



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE IBN KHALDOUN - TIARET

MEMOIRE

Présenté à :

FACULTÉ MATHÉMATIQUES ET INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

MASTER

Spécialité : Génie Informatique

Par :

Sahraoui Mehdi

Et

Fendil Mourad

Sur le thème

Vers des stratégies efficaces pour la circulation routière

Soutenu publiquement le 01/12/2020 à Tiaret devant le jury composé de :

MrMEBAREK Bendaoud

Grade Université MCA

Président

Mr Hattab Nouredine

Grade Université MAB

Encadreur

Mr Bekki Khadhir

Grade Université MCB

Examineur

Remerciements

Je remercie Dieu « ALLAH » de nous avoir donné la force, et la volonté pour bien achever ce modeste travail.

En préambule, on souhaite adresser ici tous nos remerciements aux personnes qui m'ont apportée leur aide et qui ont ainsi contribué à l'élaboration de ce mémoire.

A Monsieur Mr.HATTAB. NOUREDDINE mon encadreur pour sa supervision, ses conseils scientifiques et sa disponibilité.

Notre profond respect va aux membres du jury: Mr. Mebarek Bendaoud et Mr.Bekki. Khadhir, pour avoir accepté de juger ce travail.

J'exprime ma gratitude à tous les internautes rencontrés (virtuellement) lors de mes recherches effectuées et qui ont accepté de répondre à mes questions, spécialement « Mr. OURED. Abdelkader »

Dédicace

Je dédie ce travail à ...

A ma très chère mère

Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi.

Votre prière et votre bénédiction m'ont été d'un grand secours pour mener à bien mes études.

Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur.

*A mon défunt père **يرحمه الله** Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous.*

Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation.

A mes tendres frères et sœurs, pour tant de confiance, d'amour, de patience et d'abnégation.

*A mon, chère Binôme **Sahraoui Mehdi**, ainsi qu'à toute sa famille. Pour toutes les peines et difficultés partagées ensemble lors de l'élaboration de ce projet. A tous mes amis et en particulière **Boucharef Yahia**.*

A tous les étudiants de notre promotion et à tous les gens qui m'ont aidé et soutenu tout au long de ce projet.

Fendil Mourad

Dédicace

Je dédie ce travail à ...

A ma très chère Méré

Quoi que je fasse ou que je dise, je ne saurai point te remercier comme il se doit. Ton affection me couvre, ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force pour affronter les différents obstacles

A mon très cher père

Tu as toujours été à mes cotes pour me, et mes soutenir et m'encourager.

A mes très chers frères sœurs. Qui n'ont pas cessée de me conseiller, encourager et soutenir tout au long de mes études. Que dieu les protège et leurs offre la chance et le bonheur

*Sans oublier mon chère binôme **Fendil Mourad** pour son soutien moral, sa patience et sa compréhension tout au long de ce projet, A tous mes amis et en particulière **Boucharef Yahia, Bilal bour** et à tous les étudiants de notre promotion et à tous les gens qui m'ont aidé et soutenu tout au long de ce projet.*

Sahraoui Mehdi

ملخص

في هذا العمل نحن مهتمون بدراسة حركة المرور على الطرق. الدافع يأتي من حقيقة أن هذه الدراسة تتطلب ورًا جديدًا، مما يجعل من الممكن الحد من المشكلات الناشئة عن الإزدحام في هذا المجال مثل حظر التقاطعات (مفترق الطرق).

الهدف من هذا العمل هو دمج مناهج التعلم الآلي للتحكم في إشارات المرور لتنظيم حركة المرور الأساسية، مع مراعاة القيود البيئية المهمة "حركة المرور المدروسة".

أخيرًا ، قارنا بين ثلاث استراتيجيات التعلم الآلي الخاضع للإشراف لدراسة الحركة المرورية.

كلمات دلالية:

حركة المرور على الطرق، التعلم الآلي ، التعلم الآلي الخاضع للإشراف.

Résumé

Dans ce travail nous nous intéressons au domaine de circulation routière. La motivation vient du fait que ce secteur nécessite de nouvelles perceptions, permettant de freiner l'augmentation des problèmes issus de ce domaine tel que le blocage au niveau des intersections (carrefour).

L'objectif ultime de ce travail est d'intégrer les approches d'apprentissage automatique dans l'automatisation des feux de circulation dans une circulation routière basique, en tenant compte des contraintes environnementales de la tâche "circulation étudiée".

Finalement, nous avons comparés entre trois stratégies d'apprentissage automatique supervisé pour notre tâche de circulation étudiée.

Mots clés : Circulation routière, Apprentissage automatisée, Apprentissage automatisé supervisé.

Abstract

In this work we are interested in the field of road traffic. The motivation comes from the fact that this sector requires new perceptions, making it possible to slow down the increase in problems arising from this field such as blocking at intersections (crossroads).

The ultimate objective of this work is to integrate machine learning approaches in the automation of traffic lights in basic road traffic, taking into account the environmental constraints of the "traffic studied" task. Finally, we compared between three supervised learning strategies for our studied circulation task.

Keywords: Road Traffic, Automated learning, Supervised automated learning.

Tables Des Matières

Dédicaces

Remerciements

Résumé

Abstract

Liste des figures

Liste des tableaux

Liste des abréviations

Sommaire

Chapitre I : Introduction générale

1	Introduction générale :	10
1.1	Contexte :	10
1.2	Problématique :	10
1.3	Objective de travail :	11
1.4	Organisation du travail :	12

Chapitre II :La Circulation routière

1.1	Introduction :	14
1.2	Circulation routière :	14
1.2.1	Composantes de la circulation routier :	14
1.2.1.1	L'infrastructure :	14
1.2.1.2	Les mobiles :	14
1.2.2	Variables élémentaires du trafic routier :	15
1.2.2.1	L'écart du temps inter-véhiculaire.....	15
1.2.3	La congestion routière	16
1.2.3.1	Conséquences de la congestion de la circulation routier :	16

1.2.3.1.1	Conséquences économiques :	16
1.2.3.1.2	Conséquences sociétales :	17
1.2.3.1.3	Conséquences environnementales :	17
1.3	Des solutions pour réduire la circulation routière :	17
1.4	Le code de la route :	18
1.4.1	Règles de priorité :	18
1.4.2	Les panneaux de signalisation :	19
1.4.3	Les feux de circulation tricolores :	19
1.5	Feux de circulation intelligents :	20
1.5.1	Contrôle dynamique par l'infrastructure.....	20
1.5.1.1	Première génération : contrôle à temps fixe.....	20
1.5.1.2	Deuxième génération : contrôle à temps dynamique :.....	20
1.5.1.3	Troisième génération : contrôle à temps réel :	21
1.5.1.3.1	Contrôle réactif :	21
1.5.1.3.2	Contrôle adaptatif :	21
1.6	Conclusion :	22

Chapitre III : L'apprentissage automatique proposée pour Les feux de signalisation

1	L'apprentissage automatique proposée pour les feux de signalisation.....	24
1.1	Introduction :	24
1.2	Présentation de notre approche :	24
1.2.1	Les étapes de notre approche	25
1.2.1.1	Création des données d'apprentissage :	26
1.2.1.1.1	Smart+Connected™ Traffic de Cisco :	26
1.2.1.1.1.1	Historique de Cisco :	26
1.2.1.1.1.2	Définition de Smart+Connected™ Traffic de Cisco :	26
1.2.1.1.1.3	Principales caractéristiques de Smart+Connected™ Traffic de Cisco :	26

1.2.1.1.1.4	Architecture de Smart+Connected™ Traffic :.....	28
1.2.1.1.2	Présentation du simulateur « SimCirc » :.....	29
1.2.1.1.3	Traitement de data set	29
1.2.1.1.3.1	Traitement les données avec Foda (Feature-Oriented Domain Analysis)	30
1.2.1.1.3.1.1	Feature-Oriented Domain Analysis (Foda)	30
1.2.1.1.3.1.2	Modélisation de la variabilité :.....	30
1.2.1.2	Apprentissage automatique (Machine Learning) :.....	35
1.2.1.2.1	Les types d'apprentissage automatique.....	36
1.2.1.2.1.1	Apprentissage supervisé :	36
1.2.1.2.1.2	Apprentissage non supervisé :	36
1.2.1.2.1.3	Apprentissage par renforcement :	36
1.2.1.2.2	Classification :	37
1.2.1.2.3	Les algorithmes de la classification :	37
1.2.1.2.3.1	Forêt aléatoire (Radom Forest)	37
1.2.1.2.3.2	Machine à Vecteur de Support (SVM) :	40
1.2.1.2.3.3	Réseau Bayésien (Naïve bayes) :.....	41
1.3	Conclusion.....	42

Chapitre IV : L'implémentation

1	L'implémentation	44
1.1	Introduction	44
1.2	Environnement de développement et Les outils techniques	44
1.2.1	Dentition du langage Python en informatique :	44
1.2.2	Anaconda	44
1.2.3	Jupiter	45
1.2.4	Les bibliothèques utilisées :.....	45
1.2.4.1	Numpy :	45
1.2.4.2	Matplotlib :	45
1.2.4.3	Pandas :.....	45

1.2.4.4 Scikit-learn :	45
1.3 Apprentissage et Paramétrage des Modèles :.....	46
1.3.1 Préparation du Dataset.....	46
1.3.2 Application de PCA :	46
1.3.3 Algorithme LDA.....	48
1.4 Apprentissage du modèle	48
1.4.1 Algorithme de Radom Forest.....	48
1.4.2 Algorithme de Réseau Bayésien :	49
1.4.3 Algorithme de SVM :	51
1.5 Comparaison entre Les Trois stratégie :.....	52
1.6 Conclusion.....	55
Conclusion générale & Perspectives.....	57
Bibliographie	58

Liste des figures

Figure 1 : L'écart du temps inter-véhiculaire.....	15
Figure 2 :La congestion d'un réseau routier.....	16
Figure 3 :Exemple d'un plan de feu.....	20
Figure 4 : L'apprentissage automatique pour les feux de signalisation intelligents qui s'adaptent en temps réel.....	21
Figure 5 : Présentation de notre approche.....	24
Figure 6 : Principales caractéristiques de Smart+Connected™ Traffic de Cisco .	26
Figure 7 : Vue d'ensemble de l'architecture de Smart+Connected™ Traffic	28
Figure 8 :Sim Circ (net logo)	29
Figure 9 :Présentation Les caractéristiques de la circulation.....	30
Figure 10 : Les caractéristiques de la circulation routière	31
Figure 11 : Les caractéristiques de la route.....	31
Figure 12 : Les caractéristiques de la distance.....	32
Figure 13 : Les caractéristiques de la Véhicule	32
Figure 14 : Les caractéristiques de la vitesse.....	33
Figure 15 : Les caractéristiques de la Type.....	33
Figure 16 : Les caractéristiques des personnes	34
Figure 17 : Les caractéristiques du feu de signalisation	34
Figure 18 : L'espace de variante globale de circulation routière.....	35
Figure 19 : Apprentissage automatique	36
Figure 20 : Les types d'apprentissage automatique	37

Liste des figures

Figure 21 : Architecture générale du processus de classification.....	37
Figure 22 : l'algorithme de foret aléatoire	38
Figure 23 :Procédure de construction de foret aléatoire	39
Figure 24 : hyperplan optimal, vecteurs de support et marge maximale.....	40
Figure 25 : L'algorithme de SVM	40
Figure 26 : L'algorithme de Réseau Bayésien	42
Figure 27 : L'environnement Anaconda.....	44
Figure 28 : L'exécution de l'algorithme PCA	47
Figure 29 : L'exécution de l'algorithme LDA.....	48
Figure 30 : Comparaison entre les trois stratégies par apport Accuracy	54
Figure 31 : Comparaison entre les trois stratégies par apport la probabilité	55

Liste des tableaux

Tableau 1 : Matrice de confusion de Foret Aléatoire	50
Tableau 2 : Matrice de confusion de l'algorithme Réseau Bayésien	51
Tableau 3 : Matrice de confusion de l'algorithme SVM	53
Tableau 4 : Comparaison entre les trois stratégies par apport Accuracy, Score ..	56

Abréviations

IA	Intelligence Artificiel
AA	Apprentissage Automatique
ML	Machine Learning
IDE	Integrated Development Environment
FODA	Feature-Oriented Domain Analysis
UTCS	Urban Traffic Control Software
SVM	Machine à Vecteurs de Support
NB	Naïve Bayes
RF	Random Forest

Chapitre I:

Introduction générale

1.1 Contexte :

L'amélioration du niveau de vie moyen et du taux d'équipements des ménages a permis au plus grand nombre d'accéder au déplacement en véhicules particuliers, induisant une circulation routière de plus en plus dense. C'est ainsi que les infrastructures ont été dimensionnées pour que leurs capacités d'écoulement correspondent au mieux à la demande, sans cesse croissante, surtout très marquée aux heures les plus chargées.

1.2 Problématique :

La particularité du trafic automobile est que l'infrastructure routière est, en fait, conçue selon une demande projetée à un certain instant pour répondre à un optimum collectif alors que chaque individu cherche à optimiser son déplacement tout en satisfaisant des critères placés au niveau individuel. Cette particularité fait du trafic automobile un phénomène difficile à analyser et à optimiser; et même si les modèles de simulation constituent des outils essentiels pour son analyse et sa compréhension; ce phénomène possède un certain nombre de caractéristiques qui le rendent également difficile à modéliser et génère des problèmes au niveau de la fluidité du trafic pouvant être à l'origine de l'apparition de phénomènes de congestion et de formation des files d'attente avec les retards correspondants.

Par ailleurs, les retards causés par les congestions ont un impact important sur la qualité de vie des usagers. En effet, les conducteurs, qui fréquentent les embouteillages, sont assez souvent sujets à des problèmes de stress, de bruits, etc., et à une augmentation de risque d'accidents. Ce phénomène, largement identifié comme une source indirecte de problèmes de santé, a de plus des conséquences directes sur l'augmentation de la consommation énergétique.

Le phénomène de congestion du trafic routier est ainsi un problème socio-économique crucial qui nécessite de rechercher des solutions adaptées pouvant être mises en œuvre en pratique. Les programmes d'élargissement des infrastructures étant, dans le même temps, onéreux, techniquement et juridiquement difficiles à mettre en œuvre, l'optimisation de l'utilisation des infrastructures existantes, tant en terme de capacité qu'en terme de sécurité, s'avère donc comme étant un enjeu économique et social important. Elle n'est alors possible que par une meilleure compréhension des mécanismes régissant l'ensemble des phénomènes de trafic. Il est ainsi devenu incontournable de s'intéresser aux systèmes de régulation dynamique du trafic dont le but est d'exploiter plus efficacement en temps réel les infrastructures existantes par l'emploi de techniques qui ont fait leurs preuves. Néanmoins, la commande du trafic, présentant encore plusieurs difficultés du fait de la complexité du problème à traiter, reste à ce jour un problème d'actualité et

de nombreux centres de recherche en ont fait leur priorité. Dans ce contexte, la commande intelligente du trafic peut constituer un moyen efficace pour éliminer ou du moins alléger les effets de la congestion et de la formation des files d'attente.

La méthode classique adoptée, pour prévenir ces effets ou les réduire dans les villes modernes, est basée sur l'élaboration d'une signalisation routière adaptée. Le droit de passage est attribué ainsi aux usagers par l'utilisation de feux de couleurs standards (rouge/orange/jaune-vert) qui peut permettre de résoudre les conflits entre les différents flux de circulation aux carrefours. Le contrôle de la signalisation du trafic au niveau des carrefours se divise, généralement, en deux catégories : la stratégie de contrôle à plan de feux fixe, c'est-à-dire avec un cycle fixé, et la stratégie de contrôle adaptative, permettant des changements de durées de phases en fonction de la demande du trafic. Les deux stratégies sont généralement basées sur l'estimation du débit du trafic.

De manière schématique, un système de contrôle dynamique du trafic est un système qui dispose de capteurs lui fournissant des informations sur l'état du trafic et, notamment, le nombre de véhicules sur certaines rues ou intersections du réseau routier. Ces informations sont alors transmises à un ordinateur qui a pour rôle de donner aux commandes les valeurs de paramètres minimisant une certaine fonction exprimant la congestion du réseau routier, comme la longueur des files d'attente, le temps moyen d'attente, le temps moyen d'attente de la plus longue file d'attente et le temps d'évacuation globale.

1.3 Objectif de travail :

Dans ce mémoire, nous nous sommes intéressés à la commande des feux de carrefour en vue de résoudre les problèmes liés à la congestion ; dans ce sens, nous proposons une stratégie pour contrôler les feux de signalisation avec l'apprentissage automatique ML (Machine Learning en anglais). Le modèle de prédiction est basé sur trois algorithmes de ML : Forêt aléatoire (RF), système support vecteur machine (SVM) et réseaux bayésiens (RB). Le modèle obtenu permet de prédire la décision afin d'agir dans la circulation en fonction de plusieurs paramètres comme la rue, la voiture, les passages etc.

1.4 Organisation du travail :

Ce mémoire comporte quatre chapitres :

- **Le premier chapitre** : inclut une introduction générale et problématique et objective de travail et organisation du travail.
- **Le second chapitre** : nous présentons la circulation routière.
- **Le troisième chapitre** : concerne l'apprentissage automatique proposée pour les feux de signalisation.
- **Le quatrième chapitre** : l'implémentation.
- **Conclusion et Perspective** : Finalisé notre travail avec un résumé de ce mémoire et la rédaction de quelques perspectives envisageables de ce sujet.

Chapitre II :

La Circulation routière

1.1 Introduction :

La route occupe une place privilégiée dans le secteur du transport, et, joue un rôle central dans les déplacements des biens, et des personnes pour cela de nombreux travaux de recherche à travers le monde portent sur le transport routier. La motivation vient du fait que ce secteur nécessite de nouvelles applications de méthodes permettant de contrecarrer l'augmentation accrue des problèmes issus de la circulation routière, que ce soit, en vue de réduire le nombre d'accidents, de diminuer la circulation dans les grandes agglomérations ou de réduire la pollution. En effet, l'un des facteurs favorables à cette problématique est dû à l'accroissement du nombre exponentiel de véhicules sur la route.

1.2 Circulation routière :

Une particularité du trafic automobile est essentiellement liée à la conception de l'infrastructure qui est conçue, selon une demande projetée, pour répondre à un optimum collectif mais que chaque individu réalise sans déplacement en cherchant à atteindre son optimum individuel, ce qui est souvent antagoniste avec l'optimum collectif. Les deux définitions ci-dessous montrent bien le double aspect individuel et collectif du trafic, l'aspect individuel étant représenté par le fait que le trafic est formé par différents véhicules et l'aspect collectif par le fait que le trafic est interprété comme un fluide. [1]

Alors le Circulation routier est la circulation de nombreux véhicules sur un réseau.

1.2.1 Composantes de la circulation routier :

En général, le trafic est constitué de deux composantes majeures : l'infrastructure et les mobiles.

1.2.1.1 L'infrastructure :

Est un ensemble d'éléments structuraux interconnectés qui fournissent le cadre pour supporter la totalité de la structure du trafic. Par exemple, pour les réseaux routiers et autoroutiers, l'infrastructure contient les structures : ponts, ponceaux, signalisation et marquages, systèmes électriques et les traitements de bord (bordures, trottoirs, aménagement paysager).

1.2.1.2 Les mobiles :

Sont les individus qui utilisent l'infrastructure, principalement dit, des véhicules. Les piétons ne sont pas considérés car leurs mouvements sont supposés dépendre de celui des mobiles.

Notons que les véhicules à l'arrêt ou en stationnement, au sens défini par le code de la route, sont considérés comme faisant partie de l'infrastructure et ne sont généralement pas pris en compte. Autrement dit, seuls les mobiles en mouvement sur la route et en interaction avec la signalisation sont considérés. Dans les sous-sections suivantes, nous introduisons les variables élémentaires utilisées et les mesures usuelles pour l'analyse du trafic. [2].

1.2.2 Variables élémentaires du trafic routier :

Le développement considérable de la circulation automobile et l'accroissement constant de la motorisation ont rendu nécessaire l'élaboration des théories du trafic, capables de décrire, d'expliquer et de prévoir les interactions entre les véhicules ainsi que les mouvements d'ensemble sur les infrastructures routières.

Ces théories, déterministes ou probabilistes s'introduisent au moyen d'un certain nombre de variables, de relations ou de distributions caractéristiques.

1.2.2.1 L'écart du temps inter-véhiculaire

L'écart du temps inter-véhicule (en anglais, « The headway »). En un point donné de la route, c'est le temps qui s'écoule entre les instants de passage de deux véhicules successifs. Cette variable se révèle particulièrement utile dans les études relatives à la sécurité routière et dans de nombreux problèmes de simulation du trafic. Certains algorithmes de régulation de carrefours à feux en ville, utilisent ce temps inter-véhiculaire.

Ce paramètre permet d'appréhender les effets de la composition du trafic sur les conditions d'écoulement de la circulation.

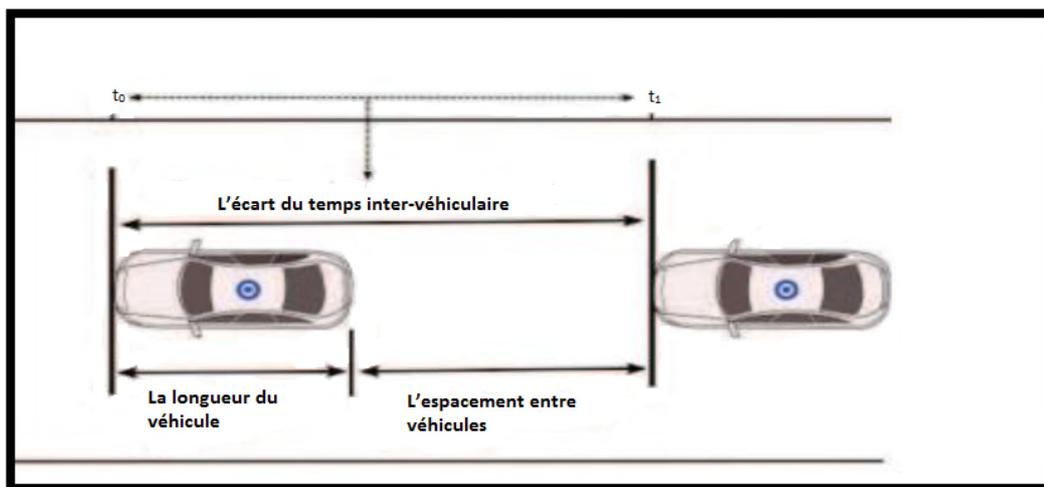


Figure 1 :L'écart du temps inter-véhiculaire

Malgré les progrès techniques et technologiques accomplis par l'homme dans tous les domaines de la connaissance, la circulation routière reste victime d'une congestion sans cesse croissante.

1.2.3 La congestion routière

La congestion du trafic est une notion intuitivement simple mais dont l'analyse rigoureuse s'avère en fait assez complexe [3]. C'est un phénomène physique concernant la façon dont des véhicules empêchent la progression des uns et des autres au fur et à mesure que la demande d'un espace routier s'approche de la capacité maximale de celui-ci [3].

La congestion d'un réseau routier est la condition dans laquelle une augmentation du trafic de véhicules provoque un ralentissement global de celui-ci. Le terme de congestion désigne la dégradation de la qualité de service quand le nombre d'utilisateurs augmente. Ce phénomène se caractérise par l'apparition de retards, voire de goulets d'étranglement en période de fort trafic, c'est-à-dire quand la capacité de l'infrastructure devient insuffisante pour réguler les flux. Le problème est fréquent localement et périodiquement, notamment dans les grandes villes et lors des grands départs pour les vacances [4].

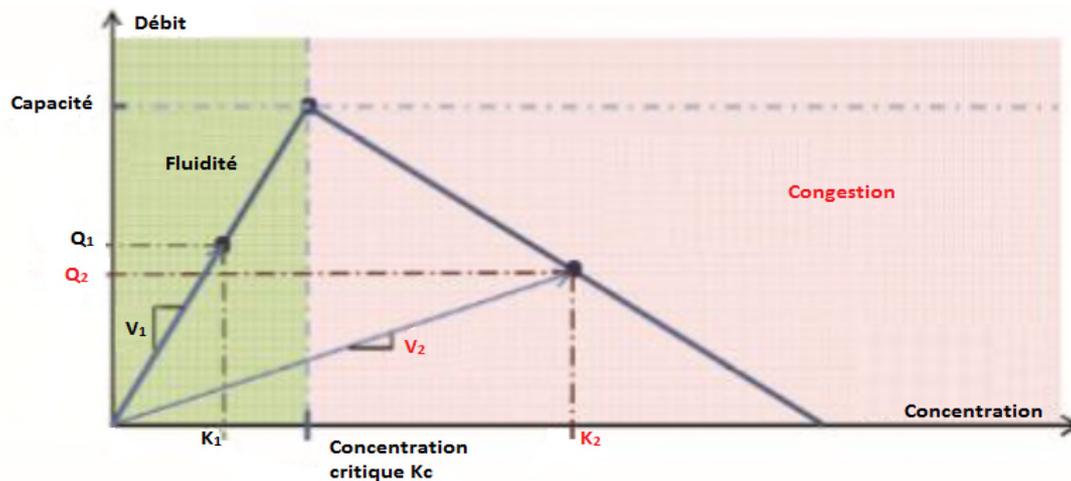


Figure 2 : La congestion d'un réseau routier.

Le trafic est dit congestionné s'il évolue dans la partie droite du diagramme, c'est-à-dire si la concentration est supérieure à la concentration critique (concentration correspondant au débit maximum) [5].

1.2.3.1 Conséquences de la congestion de la circulation routier :

Elles sont nombreuses et peuvent être classées en trois catégories : [4]

1.2.3.1.1 Conséquences économiques :

D'un point de vue économique, la congestion du trafic routier entraîne :

- Une perte de compétitivité d'une région, d'un pays ou d'un continent, du fait des congestions qui bloquent les artères de communication routière.
- Une perte de productivité, non seulement à cause des retards des personnes se rendant sur les lieux de leur travail, mais aussi en raison des livraisons tardives des approvisionnements ou des services rendus avec des retards importants voire annulés.

- En Europe, les embouteillages en ville et dans les zones urbaines coûtent chaque année près de 100 milliards d'euros, soit 1 % du PIB de l'UE1. [6].

1.2.3.1.2 Conséquences sociétales :

Les usagers de la route se plaignent amèrement des heures qu'ils perdent dans les embouteillages (plus de 1 milliard d'heures par an pour la France) parce qu'elles empiètent sur le temps qu'ils peuvent consacrer à leur travail et à leur famille.

La congestion du trafic routier aura deux impacts importants sur :

- La qualité de vie des usagers, car les conducteurs soumis à des embouteillages fréquents sont sujet à des problèmes de stress, d'anxiété et d'énervement, entraînant un accroissement du risque d'accident.
- Le pouvoir d'achat des usagers, car la congestion du trafic a des conséquences directes sur l'augmentation de la consommation énergétique. Dans le contexte actuel d'envolée des prix du pétrole, ceci engendre une nette diminution du pouvoir d'achat des usagers.
- dans 25 % des cas, les embouteillages ne sont pas récurrents, mais causés par un incident. [7]

1.2.3.1.3 Conséquences environnementales :

- Enfin, d'un point de vue environnemental, la congestion du trafic routier a un impact néfaste et négatif à cause de l'augmentation de la pollution qui engendre des coûts économiques et écologiques en croissance exponentielle.
- Une étude réalisée récemment, indique que le tiers des automobilistes passent de 1 à 2 heures dans le trafic chaque jour. Selon l'étude, alors qu'elles couvrent à peine 65 km, chaque année, «ces artères font perdre collectivement 22 millions de litres d'essence et génèrent 58 000 tonnes de Gas excédentaires».[8]

1.3 Des solutions pour réduire la circulation routière :

On parle souvent de solutions à la congestion en termes de grandes entreprises, d'autoroutes et de métros. Oui, nous devons améliorer nos infrastructures et nos systèmes de transport en commun, mais en même temps, il existe encore toute une série de solutions simples et efficaces que nous pouvons utiliser pour améliorer la circulation sur les routes et éviter les embouteillages croissants

- Dans une bretelle d'accès à l'autoroute, la façon la plus efficace de s'insérer dans une circulation dense est de le faire au dernier moment, chacun son tour, pour ne pas bloquer la voie et pour utiliser 100 % de l'espace. La longueur de la file d'attente peut ainsi être réduite de 40 % ;[8]

- Installer des paravents pour bloquer la vue aux automobilistes qui circulent à proximité d'une scène d'accident pour limiter les ralentissements et les incidents secondaires. En moyenne, les conducteurs observent une scène d'accident et ralentissent pendant 12 secondes. Avec les paravents, c'est plutôt 4 secondes. Des essais au Royaume-Uni ont permis d'établir les retombées économiques à 300 000 \$ pour chacun des incidents lors desquels les paravents ont été utilisés [8]
- Des feux de circulation intelligents qui s'adaptent en temps réel aux routines et aux incidents de circulation afin d'améliorer la fluidité.
- Respecter le code de la route.

1.4 Le code de la route :

Le code de la route est le recueil, écrit ou numérisé de toute la réglementation qui concerne la circulation sur les routes et les chaussées par tous les usages qui les empruntent.

1.4.1 Règles de priorité :

Par exemple le code de la route français : Lorsque deux conducteurs abordent une intersection par des routes différentes, le conducteur venant par la gauche est tenu de céder le passage à l'autre conducteur. Sauf dispositions différentes prévues.

Ceci implique une règle simple à respecter à chaque intersection sans signalisation souvent appelée priorité à droite.

Nous trouvons aussi : Tout conducteur abordant un carrefour à sens giratoire est tenu, quel que soit le classement de la route qu'il s'apprête à quitter, de céder le passage aux usagers circulant sur la chaussée qui ceinture le carrefour à sens giratoire. Cette règle annule donc l'effet de la première règle de priorité à droite et définit une nouvelle règle de priorité à gauche appliquée uniquement pour les carrefours à sens giratoire souvent appelés rond-point.

Ces règles ne sont pas communes à tous les pays. Elles diffèrent d'un pays à un autre et surtout dans les pays où les véhicules circulent à gauche sur la chaussée.

Par exemple, au Royaume-Uni, la priorité aux ronds-points est pour les véhicules qui viennent de droite et il n'existe pas de règle générale pour les intersections, la priorité étant généralement indiquée.

Notons que les règles de priorité du code de la route sont une forme simple de coordination réactive [09].

1.4.2 Les panneaux de signalisation :

En France, il existe des panneaux pour la gestion des intersections. Par exemple, parmi les panneaux les plus répandus nous trouvons les panneaux de signalisation de céder le passage.

A certaines intersections indiquées par une signalisation dite "cédez le passage", tout conducteur doit céder le passage aux véhicules circulant sur l'autre ou les autres routes et ne s'y engager qu'après s'être assuré qu'il peut le faire sans danger. Il est souvent utilisé pour définir une route principale par rapport à une route secondaire et n'oblige pas les usagers de la route secondaire à l'arrêt. Il existe aussi un autre panneau très répandu dans presque tous les pays du monde : le panneau (STOP). Si un conducteur trouve ce panneau implanté à une intersection sur son chemin, il doit obligatoirement marquer l'arrêt avant de continuer. Il indique que chaque véhicule sur cette voie n'est jamais prioritaire quelle que soit la direction qu'il souhaite prendre. Ce panneau est généralement utilisé pour contrôler le trafic aux intersections dangereuses, où le trafic est dense, mais pas suffisamment encombrées pour installer un système de feux de signalisation [09].

1.4.3 Les feux de circulation tricolores :

Dans les systèmes de gestion du trafic, le système le plus connu et le plus utilisé dans le monde est le système des feux de signalisation.

Un feu de circulation routière est un dispositif permettant la régulation du trafic routier entre les usagers de la route, les véhicules et les piétons. Les feux destinés aux véhicules à moteurs sont généralement de type tricolore, auxquels peuvent s'ajouter des flèches directionnelles. Ceux destinés aux piétons sont bicolores et se distinguent souvent par la reproduction d'une silhouette de piéton.

Les feux tricolores pour cyclistes se distinguent par la reproduction d'une bicyclette. Généralement, un feu tricolore est composé d'un système électronique commandé.

Il est composé de trois couleurs principales :

- La couleur rouge indique l'obligation d'arrêt aux véhicules.
- La couleur orange qui ne dure que quelques secondes signalent le passage du rouge au vert.
- La couleur verte indique aux véhicules qu'ils ont la priorité exclusive pour passer.

Ces couleurs ont été choisies parce qu'elles ont l'avantage d'être très distinctes.

Le système des feux de signalisation est le système le plus efficace pour la gestion du trafic car il évite tout malentendu entre les différents conducteurs au moment du passage [09].

Afin de régir le fonctionnement des feux de circulation, les contrôleurs utilisent des mécanismes combinatoires et temporels bien spécifiques.

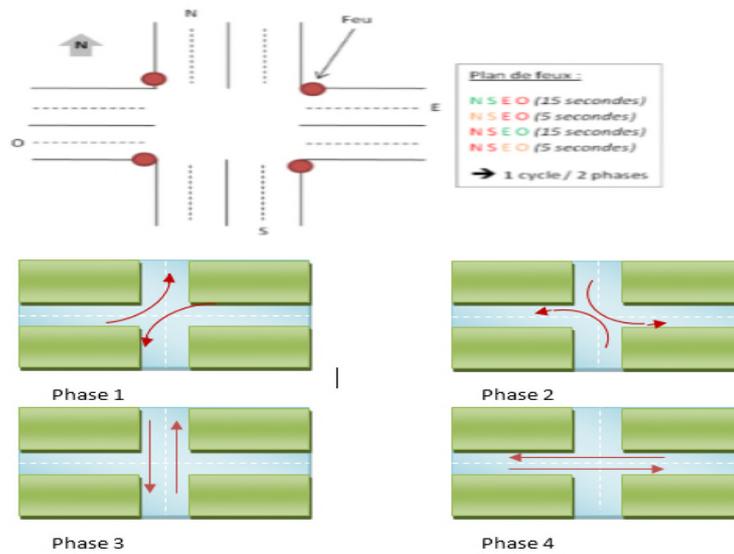


Figure 3 :Exemple d'un plan de feu

1.5 Feux de circulation intelligents :

Le développement du trafic routier urbain a amené à concevoir des systèmes intelligents. L'une des particularités de ces systèmes est leur besoin de dynamisme et de réactivité : afin de pouvoir agir sur une situation, ces derniers ont besoin de connaître des informations sur ce qu'il se déroule à des endroits bien précis, et prendre des décisions en conséquence.

1.5.1 Contrôle dynamique par l'infrastructure

La gestion des feux de circulation est un problème qui a commencé à être étudié au début des années 1970 avec l'apparition progressive de systèmes de gestion centralisés, en charge des carrefours d'une zone géographique donnée. Au fil des années, ces systèmes se sont diversifiés, et ont connu trois générations de contrôle.[10].

Aujourd'hui, ces générations peuvent être utilisées chacune en fonction des moyens mis en place sur l'infrastructure, et de la connaissance de cette dernière :

1.5.1.1 Première génération : contrôle à temps fixe

En fonction de l'heure et parfois du jour, le système va utiliser un plan de feux prédéfini.

1.5.1.2 Deuxième génération : contrôle à temps dynamique :

Des détecteurs sont utilisés afin de recueillir les données du trafic toutes les X minutes. Ces données peuvent être utilisées afin d'optimiser ou mettre en place un plan de feu.

Par exemple, une fois les phases déterminées, l'un des enjeux va être de définir un temps de feu vert pour chacune d'entre elles : ce dernier est généralement constitué d'une valeur minimale et maximale, afin de ne pas provoquer des arrêts intempestifs ou d'engendrer un cycle trop long. Le feu vert minimum suffit uniquement si pendant son temps d'exécution, aucun nouveau véhicule ne franchit un détecteur.

1.5.1.3 Troisième génération : contrôle à temps réel :

Reprend le même principe que la deuxième génération mais cette fois-ci en temps réel. Notons que les deux dernières générations, qui introduisent un caractère dynamique au système, peuvent être chacune décomposées en deux types :

1.5.1.3.1 Contrôle réactif :

En fonction des données recueillies sur le terrain périodiquement (plusieurs seconds ou cycles), le système met en place une nouvelle configuration en réponse aux informations reçues. Cette méthode est le premier niveau de dynamique, et est simple à mettre en place, mais nécessite toutefois une très bonne connaissance du système afin d'être efficace. C'est également la première méthode à être apparue aux Etats-Unis à la fin des années 1980 avec l'apparition des UTCS (Urban Traffic control software) [11].

1.5.1.3.2 Contrôle adaptatif :

Ce type de contrôle va programmer dynamiquement les plans de feux en se servant des paramètres recueillis sur le terrain, ceci en calculant des valeurs telles que le temps de cycle, des phases ou encore leur ordonnancement. L'opération va être effectuée de manière adaptative, c'est à dire en quasi-temps-réel.

L'avantage de cette méthode est qu'elle peut s'adapter à de multiples situations,

- Nous pouvons spécifier les niveaux de nos propositions de feux de signalisation dans la figure suivante :

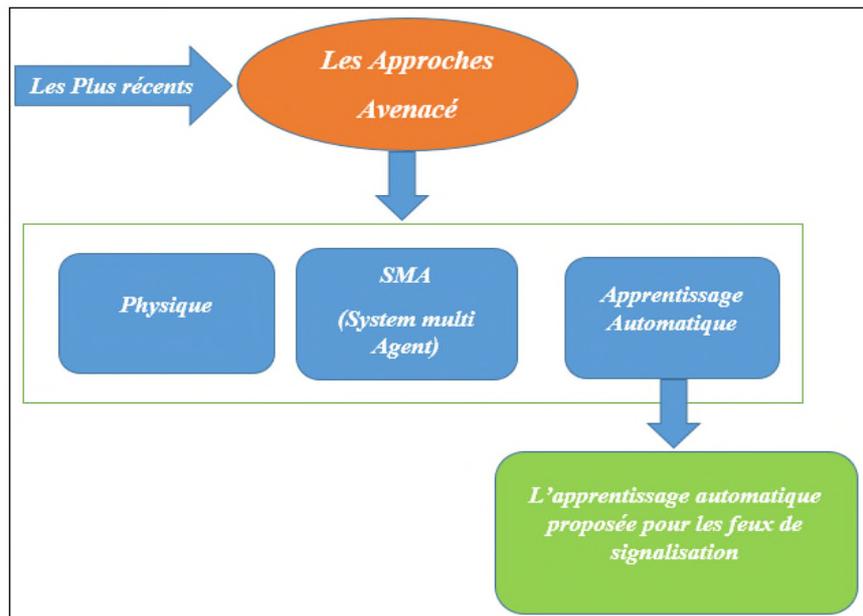


Figure 4 : L'apprentissage automatique pour les feux de signalisation intelligents qui s'adaptent en temps réel

Notre proposition exploite les données historiques de la circulation afin de proposer un modèle qui retourne la décision pour organiser la gestion de la circulation.

1.6 Conclusion :

Ce chapitre donne un aperçu du concept de circulation routière, les composants et les variables élémentaires de cette dernière ainsi il représente la congestion et les raisons qui la conduisent Sans oublier les conséquences qui résultent. Notamment les méthodes pour gérer ce problème « la circulation routière » tel que le code de la route les feux tricolores dans les intersections, et des solutions pour réduire la circulation routière. Enfin., nous donnons une solution d'apprentissage automatique pour les feux de signalisation intelligents qui s'adaptent en temps réel.

Chapitre III :

**L'apprentissage automatique
proposée pour les feux de
signalisation**

1.1 Introduction:

La gestion de carrefour dépend fortement des infrastructures routières comme principal moyen de transport, nous sommes constamment confrontés au problème des réseaux routiers encombrés et inefficaces. La majorité des recherches menées dans ce domaine ont abordé le problème de ce qui pourrait être décrit de manière unidimensionnelle, en essayant d'acheminer le trafic de manière plus efficace à l'aide de feux de signalisation intelligents.

1.2 Présentation de notre approche :

La Figure 5 illustre la vue globale de notre approche. Nous avons proposé un modèle de prédiction (système de feux de signalisation intelligence) basé sur L'apprentissage automatique. Ce système prend entrées les données réelles des différentes situations de la circulation avec contexte détaillé, ces données sont exploitées comme une base d'apprentissage par les algorithmes ML afin de générer une solution de coordination qui gères les interactions dans l'intersection.

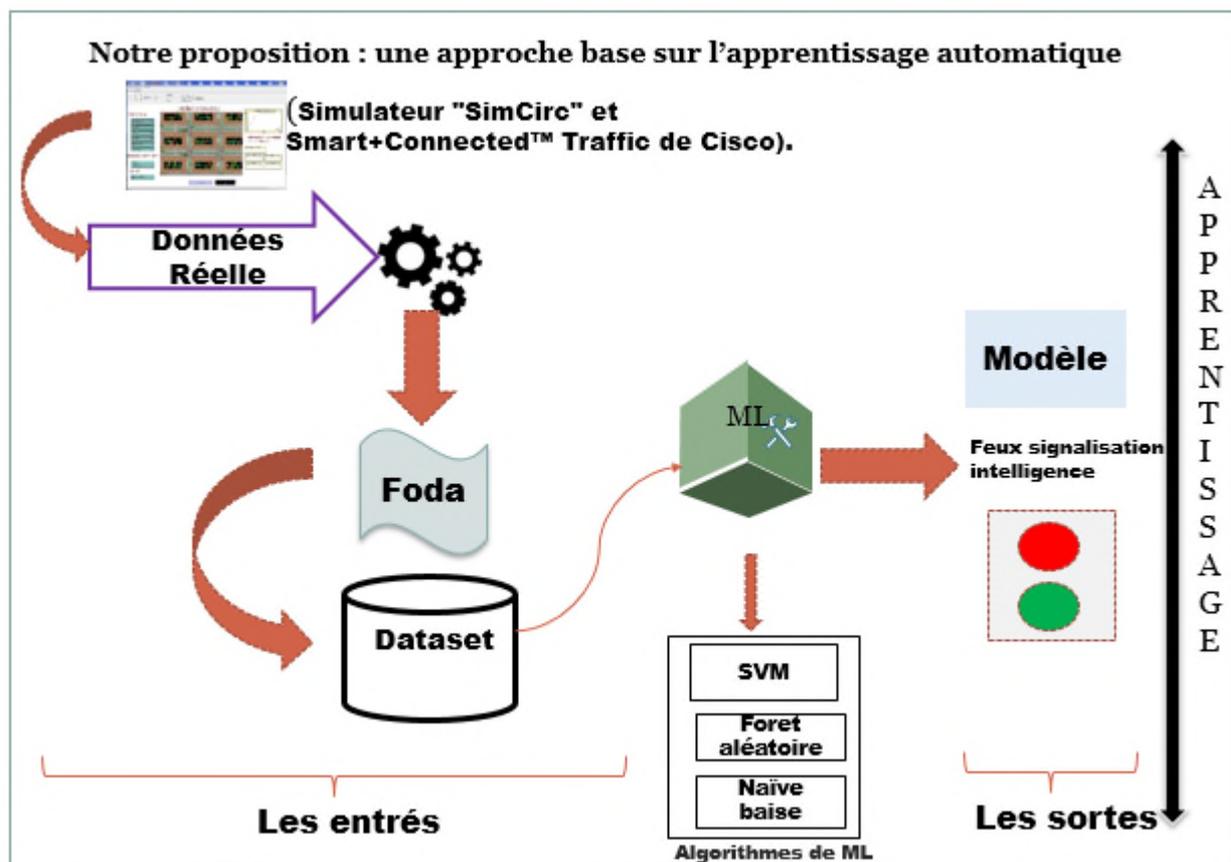


Figure 5 : Présentation de notre approche

1.2.1 Les étapes de notre approche

Le système proposé précédemment est encapsuler en six étapes fondamentales : (1) Définir le problème, (2) Création de la base d'apprentissage, (3) Préparer et prétraitement. (4) Citer les modèles de données (5) Entraîner et tester le modèle, (6) Le déploiement et la vérification.

La Figure 5ci-dessus illustre notre approche

1. Définir le problème.

- Définir les objectifs (feux signalisation intelligence).
- Identifier les objectifs d'apprentissage automatique (Modèle de classification).

2. Création de la base d'apprentissage (dataset).

- Accéder aux données nécessaires (Simulateur "SimCirc" et Smart+Connected™ Traffic de Cisco).
- Collecter et comprendre les données.

3. Extraction et préparation du data set.

- Sélectionner les données requises.
- Nettoyer / formater les données si nécessaires (Foda).

4. Choix du modèle.

- Sélectionner les algorithmes (SVM, Forêt aléatoire, Réseau Bayésien).
- Construire des modèles prédictifs :
 - stratégie1 : modèle prédictif avec SVM.
 - stratégie2 : modèle prédictif avec Forêt aléatoire (Radom Forest)
 - stratégie3 : modèle prédictif avec Réseau Bayésien (Naivbays).

5. Entraîner et tester le modèle.

- Entraîner le modèle avec des exemples d'ensembles de données (TRAIN).
- Tester et itérer le modèle (Test).

6. Le déploiement et la vérification.

- vérifier le modèle final.
- préparer la visualisation et déployer.

1.2.1.1 Création des données d'apprentissage :

Les données d'apprentissage sont générées par des équipements d'observation intelligentes qui capture les données avec des capteurs augmentés, ces données sont transmises vers des serveurs de de stockage travers des passerelles pour sera traiter par des algorithmes avancés. Nous pouvons citer une solution récente actuellement utilisée dans les pays développés comme par exemple l'appareil "Smart+Connected™ Traffic de Cisco"



1.2.1.1.1 Smart+Connected™ Traffic de Cisco :

1.2.1.1.1.1 Historique de Cisco :

Cisco Système est une entreprise informatique américaine spécialisée, à l'origine, dans le matériel réseau (routeurs et commutateurs Ethernet), et depuis 2009 dans les serveurs. Elle est dirigée par Chuck Robbins depuis juillet 2015. Son prédécesseur, John Chambers, était en fonction depuis 1995 et est président du conseil d'administration depuis novembre 2006. [12].

1.2.1.1.1.2 Définition de Smart+Connected™ Traffic de Cisco :

C'est un appareil multi fonctionnel rassemble des compétences opérationnelles, techniques et conceptuelles basées sur une connaissance approfondie et un vaste savoir-faire concernant les objectifs et les problèmes liés à la gestion de la circulation. Il s'appuie sur une suite complète de services professionnels, notamment des services de conseil, de conception, d'intégration et de mise en œuvre, mais aussi de formation, d'assistance et de maintenance. [13].

1.2.1.1.1.3 Principales caractéristiques de Smart+Connected™ Traffic de Cisco :

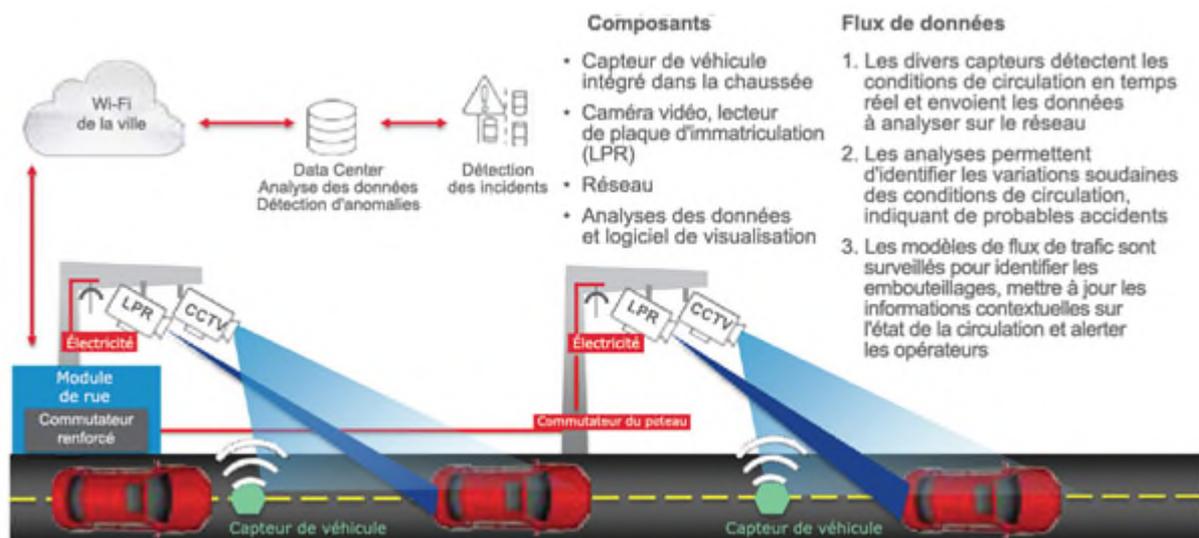


Figure 6 : Principales caractéristiques de Smart+Connected™ Traffic de Cisco [13]

A. Surveillance de la circulation :

. Prendre des Captures et vidéos dans les intersections en temps réel et affiche les conditions de circulation.

- Collecter de manière avancée les données sur la vitesse des véhicules, la densité de la circulation et le taux d'occupation des différentes voies.
- Surveiller les réseaux sociaux à la recherche d'informations pertinentes.
- Afficher les données importantes selon un système de couches pour mieux visualiser le contexte.

B. Détection et gestion des incidents :

- Identifie les accidents de la circulation en rassemblant les données de plusieurs sources, afficher leur emplacement et alerte les opérateurs.
- Créer automatiquement un rapport d'incident et recueille toutes les données pertinentes.
- Recommander des procédures pour guider les opérateurs rapidement lors des interventions.
- Permet de garantir une communication continue avec les intervenants.

C. Analyses, rapports et gestion :

- Afficher les tendances en matière d'accidents de la circulation pour identifier les problèmes de flux de circulation routière.
- Configure et gère les capteurs, l'infrastructure vidéo.

1.2.1.1.1.4 Architecture de Smart+Connected™ Traffic :

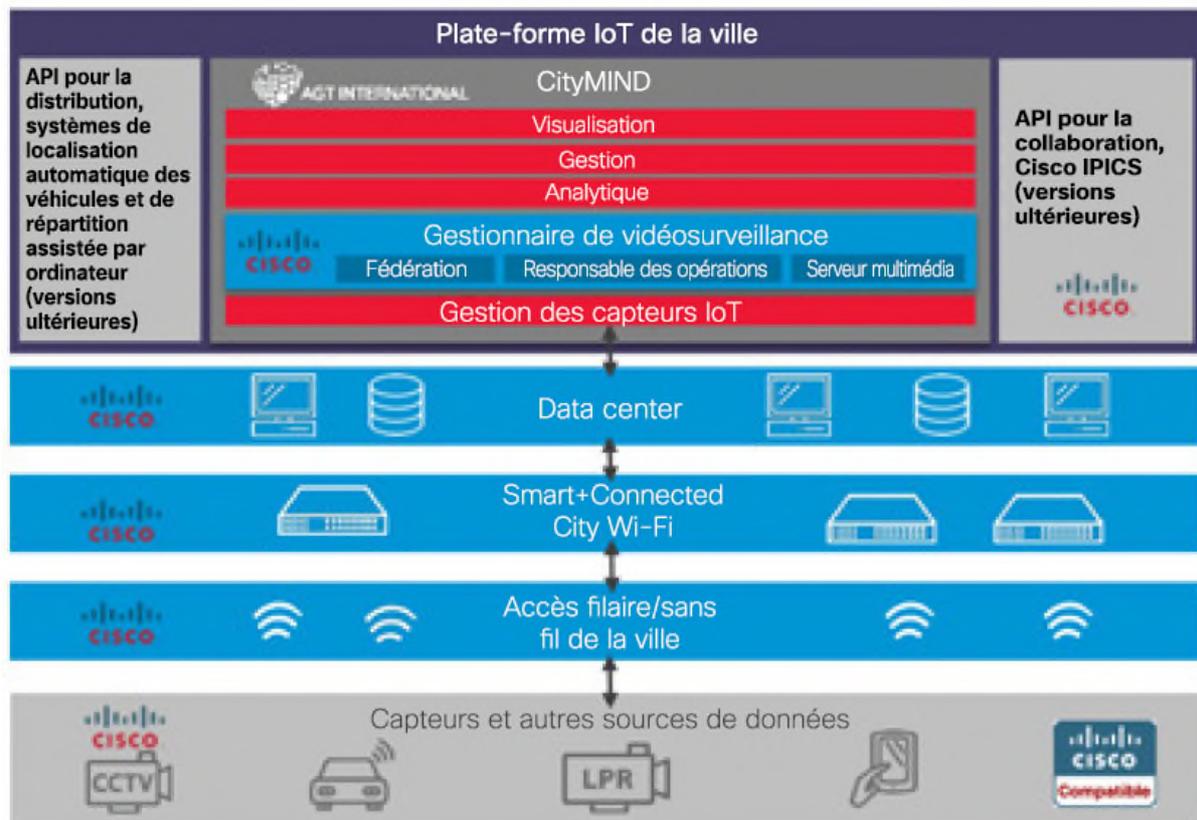


Figure 7 : Vue d'ensemble de l'architecture de Smart+Connected™ Traffic[13]

Les Composants techniques :

AGT (Analyse, Gestion, Technique) :

- Modules de visualisation, gestion, analyse et gestion des capteurs.
- Intégration de lecteurs de plaques d'immatriculation et de microphones.
- Analyses vidéo intégrées.

Cisco :

- Caméras IP intelligentes, encodeurs vidéo et gestionnaire de vidéosurveillance Cisco.
- Points d'accès sans fil renforcés et commutateurs Ethernet industriels.
- Infrastructure de data center unifiée.

Partenaire :

- Capteurs de véhicules, passerelle et application de gestion.

1.2.1.1.2 Présentation du simulateur « SimCirc » :

Un simulateur considérable trafic d'un ensemble de véhicule (virtuel), avec un but indispensable autonomiser pour Optimiser c.-à-d. Donne plus d'autonomie à nos feux de circulation afin d'améliorer la gestion des intersections de nos carrefours. Considèrent le bouton « Configuration » qui reproduit l'étape d'initialisation des différents paramètres de notre modèle, ainsi que la configuration de l'environnement de simulation (figure 8).

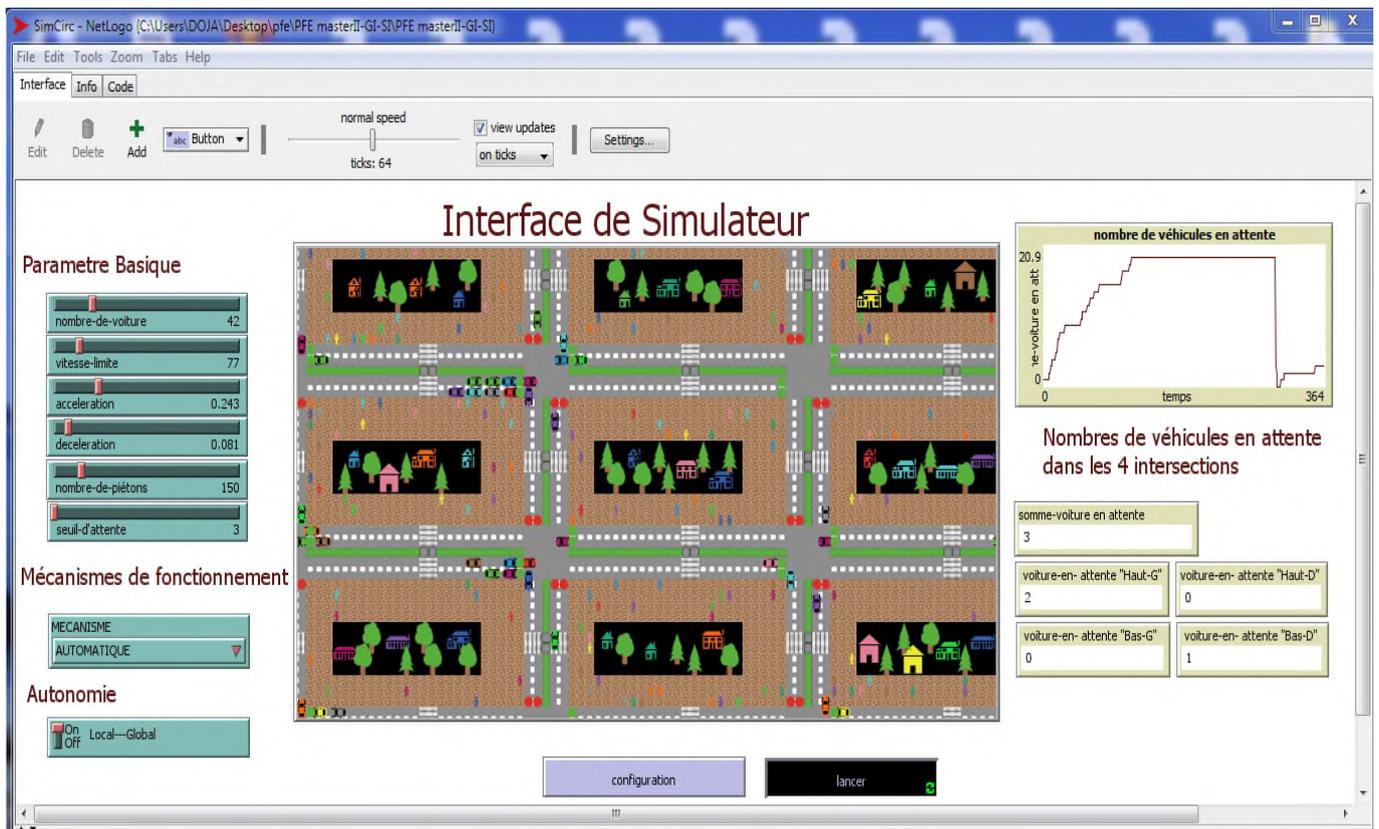


Figure8 :Simulateur "SimCirc" (net logo) [14].

1.2.1.1.3 Traitement de data set

La création de notre dataset passe par les étapes suivantes : (1) créer la structure de dataset, (2) extraction de les données, (3) alimente notre data set par des données conformes à notre Fodapropose dans la section (2.1.1.2.1.La création du datasetdans notre cas est basé sur l'observation des systèmes comme par exemple (notre cas) : Simulateur "SimCirc"(netlogo) qui a représenté dansla figure 8 et nous avons également utilisé un autre ensemble de données qui représente les résultats de la caméra Cisco(Smart+Connected™ Traffic de Cisco); Que nous avons récupéré sur le site de Kaggel.

1.2.1.1.3.1 Traitement les données avec Foda (Feature-Oriented Domain Analysis)

Pour analyser les paramètres de dataset à introduire dans la base,nous avons utilisé le diagramme de caractéristique su Foda présenté précédemment. Dans notre cas les attributs de notre data set représente les caractéristiques notre Foda par exemple : les feux de circulation, route, Véhicule, ...etc.Pour concevoir notre dataset il faut analyser le domaine nous avons utilisé le diagramme de caractéristique du Foda afin de d'écrire la variabilité.

1.2.1.1.3.1.1 Feature-Oriented Domain Analysis (Foda)

La méthode FODA (Feature-Oriented Domain Analysis) a introduit au début des années 1990 le concept des fonctionnalités et l'utilisation de diagrammes de fonctionnalités (feature diagramme, FD) pour exécuter l'analyse du domaine ciblé par un logiciel. Bien que conceptuellement simples, ces diagrammes permettent de modéliser la variabilité d'une ligne de produits au niveau de ses fonctionnalités. Afin de faciliter l'adoption du diagramme de fonctionnalités, les créateurs de Foda ont proposé des lignes directrices permettant d'incorporer l'écachement l'ingénierie par les LPL au sein d'un processus de développement logiciel [15].

1.2.1.1.3.1.2 Modélisation de la variabilité :

A. Présentation Les caractéristiques de la circulation :

L'avantage d'utiliser ce modèle réside dans la possibilité modéliser la variabilité durant la gestion de circulation de véhicule dans un carrefour. Dans la figure16, nous représentons un espace de variante globale à l'aide d'un diagramme de *fonctionnalités*. Pour une première classification des facteurs d'impact, nous définissons quatre groupes : (1) Route ; (2) Véhicule ; (3) Distance ; (4) Personne.

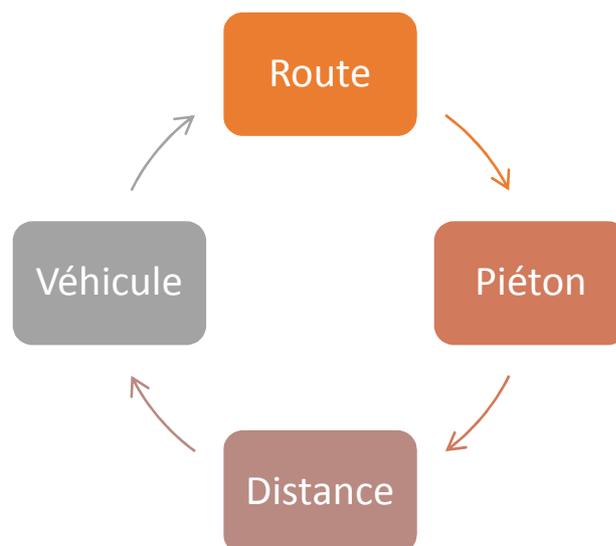


Figure 9 :Présentation Les caractéristiques de la circulation

Les quatre facteurs d'impact sont toujours présents lors de l'exécution et l'évaluation :

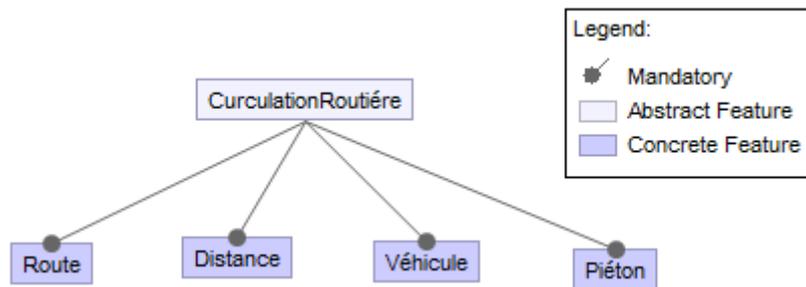


Figure 10 : Les caractéristiques de la circulation routière

A.1. Route : c'est au sens littéral une voie terrestre aménagée pour permettre la circulation de véhicule, on a basé dans la route sur la voie et l'intersection

A.1.1 .La voie

A.1 .2 L'intersection

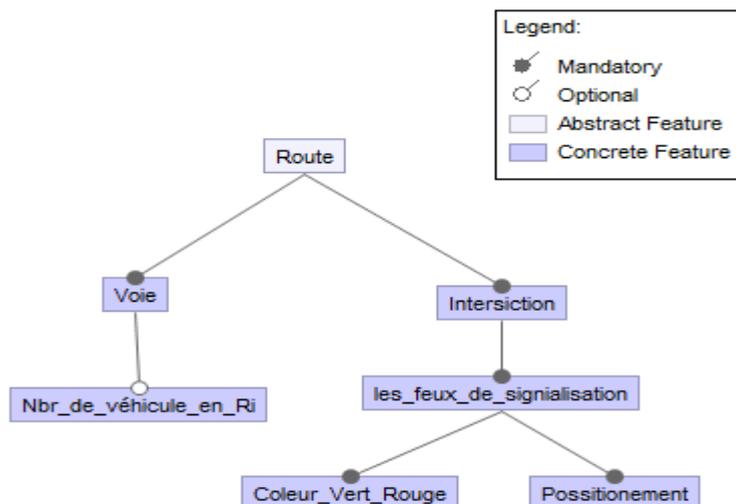


Figure11 :Les caractéristiques de la route

NB : $R(i)$: représente le nombre de route. $i : = 4$

A.2. Distance : la distance parcourue entre le temps de réaction

A.2.1. Véhicule entre Véhicule :est celle qui sépare deux véhicule roulant l'un derrière L'autre sur une même voie selon leur vitesse.

A.2.2. Véhicule entre intersection :la distance entre le carrefour et les véhicule qui y arrivent.

A.2.3. Véhicule entre personne : la distance entre la véhicule et la personne qui a traverser la route

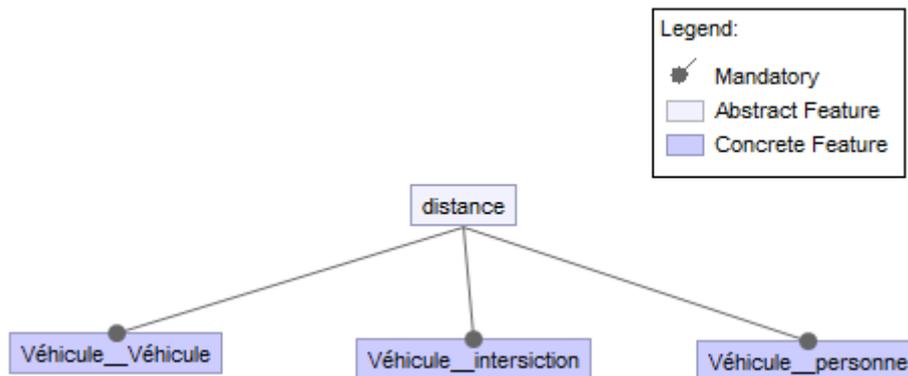


Figure 12: Les caractéristiques de la distance

A.3. Véhicule : la voiture est considérée comme l'un des moyens de transport les plus répandus à notre époque, ce que nous intéressent de la part de véhicule dans notre approche

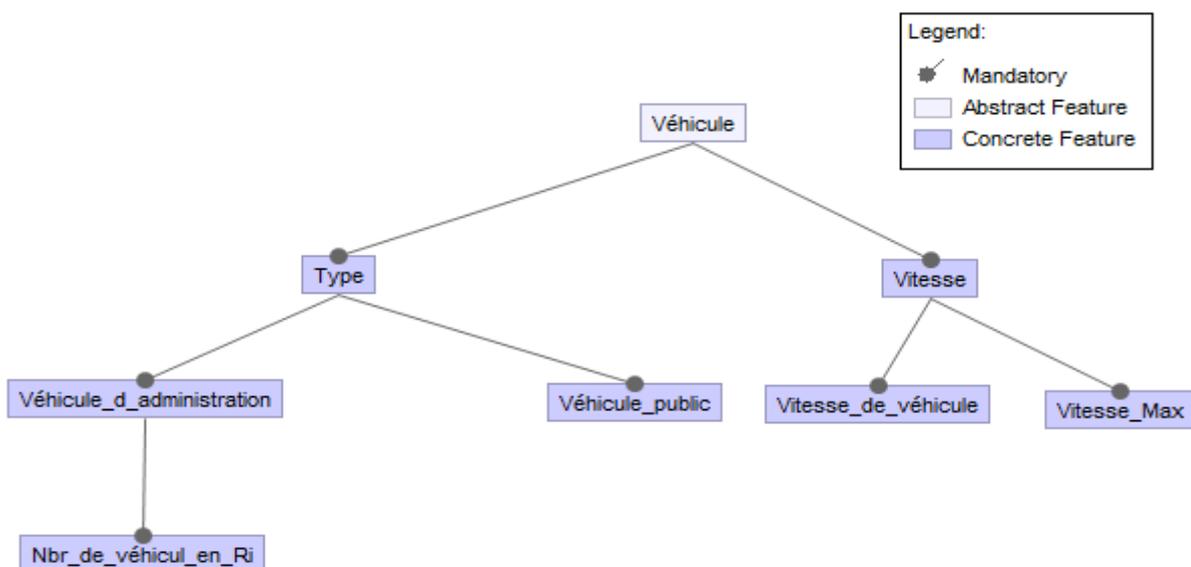


Figure 13 : Les caractéristiques de la Véhicule.

A.3.1. Vitesse : rapidité de la voiture lors de son mouvement, la vitesse de véhicule est le rapport entre la distance parcourue et la durée de parcours :

A.3.1.1. Vitesse de Véhicule : la vitesse du véhicule.

A.3.1.2. Vitesse Max : la vitesse maximes du véhicule sur la route

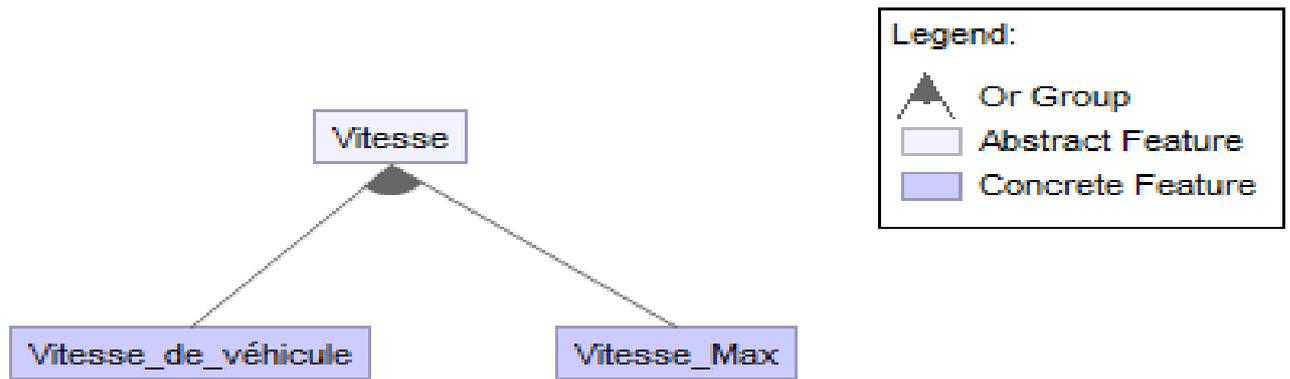


Figure 14 : Les caractéristiques de la vitesse.

A.3.2. Type : type de véhicule

A.3.2.1. Véhicule d'administration

A.3.2.2. Véhicule public :

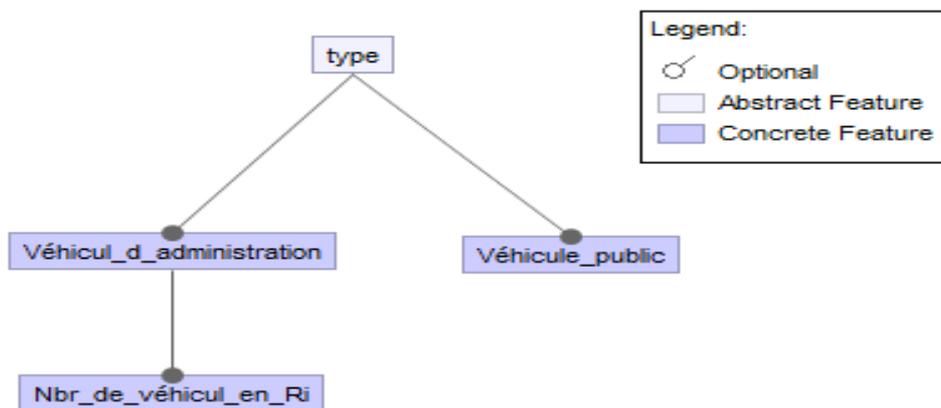


Figure 15 : Les caractéristiques de la Type

A.4. Piéton : les personnes qui traverser route

A.4.1. Personne ne traverser R(i) :

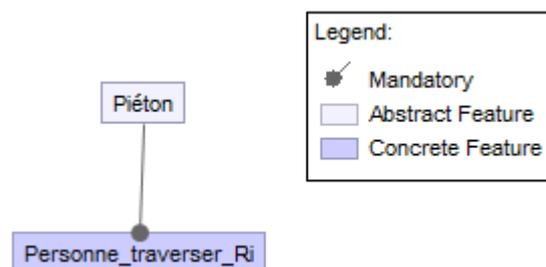


Figure 16: Les caractéristiques des personnes

B. Présentation Les caractéristiques du feu de circulation :

Les paramètres spécifiques aux feux de circulation sont : Le Feu vert et Le Feu rouge

B.1. Couleur Rouge / Vert

B.2. Positionnement

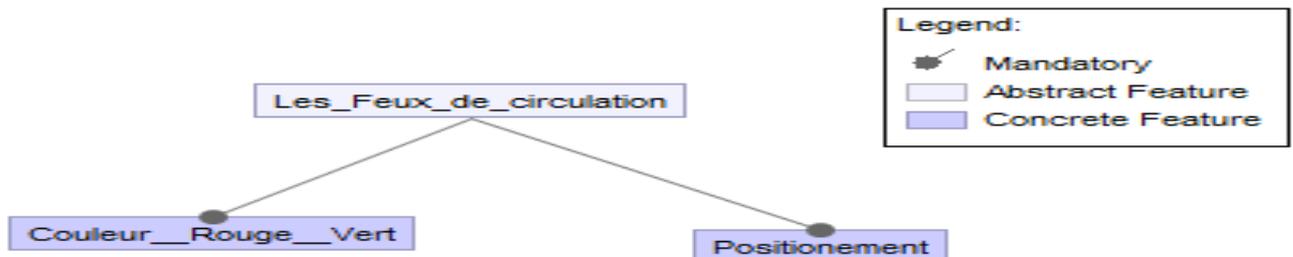


Figure 17 : Les caractéristiques du feu de signalisation.

C. L'espace de variante globale :

Pour le schéma complet des facteurs d'impact, nous fusionnons les diagrammes de fonctionnalités présentés avant dans figure 16. Dans notre modèle nous ajoutons une contrainte d'exclure certaines des variantes qui ne sont pas possible comme les différentes architectures de déploiement (Parallèle, Repartie, Distribuée, Architecture de nuage). Pour la modélisation du modèle de fonctionnalités, nous avons utilisés Feature IDE pour Eclipse.

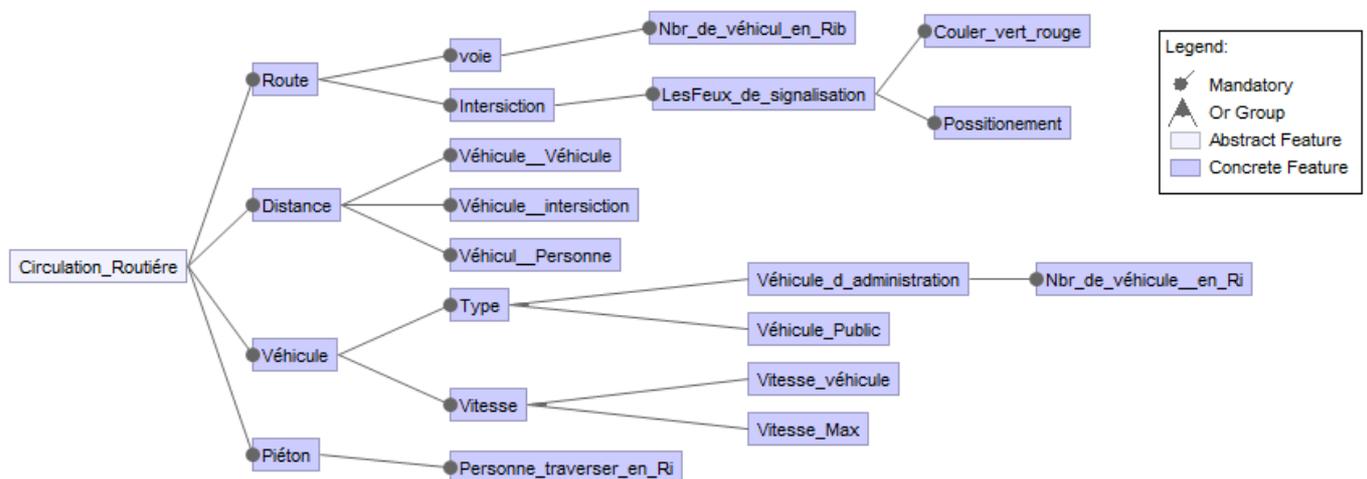


Figure 18 : L'espace de variante globale de circulation routière

NB : dans notre approche on peut ajouter d'autres composants tels que les plaques, les passages de piétons, le climat.

1.2.1.2 Apprentissage automatique (Machine Learning) :

Est un sous-domaine de l'intelligence artificielle (IA) qui permet à des applications de prédire des résultats de plus en plus précis sans être explicitement programmées en ce sens.

Les algorithmes de machine Learning utilisent des données historiques en entrée pour prédire de nouvelles valeurs en sortie.

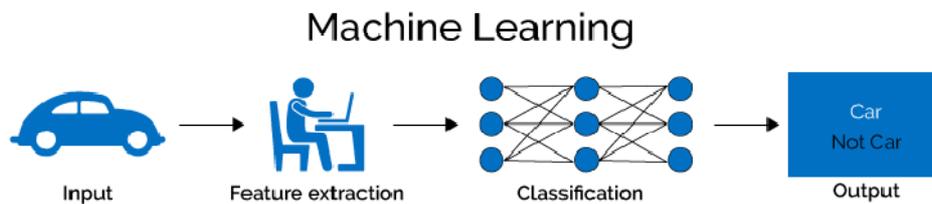


Figure 19 : Apprentissage automatique [16]

1.2.1.2.1 Les types d'apprentissage automatique

Le domaine de l'apprentissage automatique est souvent divisé en sous-domaines selon les types de problèmes abordés. On peut les classer approximativement comme suit :

1.2.1.2.1.1 Apprentissage supervisé :

à partir d'une donnée d'entrée, par exemple une photographie avec un panneau de signalisation, la tâche consiste à prédire le résultat ou l'étiquetage correct, par exemple le panneau de signalisation qui se trouve dans l'image (limitation de vitesse, panneau stop, etc.). Dans les cas les plus simples, les réponses se présentent sous la forme de oui/non (on les qualifie alors de problèmes de classification binaire).

1.2.1.2.1.2 Apprentissage non supervisé :

La tâche consiste à découvrir la structure des données : par exemple, regrouper des éléments similaires pour former des « grappes », ou réduire les données à un petit nombre de « dimensions » importantes. La visualisation des données peut aussi être considérée comme un apprentissage non supervisé.

1.2.1.2.1.3 Apprentissage par renforcement :

communément utilisé dans les situations où un agent d'IA, comme une voiture autonome, doit fonctionner dans un environnement et où un retour d'information sur les bons ou mauvais choix est disponible avec un certain retard. Également utilisé dans les jeux dont le résultat ne peut être décidé qu'à la fin de la partie.

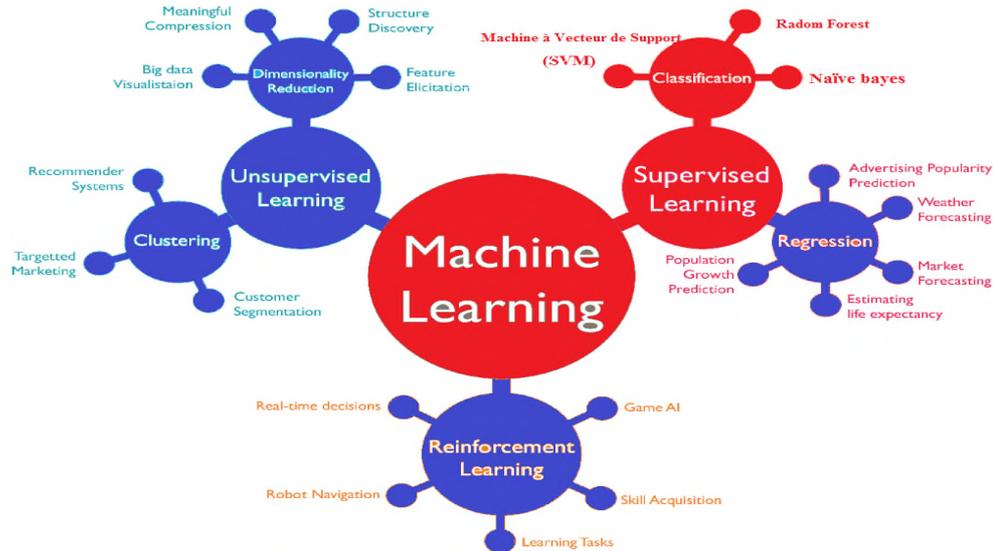


Figure 20 : Les types d'apprentissage automatique [16]

Remarque : Nous basent sur les cycle rouge (d'apprentissage automatique, Apprentissage supervisé, classification) pour résoudre notre approche.

1.2.1.2.2 Classification :

La classification est le processus de recherche ou de découverte d'un modèle (fonction) qui permet de séparer les données en plusieurs classes qualitatives. Dans la classification, l'appartenance au groupe du problème est identifiée, ce qui signifie que les données sont classées sous différentes étiquettes en fonction de certains paramètres, puis que les étiquettes sont prédites pour les données.

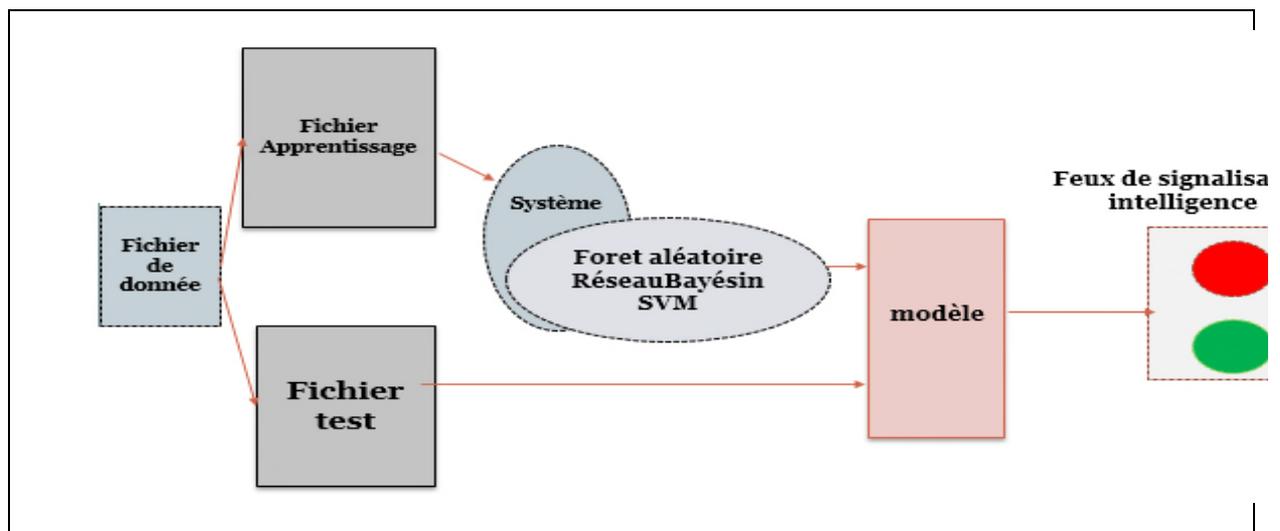


Figure 21 : Architecture générale du processus de classification

Le processus de la classification se déroule principalement en deux étapes : une phase d'apprentissage et une phase de test, Le prédicteur reçoit en entrée des données d'apprentissage qu'il utilise pour générer un modèle, il reçoit par la suite un fichier de test, dont il se servira pour prédire les nouvelles valeurs (0 : Rouge, 1 : Vert) en se basant sur le modèle préalablement généré. C'est le processus terminé, les données de test et les résultats des prédictions sont comparés pour traduire la précision,

1.2.1.2.3 Les algorithmes de la classification :

Il existe plusieurs algorithmes pour la classification Dans notre solution, on a appliqué trois algorithmes d'apprentissage automatique que nous allons comparer :Machine à Vecteur de Support (SVM),Foret aléatoire(Radom Forest)Réseau Bayésien (Nivebayse).

1.2.1.2.3.1 Foret aléatoire (RadomForest)

L'algorithme des « forêts aléatoires » (ou Radom Forest parfois aussi traduit par forêt d'arbres

Décisionnels) est un algorithme de classification qui réduit la variance des prévisions d'un arbre de

Décision seul, améliorant ainsi leurs performances. Pour cela, il combine de nombreux arbres de décisions dans une approche de type bagging. Chaque arbre de la forêt aléatoire est entraîné sur un sous ensemble aléatoire de données selon le principe du bagging, avec un sous ensemble aléatoire de

Features(caractéristiques variables des données) selon le principe des « projections aléatoires ». Les prédictions sont ensuite moyennées lorsque les données sont quantitatives ou utilisés pour un vote pour des données qualitatives, dans le cas des arbres de classification.

L'algorithme des forêts aléatoires est connu pour être un des classeurs les plus efficaces « out-of-the-box » (c'est-à-dire nécessitant peu de prétraitement des données). La [Figure 22](#) illustre la structure de l'algorithme Randon Forest.[17]

✓ Définition du Bagging

Le bagging a pour but de réduire la variance de l'estimateur, en d'autres termes de corriger l'instabilité des arbres de décision (le fait que de petites modifications dans l'ensemble d'apprentissage entraînent des arbres très différents). [18]

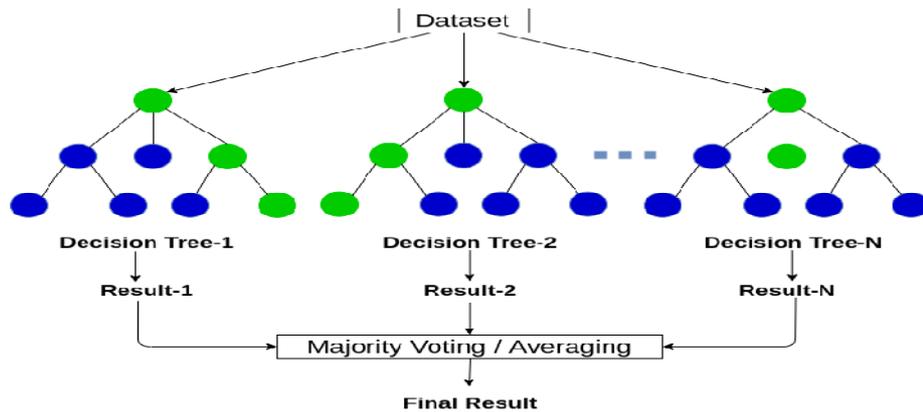


Figure 22 :l'algorithmme de foret aléatoire.

✓ **Algorithme de Foret aléatoire :**

Méthode Construire-Foret - aléatoire(D : données, N : nœud, M : méthode)

1. On applique M sur D afin de trouver le critère de division
2. En utilisant le critère de division, on divise le nœud N en sous nœuds
3. On met le nombre des sous nœuds de N dans la variable k

Si $k > 0$ **alors**

On fabrique k sous nœuds N_1, \dots, N_k de N

On divise D en D_1, \dots, D_k

Pour $i=1$ jusqu'à k **faire**

Construire- Foret - Aléatoire(D_i, N_i, M)

Fin pour

Fin Si

Fin de la procédure

Figure 23 :Procédure de construction deforet aléatoire[18]

A. Avantages :[19]

- ✓ Reconnaissance très rapide.
- ✓ Multi-classes par nature.
- ✓ Efficace sur inputs de grande dimension.
- ✓ Robustesse aux outliers.

B. Inconvénients :[19]

- ✓ Apprentissage souvent long

- ✓ Les forêts aléatoires sont très bonnes pour les problèmes de classification mais légèrement moins bonnes pour les problèmes de régression. Contrairement à la régression linéaire, la régression forestière aléatoire est incapable de faire des prédictions en dehors de ses données d'apprentissage.

- ✓ Les forêts aléatoires sont également des boîtes noires, contrairement à certains algorithmes d'apprentissage automatique traditionnels, il est difficile de regarder à l'intérieur d'un classificateur de forêt aléatoire et de comprendre les raisons de ses décisions. En outre, ils peuvent être lents à s'entraîner et à s'exécuter et produire des fichiers de grande taille.

1.2.1.2.3.2 Machine à Vecteur de Support (SVM) :

Le SVM est un classificateur dit linéaire, ça veut dire que, dans le cas parfait, les données doivent être linéairement séparables. Ainsi notre corpus est représenté comme étant un espace vectoriel, où chaque donnée représentée par un point dans ce dernier. La problématique maintenant est de trouver le meilleur séparateur (ligne, plan ou hyperplan) qui partage notre corpus en deux catégories. L'espace entre ces deux catégories est appelé marge, qui est définie par les points (Vecteurs de support) les plus proches du séparateur, de part et d'autre. Le but étant essentiellement de maximiser cette marge, plus elle est grande meilleurs est le résultat. Le classificateur se généralise bien avec les nouvelles données.

Toutefois, si les données ne sont pas linéairement séparables, la SVM peut être modifiée pour tolérer un minimum d'erreurs. Désormais, le but est de maximiser la marge et de minimiser l'erreur de classification. Une autre alternative pour parer à la non séparabilité des données, est de passer à un espace de dimension supérieur.

La figure ci-dessous illustre la façon dont SVM départage les données dans le cas où elles sont linéairement séparables, de plus, elle choisit la séparation la plus optimale où la marge est maximale.

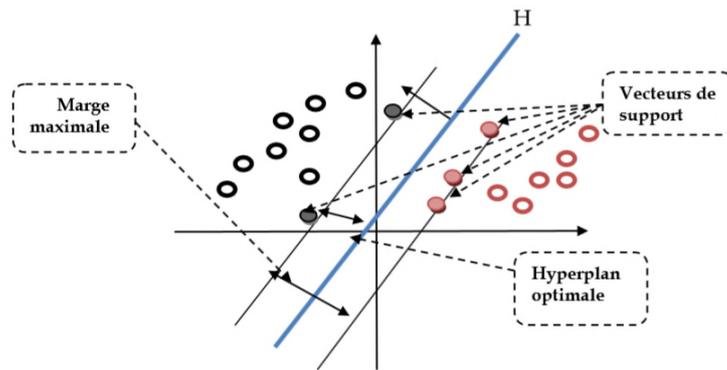


Figure 24 : hyperplan optimal, vecteurs de support et marge maximale. [20]

Méthode SVM

- 1 : Optimiser le vecteur de poids w
- 2 : Calculer le critère de rang $R_c(X_j)$ pour tout $j \in \{1, \dots, p\}$
- 3 : Éliminer la variable qui minimise $R_c(X_j)$
- 4: Répéter les étapes 1 à 3 tant qu'il reste des variables

Figure 25 : L'algorithme de SVM [21]

A. Les Avantages de SVM

Les SVM représentent plusieurs avantages parmi lesquels on peut citer par exemple [22] :

- ✓ Leur capacité à manipuler de grandes quantités de données.
- ✓ Le faible nombre d'hyper paramètres utilisés par ces méthodes.
- ✓ Elles sont bien fondées théoriquement.
- ✓ Les résultats pertinents qu'on obtient avec les SVM en pratique.

B. Les Inconvénients de SVM

Malgré leurs performances, les SVM représentent aussi des faiblesses, parmi lesquelles on peut citer :

1. Leur utilisation des fonctions mathématiques complexes pour la classification des corpus.
2. Pour trouver les meilleurs paramètres, ce type d'algorithmes demande un temps énorme pendant les phases de test. [22]

1.2.1.2.3.3 Réseau Bayésien(Naïve bayes) :

Un réseau bayésien est un graphe orienté et acyclique qui contient des nœuds connectés entre eux par des arcs représentant des dépendances probabilistes entre les variables et des distributions de probabilités conditionnelles pour chaque variable. La classification naïve bayésienne est basée sur le théorème de Bayes avec une forte indépendance des hypothèses.

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

P (A) : la probabilité que l'hypothèse h soit vraie (indépendamment des données). Ceci est connu comme la probabilité a priori de h.

P (B) : la probabilité des données (quelle que soit l'hypothèse). C'est ce qu'on appelle la probabilité a priori.

P (A | B) : la probabilité de l'hypothèse h compte tenu des données D. C'est ce qu'on appelle la probabilité postérieure (B | A) : la probabilité des données d étant donné que l'hypothèse h était vraie. C'est ce qu'on appelle la probabilité postérieure.[23]

✓ Algorithme deRéseau Bayésien :

D un ensemble de données avec leurs attribut prévues V, $p(w_k(v))$ la probabilité d'un valeur w_k tiré au hasard à partir d'un donnée quelconque de la classe v_j et $p(v_j)$ la «Prior probabilités» de la classe v_j .

1. Collecter de tous les attributs apparaissant dans l'ensemble des données.
2. Calculer les termes $p(v_j)$ et $p(w_k | v_j)$

Pour chaque valeur v_j de V faire

données_j :le sous-ensemble de données de D pour lesquels la valeur prévue est v_j .

$$p(v_j) = \frac{|doonnes_j|}{|D|}$$

Pour chaque valeur w_k apparaît dans *données*

$$p(w_k | v_j) = \frac{|n_k + 1|}{n + |D|}$$

Figure 26 : L'algorithmede Réseau Bayésien [23]

A. Les Avantages de Réseau Bayésien [24] :

- ✓ C'est relativement simple à comprendre et à construire
- ✓ Il est facile à former, même avec un petit jeu de données
- ✓ Il n'est pas sensible aux caractéristiques non pertinentes.

B. Inconvénients de Réseau Bayésien [25] :

- ✓ Il implique que chaque fonctionnalité soit indépendante, ce qui n'est pas toujours le cas.
- ✓ L'algorithme Réseau Bayésien suppose l'indépendance des variables : C'est une hypothèse forte et qui est violée dans la majorité des cas réels.

1.3 Conclusion

Notre contribution s'articule sur les approches supervisées particulièrement, Notre vision est d'utiliser le volet supervisé d'apprentissage automatique (Forêt aléatoire, Machine à Vecteur de Support, réseau Bayésien), comme des stratégies pour administrer intelligemment une circulation typique dans le chapitre prochain.

Chapitre IV :

L'Implémentation

1.1 Introduction

L'objectif de ce chapitre est de présenter les étapes de l'implémentation de l'approche proposée

Dans le cadre de transformation de modèles par l'exemples. Nous commençons tout d'abord par laPrésentation des outils technologies utilisés, et de l'environnement de développement que nous avonsUtilisés et on termine par les étapes de la réalisation et l'évaluation du modèle. Ce chapitre est composéde deux parties, l'implémentation du système et les résultats expérimentaux.

1.2 Environnement de développement et Les outils techniques

1.2.1 Dentition du langage Python en informatique :

Python est le langage de programmation open source le plus employé par les informaticiens.

Ce Langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels. En été, parmi ses qualités, Python permet notamment aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière dont ils le font. Il a libéré les développeurs des contraintes de formes qui occupaient leur temps avec les langages plus anciens. Ainsi, développer du code avec Python est plus rapide qu'avec d'autres langages [26]. On a utilisé la version Python 3.4.8 mais la dernière version c'est Python 3.9.0.

1.2.2 Anaconda



Anaconda est une distribution open source pour python et R. Il est utilisé pour la science de données, L'apprentissage automatique, l'apprentissage en profondeur, etc. Avec la disponibilité de plus de300Bibliothèques pour la science des données, il devient assez optimal pour tout programmeur de travaillerSur anaconda pour la science des données [27].

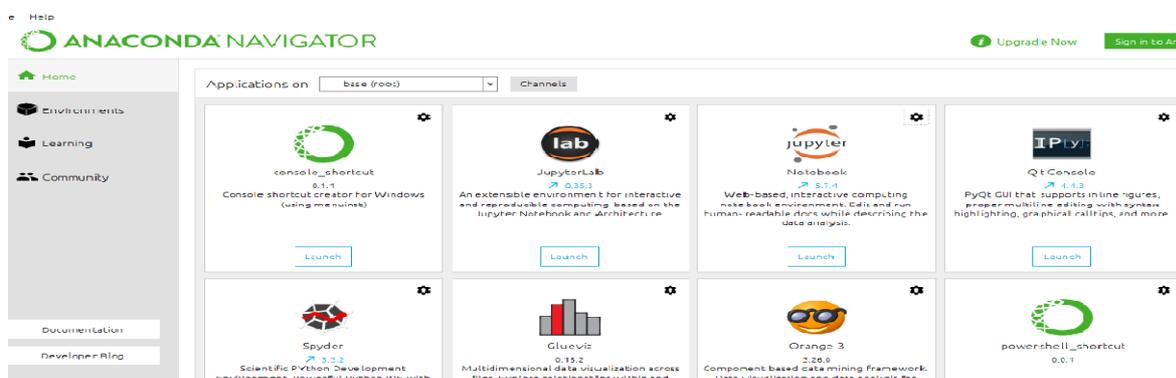


Figure 27 : L'environnement Anaconda

1.2.3 Jupiter :



Se présente comme un outil extrêmement simple à mettre en œuvre qui vous permettra de transformer vos Jupyter Notebooks en applications web ou en Dashboard quasiment automatiquement[28].

1.2.4 Les bibliothèques utilisées :

Dans le développement de notre solution, nous avons utilisé plusieurs bibliothèques : Numpy, Matplotlib, Pandas et Scikit-learn. Dans cette section, nous allons définir chaque bibliothèque utilisée.

1.2.4.1 Numpy :

Est une bibliothèque permettant d'exécuter des calculs numériques avec Python. Elle introduit une gestion facilitée des tableaux de nombres, des fonctions sophistiquées (`division`), on peut aussi l'intégrer au code C / C ++ et Fortran [29].

1.2.4.2 Matplotlib :

Est une bibliothèque de traçage pour le langage de programmation Python et son extension mathématique numérique NumPy. Il fournit une API orientée objet permettant d'incorporer des graphiques dans des applications à l'aide de kits d'outils d'interface graphique à usage général tels que Tkinter, python, Qt ou GTK + [30].

1.2.4.3 Pandas :

Pandas est une bibliothèque open source sous licence BSD fournissant des structures de données hautes performances et faciles à utiliser, ainsi que des outils d'analyse de données pour le langage de programmation Python. Pandas est un projet sponsorisé par NumFOCUS cela contribue au succès du développement de pandas en tant que projet open source de classe mondiale et permettra de faire un don au projet [31].

1.2.4.4 Scikit-learn :

Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique. Elle est développée par de nombreux contributeurs notamment dans le monde académique par des instituts français d'enseignement supérieur et de recherche comme Inria et Télécom Paris. Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions logistiques, des algorithmes de classification, et les machines à vecteurs de support. Elle est conçue pour s'harmoniser avec d'autres bibliothèques libres Python, notamment NumPy et SciPy [32].

Testez le modèle sur l'ensemble de test et évaluez la performance de notre modèle.

Avantages de la séparation train / test :

- ✓ Le modèle peut être formé et testé sur des données différentes de celles utilisées pour la formation.
- ✓ Les valeurs de réponse sont connues pour l'ensemble de données de test, donc les prédictions peuvent être évaluées.
- ✓ La précision des tests est une meilleure estimation que la précision de la formation des performances hors échantillon.

1.3 Apprentissage et Paramétrage des Modèles :

Dans cette section, nous allons présenter l'apprentissage et paramétrage des modèles

1.3.1 Préparation du Dataset

Avant toute chose, nous allons découper nos données en jeu d'entraînement (X_{train} , Y_{train}) et un jeu de test (X_{test} , Y_{test}), afin d'atteindre notre objectif, 70 des données de chacun des mélanges sont utilisées pour entraîner le classificateur, puis 30 des données ont été utilisées pour valider le modèle.

```
Entrée [24]: 1 #splitting the data test train |split
             2 from sklearn.model_selection import train_test_split
             3 x_train,x_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_state=40)
```

Avantages de la séparation train / test :

- Le modèle peut être formé et testé sur des données différentes de celles utilisées pour la formation.
- Les valeurs de réponse sont connues pour l'ensemble de données de test, donc les prédictions peuvent être évaluées.
- La précision des tests est une meilleure estimation que la précision de la formation des performances hors échantillon.

1.3.2 Application de PCA :

Pour exécuter PCA on a utilisé la bibliothèque Scikit-Learn de Python. La classe PCA est utilisée à cet effet. PCA dépend uniquement de l'ensemble de fonctionnalités et non des données d'étiquette. La réalisation de PCA à l'aide de Scikit-Learn est un processus en deux étapes :

- ✓ Initialisez la classe PCA en passant le nombre de composants au constructeur.

- ✓ Appelez l'ajustement, puis transformez les méthodes en transmettant l'ensemble de fonctionnalités à ces méthodes.
- La méthode de transformation renvoie le nombre spécifié de composants principaux, voici le code suivant :

```
1 # trying PCA
2 from sklearn.decomposition import PCA
3 pca = PCA(2)
4 new_X = pca.fit_transform(X)
5 ratio = pca.explained_variance_ratio_
6
7 # Plotting The data
8 plt.scatter(new_X[:,0] , new_X[:,1] , c = y )
9 plt.legend()
10 plt.show()
```

➤ Réduction les dimensions de dataset

Ensuite, on a créé un modèle PCA pour faire choisi le nombre optimal du dataset. Pour exécuter PCA on a utilisé la bibliothèque Scikit-Learn de Python. La classe PCA est utilisée à cet effet. PCA dépend uniquement de l'ensemble de fonctionnalités et non des données d'étiquette. La réalisation de PCA à l'aide de Scikit-Learn est un processus en deux étapes :

Initialisez la classe PCA en passant le nombre de composants au constructeur.

Appelez l'ajustement, puis transformez les méthodes en transmettant l'ensemble de fonctionnalités à ces méthodes.

La méthode de transformation renvoie le nombre spécifié de composants principaux, voici le code **Figure 28** suivant : la Figure 5.3 affiché l'exécution du l'algorithme réduction les dimensions de dataset (PCA).

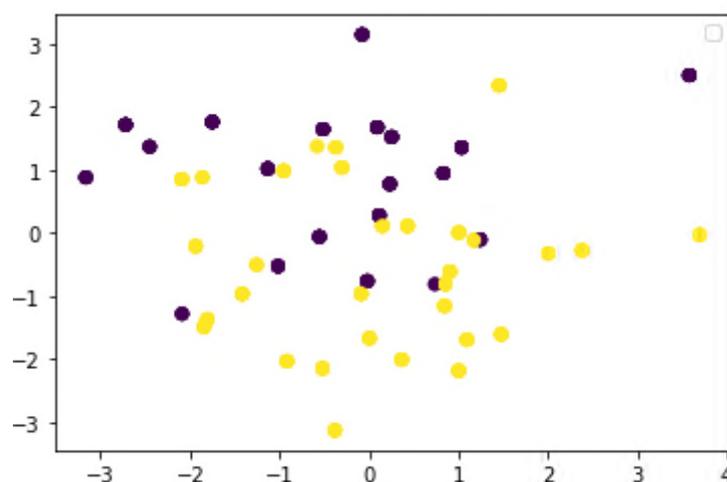


Figure28 : L'exécution de l'algorithme PCA

1.3.3 Algorithme LDA

Pseudocode de l'algorithme LDA Le code suivant l'algorithme L'analyse discriminante linéaire ou quadratique

```

1 graph_acp2(new_X, y)
2 #trying LDA
3 lda = LinearDiscriminantAnalysis(n_components=2)
4 X_iris_LDA = lda.fit(X, y).transform(X)
5 print(X_iris_LDA)
6 print("Rappel des composantes identifiées par le PCA :")
7 graph_acp2(new_X, y)
8
9 # spiting the data test train split
10 x_train , x_test , y_train , y_test = train_test_split(X,y , test_size =0.8 , random_state = 20)
11 print(x_test, y_test)
12 print(x_train, y_train)

```

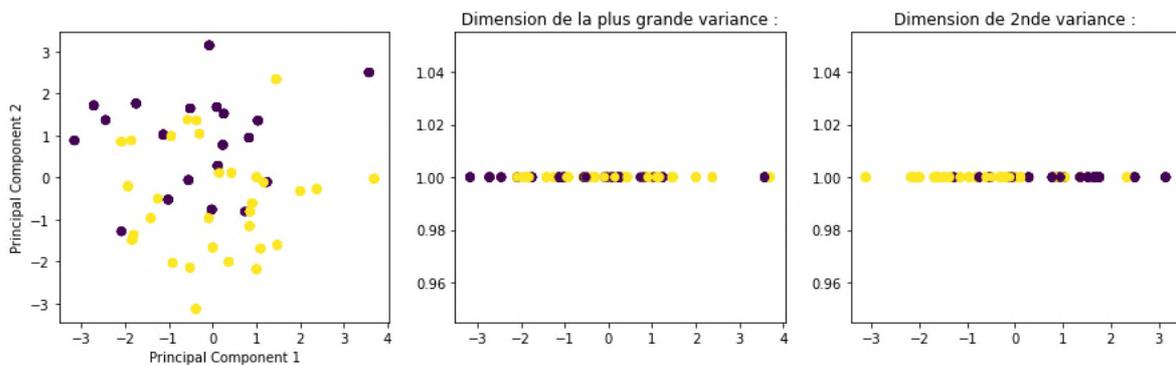


Figure29 : L'exécution de l'algorithme LDA

1.4 Apprentissage du modèle

Dans la recherche de modèles, on utilise des algorithmes d'apprentissage supervisé.

1.4.1 Algorithme de Random Forest

Pour classer dataset, on a fait la prédiction sur les données de test, et on a le résultat matrice de confusion, Score et Accuracy.

➤ Le code de l'algorithme de Forêt Aléatoire :

```

1 #####
2 from sklearn.ensemble import RandomForestClassifier
3 ##### Random Forest
4 DT = RandomForestClassifier(min_weight_fraction_leaf= 0.3 , criterion ='entropy')
5 DT.fit(x_train,y_train)
6 y_pred1 = DT.predict(x_test)
7 print('The score For Random Forest: ', f1_score(y_test , y_pred1))
8 print('The accuracy For Random Forest is ', (y_pred1 == y_test).mean())
9 cm = confusion_matrix(y_test, y_pred1)
10 print('Confusion Matrix is :', cm)
11 disp = plot_confusion_matrix(DT, x_test, y_test)
12 disp.ax_.set_title("plot")
13
14 plt.show()
15

```

➤ **Score et Accuracy :**

Voilà le Score et Accuracy :

```
The score For Random Forest: 0.8615384615384616
The accuracy For Random Forest is 0.82
```

➤ **Matrice de confusion de Foret Aléatoire :**

Le Tableau 1 explique la performance du modèle. Comme on peut l'observer, une matrice de confusion à en colonnes et en lignes les mêmes intitulés :

- ✓ En ligne : on lit les labels des individus.
- ✓ En colonne : les labels prédits par le modèle Ainsi, on peut conclure par exemple que dans 35 situation Où il appartient à la classe 0, le modèle a bien prédit les classes 1 et -1 déduite par le modèle contient 8 observations bien classés.

La figure suivante illustre l'affichage de l'exécution du code précédent

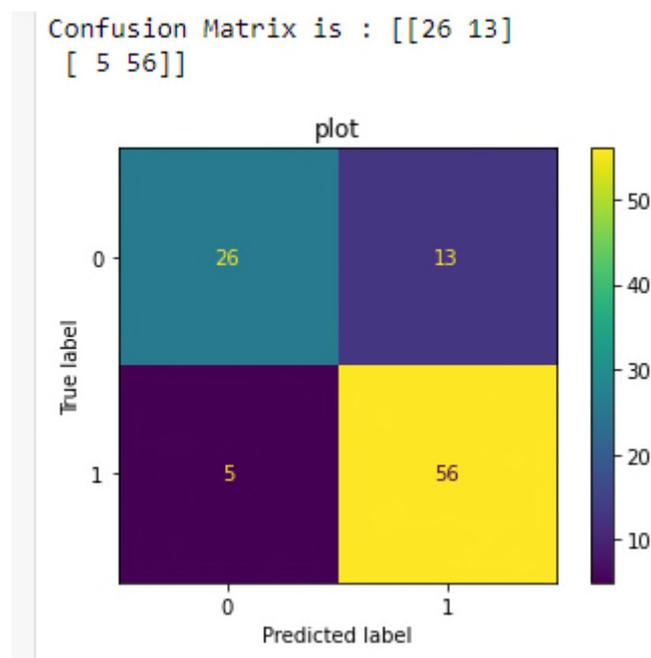


Tableau 1 : Matrice de confusion de Foret Aléatoire

1.4.2 Algorithme de Réseau Bayésien :

Pour classer dataset, on a fait la prédiction sur les données de test, et on a le résultat matrice de confusion, Score et Accuracy.

➤ **Le code de l'algorithme de Réseau Bayésien :**

```
1 from sklearn.naive_bayes import GaussianNB
2 import matplotlib.pyplot as plt
3
4
5 ##### Naive Bayes
6 NB = GaussianNB(var_smoothing = 6)
7 NB.fit(x_train,y_train)
8 y_pred3 = NB.predict(x_test)
9 print("nb",NB.predict(test))
10 print("Accuracy Naive Bayes :",accuracy_score(y_test, y_pred3))
11 print("scort Naive Bayes:", f1_score(y_test, y_pred3))
12
13 disp = plot_confusion_matrix(NB, x_test, y_test)
14 disp.ax_.set_title("plot")
15
16 plt.show()
```

➤ **Score et Accuracy :**

Voilà le Score et Accuracy :

```
Accuracy Naive Bayes : 0.66
scort Naive Bayes: 0.782051282051282
```

➤ **Matrice de confusion de Réseau Bayésien :**

Le Tableau 2explique la performance du modèle. Comme on peut l'observer, une matrice de confusion a en colonnes et en lignes les mêmes intitulés :

— En ligne : on lit les labels des individus.

— En colonne : les labels prédits par le modèle Ainsi, on peut conclure par exemple que dans 35 situation Où il appartient à la classe 0, le modèle a bien prédit les classes 1 et -1 déduite par le modèle contient 8 observations bien classés. La figure suivante illustre la matrice de confusion de l'algorithme Réseau Bayésien :

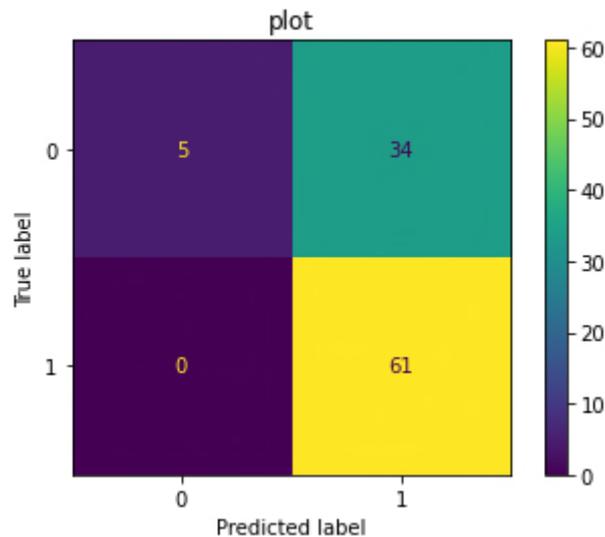


Tableau 2 :Matrice de confusion de l'algorithme Réseau Bayésien

1.4.3 Algorithme de SVM :

Pour classer dataset, on a fait la prédiction sur les données de test, et on a le résultat matrice de confusion, Score et Accuracy.

➤ Le code de l'algorithme de SVM :

```

1 from sklearn.svm import SVC
2 ##### SVM
3 svc = SVC(kernel = 'sigmoid' , random_state = 12 , gamma =0.9 , C =12)
4 svc.fit(x_train, y_train)
5 y_pred2 = svc.predict(x_test)
6
7 print("SVM",svc.predict(test))
8 print("Accuracy SVM:",accuracy_score(y_test, y_pred2))
9 print(" scort SVM:", f1_score(y_test, y_pred2))
10 cm = confusion_matrix(y_test,y_pred2)
11 print('Confusion Matrix is :', cm)
12 disp = plot_confusion_matrix(svc, x_test, y_test)
13 disp.ax_.set_title("plot")
14
15 plt.show()
16 print('#####')
17

```

➤ Score et Accuracy :

Voilà le Score et Accuracy :

```

Accuracy SVM: 0.76
scort SVM: 0.8208955223880597

```

➤ **Matrice de confusion de SVM :**

Le Tableau 3 explique la performance du modèle. Comme on peut l'observer, une matrice de confusion a en colonnes et en lignes les mêmes intitulés :

— En ligne : on lit les labels des individus.

— En colonne : les labels prédits par le modèle. Ainsi, on peut conclure par exemple que dans 35 situations où il appartient à la classe 0, le modèle a bien prédit les classes 1 et -1 déduite par le modèle contient 8 observations bien classés. La figure suivante illustre la matrice de confusion de l'algorithme SVM :

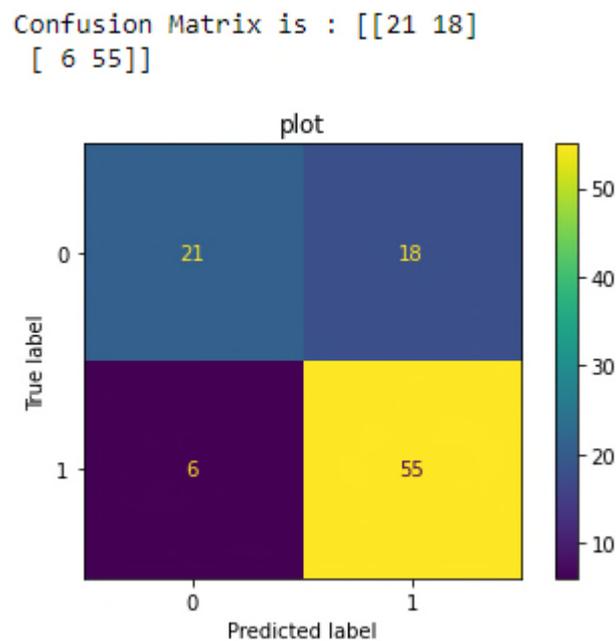


Tableau 3 :Matrice de confusion de l'algorithme SVM

1.5 Comparaison entre Les Trois stratégie :

Nous avons comparé les trois stratégies de classification SVM, Reaseaubaysien et Foret aléatoire. Cette comparaison est basée sur les métrique de performances.

A. Comparaison entre les trois stratégies qui nous utilisent :par apport Accuracy.

Notre objectif dans ce travail est de construire un classifieur capable d'identifier la décision du tableau de signalisation à partir des informations qui exprimé en détail la situation de la circulation.

Nous avons utilisé notre ensemble de données proposé pour former différents algorithmes d'apprentissage et comparer leurs performances dans différents contextes expérimentaux.

Dans cette section, nous présentons les paramètres de l'expérience et les résultats de l'utilisation de différents algorithmes d'apprentissage. Pour entraîner notre modèle, nous avons mis nos données de test dans un pipeline afin de générer des prédictions. Ensuite, nous calculons les métriques Précision ('accuracy' et 'score') pour évaluer la précision de classification en fonction de la configuration du tableau de signalisation.

Nos entrées d'apprentissage sont l'ensembles de données globaux. Nous l'avons divisé au hasard en 70 \% pour la formation et les 30 \% restants pour l'évaluation (c'est-à-dire les tests).

Les métriques de précision ('accuracy' et 'score') sont utilisées pour évaluer les performances de différents algorithmes d'apprentissage, elles sont calculées comme illustré dans l'équation du chapitre précédent.

Nous évaluons différents algorithmes d'apprentissage, y compris la Support Vector Machine (SVM), Naïve Bayes (NB) et Radom Forest (RF), qui sont tous largement utilisés dans classification supervisée.

Nous avons d'abord implémenté et construit le modèle ML pour produire trois classificateurs.

Ensuite, nous avons comparé la qualité du modèle en termes de la précision.

Les résultats sont affichés dans la figure suivante.

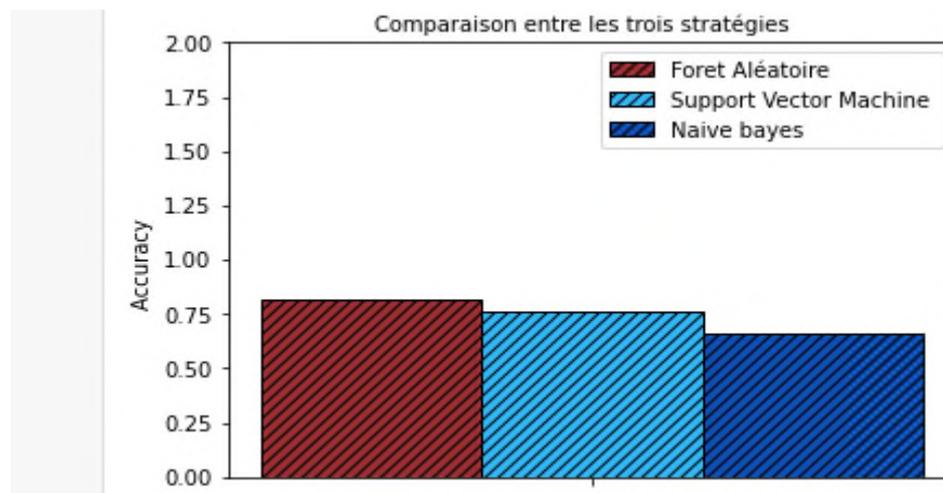


Figure 30: Comparaison entre les trois stratégies par apportAccuracy

Les modèles qui capturent plus de paramètres de contexte ont le potentiel de mieux généraliser sur l'identification du couleur de feu. C'est le cas à la fois pour RF et NB, mais pas pour SVM.

Dans cette expérience, Forêt Aléatoire a montré la plus grande précision (81 %) dans les valeurs moyennes de précision de classification que Support Vector Machine (78 %) et Naïve Bayes (71%), la différence est statistiquement significative. Le modèle Naïve Bayes est quasiment impossible de détecter le contexte correct à moins qu'un autre algorithme d'apprentissage SVM et RF.

B. Comparaison entre les trois stratégies qui nous utilisent : par apport la probabilité.

Nous avons utilisé le notre dataset et comparer les trois algorithmes en terme de probabilité.

Les résultats sont illustrés dans les figures suivante en précision et en score, respectivement.

Les résultats expérimentaux montrent que SVM a obtenu de meilleures performances que NB et RF dans l'intervalle [0,06]. Par exemple, l'algorithme RF a acquis une précision de 86 % lors de l'entraînement de notre ensemble de données. Cette précision a diminué à 6 % en utilisant SVM. À partir des résultats, nous pouvons également voir que l'algorithme RF a un Score d'une précision de 90 % dans l'intervalle [0.8,1].

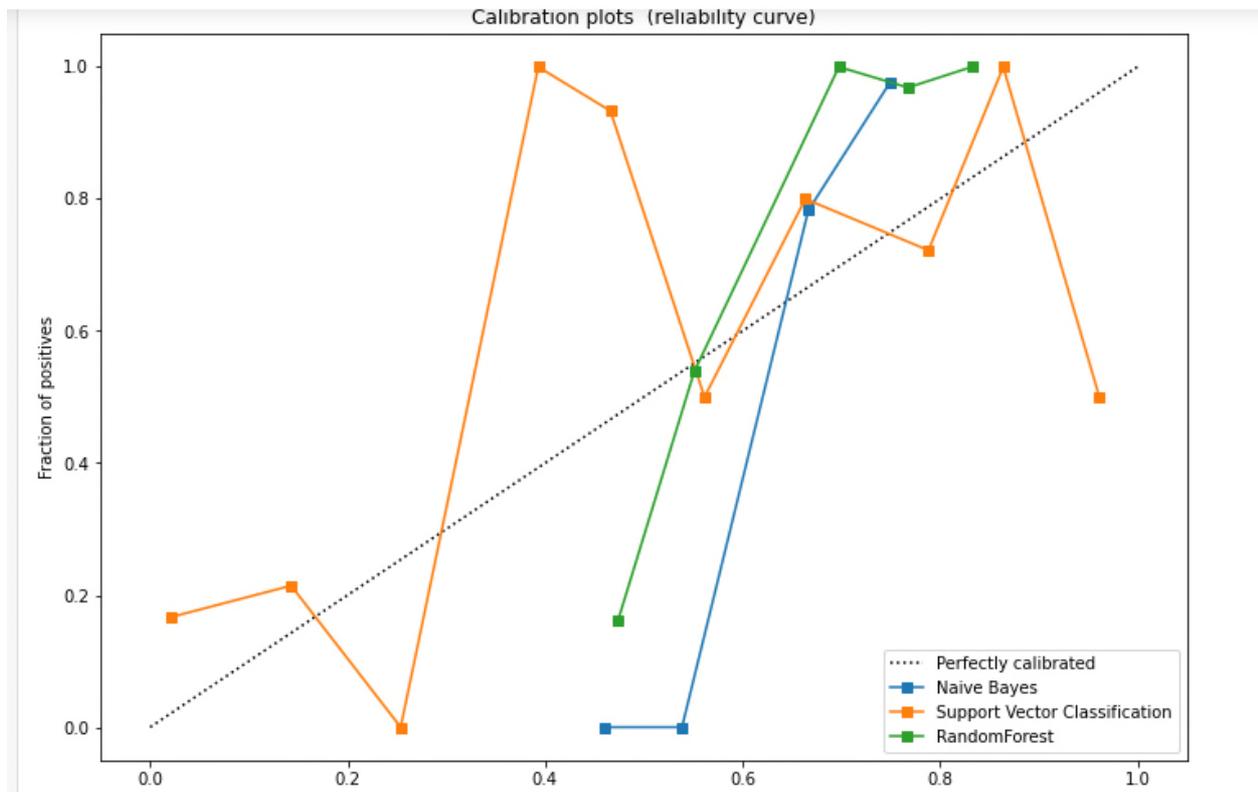


Figure 31 : Comparaison entre les trois stratégies par apport la probabilité.

C. Comparaison entre les trois stratégies qui nous utilisent :par apport Accuracy, Score.

Stratégie	Accuracy	Score
Foret Aléatoire	0.86	0.82
Réseau Bayésien	0.66	0.78
SVM	0.76	0.82

Tableau 4 :Comparaison entre les trois stratégies par apport Accuracy, Score

Nous constatons que le résultat obtenu avec le modèle : Stratégie Foret Aléatoire est le meilleur.

1.6 Conclusion

Dans ce chapitre nous avons, en premier lieu, présenté les différents outils et langages que nous avons utilisés pour implémenter notre model. Ainsi une approche basée sur les différents algorithmes d'apprentissage supervisé (Foret Aléatoire, SVM, Réseau Bayésien). On a aussi cité les différentes étapes de l'approche proposée dans le cadre de transformation de modèles, et les résultats obtenus sont créer dans un tableau.

Conclusion

Générale

Conclusion générale & Perspectives :

Dans ce mémoire, nous nous sommes intéressés à la commande des feux de signalisation en vue de résoudre les problèmes liés à la congestion ; dans ce sens, nous proposons une stratégie pour contrôler les feux de signalisation intelligente avec l'apprentissage automatique. Le modèle de prédiction est basé sur trois algorithmes de ML : forêt aléatoire (RF), support vecteur machine (SVM) et réseaux bayésiens (RB). Les solutions obtenues permettent de prédire la décision afin d'agir dans la circulation en fonction de plusieurs paramètres comme la route, voitures, les passagers etc.

En premier lieu, nous avons commencé par le contexte, problématique, l'objectif et l'organisation de notre travail.

Dans le deuxième chapitre, nous avons présenté le domaine de la circulation routière, les composants et les variables élémentaires de ce domaine, la congestion, les méthodes pour gérer quelques problèmes dans ce domaine « la circulation routière » tel que les feux signalisation dans les intersections.

Dans le troisième chapitre, nous avons proposé notre contribution qui s'articule particulièrement sur les approches supervisées, notre vision est d'utiliser le volet superviser des approches de l'apprentissage automatique (Forêt aléatoire, Machine à Vecteur de Support, réseau Bayésien), comme des stratégies pour ordonnancer intelligemment une circulation typique.

Dans l'étape de réalisation, nous avons eu l'opportunité de nous familiariser avec une panoplie d'outils récents telles que, Python, Anaconda, Jupiters, et des bibliothèques utilisées Numpy, Matplotlib, Pandas et Scikit-learn.

Notre effort sert à ouvrir une fenêtre sur les problèmes de gestion de circulation routière existante et de contribuer avec des solutions basées sur l'apprentissage automatique, une contribution qui nous semblent prometteurs et encourageants pour une application réelle.

Comme perspective de notre travail, il serait intéressant d'approfondir cette recherche à court terme avec d'autre volet d'apprentissage automatique tels que les l'approches non supervisés. et à long terme avec l'apprentissage en profondeur (Deep-Learning).

Bibliographie

Bibliographie :

- [1] - Mécanisme de coordination multi-agent fondé sur des jeux : application à la simulation comportementale de trafic routier en situation de carrefour, 2003.
- [2] - Contribution à la modélisation et à la régulation du trafic aux intersections : Intégration des communications Véhicule-Infrastructure ,2012
- [3] - Gérer la congestion urbaine Par European Conference of Ministers of Transport
- [4] - www.mediatheque.snbpe.org
- [5] - www.moodle.epfl.ch
- [6] - « Feuille de route pour un espace européen unique des transports », Livre blanc de 2011 sur les transports.
- [7] - Federal Highway Administration
- [08]-dix-solutions-pour-reduire-la-congestion-routiere-1/
<https://www.journaldequebec.com.2018/08/29>
- [09] - Coordination locale et optimisation distribuée du trafic de véhicules autonomes dans un réseau routier ,2015
- [10] - M. Abbas, H. Charara, N. Chaudhary, and Y. s. Jung. Distributed architecture and algorithm for robust real-time progress evaluation and improvement. Texas Transportation Institute, Texas A & M University System, 2006.
- [11].https://www.researchgate.net/publication/281598421_Controlle_du_trafic_routier_urbain_par_un_reseau_fixe_de_capteurs_sans_fil.
- [12] <https://techsecuritenews.com/cisco-rv320-rv325-exploit-menace-routeurs/29janvier2019>
- [13] [https://www.cisco.com/c/dam/m/fr_fr/neverbetter/assets/pdfs/cisco_nbabt_paris-smart-city_traffic_incident_management_solution-aag-cte-fr.pdf\(2019\)](https://www.cisco.com/c/dam/m/fr_fr/neverbetter/assets/pdfs/cisco_nbabt_paris-smart-city_traffic_incident_management_solution-aag-cte-fr.pdf(2019))
- [14] - *KADI Nedjma et THAMER Nour el houda*, « Développement des mécanismes et des stratégies a base des SMA pour l'optimisation d'une circulation routière »,2016.
- [15] - David Broneske. On the Impact of Hardware on Relational Join Processing. Master thesis University of Magdeburg 19 aout 2013.
- [16] - D. G. IONOS. fonctionne le machine learnings.

Bibliographie

- [17] - Kontonatsios, G., Korkontzelos, Y., Tsujii, J. I., & Ananiadou, S. Using a random forest classifier to compile bilingual dictionaries of technical terms from comparable corpora. In 14th Conference of the European Chapter of the Association for Computational Linguistics, (2014, April).
- [18] - Baptiste Gregorutti, Forêts aléatoires et sélection de variables : analyse des données des enregistreurs de vol pour la sécurité aérienne, 11 mars 2015.
- [19] - P. F. Moutarde. Arbres de Décision et Forêts Aléatoires. Paris : MINES ParisTech, PSL, 2017 (cf. p. 48, 49).
- [20] - Mohamadally Hasan et Fomani Boris, " SVM : Machines à Vecteurs de Support ou Séparateurs à Vastes Marges ". Thèse. Versailles St Quentin, France janvier 2006.
- [21] – Classification Supervisée de documents étude comparative-lahlououchiha- janvier 2016.
- [22]- Stéphane TUFFÉRY. « Data Mining et statistique décisionnelle, L'intelligence des données » 2009.
- [23] – hassanehilali, Application de la classification textuelle pour l'extraction des règles d'association maximales, avril 2009.
- [24] - Graham-Cumming, J. "Interview de John Graham-Cumming, l'auteur du logiciel antispam PopFile ", 2006.
- [25] - Denoue, L, Classification supervisée de documents, 2003.
- [26] -Le tutoriel Python.
- [27] - Anaconda Individual Edition <https://www.anaconda.com/products/individual>.
- [28] -Voilà, des dashboards à partir de vos JupyterNotebooks.
- [29] -D. Harel et A. Pnueli. « On the development of reactive systems ». In:Logics and models of concurrent systems. 1985, p. 477–498.
- [30]-H. Hilali. « Application de la classification textuelle pour l'extraction des règles d'association maximales ». Thèse de doct. Université du Québec à Trois-Rivières, 2009.
- [31] –B. Cayla. Python Pandas – Tuto <https://pandas.pydata.org/> (Partie N°1).
- [32] –E. F.M. D. John Hunter Darren Dale. Lancement de l'initiative scikit-learn, bibliothèque logicielle de référence en machine learning (cf. p. 57).