptation	Accept Test	Accept Ex
C-108/97	Avis				
C-342/97	Avis				
C-299/99		Avis			
C-48/09	Rejet		Accepter		Accepter
C-344/10	Annulation		Accepter	Accepter	
C-90/11	Avis		Refuser		Refuser
C-98/11	Rejet		Refuser	Refuser	

FIGURE 4.8 – Exemple de la base de données repris à l'annexe G à la page 95. La colonne "Liste des arrêts" reprend les références de chaque arrêt qui se trouvent dans au moins un jeu de données. Les autres colonnes nous indiquent sur la présence ou non de la référence et de la valeur finale attendue dans ce jeu de données.

7. traduction de "stochastic gradient descent"

8. traduction du widget "tree"

Le tableau reprend, dans sa première colonne, la référence de l'arrêt comme indiqué à l'annexe F à la page 93 et les autres colonnes indiquent dans quelle base de données la référence est reprise pour effectuer les tests ainsi que la réponse que l'on attend du test pour cette référence.

Prenons un exemple, pour l'arrêt C-48/09, dans les tests sur la conclusion finale du pourvoi, on s'attend à avoir comme réponse "Rejet". Car le pourvoi de l'arrêt C-48/09 a été rejeté. Il est repris dans le jeu de données servant d'entraînement aux modèles de prédiction du programme Orange [Borondics et al., 2020]. Dans la colonne "Acceptation", il est considéré comme "Accepter" et il fait partie du jeu de données de test des modèles de prédiction.

Après création des bases de données, les valeurs sont importées dans le programme Orange [Borondics et al., 2020] via le module d'importation de document⁹ qui se trouve en début de schéma (exemple, la figure 4.9 à la page 36). Le résultat du test est récupéré dans les modules applications de test et résultat¹⁰ et dans la matrice de confusion¹¹.

Nous commençons par détailler les jeux de données qui se composent des textes de jurisprudence du domaine des marques fournies par les professeurs Benoit Michaux et Zorana Rosic et récupérées sur le site <http://curia.europa.eu>. Le détail de la liste des arrêts se trouve à l'annexe F à la page 93.

Pourquoi utiliser la jurisprudence ? Car elle détermine l'interprétation finale des articles qui ne sont pas écrits dans un langage formel. Un langage mathématique est un langage formel, c'est-à-dire qu'il est précis. Exemple :

$$\forall a \in \{0, 1, 2, 3\} \Rightarrow b = 9$$

ce qui signifie que si a est égal à 0, 1, 2 ou 3 alors b est égal à 9.

En droit, les textes sont sujets à interprétation avant qu'ils ne soient éclaircis par une jurisprudence qui donne l'orientation de la loi au moment de son écriture. Le langage formel est difficilement applicable, car il faudrait penser à toutes les options qui pourraient intervenir dans le futur ou attendre indéfiniment que tous les cas se présentent au législateur.

Le jeu de données qu'on appellera "Conclusion" concerne les tests que l'on effectue sur les textes pour obtenir un schéma prédisant le résultat final du pourvoi. Il se compose de 3 catégories.

- * "AVIS" comprenant 19 textes, lorsque la cour ne donne qu'un avis sur une question posée par les parties en cause.
- * "ANNULLATION" comprenant 17 textes, lorsque la cour annule la décision attaquée.
- * "REJET" comprenant 29 textes, lorsque le pourvoi est rejeté.

Ce jeu de données possède un jeu de test pour le schéma de prédiction qui sera utilisé avec le module de prédiction¹². Ce jeu complémentaire est composé de

9. traduction du widget "Import Documents"

10. traduction du widget "Test and Score"

11. traduction du widget "Confusion Matrix"

12. traduction de "predictions"

- * "AVIS" comprenant 1 texte.
- * "ANNULLATION" comprenant 5 textes.
- * "REJET" comprenant 5 textes.

Le but des tests sur le jeu de données "Conclusion" est de prédire la réponse que pourrait avoir un pourvoi d'une décision de justice. Les tests pourraient être utilisés par des chercheurs spécialisés dans le domaine de la justice afin de comparer la cohérence des conclusions entre les jugements. Ils pourraient aussi servir aux juges désirant vérifier la cohérence de leurs décisions avec les décisions qui ont été prises auparavant. Un cabinet d'avocats pourrait tenter de prédire la décision d'un pourvoi contre une décision en créant un jugement qui prendrait les faits et les conclusions que les avocats apporteraient lors des séances plénières futures et ainsi voir la décision qui se rapprocherait le plus du cas étudié.

Le jeu de données que nous appellerons "Acceptation" reprend les textes classés en fonction d'une interprétation personnelle sur l'acceptation en temps que marque du signe mis en cause. Il se compose de 2 catégories.

- * "ACCEPTER" comprenant 17 textes.
- * "REFUSER" comprenant 32 textes.

Ce jeu de données sera divisé en 2 jeux de données pour effectuer les tests avec le module de prédiction^{12 p34}. Le jeu de données qui servira d'entraînement est composé de

- * "ACCEPTER" comprenant 12 textes.
- * "REFUSER" comprenant 22 textes.

Et le jeu servant de test est composé de

- * "ACCEPTER" comprenant 5 textes.
- * "REFUSER" comprenant 10 textes.

Pour les tests sur le jeu de données "Acceptation", nous recherchons plutôt à prédire si un signe peut être enregistré en tant que marque.

Maintenant, nous allons analyser les schémas qui serviront aux tests des textes. Le schéma de la figure 4.9 représente le traitement des textes réalisés avec le programme Orange [Borondics et al., 2020] afin d'analyser les scores des différents modèles dans le cas d'une recherche de la réponse finale du pourvoi. Nous utiliserons le même schéma pour réaliser les tests sur l jeu de données "Acceptation" en vue d'essayer de prédire en fonction de texte de jurisprudence si un signe peut être enregistré comme une marque.

Le parcours de l'information commence par l'importation des données via le module d'importation de document^{9 p34}. Il récupère les fichiers texte classés dans les 3 répertoires du répertoire "Conclusion". Le module de pré-traitement du texte¹³ traite alors chaque texte pour enlever la ponctuation et les termes qui pourraient les parasiter. Nous lui donnons comme liste de mots parasite la liste repris à l'annexe D à la page 85. Les mots parasite sont les mots qui n'apportent rien à la recherche d'une réponse, ils ne représentent que du bruit. Le module de pré-traitement du texte sera utilisé avec différents paramètres durant les tests. Le premier paramètre sera les mots à prendre en compte pour le comptage, il aura le choix entre tous les mots des textes ou une liste de mots repris à l'annexe E à la page 86. Cette liste comprend les mots que nous estimons être important pour la prédiction du résultat désiré. Le second paramètre permet de

13. traduction du widget "Preprocess Text"

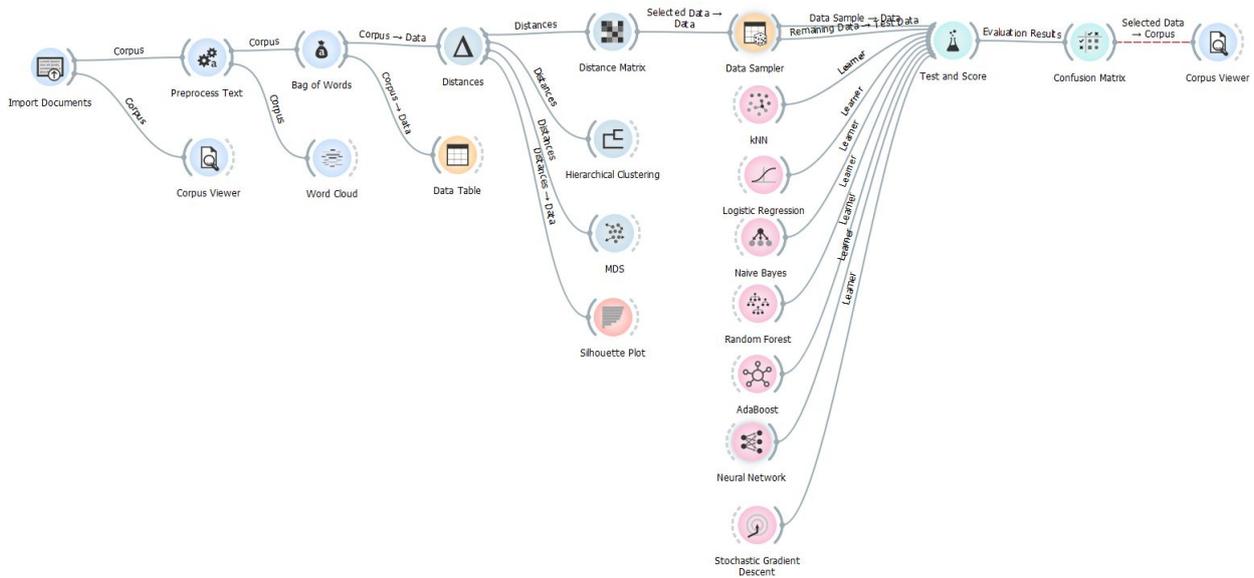


FIGURE 4.9 – Schéma représentant les widgets utilisés et le parcours de l'information.

déterminer les suites de mots à prendre dans les textes. Mais quel est la taille des suites mots que l'on peut prendre en compte ?

Durant l'analyse, nous avons constaté que plus la taille des suites à prendre en compte était grande plus la durée de traitement était grande (voir le tableau ci-dessous) et cela sans apporter d'amélioration aux résultats de performance dans les modèles algorithmiques

Mots	Suite de mots	Nombre de mots ou suite de mots	Durée de l'analyse
Mots du texte	1-1	8.835	+/- 1 minutes
	1-2	78.233	+/- 11 minutes
	1-3	186.037	+/- 45 minutes
	1-4	313.000	+/- 120 minutes
Liste de mots	1-1	90	+/- 9 secondes
	1-2	2.436	+/- 25 secondes
	1-3	10.617	+/- 2 minutes
	1-5	40.851	+/- 6 minutes

Un autre problème apparait pour l'analyse lors de l'utilisation de la liste des mots même si le nombre de mots ou couples de mots différents est moins élevé, le nombre de données parasite augmente très fortement, car l'application de pré-traitement du texte^{13 p35} associe les mots de la liste sans tenir compte des autres mots entre eux. Illustrons ces propos

Le jeu de hasard est interdit au moins de 18 ans

et nous la combinons avec la liste de mots ci-dessous

jeu, interdit, 18, ans

ce qui nous donnera après le pré-traitement du texte^{13 p35} avec l'option "1-2" pour le nombre de combinaison possible, la base de travail suivante :

module distance¹⁹. Le module "Silhouette Plot" affiche la cohérence dans les groupes de données, elle se calcule en effectuant la différence entre la distance moyenne avec les textes du même groupe et la distance moyenne avec les autres groupes [Wikipédia, 2020m].

La module de classification hiérarchique²⁰ affiche le regroupement qu'il a effectué des différents textes, chaque groupe rassemble les textes ayant la plus grande similarité en fonction de la distance cosinus qui a été calculée.

Le module d'échelle multidimensionnelle²¹ affiche en 2 dimensions la distance entre les points.

Et le module de distance matricielle²² compare les distances 2 à 2 entre chaque texte. Nous utiliserons le module "distance Matrix"²².

Après avoir fait le prétraitement du test, nous passons à la partie de prédiction en commençant par le module d'échantillonneur de données²³ qui va séparer les données en 2 groupes, le premier servant de groupe d'apprentissage (représentant 70% de la totalité des cas, valeur par défaut pour tout les tests) et le second de groupe de test (représentant 30% de la totalité des cas).

Le module de test et de résultat^{10 p34} va tester les données reçues pour le test avec chaque modèles algorithmiques d'apprentissage, que nous avons vu à la section 4.1.2 à la page 31 et va évaluer les résultats pour chaque modèle. Il affiche les valeurs pour l'aire sous la courbe, la précision de la classification, la valeur F-1, la valeur prédictive positive et la sensibilité de chaque modèle d'apprentissage. Dans la figure suivante, nous reprenons un exemple des valeurs obtenues.

Commençons par interpréter les valeurs fournies au moyen du tableau ci dessous.

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.964	0.867	0.863	0.893	0.867
Random Forest	0.768	0.667	0.667	0.673	0.667
Neural Network	0.804	0.733	0.707	0.822	0.733
Naive Bayes	0.821	0.733	0.733	0.733	0.733
SVM	0.875	0.800	0.798	0.804	0.800
AdaBoost	0.866	0.867	0.867	0.867	0.867
kNN	0.964	0.867	0.867	0.867	0.867
Tree	0.920	0.867	0.867	0.867	0.867

FIGURE 4.11 – Exemple des valeurs obtenus avec réalisation d'un test

Interprétation du résultat [Thomas G. Tape,]

20. traduction du widget "Hierarchical clustering"

21. traduction de "Multidimensional Scaling" qui est l'abréviation du widget "MDS"

22. traduction du widget "Distance Matrix"

23. traduction du widget "Data Sampler"

Aire sous ROC	Interprétation
0,90 à 1	excellent
0,80 à 0,90	bon
0,70 à 0,80	passable
0,60 à 0,70	médiocre
0,50 à 0,60	échec

Nous regardons maintenant la signification de chaque valeur [University of Ljubljana, 2020] L'aire sous la courbe ROC (AUC) ou *la fonction d'efficacité du récepteur* [Wikipédia, 2019a], permet de mesurer les performances d'un modèle prédictif.

La précision de la classification (CA) est *la proportion d'exemples correctement classés*²⁴ [University of Ljubljana, 2020].

La valeur F-1 (F1) est *est une moyenne harmonique pondérée de précision et de rappel*²⁵ [University of Ljubljana, 2020, Wikipédia, 2019b].

La précision ou *valeur prédictive positive* [Wikipédia, 2020k] est la proportion des vrais positifs sur l'ensemble des valeurs étiquetées positives (vrais positifs + faux positifs) [University of Ljubljana, 2020, Wikipédia, 2020k].

La *sensibilité* [Wikipédia, 2020k] (Recall) est la proportion des vrais positifs sur l'ensemble des valeurs positives ou qui auront dû être étiquetées positives (vrais positifs + faux négatifs) [University of Ljubljana, 2020, Wikipédia, 2020k].

Enfin, nous terminons par le module de matrice de confusion^{11 p34} qui affiche, pour le modèle d'algorithme sélectionné, le résultat des données test en affichant un tableau représentant le pourcentage pour chaque réponse possible, le taux de vrai ou de faux positif ou négatif.

Nous passons à l'analyse du parcours de l'information pour le schéma de la figure 4.12. Nous nous arrêterons que sur les modifications par rapport au schéma de la figure 4.9 à la page 36 car les modules utilisés sont identiques ainsi que leur fonctionnement.

Le cheminement de l'information s'effectue maintenant via 2 chemins distincts jusqu'au regroupement dans le module de prédiction^{12 p34}. Le premier chemin prendra les documents servant de test au module de prédiction^{12 p34}. Tandis que le second chemin s'occupera des documents servant d'apprentissage pour les différents modèles algorithmiques.

Le module de prédiction^{12 p34} utilise les algorithmes qu'il a reçu des différents modèles d'apprentissage et les applique aux données du test. Il affiche le résultat pour chaque algorithme comme pour le module de test et de résultat^{10 p34} mais il affiche aussi pour chaque données du test les résultats que chaque algorithme lui a attribué.

Dans les analyses précédentes, on essayait de déterminer si une marque pourrait être acceptée ou non. Pour ce faire, on fournissait aux différents modèles algorithmiques, une série de jeux de données. Les données étant composées de textes qui représentaient les cas positifs ou négatifs. Ces mêmes textes étaient alors transformés en données numériques via le module de sac de mots^{15 p37} et, sur lesquelles, on calculait la distance cosinus.

Mais si, au lieu de partir d'un jeu de données composé de textes, nous fournissions directement des valeurs numériques au programme Orange [Borondics et al., 2020].

24. traduction de "the proportion of correctly classified examples"

25. traduction de "a weighted harmonic mean of precision and recall"

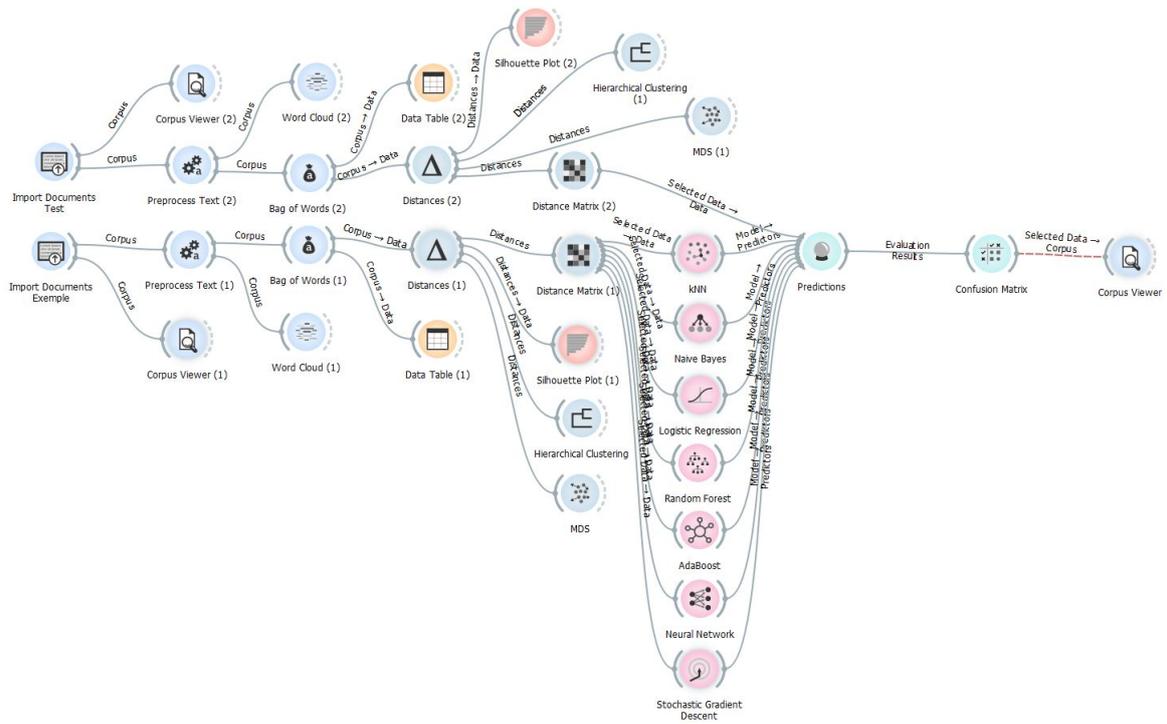


FIGURE 4.12 – Schéma représentant les widgets utilisés et le parcours de l'information en utilisant la prédiction.

Sur base de l'analyse des articles 7 (annexe B à la page 80) et 4(annexe A à la page 78), nous avons établi les cas généraux qui établissent qu'un signe peut être une marque. Et grâce à l'analyse des arrêts de jurisprudence(voir annexe F à la page93), nous avons obtenu les cas spéciaux. Tous ces cas ont été repris dans le tableau à l'annexe H à la page 96. Dans ce tableau, chaque cas répond a des affirmations par oui (1) ou par non (-1) ou par nul (0), nul signifiant que la réponse n'est pas importante dans la détermination du résultat. Ce qui permet d'obtenir, pour chaque cas, une liste de de valeur comprise entre -1 et 1. Le schéma de la figure 4.13 ci-dessous représente le parcours des données.

À nouveau, nous ne détaillerons que la modification apportée par rapport aux schémas précédents. Ce schéma est de nouveau composé de 2 chemins. Mais la différence tient dans l'importation des données qui est réalisée par 2 modules appelés fichier²⁶. Le premier module ne prenant que les données numériques des cas de jurisprudence et le second module les valeurs des cas généraux. Les valeurs des cas de jurisprudence serviront en partie (30%) pour effectuer les tests des algorithmes de prédiction. Les cas qui n'auront pas été sélectionnés pour les tests seront regroupés avec les cas généraux via le module de concaténation²⁷.

Le reste du chemin correspond au chemin parcouru dans le schéma repris à la figure 4.12 à la page 40.

Ayant vu en détail les tests réalisés avec le programme Orange [Borondics et al., 2020], nous pouvons passer à l'analyse des résultats pour chaque test qui a été réalisé.

26. traduction du widget "file"

27. traduction du widget "concatenate"

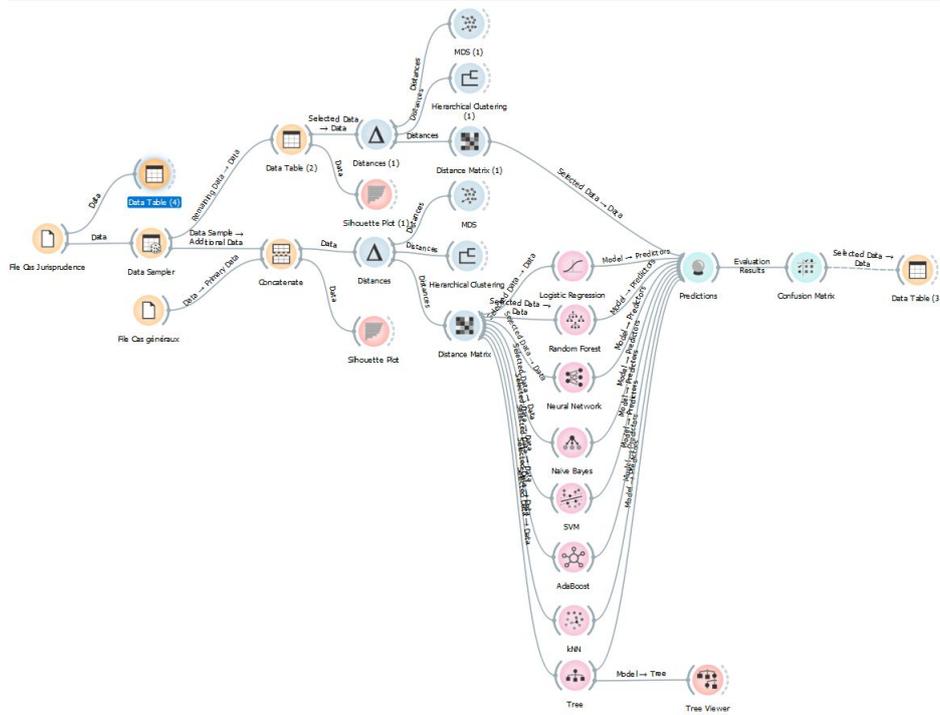


FIGURE 4.13 – Schéma représentant les modules utilisés et le parcours de l'information en utilisant la prédiction avec l'utilisation d'un jeu de données numérique.

4.2 Analyse des résultats

4.2.1 Analyse des résultats sur le jeu de données "Conclusion" en utilisant le module de test et résultat^{10 p34}

Pour rappel, nous cherchons à déterminer s'il serait possible de prédire la décision finale d'un arrêt de jurisprudence dans le domaine du droit des marques.

Dans un premier temps, nous analyserons les graphiques obtenus avec les modules de classification hiérarchique^{20 p38}, de représentation graphique de l'échelle multidimensionnelle^{21 p38} et de "silhouette plot" dans le schéma repris à la figure 4.9 page 36.

Les graphiques de la cohérence dans les groupes de données (figure 4.14) nous indiquent que le groupe "AVIS" est cohérent, car les valeurs positives indiquent que le texte est plus proche des voisins de leur groupe que des voisins des autres groupes. Et inversement, les textes ayant une valeur négative signifient qu'ils sont classés dans un mauvais groupe. Les graphiques sont semblables que ce soit en mode binaire (le graphe de droite) ou en mode comptage (le graphe de gauche).

L'analyse des graphes de classification hiérarchique^{20 p38} à la figure 4.15 nous montrent bien un groupe ne comprenant que des cas classés en "AVIS".

En analysant les représentations graphiques de l'échelle multidimensionnelle^{21 p38} à la figure 4.16, nous constatons très vite que les points verts représentent un groupe homogène.

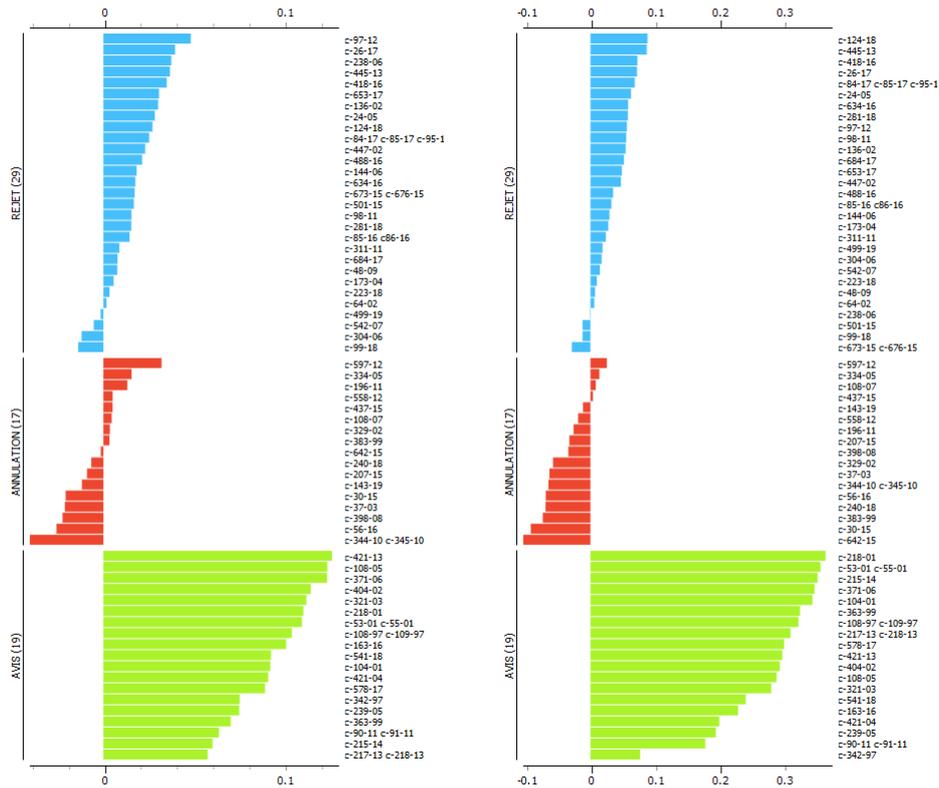


FIGURE 4.14 – Représentation du graphique de la cohérence dans les groupes de données dans le schéma de la figure 4.9 à la page 36 en mode binaire et comptage

Nous pouvons constater que les textes du groupe "AVIS" ont une plus grande similarité que les 2 groupes pour lesquels il est plus difficile de déterminer une similarité dans un même groupe. Nous pouvons déjà dire que les modèles algorithmiques devraient nous fournir de bon résultat de prédiction et qu'il ne devrait pas y avoir d'erreur dans la prédiction des textes catalogués dans "AVIS".

Nous passons à l'analyse des résultats fournis par le module de test et des résultats^{10 p34} repris à l'annexe I à la page 99 et fournis par les statistiques de performance des modèles algorithmiques.

En regardant les résultats obtenus en mode binaire, nous constatons que les modèles algorithmiques ont tous une bonne performance au niveau de l'aire sous la courbe ROC et même une excellente performance pour le modèle des k plus proches voisins^{2 p31}. Le modèle des forêts aléatoires^{6 p32} n'obtient qu'un score de passable.

En mode comptage, nous remarquons que les modèles algorithmiques sont légèrement moins performants sauf pour le modèle des forêts aléatoires^{6 p32} qui passe d'un score de passable à excellent. Les 2 moins bons résultats sont ceux des modèles de boosting adaptatif^{1 p31} et de descente stochastique du gradient^{7 p33}.

Comment pourrions nous expliquer la différence des résultats entre les 2 tests ? Nous savons que le module d'échantillonneur de données^{23 p38} effectue la séparation des jeux de données

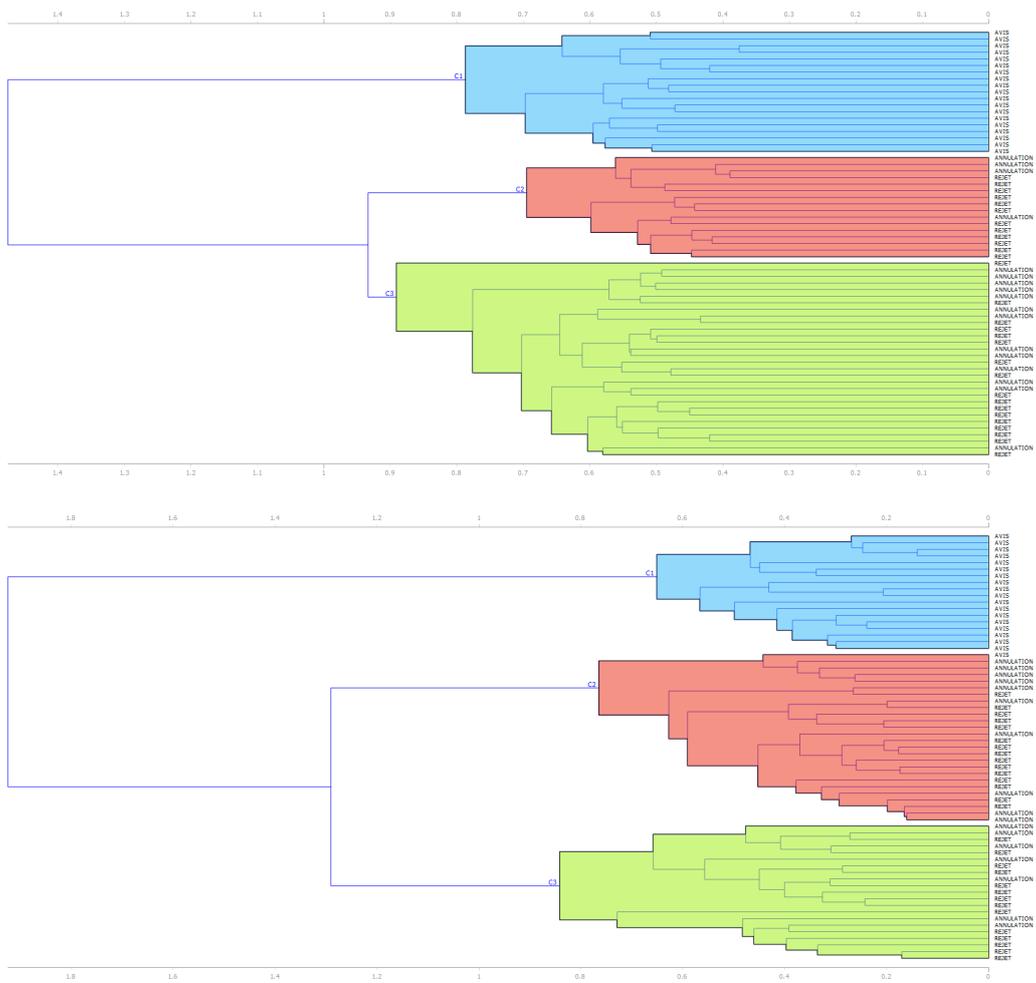


FIGURE 4.15 – Représentation graphique de classification hiérarchique²⁰ p^{38} fourni par le schéma de la figure 4.9 à la page 36 en mode binaire et comptage.

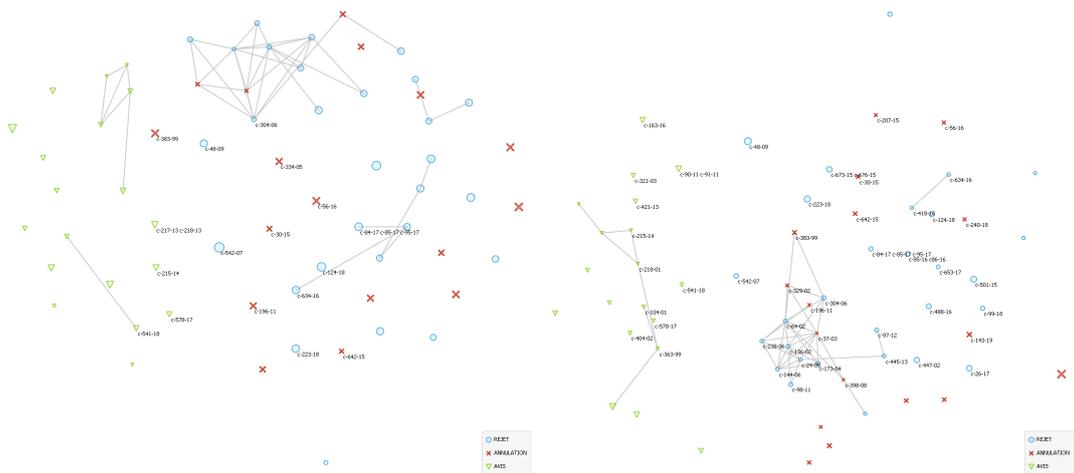


FIGURE 4.16 – Représentation graphique de l'échelle multidimensionnelle²¹ p^{38} fourni par le schéma de la figure 4.9 à la page 36 en mode binaire et comptage.

d'entraînement et de test de manière aléatoire. Ce qui signifie que les modèles algorithmiques ont, peut-être, été entraînés sur des exemples différents. Cela pourrait expliquer la différence dans les résultats des tests.

Nous pourrions remédier au problème en séparant les données de test et d'entraînement manuellement en créant 2 répertoires différents comme nous l'avons fait pour les tests avec le module de prédiction^{12 p34}.

Nous pouvons constater que les textes classés dans le groupe "AVIS" sont tous correctement prédits sauf avec les modèles de boosting adaptatif^{1 p31} et des forêts aléatoires^{6 p32}. Les textes en erreur sont C-90/11 et C-163/16 cela peut être dû à des textes plus petit. Dans les autres textes mal classés, nous allons analyser les textes qui sont fréquemment en erreur sur l'ensemble des tests sur le jeu de données "Conclusion".

Le texte C-499/19 est en règle générale mal prédit, cela s'explique peut-être par le fait que ce texte est très petit. Il y a, dès lors, moins de mots pour appliquer la prédiction.

L'analyse des textes C-173/04, C-143/19 et C-56/16, ne permet pas de déterminer la raison de leurs mauvais classements dans la prédiction, en analysant les graphiques de cohérence dans les groupes de données à la figure 4.14 nous constatons déjà que les textes ont une mauvaise cohérence dans leurs groupes respectifs.

4.2.2 Analyse des résultats sur le jeu de données "Acceptation" en utilisant le module de test et résultat^{10 p34}

Nous recherchons, maintenant, à prédire si un signe peut être enregistré en tant que marque seulement en analysant un arrêt de jurisprudence du droit des marques.

Nous analyserons d'abord les graphiques obtenus avec les modules de classification hiérarchique^{20 p38}, de représentation graphique de l'échelle multidimensionnelle^{21 p38} et de "silhouette plot" dans le schéma repris à la figure 4.9 page 36.

Les distances des 3 premiers graphes de l'analyse de la figure 4.17 ont été calculées avec des coordonnées en mode binaire et avec l'ensemble des mots sur une combinaison de mots limitée à 1-1.

Les 3 graphes suivants, repris à la figure 4.18 utilisent des distances calculées avec des coordonnées en mode comptage et avec l'ensemble des mots des textes repris dans le jeu de données "Acceptation" limité à une combinaison de mots 1-2.

Sur les graphiques de classification hiérarchique^{20 p38}, nous constatons directement que le programme Orange [Borondics et al., 2020] ne parvient pas à trouver un groupe représentant le même résultat. Nous voyons également sur les représentations graphiques de l'échelle multidimensionnelle^{21 p38} que aucun groupe de points n'est homogène. L'analyse des graphes de "silhouette plot" nous indique une petite cohérence entre certains textes catalogués comme "RE-FUSER" mais les textes repris dans le groupe "ACCEPTER" ont tendance à être plus proche de l'autre groupe.

Nous pouvons déjà estimer qu'au vu de l'analyse des graphes, les modèles algorithmiques

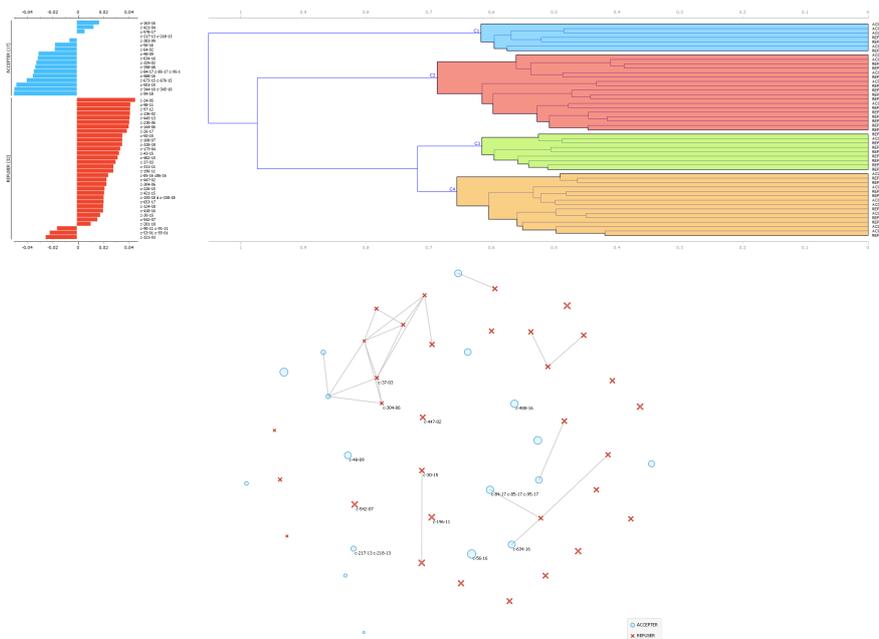


FIGURE 4.17 – Représentation des graphes en mode binaire pour le schéma de la figure 4.9 à la page 36 avec une combinaison de mots limitée à 1-1

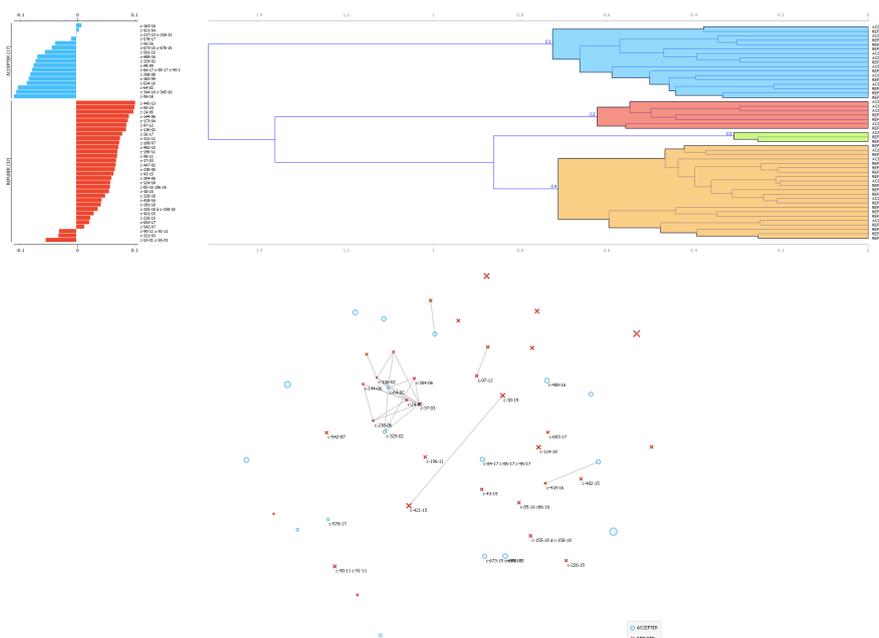


FIGURE 4.18 – Représentation des graphes en mode comptage pour le schéma de la figure 4.9 à la page 36 avec une combinaison de mots limitée à 1-2

auront des difficultés à réaliser un algorithme offrant des résultats satisfaisants.

L'analyse des résultats des algorithmes de prédiction nous indique que le test ne fonctionne pas, seul le modèle des k plus proches voisins² p^{31} obtient un score de passable. L'analyse des textes mal classés est superflue, car il est très difficile de déterminer, avec les mauvais scores

des algorithmes, pourquoi ils ont été mal classés. La suite de l'analyse sur le jeu de données "Acceptation" sera réalisée avec une sélection de mots prédéfinis à l'avance. Nous espérons améliorer les résultats des algorithmes en modifiant les mots utilisés pour effectuer l'analyse.

Les 2 dernières séries de graphes utilisent non plus l'ensemble des mots des textes mais se limite à une liste de mots (repris à l'annexe E à la page 86. Cette liste de mots est composée de tous les mots qui caractérisent le droit des marques. A nouveau, l'analyse se portera sur une combinaison des mots limitée à 1-1 et à 1-2. En analysant les graphes repris à la figure 4.19

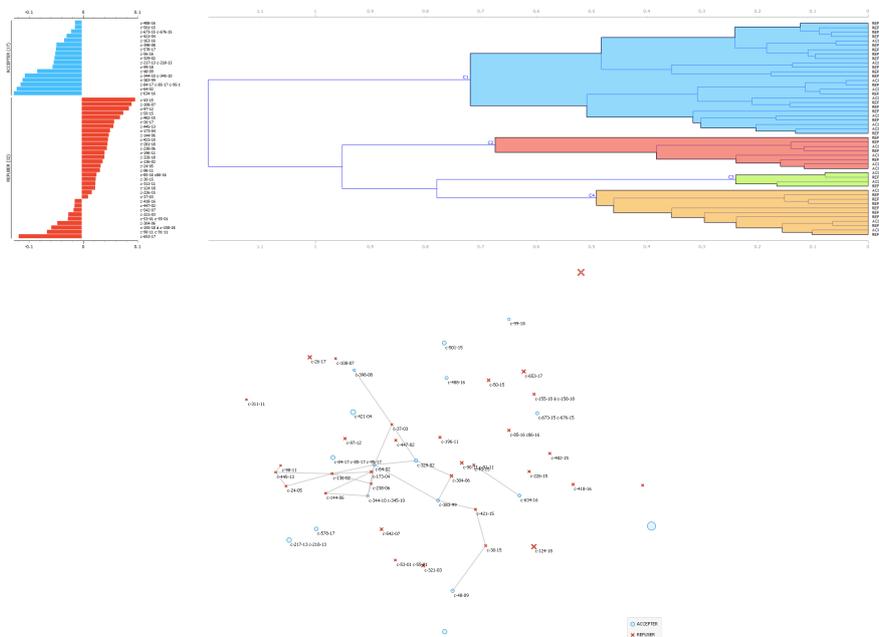


FIGURE 4.19 – Représentation des graphes en mode comptage pour le schéma de la figure 4.9 à la page 36 avec une combinaison de mots limitée à 1-1 et la liste des mots repris à l'annexe E à la page 86

et à la figure 4.20, on constate que l'application à beaucoup de difficultés à faire la distinction dans les groupes. Il n'y a pas d'amélioration sur la cohérence dans les groupes en utilisant la liste de mots.

Les résultats des modèles algorithmiques ne sont pas améliorés. Il probable qu'il ne sera pas possible de prédire l'acceptation de l'enregistrement d'une marque sur base des textes brutes.

La suite des tests s'effectueront en utilisant le schéma repris à la figure 4.12 à la page 40 avec le module de prédiction^{12 p34}.

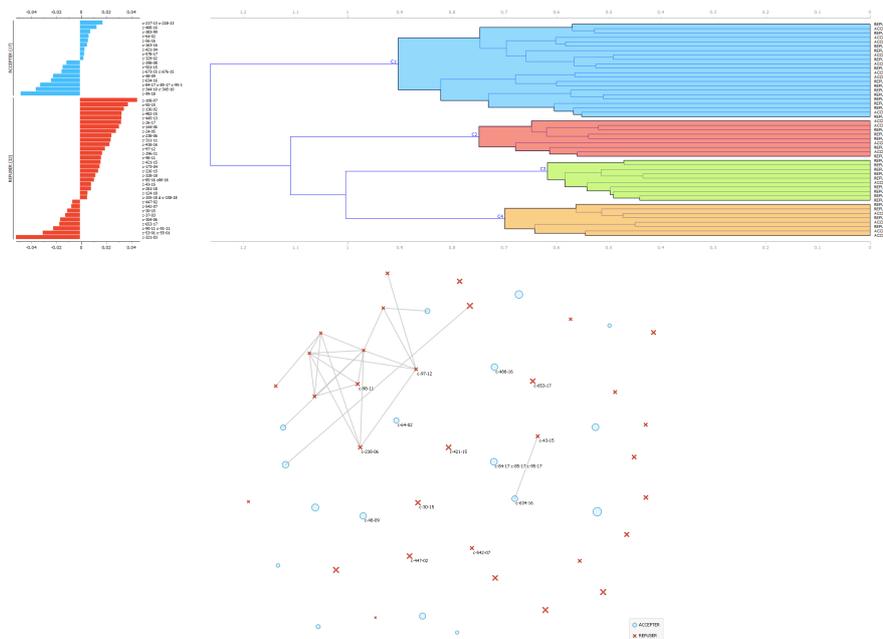


FIGURE 4.20 – Représentation des graphes en mode binaire pour le schéma de la figure 4.9 à la page 36 avec une combinaison de mots limitée à 1-2 et la liste des mots repris à l'annexe E à la page 86

4.2.3 Analyse des résultats sur le jeu de données "Conclusion" en utilisant le module de prédiction^{12 p34}

Les résultats obtenus avec le module de test et résultat^{10 p34} sur le jeu de données "Conclusion", étaient bon, les algorithmes fournis permettaient d'obtenir de très bonne prédiction sur la conclusion des textes étudiés.

Nous regardons à présent si les algorithmes obtenus avec le module de prédiction^{12 p34} nous fournissent aussi des bons résultats.

Dans ces tests, le jeu d'entraînements et le jeu d'exemples resteront identiques pour tout le test. Ce qui nous permettra une meilleur comparaison entre les différents paramètres utilisés.

Les graphiques obtenus avec les modules de classification hiérarchique^{20 p38}, de représentation graphique de l'échelle multidimensionnelle^{21 p38} et de "silhouette plot" dans le schéma repris à la figure 4.12 page 40 sont donnés à titre informatif à l'annexe J page 112 car ceux sont semblable aux graphiques obtenus aux figures 4.14, 4.15 et 4.16.

Dès lors, nous pouvons passer directement à l'analyse des résultats repris sur la figure 4.21.

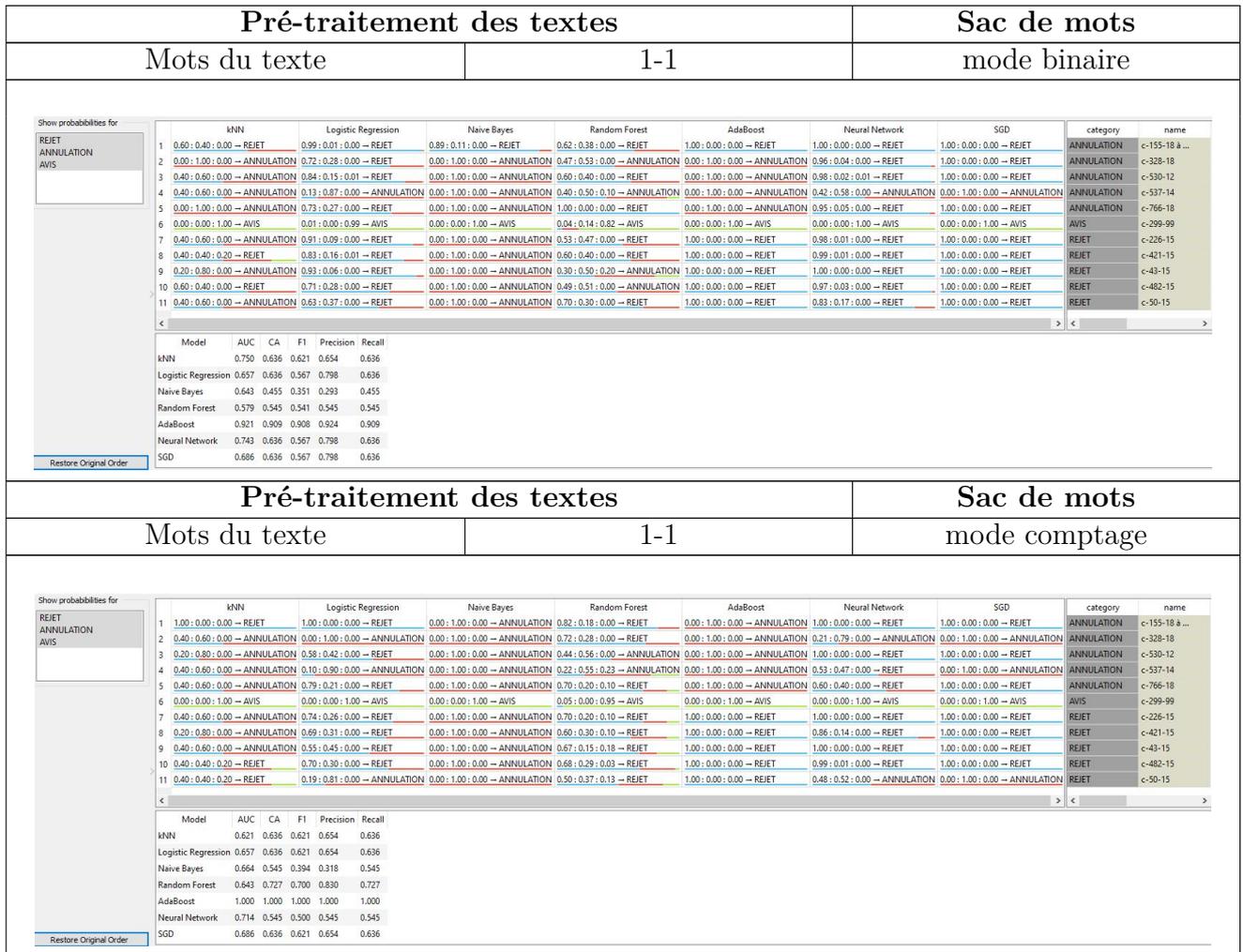


FIGURE 4.21 – Résultats obtenus sur le schéma de la figure 4.12 à la page 40

La principale remarque que nous nous faisons directement est que les performances des algorithmiques sont en général moins bonnes mais on constate aussi que le modèle de boosting adaptatif¹ p^{31} obtient d'excellent score dans les 2 tests. Le texte repris dans la catégorie "AVIS" est d'office prédit correctement, comme pour les tests avec le module de test et résultat¹⁰ p^{34} . Et le modèle de bayésien naïf⁴ p^{31} met toutes les valeurs à "ANNULATION".

Nous pouvons passer à l'analyse des textes C-155/18,C-766/18 et C-50/15 qui sont souvent mal classé par les algorithmes. Après relecture des textes, ceux-ci ne sont pas mal classés, ce qui amène à la même conclusion que dans la section précédente qu'il est difficile en analysant les textes de déterminer pourquoi ils sont mal classés.

Il est probable qu'en modifiant le jeu d'entraînement, il soit possible d'améliorer certains algorithmes. Il n'est pas étonnant d'obtenir de bons résultats que ce soit avec le module de test et résultat¹⁰ p^{34} ou de prédiction¹² p^{34} car chaque règle qui compose l'algorithme est la règle ayant obtenu la valeur d'erreur la plus faible du niveau.

Dans la sous-section suivante, Nous analyserons les résultats obtenus sur le jeu de données

"Acceptation" en utilisant le module de prédiction^{12 p34}.

4.2.4 Analyse des résultats sur le jeu de données "Acceptation" en utilisant le module de prédiction^{12 p34}

Dans cette analyse, nous cherchons à déterminer s'il est possible de prédire, avec le module de prédiction^{12 p34}, l'acceptation ou non d'un signe à l'enregistrement en tant que marque. Les résultats que nous avons déjà analysés précédemment avec le module de test et résultat^{10 p34} n'était pas concluant.

Tout comme la sous-section précédente, les graphiques obtenus avec les modules de classification hiérarchique^{20 p38}, de représentation graphique de l'échelle multidimensionnelle^{21 p38} et de "silhouette plot" dans le schéma repris à la figure 4.12 page 40 sont donnés à titre informatif à l'annexe K page 114 car ceux-ci sont semblables aux graphiques obtenus aux figures précédentes.

Nous commencerons directement par analyser les résultats du module de prédiction^{12 p34}.

Pré-traitement des textes							Sac de mots																																																																																																																																																																		
Mots des textes				1-1			mode comptage																																																																																																																																																																		
<div style="display: flex; justify-content: space-between;"> <div style="width: 15%;"> <p>Show probabilities for</p> <p>ACCEPTER</p> <p>REFUSER</p> </div> <table border="1" style="width: 80%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>kNN</th> <th>Logistic Regression</th> <th>Naive Bayes</th> <th>Random Forest</th> <th>AdaBoost</th> <th>Neural Network</th> <th>SGD</th> <th>category</th> <th>name</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.40 : 0.60 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-344-10 ...</td></tr> <tr><td>2</td><td>0.40 : 0.60 → REFUSER</td><td>0.97 : 0.03 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.02 : 0.98 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>ACCEPTER</td><td>c-398-08</td></tr> <tr><td>3</td><td>0.40 : 0.60 → REFUSER</td><td>0.60 : 0.40 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.33 : 0.68 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.01 : 0.99 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-501-15</td></tr> <tr><td>4</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.99 : 0.01 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.12 : 0.88 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>ACCEPTER</td><td>c-56-16</td></tr> <tr><td>5</td><td>0.40 : 0.60 → REFUSER</td><td>0.82 : 0.18 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.30 : 0.70 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.66 : 0.34 → ACCEPTER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-578-17</td></tr> <tr><td>6</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.00 : 1.00 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.30 : 0.70 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-124-18</td></tr> <tr><td>7</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.91 : 0.09 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.53 : 0.47 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.98 : 0.02 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>REFUSER</td><td>c-226-15</td></tr> <tr><td>8</td><td>0.20 : 0.80 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-311-11</td></tr> <tr><td>9</td><td>0.40 : 0.60 → REFUSER</td><td>0.30 : 0.70 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.40 : 0.60 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.89 : 0.11 → ACCEPTER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-421-15</td></tr> <tr><td>10</td><td>0.40 : 0.60 → REFUSER</td><td>0.05 : 0.95 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.40 : 0.60 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.17 : 0.83 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-482-15</td></tr> <tr><td>11</td><td>0.40 : 0.60 → REFUSER</td><td>0.03 : 0.97 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.38 : 0.62 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.08 : 0.92 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-50-15</td></tr> <tr><td>12</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.92 : 0.08 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.50 : 0.50 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.61 : 0.39 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>REFUSER</td><td>c-542-07</td></tr> <tr><td>13</td><td>0.80 : 0.20 → ACCEPTER</td><td>0.93 : 0.07 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.33 : 0.68 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.16 : 0.84 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>REFUSER</td><td>c-653-17</td></tr> <tr><td>14</td><td>0.40 : 0.60 → REFUSER</td><td>0.02 : 0.98 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.05 : 0.95 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.09 : 0.91 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-85-16 c86-16</td></tr> <tr><td>15</td><td>0.40 : 0.60 → REFUSER</td><td>0.27 : 0.73 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.30 : 0.70 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-98-11</td></tr> </tbody> </table> </div> <div style="width: 15%; text-align: right;"> <p>category</p> <p>name</p> </div>											kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	1	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...	2	0.40 : 0.60 → REFUSER	0.97 : 0.03 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	1.00 : 0.00 → ACCEPTER	ACCEPTER	c-398-08	3	0.40 : 0.60 → REFUSER	0.60 : 0.40 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.33 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15	4	0.60 : 0.40 → ACCEPTER	0.99 : 0.01 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.12 : 0.88 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTER	1.00 : 0.00 → ACCEPTER	ACCEPTER	c-56-16	5	0.40 : 0.60 → REFUSER	0.82 : 0.18 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.66 : 0.34 → ACCEPTER	0.00 : 1.00 → REFUSER	ACCEPTER	c-578-17	6	0.60 : 0.40 → ACCEPTER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18	7	0.60 : 0.40 → ACCEPTER	0.91 : 0.09 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.53 : 0.47 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.98 : 0.02 → ACCEPTER	1.00 : 0.00 → ACCEPTER	REFUSER	c-226-15	8	0.20 : 0.80 → REFUSER	0.01 : 0.99 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11	9	0.40 : 0.60 → REFUSER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTER	0.89 : 0.11 → ACCEPTER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15	10	0.40 : 0.60 → REFUSER	0.05 : 0.95 → REFUSER	1.00 : 0.00 → ACCEPTER	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTER	0.17 : 0.83 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15	11	0.40 : 0.60 → REFUSER	0.03 : 0.97 → REFUSER	1.00 : 0.00 → ACCEPTER	0.38 : 0.62 → REFUSER	1.00 : 0.00 → ACCEPTER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15	12	0.60 : 0.40 → ACCEPTER	0.92 : 0.08 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.50 : 0.50 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.61 : 0.39 → ACCEPTER	1.00 : 0.00 → ACCEPTER	REFUSER	c-542-07	13	0.80 : 0.20 → ACCEPTER	0.93 : 0.07 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.33 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTER	0.16 : 0.84 → REFUSER	1.00 : 0.00 → ACCEPTER	REFUSER	c-653-17	14	0.40 : 0.60 → REFUSER	0.02 : 0.98 → REFUSER	1.00 : 0.00 → ACCEPTER	0.05 : 0.95 → REFUSER	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16	15	0.40 : 0.60 → REFUSER	0.27 : 0.73 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11
	kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name																																																																																																																																																																
1	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...																																																																																																																																																																
2	0.40 : 0.60 → REFUSER	0.97 : 0.03 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	1.00 : 0.00 → ACCEPTER	ACCEPTER	c-398-08																																																																																																																																																																
3	0.40 : 0.60 → REFUSER	0.60 : 0.40 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.33 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15																																																																																																																																																																
4	0.60 : 0.40 → ACCEPTER	0.99 : 0.01 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.12 : 0.88 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTER	1.00 : 0.00 → ACCEPTER	ACCEPTER	c-56-16																																																																																																																																																																
5	0.40 : 0.60 → REFUSER	0.82 : 0.18 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.66 : 0.34 → ACCEPTER	0.00 : 1.00 → REFUSER	ACCEPTER	c-578-17																																																																																																																																																																
6	0.60 : 0.40 → ACCEPTER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18																																																																																																																																																																
7	0.60 : 0.40 → ACCEPTER	0.91 : 0.09 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.53 : 0.47 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.98 : 0.02 → ACCEPTER	1.00 : 0.00 → ACCEPTER	REFUSER	c-226-15																																																																																																																																																																
8	0.20 : 0.80 → REFUSER	0.01 : 0.99 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11																																																																																																																																																																
9	0.40 : 0.60 → REFUSER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTER	0.89 : 0.11 → ACCEPTER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15																																																																																																																																																																
10	0.40 : 0.60 → REFUSER	0.05 : 0.95 → REFUSER	1.00 : 0.00 → ACCEPTER	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTER	0.17 : 0.83 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15																																																																																																																																																																
11	0.40 : 0.60 → REFUSER	0.03 : 0.97 → REFUSER	1.00 : 0.00 → ACCEPTER	0.38 : 0.62 → REFUSER	1.00 : 0.00 → ACCEPTER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15																																																																																																																																																																
12	0.60 : 0.40 → ACCEPTER	0.92 : 0.08 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.50 : 0.50 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.61 : 0.39 → ACCEPTER	1.00 : 0.00 → ACCEPTER	REFUSER	c-542-07																																																																																																																																																																
13	0.80 : 0.20 → ACCEPTER	0.93 : 0.07 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.33 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTER	0.16 : 0.84 → REFUSER	1.00 : 0.00 → ACCEPTER	REFUSER	c-653-17																																																																																																																																																																
14	0.40 : 0.60 → REFUSER	0.02 : 0.98 → REFUSER	1.00 : 0.00 → ACCEPTER	0.05 : 0.95 → REFUSER	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16																																																																																																																																																																
15	0.40 : 0.60 → REFUSER	0.27 : 0.73 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11																																																																																																																																																																

Model	AUC	CA	F1	Precision	Recall
kNN	0.430	0.467	0.467	0.467	0.467
Logistic Regression	0.680	0.733	0.741	0.774	0.733
Naive Bayes	0.560	0.333	0.167	0.111	0.333
Random Forest	0.230	0.533	0.464	0.410	0.533
AdaBoost	0.350	0.333	0.345	0.407	0.333
Neural Network	0.460	0.600	0.600	0.600	0.600
SGD	0.550	0.600	0.600	0.600	0.600

Restore Original Order

Pré-traitement des textes	Sac de mots
Mots des textes	mode binaire

Show probabilities for <input type="checkbox"/> ACCEPTER <input checked="" type="checkbox"/> REFUSER	<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>kNN</th> <th>Logistic Regression</th> <th>Naive Bayes</th> <th>Random Forest</th> <th>AdaBoost</th> <th>Neural Network</th> <th>SGD</th> <th>category</th> <th>name</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.20 : 0.80 → REFUSER</td><td>0.03 : 0.97 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.37 : 0.63 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-344-10 ...</td></tr> <tr><td>2</td><td>0.40 : 0.60 → REFUSER</td><td>0.19 : 0.81 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.53 : 0.47 → ACCEPTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-398-08</td></tr> <tr><td>3</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.10 : 0.90 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.31 : 0.69 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-501-15</td></tr> <tr><td>4</td><td>0.20 : 0.80 → REFUSER</td><td>0.36 : 0.64 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.30 : 0.70 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-56-16</td></tr> <tr><td>5</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.48 : 0.52 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.36 : 0.64 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.17 : 0.83 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-578-17</td></tr> <tr><td>6</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.07 : 0.93 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.37 : 0.63 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-124-18</td></tr> <tr><td>7</td><td>0.40 : 0.60 → REFUSER</td><td>0.25 : 0.75 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.28 : 0.72 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.02 : 0.98 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-226-15</td></tr> <tr><td>8</td><td>0.20 : 0.80 → REFUSER</td><td>0.07 : 0.93 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.28 : 0.72 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-311-11</td></tr> <tr><td>9</td><td>0.40 : 0.60 → REFUSER</td><td>0.30 : 0.70 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.22 : 0.78 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.03 : 0.97 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-421-15</td></tr> <tr><td>10</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.37 : 0.63 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.31 : 0.69 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-482-15</td></tr> <tr><td>11</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.31 : 0.69 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.32 : 0.68 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.07 : 0.93 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-50-15</td></tr> <tr><td>12</td><td>0.40 : 0.60 → REFUSER</td><td>0.28 : 0.72 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.50 : 0.50 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.03 : 0.97 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-542-07</td></tr> <tr><td>13</td><td>0.40 : 0.60 → REFUSER</td><td>0.12 : 0.88 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.33 : 0.67 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-653-17</td></tr> <tr><td>14</td><td>0.40 : 0.60 → REFUSER</td><td>0.20 : 0.80 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.23 : 0.77 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-85-16 c86-16</td></tr> <tr><td>15</td><td>0.00 : 1.00 → REFUSER</td><td>0.06 : 0.94 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.38 : 0.62 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-98-11</td></tr> </tbody> </table>		kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	1	0.20 : 0.80 → REFUSER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	0.37 : 0.63 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...	2	0.40 : 0.60 → REFUSER	0.19 : 0.81 → REFUSER	1.00 : 0.00 → ACCEPTER	0.53 : 0.47 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-398-08	3	0.60 : 0.40 → ACCEPTER	0.10 : 0.90 → REFUSER	1.00 : 0.00 → ACCEPTER	0.31 : 0.69 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15	4	0.20 : 0.80 → REFUSER	0.36 : 0.64 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-56-16	5	0.60 : 0.40 → ACCEPTER	0.48 : 0.52 → REFUSER	1.00 : 0.00 → ACCEPTER	0.36 : 0.64 → REFUSER	1.00 : 0.00 → ACCEPTER	0.17 : 0.83 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-578-17	6	0.60 : 0.40 → ACCEPTER	0.07 : 0.93 → REFUSER	1.00 : 0.00 → ACCEPTER	0.37 : 0.63 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18	7	0.40 : 0.60 → REFUSER	0.25 : 0.75 → REFUSER	1.00 : 0.00 → ACCEPTER	0.28 : 0.72 → REFUSER	1.00 : 0.00 → ACCEPTER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15	8	0.20 : 0.80 → REFUSER	0.07 : 0.93 → REFUSER	1.00 : 0.00 → ACCEPTER	0.28 : 0.72 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11	9	0.40 : 0.60 → REFUSER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.22 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15	10	0.60 : 0.40 → ACCEPTER	0.37 : 0.63 → REFUSER	1.00 : 0.00 → ACCEPTER	0.31 : 0.69 → REFUSER	0.00 : 1.00 → REFUSER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15	11	0.60 : 0.40 → ACCEPTER	0.31 : 0.69 → REFUSER	1.00 : 0.00 → ACCEPTER	0.32 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTER	0.07 : 0.93 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15	12	0.40 : 0.60 → REFUSER	0.28 : 0.72 → REFUSER	1.00 : 0.00 → ACCEPTER	0.50 : 0.50 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-542-07	13	0.40 : 0.60 → REFUSER	0.12 : 0.88 → REFUSER	1.00 : 0.00 → ACCEPTER	0.33 : 0.67 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-653-17	14	0.40 : 0.60 → REFUSER	0.20 : 0.80 → REFUSER	1.00 : 0.00 → ACCEPTER	0.23 : 0.77 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16	15	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	1.00 : 0.00 → ACCEPTER	0.38 : 0.62 → REFUSER	1.00 : 0.00 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11	
	kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name																																																																																																																																																									
1	0.20 : 0.80 → REFUSER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	0.37 : 0.63 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...																																																																																																																																																									
2	0.40 : 0.60 → REFUSER	0.19 : 0.81 → REFUSER	1.00 : 0.00 → ACCEPTER	0.53 : 0.47 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-398-08																																																																																																																																																									
3	0.60 : 0.40 → ACCEPTER	0.10 : 0.90 → REFUSER	1.00 : 0.00 → ACCEPTER	0.31 : 0.69 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15																																																																																																																																																									
4	0.20 : 0.80 → REFUSER	0.36 : 0.64 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-56-16																																																																																																																																																									
5	0.60 : 0.40 → ACCEPTER	0.48 : 0.52 → REFUSER	1.00 : 0.00 → ACCEPTER	0.36 : 0.64 → REFUSER	1.00 : 0.00 → ACCEPTER	0.17 : 0.83 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-578-17																																																																																																																																																									
6	0.60 : 0.40 → ACCEPTER	0.07 : 0.93 → REFUSER	1.00 : 0.00 → ACCEPTER	0.37 : 0.63 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18																																																																																																																																																									
7	0.40 : 0.60 → REFUSER	0.25 : 0.75 → REFUSER	1.00 : 0.00 → ACCEPTER	0.28 : 0.72 → REFUSER	1.00 : 0.00 → ACCEPTER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15																																																																																																																																																									
8	0.20 : 0.80 → REFUSER	0.07 : 0.93 → REFUSER	1.00 : 0.00 → ACCEPTER	0.28 : 0.72 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11																																																																																																																																																									
9	0.40 : 0.60 → REFUSER	0.30 : 0.70 → REFUSER	1.00 : 0.00 → ACCEPTER	0.22 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15																																																																																																																																																									
10	0.60 : 0.40 → ACCEPTER	0.37 : 0.63 → REFUSER	1.00 : 0.00 → ACCEPTER	0.31 : 0.69 → REFUSER	0.00 : 1.00 → REFUSER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15																																																																																																																																																									
11	0.60 : 0.40 → ACCEPTER	0.31 : 0.69 → REFUSER	1.00 : 0.00 → ACCEPTER	0.32 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTER	0.07 : 0.93 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15																																																																																																																																																									
12	0.40 : 0.60 → REFUSER	0.28 : 0.72 → REFUSER	1.00 : 0.00 → ACCEPTER	0.50 : 0.50 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-542-07																																																																																																																																																									
13	0.40 : 0.60 → REFUSER	0.12 : 0.88 → REFUSER	1.00 : 0.00 → ACCEPTER	0.33 : 0.67 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-653-17																																																																																																																																																									
14	0.40 : 0.60 → REFUSER	0.20 : 0.80 → REFUSER	1.00 : 0.00 → ACCEPTER	0.23 : 0.77 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16																																																																																																																																																									
15	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	1.00 : 0.00 → ACCEPTER	0.38 : 0.62 → REFUSER	1.00 : 0.00 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11																																																																																																																																																									
<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Model</th> <th>AUC</th> <th>CA</th> <th>F1</th> <th>Precision</th> <th>Recall</th> </tr> </thead> <tbody> <tr><td>kNN</td><td>0.490</td><td>0.600</td><td>0.600</td><td>0.600</td><td>0.600</td></tr> <tr><td>Logistic Regression</td><td>0.520</td><td>0.667</td><td>0.533</td><td>0.444</td><td>0.667</td></tr> <tr><td>Naive Bayes</td><td>0.520</td><td>0.267</td><td>0.140</td><td>0.095</td><td>0.267</td></tr> <tr><td>Random Forest</td><td>0.660</td><td>0.667</td><td>0.617</td><td>0.628</td><td>0.667</td></tr> <tr><td>AdaBoost</td><td>0.350</td><td>0.400</td><td>0.411</td><td>0.426</td><td>0.400</td></tr> <tr><td>Neural Network</td><td>0.420</td><td>0.667</td><td>0.533</td><td>0.444</td><td>0.667</td></tr> <tr><td>SGD</td><td>0.500</td><td>0.667</td><td>0.533</td><td>0.444</td><td>0.667</td></tr> </tbody> </table>										Model	AUC	CA	F1	Precision	Recall	kNN	0.490	0.600	0.600	0.600	0.600	Logistic Regression	0.520	0.667	0.533	0.444	0.667	Naive Bayes	0.520	0.267	0.140	0.095	0.267	Random Forest	0.660	0.667	0.617	0.628	0.667	AdaBoost	0.350	0.400	0.411	0.426	0.400	Neural Network	0.420	0.667	0.533	0.444	0.667	SGD	0.500	0.667	0.533	0.444	0.667																																																																																																									
Model	AUC	CA	F1	Precision	Recall																																																																																																																																																													
kNN	0.490	0.600	0.600	0.600	0.600																																																																																																																																																													
Logistic Regression	0.520	0.667	0.533	0.444	0.667																																																																																																																																																													
Naive Bayes	0.520	0.267	0.140	0.095	0.267																																																																																																																																																													
Random Forest	0.660	0.667	0.617	0.628	0.667																																																																																																																																																													
AdaBoost	0.350	0.400	0.411	0.426	0.400																																																																																																																																																													
Neural Network	0.420	0.667	0.533	0.444	0.667																																																																																																																																																													
SGD	0.500	0.667	0.533	0.444	0.667																																																																																																																																																													

Pré-traitement des textes	Sac de mots
Mots des textes	mode comptage

Show probabilities for <input checked="" type="checkbox"/> REFUSER <input type="checkbox"/> ACCEPTER	<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>kNN</th> <th>Logistic Regression</th> <th>Naive Bayes</th> <th>Random Forest</th> <th>AdaBoost</th> <th>Neural Network</th> <th>SGD</th> <th>category</th> <th>name</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.60 : 0.40 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.98 : 0.02 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>1.00 : 0.00 → REFUSER</td><td>ACCEPTER</td><td>c-344-10 ...</td></tr> <tr><td>2</td><td>0.80 : 0.20 → REFUSER</td><td>0.16 : 0.84 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.78 : 0.22 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>ACCEPTER</td><td>c-398-08</td></tr> <tr><td>3</td><td>0.60 : 0.40 → REFUSER</td><td>0.51 : 0.49 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.80 : 0.20 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>ACCEPTER</td><td>c-501-15</td></tr> <tr><td>4</td><td>0.60 : 0.40 → REFUSER</td><td>0.22 : 0.78 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.60 : 0.40 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>ACCEPTER</td><td>c-56-16</td></tr> <tr><td>5</td><td>0.60 : 0.40 → REFUSER</td><td>0.26 : 0.74 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.70 : 0.30 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>1.00 : 0.00 → REFUSER</td><td>ACCEPTER</td><td>c-578-17</td></tr> <tr><td>6</td><td>0.60 : 0.40 → REFUSER</td><td>0.97 : 0.03 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.40 : 0.60 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>1.00 : 0.00 → REFUSER</td><td>REFUSER</td><td>c-124-18</td></tr> <tr><td>7</td><td>0.40 : 0.60 → ACCEPTER</td><td>0.05 : 0.95 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.60 : 0.40 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>REFUSER</td><td>c-226-15</td></tr> <tr><td>8</td><td>0.80 : 0.20 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.70 : 0.30 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>1.00 : 0.00 → REFUSER</td><td>REFUSER</td><td>c-311-11</td></tr> <tr><td>9</td><td>0.60 : 0.40 → REFUSER</td><td>0.60 : 0.40 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.73 : 0.27 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>REFUSER</td><td>c-421-15</td></tr> <tr><td>10</td><td>0.60 : 0.40 → REFUSER</td><td>0.92 : 0.08 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.53 : 0.47 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>REFUSER</td><td>c-482-15</td></tr> <tr><td>11</td><td>0.60 : 0.40 → REFUSER</td><td>0.98 : 0.02 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.83 : 0.17 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>1.00 : 0.00 → REFUSER</td><td>REFUSER</td><td>c-50-15</td></tr> <tr><td>12</td><td>0.40 : 0.60 → ACCEPTER</td><td>0.07 : 0.93 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.70 : 0.30 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>REFUSER</td><td>c-542-07</td></tr> <tr><td>13</td><td>0.40 : 0.60 → ACCEPTER</td><td>0.13 : 0.87 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.83 : 0.17 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>REFUSER</td><td>c-653-17</td></tr> <tr><td>14</td><td>0.60 : 0.40 → REFUSER</td><td>0.99 : 0.01 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.82 : 0.18 → REFUSER</td><td>1.00 : 0.00 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>REFUSER</td><td>c-85-16 c86-16</td></tr> <tr><td>15</td><td>0.60 : 0.40 → REFUSER</td><td>0.84 : 0.16 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.78 : 0.22 → REFUSER</td><td>0.00 : 1.00 → ACCEPTER</td><td>0.00 : 1.00 → ACCEPTER</td><td>1.00 : 0.00 → REFUSER</td><td>REFUSER</td><td>c-98-11</td></tr> </tbody> </table>		kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	1	0.60 : 0.40 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.98 : 0.02 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	ACCEPTER	c-344-10 ...	2	0.80 : 0.20 → REFUSER	0.16 : 0.84 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.78 : 0.22 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	ACCEPTER	c-398-08	3	0.60 : 0.40 → REFUSER	0.51 : 0.49 → REFUSER	0.00 : 1.00 → ACCEPTER	0.80 : 0.20 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	ACCEPTER	c-501-15	4	0.60 : 0.40 → REFUSER	0.22 : 0.78 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.60 : 0.40 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	ACCEPTER	c-56-16	5	0.60 : 0.40 → REFUSER	0.26 : 0.74 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.70 : 0.30 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	ACCEPTER	c-578-17	6	0.60 : 0.40 → REFUSER	0.97 : 0.03 → REFUSER	0.00 : 1.00 → ACCEPTER	0.40 : 0.60 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-124-18	7	0.40 : 0.60 → ACCEPTER	0.05 : 0.95 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.60 : 0.40 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-226-15	8	0.80 : 0.20 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.70 : 0.30 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-311-11	9	0.60 : 0.40 → REFUSER	0.60 : 0.40 → REFUSER	0.00 : 1.00 → ACCEPTER	0.73 : 0.27 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-421-15	10	0.60 : 0.40 → REFUSER	0.92 : 0.08 → REFUSER	0.00 : 1.00 → ACCEPTER	0.53 : 0.47 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-482-15	11	0.60 : 0.40 → REFUSER	0.98 : 0.02 → REFUSER	0.00 : 1.00 → ACCEPTER	0.83 : 0.17 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-50-15	12	0.40 : 0.60 → ACCEPTER	0.07 : 0.93 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.70 : 0.30 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-542-07	13	0.40 : 0.60 → ACCEPTER	0.13 : 0.87 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.83 : 0.17 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-653-17	14	0.60 : 0.40 → REFUSER	0.99 : 0.01 → REFUSER	0.00 : 1.00 → ACCEPTER	0.82 : 0.18 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-85-16 c86-16	15	0.60 : 0.40 → REFUSER	0.84 : 0.16 → REFUSER	0.00 : 1.00 → ACCEPTER	0.78 : 0.22 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-98-11	
	kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name																																																																																																																																																									
1	0.60 : 0.40 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.98 : 0.02 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	ACCEPTER	c-344-10 ...																																																																																																																																																									
2	0.80 : 0.20 → REFUSER	0.16 : 0.84 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.78 : 0.22 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	ACCEPTER	c-398-08																																																																																																																																																									
3	0.60 : 0.40 → REFUSER	0.51 : 0.49 → REFUSER	0.00 : 1.00 → ACCEPTER	0.80 : 0.20 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	ACCEPTER	c-501-15																																																																																																																																																									
4	0.60 : 0.40 → REFUSER	0.22 : 0.78 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.60 : 0.40 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	ACCEPTER	c-56-16																																																																																																																																																									
5	0.60 : 0.40 → REFUSER	0.26 : 0.74 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.70 : 0.30 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	ACCEPTER	c-578-17																																																																																																																																																									
6	0.60 : 0.40 → REFUSER	0.97 : 0.03 → REFUSER	0.00 : 1.00 → ACCEPTER	0.40 : 0.60 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-124-18																																																																																																																																																									
7	0.40 : 0.60 → ACCEPTER	0.05 : 0.95 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.60 : 0.40 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-226-15																																																																																																																																																									
8	0.80 : 0.20 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.70 : 0.30 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-311-11																																																																																																																																																									
9	0.60 : 0.40 → REFUSER	0.60 : 0.40 → REFUSER	0.00 : 1.00 → ACCEPTER	0.73 : 0.27 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-421-15																																																																																																																																																									
10	0.60 : 0.40 → REFUSER	0.92 : 0.08 → REFUSER	0.00 : 1.00 → ACCEPTER	0.53 : 0.47 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-482-15																																																																																																																																																									
11	0.60 : 0.40 → REFUSER	0.98 : 0.02 → REFUSER	0.00 : 1.00 → ACCEPTER	0.83 : 0.17 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-50-15																																																																																																																																																									
12	0.40 : 0.60 → ACCEPTER	0.07 : 0.93 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.70 : 0.30 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-542-07																																																																																																																																																									
13	0.40 : 0.60 → ACCEPTER	0.13 : 0.87 → ACCEPTER	0.00 : 1.00 → ACCEPTER	0.83 : 0.17 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-653-17																																																																																																																																																									
14	0.60 : 0.40 → REFUSER	0.99 : 0.01 → REFUSER	0.00 : 1.00 → ACCEPTER	0.82 : 0.18 → REFUSER	1.00 : 0.00 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	REFUSER	c-85-16 c86-16																																																																																																																																																									
15	0.60 : 0.40 → REFUSER	0.84 : 0.16 → REFUSER	0.00 : 1.00 → ACCEPTER	0.78 : 0.22 → REFUSER	0.00 : 1.00 → ACCEPTER	0.00 : 1.00 → ACCEPTER	1.00 : 0.00 → REFUSER	REFUSER	c-98-11																																																																																																																																																									
<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Model</th> <th>AUC</th> <th>CA</th> <th>F1</th> <th>Precision</th> <th>Recall</th> </tr> </thead> <tbody> <tr><td>kNN</td><td>0.330</td><td>0.467</td><td>0.424</td><td>0.389</td><td>0.467</td></tr> <tr><td>Logistic Regression</td><td>0.560</td><td>0.667</td><td>0.673</td><td>0.685</td><td>0.667</td></tr> <tr><td>Naive Bayes</td><td>0.500</td><td>0.333</td><td>0.167</td><td>0.111</td><td>0.333</td></tr> <tr><td>Random Forest</td><td>0.400</td><td>0.600</td><td>0.500</td><td>0.429</td><td>0.600</td></tr> <tr><td>AdaBoost</td><td>0.450</td><td>0.467</td><td>0.481</td><td>0.512</td><td>0.467</td></tr> <tr><td>Neural Network</td><td>0.440</td><td>0.333</td><td>0.167</td><td>0.111</td><td>0.333</td></tr> <tr><td>SGD</td><td>0.500</td><td>0.467</td><td>0.476</td><td>0.556</td><td>0.467</td></tr> </tbody> </table>										Model	AUC	CA	F1	Precision	Recall	kNN	0.330	0.467	0.424	0.389	0.467	Logistic Regression	0.560	0.667	0.673	0.685	0.667	Naive Bayes	0.500	0.333	0.167	0.111	0.333	Random Forest	0.400	0.600	0.500	0.429	0.600	AdaBoost	0.450	0.467	0.481	0.512	0.467	Neural Network	0.440	0.333	0.167	0.111	0.333	SGD	0.500	0.467	0.476	0.556	0.467																																																																																																									
Model	AUC	CA	F1	Precision	Recall																																																																																																																																																													
kNN	0.330	0.467	0.424	0.389	0.467																																																																																																																																																													
Logistic Regression	0.560	0.667	0.673	0.685	0.667																																																																																																																																																													
Naive Bayes	0.500	0.333	0.167	0.111	0.333																																																																																																																																																													
Random Forest	0.400	0.600	0.500	0.429	0.600																																																																																																																																																													
AdaBoost	0.450	0.467	0.481	0.512	0.467																																																																																																																																																													
Neural Network	0.440	0.333	0.167	0.111	0.333																																																																																																																																																													
SGD	0.500	0.467	0.476	0.556	0.467																																																																																																																																																													

Pré-traitement des textes	1-2	Sac de mots
Mots des textes		mode binaire

Show probabilities for <input type="checkbox"/> ACCEPTER <input type="checkbox"/> REFUSER		kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	
	1	0.60 : 0.40 → ACCEPTER	0.03 : 0.97 → REFUSER	1.00 : 0.00 → ACCEPTER	0.18 : 0.82 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...
	2	0.60 : 0.40 → ACCEPTER	0.15 : 0.85 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.28 : 0.72 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-398-08
	3	0.40 : 0.60 → REFUSER	0.12 : 0.88 → REFUSER	1.00 : 0.00 → ACCEPTER	0.10 : 0.90 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15
	4	0.60 : 0.40 → ACCEPTER	0.26 : 0.74 → REFUSER	1.00 : 0.00 → ACCEPTER	0.49 : 0.51 → REFUSER	0.00 : 1.00 → REFUSER	0.67 : 0.33 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-56-16
	5	0.60 : 0.40 → ACCEPTER	0.56 : 0.44 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.35 : 0.65 → REFUSER	1.00 : 0.00 → ACCEPTER	0.93 : 0.07 → ACCEPTER	1.00 : 0.00 → ACCEPTER	1.00 : 0.00 → ACCEPTER	ACCEPTER	c-578-17
	6	0.60 : 0.40 → ACCEPTER	0.20 : 0.80 → REFUSER	1.00 : 0.00 → ACCEPTER	0.35 : 0.65 → REFUSER	0.00 : 1.00 → REFUSER	0.07 : 0.93 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18
	7	0.40 : 0.60 → REFUSER	0.25 : 0.75 → REFUSER	1.00 : 0.00 → ACCEPTER	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTER	0.29 : 0.71 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15
	8	0.40 : 0.60 → REFUSER	0.06 : 0.94 → REFUSER	1.00 : 0.00 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11
	9	0.60 : 0.40 → ACCEPTER	0.27 : 0.73 → REFUSER	1.00 : 0.00 → ACCEPTER	0.37 : 0.63 → REFUSER	0.00 : 1.00 → REFUSER	0.91 : 0.09 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15
	10	0.60 : 0.40 → ACCEPTER	0.20 : 0.80 → REFUSER	1.00 : 0.00 → ACCEPTER	0.10 : 0.90 → REFUSER	0.00 : 1.00 → REFUSER	0.32 : 0.68 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15
	11	0.40 : 0.60 → REFUSER	0.16 : 0.84 → REFUSER	1.00 : 0.00 → ACCEPTER	0.10 : 0.90 → REFUSER	1.00 : 0.00 → ACCEPTER	0.04 : 0.96 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15
	12	0.60 : 0.40 → ACCEPTER	0.27 : 0.73 → REFUSER	1.00 : 0.00 → ACCEPTER	0.10 : 0.90 → REFUSER	1.00 : 0.00 → ACCEPTER	0.47 : 0.53 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-542-07
	13	0.60 : 0.40 → ACCEPTER	0.13 : 0.87 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-653-17
	14	0.60 : 0.40 → ACCEPTER	0.16 : 0.84 → REFUSER	1.00 : 0.00 → ACCEPTER	0.09 : 0.91 → REFUSER	0.00 : 1.00 → REFUSER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16
	15	0.60 : 0.40 → ACCEPTER	0.07 : 0.93 → REFUSER	1.00 : 0.00 → ACCEPTER	0.10 : 0.90 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11

Model	AUC	CA	F1	Precision	Recall
kNN	0.550	0.467	0.452	0.621	0.467
Logistic Regression	0.460	0.733	0.667	0.810	0.733
Naive Bayes	0.520	0.333	0.167	0.111	0.333
Random Forest	0.680	0.667	0.533	0.444	0.667
AdaBoost	0.450	0.533	0.519	0.508	0.533
Neural Network	0.520	0.733	0.712	0.722	0.733
SGD	0.600	0.733	0.667	0.810	0.733

Pré-traitement des textes	1-1	Sac de mots
Mots de la liste (annexe E p86)		mode comptage

Show probabilities for <input type="checkbox"/> ACCEPTER <input type="checkbox"/> REFUSER		kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	
	1	0.40 : 0.60 → REFUSER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	0.12 : 0.88 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...
	2	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	0.63 : 0.37 → ACCEPTER	0.12 : 0.88 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-398-08			
	3	0.40 : 0.60 → REFUSER	0.77 : 0.23 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.72 : 0.28 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15
	4	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	0.35 : 0.65 → REFUSER	0.60 : 0.40 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-56-16
	5	0.20 : 0.80 → REFUSER	0.15 : 0.85 → REFUSER	0.78 : 0.22 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-578-17			
	6	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	0.90 : 0.10 → ACCEPTER	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18
	7	0.40 : 0.60 → REFUSER	0.02 : 0.98 → REFUSER	0.44 : 0.56 → REFUSER	0.10 : 0.90 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15			
	8	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.55 : 0.45 → ACCEPTER	0.05 : 0.95 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11			
	9	0.40 : 0.60 → REFUSER	0.94 : 0.06 → ACCEPTER	0.98 : 0.02 → ACCEPTER	0.33 : 0.67 → REFUSER	1.00 : 0.00 → ACCEPTER	0.32 : 0.68 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15
	10	0.40 : 0.60 → REFUSER	0.01 : 0.99 → REFUSER	1.00 : 0.00 → ACCEPTER	0.35 : 0.65 → REFUSER	1.00 : 0.00 → ACCEPTER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15
	11	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	0.92 : 0.08 → ACCEPTER	0.05 : 0.95 → REFUSER	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15
	12	0.80 : 0.20 → ACCEPTER	1.00 : 0.00 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.60 : 0.40 → ACCEPTER	1.00 : 0.00 → ACCEPTER	REFUSER	c-542-07			
	13	0.60 : 0.40 → ACCEPTER	1.00 : 0.00 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.47 : 0.53 → REFUSER	0.00 : 1.00 → REFUSER	0.87 : 0.13 → ACCEPTER	1.00 : 0.00 → ACCEPTER	1.00 : 0.00 → ACCEPTER	REFUSER	c-653-17
	14	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	0.69 : 0.31 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.03 : 0.97 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16
	15	0.40 : 0.60 → REFUSER	0.98 : 0.02 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.53 : 0.47 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.13 : 0.87 → REFUSER	1.00 : 0.00 → ACCEPTER	1.00 : 0.00 → ACCEPTER	REFUSER	c-98-11

Model	AUC	CA	F1	Precision	Recall
kNN	0.370	0.533	0.464	0.410	0.533
Logistic Regression	0.460	0.467	0.467	0.467	0.467
Naive Bayes	0.220	0.267	0.220	0.306	0.267
Random Forest	0.460	0.600	0.568	0.556	0.600
AdaBoost	0.350	0.400	0.411	0.426	0.400
Neural Network	0.360	0.600	0.568	0.556	0.600
SGD	0.350	0.467	0.424	0.389	0.467

Pré-traitement des textes	1-1	Sac de mots
Mots de la liste (annexe E p86)		mode binaire

Show probabilities for <input type="checkbox"/> ACCEPTTER <input type="checkbox"/> REFUSER	<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>kNN</th> <th>Logistic Regression</th> <th>Naive Bayes</th> <th>Random Forest</th> <th>AdaBoost</th> <th>Neural Network</th> <th>SGD</th> <th>category</th> <th>name</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.00 : 1.00 → REFUSER</td><td>0.13 : 0.87 → REFUSER</td><td>0.02 : 0.98 → REFUSER</td><td>0.08 : 0.92 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.04 : 0.96 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-344-10 ...</td></tr> <tr><td>2</td><td>0.40 : 0.60 → REFUSER</td><td>0.26 : 0.74 → REFUSER</td><td>0.68 : 0.32 → ACCEPTTER</td><td>0.47 : 0.53 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.35 : 0.65 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-398-08</td></tr> <tr><td>3</td><td>0.00 : 1.00 → REFUSER</td><td>0.06 : 0.94 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.15 : 0.85 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-501-15</td></tr> <tr><td>4</td><td>0.60 : 0.40 → ACCEPTTER</td><td>0.31 : 0.69 → REFUSER</td><td>0.46 : 0.54 → REFUSER</td><td>0.55 : 0.45 → ACCEPTTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.56 : 0.44 → ACCEPTTER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-56-16</td></tr> <tr><td>5</td><td>0.00 : 1.00 → REFUSER</td><td>0.08 : 0.92 → REFUSER</td><td>0.40 : 0.60 → REFUSER</td><td>0.17 : 0.83 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.02 : 0.98 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-578-17</td></tr> <tr><td>6</td><td>0.00 : 1.00 → REFUSER</td><td>0.06 : 0.94 → REFUSER</td><td>0.08 : 0.92 → REFUSER</td><td>0.33 : 0.67 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-124-18</td></tr> <tr><td>7</td><td>0.00 : 1.00 → REFUSER</td><td>0.06 : 0.94 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.21 : 0.79 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-226-15</td></tr> <tr><td>8</td><td>0.00 : 1.00 → REFUSER</td><td>0.09 : 0.91 → REFUSER</td><td>0.03 : 0.97 → REFUSER</td><td>0.16 : 0.84 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-311-11</td></tr> <tr><td>9</td><td>0.20 : 0.80 → REFUSER</td><td>0.11 : 0.89 → REFUSER</td><td>0.22 : 0.78 → REFUSER</td><td>0.36 : 0.64 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-421-15</td></tr> <tr><td>10</td><td>0.00 : 1.00 → REFUSER</td><td>0.26 : 0.74 → REFUSER</td><td>0.13 : 0.87 → REFUSER</td><td>0.31 : 0.69 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-482-15</td></tr> <tr><td>11</td><td>0.40 : 0.60 → REFUSER</td><td>0.25 : 0.75 → REFUSER</td><td>0.50 : 0.50 → REFUSER</td><td>0.38 : 0.62 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.08 : 0.92 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-50-15</td></tr> <tr><td>12</td><td>0.80 : 0.20 → ACCEPTTER</td><td>0.64 : 0.36 → ACCEPTTER</td><td>0.98 : 0.02 → ACCEPTTER</td><td>0.66 : 0.35 → ACCEPTTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.06 : 0.94 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>REFUSER</td><td>c-542-07</td></tr> <tr><td>13</td><td>0.60 : 0.40 → ACCEPTTER</td><td>0.37 : 0.63 → REFUSER</td><td>0.81 : 0.19 → ACCEPTTER</td><td>0.56 : 0.44 → ACCEPTTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.08 : 0.92 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-653-17</td></tr> <tr><td>14</td><td>0.40 : 0.60 → REFUSER</td><td>0.19 : 0.81 → REFUSER</td><td>0.10 : 0.90 → REFUSER</td><td>0.38 : 0.62 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.02 : 0.98 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-85-16 c86-16</td></tr> <tr><td>15</td><td>0.40 : 0.60 → REFUSER</td><td>0.23 : 0.77 → REFUSER</td><td>0.51 : 0.49 → ACCEPTTER</td><td>0.32 : 0.68 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.08 : 0.92 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-98-11</td></tr> </tbody> </table>		kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	1	0.00 : 1.00 → REFUSER	0.13 : 0.87 → REFUSER	0.02 : 0.98 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	0.04 : 0.96 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-344-10 ...	2	0.40 : 0.60 → REFUSER	0.26 : 0.74 → REFUSER	0.68 : 0.32 → ACCEPTTER	0.47 : 0.53 → REFUSER	0.00 : 1.00 → REFUSER	0.35 : 0.65 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-398-08	3	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.01 : 0.99 → REFUSER	0.15 : 0.85 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-501-15	4	0.60 : 0.40 → ACCEPTTER	0.31 : 0.69 → REFUSER	0.46 : 0.54 → REFUSER	0.55 : 0.45 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.56 : 0.44 → ACCEPTTER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-56-16	5	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.40 : 0.60 → REFUSER	0.17 : 0.83 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-578-17	6	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.08 : 0.92 → REFUSER	0.33 : 0.67 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18	7	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.01 : 0.99 → REFUSER	0.21 : 0.79 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15	8	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.03 : 0.97 → REFUSER	0.16 : 0.84 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11	9	0.20 : 0.80 → REFUSER	0.11 : 0.89 → REFUSER	0.22 : 0.78 → REFUSER	0.36 : 0.64 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15	10	0.00 : 1.00 → REFUSER	0.26 : 0.74 → REFUSER	0.13 : 0.87 → REFUSER	0.31 : 0.69 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15	11	0.40 : 0.60 → REFUSER	0.25 : 0.75 → REFUSER	0.50 : 0.50 → REFUSER	0.38 : 0.62 → REFUSER	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15	12	0.80 : 0.20 → ACCEPTTER	0.64 : 0.36 → ACCEPTTER	0.98 : 0.02 → ACCEPTTER	0.66 : 0.35 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-542-07	13	0.60 : 0.40 → ACCEPTTER	0.37 : 0.63 → REFUSER	0.81 : 0.19 → ACCEPTTER	0.56 : 0.44 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-653-17	14	0.40 : 0.60 → REFUSER	0.19 : 0.81 → REFUSER	0.10 : 0.90 → REFUSER	0.38 : 0.62 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16	15	0.40 : 0.60 → REFUSER	0.23 : 0.77 → REFUSER	0.51 : 0.49 → ACCEPTTER	0.32 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11	
	kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name																																																																																																																																																									
1	0.00 : 1.00 → REFUSER	0.13 : 0.87 → REFUSER	0.02 : 0.98 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	0.04 : 0.96 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-344-10 ...																																																																																																																																																									
2	0.40 : 0.60 → REFUSER	0.26 : 0.74 → REFUSER	0.68 : 0.32 → ACCEPTTER	0.47 : 0.53 → REFUSER	0.00 : 1.00 → REFUSER	0.35 : 0.65 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-398-08																																																																																																																																																									
3	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.01 : 0.99 → REFUSER	0.15 : 0.85 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-501-15																																																																																																																																																									
4	0.60 : 0.40 → ACCEPTTER	0.31 : 0.69 → REFUSER	0.46 : 0.54 → REFUSER	0.55 : 0.45 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.56 : 0.44 → ACCEPTTER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-56-16																																																																																																																																																									
5	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.40 : 0.60 → REFUSER	0.17 : 0.83 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-578-17																																																																																																																																																									
6	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.08 : 0.92 → REFUSER	0.33 : 0.67 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18																																																																																																																																																									
7	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.01 : 0.99 → REFUSER	0.21 : 0.79 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15																																																																																																																																																									
8	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.03 : 0.97 → REFUSER	0.16 : 0.84 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11																																																																																																																																																									
9	0.20 : 0.80 → REFUSER	0.11 : 0.89 → REFUSER	0.22 : 0.78 → REFUSER	0.36 : 0.64 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15																																																																																																																																																									
10	0.00 : 1.00 → REFUSER	0.26 : 0.74 → REFUSER	0.13 : 0.87 → REFUSER	0.31 : 0.69 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15																																																																																																																																																									
11	0.40 : 0.60 → REFUSER	0.25 : 0.75 → REFUSER	0.50 : 0.50 → REFUSER	0.38 : 0.62 → REFUSER	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15																																																																																																																																																									
12	0.80 : 0.20 → ACCEPTTER	0.64 : 0.36 → ACCEPTTER	0.98 : 0.02 → ACCEPTTER	0.66 : 0.35 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-542-07																																																																																																																																																									
13	0.60 : 0.40 → ACCEPTTER	0.37 : 0.63 → REFUSER	0.81 : 0.19 → ACCEPTTER	0.56 : 0.44 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-653-17																																																																																																																																																									
14	0.40 : 0.60 → REFUSER	0.19 : 0.81 → REFUSER	0.10 : 0.90 → REFUSER	0.38 : 0.62 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16																																																																																																																																																									
15	0.40 : 0.60 → REFUSER	0.23 : 0.77 → REFUSER	0.51 : 0.49 → ACCEPTTER	0.32 : 0.68 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11																																																																																																																																																									
Restore Original Order	<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Model</th> <th>AUC</th> <th>CA</th> <th>F1</th> <th>Precision</th> <th>Recall</th> </tr> </thead> <tbody> <tr><td>kNN</td><td>0.420</td><td>0.600</td><td>0.568</td><td>0.556</td><td>0.600</td></tr> <tr><td>Logistic Regression</td><td>0.480</td><td>0.600</td><td>0.500</td><td>0.429</td><td>0.600</td></tr> <tr><td>Naive Bayes</td><td>0.420</td><td>0.533</td><td>0.519</td><td>0.508</td><td>0.533</td></tr> <tr><td>Random Forest</td><td>0.340</td><td>0.600</td><td>0.568</td><td>0.556</td><td>0.600</td></tr> <tr><td>AdaBoost</td><td>0.400</td><td>0.533</td><td>0.464</td><td>0.410</td><td>0.533</td></tr> <tr><td>Neural Network</td><td>0.700</td><td>0.733</td><td>0.667</td><td>0.810</td><td>0.733</td></tr> <tr><td>SGD</td><td>0.450</td><td>0.600</td><td>0.500</td><td>0.429</td><td>0.600</td></tr> </tbody> </table>	Model	AUC	CA	F1	Precision	Recall	kNN	0.420	0.600	0.568	0.556	0.600	Logistic Regression	0.480	0.600	0.500	0.429	0.600	Naive Bayes	0.420	0.533	0.519	0.508	0.533	Random Forest	0.340	0.600	0.568	0.556	0.600	AdaBoost	0.400	0.533	0.464	0.410	0.533	Neural Network	0.700	0.733	0.667	0.810	0.733	SGD	0.450	0.600	0.500	0.429	0.600																																																																																																																	
Model	AUC	CA	F1	Precision	Recall																																																																																																																																																													
kNN	0.420	0.600	0.568	0.556	0.600																																																																																																																																																													
Logistic Regression	0.480	0.600	0.500	0.429	0.600																																																																																																																																																													
Naive Bayes	0.420	0.533	0.519	0.508	0.533																																																																																																																																																													
Random Forest	0.340	0.600	0.568	0.556	0.600																																																																																																																																																													
AdaBoost	0.400	0.533	0.464	0.410	0.533																																																																																																																																																													
Neural Network	0.700	0.733	0.667	0.810	0.733																																																																																																																																																													
SGD	0.450	0.600	0.500	0.429	0.600																																																																																																																																																													

Pré-traitement des textes	1-2	Sac de mots
Mots de la liste (annexe E p86)		mode comptage

Show probabilities for <input type="checkbox"/> ACCEPTTER <input type="checkbox"/> REFUSER	<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>kNN</th> <th>Logistic Regression</th> <th>Naive Bayes</th> <th>Random Forest</th> <th>AdaBoost</th> <th>Neural Network</th> <th>SGD</th> <th>category</th> <th>name</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.40 : 0.60 → REFUSER</td><td>0.22 : 0.78 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-344-10 ...</td></tr> <tr><td>2</td><td>0.40 : 0.60 → REFUSER</td><td>0.70 : 0.30 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.19 : 0.81 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>ACCEPTTER</td><td>c-398-08</td></tr> <tr><td>3</td><td>0.60 : 0.40 → ACCEPTTER</td><td>0.99 : 0.01 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.24 : 0.76 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>ACCEPTTER</td><td>c-501-15</td></tr> <tr><td>4</td><td>0.40 : 0.60 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.50 : 0.50 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.99 : 0.01 → ACCEPTTER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-56-16</td></tr> <tr><td>5</td><td>0.20 : 0.80 → REFUSER</td><td>0.18 : 0.82 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.48 : 0.52 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.30 : 0.70 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTTER</td><td>c-578-17</td></tr> <tr><td>6</td><td>0.40 : 0.60 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.11 : 0.89 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.09 : 0.91 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>REFUSER</td><td>c-124-18</td></tr> <tr><td>7</td><td>0.40 : 0.60 → REFUSER</td><td>0.74 : 0.26 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.36 : 0.64 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.05 : 0.95 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>REFUSER</td><td>c-226-15</td></tr> <tr><td>8</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.06 : 0.94 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.04 : 0.96 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-311-11</td></tr> <tr><td>9</td><td>0.40 : 0.60 → REFUSER</td><td>0.95 : 0.05 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.10 : 0.90 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.97 : 0.03 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>REFUSER</td><td>c-421-15</td></tr> <tr><td>10</td><td>0.60 : 0.40 → ACCEPTTER</td><td>0.08 : 0.92 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.16 : 0.84 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.11 : 0.89 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-482-15</td></tr> <tr><td>11</td><td>0.40 : 0.60 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.30 : 0.70 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.05 : 0.95 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-50-15</td></tr> <tr><td>12</td><td>0.60 : 0.40 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.23 : 0.78 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>REFUSER</td><td>c-542-07</td></tr> <tr><td>13</td><td>0.40 : 0.60 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.22 : 0.78 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.84 : 0.16 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>REFUSER</td><td>c-653-17</td></tr> <tr><td>14</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.23 : 0.78 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-85-16 c86-16</td></tr> <tr><td>15</td><td>0.40 : 0.60 → REFUSER</td><td>0.98 : 0.02 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>0.13 : 0.87 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.88 : 0.12 → ACCEPTTER</td><td>1.00 : 0.00 → ACCEPTTER</td><td>REFUSER</td><td>c-98-11</td></tr> </tbody> </table>		kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	1	0.40 : 0.60 → REFUSER	0.22 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-344-10 ...	2	0.40 : 0.60 → REFUSER	0.70 : 0.30 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.19 : 0.81 → REFUSER	1.00 : 0.00 → ACCEPTTER	ACCEPTTER	c-398-08	3	0.60 : 0.40 → ACCEPTTER	0.99 : 0.01 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.24 : 0.76 → REFUSER	1.00 : 0.00 → ACCEPTTER	ACCEPTTER	c-501-15	4	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.50 : 0.50 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.99 : 0.01 → ACCEPTTER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-56-16	5	0.20 : 0.80 → REFUSER	0.18 : 0.82 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.48 : 0.52 → REFUSER	0.00 : 1.00 → REFUSER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-578-17	6	0.40 : 0.60 → REFUSER	0.01 : 0.99 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.11 : 0.89 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.09 : 0.91 → REFUSER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-124-18	7	0.40 : 0.60 → REFUSER	0.74 : 0.26 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.36 : 0.64 → REFUSER	0.00 : 1.00 → REFUSER	0.05 : 0.95 → REFUSER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-226-15	8	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.06 : 0.94 → REFUSER	0.00 : 1.00 → REFUSER	0.04 : 0.96 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11	9	0.40 : 0.60 → REFUSER	0.95 : 0.05 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.10 : 0.90 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.97 : 0.03 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-421-15	10	0.60 : 0.40 → ACCEPTTER	0.08 : 0.92 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.16 : 0.84 → REFUSER	0.00 : 1.00 → REFUSER	0.11 : 0.89 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15	11	0.40 : 0.60 → REFUSER	0.01 : 0.99 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.05 : 0.95 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15	12	0.60 : 0.40 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.23 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-542-07	13	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.22 : 0.78 → REFUSER	0.00 : 1.00 → REFUSER	0.84 : 0.16 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-653-17	14	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.23 : 0.78 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16	15	0.40 : 0.60 → REFUSER	0.98 : 0.02 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.13 : 0.87 → REFUSER	0.00 : 1.00 → REFUSER	0.88 : 0.12 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-98-11	
	kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name																																																																																																																																																									
1	0.40 : 0.60 → REFUSER	0.22 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-344-10 ...																																																																																																																																																									
2	0.40 : 0.60 → REFUSER	0.70 : 0.30 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.19 : 0.81 → REFUSER	1.00 : 0.00 → ACCEPTTER	ACCEPTTER	c-398-08																																																																																																																																																									
3	0.60 : 0.40 → ACCEPTTER	0.99 : 0.01 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.24 : 0.76 → REFUSER	1.00 : 0.00 → ACCEPTTER	ACCEPTTER	c-501-15																																																																																																																																																									
4	0.40 : 0.60 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.50 : 0.50 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.99 : 0.01 → ACCEPTTER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-56-16																																																																																																																																																									
5	0.20 : 0.80 → REFUSER	0.18 : 0.82 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.48 : 0.52 → REFUSER	0.00 : 1.00 → REFUSER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTTER	c-578-17																																																																																																																																																									
6	0.40 : 0.60 → REFUSER	0.01 : 0.99 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.11 : 0.89 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.09 : 0.91 → REFUSER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-124-18																																																																																																																																																									
7	0.40 : 0.60 → REFUSER	0.74 : 0.26 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.36 : 0.64 → REFUSER	0.00 : 1.00 → REFUSER	0.05 : 0.95 → REFUSER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-226-15																																																																																																																																																									
8	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.06 : 0.94 → REFUSER	0.00 : 1.00 → REFUSER	0.04 : 0.96 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11																																																																																																																																																									
9	0.40 : 0.60 → REFUSER	0.95 : 0.05 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.10 : 0.90 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.97 : 0.03 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-421-15																																																																																																																																																									
10	0.60 : 0.40 → ACCEPTTER	0.08 : 0.92 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.16 : 0.84 → REFUSER	0.00 : 1.00 → REFUSER	0.11 : 0.89 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15																																																																																																																																																									
11	0.40 : 0.60 → REFUSER	0.01 : 0.99 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.05 : 0.95 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15																																																																																																																																																									
12	0.60 : 0.40 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.23 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-542-07																																																																																																																																																									
13	0.40 : 0.60 → REFUSER	1.00 : 0.00 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.22 : 0.78 → REFUSER	0.00 : 1.00 → REFUSER	0.84 : 0.16 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-653-17																																																																																																																																																									
14	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	1.00 : 0.00 → ACCEPTTER	0.23 : 0.78 → REFUSER	0.00 : 1.00 → REFUSER	0.01 : 0.99 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16																																																																																																																																																									
15	0.40 : 0.60 → REFUSER	0.98 : 0.02 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	0.13 : 0.87 → REFUSER	0.00 : 1.00 → REFUSER	0.88 : 0.12 → ACCEPTTER	1.00 : 0.00 → ACCEPTTER	REFUSER	c-98-11																																																																																																																																																									
Restore Original Order	<table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Model</th> <th>AUC</th> <th>CA</th> <th>F1</th> <th>Precision</th> <th>Recall</th> </tr> </thead> <tbody> <tr><td>kNN</td><td>0.500</td><td>0.600</td><td>0.568</td><td>0.556</td><td>0.600</td></tr> <tr><td>Logistic Regression</td><td>0.460</td><td>0.467</td><td>0.481</td><td>0.512</td><td>0.467</td></tr> <tr><td>Naive Bayes</td><td>0.400</td><td>0.333</td><td>0.167</td><td>0.111</td><td>0.333</td></tr> <tr><td>Random Forest</td><td>0.600</td><td>0.667</td><td>0.533</td><td>0.444</td><td>0.667</td></tr> <tr><td>AdaBoost</td><td>0.450</td><td>0.533</td><td>0.519</td><td>0.508</td><td>0.533</td></tr> <tr><td>Neural Network</td><td>0.540</td><td>0.467</td><td>0.467</td><td>0.467</td><td>0.467</td></tr> <tr><td>SGD</td><td>0.400</td><td>0.400</td><td>0.416</td><td>0.464</td><td>0.400</td></tr> </tbody> </table>	Model	AUC	CA	F1	Precision	Recall	kNN	0.500	0.600	0.568	0.556	0.600	Logistic Regression	0.460	0.467	0.481	0.512	0.467	Naive Bayes	0.400	0.333	0.167	0.111	0.333	Random Forest	0.600	0.667	0.533	0.444	0.667	AdaBoost	0.450	0.533	0.519	0.508	0.533	Neural Network	0.540	0.467	0.467	0.467	0.467	SGD	0.400	0.400	0.416	0.464	0.400																																																																																																																	
Model	AUC	CA	F1	Precision	Recall																																																																																																																																																													
kNN	0.500	0.600	0.568	0.556	0.600																																																																																																																																																													
Logistic Regression	0.460	0.467	0.481	0.512	0.467																																																																																																																																																													
Naive Bayes	0.400	0.333	0.167	0.111	0.333																																																																																																																																																													
Random Forest	0.600	0.667	0.533	0.444	0.667																																																																																																																																																													
AdaBoost	0.450	0.533	0.519	0.508	0.533																																																																																																																																																													
Neural Network	0.540	0.467	0.467	0.467	0.467																																																																																																																																																													
SGD	0.400	0.400	0.416	0.464	0.400																																																																																																																																																													

Pré-traitement des textes							Sac de mots																																																																																																																																																																																																																		
Mots de la liste (annexe E p86)				1-2			mode binaire																																																																																																																																																																																																																		
<div style="display: flex; justify-content: space-between;"> <div style="width: 15%;"> <p>Show probabilities for</p> <p>ACCEPTER REFUSER</p> </div> <table border="1" style="width: 80%; text-align: center;"> <thead> <tr> <th></th> <th>kNN</th> <th>Logistic Regression</th> <th>Naive Bayes</th> <th>Random Forest</th> <th>AdaBoost</th> <th>Neural Network</th> <th>SGD</th> <th>category</th> <th>name</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.40 : 0.60 → REFUSER</td><td>0.05 : 0.95 → REFUSER</td><td>0.01 : 0.99 → REFUSER</td><td>0.24 : 0.76 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-344-10 ...</td></tr> <tr><td>2</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.23 : 0.77 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.39 : 0.61 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.08 : 0.92 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-398-08</td></tr> <tr><td>3</td><td>0.80 : 0.20 → ACCEPTER</td><td>0.28 : 0.72 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.06 : 0.94 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-501-15</td></tr> <tr><td>4</td><td>0.80 : 0.20 → ACCEPTER</td><td>0.51 : 0.49 → ACCEPTER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.37 : 0.63 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.21 : 0.79 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-56-16</td></tr> <tr><td>5</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.23 : 0.77 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.37 : 0.63 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.19 : 0.81 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>ACCEPTER</td><td>c-578-17</td></tr> <tr><td>6</td><td>0.40 : 0.60 → REFUSER</td><td>0.31 : 0.69 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.30 : 0.70 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.07 : 0.93 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-124-18</td></tr> <tr><td>7</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.33 : 0.67 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.19 : 0.81 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.15 : 0.85 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-226-15</td></tr> <tr><td>8</td><td>0.40 : 0.60 → REFUSER</td><td>0.06 : 0.94 → REFUSER</td><td>0.02 : 0.98 → REFUSER</td><td>0.27 : 0.73 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.02 : 0.98 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-311-11</td></tr> <tr><td>9</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.31 : 0.69 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.33 : 0.67 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.29 : 0.71 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-421-15</td></tr> <tr><td>10</td><td>0.80 : 0.20 → ACCEPTER</td><td>0.26 : 0.74 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.12 : 0.88 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.26 : 0.74 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-482-15</td></tr> <tr><td>11</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.22 : 0.78 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.15 : 0.85 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.09 : 0.91 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-50-15</td></tr> <tr><td>12</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.46 : 0.54 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.42 : 0.58 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.23 : 0.77 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-542-07</td></tr> <tr><td>13</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.42 : 0.58 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.30 : 0.70 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.22 : 0.78 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-653-17</td></tr> <tr><td>14</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.23 : 0.77 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.39 : 0.61 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.04 : 0.96 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-85-16 c86-16</td></tr> <tr><td>15</td><td>0.60 : 0.40 → ACCEPTER</td><td>0.20 : 0.80 → REFUSER</td><td>1.00 : 0.00 → ACCEPTER</td><td>0.20 : 0.80 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>0.09 : 0.91 → REFUSER</td><td>0.00 : 1.00 → REFUSER</td><td>REFUSER</td><td>c-98-11</td></tr> </tbody> </table> </div> <div style="width: 15%; text-align: center;"> <table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th>Model</th> <th>AUC</th> <th>CA</th> <th>F1</th> <th>Precision</th> <th>Recall</th> </tr> </thead> <tbody> <tr><td>kNN</td><td>0.620</td><td>0.400</td><td>0.362</td><td>0.556</td><td>0.400</td></tr> <tr><td>Logistic Regression</td><td>0.460</td><td>0.733</td><td>0.667</td><td>0.810</td><td>0.733</td></tr> <tr><td>Naive Bayes</td><td>0.420</td><td>0.333</td><td>0.259</td><td>0.436</td><td>0.333</td></tr> <tr><td>Random Forest</td><td>0.630</td><td>0.667</td><td>0.533</td><td>0.444</td><td>0.667</td></tr> <tr><td>AdaBoost</td><td>0.500</td><td>0.600</td><td>0.568</td><td>0.556</td><td>0.600</td></tr> <tr><td>Neural Network</td><td>0.340</td><td>0.667</td><td>0.533</td><td>0.444</td><td>0.667</td></tr> <tr><td>SGD</td><td>0.500</td><td>0.667</td><td>0.533</td><td>0.444</td><td>0.667</td></tr> </tbody> </table> </div>											kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name	1	0.40 : 0.60 → REFUSER	0.05 : 0.95 → REFUSER	0.01 : 0.99 → REFUSER	0.24 : 0.76 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...	2	0.60 : 0.40 → ACCEPTER	0.23 : 0.77 → REFUSER	1.00 : 0.00 → ACCEPTER	0.39 : 0.61 → REFUSER	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-398-08	3	0.80 : 0.20 → ACCEPTER	0.28 : 0.72 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15	4	0.80 : 0.20 → ACCEPTER	0.51 : 0.49 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.37 : 0.63 → REFUSER	1.00 : 0.00 → ACCEPTER	0.21 : 0.79 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-56-16	5	0.60 : 0.40 → ACCEPTER	0.23 : 0.77 → REFUSER	1.00 : 0.00 → ACCEPTER	0.37 : 0.63 → REFUSER	0.00 : 1.00 → REFUSER	0.19 : 0.81 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-578-17	6	0.40 : 0.60 → REFUSER	0.31 : 0.69 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.07 : 0.93 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18	7	0.60 : 0.40 → ACCEPTER	0.33 : 0.67 → REFUSER	1.00 : 0.00 → ACCEPTER	0.19 : 0.81 → REFUSER	1.00 : 0.00 → ACCEPTER	0.15 : 0.85 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15	8	0.40 : 0.60 → REFUSER	0.06 : 0.94 → REFUSER	0.02 : 0.98 → REFUSER	0.27 : 0.73 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11	9	0.60 : 0.40 → ACCEPTER	0.31 : 0.69 → REFUSER	1.00 : 0.00 → ACCEPTER	0.33 : 0.67 → REFUSER	0.00 : 1.00 → REFUSER	0.29 : 0.71 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15	10	0.80 : 0.20 → ACCEPTER	0.26 : 0.74 → REFUSER	1.00 : 0.00 → ACCEPTER	0.12 : 0.88 → REFUSER	1.00 : 0.00 → ACCEPTER	0.26 : 0.74 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15	11	0.60 : 0.40 → ACCEPTER	0.22 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTER	0.15 : 0.85 → REFUSER	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15	12	0.60 : 0.40 → ACCEPTER	0.46 : 0.54 → REFUSER	1.00 : 0.00 → ACCEPTER	0.42 : 0.58 → REFUSER	0.00 : 1.00 → REFUSER	0.23 : 0.77 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-542-07	13	0.60 : 0.40 → ACCEPTER	0.42 : 0.58 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.22 : 0.78 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-653-17	14	0.60 : 0.40 → ACCEPTER	0.23 : 0.77 → REFUSER	1.00 : 0.00 → ACCEPTER	0.39 : 0.61 → REFUSER	0.00 : 1.00 → REFUSER	0.04 : 0.96 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16	15	0.60 : 0.40 → ACCEPTER	0.20 : 0.80 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11	Model	AUC	CA	F1	Precision	Recall	kNN	0.620	0.400	0.362	0.556	0.400	Logistic Regression	0.460	0.733	0.667	0.810	0.733	Naive Bayes	0.420	0.333	0.259	0.436	0.333	Random Forest	0.630	0.667	0.533	0.444	0.667	AdaBoost	0.500	0.600	0.568	0.556	0.600	Neural Network	0.340	0.667	0.533	0.444	0.667	SGD	0.500	0.667	0.533	0.444	0.667
	kNN	Logistic Regression	Naive Bayes	Random Forest	AdaBoost	Neural Network	SGD	category	name																																																																																																																																																																																																																
1	0.40 : 0.60 → REFUSER	0.05 : 0.95 → REFUSER	0.01 : 0.99 → REFUSER	0.24 : 0.76 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-344-10 ...																																																																																																																																																																																																																
2	0.60 : 0.40 → ACCEPTER	0.23 : 0.77 → REFUSER	1.00 : 0.00 → ACCEPTER	0.39 : 0.61 → REFUSER	0.00 : 1.00 → REFUSER	0.08 : 0.92 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-398-08																																																																																																																																																																																																																
3	0.80 : 0.20 → ACCEPTER	0.28 : 0.72 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.06 : 0.94 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-501-15																																																																																																																																																																																																																
4	0.80 : 0.20 → ACCEPTER	0.51 : 0.49 → ACCEPTER	1.00 : 0.00 → ACCEPTER	0.37 : 0.63 → REFUSER	1.00 : 0.00 → ACCEPTER	0.21 : 0.79 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-56-16																																																																																																																																																																																																																
5	0.60 : 0.40 → ACCEPTER	0.23 : 0.77 → REFUSER	1.00 : 0.00 → ACCEPTER	0.37 : 0.63 → REFUSER	0.00 : 1.00 → REFUSER	0.19 : 0.81 → REFUSER	0.00 : 1.00 → REFUSER	ACCEPTER	c-578-17																																																																																																																																																																																																																
6	0.40 : 0.60 → REFUSER	0.31 : 0.69 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.07 : 0.93 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-124-18																																																																																																																																																																																																																
7	0.60 : 0.40 → ACCEPTER	0.33 : 0.67 → REFUSER	1.00 : 0.00 → ACCEPTER	0.19 : 0.81 → REFUSER	1.00 : 0.00 → ACCEPTER	0.15 : 0.85 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-226-15																																																																																																																																																																																																																
8	0.40 : 0.60 → REFUSER	0.06 : 0.94 → REFUSER	0.02 : 0.98 → REFUSER	0.27 : 0.73 → REFUSER	0.00 : 1.00 → REFUSER	0.02 : 0.98 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-311-11																																																																																																																																																																																																																
9	0.60 : 0.40 → ACCEPTER	0.31 : 0.69 → REFUSER	1.00 : 0.00 → ACCEPTER	0.33 : 0.67 → REFUSER	0.00 : 1.00 → REFUSER	0.29 : 0.71 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-421-15																																																																																																																																																																																																																
10	0.80 : 0.20 → ACCEPTER	0.26 : 0.74 → REFUSER	1.00 : 0.00 → ACCEPTER	0.12 : 0.88 → REFUSER	1.00 : 0.00 → ACCEPTER	0.26 : 0.74 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-482-15																																																																																																																																																																																																																
11	0.60 : 0.40 → ACCEPTER	0.22 : 0.78 → REFUSER	1.00 : 0.00 → ACCEPTER	0.15 : 0.85 → REFUSER	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-50-15																																																																																																																																																																																																																
12	0.60 : 0.40 → ACCEPTER	0.46 : 0.54 → REFUSER	1.00 : 0.00 → ACCEPTER	0.42 : 0.58 → REFUSER	0.00 : 1.00 → REFUSER	0.23 : 0.77 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-542-07																																																																																																																																																																																																																
13	0.60 : 0.40 → ACCEPTER	0.42 : 0.58 → REFUSER	1.00 : 0.00 → ACCEPTER	0.30 : 0.70 → REFUSER	0.00 : 1.00 → REFUSER	0.22 : 0.78 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-653-17																																																																																																																																																																																																																
14	0.60 : 0.40 → ACCEPTER	0.23 : 0.77 → REFUSER	1.00 : 0.00 → ACCEPTER	0.39 : 0.61 → REFUSER	0.00 : 1.00 → REFUSER	0.04 : 0.96 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-85-16 c86-16																																																																																																																																																																																																																
15	0.60 : 0.40 → ACCEPTER	0.20 : 0.80 → REFUSER	1.00 : 0.00 → ACCEPTER	0.20 : 0.80 → REFUSER	0.00 : 1.00 → REFUSER	0.09 : 0.91 → REFUSER	0.00 : 1.00 → REFUSER	REFUSER	c-98-11																																																																																																																																																																																																																
Model	AUC	CA	F1	Precision	Recall																																																																																																																																																																																																																				
kNN	0.620	0.400	0.362	0.556	0.400																																																																																																																																																																																																																				
Logistic Regression	0.460	0.733	0.667	0.810	0.733																																																																																																																																																																																																																				
Naive Bayes	0.420	0.333	0.259	0.436	0.333																																																																																																																																																																																																																				
Random Forest	0.630	0.667	0.533	0.444	0.667																																																																																																																																																																																																																				
AdaBoost	0.500	0.600	0.568	0.556	0.600																																																																																																																																																																																																																				
Neural Network	0.340	0.667	0.533	0.444	0.667																																																																																																																																																																																																																				
SGD	0.500	0.667	0.533	0.444	0.667																																																																																																																																																																																																																				

En examinant tous les résultats fournis par le module de prédiction^{12 p34}. Nous constatons qu'aucun des modèles algorithmiques offrent une solution acceptable.

Nous pouvons rapidement conclure que le travail sur les textes bruts de jurisprudence ne fonctionne pas si nous désirons savoir si un signe peut-être enregistré comme une marque. Nous devons changer la méthode d'analyse.

La sous-section suivante analyse les résultats utilisant des données comprises entre 1 et -1. Ces données sont issues d'une analyse des règles du droit des marques et textes de jurisprudence du droit des marques. Ces données ont été compilées dans le fichier *valeurTest.csv*.

4.2.5 Analyse des résultats obtenus en utilisant les données reprises dans le fichier *valeurTest.csv* en utilisant le module de prédiction^{12 p34}

L'analyse portera sur le résultat obtenu sur base de données numériques et non plus sur des données transformées en données numériques au départ de texte brut grâce aux calculs des poids de chaque mot dans ce même texte.

Ces données numériques sont issues de l'analyse des textes de jurisprudence et des textes du droit des marques. Elles se présentent sous une forme de liste numérique ayant comme valeur -1, 0 et 1. Le schéma utilisé se trouve à la figure 4.13 à la page 41.

Nous débuterons en analysant les graphiques repris aux figures 4.22,4.23 et 4.24 obtenus avec les modules de classification hiérarchique^{20 p38}, de représentation graphique de l'échelle multidimensionnelle^{21 p38} et de "silhouette plot" dans le schéma repris à la figure 4.13 page 41.

En examinant les graphes, nous pouvons constater que les résultats sont meilleurs que dans les analyses précédentes pour le jeu de données "Acceptation".

Plus particulièrement, nous pouvons constater que :

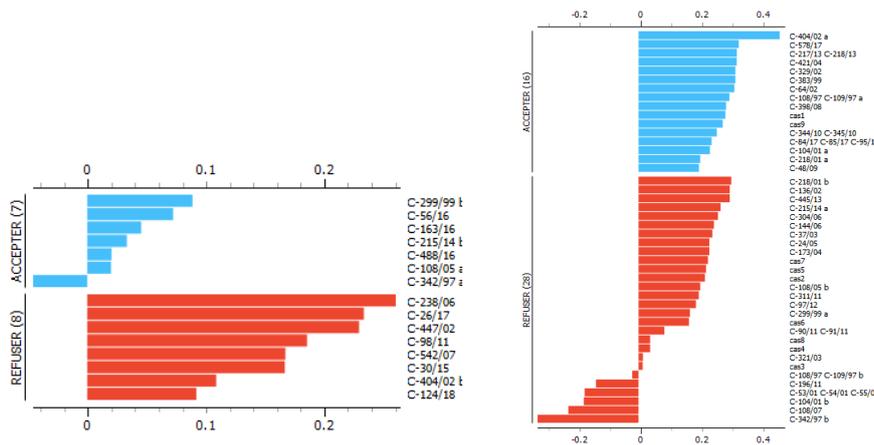


FIGURE 4.22 – Graphiques représentant la cohérence des groupes dans le jeu de données test (à gauche) et d'exemple (à droite) pour le schéma repris à la figure 4.13 à la page 4.13



FIGURE 4.23 – Graphiques représentant la position des points sur une échelle multidimensionnelle^{21 p38} des cas pour le jeu de données test (à gauche) et d'exemple (à droite) pour le schéma repris à la figure 4.13 à la page 4.13.

- dans les graphes de "silhouette plot", il existe une bonne cohérence dans les groupes des différents jeux de donnée. Pour le jeu de test seul un élément tend à se dissocier du groupe, il s'agit du cas C-342/97 a.
- dans les graphiques d'échelle multidimensionnelle^{21 p38}, les données de test sont bien regroupées ce qui paraît être visuellement moins le cas pour les données d'exemple, mais il s'agit peut être que d'un effet dimensionnel.
- dans les graphes de classification hiérarchique^{20 p38}, des groupes de cas se forment rapidement, nous pouvons remarquer le groupe C2 et C3 dans le jeu de test et le groupe C1, C3 et C4 dans le jeu d'exemple.

Nous pouvons déjà dire que les résultats du module de prédiction^{12 p34} devraient être bons, car nous voyons bien la formation de 2 groupes homogènes.

Passons à l'analyse des résultats fournis par le module de prédiction^{12 p34} sur le schéma 4.13 à la page 41.

Les algorithmes de prédiction des scores compris entre passable et excellent. Les 2 meilleurs

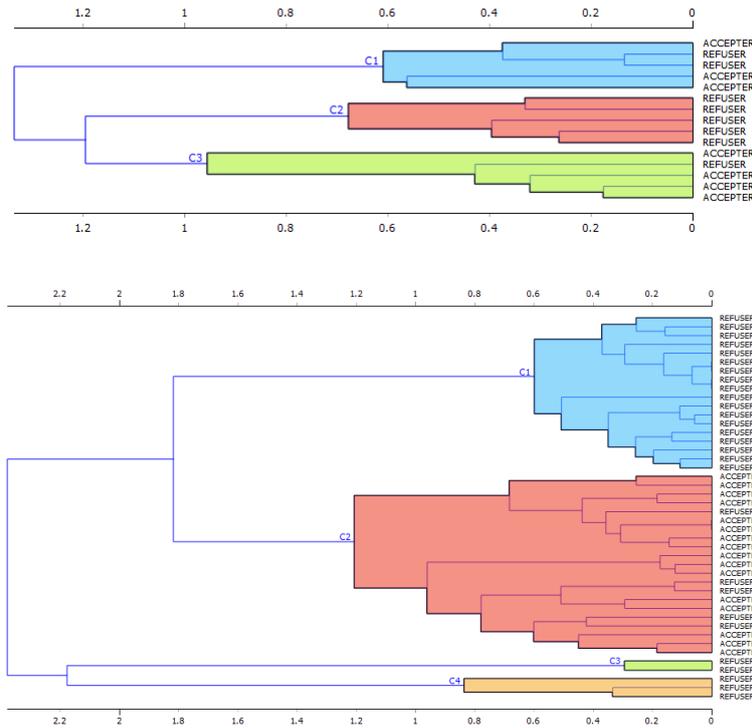


FIGURE 4.24 – Graphiques représentant la classification hiérarchique^{20 p38} des cas pour le jeu de données test (au-dessus) et d'exemple (en-dessous) pour le schéma repris à la figure 4.13 à la page 4.13.

	Logistic Regression	Random Forest	Neural Network	Naive Bayes	SVM	AdaBoost	kNN	Tree	Resultat	Jurisprudence
1	0.71 : 0.29 – ACCEPTER	0.53 : 0.47 – ACCEPTER	0.66 : 0.34 – ACCEPTER	0.84 : 0.16 – ACCEPTER	0.75 : 0.25 – ACCEPTER	1.00 : 0.00 – ACCEPTER	0.80 : 0.20 – ACCEPTER	1.00 : 0.00 – ACCEPTER	ACCEPTER	C-108/05 a
2	0.75 : 0.25 – ACCEPTER	0.70 : 0.30 – ACCEPTER	0.30 : 0.70 – REFUSER	1.00 : 0.00 – ACCEPTER	0.72 : 0.28 – ACCEPTER	1.00 : 0.00 – ACCEPTER	1.00 : 0.00 – ACCEPTER	1.00 : 0.00 – ACCEPTER	ACCEPTER	C-56/16
5	0.87 : 0.13 – ACCEPTER	0.67 : 0.33 – ACCEPTER	1.00 : 0.00 – ACCEPTER	1.00 : 0.00 – ACCEPTER	0.85 : 0.15 – ACCEPTER	1.00 : 0.00 – ACCEPTER	0.80 : 0.20 – ACCEPTER	1.00 : 0.00 – ACCEPTER	ACCEPTER	C-488/16
6	0.92 : 0.08 – ACCEPTER	0.88 : 0.12 – ACCEPTER	0.94 : 0.06 – ACCEPTER	0.99 : 0.01 – ACCEPTER	0.93 : 0.07 – ACCEPTER	1.00 : 0.00 – ACCEPTER	1.00 : 0.00 – ACCEPTER	1.00 : 0.00 – ACCEPTER	ACCEPTER	C-163/16
7	0.36 : 0.64 – REFUSER	0.20 : 0.80 – REFUSER	0.00 : 1.00 – REFUSER	0.01 : 0.99 – REFUSER	0.27 : 0.73 – REFUSER	1.00 : 0.00 – ACCEPTER	0.20 : 0.80 – REFUSER	1.00 : 0.00 – ACCEPTER	ACCEPTER	C-299/99 b
13	0.63 : 0.37 – ACCEPTER	0.69 : 0.31 – ACCEPTER	0.13 : 0.87 – REFUSER	0.70 : 0.30 – ACCEPTER	0.72 : 0.28 – ACCEPTER	1.00 : 0.00 – ACCEPTER	0.80 : 0.20 – ACCEPTER	1.00 : 0.00 – ACCEPTER	ACCEPTER	C-215/14 b
14	0.27 : 0.73 – REFUSER	0.32 : 0.68 – REFUSER	0.05 : 0.95 – REFUSER	0.03 : 0.97 – REFUSER	0.43 : 0.57 – REFUSER	0.00 : 1.00 – REFUSER	0.80 : 0.20 – ACCEPTER	0.00 : 1.00 – REFUSER	ACCEPTER	C-342/97 a
3	0.19 : 0.81 – REFUSER	0.62 : 0.38 – ACCEPTER	0.42 : 0.58 – REFUSER	0.19 : 0.81 – REFUSER	0.37 : 0.63 – REFUSER	0.00 : 1.00 – REFUSER	0.60 : 0.40 – ACCEPTER	0.00 : 1.00 – REFUSER	REFUSER	C-404/02 b
4	0.33 : 0.67 – REFUSER	0.12 : 0.88 – REFUSER	0.01 : 0.99 – REFUSER	0.14 : 0.86 – REFUSER	0.33 : 0.67 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	REFUSER	C-30/15
8	0.02 : 0.98 – REFUSER	0.07 : 0.93 – REFUSER	0.01 : 0.99 – REFUSER	0.00 : 1.00 – REFUSER	0.14 : 0.86 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	REFUSER	C-238/06
9	0.04 : 0.96 – REFUSER	0.00 : 1.00 – REFUSER	0.01 : 0.99 – REFUSER	0.00 : 1.00 – REFUSER	0.08 : 0.92 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	REFUSER	C-98/11
10	0.01 : 0.99 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	0.30 : 0.70 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	REFUSER	C-26/17
11	0.34 : 0.66 – REFUSER	0.32 : 0.68 – REFUSER	0.05 : 0.95 – REFUSER	0.70 : 0.30 – ACCEPTER	0.44 : 0.56 – REFUSER	0.00 : 1.00 – REFUSER	0.40 : 0.60 – REFUSER	0.50 : 0.50 – ACCEPTER	REFUSER	C-124/18
12	0.06 : 0.94 – REFUSER	0.52 : 0.48 – ACCEPTER	0.02 : 0.98 – REFUSER	0.15 : 0.85 – REFUSER	0.25 : 0.75 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	REFUSER	C-542/07
15	0.13 : 0.87 – REFUSER	0.87 : 0.13 – ACCEPTER	0.10 : 0.90 – REFUSER	0.64 : 0.36 – ACCEPTER	0.71 : 0.29 – ACCEPTER	1.00 : 0.00 – ACCEPTER	0.00 : 1.00 – REFUSER	0.00 : 1.00 – REFUSER	REFUSER	C-547/02

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.964	0.867	0.863	0.893	0.867
Random Forest	0.768	0.667	0.667	0.673	0.667
Neural Network	0.804	0.733	0.707	0.822	0.733
Naive Bayes	0.821	0.733	0.733	0.733	0.733
SVM	0.875	0.800	0.798	0.804	0.800
AdaBoost	0.866	0.867	0.867	0.867	0.867
kNN	0.964	0.867	0.867	0.867	0.867
Tree	0.920	0.867	0.867	0.867	0.867

FIGURE 4.25 – Résultat du schéma repris à la figure 4.13 à la page 41 en utilisant le jeu de données numérique.

modèles étant le modèle de régression logique^{3 p31} et le modèle de boosting adaptatif^{1 p31}. L'algorithme qui n'obtient qu'un score de passable est le modèle des forêts aléatoires^{6 p32}.

Pour poursuivre l'analyse, prenons le modèle algorithmique des k plus proches voisins^{2 p31} pour analyser les erreurs de prédiction sur les cas de jurisprudence. Le cas C-299/99 b est en erreur dans tous les modèles algorithmiques sauf dans le modèle des arbres de décision^{8 p33}. Il s'agit d'une demande de marque de forme utilisant un signe descriptif du produit signifiant qu'il devrait être refusé, sauf si la forme du produit est protégé par un brevet par exemple. Le cas C-299/99 b est en erreur car il n'y avait pas d'autres cas dans le jeu de données d'entraînement des algorithmes.

Pour le cas C-404/02 b, il est en erreur que dans le modèle des k plus proches voisins^{2 p31} et dans le modèle des forêts aléatoires^{6 p32}. Dans ce cas, nous sommes dans l'exemple d'une demande d'enregistrement d'une marque constituée d'un patronyme répandu sans avoir de caractère distinctif. Ici l'erreur vient probablement que la liste des valeurs du cas est très semblable aux valeurs du cas C-404/02 a qui avait comme valeur finale ACCEPTER.

Pour le cas C-342/97, il est en erreur dans tous les modèles algorithmiques sauf dans le modèles des k plus proches voisins^{2 p31} mais, en examinant la figure 4.22, il est le seul dans le jeu de test à ne pas être en cohérence avec son groupe.

Les autres cas en erreur découlent des mêmes conclusions que les cas repris ci-dessus.

4.3 Conclusion

Concluons ce chapitre en disant que l'utilisation du programme Orange [Borondics et al., 2020] a grandement facilité le travail sur les différents tests réalisés dans le cadre du travail sur la justice prédictive appliqué au droit des marques.

Les textes utilisés pour les tests sont des arrêts de jurisprudence liés aux droits des marques. Dans un premier temps, nous nous posons la question de savoir si nous pouvions prédire le résultat d'un arrêt de jurisprudence avec la distance cosinus calculée entre les textes et les modèles algorithmiques fournis par le logiciel Orange [Borondics et al., 2020]. L'analyse des résultats nous a prouvé qu'il est possible de prédire le résultat d'un pourvoi en appliquant les algorithmes fournis par le logiciel Orange [Borondics et al., 2020].

Par la suite, avec ces mêmes outils et en reclassant les différents arrêts en fonction de l'acceptation ou du refus de la marque en question, nous recherchions à prédire si une marque pouvait être enregistrée conformément au droit des marques. Les résultats obtenus nous ont indiqué qu'il n'était pas possible de prédire l'acceptation d'une marque en se basant sur les textes bruts.

La solution trouvée consistant en la réalisation d'un tableau reprenant pour chaque arrêt une liste de valeur comprise entre -1 et 1 qui correspond à des réponses aux questions qui tentent de cerner les cas où un signe a été accepté en tant que marque ou non. Les résultats obtenus avec les algorithmes du logiciel Orange [Borondics et al., 2020] ont démontré que l'on pouvait avec cette méthode dire qu'un signe pouvait être enregistré en tant que marque.

Même si tous les tests n'ont pas aboutis aux réponses souhaitées, ils ont permis d'y voir plus clair dans ce qu'il est possible de faire ou non avec les textes de jurisprudence du domaine du droit des marques. Nous pouvons dire entre autre que les tests sur le jeu de données "Conclusion" permettent de vérifier la cohérence des conclusions entre les textes et avec un peu plus de

travail d'élaboration, nous pourrions réaliser une prédiction sur la conclusion d'un pourvoi futur.

Les principaux outils utilisés sont les outils fournis par le logiciel Orange [Borondics et al., 2020]. Le site de *Wikipedia*, nous a permis de comprendre en grande partie les théories que nous n'avions pas encore acquises lors de notre cursus scolaire.

Le chapitre suivant apportera une nouvelle méthode pour prédire l'acceptation d'un signe en tant que marque. Nous passerons en revue la théorie sur la programmation logique avant de la mettre en application en codant un programme logique de manière déductive.

Chapitre 5

La programmation logique

Dans le chapitre 4, nous avons travaillé sur l'analyse des textes pour trouver des similarités entre eux. Dans ce chapitre, nous abordons une autre approche pour robotiser le droit des marques, nous allons créer un programme logique.

Dans un premier temps nous définirons et expliquerons la programmation logique. Ensuite, nous élaborerons le programme logique de manière déductive sous forme de bot. Le bot qui, via une série de question, tentera de répondre à la question final *"Un signe pourra-t-il être accepté en tant que marque en regard du droit des marques ?"*.

5.1 Théorie sur la programmation logique

"La programmation logique est une forme de programmation qui définit les applications à l'aide d'un ensemble de faits élémentaires les concernant et de règles de logique leur associant des conséquences plus ou moins directes. Ces faits et ces règles sont exploités par un démonstrateur de théorème ou moteur d'inférence, en réaction à une question ou requête" [Wikipédia, 2020].

Le langage Prolog, développé par Colmerauer et Roussel en 1972, utilise la programmation logique[Wikipédia, 2020].

Le logiciel "SWI-Prolog version 7.6.4" de Jan Wielemaker (jan@swi-prolog.org) sera utilisé pour exécuter des programmes Prolog, en particulier pour exécuter le programme *"marque.pl"*, réalisé dans le cadre du projet de prédire l'acceptation d'une demande de marque.

Prenons la figure 5.1 ci-dessous. C'est un exemple simple de généalogie. En généalogie, on sait qu'un enfant a un père et une mère.

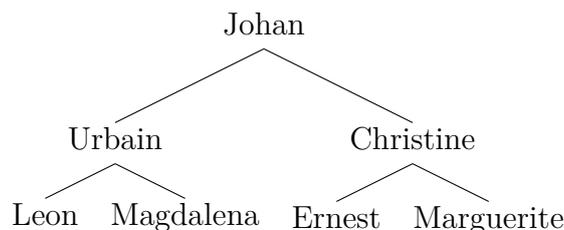


FIGURE 5.1 – Généalogie reprenant mes parents et mes grand-parents

Donc on peut dire que :

père(Urbain, Johan)
mère(Christine, Johan)
père(Leon, Urbain)
mère(Magdalena, Urbain)
père(Ernest, Christine)
mère(Marguerite, Christine)

qui se lit :

Urbain est le père de Johan
Christine est la mère de Johan
Leon est le père de Urbain
Magdalena est la mère de Urbain
Ernest est le père de Christine
Marguerite est la mère de Christine

on en déduit les 2 prédicats simples suivants

père(X, Y)
mère(X, Y)

ce qui est équivalent à dire que "*X* est le père de *Y*" ou "*X* est la mère de *Y*".

Sur base des 2 prédicats précédents et de l'arbre généalogique repris à la figure 5.1 à la page 58, nous pouvons facilement déduire 4 nouvelles règles de base

grandpère(X, Y) ← père(X, Z), père(Z, Y)
grandpère(X, Y) ← père(X, Z), mère(Z, Y)
grandmère(X, Y) ← mère(X, Z), père(Z, Y)
grandmère(X, Y) ← mère(X, Z), mère(Z, Y)

qui se lit

X est le grand père de *Y* si *X* est le père de *Z* et *Z* est le père ou la mère de *Y*

ou

X est la grand mère de *Y* si *X* est la mère de *Z* et *Z* est le père ou la mère de *Y*.

Après avoir vu les bases de la programmation logique, nous allons, dans la section suivante, comparer la programmation logique déductive à la programmation logique inductive.

5.2 La programmation logique inductive

La programmation logique inductive est

- *un ensemble de techniques permettant de construire des programmes logiques à partir d'exemples*¹ [Evans and Grefenstette, 2018], selon "Learning Explanatory Rules from Noisy Data" (de Richard Evans et Edward Grefenstette).
- *l'intersection de l'apprentissage inductif et de la programmation logique. Ainsi, ILP utilise des techniques d'apprentissage automatique et de programmation logique*² [Muggleton and Raedt, 1994], selon "INDUCTIVE LOGIC PROGRAMMING : THEORY AND METHODS" (de Stephen Muggleton et Luc De Raedt).

La théorie de l'apprentissage automatique inductive a été développée par Gordon Plotkin et Ehud Shapiro [Wikipédia, 2019d].

Muggleton utilisait en 1991 le terme de programmation logique inductive dans l'article "Inductive logic programming", *New Generation Computing*, vol. 8, no 4, 1991, p. 295–318 [Wikipédia, 2019d].

En résumé, nous pouvons dire qu'un système de programmation logique inductive va au départ de règle de base, d'exemples positifs et négatifs et en se basant sur la programmation logique déduire un "*programme logique hypothétique qui confirme les exemples positifs et infirme les exemples négatifs*" [Wikipédia, 2019d].

Sur base d'un exemple de généalogie tiré de [Muggleton and Raedt, 1994] à la page 631, appliquons la théorie à la figure 5.1 à la page 58.

Nous avons, dès lors, les relations suivantes

père(Urbain, Johan)
mère(Christine, Johan)
père(Leon, Urbain)
mère(Magdalena, Urbain)
père(Ernest, Christine)
mère(Marguerite, Christine)

Auxquelles on rajoute 2 règles de base

grandpère(X, Y) ← père(X, Z), parent(Z, Y)
grandmère(X, Y) ← mère(X, Z), parent(Z, Y)

Comme en programmation logique inductive, il nous faut des exemples positifs et négatifs. Prenons les exemples positifs (tiré de la figure 5.1 à la page 58)

grandpère(Leon, Johan)
grandpère(Ernest, Johan)
grandmère(Magdalena, Johan)
grandmère(Marguerite, Johan)

1. traduction de "is a collection of techniques for constructing logic programs from examples" avec google translate

2. traduction de "as the intersection of inductive learning and logic programming. Thus, ILP employs techniques from both machine learning and logic programming" avec google translate

et les exemples négatifs

```
grandpère(Magdalena, Johan)
grandpère(Johan, Ernest)
grandmère(Leon, Johan)
grandmère(Johan, Marguerite)
```

L'hypothèse que nous trouvons en appliquant la programmation logique inductive est

```
parent(X, Y) ← père(X, Y)
parent(X, Y) ← mère(X, Y)
```

Ce qui peut paraître facile au vu de l'exemple mais qui se complexifie rapidement en fonction des données qui lui sont transmises.

La complexité de la programmation logique inductive réside dans la création des règles de bases. Alors qu'en programmation logique déductive, nous connaissons, grâce au exemple en notre possession, la réponse que nous désirons trouver, il ne nous reste plus qu'à effectuer le codage pour y arriver. C'est pour cette raison que nous avons réalisé le programme en programmation logique déductive.

Dans la section suivante, nous produirons un programme logique utilisant la déduction pour répondre à la question "*Un signe pourra-t-il être accepté en tant que marque en regard du droit des marques ?*" en se basant sur les cas de jurisprudence comme exemple.

5.3 Mise en pratique de la théorie de la programmation logique

Le programme *marque.pl* utilise le programme "*Legrandgousier*" du Professeur Jean-Marie Jacquet, fournit comme travail dans le cadre du cours de technique d'intelligence artificielle à ses étudiants pour l'année académique 2018-2019, comme base pour la transformation des entrées et des sorties d'un format texte vers un format de liste et inversement.

En se basant sur la théorie du droit des marques repris à la page 18, nous déduisons qu'une marque est licite si elle respecte les points suivants :

- une marque doit être reconnue par un public pertinent, le public pertinent est déterminé en fonction de la classe de produit ou du service demandé et de la taille du marché visé.
- une marque ne doit pas être contraire à l'intérêt général. Cela signifie que l'on ne peut pas prendre un signe qui est utilisé habituellement dans la zone géographique et pour les classes de produit ou de service de la demande.
- une marque n'est pas une description du produit ou du service ou du contenant du produit.
- une marque ne possède donc pas de caractère descriptif mais doit posséder un caractère distinctif, qui la distingue des autres produits de la concurrence dans les classes de produits ou de services ou dans la zone géographique de la demande.

En programmation logique, on aura les prédicats suivants :

- Un public pertinent

$$publicPertinent(\mathbf{Public}) \begin{cases} \text{vrai, si le } \mathbf{Public} \text{ reconnaît la marque} \\ \text{faux, sinon} \end{cases}$$

- L'intérêt général

$$interetGeneral(\mathbf{Interet}) \begin{cases} \text{vrai, si la marque n'est pas contraire à l'Interet général} \\ \text{faux, sinon} \end{cases}$$

- La marque n'est pas descriptive du produit ou du service

$$caractereDescriptif(\mathbf{CaractDescript}) \begin{cases} \text{vrai, si la marque n'est pas descriptive} \\ \text{ou si } \mathbf{CaractDescript} \text{ est faux} \\ \text{faux, sinon} \end{cases}$$

- La marque a un caractère distinctif pour la classe de produits et de services

$$caractereDistinctif(\mathbf{CaractDistinct}) \begin{cases} \text{vrai, si la marque a un } \mathbf{CaractDistinct} \\ \text{faux, sinon} \end{cases}$$

Ceux-ci ne sont que des prédicats de base.

En reprenant la liste des arrêts de jurisprudence repris à l'annexe F à la page 93 et dans lequel nous avons sélectionné un certain nombre d'affaires représentant la plupart des cas rencontrés, nous obtenons une liste d'arrêts que nous allons utiliser pour mettre en évidence des cas particuliers ou des cas généraux qui nous serviront à réaliser le programme.

Réalisons une liste de questions qui nous permettra de classer les différents cas. Voici quelques questions qui s'appliquent à toutes les affaires étudiées.

- Quel est le format de la marque ?
- Quels sont les pays concernés dans la demande ?
- Quels sont les classes de produits et de services de la demande ?
- Un public reconnaît-il la marque ?
- N'y a-t-il pas de risques de conflits avec l'intérêt général ?
- Le signe est un caractère descriptif du produit ou du service ?
- La marque est-elle suffisamment décrite ?

L'annexe L à la page 116 reprend toutes les questions qui ont servi à créer le programme logique permettant de répondre à la question principale, à savoir "Un signe pourra-t-il être accepté en temps que marque en regard du droit des marques ?". Elle reprend aussi pour chaque affaire la réponse aux questions et la réponse finale que l'on devrait obtenir. Les réponses aux questions serviront d'argument pour les prédicats comme pour la question "Un public reconnaît-il la marque ?", la réponse est soit oui, ou soit non ce qui nous donne le prédicat suivant

$$\begin{aligned} publicPertinent("oui") &= true \\ publicPertinent("non") &= false \end{aligned}$$

Nous allons commencer à programmer en utilisant les questions générales. La programmation est réalisée de manière déductive, après avoir encodé les prédicats de base, nous effectuerons des tests utilisateurs³ ou des tests utilisant un fichier *test.csv*. Par la suite, nous recommencerons la procédure en rajoutant de nouvelles règles au programme et en vérifiant qu'il répond toujours correctement aux différentes questions. L'idée est d'arriver par raffinement à un programme qui donne le résultat final souhaité.

Une des questions était "La marque est-elle suffisamment décrite?". Une marque doit être correctement décrite pour pouvoir déterminer si elle possède un caractère distinctif. L'exemple C-124/18, annexe F à la page 93, nous confirme que si nous décrivons pas suffisamment le signe, elle sera refusée. Dans cette affaire, le signe était très évasif quant à la position des couleurs et à leur proportion dans le signe. Ce qui aurait pu être utilisé pour empêcher d'autres producteurs d'un produit similaire à utiliser ces couleurs. Ce qui nous donne le prédicat suivant.

$$\text{marqueDecrit}(\mathbf{Decrit}) \begin{cases} \text{vrai, si la marque est correctement décrite}(\mathbf{Decrit} \text{ donné en argument}) \\ \text{faux, sinon} \end{cases}$$

Mais pour une marque de couleur, il faut être plus complet en lui rajoutant un code d'identification des couleurs utilisées. Prenons comme exemple, le cas C-104/01, annexe F à la page 93. Dans cette affaire, le signe déposé était la couleur "orange" sans indication d'un code d'identification. Si le signe venait à être accepté, l'entreprise possédant la marque aurait pu faire interdire toutes les marques utilisant de l'orange clair ou une nuance d'orange.

Pour le cas suivant prenons le risque de conflit avec une marque ayant une renommée. Pour rappel, l'utilisation d'un signe déjà en circulation est autorisé, pour autant qu'il soit utilisé dans une classe de produit ou de service différente ou sur une zone territoriale différente que la marque existante.

Mais cette règle ne s'applique pas lorsque la marque existante a une renommée (voir le cas de Jaguar contre Jaguar, voir 3.2 à la page 23).

$$\text{marqueRenommee}(\text{Renommee}) \begin{cases} \text{vrai, si la marque n'est pas semblable à une marque} \\ \text{ayant une renommée} \\ \text{faux, sinon} \end{cases}$$

Maintenant, prenons le cas des marques ayant un caractère descriptif. La règle veut que celles-ci soient toujours refusées. Sauf si elles ont acquis un caractère distinctif par l'usage au moment de la demande d'enregistrement de la marque, ou encore que la marque, même si elle possède un caractère descriptif du produit ou du service de la demande, est reconnu par les marchés comme appartenant à une entreprise précise. Les affaires C-215/14, C-108/05 ou C-64/02, annexe F à la page 93, illustrent le caractère distinctif acquis à l'usage. Dans le cas C-64/02, le signe "DAS PRINZIP DER BEGUEMLICHKEIT" est descriptif du service fournit par la marque mais a acquis un caractère distinctif à l'usage au moment de la demande.

$$\text{caractDistUsage}(\text{CaractDistUsage}) \begin{cases} \text{vrai, si la marque a acquis} \\ \text{un caractère distinctif à l'usage} \\ \text{faux, sinon} \end{cases}$$

Mais l'acquisition d'un caractère distinctif par l'usage ne pourra pas s'appliquer aux marques tridimensionnelles ou figuratives.

3. ou user case, ces tests sont une simulation d'utilisation par l'utilisateur final

Nous passons maintenant aux cas s'appliquant exclusivement aux marques verbales ou semi-figuratives.

Pour les marques verbales et semi-figuratives, nous distinguons plusieurs cas utilisant soit un terme géographique, soit un patronyme, soit un slogan publicitaire, soit une abréviation, soit une marque qui est semblable de manière auditive à une autre marque de la même classe de produit ou de service.

- Pour les marques utilisant un terme géographique, nous pouvons utiliser un arbre de décisions (voir la figure 5.2) pour afficher les possibilités que l'on rencontre.

Pour résumer l'arbre de la figure 5.2, nous pouvons dire que

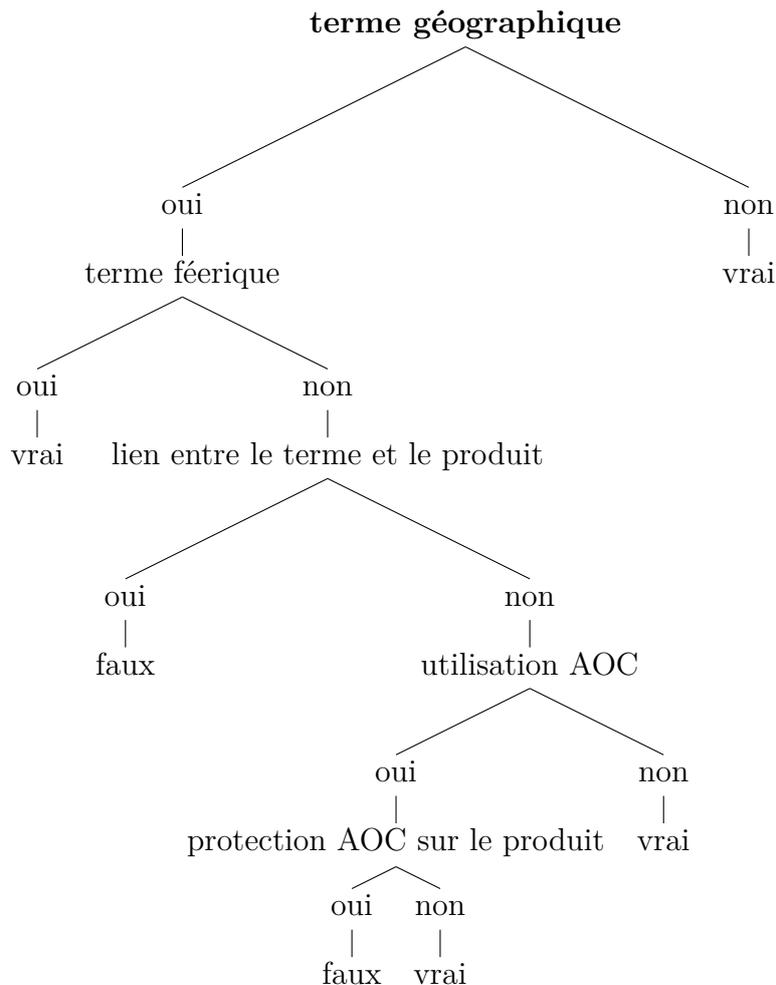


FIGURE 5.2 – Arbre de décision pour une marque ayant un terme géographique

- * une marque ayant un terme géographique féerique sera acceptée (exemple C-488/16, annexe F à la page 93).
- * une marque ayant un terme géographique pour lequel la concurrence reconnaît un lien avec les produits sera refusé (exemple C-108/97, annexe F à la page 93).
- * une marque utilisant une appellation d'origine contrôlée mais pour un produit qui n'est pas repris par l'appellation d'origine contrôlée sera acceptée (exemple C-402/02, annexe F à la page 93) et inversement.

- L'acceptation d'une marque utilisant un patronyme dépendra de la fréquence du patronyme et de la taille du marché visée dans la zone territoriale de référence.
Pour résumer l'arbre de la figure 5.3, nous pouvons dire qu'une marque utilisant un pa-

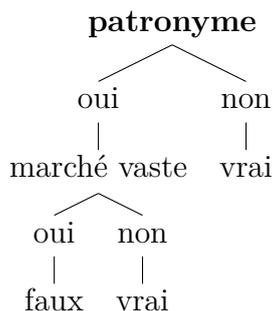


FIGURE 5.3 – Arbre de décision pour un terme patronymique

tronyme sera refusé si le patronyme est répandu et que le public ne pourra pas reconnaître le produit dans un marché trop vaste (exemple C-56/16, annexe F à la page 93).

- Une marque qui est semblable à une autre marque existante prend un risque d'être refusée proportionnellement à la résonance sur le marché de la marque antérieure (exemple C-342/97, annexe F à la page 93).
- Quand on utilise un slogan publicitaire comme marque, il faut prendre en compte le caractère distinctif de l'ensemble du slogan et pas juste l'accumulation des caractères distinctifs de chaque mot (exemple C-311/11, C-398/08 et 64/02, annexe F à la page 93).
- L'utilisation d'abréviations comme signe n'induit pas une reconnaissance automatique du caractère distinctif, il faut déterminer le caractère distinctif de l'ensemble des mots derrière l'abréviation (exemple C-90/11, annexe F à la page 93).

Maintenant, nous analysons le cas particulier des marques tridimensionnelles, qui en règle générale ne peuvent acquérir un caractère distinctif par l'usage. Néanmoins il existe certaine exception à cette règle.

Pour résumer l'arbre de la figure 5.4, nous pouvons dire que :

- * une marque tridimensionnelle ayant un caractère descriptif, mais dont le signe diverge de la description habituelle des produits des classes de produit de la demande, pourrait avoir un caractère distinctif (exemple C-344/10, annexe F à la page 93).
- * par contre, une marque tridimensionnelle, ayant un caractère descriptif donc la forme de la marque est, ou pourrait être, utilisée par la concurrence pour produire ou pour décrire un produit ou un service dans la même classe de produit ou de service, ne pourra pas avoir de caractère distinctif (exemple C-136/02, C-238/06, C-30/15, ..., annexe F à la page 93) sauf si le signe en question est protégé par un brevet de conception (exemple C-299/99, annexe F à la page 93).

marque tridimensionnelle avec un caractère descriptif

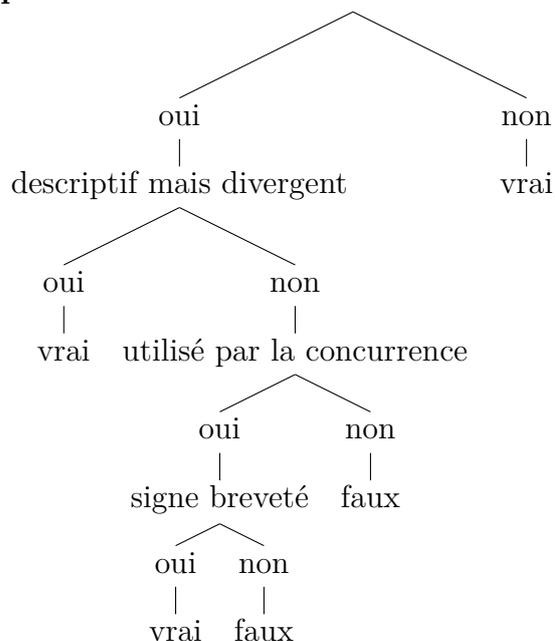


FIGURE 5.4 – Arbre de décision d'une marque tridimensionnelle ayant un caractère descriptif

En ajoutant les nouveaux prédicats aux prédicats de base, nous obtenons un programme qui devrait accepter les cas positifs et refuser les cas négatifs.

Voici quelques cas positifs repris dans l'annexe L à la page 116.

- + C-64/02, "*das prinzip der bequemlichkeit*", marque verbale contenant un slogan avec un caractère distinctif.
- + C-56/16, "*Port Charlotte*", marque verbale pour un whisky contenant un terme géographique avec une appellation d'origine contrôlée pour du porto.
- + C-421/04, "*Matratzen*", marque verbale pour l'Espagne mais ayant un caractère descriptif en allemand.
- + C-163/16, marque complexe de chaussures à haut talon avec une semelle rouge ayant un caractère distinctif acquis à l'usage.
- + C-383/99, "*Baby-Dry*", marque verbale ayant un caractère distinctif acquis par l'usage.

Et voici quelques cas négatifs repris dans l'annexe L à la page 116.

- C-98/11, marque figurative ayant un caractère descriptif.
- C-542/07, "*Pure Digital*", marque verbale ayant un caractère descriptif mais sans avoir acquis de caractère distinctif à l'usage.
- C-37/03, "*BioID*", marque verbale ayant un caractère descriptif mais sans avoir acquis de caractère distinctif à l'usage.
- C-321/03, marque tridimensionnelle d'aspirateur ayant un caractère descriptif du produit.
- C-238/06, marque tridimensionnelle représentant le contenant du produit représenté par la marque.

Le programme est repris dans le fichier *marque.pl*. Le langage utilisé est le Prolog. Un programme qui, sous forme de questions-réponses, conduit l'utilisateur vers une prédiction de l'acceptation ou du refus de l'enregistrement d'un signe en tant que marque. La figure 5.5 reprend le morceau du code qui sert de squelette pour l'importation et le test d'un fichier *test.csv* du programme *marque.pl*.

```

marque_test:-
    importation(L),
    length(L,T),
    produire_question(Q),
    lire_fichier_test(L,Q,T,[],L1),
    length(L1,T),nl,nl,nl,
    write_tab_result(L1,0,T).

lire_fichier_test(_,_,0,L,L).

lire_fichier_test([A|L],[Q1,Q2,Q3,Q4,Q5,Q6,Q8,Q9],T,L1,L3):-
    valeurReference(A,Ref),
    valeurNom(A,Nom),
    valeurType(A,Type),
    choixMarque(Type),
    produire_question_spec(Type,[Q7,Q10,Q16,Q26]),
    valeurPays(A,Pays),
    langueMarque(A,[Nom],Type,Lm,Lp),
    valeurClasse(A,C),
    read_list(C,Classe),
    coherentClasseNice(Classe),
    valeurDescription(A,Decrit),
    marqueDecrit(Decrit,[],T1),
    evalTypeVerb(A,Type,T1,T1a),
    valeurReconPublic(A,Public),
    publicPertinent(Public,T1a,T2),
    valeurRisqueConflit(A,Interet),
    interetGeneral(Interet,T2,T3),
    valeurRenommee(A,Renommee),
    marqueRenommee(Renommee,T3,T6),
    couleurOrginal(A,Type,Q26,Original),
    marqueCouleur(Original,T6,T7),
    slogan(A,Type,Q16,Q26,Slogan,CaractDiscEns),
    marquePub(Slogan,CaractDiscEns,T7,T8),
    abreviation(A,Type,T8,T11,SigneAbrDist),
    valeurClasseDistinct(A,D),
    read_list(D,DistClasse),
    length(DistClasse,Taille1),
    classeProdServTest(Type,DistClasse,CaractDiscEns,
        SigneAbrDist,Taille1,0,Distinctif),
    valeurCaractDescrip(A,CaractDesc),
    caractereDescriptif(Distinctif,CaractDesc,Type,A,Lm,Lp,T11,T13),
    V is T-1,

```

```

valeurResult (A, ResultReel) ,
resultat (T13, ResultObt) ,
tabResultat ([ Ref, ResultObt, ResultReel ], L1, L2) ,
lire_fichier_test (L, [ Q1, Q2, Q3, Q4, Q5, Q6, Q8, Q9 ], V, L2, L3) .

```

FIGURE 5.5 – Portion de code du programme *marque.pl* réalisant le test sur le fichier *valeur-Test.csv* (repris à l'annexe L à la page 116), le langage utilisé est le Prolog.

Le morceau de code importe le fichier *valeurTest.csv* qui est repris à l'annexe L à la page 116 et teste si le résultat obtenu correspond au résultat attendu. Le résultat fourni par SWI-Prolog est repris à la figure 5.6.

Expliquons les prédicats de ce bout de code. Le prédicat *"importation(Liste)"* va importer le fichier *valeur_Test.csv* est le convertir en une liste.

Le prédicat *"produire_question(Question)"* produit la liste de questions qui serviront à interroger l'utilisateur ou le fichier *valeurTest.csv*.

Le fonctionnement du prédicat *lire_fichier_test(L, Q, T, [], L1)* suit la même logique, pour chaque ligne de test, il récupère une réponse à une question de la liste des valeurs et évalue cette réponse. Ensuite, il passe à la question suivante, après avoir passé en revue toutes les questions, il enregistre la réponse finale et recommence pour le test suivant.

A la fin du fichier test, il renvoie la liste des réponses finale et l'affiche via le prédicat *"write_tab_result(Liste finale, 0, Taille de la liste finale)"*. Le résultat obtenu est repris à la figure 5.6. Dans le prédicat *"lire_fichier_test(Liste, Question, Longueur de la liste, [], Liste avec les valeurs finales)"*, nous trouvons les prédicats commençant par *valeur...(Liste, Réponse)* qui interrogent la liste avec les valeurs de tests et récupèrent la valeur désirée pour le premier test.

Le prédicat *"choixMarque(Type)"* vérifie que le type de la marque est conforme aux types possibles pour une marque. Le prédicat *"evalTypeVerb(Liste, Type, Liste des réponses précédentes, Liste des réponses précédentes avec la nouvelle réponse)"* va évaluer la marque si celle-ci est de type "verbale" ou "semi-figurative". L'évaluation porte sur les marques utilisant un terme géographique ou un patronyme ou un slogan publicitaire ou des abréviations.

Le prédicat *caractereDescriptif(Caractère distinctif de la marque, Réponse à la question sur le caractère descriptif, Type, Liste, Langue utilisé pour la marque, Langues utilisés dans les zones géographiques de la demande, Liste des réponses précédentes, Liste des réponses précédentes avec la nouvelle réponse)"* évalue si une marque qui n'a pas de caractère distinctif possède un caractère descriptif acquis par l'usage en fonction des règles qui peuvent être différentes pour les différents types marques.

Nous testons le programme *marque.pl* avec *SWI-Prolog* avec une liste de 5 arrêts de jurisprudence qui n'ont pas servi à réaliser le programme. Nous avons répondu pour chaque arrêts à la même liste de questions que pour la réalisation du programme (voir la figure 5.7). Le fichier test obtenu a été importé dans le programme *marque.pl* et avec *SWI-Prolog* Nous avons obtenu le résultat repris à la figure 5.8.

```

Resultat pour C-99/18 est correcte : ACCEPTER
Resultat pour C-43/15 est correcte : REFUSER
Resultat pour C-328/18 est incorrect : ACCEPTER/REFUSER
À chercher
Resultat pour C-281/18 est correcte : REFUSER
Resultat pour C-155/18 est correcte : ACCEPTER
Total : 4/5
true ■

```

```

Resultat pour C-98/11 est correcte : REFUSER
Resultat pour C-97/12 est correcte : REFUSER
Resultat pour C-90/11 est correcte : REFUSER
Resultat pour C-64/02 est correcte : ACCEPTER
Resultat pour C-56/16 est correcte : ACCEPTER
Resultat pour C-542/07 est correcte : REFUSER
Resultat pour C-53/01 est correcte : REFUSER
Resultat pour C-488/16 est correcte : ACCEPTER
Resultat pour C-447/02 est correcte : REFUSER
Resultat pour C-445/13 est correcte : REFUSER
Resultat pour C-421/04 est correcte : ACCEPTER
Resultat pour C-404/02 b est correcte : REFUSER
Resultat pour C-404/02 a est correcte : ACCEPTER
Resultat pour C-398/08 est correcte : ACCEPTER
Resultat pour C-383/99 est correcte : ACCEPTER
Resultat pour C-37/03 est correcte : REFUSER
Resultat pour C-344/10 est correcte : ACCEPTER
Resultat pour C-342/97 b est correcte : REFUSER
Resultat pour C-342/97 a est correcte : ACCEPTER
Resultat pour C-329/02 est correcte : ACCEPTER
Resultat pour C-321/03 est correcte : REFUSER
Resultat pour C-311/11 est correcte : REFUSER
Resultat pour C-304/06 est correcte : REFUSER
Resultat pour C-30/15 est correcte : REFUSER
Resultat pour C-299/99 est correcte : ACCEPTER
Resultat pour C-26/17 est correcte : REFUSER
Resultat pour C-24/05 est correcte : REFUSER
Resultat pour C-238/06 est correcte : REFUSER
Resultat pour C-218/01 b est correcte : REFUSER
Resultat pour C-218/01 a est correcte : ACCEPTER
Resultat pour C-217/13 est correcte : ACCEPTER
Resultat pour C-215/14 b est correcte : ACCEPTER
Resultat pour C-215/14 a est correcte : REFUSER
Resultat pour C-196/11 est correcte : REFUSER
Resultat pour C-173/04 est correcte : REFUSER
Resultat pour C-163/16 est correcte : ACCEPTER
Resultat pour C-144/06 est correcte : REFUSER
Resultat pour C-136/02 est correcte : REFUSER
Resultat pour C-124/18 est correcte : REFUSER
Resultat pour C-108/97b est correcte : REFUSER
Resultat pour C-108/97a est correcte : ACCEPTER
Resultat pour C-108/07 est correcte : REFUSER
Resultat pour C-108/05 b est correcte : REFUSER
Resultat pour C-108/05 a est correcte : ACCEPTER
Resultat pour C-104/01 b est correcte : REFUSER
Resultat pour C-104/01 a est correcte : ACCEPTER
Total : 46/46
true .
?- ■

```

FIGURE 5.6 – Résultat obtenue en ouvrant *marque.pl* avec le programme *SWI-Prolog version 7.6.4* et en lançant l'atome *marque_test* avec le fichier test *valeurTest.csv*(repris à l'annexe L à la page 116)

FIGURE 5.8 – Résultat obtenu en important le fichier *valeurTest2.csv* dans le programme *marque.pl* avec *SWI-Prolog*

Cependant tout à l'air correct sauf un test. Nous allons analyser ce test incorrect, le cas C-328/18.

L'affaire C-328/18 concerne la demande d'enregistrement d'une marque dans la classe de produit n° 3, "les parfums", pour la zone de l'Union Européenne. La marque est rejetée car elle possède des similitudes avec une autre marque de parfum en République Tchèque, en Hongrie, en Pologne et en Slovénie, représentant un risque de confusion pour le public de ces régions. L'erreur provient du fait que dans notre questionnaire, nous ne demandons pas si la marque est similaire à une marque antérieure dans l'une des zones et dans l'une des classes de produit et de service de la demande. Nous aurions pu répondre à la question "Un public reconnaît-il la marque?" par "non" car s'il y a un risque de confusion entre 2 marques, nous ne pouvons pas

référence	Quel sera le nom de la marque ou de sa référence	Quel est le format de la marque	Quel est la langue utilisée pour la marque	Quels sont les pays concernés dans la demande	Quels sont les langues utilisées dans la zone territoriale de la demande	Quels sont les classes de produit ou de service de la demande	La marque a un caractère distinctif dans la classe de produit ou de service	Un public pertinent reconnaît-il la marque
C-43/15	compressor technology	verbale	indéterminé	UE	tous	7, 9, 11	non	non
C-281/18	REPOWER	verbale	anglais	UE	tous		9	non
C-155/18	BURLINGTON	verbale	indéterminé	UE	tous		35	non
C-99/18	C-99/18	figurative	null	UE	null	16, 39, 43	oui	non
C-328/18	C-328/18	figurative	null	UE	null		3	oui

référence	N y a t il pas de risque de conflit avec l'intérêt général	Le signe est un caractère descriptif de la classe de produit ou du service contenant	Le signe est une marque dans une autre classe de produit ou de service	La marque est identique ou partiellement identique à une marque acquise un caractère distinctif par l'usage avant la demande d'enregistrement	La marque est suffisamment décrite	La marque utilise un terme géographique	Le terme géographique est-il imagé ou féérique	Y a t il un lien entre le terme géographique et la classe de produit ou de service
C-43/15	null	non	non	null	oui	non	null	null
C-281/18	null	oui	non	non	oui	non	null	null
C-155/18	null	non	non	null	oui	non	null	null
C-99/18	null	non	non	null	oui	null	null	null
C-328/18	null	non	non	null	oui	null	null	null

référence	Le marque utilise une appellation d'origine contrôlée	La protection de l'AOC s'applique à la classe de produit	La marque verbale est un slogan publicitaire	La marque est la même classe de produit ou service	La marque est auditive et similaire à une marque antérieure dans la même classe de produit ou service	La marque possède un caractère distinctif fort	La marque utilise un patronyme répandu dans la zone territoriale de la demande	La taille du marché de la demande est trop vaste pour permettre au public de reconnaître la marque comme appartenant à une seule entité	La marque utilise des abréviations d'un signe
C-43/15	non	null	non	oui	oui	non	null	non	
C-281/18	non	null	non	non	null	non	null	non	
C-155/18	non	null	non	non	null	non	null	non	
C-99/18	null	null	null	null	null	null	null	null	
C-328/18	null	null	null	null	null	null	null	null	

référence	Le signe représenté par les abréviations possède un caractère distinctif	La marque tridimensionnelle diverge de la norme ou des habitudes du secteur	Le signe peut être utilisé par la concurrence de manière habituelle pour décrire ou produire le même produit ou service	La forme de la marque est protégée par un brevet de conception	La marque est vue dans son ensemble et possède un caractère distinctif	Résultat
C-43/15	null	null	null	null	non	REFUSER
C-281/18	null	null	null	null	non	REFUSER
C-155/18	null	null	null	null	oui	ACCEPTER
C-99/18	null	null	null	null	oui	ACCEPTER
C-328/18	null	null	null	null	oui	REFUSER

FIGURE 5.7 – Listing des questions-réponses pour les 5 arrêts de jurisprudence qui serviront de test pour le programme *marque.pl* avec *SWI-Prolog*

dire que la marque soit reconnue par un public pertinent.

Le programme *marque.pl* a deux usages. Le premier usage permet de tester la cohérence des décisions d'acceptation d'une marque en utilisant une liste de marques comme repris à l'annexe 5.7 à la page 70. Le second usage, qui était le but principal du programme, est de

```

Reference de la marque : C-328/18.
quel sera le nom de la marque ou de ca reference ?
Nom de la marque : C-328/18.
quel est le format de la marque ?
Type de la marque : semifigurative.
quels sont les pays concernes dans la demande ?
Zone territoriale de la marque : UE.
Quel est la langue utilise pour la marque C-328/18?
La langue utilisee pour la marque : null
Quels sont les langues utilisees dans la zone territoriale de la demande?
Les langues utilisees dans la zone territoriale de la demande : null
quels sont les classes de produit ou de service de la demande ?
Les classes de la marque : [3].
la marque est suffisamment decrite ?
Marque suffisamment decrite: oui.
la marque utilise un terme geographique ?
Marque ayant un terme geographique: non.
la marque auditivement semblable a une marque anterieur ?
Marque est auditivement semblable : non.
la marque utilise un patronyme rependu dans la zone territoriale de la demande ?
Utilisation d un patronyme : non.
un public pertinent reconnait il la marque ?
Public pertinent de la marque : null.
n y a t il pas de risque de conflit avec l interet general ?
L interet general de la marque : null.
la marque est identique ou partiellement identique a une marque ayant une renommee dans une autre classe de produit ou de service ?
Conflit avec marque renommee : non.
la marque est un slogan publicitaire ?
Marque est un slogan : non.
la marque utilise les abreviations de la marque ?
Utilisation d abreviation : non.
Le caractere distinctif des classes de la marque : [oui].
le signe est un caractere descriptif du produit ou du service dans la langue de la marque ?
Caractere descriptif : non.
Resultat voulu : REFUSER.

```

FIGURE 5.9 – Parcours de l’affaire C-328/18 obtenu en important le fichier *valeurTest2.csv* dans le programme *marque.pl* avec *SWI-Prolog*

permettre à un utilisateur de pouvoir déterminer si la demande d’enregistrement d’un signe qu’il pourrait introduire auprès d’une institution agréée serait acceptée ou s’il existe un risque de refus d’enregistrement. Le code dans *marque.pl* remplit ces fonctions sous forme de questions et d’évaluation de la réponse à la question, il agit comme un arbre de décision où chaque question conduit vers une branche en fonction de la réponse données par l’utilisateur ou vers une feuille lorsque le programme donne l’appréciation finale à l’utilisateur.

5.4 Conclusion

Dans ce chapitre, nous avons abordé la programmation logique déductive pour pouvoir réaliser le programme qui prédit l’acceptation d’un signe en tant que marque.

Les résultats obtenus sont bons mais les questions et l’évaluation de ceux-ci devraient être revus et corrigés par des experts du droit des marques . Il faudrait aussi augmenter le nombre de cas afin de couvrir toutes les possibilités qui pourraient se présenter lors de la demande d’enregistrement d’une marque.

Le langage utilisé est le *Prolog* qui est un langage de programmation logique. Le programme *marque.pl* s’appuie sur le programme "*Legrandgousier*" du Professeur Jean-Marie Jacquet, fournit comme travail dans le cadre du cours de technique d’intelligence artificielle à ses étudiants pour l’année académique 2018-2019, pour réaliser la transformation des textes en liste et inversement. Le logiciel "*SWI-Prolog version 7.6.4*" de Jan Wielemaker a été utilisé pour faire fonctionner le programme *marque.pl*.

Chapitre 6

Conclusion

En Belgique, la justice est l'un des dernier domaine dans lequel l'intelligence artificielle n'est pas encore utilisées. Pourtant, elle pourrait ,entre autre, apporter une diminution des coûts ou une diminution des délais. Cependant elle n'est pas sans risque. Nous pouvons citer le risque d'une privatisation de le justice ou encore le risque d'une diminution de l'accessibilité aux personnes les plus fragiles. Ce sont ces risques qui font peur, à une partie du monde de la magistrature. Malgré tout elle n'a pas que des mauvais côtés, elle pourrait être utile dans le domaine de la prévention, par exemple en ayant une police prédictive, une police qui pourrait prévenir au lieu de subir.

Le travail que nous avons réalisé portait sur la justice prédictive appliquée au droit des marques. Ceci consiste à déterminer si nous pouvons prédire l'acceptation d'un signe en tant qu'une marque en appliquant le droit des marques.

La première méthode consiste en l'utilisation d'outils d'analyse de textes fournis par le programme Orange [Borondics et al., 2020]. Les textes utilisés sont des arrêts de jurisprudence liés aux droits des marques. Dans un premier temps, nous nous posons la question de savoir si nous pouvons prédire le résultat d'un arrêt de jurisprudence avec la distance cosinus calculée entre les textes et les modèles algorithmiques fournis par le logiciel Orange [Borondics et al., 2020]. Par la suite, avec ces mêmes outils et en reclassant les différents arrêts en fonction de l'acceptation ou du refus de la marque en question, nous avons pu prédire si une marque pouvait être enregistrée conformément aux droits des marques.

L'analyse des résultats, se rapportant à la première question, nous a prouvé qu'il est possible de prédire le résultat d'un pourvoi en appliquant les algorithmes fournis par le logiciel Orange [Borondics et al., 2020]. Les résultats obtenus à la deuxième question, nous ont indiqué qu'il n'était pas possible de prédire l'acceptation d'une marque en se basant sur les textes bruts. La solution trouvée consistant en la réalisation d'un tableau reprenant pour chaque arrêt une liste de valeur comprise entre -1 et 1 qui correspond à des réponses aux questions qui tentent de cerner les cas où un signe a été accepté en tant que marque ou non. Les résultats obtenus avec les algorithmes du logiciel Orange [Borondics et al., 2020] ont démontré que l'on pouvait avec cette méthode dire qu'un signe pouvait être enregistré en tant que marque.

La seconde méthode a consisté en l'élaboration d'un programme utilisant la programmation logique déductive pour répondre à question *"Un signe pourra-t-il être accepté en tant que marque en regard du droit des marques ?"*. Les résultats de la méthode nous ont permis de dire qu'elle répondait bien à la question posée.

Enfin, les tests réalisés avec le logiciel Orange [Borondics et al., 2020] sur le jeu de données "Conclusion" sont positifs mais en prenant un peu de recul, nous constatons qu'ils seront plus utiles pour vérifier la cohérence entre les jugements entre autre à des fins de recherche. Un justiciable n'a pas besoin de prédire si un pourvoi en cours a des chances d'être gagné mais bien de prédire le résultat de l'enregistrement d'une marque. D'où l'utilité des derniers tests réalisés avec le logiciel Orange [Borondics et al., 2020] ou du programme *marque.pl* qui permettent, après une série de questions, de donner une prédiction de la réponse que la cour pourrait donner sur l'enregistrement ou le refus de la marque.

Le travail pourrait être complété, par l'augmentation des jeux de données ce qui permettrait de faire des test sur une plus grande quantité de données ainsi, garantir la prise en charge de tous les cas possibles du droit des marques. Par la suite, les questions devront être aussi complétées et corrigées avec la participation d'expert du droit des marques.

Nous pourrions aussi affiner les tests sur les textes bruts en adaptant les paramètres de tests pour chaque modèle algorithmique de prédiction afin de dégager le modèle répondant le mieux à la recherche. Le programme *marque.pl* pourrait être refait en utilisant, cette fois-ci, la programmation logique inductive afin de le comparer avec le programme déjà réalisé dans ce travail.

Bibliographie

- [Lar, 2016] (2016). Le petit Larousse Illustré 2017. Larrousse.
- [AdM avec A.Ca., 2020] AdM avec A.Ca. (2020). Salon de l’auto : Coyote présente son nouveau boîtier avec commande vocale et système de "sécurité prédictive". La Dernière Heure. Consulté en date du 9 janvier 2020.
- [Balboni, 2019] Balboni, J. (2019). Bientôt jugés par un robot ? la justice pourrait utiliser l’intelligence artificielle. Page consulté en date du 11 décembre 2019.
- [Bastien, 2019] Bastien, L. (2019). Réseau de neurones artificiels : qu’est-ce que c’est et à quoi ça sert ? Consulté en date du 28 avril 2020.
- [Belga, 2020] Belga (2020). La nasa découvre une autre planète de la taille de la terre dans une "zone habitable". Consulté en date du 7 janvier 2020.
- [Benzaki, 2017] Benzaki, Y. (2017). Naive bayes classifier pour machine learning. Consulté en date du 28 avril 2020.
- [Borondics et al., 2020] Borondics, Vitaliet, and Shaulsky (2020). Orange widget-catalog.
- [Chabaud, 2020] Chabaud, C. (2020). Quand le robot menace d’éclipser le juge et l’avocat. Consulté en date du 26 juillet 2020.
- [Dachez, 2019] Dachez, K. (2019). Apple card : la carte de paiement accusée de sexisme. Page consulté en date du 11 décembre 2019.
- [Desjardins, 2018] Desjardins, C. (2018). Il y a cinq niveaux d’intelligence artificielle. Consulté en date du 3 mars 2020.
- [euipo.europa.eu, 2020a] euipo.europa.eu (2015-2020a). Ce que peut être une marque de l’ue. Consulté en date du 16 octobre 2019.
- [euipo.europa.eu, 2020b] euipo.europa.eu (2015-2020b). Définition d’une marque. Consulté en date du 13 octobre 2019.
- [Evans and Grefenstette, 2018] Evans, R. and Grefenstette, E. (2018). Learning explanatory rules from noisy data. J. Artif. Int. Res., 61(1) :1–64.
- [Gillet et al., 2011] Gillet, A., Brostaux, Y., and Palm, R. (2011). Principaux modèles utilisés en régression logistique. BASE[En ligne], 15(3) :425–426. consulté en date du 27 avril 2020.
- [Hubin et al., 2019] Hubin, J., Jacquemin, H., and Michaux, B. (2019). Le juge et l’algorithme : juges augmentés ou justice diminuée ? Collection du CRIDS. Éditions Larcier.
- [Maitre Frédéric Lejeune, 2019] Maitre Frédéric Lejeune, a. e. p. i. a. b. d. B. (2019). Droit des marques : le cas jaguar et la marque renommée. Consulté en date du 16 novembre 2019.
- [McCaffrey, 2016] McCaffrey, J. (2016). Classification et prédiction par boosting adaptatif.
- [Muggleton and Raedt, 1994] Muggleton, S. and Raedt, L. D. (1994). Inductive logic programming : Theory and methods. JOURNAL OF LOGIC PROGRAMMING, 19(20) :629–679.

- [Negre, 2013] Negre, E. (2013). Comparaison de textes : quelques approches... Cahier du Lamsade, pages 5–8.
- [Parlement Européen et Conseil de l'Union Européenne, 2015] Parlement Européen et Conseil de l'Union Européenne (2015). Article 4 paragraphe 1 de la directive (UE) 2015/2436 du parlement européen et du conseil.
- [Parlement Européen et Conseil de l'Union Européenne, 2017] Parlement Européen et Conseil de l'Union Européenne (2017). Article 7, paragraphe 1 du règlement EU 2017/1001 du parlement européen et du conseil.
- [press@just.fgov.be, 2019] press@just.fgov.be (2019). Crossborder récompensé à la seconde édition des cpl awards. Consulté en date du 10 aout 2020.
- [Rouvière, 2018] Rouvière, F. (2018). Le raisonnement par algorithmes : le fantasme du juge-robot. RTDCiv. : Revue trimestrielle de droit civil, 02 :530.
- [Sanderson, 2017] Sanderson, G. (2017). Mais *qu'est-ce* qu'un réseau de neurones ? | apprentissage profond, chapitre 1. Consulté en date du 8 aout 2020.
- [Signoret, 2018] Signoret, P. (2018). Amazon a dû désactiver une ia qui discriminait les candidatures de femmes à l'embauche. Page consulté en date du 11 décembre 2019.
- [SPF Economie, 2019] SPF Economie (2019). Marques : conditions de protection. Consulté en date du 16 octobre 2019.
- [Thomas G. Tape,] Thomas G. Tape, M. The area under an roc curve. Consulté en date du 29 avril 2020.
- [University of Ljubljana, 2020] University of Ljubljana (2020). Widget test and score. Consulté en date du 29 avril 2020.
- [van den Branden et al., 2019] van den Branden, A., Alexandre, L., Garapon, A., and Buyle, J. (2019). Les robots à l'assaut de la justice : L'intelligence artificielle au service des justiciables. Bruylant.
- [Wikipédia, 2018] Wikipédia (2018). Distance entre deux points sur le plan cartésien — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 27 décembre 2018 ; Page consulté le 9 mai 2020].
- [Wikipédia, 2019a] Wikipédia (2019a). Courbe roc — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 25 août 2019 ; Page consulté le 30 avril 2020].
- [Wikipédia, 2019b] Wikipédia (2019b). F1 score — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 25 avril 2019 ; Page consulté le 30 avril 2020].
- [Wikipédia, 2019c] Wikipédia (2019c). Prédiction — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 29-septembre-2019 ; Consuté en date du 21 juin 2020].
- [Wikipédia, 2019d] Wikipédia (2019d). Programmation logique inductive — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 26 juin 2019 ; Page consulté le 11 novembre 2019].
- [Wikipédia, 2019e] Wikipédia (2019e). Régression logistique — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 21 novembre 2019 ; Consulté en date du 27 avril 2020].
- [Wikipédia, 2020a] Wikipédia (2020a). Adaboost — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 24 février 2020 ; Consulté en date du 27 avril 2020].
- [Wikipédia, 2020b] Wikipédia (2020b). A.i. intelligence artificielle — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 15-juin-2020 ; Consulté en date du 6 juillet 2020].

- [Wikipédia, 2020c] Wikipédia (2020c). Algorithme du gradient stochastique — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 21 avril 2020 ; Page consulté en date du 28 avril 2020].
- [Wikipédia, 2020d] Wikipédia (2020d). Alphago — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 2-mai-2020 ; Consulté en date du 5 juillet 2020].
- [Wikipédia, 2020e] Wikipédia (2020e). Arbre de décision — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 16 avril 2020 ; Page consulté le 28 avril 2020].
- [Wikipédia, 2020f] Wikipédia (2020f). Arbre de décision (apprentissage) — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 20 avril 2020 ; Page consulté le 28 avril 2020].
- [Wikipédia, 2020g] Wikipédia (2020g). Distance (mathématiques) — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 25 mars 2020 ; Page consulté le 9 mai 2020].
- [Wikipédia, 2020h] Wikipédia (2020h). Intelligence artificielle — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 5-juillet-2020 ; Consulté en date du 5 juillet 2020].
- [Wikipédia, 2020i] Wikipédia (2020i). Justice — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 20-mai-2020 ; Consulté en date du 21 juin 2020].
- [Wikipédia, 2020j] Wikipédia (2020j). Méthode des k plus proches voisins — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 5 février 2020 ; Consulté en date du 27 avril 2020].
- [Wikipédia, 2020k] Wikipédia (2020k). Précision et rappel — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 22 avril 2020 ; Page consulté le 30 avril 2020].
- [Wikipédia, 2020l] Wikipédia (2020l). Programmation logique — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 24 mars 2020 ; Consulté le 12 mai 2020].
- [Wikipédia, 2020m] Wikipédia (2020m). Silhouette (clustering) — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 12-août-2020 ; Consulté en date du 15 août 2020].
- [Wikipédia, 2020n] Wikipédia (2020n). Similarité cosinus — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 20 février 2020 ; Page consulté le 9 mai 2020].
- [Wikipédia, 2020o] Wikipédia (2020o). Watson (intelligence artificielle) — wikipédia, l'encyclopédie libre. [En ligne ; Page disponible le 25-mai-2020 ; Page consulté en date du 27 juin 2020].

Résumé de la matière :

- "Liste d'arrêts CJUE – caractère distinctif (DM) Bref résumé de la matière pour l'ILP à travers la jurisprudence de la Cour de justice de l'Union européenne (CJUE)" du 4 octobre 2019, auteur Zorana Rosic

Annexe A

Article 4, paragraphe 1 de la directive 2015/2436

Motifs absolus de refus ou de nullité

1. *Sont refusés à l'enregistrement ou sont susceptibles d'être déclarés nuls s'ils sont enregistrés :*

- a) les signes qui ne peuvent constituer une marque ;*
- b) les marques qui sont dépourvues de caractère distinctif ;*
- c) les marques qui sont composées exclusivement de signes ou d'indications pouvant servir, dans le commerce, à désigner l'espèce, la qualité, la quantité, la destination, la valeur, la provenance géographique ou l'époque de la production du produit ou de la prestation du service, ou d'autres caractéristiques de ceux-ci ;*
- d) les marques qui sont composées exclusivement de signes ou d'indications devenus usuels dans le langage courant ou dans les habitudes loyales et constantes du commerce ;*
- e) les signes constitués exclusivement :*
 - i) par la forme ou une autre caractéristique imposée par la nature même du produit ;*
 - ii) par la forme ou une autre caractéristique du produit qui est nécessaire à l'obtention d'un résultat technique ;*
 - iii) par la forme ou une autre caractéristique qui donne une valeur substantielle au produit ;*
- f) les marques qui sont contraires à l'ordre public ou aux bonnes mœurs ;*
- g) les marques qui sont de nature à tromper le public, par exemple, sur la nature, la qualité ou la provenance géographique du produit ou du service ; 23.12.2015 L 336/7 Journal officiel de l'Union européenne FR*
- h) les marques qui, à défaut d'autorisation des autorités compétentes, sont à refuser ou à invalider en vertu de l'article 6 ter de la convention de Paris ;*
- i) les marques exclues de l'enregistrement en application de la législation de l'Union ou du droit national de l'État membre concerné, ou d'accords internationaux auxquels l'Union ou l'État membre concerné est partie, qui prévoient la protection des appellations d'origine et des indications géographiques ;*

- j) les marques exclues de l'enregistrement en application de la législation de l'Union ou d'accords internationaux auxquels l'Union est partie qui prévoient la protection des mentions traditionnelles pour les vins ;*
- k) les marques exclues de l'enregistrement en application de la législation de l'Union ou d'accords internationaux auxquels l'Union est partie qui prévoient la protection des spécialités traditionnelles garanties ;*
- l) les marques qui consistent en une dénomination d'une variété végétale antérieure enregistrée conformément à la législation de l'Union ou au droit national de l'État membre concerné, ou d'accords internationaux auxquels l'Union ou l'État membre concerné est partie, qui prévoient la protection des droits d'obtention végétale, ou la reproduisent dans leurs éléments essentiels, et qui portent sur des variétés végétales de la même espèce ou d'une espèce étroitement liée.*

Annexe B

Article 7, paragraphe 1 du règlement EU 2017/1001

Motifs absolus de refus

1. Sont refusés à l'enregistrement :

- a) les signes qui ne sont pas conformes à l'article 4 ;*
- b) les marques qui sont dépourvues de caractère distinctif ;*
- c) les marques qui sont composées exclusivement de signes ou d'indications pouvant servir, dans le commerce, à désigner l'espèce, la qualité, la quantité, la destination, la valeur, la provenance géographique ou l'époque de la production du produit ou de la prestation du service, ou d'autres caractéristiques de ceux-ci ;*
- d) les marques qui sont composées exclusivement de signes ou d'indications devenus usuels dans le langage courant ou dans les habitudes loyales et constantes du commerce ;*
- e) les signes constitués exclusivement :*
 - i) par la forme, ou une autre caractéristique, imposée par la nature même du produit ;*
 - ii) par la forme, ou une autre caractéristique du produit, nécessaire à l'obtention d'un résultat technique ;*
 - iii) par la forme, ou une autre caractéristique du produit, qui donne une valeur substantielle au produit ;*
- f) les marques qui sont contraires à l'ordre public ou aux bonnes moeurs ; 16.6.2017 L 154/8 Journal officiel de l'Union européenne FR*
- g) les marques qui sont de nature à tromper le public, par exemple sur la nature, la qualité ou la provenance géographique du produit ou du service ;*
- h) les marques qui, à défaut d'autorisation des autorités compétentes, sont à refuser en vertu de l'article 6 ter de la convention de Paris pour la protection de la propriété industrielle (ci-après dénommée "convention de Paris") ;*
- i) les marques qui comportent des badges, emblèmes ou écussons autres que ceux visés par l'article 6 ter de la convention de Paris et présentant un intérêt public particulier, à moins que leur enregistrement n'ait été autorisé par l'autorité compétente ;*

- j) les marques exclues de l'enregistrement en application de la législation de l'Union ou du droit national ou d'accords internationaux auxquels l'Union ou l'État membre concerné est partie, qui prévoient la protection des appellations d'origine et des indications géographiques ;*
- k) les marques exclues de l'enregistrement en application de la législation de l'Union ou d'accords internationaux auxquels l'Union est partie, qui prévoient la protection des mentions traditionnelles pour les vins ;*
- l) les marques exclues de l'enregistrement en application de la législation de l'Union ou d'accords internationaux auxquels l'Union est partie, qui prévoient la protection des spécialités traditionnelles garanties ; m) les marques qui consistent en une dénomination d'une variété végétale antérieure enregistrée conformément à la législation de l'Union ou au droit national ou aux accords internationaux auxquels l'Union ou l'État membre concerné est partie, qui prévoient la protection des droits d'obtention végétale, ou la reproduisent dans leurs éléments essentiels, et qui portent sur des variétés végétales de la même espèce ou d'une espèce étroitement liée.*

Annexe C

Résumé des petites applications "widgets" utilisé dans le programme Orange, version 3.24.1 (data mining) [Borondics et al., 2020]

Import Documents



Entrée : Néant.

Sortie : { Une liste de textes contenant la catégorie dans laquelle le fichier a été classé.
(en fonction des sous répertoires du répertoire principale), l'adresse, le nom
et le contenu du fichier.

Fonction : Permettre d'importer un répertoire contenant les fichiers textes à analyser.

Corpus Viewer



Entrée : Une liste de textes.

Sortie : Néant.

Fonction : Afficher les textes avec leurs nouveaux attributs.

Preprocess Text



Entrée : Une liste de textes.

Sortie : Une liste de textes modifiés.

Fonction : Effectuer un traitement personnalisable sur les différents textes.

Word Cloud



Entrée : Une liste des textes.

Sortie : Néant.

Fonction : { Afficher sous forme de nuage l'ensemble des mots et signes repris
dans les textes en fonction du poids de chacun d'eux.

Bag of Words



Entrée : Une liste des textes.

Sortie : Une liste des textes avec pour chaque texte une liste des mots et leurs poids.

Fonction : Ajouter une liste des mots et de leurs fréquences pour chaque textes.

Distances



Entrée : Une liste de données.
Sortie : Une matrice de distances.
Fonction : Calculer une distance entre les données.

Data Table



Entrée : Une liste de données.
Sortie : Néant.
Fonction : Afficher les données de la liste.

Hierarchical Clustering



Entrée : Une matrice de distances.
Sortie : Néant.
Fonction : Afficher un regroupement hiérarchique des distances.

MDS



Entrée : Un ensemble de données.
Sortie : Néant.
Fonction : Afficher les distances entre les points.

Distance Matrix



Entrée : Une matrice de distance.
Sortie : Un tableau de mesure de distance.
Fonction : Afficher un tableau mesurant les distances entre les points 2 à 2.

Silhouette Plot



Entrée : Une liste de données.
Sortie : Néant.
Fonction : Afficher le coefficient de silhouette des données.

Data Sampler



Entrée : Une liste de données.
Sortie : Deux listes $\left\{ \begin{array}{l} - \text{Une liste de données d'exemple.} \\ - \text{Une liste de données de test.} \end{array} \right.$
Fonction : Séparer la liste de données en 2 listes distinguées.

Confusion Matrix



Entrée : Liste de résultats.
Sortie : Néant.
Fonction : Afficher une matrice comparant les résultats entre la prédiction et la réalité.

AdaBoost



- Entrée : {
- Une liste de données de test.
- Néant.
- Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.
- Fonction : {
Fournir un algorithme d'apprentissage ou un modèle basé sur
"Adaptive Boosting" de Yoav Freund et Robert Schapire.

kNN



- Entrée : {
- Une liste de données de test.
- Néant.
- Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.
- Fonction : {
Fournir un algorithme d'apprentissage ou un modèle basé sur
les k plus proche voisin ("k-nearest neighbors").

Logistic Regression



- Entrée : {
- Une liste de données de test.
- Néant.
- Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.
- Fonction : {
Fournir un algorithme d'apprentissage ou un modèle basé sur
la régression logistique.

Naive Bayes



- Entrée : {
- Une liste de données de test.
- Néant.
- Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.
- Fonction : Fournir un algorithme d'apprentissage ou un modèle basé sur la théorie de Bayes.

Neural Network



- Entrée : {
- Une liste de données de test.
- Néant.
- Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.
- Fonction : {
Fournir un algorithme d'apprentissage ou un modèle basé sur
les réseaux de neurones artificiels.

Random Forest



- Entrée : {
- Une liste de données de test.
- Néant.
- Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.
- Fonction : {
Fournir un algorithme d'apprentissage ou un modèle basé sur
les forêts d'arbre décisionnel ("Random Forest").

Stochastic Gradient Descent



Entrée : {
- Une liste de données de test.
- Néant.

Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.

Fonction : {
Fournir un algorithme d'apprentissage ou un modèle basé sur
l'algorithme du gradient stochastique ("Stochastic Gradient Descent").

Tree



Entrée : {
- Une liste de données de test.
- Néant.

Sortie : {
- Un modèle prédictif.
- Un modèle algorithmique d'apprentissages.

Fonction : {
Fournir un algorithme d'apprentissage ou un modèle basé sur
des arbres de décisions.

Test and Score



Entrée : {
- Une liste de données d'exemple.
- Une liste de données de test.
- Un ou plusieurs modèles algorithmiques d'apprentissages.

Sortie : Les résultats pour chaque algorithme d'apprentissage.

Fonction : {
Afficher une évaluation des résultats pour chaque algorithme
d'apprentissage sur les données de test.

Predictions



Entrée : {
Une liste de données de test.
Un ou plusieurs modèles prédictifs.

Sortie : Une liste de données comprenant les résultats des prédictions

Fonction : {
Afficher une évaluation des résultats pour chaque modèle prédictifs
et le résultat sur la base de données de test.

Data Concatenate



Entrée : {
Un jeu de données primaires.
Un ou plusieurs jeux de données secondaires.

Sortie : Un jeu de données comprenant les données reçues en entrée

Fonction : Regrouper l'ensemble des jeux de données reçus en un seul jeu de données

Annexe D

Liste de mots à enlever lors de la phase de pré-traitement des textes

- a
- est
- être
- si
- sous
- doit
- fait
- ci
- tant
- ceux
- autres
- peut
- ainsi
- tout
- selon
- donc
- tel
- où
- sans
- cette
- cet
- dès
- dont
- lors
- celle
- celui
- no

Annexe E

Liste de mots à comptabiliser lors de la phase de pré-traitement du texte

- caractères
- distinctifs
- caractère
- distinctif
- couleur
- verbale
- figuratif
- figurative
- figuratifs
- figuratives
- verbaux
- forme
- position
- motif
- couleur
- couleurs
- sonore
- mouvement
- multimédia
- hologramme
- tridimensionnel
- article
- articles
- paragraphe
- paragraphes
- territoire
- intérêt
- général
- intérêts
- généraux
- dépôt
- demande
- enregistrement
- ordre
- public
- mœurs
- tromper
- géographique
- 1
- 3
- 4
- 6
- 7
- produit
- produits
- services
- service
- perception
- perceptions
- descriptif
- descriptifs
- usage
- usages
- application
- stricte
- demande
- critère
- applications
- strictes
- demandes
- critères
- rejet
- rejeter
- annulation
- annuler
- annulé
- rejeté
- pourvoi
- admis
- affaire
- renvoyée
- public
- publics
- pertinent
- pertinents
- consommateur
- consommateurs
- confusion
- confusions
- générique
- génériques
- similitude
- similitudes
- visuelle
- visuel
- visuels
- visuelles
- phonétique
- phonétiques
- antérieur
- identité
- origine
- protection
- appellation
- appellations
- 53
- 8
- renommée

Annexe F

Liste des arrêts utilisés dans la base de données

Tous les arrêts utilisés pour la base de données ainsi que les détails de la liste ci-dessous sont issues du site <http://curia.europa.eu>

- **C-108/97 C-109/97** : Arrêt de la Cour du 4 mai 1999.
Windsurfing Chiemsee Produktions- und Vertriebs GmbH (WSC) contre Boots- und Segelzubehör Walter Huber et Franz Attenberger.
Identifiant ECLI : ECLI :EU :C :1999 :230
- **C-342/97** : Arrêt de la Cour du 22 juin 1999.
Lloyd Schuhfabrik Meyer & Co. GmbH contre Klijsen Handel BV.
Identifiant ECLI : ECLI :EU :C :1999 :323
- **C-299/99** : Arrêt de la Cour du 18 juin 2002.
Koninklijke Philips Electronics NV contre Remington Consumer Products Ltd.
Demande de décision préjudicielle : Court of Appeal (England & Wales) (Civil Division) - Royaume-Uni.
Identifiant ECLI : ECLI :EU :C :2002 :377
- **C-363/99** : Arrêt de la Cour (sixième chambre) du 12 février 2004.
Koninklijke KPN Nederland NV contre Benelux-Merkenbureau.
Identifiant ECLI : ECLI :EU :C :2004 :86
- **C-383/99** : Arrêt de la Cour du 20 septembre 2001.
Procter & Gamble Company contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles).
Identifiant ECLI : ECLI :EU :C :2001 :461
- **C-53/01 à C-55/01** : Arrêt de la Cour du 8 avril 2003.
Linde AG (C-53/01), Winward Industries Inc. (C-54/01) et Rado Uhren AG (C-55/01).
Identifiant ECLI : ECLI :EU :C :2003 :206
- **C-104/01** : Arrêt de la Cour du 6 mai 2003.
Libertel Groep BV contre Benelux-Merkenbureau.
Identifiant ECLI : ECLI :EU :C :2003 :244
- **C-218/01** : Arrêt de la Cour (sixième chambre) du 12 février 2004.
Henkel KGaA.
Identifiant ECLI : ECLI :EU :C :2004 :88
- **C-64/02** : Arrêt de la Cour (deuxième chambre) du 21 octobre 2004.

- Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI) contre Erpo Möbelwerk GmbH.
Identifiant ECLI : ECLI :EU :C :2004 :645
- **C-136/02** : Arrêt de la Cour (deuxième chambre) du 7 octobre 2004.
Mag Instrument Inc. contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles).
Identifiant ECLI : ECLI :EU :C :2004 :592
 - **C-329/02** : Arrêt de la Cour (deuxième chambre) du 16 septembre 2004.
SAT.1 SatellitenFernsehen GmbH contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2004 :532
 - **C-404/02** : Arrêt de la Cour (deuxième chambre) du 16 septembre 2004.
Nichols plc contre Registrar of Trade Marks.
Identifiant ECLI : ECLI :EU :C :2004 :538
 - **C-447/02** : Arrêt de la Cour (deuxième chambre) du 21 octobre 2004.
KWS Saat AG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2004 :649
 - **C-37/03** : Arrêt de la Cour (troisième chambre) du 15 septembre 2005.
BioID AG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2005 :547
 - **C-321/03** : Arrêt de la Cour (troisième chambre) du 25 janvier 2007.
Dyson Ltd contre Registrar of Trade Marks.
Identifiant ECLI : ECLI :EU :C :2007 :51
 - **C-173/04** : Arrêt de la Cour (première chambre) du 12 janvier 2006.
Deutsche SiSi-Werke GmbH & Co. Betriebs KG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2006 :20
 - **C-421/04** : Arrêt de la Cour (première chambre) du 9 mars 2006.
Matratzen Concord AG contre Hukla Germany SA.
Identifiant ECLI : ECLI :EU :C :2006 :164
 - **C-24/05** : Arrêt de la Cour (première chambre) du 22 juin 2006.
August Storck KG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2006 :421
 - **C-108/05** : Arrêt de la Cour (première chambre) du 7 septembre 2006.
Bovemij Verzekeringen NV contre Benelux-Merkenbureau.
Identifiant ECLI : ECLI :EU :C :2006 :530
 - **C-239/05** : Arrêt de la Cour (deuxième chambre) du 15 février 2007.
BVBA Management, Training en Consultancy contre Benelux-Merkenbureau.
Identifiant ECLI : ECLI :EU :C :2007 :99
 - **C-334/05** : Arrêt de la Cour (troisième chambre) du 12 juin 2007.
Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI) contre Shaker di L. Laudato & C. Sas.
Identifiant ECLI : ECLI :EU :C :2007 :333

- **C-144/06** : Arrêt de la Cour (huitième chambre) du 4 octobre 2007.
Henkel KgaA contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2007 :577
- **C-238/06** : Arrêt de la Cour (huitième chambre) du 25 octobre 2007.
Develey Holding GmbH & Co. Beteiligungs KG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2007 :635
- **C-304/06** : Arrêt de la Cour (première chambre) du 8 mai 2008.
Eurohypo AG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2008 :261
- **C-371/06** : Arrêt de la Cour (sixième chambre) du 20 septembre 2007.
Benetton Group SpA contre G-Star International BV.
Identifiant ECLI : ECLI :EU :C :2007 :542
- **C-108/07** : Arrêt de la Cour (première chambre) du 17 avril 2008.
Ferrero Deutschland GmbH contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2008 :234
- **C-542/07** : Arrêt de la Cour (cinquième chambre) du 11 juin 2009.
Imagination Technologies Ltd contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2009 :362
- **C-398/08** : Arrêt de la Cour (première chambre) du 21 janvier 2010.
Audi AG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2010 :29
- **C-48/09** : Arrêt de la Cour (grande chambre) du 14 septembre 2010.
Lego Juris A/S contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2010 :516
- **C-344/10 C-345/10** : Arrêt de la Cour (troisième chambre) du 20 octobre 2011.
Freixenet, SA contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2011 :680
- **C-90/11 C-91/11** : Arrêt de la Cour (troisième chambre) du 15 mars 2012.
Alfred Strigl - Deutsches Patent- und Markenamt et Securvita Gesellschaft zur Entwicklung alternativer Versicherungskonzepte mbH contre Öko-Invest Verlagsgesellschaft mbH.
Identifiant ECLI : ECLI :EU :C :2012 :147
- **C-98/11** : Arrêt de la Cour (quatrième chambre) du 24 mai 2012.
Chocoladefabriken Lindt & Sprüngli AG contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2012 :307
- **C-196/11** : Arrêt de la Cour (troisième chambre) du 24 mai 2012.

- Formula One Licensing BV contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI) et Global Sports Media Ltd.
Identifiant ECLI : ECLI :EU :C :2012 :314
- **C-311/11** : Arrêt de la Cour (cinquième chambre) du 12 juillet 2012.
Smart Technologies ULC contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2012 :460
 - **C-97/12** : Arrêt de la Cour (huitième chambre) du 15 mai 2014.
Louis Vuitton Malletier contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2014 :324
 - **C-530/12** : Arrêt de la Cour (première chambre) du 27 mars 2014.
Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI) contre National Lottery Commission.
Identifiant ECLI : ECLI :EU :C :2014 :186
 - **C-558/12** : Arrêt de la Cour (sixième chambre) du 23 janvier 2014.
Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI) contre riha WeserGold Getränke GmbH & Co. KG.
Identifiant ECLI : ECLI :EU :C :2014 :22
 - **C-597/12** : Arrêt de la Cour (septième chambre) du 17 octobre 2013.
Isdin SA contre Bial-Portela & C² SA.
Identifiant ECLI : ECLI :EU :C :2013 :672
 - **C-217/13 C-218/13** : Arrêt de la Cour (troisième chambre) du 19 juin 2014.
Oberbank AG e.a. contre Deutscher Sparkassen- und Giroverband eV.
Identifiant ECLI : ECLI :EU :C :2014 :2012
 - **C-421/13** : Arrêt de la Cour (troisième chambre) du 10 juillet 2014.
Apple Inc. contre Deutsches Patent- und Markenamt.
Identifiant ECLI : ECLI :EU :C :2014 :2070
 - **C-445/13** : Arrêt de la Cour (sixième chambre) du 7 mai 2015.
Voss of Norway ASA contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2015 :303
 - **C-215/14** : Arrêt de la Cour (première chambre) du 16 septembre 2015.
Société de Produits Nestlé SA contre Cadbury UK Ltd.
Identifiant ECLI : ECLI :EU :C :2015 :604
 - **C-537/14** : Arrêt de la Cour (huitième chambre) du 27 octobre 2016.
Debonair Trading Internacional Lda contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2016 :814
 - **C-30/15** : Arrêt de la Cour (première chambre) du 10 novembre 2016.
Simba Toys GmbH & Co. KG contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2016 :849
 - **C-43/15** : Arrêt de la Cour (grande chambre) du 8 novembre 2016.
BSH Bosch und Siemens Hausgeräte GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).

Identifiant ECLI : ECLI :EU :C :2016 :837

- **C-50/15** : Arrêt de la Cour (sixième chambre) du 21 janvier 2016.
Kurt Hesse contre Office de l'harmonisation dans le marché intérieur (marques, dessins et modèles) (OHMI).
Identifiant ECLI : ECLI :EU :C :2016 :34
- **C-207/15** : Arrêt de la Cour (cinquième chambre) du 22 juin 2016.
Nissan Jidosha KK contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2016 :465
- **C-226/15** : Arrêt de la Cour (quatrième chambre) du 21 juillet 2016.
Apple and Pear Australia Ltd et Star Fruits Diffusion contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2016 :582
- **C-421/15** : Arrêt de la Cour (cinquième chambre) du 11 mai 2017.
Yoshida Metal Industry Co. Ltd contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2017 :360
- **C-437/15** : Arrêt de la Cour (troisième chambre) du 17 mai 2017.
Office de l'Union européenne pour la propriété intellectuelle (EUIPO) contre Deluxe Entertainment Services Group Inc.
Identifiant ECLI : ECLI :EU :C :2017 :380
- **C-482/15** : Arrêt de la Cour (dixième chambre) du 26 octobre 2016.
Westermann Lernspielverlage GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2016 :805
- **C-501/15** : Arrêt de la Cour (quatrième chambre) du 11 octobre 2017.
Office de l'Union européenne pour la propriété intellectuelle (EUIPO) contre Cactus SA.
Identifiant ECLI : ECLI :EU :C :2017 :750
- **C-642/15** : Arrêt de la Cour (troisième chambre) du 1er décembre 2016.
Toni Klement contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2016 :918
- **C-673/15 C-676/15** : Arrêt de la Cour (deuxième chambre) du 20 septembre 2017.
The Tea Board contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2017 :702
- **C-56/16** : Arrêt de la Cour (deuxième chambre) du 14 septembre 2017.
Office de l'Union européenne pour la propriété intellectuelle (EUIPO) contre Instituto dos Vinhos do Douro e do Porto IP.
Identifiant ECLI : ECLI :EU :C :2017 :693
- **C-85/16 C-86/16** : Arrêt de la Cour (dixième chambre) du 30 mai 2018.
Kenzo Tsujimoto contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2018 :349
- **C-163/16** : Arrêt de la Cour (grande chambre) du 12 juin 2018.
Christian Louboutin et Christian Louboutin SAS contre Van Haren Schoenen BV.

Identifiant ECLI : ECLI :EU :C :2018 :423

- **C-418/16** : Arrêt de la Cour (première chambre) du 28 février 2018.
mobile.de GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2018 :128
- **C-488/16** : Arrêt de la Cour (cinquième chambre) du 6 septembre 2018.
Bundesverband Souvenir – Geschenke – Ehrenpreise eV contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2018 :673
- **C-634/16** : Arrêt de la Cour (première chambre) du 24 janvier 2018.
Office de l'Union européenne pour la propriété intellectuelle (EUIPO) contre European Food SA.
Identifiant ECLI : ECLI :EU :C :2018 :30
- **C-26/17** : Arrêt de la Cour (dixième chambre) du 13 septembre 2018.
Birkenstock Sales GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2018 :714
- **C-84/17 C-85/17 C-95/17** : Arrêt de la Cour (troisième chambre) du 25 juillet 2018.
Société des produits Nestlé SA e.a. contre Mondelez UK Holdings & Services Ltd.
Identifiant ECLI : ECLI :EU :C :2018 :596
- **C-578/17** : Arrêt de la Cour (quatrième chambre) du 27 mars 2019.
Procédure engagée par Oy Hartwall Ab.
Identifiant ECLI : ECLI :EU :C :2019 :261
- **C-653/17** : Arrêt de la Cour (cinquième chambre) du 15 mai 2019.
VM Vermögens-Management GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2019 :406
- **C-684/17** : Ordonnance de la Cour (dixième chambre) du 17 mai 2018.
Banca Monte dei Paschi di Siena SpA et Wise Dialog Bank SpA (Banca Widiba SpA) contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2018 :332
- **C-99/18** : Arrêt de la Cour (dixième chambre) du 4 juillet 2019.
FTI Touristik GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2019 :565
- **C-124/18** : Arrêt de la Cour (dixième chambre) du 29 juillet 2019.
Red Bull GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2019 :641
- **C-155/18 à C-158/18** : Arrêt de la Cour (quatrième chambre) du 4 mars 2020.
Tulliallan Burlington Ltd contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2020 :151
- **C-223/18** : Arrêt de la Cour (sixième chambre) du 6 juin 2019.
Deichmann SE contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).

Identifiant ECLI : ECLI :EU :C :2019 :471

- **C-240/18** : Arrêt de la Cour (cinquième chambre) du 27 février 2020.
Constantin Film Produktion GmbH contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2020 :118
- **C-281/18** : Arrêt de la Cour (deuxième chambre) du 31 octobre 2019.
Repower AG contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2019 :916
- **C-328/18** : Arrêt de la Cour (quatrième chambre) du 4 mars 2020.
Office de l'Union européenne pour la propriété intellectuelle (EUIPO) contre
Equivalenza Manufactory SL.
Identifiant ECLI : ECLI :EU :C :2020 :156
- **C-541-18** : Arrêt de la Cour (cinquième chambre) du 12 septembre 2019.
AS contre Deutsches Patent- und Markenamt.
Identifiant ECLI : ECLI :EU :C :2019 :725
- **C-766/18** : Arrêt de la Cour (cinquième chambre) du 5 mars 2020.
Foundation for the Protection of the Traditional Cheese of Cyprus named Halloumi
contre Office de l'Union européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2020 :170
- **C-143/19** : Arrêt de la Cour (cinquième chambre) du 12 décembre 2019.
Der Grüne Punkt – Duales System Deutschland GmbH contre Office de l'Union
européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2019 :1076
- **C-499/19** : Arrêt de la Cour (cinquième chambre) du 12 décembre 2019.
Der Grüne Punkt – Duales System Deutschland GmbH contre Office de l'Union
européenne pour la propriété intellectuelle (EUIPO).
Identifiant ECLI : ECLI :EU :C :2019 :1076

Annexe G

Différentes Bases de données et leurs contenus

Liste des arrêts	Base de donnée				
	Conclusion	Concl Test	Acceptation	Accept Test	Accept Ex
C-108/97	Avis				
C-342/97	Avis				
C-299/99		Avis			
C-363/99	Avis				
C-383/99	Annulation		Accepter		Accepter
C-53/01	Avis		Refuser		Refuser
C-104/01	Avis				
C-218/01	Avis				
C-64/02	Rejet		Accepter		Accepter
C-136/02	Rejet		Refuser		Refuser
C-329/02	Annulation		Accepter		Accepter
C-404/02	Avis				
C-447/02	Rejet		Refuser		Refuser
C-37/03	Annulation		Refuser		Refuser
C-321/03	Avis		Refuser		Refuser
C-173/04	Rejet		Refuser		Refuser
C-421/04	Avis		Accepter		Accepter
C-24/05	Rejet		Refuser		Refuser
C-108/05	Avis				
C-239/05	Avis				
C-334/05	Annulation				
C-144/06	Rejet		Refuser		Refuser
C-238/06	Rejet		Refuser		Refuser
C-304/06	Rejet		Refuser		Refuser
C-371/06	Avis				
C-108/07	Annulation		Refuser		Refuser
C-542/07	Rejet		Refuser	Refuser	
C-398/08	Annulation		Accepter	Accepter	
C-48/09	Rejet		Accepter		Accepter
C-344/10	Annulation		Accepter	Accepter	
C-90/11	Avis		Refuser		Refuser
C-98/11	Rejet		Refuser	Refuser	

Liste des arrêts	Base de donnée				
	Conclusion	Concl Test	Acceptation	Accept Test	Accept Ex
C-196/11	Annulation		Refuser		Refuser
C-311/11	Rejet		Refuser	Refuser	
C-97/12	Rejet		Refuser		Refuser
C-530/12		Annulation			
C-558/12	Annulation				
C-597/12	Annulation				
C-217/13	Avis		Accepter		Accepter
C-421/13	Avis				
C-445/13	Rejet		Refuser		Refuser
C-215/14	Avis				
C-537/14		Annulation			
C-30/15	Annulation		Refuser		Refuser
C-43/15		Rejet	Refuser		Refuser
C-50/15		Rejet	Refuser	Refuser	
C-207/15	Annulation				
C-226/15		Rejet	Refuser	Refuser	
C-421/15		Rejet	Refuser	Refuser	
C-437/15	Annulation				
C-482/15		Rejet	Refuser	Refuser	
C-501/15	Rejet		Accepter	Accepter	
C-642/15	Annulation				
C-673/15	Rejet		Accepter		Accepter
C-56/16	Annulation		Accepter	Accepter	
C-85/16	Rejet		Refuser	Refuser	
C-163/16	Avis		Accepter		Accepter
C-418/16	Rejet		Refuser		Refuser
C-488/16	Rejet		Accepter		Accepter
C-634/16	Rejet		Accepter		Accepter
C-26/17	Rejet		Refuser		Refuser
C-84/17	Rejet		Accepter		Accepter
C-578/17	Avis		Accepter	Accepter	
C-653/17	Rejet		Refuser	Refuser	
C-684/17	Rejet				
C-99/18	Rejet		Accepter		Accepter
C-124/18	Rejet		Refuser	Refuser	
C-155/18		Annulation	Refuser		Refuser
C-223/18	Rejet				
C-240/18	Annulation				
C-281/18	Rejet		Refuser		Refuser
C-328/18		Annulation	Refuser		Refuser
C-541-18	Avis				
C-766/18		Annulation			
C-143/19	Annulation				
C-499/19	Rejet				

Annexe H

Résultats des analyses des textes sous forme de questions

Identifiant	Description	La marque antérieure		Le signe est																		
		La marque antérieure	Le signe est																			
C-383/99	Marque verbale composée de plusieurs termes qui possèdent un caractère distinctif dans son ensemble	1	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-398/08	Marque verbale constituée de slogan publicitaire ayant acquis un caractère distinctif	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-404/02 a	Marque verbale composée d'un patronyme courant avec un caractère distinctif acquis à l'usage à la hauteur de la classe voulue	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-404/02 b	Marque verbale composée d'un patronyme courant avec un caractère distinctif acquis à l'usage à la hauteur de la classe voulue	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-421/04	Marque verbale composée de terme qui dans une autre langue (qui ne n'est pas compris par la majorité des consommateurs de la zone d'enregistrement) de la zone de l'Union Européenne possède un caractère descriptif	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-445/13	Marque tridimensionnelle descriptif d'un emballage sans caractère distinctif acquis à l'usage	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-447/02	Marque de couleur n'ayant pas de caractère distinctif dans les classes de produit et de service soustaité	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-48/09	Marque tridimensionnelle descriptif du produit ou du service	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-53/01 C-53/01 C-54/01 C-55/01	Marque verbale d'un lieu image et original	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-54/07	Marques tridimensionnelles représentant la forme du produit ou du service	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-542/07	Marque verbale sans caractère distinctif au moment de l'enregistrement de la demande	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-56/16	Marque verbale ayant un signe lié à une appellation d'origine pour un produit différent de l'appellation (whisky et vin de Porto)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-578/17	Marque de couleur ayant un caractère distinctif reconnue en temps que marque figurative	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-64/02	Marque verbale composée d'un slogan publicitaire	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-84/17 C-85/17 C-85/17	Marque tridimensionnelle représentant la forme du produit ou service	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-90/11 C-91/11	Marques verbales représentant les initiales de la description du produit ou du service	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-97/12	Marque figurative représentant la description du produit ou du service	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C-98/11	Marque figurative représentant la description du produit ou du service	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Annexe I

Résultats des tests sur le schéma repris à la figure 4.9 de la page 36

Rapports des résultats édités avec le logiciel Orange [Borondics et al., 2020] sur les textes séparés

- * en 3 catégories (jeu de données "Conclusion" annexe G à la page 95) séparées en fonction de la fréquence des termes :
 - en mode comptage
 - en mode binaire
- * en 2 catégories (jeu de données "Acceptation" annexe G à la page 95) séparées en fonction des combinaison de mots et en fonction de la fréquence des termes :
 - en 1-1 (combinaison de 1 mot à 1 mot)
 - en 1-2 (combinaison de 1 mot à 2 mots)

Test and Score

Sat Aug 15 20, 11:58:42

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
kNN	0.9201680672268908	0.7894736842105263	0.7894736842105263	0.8135338345864662	0.7894736842105263
Logistic Regression	0.8907563025210086	0.7894736842105263	0.7763157894736842	0.7894736842105263	0.7894736842105263
AdaBoost	0.8802521008403362	0.8421052631578947	0.83778677462888	0.8399122807017544	0.8421052631578947
Neural Network	0.8697478991596639	0.6842105263157895	0.5797213622291022	0.5035087719298246	0.6842105263157895
SGD	0.8676470588235294	0.8421052631578947	0.822202565236621	0.8894736842105262	0.8421052631578947
Naive Bayes	0.865546218487395	0.631578947368421	0.5232198142414861	0.47807017543859653	0.631578947368421
Random Forest	0.7689075630252102	0.631578947368421	0.6225887115979994	0.6403508771929824	0.631578947368421

Résultat des tests sur Le jeu de donnée "Conclusion"
 Preprocess text : tous Les mots sauf La liste des mots "KO" en 1-1
 Bag of word : en binaire

Confusion Matrix

Sat Aug 15 20, 11:59:25

Confusion matrix for kNN (showing number of instances)

		Predicted			
		REJET	ANNULATION	AVIS	Σ
Actual	REJET	4	3	0	7
	ANNULATION	1	4	0	5
	AVIS	0	0	7	7
	Σ	5	7	7	19

Confusion Matrix

Sat Aug 15 20, 11:59:37

Confusion matrix for Logistic Regression (showing number of instances)

		Predicted			
		REJET	ANNULATION	AVIS	Σ
Actual	REJET	6	1	0	7
	ANNULATION	3	2	0	5
	AVIS	0	0	7	7
	Σ	9	3	7	19

Confusion Matrix

Confusion matrix for AdaBoost (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	5	1	1	7
	ANNULATION	1	4	0	5
	AVIS	0	0	7	7
	Σ	6	5	8	19



Confusion Matrix

Confusion matrix for Neural Network (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	6	0	1	7
	ANNULATION	3	0	2	5
	AVIS	0	0	7	7
	Σ	9		10	19



Confusion Matrix

Confusion matrix for SGD (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	7	0	0	7
	ANNULATION	3	2	0	5
	AVIS	0	0	7	7
	Σ	10	2	7	19



Confusion Matrix

Confusion matrix for Naive Bayes (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	0	7	0	7
	ANNULATION	0	5	0	5
	AVIS	0	0	7	7
	Σ		12	7	19



Confusion Matrix

Sat Aug 15 20, 12:00:34

Confusion matrix for Random Forest (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	5	2	0	7
	ANNULATION	4	1	0	5
	AVIS	1	0	6	7
	Σ	10	3	6	19



Test and Score

Sat Aug 15 20, 10:08:28

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
Random Forest	0.9075630252100841	0.6842105263157895	0.5894736842105263	0.5233253588516746	0.6842105263157895
Naive Bayes	0.8865546218487397	0.631578947368421	0.5232198142414861	0.47807017543859653	0.631578947368421
Neural Network	0.865546218487395	0.5263157894736842	0.4351503759398497	0.3742690058479532	0.5263157894736842
Logistic Regression	0.8613445378151261	0.7894736842105263	0.7763157894736842	0.7894736842105263	0.7894736842105263
kNN	0.846638655462185	0.7368421052631579	0.7309941520467836	0.7302631578947368	0.7368421052631579
AdaBoost	0.7962184873949582	0.7368421052631579	0.742518059855521	0.7868421052631579	0.7368421052631579
SGD	0.796218487394958	0.7368421052631579	0.7309941520467836	0.7302631578947368	0.7368421052631579

Résultat des tests sur Le jeu de donnée "Conclusion"
 Preprocess text : tous Les mots sauf La liste des mots "KO" en 1-1
 Bag of word : en comptage

Confusion Matrix

Sat Aug 15 20, 10:10:51

Confusion matrix for Random Forest (showing number of instances)

		Predicted			
		REJET	ANNULATION	AVIS	Σ
Actual	REJET	6	0	1	7
	ANNULATION	5	0	0	5
	AVIS	0	0	7	7
	Σ	11		8	19

Confusion Matrix

Sat Aug 15 20, 10:15:15

Confusion matrix for Naive Bayes (showing number of instances)

		Predicted			
		REJET	ANNULATION	AVIS	Σ
Actual	REJET	0	7	0	7
	ANNULATION	0	5	0	5
	AVIS	0	0	7	7
	Σ		12	7	19

Confusion Matrix

Sat Aug 15 20, 10:16:41

Confusion matrix for Neural Network (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	0	6	1	7
	ANNULATION	1	3	1	5
	AVIS	0	0	7	7
	Σ	1	9	9	19

Confusion Matrix

Sat Aug 15 20, 10:16:56

Confusion matrix for Logistic Regression (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	6	1	0	7
	ANNULATION	3	2	0	5
	AVIS	0	0	7	7
	Σ	9	3	7	19

Confusion Matrix

Sat Aug 15 20, 10:17:12

Confusion matrix for kNN (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	5	2	0	7
	ANNULATION	3	2	0	5
	AVIS	0	0	7	7
	Σ	8	4	7	19

Confusion Matrix

Sat Aug 15 20, 10:17:22

Confusion matrix for AdaBoost (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	6	1	0	7
	ANNULATION	2	3	0	5
	AVIS	2	0	5	7
	Σ	10	4	5	19

Confusion Matrix

Sat Aug 15 20, 10:17:33

Confusion matrix for SGD (showing number of instances)

		Predicted			Σ
		REJET	ANNULATION	AVIS	
Actual	REJET	5	2	0	7
	ANNULATION	3	2	0	5
	AVIS	0	0	7	7
	Σ	8	4	7	19

Test and Score

Sat Aug 15 20, 17:54:30

Settings**Sampling type:** No sampling, test on testing data**Target class:** Average over classes**Scores**

Model	AUC	CA	F1	Precision	Recall
kNN	0.7121212121212122	0.7142857142857143	0.7357142857142857	0.7841269841269841	0.7142857142857143
AdaBoost	0.6515151515151516	0.6428571428571429	0.6741854636591479	0.7589285714285714	0.6428571428571429
Logistic Regression	0.48484848484848486	0.6428571428571429	0.6149068322981366	0.5892857142857143	0.6428571428571429
Random Forest	0.40909090909090906		0.5	0.5238095238095238	0.5499999999999999
Neural Network	0.39393939393939394	0.42857142857142855	0.47936507936507944	0.5918367346938777	0.42857142857142855
SGD	0.36363636363636365	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
Naive Bayes	0.18181818181818182	0.21428571428571427	0.07563025210084033	0.04591836734693877	0.21428571428571427

Résultat des tests sur Le jeu de donnée "Acceptation"
 Preprocess text : tous Les mots sauf La Liste des mots "k0" en 1-1
 Bag of word : en comptage

Confusion Matrix

Sat Aug 15 20, 17:56:43

Confusion matrix for kNN (showing number of instances)

		Predicted		Σ
		ACCEPTER	REFUSER	
Actual	ACCEPTER	2	1	3
	REFUSER	3	8	11
Σ		5	9	14

Confusion Matrix

Sat Aug 15 20, 17:56:55

Confusion matrix for AdaBoost (showing number of instances)

		Predicted		Σ
		ACCEPTER	REFUSER	
Actual	ACCEPTER	2	1	3
	REFUSER	4	7	11
Σ		6	8	14

Test and Score

Sat Aug 15 20, 18:14:53

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
kNN	0.6666666666666667	0.6428571428571429	0.6741854636591479	0.7589285714285714	0.6428571428571429
AdaBoost	0.6060606060606062	0.5714285714285714	0.6095238095238094	0.7346938775510204	0.5714285714285714
Logistic Regression	0.4545454545454546	0.6428571428571429	0.6149068322981366	0.5892857142857143	0.6428571428571429
Random Forest	0.42424242424242425	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
SGD	0.4090909090909091	0.6428571428571429	0.6149068322981366	0.5892857142857143	0.6428571428571429
Naive Bayes	0.3030303030303031	0.21428571428571427	0.07563025210084033	0.04591836734693877	0.21428571428571427
Neural Network		0.7857142857142857	0.6914285714285714	0.6173469387755102	0.7857142857142857

Résultat des tests sur Le jeu de donnée "Acceptation"
 Preprocess text : tous Les mots sauf La Liste des mots "k0" en 1-1
 Bag of word : en comptage

Confusion Matrix

Sat Aug 15 20, 18:15:31

Confusion matrix for kNN (showing number of instances)

		Predicted		Σ
		ACCEPTER	REFUSER	
Actual	ACCEPTER	2	1	3
	REFUSER	4	7	11
Σ		6	8	14

Confusion Matrix

Sat Aug 15 20, 18:15:47

Confusion matrix for AdaBoost (showing number of instances)

		Predicted		Σ
		ACCEPTER	REFUSER	
Actual	ACCEPTER	2	1	3
	REFUSER	5	6	11
Σ		7	7	14

Test and Score

Sat Aug 15 20, 18:22:09

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
AdaBoost	0.36363636363636365	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
SGD	0.36363636363636365	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
kNN	0.21212121212121213	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
Naive Bayes	0.21212121212121213	0.21428571428571427	0.17802197802197803	0.42857142857142855	0.21428571428571427
Logistic Regression	0.18181818181818182	0.6428571428571429	0.6149068322981366	0.5892857142857143	0.6428571428571429
Neural Network	0.18181818181818182	0.2857142857142857	0.3303571428571428	0.4952380952380952	0.2857142857142857
Random Forest	0.12121212121212123	0.42857142857142855	0.4714285714285714	0.5238095238095238	0.42857142857142855

Résultat des tests sur Le jeu de donnée "Acceptation"
 Preprocess text : tous les mots sauf la liste des mots "KO" en 1-1
 Bag of word : en mode binaire

Test and Score

Sat Aug 15 20, 18:34:29

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
Random Forest	0.6818181818181819	0.7142857142857143	0.6547619047619048	0.6043956043956044	0.7142857142857143
kNN	0.4848484848484849	0.42857142857142855	0.46428571428571436	0.6761904761904762	0.42857142857142855
AdaBoost	0.48484848484848486	0.5714285714285714	0.6035714285714286	0.653968253968254	0.5714285714285714
SGD	0.45454545454545453	0.7142857142857143	0.6547619047619048	0.6043956043956044	0.7142857142857143
Naive Bayes	0.30303030303030304	0.21428571428571427	0.07563025210084033	0.04591836734693877	0.21428571428571427
Logistic Regression	0.27272727272727276	0.7142857142857143	0.6547619047619048	0.6043956043956044	0.7142857142857143
Neural Network	0.18181818181818182	0.35714285714285715	0.40870893812070286	0.550595238095238	0.35714285714285715

Résultat des tests sur Le jeu de donnée "Acceptation"
 Preprocess text : tous les mots sauf la liste des mots "KO" en 1-2
 Bag of word : en mode binaire

Confusion Matrix

Sat Aug 15 20, 18:35:12

Confusion matrix for Random Forest (showing number of instances)

		Predicted		Σ
		ACCEPTER	REFUSER	
Actual	ACCEPTER	0	3	3
	REFUSER	1	10	11
Σ		1	13	14

Test and Score

Sat Aug 15 20, 19:19:43

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.4848484848484849	0.7142857142857143	0.7357142857142857	0.7841269841269841	0.7142857142857143
AdaBoost	0.43939393939393945		0.5	0.543859649122807	0.625
SGD	0.36363636363636365	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
Random Forest	0.24242424242424243	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
kNN	0.16666666666666669	0.35714285714285715	0.41353383458646614	0.49107142857142855	0.35714285714285715
Logistic Regression	0.12121212121212123		0.5	0.5238095238095238	0.5499999999999999
Naive Bayes	0.09090909090909091	0.21428571428571427	0.27731092436974786	0.39285714285714285	0.21428571428571427

Résultat des tests sur Le jeu de donnée "Acceptation"
 Preprocess text : tous Les mots de La liste "OK" en 1-1
 Bag of word : en mode comptage

Test and Score

Sat Aug 15 20, 19:22:20

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
AdaBoost	0.5151515151515152	0.42857142857142855	0.46428571428571436	0.6761904761904762	0.42857142857142855
SGD	0.48484848484848486	0.5714285714285714	0.6035714285714286	0.653968253968254	0.5714285714285714
kNN	0.3484848484848485		0.5	0.5238095238095238	0.5499999999999999
Naive Bayes	0.33333333333333337	0.21428571428571427	0.07563025210084033	0.04591836734693877	0.21428571428571427
Neural Network	0.27272727272727276	0.6428571428571429	0.6598639455782314	0.6821428571428572	0.6428571428571429
Random Forest	0.24242424242424243	0.42857142857142855	0.4714285714285714	0.5238095238095238	0.42857142857142855
Logistic Regression	0.21212121212121215	0.5714285714285714	0.6035714285714286	0.653968253968254	0.5714285714285714

Résultat des tests sur Le jeu de donnée "Acceptation"
 Preprocess text : tous Les mots de La liste "OK" en 1-2
 Bag of word : en mode comptage

Confusion Matrix

Sat Aug 15 20, 19:23:42

Confusion matrix for AdaBoost (showing number of instances)

		Predicted		Σ
		ACCEPTER	REFUSER	
Actual	ACCEPTER	2	1	3
	REFUSER	7	4	11
Σ		9	5	14

Test and Score

Sat Aug 15 20, 18:38:15

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
kNN	0.7424242424242424	0.6428571428571429	0.6598639455782314	0.6821428571428572	0.6428571428571429
AdaBoost	0.48484848484848486	0.5714285714285714	0.6035714285714286	0.653968253968254	0.5714285714285714
SGD	0.4090909090909091	0.6428571428571429	0.6149068322981366	0.5892857142857143	0.6428571428571429
Random Forest	0.3939393939393939	0.5714285714285714	0.6035714285714286	0.653968253968254	0.5714285714285714
Naive Bayes	0.18181818181818182	0.42857142857142855	0.4714285714285714	0.5238095238095238	0.42857142857142855
Neural Network	0.18181818181818182	0.6428571428571429	0.6149068322981366	0.5892857142857143	0.6428571428571429
Logistic Regression	0.12121212121212122	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714

Résultat des tests sur le jeu de donnée "Acceptation"
 Preprocess text : tous Les mots de la liste "OK" en 1-1
 Bag of word : en mode binaire

Confusion Matrix

Sat Aug 15 20, 18:38:56

Confusion matrix for kNN (showing number of instances)

		Predicted		Σ
		ACCEPTER	REFUSER	
Actual	ACCEPTER	1	2	3
	REFUSER	3	8	11
Σ		4	10	14

Test and Score

Sat Aug 15 20, 18:41:38

Settings

Sampling type: No sampling, test on testing data
Target class: Average over classes

Scores

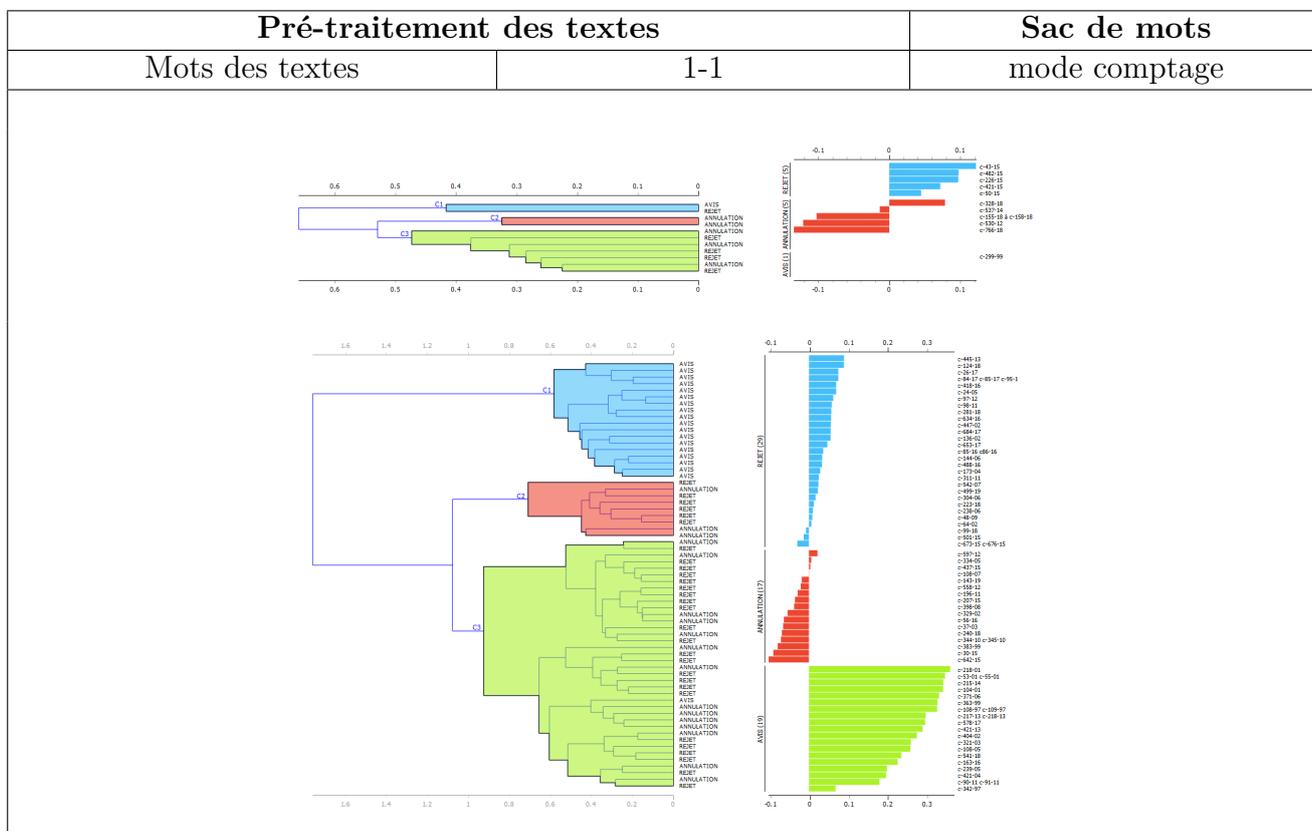
Model	AUC	CA	F1	Precision	Recall
Random Forest	0.5	0.7142857142857143	0.6547619047619048	0.6043956043956044	0.7142857142857143
SGD	0.5	0.7857142857142857	0.6914285714285714	0.6173469387755102	0.7857142857142857
kNN	0.42424242424242425	0.6428571428571429	0.6598639455782314	0.6821428571428572	0.6428571428571429
Naive Bayes	0.42424242424242425	0.14285714285714285	0.05357142857142857	0.03296703296703297	0.14285714285714285
AdaBoost	0.36363636363636365	0.5714285714285714	0.5714285714285714	0.5714285714285714	0.5714285714285714
Logistic Regression	0.33333333333333337	0.7142857142857143	0.6547619047619048	0.6043956043956044	0.7142857142857143
Neural Network	0.18181818181818182	0.7142857142857143	0.6547619047619048	0.6043956043956044	0.7142857142857143

*Résultat des tests sur Le jeu de donnée "Acceptation"
Preprocess text : tous Les mots de La liste "OK" en 1-2
Bag of word : en mode binaire*

Annexe J

Représentation graphique des résultats obtenus sur le schéma repris à la figure 4.12 de la page 40 avec les textes repris dans la base de données "Conclusion" et "Concl Test" (voir l'annexe G de la page 95

Représentation des graphes en mode comptage et binaire avec une combinaison limitée à 1-1 sur les mots des textes pour le schéma repris à la figure 4.12 à la page 40 avec 2 jeux de données séparés en 3 catégories.



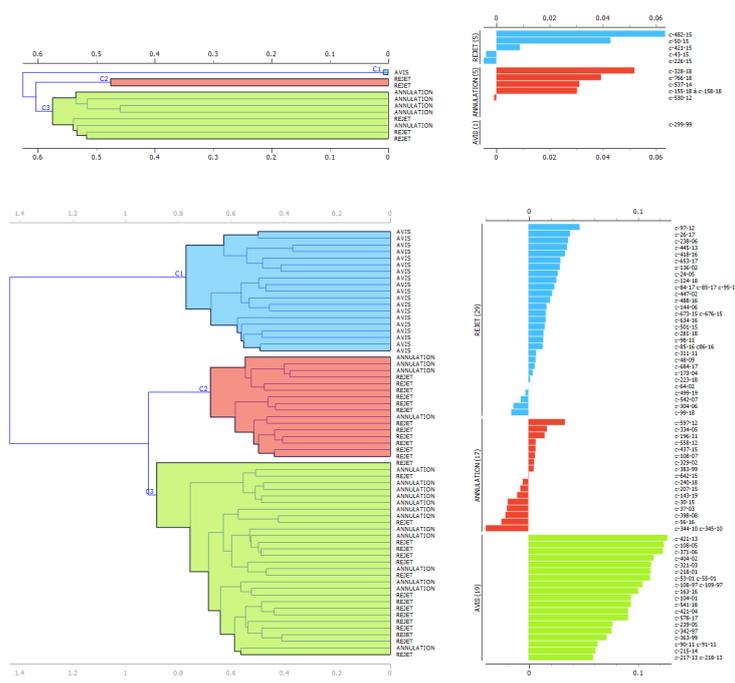
Pré-traitement des textes

Sac de mots

Mots des textes

1-1

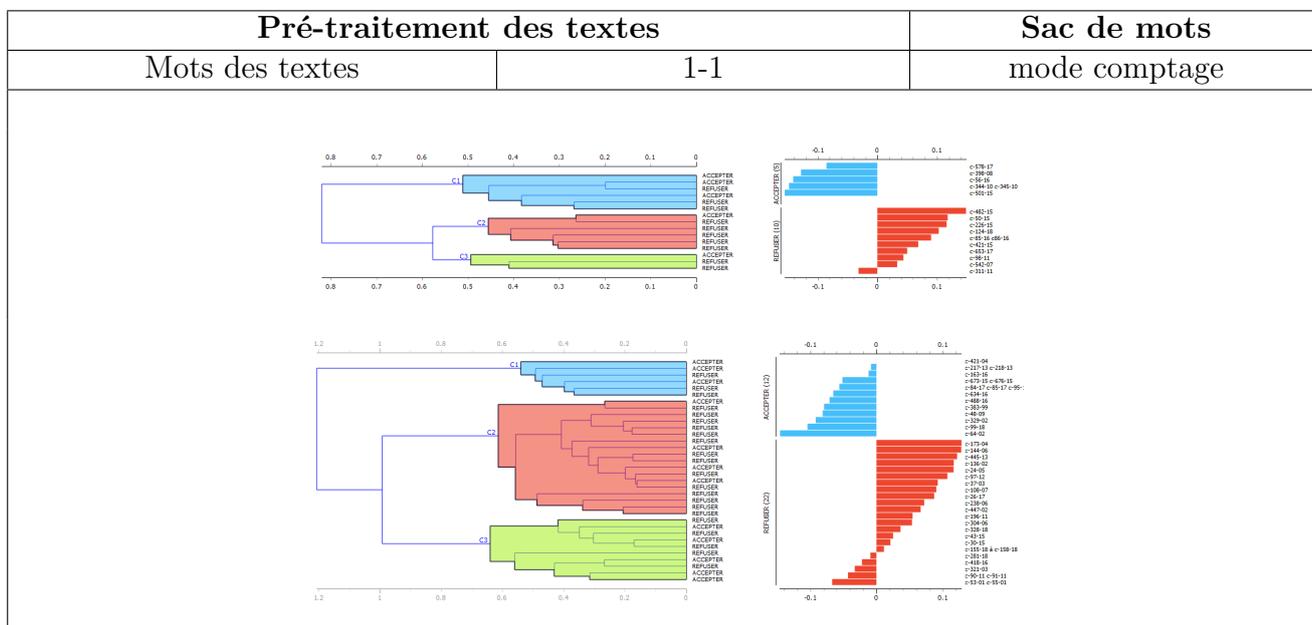
mode binaire

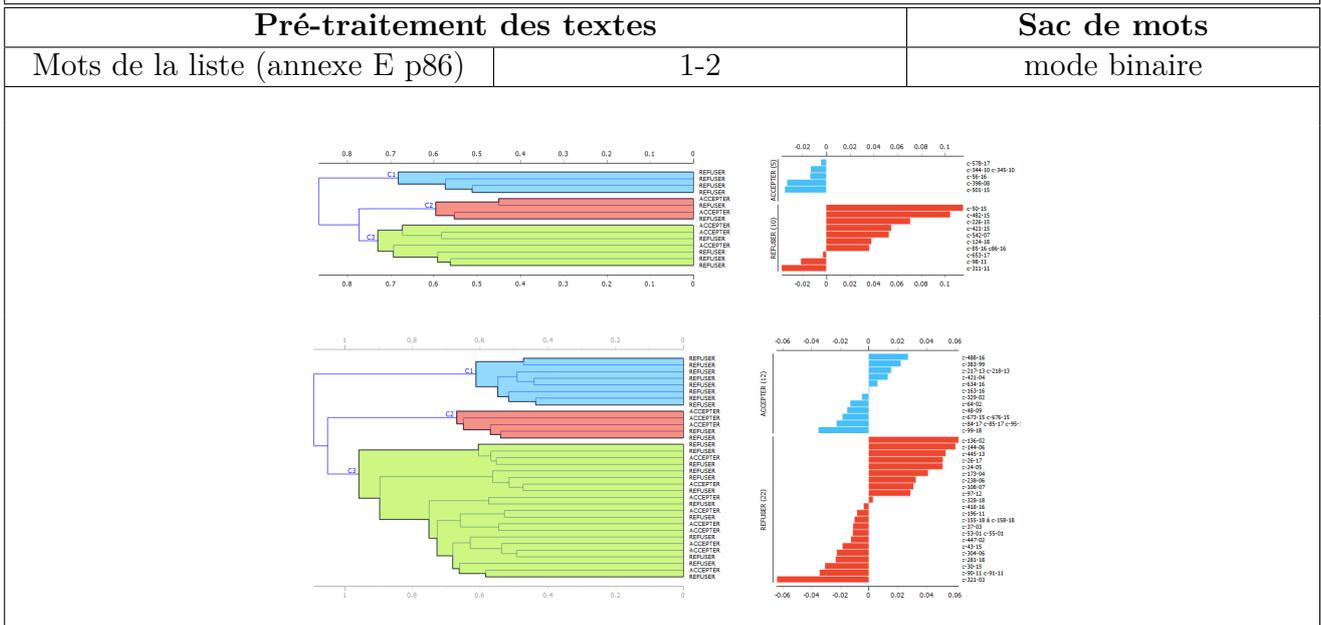
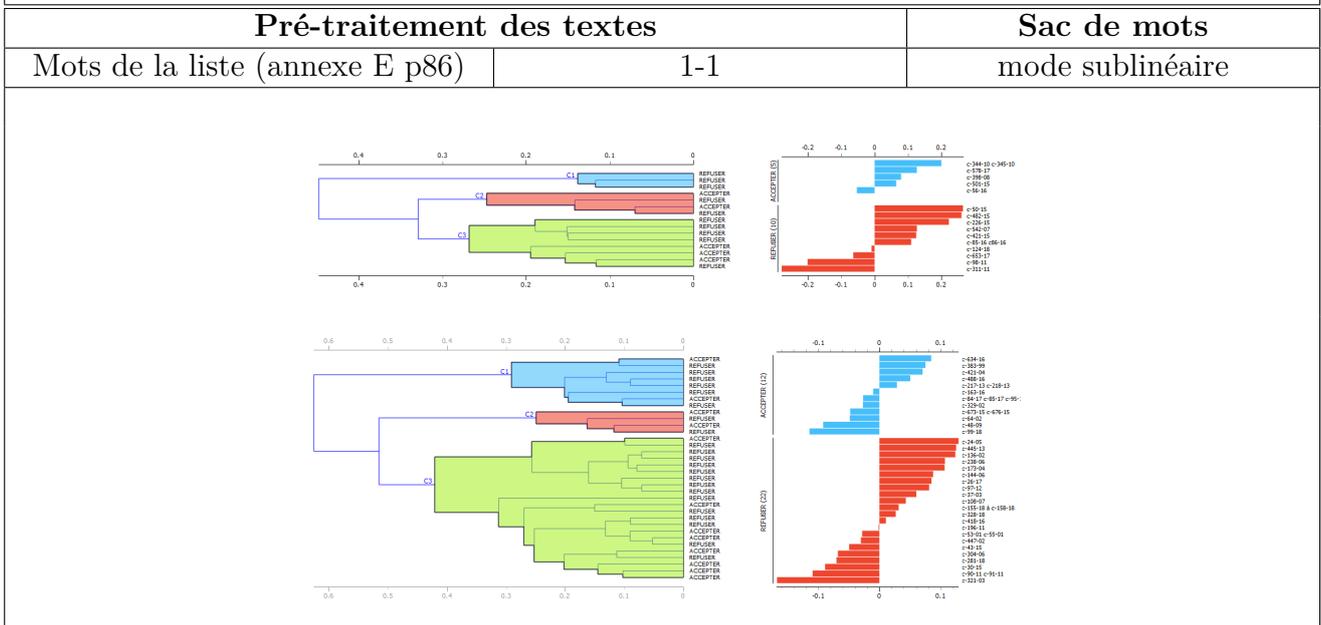
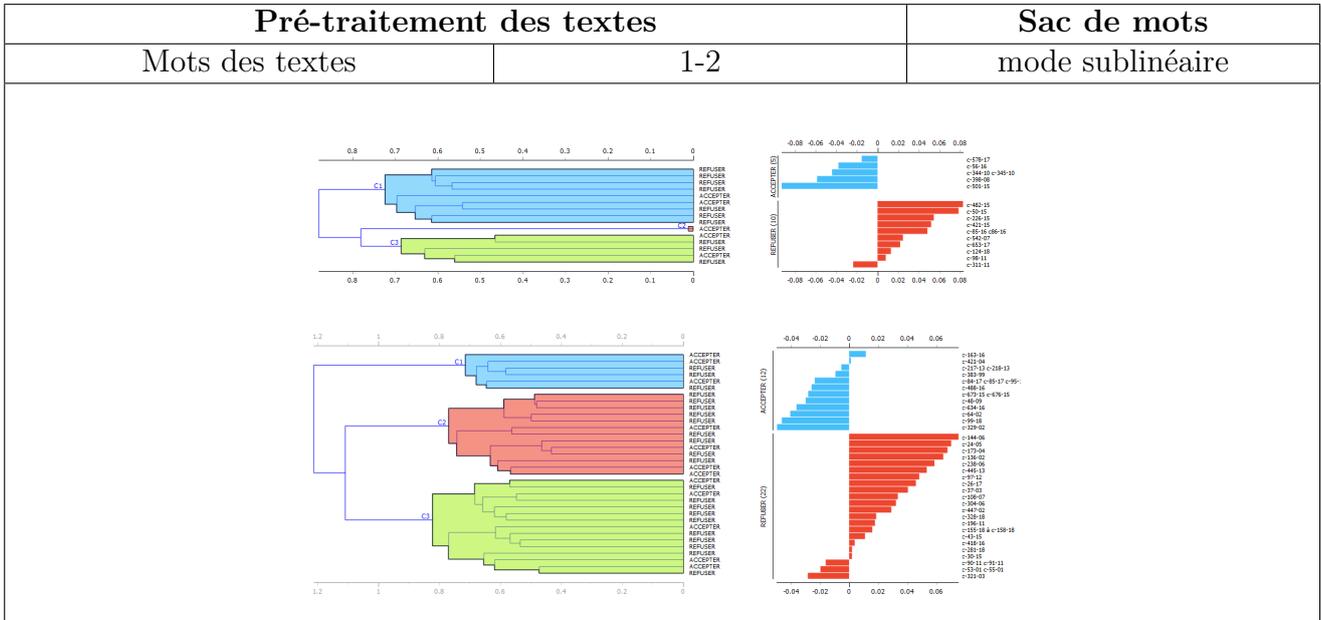


Annexe K

Représentation graphique des résultats obtenus sur le schéma repris à la figure 4.12 de la page 40 avec les textes repris dans la base de données "Accept Test" et "Accept Ex" (voir l'annexe G de la page 95)

Représentation des graphes en mode comptage et binaire avec une combinaison limitée à 1-1 et 1-2 sur les mots des textes et sur la liste de mots repris à l'annexe E à la page 86 pour le schéma repris à la figure 4.12 à la page 40 avec 2 jeux de données séparés en 2 catégories.





Annexe L

Affichage des valeurs reprises dans le fichier *valeurTest.csv*

référence	Quel sera le nom de la marque ou de sa référence	Quel est le format de la marque	Quelle est la langue utilisée pour la marque	Quels sont les pays concernés dans la demande	Quels sont les langues utilisées dans la zone territoriale de la demande	Quels sont les classes de produit ou de service de la demande	La marque a un caractère distinctif dans la classe de produit ou de service	Un public pertinent reconnaît-il la marque	N'y a-t-il pas de risque de conflit d'intérêt général	Le signe est-il identique ou partiellement identique à une marque a caractère descriptif de la classe de produit ou du contenant	Le caractère avant une demande de service	Le caractère distinctif par usage de produit ou de service	Le caractère distinctif par usage de produit ou de service	La marque utilise-t-elle un terme géographique	La marque utilise-t-elle un terme géographique	Le terme géographique est-il un produit ou un service	Le terme géographique utilise-t-il une appellation d'origine contrôlée
C-104/01 a	C-104/01 a	couleur	neerlandais	benelux	français neerlandais	9, 35, 38 non non non	oui	non	non	non	non	non	non	non	non	non	non
C-104/01 b	C-104/01 b	couleur	neerlandais	benelux	français neerlandais	9, 35, 38 non non non	oui	non	non	non	non	non	non	non	non	non	non
C-108/05 a	EUROPOLIS	verbe	neerlandais	benelux	français neerlandais	36, 39, non non	oui	non	non	non	non	non	non	non	non	non	non
C-108/07	FERRO	verbe	incendu	UE	tous	30 non	oui	non	non	non	non	non	non	non	non	non	non
C-108/97a	Chiemsee	semifigurative	aucune	Allemagne	allemand	25 oui	oui	non	non	non	non	non	non	non	non	non	non
C-108/97b	Chiemsee	semifigurative	aucune	Allemagne	allemand	25 non	oui	non	non	non	non	non	non	non	non	non	non
C-124/18	C-124/18	couleur	neerlandais	UE	tous	32 non	oui	non	non	non	non	non	non	non	non	non	non
C-136/02	C-136/02	tridimensionnelle	neerlandais	UE	tous	9, 11, non, non	oui	non	non	non	non	non	non	non	non	non	non
C-144/06	C-144/06	figurative	neerlandais	UE	tous	1, 3, 21 non non non	oui	non	non	non	non	non	non	non	non	non	non
C-163/16	C-163/16	complexe	neerlandais	UE	tous	25 non	oui	non	non	non	non	non	non	non	non	non	non
C-173/04	C-173/04	tridimensionnelle	neerlandais	UE	tous	32 non	oui	non	non	non	non	non	non	non	non	non	non
C-196/11	C-196/11	figurative	neerlandais	UE	tous	16, 38, 41 non non non	oui	non	non	non	non	non	non	non	non	non	non
C-215/14 a	C-215/14 a	complexe	neerlandais	Royaume-Unis	anglais	30 non	oui	non	non	non	non	non	non	non	non	non	non
C-215/14 b	C-215/14 b	complexe	neerlandais	Royaume-Unis	anglais	30 non	oui	non	non	non	non	non	non	non	non	non	non
C-217/13	C-217/13	couleur	neerlandais	UE	tous	36 non	oui	non	non	non	non	non	non	non	non	non	non
C-218/01 a	C-218/01 a	complexe	neerlandais	UE	tous	2 non	oui	non	non	non	non	non	non	non	non	non	non
C-218/01 b	C-218/01 b	complexe	neerlandais	UE	tous	2 non	oui	non	non	non	non	non	non	non	non	non	non
C-238/06	C-238/06	tridimensionnelle	neerlandais	UE	tous	29, 30, 32 non, non, non	oui	non	non	non	non	non	non	non	non	non	non
C-24/05	C-24/05	tridimensionnelle	neerlandais	UE	tous	30 non	oui	non	non	non	non	non	non	non	non	non	non
C-26/17	C-26/17	figurative	neerlandais	UE	tous	10, 18, 28 non, non, non	oui	non	non	non	non	non	non	non	non	non	non
C-299/99	C-299/99	tridimensionnelle	neerlandais	UE	tous	3, 8, 21 non, non, non	oui	non	non	non	non	non	non	non	non	non	non
C-300/15	C-300/15	tridimensionnelle	neerlandais	UE	tous	28 non	oui	non	non	non	non	non	non	non	non	non	non
C-304/06	EUROHYPO	verbe	allemand	UE	tous	36 non	oui	non	non	non	non	non	non	non	non	non	non
C-311/11	WIR MACHEN DAS BESONDERE EINFACHE	verbe	allemand	UE	tous	9 non	oui	non	non	non	non	non	non	non	non	non	non
C-321/03	C-321/03	tridimensionnelle	allemand	UE	tous	9 non	oui	non	non	non	non	non	non	non	non	non	non
C-329/02	SAT 2	verbe	allemand	UE	tous	30, 35, 38, 41, 42 oui oui oui c null	oui	non	non	non	non	non	non	non	non	non	non
C-342/97 a	Lloyd contre Loint	verbe	incendu	Allemagne	allemand	25 oui	oui	non	non	non	non	non	non	non	non	non	non
C-342/97 b	Lloyd contre Loint	verbe	incendu	Allemagne	allemand	25 oui	oui	non	non	non	non	non	non	non	non	non	non
C-344/70	C-344/70	tridimensionnelle	allemand	UE	tous	33 non	oui	non	non	non	non	non	non	non	non	non	non
C-37/03	BlöD	semifigurative	allemand	UE	tous	25 non	oui	non	non	non	non	non	non	non	non	non	non
C-383/99	Baby-dry	verbe	anglais	UE	tous	9, 38, 42 non non non	oui	non	non	non	non	non	non	non	non	non	non
C-398/08	Vorsprung durch Technik	verbe	allemand	UE	tous	24 non	oui	non	non	non	non	non	non	non	non	non	non
C-404/02 a	Nichols	verbe	allemand	UE	tous	40, 41, 42, 43, 45 non non non non non	oui	non	non	non	non	non	non	non	non	non	non
C-404/02 b	Nichols	verbe	allemand	Royaume-Unis	anglais	7 oui	oui	non	non	non	non	non	non	non	non	non	non
C-42/04	MATRATZEN	verbe	allemand	Royaume-Unis	anglais	1 non	oui	non	non	non	non	non	non	non	non	non	non
C-445/13	C-445/13	tridimensionnelle	allemand	Espagne	espagnole	20 non	oui	non	non	non	non	non	non	non	non	non	non
C-447/02	C-447/02	couleur	allemand	UE	tous	32, 33, non, non	oui	non	non	non	non	non	non	non	non	non	non
C-488/16	NEUSCHWANSTEIN	verbe	allemand	UE	tous	7, 11, 31, 42 non non non non non	oui	non	non	non	non	non	non	non	non	non	non
C-53/01	C-53/01	tridimensionnelle	allemand	UE	tous	34, 35, 36, 38, 44 oui oui oui c null	oui	non	non	non	non	non	non	non	non	non	non
C-542/07	PURE DIGITAL	verbe	anglais	UE	tous	1 non	oui	non	non	non	non	non	non	non	non	non	non
C-56/16	PORT CHARLOTTE	verbe	français	UE	tous	9, 38 non non	oui	non	non	non	non	non	non	non	non	non	non
C-64/02	DAS PRINZIP DER BEQUEMLICHKEIT	verbe	allemand	UE	tous	33 oui	oui	non	non	non	non	non	non	non	non	non	non
C-90/11	Multi Markets Fund MMF	verbe	allemand	UE	tous	8, 12, 20 non non non	oui	non	non	non	non	non	non	non	non	non	non
C-97/12	C-97/12	figurative	allemand	UE	tous	36 non	oui	non	non	non	non	non	non	non	non	non	non
C-98/11	C-98/11	figurative	allemand	UE	tous	9, 14, 18, 25 non non non non non	oui	non	non	non	non	non	non	non	non	non	non

référence	Quel sera le nom de la marque ou de sa référence	Quel est le format de la marque	La protection de l'AOC s'applique à la classe de produit	La marque est-elle similaire à une marque antérieure dans la même classe de produit ou service	La marque utilise un patronyme répandu dans la zone territoriale de la demande	La marque est-elle vaste pour permettre au public de reconnaître la marque comme appartenant à une seule entité	La taille de la demande est-elle trop petite	Le signe peut-il être utilisé par le concurrent	La marque est-elle de dimensionnelle divergente de la norme ou des habitudes du secteur	Le signe est-il représenté par les abréviations caractéristiques d'un signe	Le signe est-il habituel pour décrire le produit ou le service	La forme de la marque est-elle vue dans son ensemble possédant un caractère distinctif	Résultat
C-104/01 a	C-104/01 a	couleur	verbal	non	non	non	non	non	non	non	non	non	ACCEPTER
C-104/01 b	C-104/01 b	couleur	verbal	non	non	non	non	non	non	non	non	non	REFUSER
C-108/05 a	EUROPOLIS	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-108/05 b	EUROPOLIS	verbal	non	non	non	non	non	non	non	non	non	non	REFUSER
C-108/07	FERRO	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-108/97a	Chiemsee	semifigurative	non	non	non	non	non	non	non	non	non	non	REFUSER
C-108/97b	Chiemsee	semifigurative	non	non	non	non	non	non	non	non	non	non	REFUSER
C-124/18	Chiemsee	semifigurative	non	non	non	non	non	non	non	non	non	non	REFUSER
C-136/02	C-136/02	couleur	tridimensionnelle	non	non	non	non	non	non	non	non	non	REFUSER
C-144/06	C-144/06	figurative	non	non	non	non	non	non	non	non	non	non	REFUSER
C-163/16	C-163/16	complexe	tridimensionnelle	non	non	non	non	non	non	non	non	non	REFUSER
C-173/04	C-173/04	figurative	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-196/11	C-196/11	figurative	non	non	non	non	non	non	non	non	non	non	REFUSER
C-215/14 a	C-215/14 a	complexe	non	non	non	non	non	non	non	non	non	non	REFUSER
C-215/14 b	C-215/14 b	complexe	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-217/13	C-217/13	couleur	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-218/01 a	C-218/01 a	complexe	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-218/01 b	C-218/01 b	complexe	non	non	non	non	non	non	non	non	non	non	REFUSER
C-238/06	C-238/06	tridimensionnelle	non	non	non	non	non	non	non	non	non	non	REFUSER
C-24/05	C-24/05	tridimensionnelle	non	non	non	non	non	non	non	non	non	non	REFUSER
C-26/17	C-26/17	figurative	non	non	non	non	non	non	non	non	non	non	REFUSER
C-299/99	C-299/99	figurative	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-300/15	C-300/15	tridimensionnelle	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-304/06	EUROHYPO	verbal	non	non	non	non	non	non	non	non	non	non	REFUSER
C-311/11	WIR MACHEN DAS BESONDERE EINFACHER	verbal	non	non	non	non	non	non	non	non	non	non	REFUSER
C-321/03	C-321/03	tridimensionnelle	non	non	non	non	non	non	non	non	non	non	REFUSER
C-329/02	SAT 2	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-342/97 a	Lloyd contre Loint	verbal	non	oui	non	non	non	non	non	non	non	non	ACCEPTER
C-342/97 b	Lloyd contre Loint	verbal	non	oui	non	non	non	non	non	non	non	non	REFUSER
C-344/70	C-344/70	tridimensionnelle	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-37/03	BloID	semifigurative	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-383/99	Baby-dry	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-398/08	Vorsprung durch Technik	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-404/02 a	Nichols	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-404/02 b	Nichols	verbal	non	non	non	non	non	non	non	non	non	non	REFUSER
C-42/104	MATRATZEN	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-445/13	C-445/13	tridimensionnelle	non	non	non	non	non	non	non	non	non	non	REFUSER
C-447/02	C-447/02	couleur	non	non	non	non	non	non	non	non	non	non	REFUSER
C-488/16	NEUSCHWANSTEIN	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-53/01	C-53/01	tridimensionnelle	non	non	non	non	non	non	non	non	non	non	REFUSER
C-542/07	PURE DIGITAL	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-56/16	PORT CHARLOTTE	verbal	non	non	non	non	non	non	non	non	non	non	REFUSER
C-64/02	DAS PRINZIP DER BEQUEMLICHKEIT	verbal	non	non	non	non	non	non	non	non	non	non	ACCEPTER
C-90/11	Multi Markets Fund MMF	verbal	non	non	non	non	non	non	non	non	non	non	REFUSER
C-97/12	C-97/12	figurative	non	non	non	non	non	non	non	non	non	non	REFUSER
C-98/11	C-98/11	figurative	non	non	non	non	non	non	non	non	non	non	REFUSER