

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOHAMED BOUDIAF - M'SILA

Faculté des Mathématiques et
de l'Informatique

Département d'Informatique

N° :.....

DOMAINE : Mathématiques
et Informatique

FILIERE : Informatique

OPTION : Intelligence
Artificielle



Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par : Debih Meriem
Bouguerra Cheyma

Intitulé

Un modèle de prévision d'incendie des forêts
Etude de cas : Wilaya de M'sila

Soutenu devant le jury composé de :

Bouguerra Abdel Baki

Université de M'sila

Président

Tahar Mehenni

Université de M'sila

Rapporteur

Hammani Mohammed Said

Université de M'sila

Examineur

Année universitaire : 2021 / 2022

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOHAMED BOUDIAF - M'SILA

Faculté des Mathématiques et
de l'Informatique

Département d'Informatique

N° :.....

DOMAINE : Mathématiques
et Informatique

FILIERE : Informatique

OPTION : Intelligence
Artificielle



Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par : Debih Meriem
Bouguerra Cheyma

Intitulé

Un modèle de prévision d'incendie des forêts
Etude de cas : Wilaya de M'sila

Soutenu devant le jury composé de :

Bouguerra Abdel Baki

Université de M'sila

Président

Tahar Mehenni

Université de M'sila

Rapporteur

Hammani Mohammed Said

Université de M'sila

Examineur

Année universitaire : 2021 / 2022

Dédicaces

A Mes très chers et adorables parents, en témoignage et en gratitude de leur dévouement, de leur soutien permanent durant toutes mes années d'études.

A mes frères, au nom de leur amour et en témoignage à leur encouragement. Sans toutefois oublier mes amis qui m'ont encouragé et à tous ceux qui m'ont aidé de près ou de loin.

B. Cheyma

Dédicaces

Je dédie ce travail :

Pour mes frères et mes sœurs, qui ont toujours été mon soutien indéfectible et à toutes épreuves, ces êtres chers qui ont toujours été là pour me protéger et évacuer toutes les embûches qui entravent mon chemin vers le succès.

A mon tout cher père, l'incalculable appui et inépuisable source d'assurance et de savoir.

A ma toute chère mère, ma plus grande école de la vie, l'incarnation de l'espoir et de la bonté, la clé de tous mes succès réalisés, que Allah vous prête longue vie.

A mes amis qui m'ont encouragé et aidé à réaliser ce modeste travail

D.Meriem



Remerciements

En premier lieu, nous remercions ALLAH le Tout Puissant de nous avoir donné la force et la volonté durant nos études et pendant la réalisation de ce projet.

Nous voudrions remercier aussi notre encadreur Mr Tahar Mehenni pour son encadrement.

Comme nous tenons à remercier les membres du jury d'avoir accepté

D'examiner et juger notre travail.

Nous remercions la Conservation des Forêts de la Wilaya de M'sila qui nous ont permis d'obtenir les données de base pour notre étude.

Un remerciement particulier à nos familles pour nous avoir soutenus, accompagnés et nous avoir permis d'en arriver là.

Enfin, nous remercions tous nos amis pour leur aide et leurs encouragements et tous ceux qui ont contribué de près ou

De loin à la réalisation de ce projet.

Table des matières

Table des matières.....	I
Liste des figures	Erreur ! Signet non défini.
Liste des tableaux	Erreur ! Signet non défini.
Liste des abréviations.....	II
Introduction Générale	1
CHAPITRE 1 FEU DE FORÊT	1
Introduction.....	1
1. Définitions.....	1
2. Présentation générale des incendies	2
3. Les différents types de feux.....	3
3.1. Les feux de profondeur (sol).....	3
3.2. Les feux de surface	3
3.3. Les feux de cime	3
3.4. Les sautes de feu (braises)	3
4. Les causes d'incendie.....	4
4.1. Les causes naturelles.....	4
4.2. Les causes humaines	4
5. Les facteurs influents dans la propagation du feu	5
5.1. Les caractéristiques de la végétation	5
5.2. Les conditions climatiques.....	5
5.3. Les facteurs topographiques :	6
6. Conséquences d'un incendie	6
7. Méthodes de prévention	7
8. Incendies de forêt en Algérie.....	8
Conclusion.....	9
CHAPITRE 2 APPRENTISAGE AUTOMTIQUE	10
Introduction.....	10

1. L'intelligence Artificielle	10
2. Apprentissage automatique (Machine Learning)	11
2.1. Apprentissage Supervisé (Supervised Learning)	12
2.1.1. La Régression	12
2.1.1.1. Régression linéaire.....	12
2.1.2. La Classification	14
2.1.2.1. Le Classifieur Bayésien Naïf (Naïve Bayes)	15
2.1.2.2. L'Arbre de Décision	16
2.1.2.3. Machines à Vecteurs de Support (Support Vector Machines).....	18
2.2. Apprentissage Non Supervisé (Unsupervised Learning).....	19
2.3. Apprentissage par Renforcement (Reinforcement Learning).....	21
2.4. Apprentissage en Profondeur (Deep Learning)	21
2.5. Apprentissage Ensembliste (Ensemble Learning)	22
Conclusion.....	23
CHAPITER 3 L'APPROCHE PROPOSÉE	24
Introduction	24
1. L'état de l'art	24
2. Conservation des forêts de M'sila	25
2.1. Définition	25
.3 Lieu d'étude.....	26
4. Méthodologie et préparation des données	26
5. Les outils et détails d'implémentation	28
5.1. Langage de programmation	28
5.1.1. Python	28
5.1.2. Pourquoi Python ?.....	29
5.2. Environnement de développement.....	29
5.2.1. Jupyter.....	29
5.2.2. Sklearn	30
5.2.3. Pandas	30
5.2.4. Matplotlib.....	30
5.2.5. NumPy	30

Table des matières

5.2.6. Matrice de Confusion.....	31
6. L'apprentissage du modèle proposé (Classification).....	32
6.1. Présentation des ensembles d'apprentissage utilisés	32
6.2. L'importation et le partitionnement de la base de données	35
6.3. Naïf Bayésien.....	37
6.4. Support Vector Machine (SVC) :	39
6.5. Arbre de décision (Decision tree	41
Conclusion.....	45
CHAPITRE 4 RÉSULTATS ET DISCUSSION.....	46
Introduction :.....	46
1. Résultats d'évaluation des approches d'apprentissage automatique sur les trois classes de donnée :.....	46
1.1. Résultat de 1 ^{er} Ensemble d'apprentissage (BDD 1).....	46
1.2. Résultat de 2 ^{-ème} Ensemble d'apprentissage (BDD 2)	48
1.3. Résultat de 3 ^{-ème} Ensemble d'apprentissage (BDD 3) :	49
1.4. Résultat de les 3 Ensembles d'apprentissage.....	50
2. Limites de la recherche.....	50
Conclusion :	50
Conclusion générale	52
Bibliographie	54

Liste des figures

Figure 1 1: Représente Triangle du feu.[3]	2
Figure 1 2: Types de feux de forêts [5].	3
Figure 1 3: Deux exemples de différentes combinaisons de forces d’alignement pour une même forêt. [6]	6
Figure 2 1: Apprentissage supervisé [14]	12
Figure 2 2: Un modèle de régression linéaire simple [16]	14
Figure 2 3: Organigramme de l’algorithme Naïve Bayes.[19]	16
Figure 2 4: Exemple de l’arbre de décision.....	17
Figure 2 5: Méthode à noyau SVM. [22]	18
Figure 2 6: Organigramme de l’algorithme SVM.	19
Figure 2 7: Apprentissage non supervisé [14]	20
Figure 2 8: Les sous-branches de l’intelligence artificiel [13].....	22
Figure 3 1: nombre d’incendies dans M’sila (2011-2021)	27
Figure 3 2: Extrait de la Base de données en Excel.....	28
Figure 3 3: Logo de l’IDE Jupyter.....	30
Figure 3 4: Logo de bibliothèque Sklearn.....	30
Figure 3 5: les graphiques des variable indépendantes.....	33
Figure 3 6: Extrait de l’ensemble d’apprentissage pour la prédiction d’incendie (BDD1).	34
Figure 3 7: Graphique de la variable dépendante Incendie ou non.....	34
Figure 3 8: Extrait de l’ensemble d’apprentissage pour la prédiction d’incendie des maquis (BDD2).....	35
Figure 3 9: Graphique de la variable dépendante Maquis ou autres.	35
Figure 3 10: Extrait de l’ensemble d’apprentissage pour la prédiction d’incendie des forêts (BDD3).....	36
Figure 3 11: Graphique de la variable dépendante Forêt ou autres.	36
Figure 3 12: L’importation de l’ensemble d’apprentissage et la définition des variables.	36
Figure 3 13: Code de Partitionnement de l’ensemble d’apprentissage.	37
Figure 3 14: Lignes des codes pour obtenu la courbe de ROC.	37
Figure 3 15: modèle Naïve Bayes et matrice de confusion.	38
Figure 3 16: Les résultats de matrice de confusion BDD1.	38
Figure 3 17: Les résultats de matrice pour BDD 2.	39
Figure 3 18: Les résultats de matrice pour BDD 3.....	39
Figure 3 19: le code de module GridSearchCV.	41

Liste des figures

Figure 3 20: La matrice de confusion.	41
Figure 3 21: Les résultats de matrice de confusion SVC BDD1.	42
Figure 3 22: Les résultats de matrice pour BDD 2.	42
Figure 3 23: Les résultats de matrice pour BDD 3.	42
Figure 3 24: Le model de l'Arbre de Décision.	44
Figure 3 25: Les résultats de matrice de confusion Arbre Décision BDD1.....	44
Figure 3 26: Les résultats de matrice de confusion Arbre Décision BDD2.....	45
Figure 3 27: Les résultats de matrice de confusion Arbre Décision BDD3....	45
Figure 3 28: un code pour dessinée l'Arbre de Décision.....	45
Figure 3 29: le graphe de arbre de décision (BDD1).....	46
Figure 3 30: Le graphe d'Arbre de Décision (BDD2).....	46
Figure 3 31: Le graphe de arbre de décision (BDD3).....	47
Figure 4 1: Graph de Accuracy pour BDD1.....	50
Figure 4 2: La courbe de caractéristique de fonctionnement du récepteur (ROC) de SVC pour BDD1.....	50
Figure 4 3: Graph de Accuracy pour BDD2.....	51
Figure 4 4 : La courbe de caractéristique de fonctionnement du récepteur (ROC) de Naive Bayes pour BDD2.....	52
Figure 4 5: Graph de Accuracy pour BDD 3.....	52
Figure 4 6 : La courbe de caractéristique de fonctionnement du récepteur (ROC) de SVC pour BDD3.....	53
Figure 4 7 : Performance des algorithmes d'apprentissage sur les ensembles d'apprentissage (BDD1, BDD2, BDD3).....	53

Liste des tableaux

Table 1 1: Répartition des superficies brûlées (ha) en Algérie par types d'occupation du sol en 2021. [8].....	9
Table 3 1: L'état de l'art.....	25
Table 3 2: Attributs de l'ensemble de données.....	28
Table 3 3: Exemple de Matrice de Confusion.....	31
Table 3 4: Résultats obtenus par le modèle Naïve bayes.....	40
Table 3 5: Le modèle de SVC.....	40
Table 3 6: Les résultats obtenus en SVC.....	43
Table 3 7: Les résultats obtenus en Arbre de Décision.....	47
Table 4 1: Pourcentage de Accuracy pour BDD1.....	49
Table 4 2: Pourcentage de Accuracy pour BDD 2.....	51
Table 4 3: Pourcentage de Accuracy pour BDD 3.....	52

Liste des abréviations

FAO: Organisation des Nations unies pour l'alimentation et l'agriculture (en anglais Food and Agriculture Organization of the United Nations)

IA : Intelligence Artificiel

ML : Machine Learning

DT: Decision Tree

SVM : Support Vector Machines

SVC : Support Vector Classification

PCA : Analyse des Composants Principaux

LDA : Analyse discriminante linéaire

ACC : Accuracy

RBF : RADIAL BASIS FUNCTION

BDD : Base de données

Introduction Générale

Les crises naturelles tels que les tremblements de terre, les tornades, les inondations et les incendies de forêt ne causent pas seulement une détérioration du terrain, mais menacent aussi la vie des êtres humains, directement ou indirectement. L'incendie de forêt est l'une des catastrophes naturelles les plus complexes auxquelles est confrontée notre communauté.

Les études statistiques mondiales ont montré que la fréquence des incendies augmente jour après jour en conduisant à des résultats catastrophiques sur la vie humaine et la biodiversité. Ces conséquences ont attiré l'attention des gouvernements, des donateurs et des organisations non gouvernementales pour mettre en œuvre de nouvelles stratégies et donc trouver des solutions pour cette crise.

La prévention des incendies est considérée comme l'une des techniques les plus importantes qui aident à la réduction des conséquences dangereuses d'incendie.

Pour ces raisons, les scientifiques ont travaillé dur pour trouver des stratégies et des politiques efficaces qui aident à la prédiction de l'incidence du feu et le traitement de sa diffusion afin de limiter les dégâts et réduire son danger sur nos sociétés.

L'intelligence artificielle et apprentissage automatique font partie des stratégies que les scientifiques utilisent pour prédire les incendies de forêt en raison de son importance effective dans l'accomplissement rapide de nombreuses tâches et la réduction de nombreux efforts humains.

Dans ce contexte, nous proposerons, dans ce mémoire, une approche basée sur l'apprentissage automatique supervisé, pour prédire les incendies de forêts. Ce modèle est évalué sur la base d'une série d'incendies de forêt de M'sila pris au cours des dernières années.

Ce mémoire est organisé comme suit :

- Dans le chapitre 1, nous allons examiner en détail le feu de forêt, sa définition, ses types, ses caractéristiques, ses causes, ses facteurs, ses résultats et ses méthodes de prévention. Ensuite, nous présenterons également les statistiques relatives à cette catastrophe, notamment en Algérie.
- Au début du chapitre 2, nous allons présenter le domaine de l'intelligence artificielle, l'apprentissage automatique et les différentes branches et caractéristiques

Introduction Générale

d'apprentissage automatique . A la fin du chapitre, nous discuterons quelques travaux relatifs à notre problématique.

- Le Chapitre 3 sera consacré au détail de notre étude. Notamment une description détaillée de l'approche et l'architecture proposée, ensuite une explication détaillée de la méthode de fonctionnement de notre proposition.
- En fin le chapitre 4 donnera Analyse des résultats de performances de notre modèle proposé et une comparaison des résultats obtenus
- Le mémoire s'achève par une conclusion, qui récapitule notre approche, et les perspectives de recherche.

CHAPITRE 1

FEU DE FORET

Introduction

Dans ce chapitre, nous aborderons en détail les incendies de forêt, les causes et les facteurs de ce phénomène, et nous présenterons également quelques statistiques liées à ce phénomène en Algérie.

1. Définitions

Le feu

Le feu est défini comme étant un dégagement simultanément de chaleur, de lumière et de la flamme produite par la combustion vive dans certains corps (bois, feuille, tapis herbacé ...etc.). [1]

L'incendie

L'incendie est une combustion qui se développe sans contrôle dans le temps et l'espace, contrairement au feu qui est une combustion maîtrisée [2].

L'incendie de forêt

Selon l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture [1], la définition des feux de forêts varie selon les pays en fonction de celle de la forêt, mais aussi selon des critères annexes : lieu d'éclosion et distance par rapport à la forêt, nature du propriétaire, taille du feu...

D'après le Ministère de l'Écologie et du Développement Durable de France [3], l'incendie de forêt est le feu concerne une surface minimale d'un hectare d'un seul tenant et qu'une partie au moins des étages arbustif et/ou arboré (parties hautes) est détruite. Un incendie est un phénomène qui échappe au contrôle de l'Homme, tant en durée qu'en étendue.

Pour qu'il y ait inflammation et combustion, trois facteurs doivent être réunis, chacun en proportions convenables : un combustible, qui peut être n'importe quel matériau pouvant

brûler, une source externe de chaleur (flamme ou étincelle) et de l'oxygène, nécessaire pour alimenter le feu.



Figure 1-1: Représente Triangle du feu.[3]

La forêt

La FAO, définit les forêts comme des terres occupant une superficie de plus de 0,5 hectare (5000 m²), avec des arbres atteignant une hauteur supérieure à 5 mètres et un couvert forestier de plus de 10%. Cette définition exclut les terres dont la vocation prédominante est agricole ou urbaine. (4)

2. Présentation générale des incendies

La forêt, dans son intégralité, doit être considérée comme un combustible potentiel. Les flammes peuvent en effet parcourir indifféremment la végétation vivante (branches, feuilles) ou morte (aiguilles, arbres morts sur pied), tout comme les infrastructures humaines implantées en zone forestière.

Les incendies concernent, outre les forêts au sens strict, des formations sèches forestières de petite taille :

- Le maquis est une formation fermée et dense, poussant sur un sol siliceux ;
- La garrigue est une formation plutôt ouverte, poussant sur un sol calcaire ;
- Les landes sont des formations sur sols acides, composées de genêts et de petits arbustes [3].

3. Les différents types de feux

Selon les caractéristiques végétales et les conditions climatiques dans lesquelles le feu de forêt se développe, quatre types de feux se distinguent.



Figure 1-2: Types de feux de forêts [5].

3.1. Les feux de profondeur (sol)

Le feu se propage grâce à des racines et des matières organiques dans des conditions particulières (écosystèmes de tourbière). Même si les flammes ne sont pas visibles, le feu peut rester actif pendant de longues périodes [6].

3.2. Les feux de surface

Brûlent les strates basses et contiguës au sol (litière, tapis herbacé, broussailles). Ce sont les plus communs. Ils se propagent rapidement, en dégageant beaucoup de flammes et de chaleur [1]

3.3. Les feux de cime

Feu qui se déplace sur le faîte des arbres ou des arbustes, plus ou moins indépendamment du feu de surface et leur vitesse de propagation est très élevée [7].

3.4. Les sautes de feu (braises)

Les feux les plus graves sont les « feux totaux » qui brûlent tant la surface que les cimes. Lorsqu'ils atteignent des niveaux d'intensité élevés, les incendies de forêt sont capables de transporter des matières en ignition ou incandescentes (brandons) sur plusieurs centaines de mètres depuis le front du feu et d'allumer des foyers secondaires [6].

4. Les causes d'incendie

L'origine d'un incendie est souvent difficile à déterminer du fait de l'absence de preuves matérielles concrètes ; il en résulte que le pourcentage de causes inconnues peut être très important. La cause d'un incendie peut être connue selon différents niveaux de certitudes.

4.1. Les causes naturelles

La végétation ne s'enflammant pas seule, même par forte sécheresse, l'unique cause naturelle connue dans le Bassin Méditerranéen est la foudre. Ce phénomène- ne, très répandu en forêt boréale ("orages secs"), est relativement rare en région méditerranéenne où il ne concerne que 1 à 5 % des cas d'incendies. Des exceptions peuvent toutefois être observées, notamment en Espagne, où, dans certaines régions, la foudre représente 30 % des départs de feu (Aragon : 38 % et Castille-la Manche : 29 %). Les éruptions volcaniques peuvent également être à l'origine d'incendies de forêt. Ce phénomène est cependant exceptionnel dans le Bassin Méditerranéen. [1]

4.2. Les causes humaines

Elles représentent l'essentiel des origines des incendies de forêts. Globalement, pour l'ensemble des pays du Bassin Méditerranéen, on retrouve des causes involontaires et des causes volontaires. Leur répartition dépend étroitement du contexte social, économique, politique et législatif de chaque pays.

- **Les causes accidentelles** provenant des infrastructures de transport comme les chemins de fer ou les automobiles, des lignes électriques et des dépôts d'ordures, dans lesquels des feux peuvent couvrir.
- **Imprudence** par des jets de mégots, des pique-niques en forêt ou des jeux d'enfants.
- **La déprise agricole** : le déclin des activités agricoles et pastorales et l'abandon des espaces ruraux favorisent l'extension des friches, zones potentielles de départs d'incendie.
- **L'urbanisation** : le mitage, par l'expansion des habitations aux abords des zones boisées, voire au sein même de la forêt, multiplie les points potentiels de départs d'incendies accidentels.

5. Les facteurs influents dans la propagation du feu

Les conditions météorologiques et les caractéristiques de la végétation conditionnent le développement des incendies, les premières pouvant avoir une influence non négligeable sur les secondes. Dans certaines situations (forts vents par exemple), la topographie du site peut également favoriser le développement des incendies.

5.1. Les caractéristiques de la végétation

La prédisposition de la végétation aux incendies est souvent liée à sa teneur en eau, elle-même déterminée par les conditions météorologiques. L'état général de la zone forestière, c'est-à-dire les conditions de peuplement de la forêt (disposition des strates, état d'entretien, densité, essences présentes, passage récent d'un incendie) et la composition chimique de la végétation (teneur en essences volatiles ou en résines), jouent également un rôle déterminant dans l'éclosion des incendies. [3]

La composition floristique et l'état biologique des peuplements végétaux déterminent leur degré d'inflammabilité et de combustibilité. Ainsi quelques forêts brûlent plus facilement que d'autres grâce à leur richesse en conifères qui contient une forte teneur en résine très inflammable et une faible teneur en eau.

5.2. Les conditions climatiques

Les conditions climatiques jouent un rôle clé dans le niveau de risque d'incendie. La capacité de propagation du feu s'accroît avec l'augmentation de la température, de la vitesse du vent et une baisse de l'humidité de l'air ambiant. Les notions de puissance et de vitesse de propagation du feu dépendent largement de ces paramètres [6].

Le jargon technique il est souvent fait allusion au principe des « trois tiers » pour désigner les conditions susceptibles de donner lieu à un grand incendie de forêt : des températures supérieures à 30°C, une vitesse du vent supérieure à 30 km/h et une humidité relative inférieure à 30%. Plus les conditions s'approchent de ces valeurs et plus elles perdurent (avec un impact direct sur la teneur en eau du combustible), plus le risque est élevé [6].

5.3. Les facteurs topographiques :

Les variations topographiques influent sur les précipitations, les eaux de ruissellement, la température, le vent et le rayonnement solaire, qui à leur tour ont une incidence sur l'inflammabilité à travers la production du combustible et sa teneur en eau. L'inclinaison d'une pente amplifie l'effet de radiation et de convection et influence aussi la vitesse de propagation du feu, surtout lorsque le vent rajoute son action en produisant l'inclinaison des flammes.

Dans une zone sans relief, un départ de feu est facilement soumis à l'accélération du vent. En zone de relief irrégulier, la progression du feu est accélérée dans les montées et ralentie dans les descentes. [2]

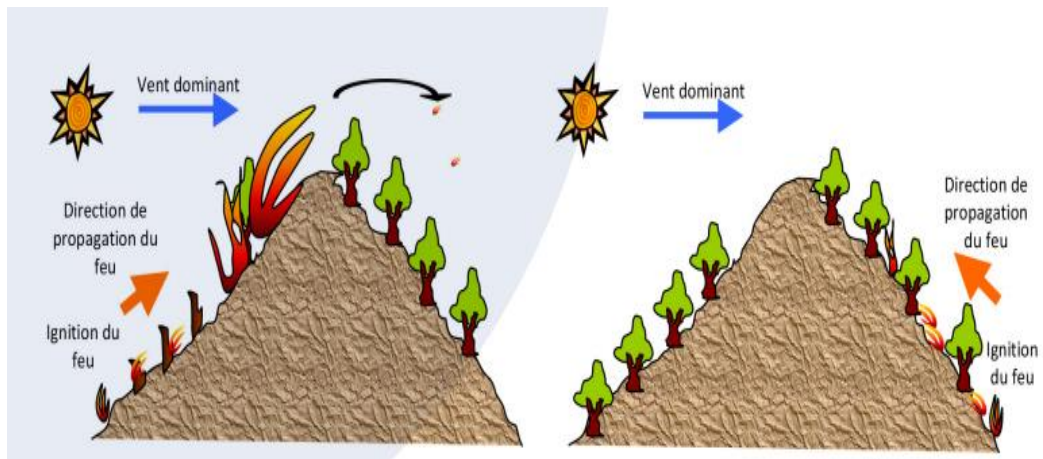


Figure 1-3: Deux exemples de différentes combinaisons de forces d'alignement pour une même forêt. [6]

6. Conséquences d'un incendie

Sur les plantes :

Le passage d'un feu se traduit par l'altération plus ou moins poussée d'organes vitaux du végétal, au niveau du feuillage, du tronc et des racines. Il en découle une perte de vigueur de l'arbre pouvant entraîner sa mort.

Le degré d'altération est fonction de la combinaison des dégâts sur les différentes parties de l'arbre (feuillage, tronc, racines), résultant de la nature (feu de surface, feu de cime) et de l'intensité du feu, ainsi que de la sensibilité au feu de l'espèce. Un feu rapide provoque beaucoup moins de dommages qu'un feu lent. [1]

Sur Le Sol :

Le passage du feu agit sur la structure et la composition du sol et sur les microorganismes présents.

- Sur la structure Le feu induit une diminution de la stabilité des agrégats conduisant à une structure particulière. Cette transformation entraîne la réduction de la capacité de rétention en eau et du taux d'infiltration de l'eau, la porosité des sols étant comblée par des particules.
- Sur la composition Le feu entraîne une minéralisation rapide de la biomasse, des litières et des humus, qui court-circuite le cycle biologique normal, au détriment des décomposeurs de l'écosystème (champignons, bactéries, microfaune).

Sur les animaux :

L'incendie affecte de façon différente les divers groupes faunistiques : certains ne survivent pas, brûlés ou asphyxiés par les fumées ; d'autres échappent au feu en fuyant. Le feu perturbe en outre de façon indirecte les cycles biologiques des animaux.

Sur le paysage :

L'incendie entraîne un changement brutal du paysage transformant le cadre de vie de la population en un environnement calciné. La disparition d'une végétation basse semble toutefois plus facile à accepter que celle des arbres d'une forêt.

Sur l'économie :

Les différents éléments qui constituent le coût d'un incendie sont :

- Les coûts directs : lutte contre le feu, équipements détruits (habitations, infrastructures, véhicules), forêts détruites.
- Les coûts indirects : perte des usages, restauration de la végétation et des paysages, incidence sur l'économie du tourisme et des loisirs.

7. Méthodes de prévention

Afin de s'adapter aux incendies qui seront éventuellement plus fréquents et plus intenses, une stratégie de prévention des incendies doit d'abord être élaborée. Cette stratégie doit être pensée comme un système qui intègre différents secteurs. La sensibilisation par l'implication de la société civile est essentielle. La cartographie des zones sensibles aux

incendies de forêt aidera à prévoir les risques. Le nettoyage saisonnier et l'exploitation forestière linéaire (barrières coupe-feu) sont également des interventions importantes avant le début de la saison sèche.

8. Incendies de forêt en Algérie

En 2021, la superficie brûlée totale dans le monde est de 1 113 464 ha, un total similaire à celui cartographié en 2020.

La superficie brûlée totale cartographiée en Afrique du Nord et au Moyen-Orient était très similaire à celle de 2020 et un peu pire que la moyenne à long terme, mais avec de grandes différences au sein des nations individuelles. La Tunisie a connu une année pire qu'en 2020, tandis que la saison de la Libye était meilleure. Le pays le plus touché de la région a été l'Algérie, représentant 69 % de la superficie brûlée totale. [8]

En Algérie, La superficie totale brûlée cartographiée était la plus élevée depuis 2012. 295 incendies ont été cartographiés, soit une superficie brûlée totale de 31 275 ha, dont les deux tiers sur des terres agricoles. Le premier incendie de la saison a été cartographié en février et le dernier en novembre, mais 85 % des dégâts se sont produits en août. Le plus grand incendie de la saison a dépassé 25 000 ha, et il y a eu 20 autres incendies de plus de 1 000 ha et 15 de plus de 500 ha[8].

Couverture terrestre	Superficie brûlée (ha)	% Du total
Forêt de feuillus	20466	15,2
Forêt de conifères	3171	2,4
Forêt mixte	105	0,1
Transitionnel	19711	14,7
Autres terres naturelles	4720	3,5
Agriculture	85934	64,0
Autre couverture terrestre	6	0,0
Surfaces artificielles	161	0,1

Table 1-1: Répartition des superficies brûlées (ha) en Algérie par types d'occupation du sol en 2021. [8]

Conclusion

Une nette tendance vers une augmentation de la fréquence des incendies est notée lors des dernières décennies. La saison des incendies 2021 a été marquée par un certain nombre de très grands incendies en août, en particulier dans les pays européens et en Algérie [8], faisant de ce mois, le mois le plus dévastateur et représentant près de la moitié de la superficie annuelle totale brûlée.

Cela aura des conséquences graves sur un écosystème déjà fragile, si on ne maîtrise pas ce facteur écologique qu'est le feu. Cette propagation doit être arrêtée, c'est pourquoi dans le chapitre suivant, nous définirons les méthodes et techniques de prédiction les plus importantes en apprentissage automatique.

CHAPITRE 2

APPRENTISAGE AUTOMTIQUE

Introduction

Actuellement, l'informatique est presque présente dans tous les domaines : la santé, l'éducation, l'économie et la cosmologie. Cette présence se reflète dans la vie quotidienne de l'individu et elle a permis des facilités d'utilisation et de compréhension de plusieurs complexes domaines. Un des domaines les plus importants qui ont été touchés par ce terrible développement de l'informatique est le domaine de Protection de l'environnement. Le développement de l'informatique et de la technologie continue à prendre une place de plus en plus importante dans le domaine de Protection de l'environnement, ce qui a permis le développement des logiciels de surveillance qui augmentent la précision des résultats. Ce dernier a donné une grande attention par les scientifiques, en particulier les spécialistes dans les domaines : de l'intelligence artificielle, les systèmes experts, l'apprentissage automatique « Machine Learning ».

Dans ce chapitre, nous allons présenter les techniques d'apprentissage automatique et, nous discuterons à la fin du chapitre de quelques travaux basés sur l'apprentissage automatique « Machine Learning » pour la prédiction des feux de forêt.

1. L'intelligence Artificielle

L'intelligence artificielle est née dans les années 50, quand une poignée de pionniers du domaine naissant de l'informatique, ont commencé à se demander si les ordinateurs pouvaient être amenés à « penser », une question dont nous explorons encore aujourd'hui les ramifications. Une définition concise du champ serait la suivante : l'effort d'automatiser les tâches intellectuelles normalement effectuées par les humains. En tant que tel, l'IA est un domaine général qui englobe l'apprentissage automatique et l'apprentissage en profondeur, mais qui comprend également beaucoup plus d'approches qui n'impliquent aucun apprentissage.

Les programmes d'échecs initiaux, par exemple, ne concernaient que des règles codées en dur élaborées par des programmeurs et ne se qualifiaient pas comme apprentissage automatique. Pendant un temps assez long, de nombreux experts ont estimé que l'intelligence

artificielle au niveau humain, pouvait être obtenue en faisant en sorte que les programmeurs fabriquent à la main un ensemble suffisamment large de règles explicites pour manipuler les connaissances. Cette approche est connue sous le nom d'IA symbolique et elle était le paradigme dominant de l'IA des années 50 et à la fin des années 80. Elle a atteint son pic de popularité durant le boom des systèmes experts des années 80. Bien que l'IA symbolique se soit révélée appropriée pour résoudre des problèmes logiques bien Définis, comme jouer aux échecs, il était difficile de trouver des règles explicites pour résoudre des problèmes flous plus complexes, tels que la classification des images, la reconnaissance de la parole et la traduction. Une nouvelle approche est apparue pour prendre la place de l'IA symbolique : l'apprentissage automatique. [9]

2. Apprentissage automatique (Machine Learning)

Le machine Learning (ML) est une branche de l'intelligence artificielle qui permet aux ordinateurs d'apprendre sans être explicitement programmés pour cela [10]. Le ML est constitué d'un ensemble d'algorithmes avancés par lesquels les résultats peuvent être prédits sur la base d'un modèle préparé et formé sur des données d'entrée historiques et son comportement de sortie [11]. Cependant, selon l'approche machine Learning choisie, les performances et les avantages peuvent varier. De nos jours, les techniques du ML sont appliquées dans plusieurs domaines de fabrication tels que la maintenance, l'optimisation, le dépannage et le contrôle [12].

Durant ces dernières années, le ML est devenu de plus en plus important dans le domaine informatique car les données peuvent être collectées et stockées beaucoup plus facilement. Les données collectées sont souvent si volumineuses qu'il n'est pas pratique de les analyser manuellement. Dans un tel scénario, la technique de ML joue un rôle clé. Une autre raison de la popularité croissante du ML est la réduction des coûts de calcul. Avec l'évolution du matériel ces dernières années, l'utilisation d'approches du ML en maintenance industrielle est devenue efficace en termes de temps et d'argent, notamment pour la détection des types de pannes et pour la maintenance prédictive.

Objectif : Le ML possède de nombreuses familles d'algorithmes qui répondent souvent à des questions précises :

- Explorer ou valider, représenter, décrire des variables et des observations d'un échantillon.

- Expliquer ou tester les effets de variables ou de facteurs du modèle supposés connus a priori.
- Prédire un phénomène et choisir un ensemble de meilleurs prédicteurs.

2.1. Apprentissage Supervisé (Supervised Learning)

L'apprentissage supervisé commence généralement par un ensemble établi de données et une certaine compréhension de la façon dont ces données sont classées. L'apprentissage supervisé vise à trouver des modèles dans les données qui peuvent être appliqués à un processus d'analyse. Ces données ont des caractéristiques étiquetées qui définissent la signification des données. Par exemple, il pourrait y avoir des millions d'images d'animaux et inclure une explication de ce qu'est chaque animal, puis vous pouvez créer une application d'apprentissage automatique qui distingue un animal d'un autre [31]. (Voir figure 2.1).

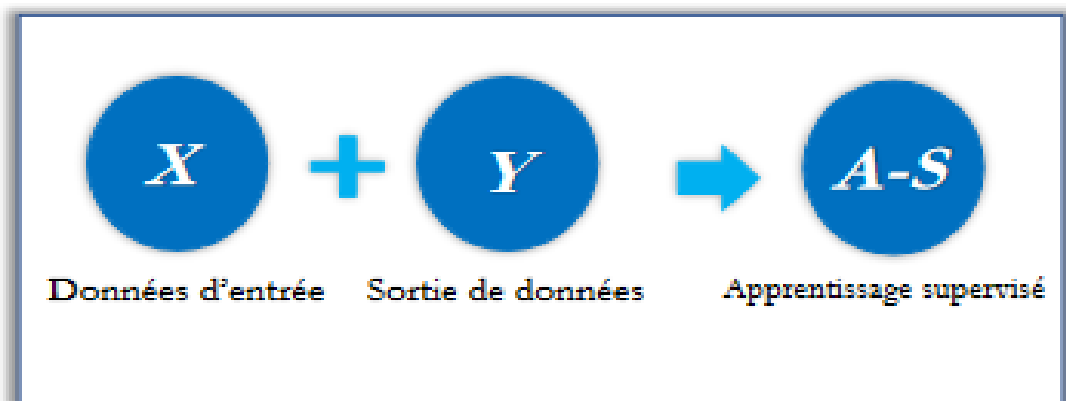


Figure 2-1 : Apprentissage supervisé [14]

2.1.1. La Régression

C'est la tâche d'apprentissage d'un modèle dont les variables dépendantes (ou variables de sortie) sont continues. C'est l'un des problèmes les plus étudiés avec des applications dans presque tous les domaines et ainsi, de nombreux modèles de régression ont été proposés, tels que la régression linéaire avec régularisation, la régression à vecteurs de support, la régression polynomiale, ...etc. [15].

2.1.1.1. Régression linéaire

La régression linéaire est l'une des plus anciennes techniques de prédiction en statistique. L'objectif de la régression linéaire est d'adapter un modèle linéaire entre la réponse et les variables indépendantes, et de l'utiliser pour prédire le résultat compte tenu d'un ensemble

de variables indépendantes observées. Un modèle de régression linéaire simple utilise une formule de la forme :

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + \dots + \epsilon \quad (2.1)$$

Où :

- Y est la variable de réponse (c'est-à-dire le résultat qu'on essaie de prédire).
- X1, X2, X3, etc. sont les variables indépendantes utilisées pour prédire le résultat.
- b0 est une constante qui est l'intersection de la droite de régression.
- b1, b2, b3, etc. sont les coefficients des variables indépendantes. Il s'agit des pentes partielles de chaque variable. - ϵ est l'erreur ou le bruit associé à la variable de réponse qui ne peut pas être expliqué par les variables indépendantes X1, X2, et X3.

Un modèle de régression linéaire a deux composantes : une partie déterministe (i.e. $b_1X_1 + b_2X_2 + \dots$) et une partie aléatoire (i.e. l'erreur, ϵ). On peut considérer ces deux composantes comme le signal et le bruit dans le modèle. Si On n'a qu'une seule variable d'entrée X, le modèle de régression est la meilleure ligne qui correspond aux données. La figure 2.2 montre un exemple de modèle de régression linéaire simple. Avec deux variables d'entrée, la régression linéaire est le meilleur plan qui s'adapte à un ensemble de points de données dans un espace 3D. Les coefficients des variables (i.e. b1, b2, b3, etc.) sont les pentes partielles de chaque variable. Si on maintient toutes les autres variables constantes, le résultat Y augmentera de b1 lorsque la variable X1 augmentera de 1. C'est pourquoi les économistes utilisent généralement l'expression « ceteris partibus » ou « toutes choses étant égales par ailleurs » pour décrire l'effet d'une variable indépendante sur un résultat donné.

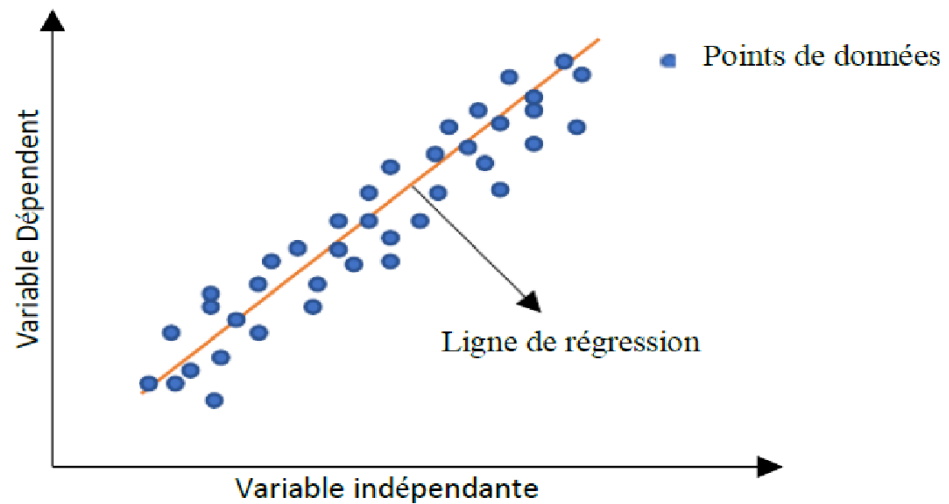


Figure 2-2: Une relation linéaire simple entre deux variables [16]

La régression linéaire utilise les méthodes des moindres carrés ou de descente de gradient pour trouver les meilleurs coefficients de modèle pour un ensemble de données fournies. La méthode des moindres carrés y parvient en minimisant la somme de l'erreur quadratique entre les valeurs ajustées et réelles de chaque observation dans les données d'apprentissage. La descente de gradient trouve les coefficients de modèle optimaux en mettant à jour les coefficients à chaque itération. Les mises à jour vont dans le sens où la somme des erreurs entre le modèle ajusté et les valeurs réelles des données d'apprentissage est réduite. Par plusieurs itérations, il trouve le minimum local en se déplaçant dans le sens du gradient négatif [16].

2.1.2. La Classification

Dans ML et les statistiques, la classification est le problème qui consiste à identifier à quel groupe de catégories (sous-populations) une nouvelle observation appartient, à partir d'un ensemble d'apprentissages contenant des données dont l'appartenance à une catégorie est connue. Par exemple, attribuer un courrier électronique donné à la classe « spam » ou « non-spam » et attribuer un diagnostic à un patient donné en fonction des caractéristiques observées du patient (sexe, pression artérielle, présence ou non de certains symptômes, ...etc.).

La classification est considérée comme un exemple d'apprentissage supervisé et de reconnaissance de formes, c'est-à-dire un apprentissage dans lequel un ensemble de

formations d'observations correctement identifiées est disponible. La procédure non supervisée correspondante est connue sous le nom de *Clustering*. Elle consiste à regrouper des données en catégories en fonction d'une mesure de la similarité inhérente ou de la distance [17].

Les modèles de classification sont utilisés pour affecter des éléments à un groupe ou une classe discrète en fonction d'un ensemble spécifique de fonctionnalités. Chaque modèle a ses propres forces et faiblesses dans un scénario donné. Le choix d'un modèle de classification des données est étroitement lié à l'analyse de rentabilisation et à une solide compréhension de ce qu'on essaye d'accomplir. Le choix de la bonne approche nécessite souvent des essais et des erreurs pour atteindre le bon équilibre entre complexité, performances et précision. Les sections suivantes décrivent certains des modèles courants qui sont utiles à connaître [18].

2.1.2.1. Le Classifieur Bayésien Naïf (Naïve Bayes)

La classification naïve bayésienne repose sur l'hypothèse que toutes les caractéristiques sont conditionnellement indépendantes les unes des autres. Cette méthode est basée sur le théorème de Bayes qui calcule la probabilité d'un événement à l'aide de la connaissance au préalable des conditions connexes. Ce théorème a été découvert par un statisticien anglais, Thomas Bayes, au 18ème siècle mais il n'a jamais publié son travail. Après son décès, ses notes ont été éditées et publiées par le mathématicien Richard Price [19]. Le théorème est donné par la formule suivante :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.2)$$

- A et B sont des événements.
- $P(A)$ est la probabilité d'observer l'événement A
- $P(B)$ est la probabilité d'observer l'événement B .
- $P(A|B)$ est la probabilité conditionnelle d'observer A , sachant qu'un autre événement B de probabilité non nulle s'est réalisé.

Dans un problème de classification, notre tâche est de trouver l'étiquette la plus probable A , étant donné les caractéristiques B , le théorème de Bayes devient :

$$P(y | x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n | y) P(y)}{P(x_1, \dots, x_n)} \quad (2.3)$$

Où n représente le nombre de caractéristiques, y est l'évènement qu'on cherche à classer.

Par conséquent, en tenant compte de l'hypothèse d'indépendance, Bayes prédit la classe qui constitue la probabilité la plus élevée [19] :

$$\hat{y} = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(x_i | y) \quad (2.4)$$

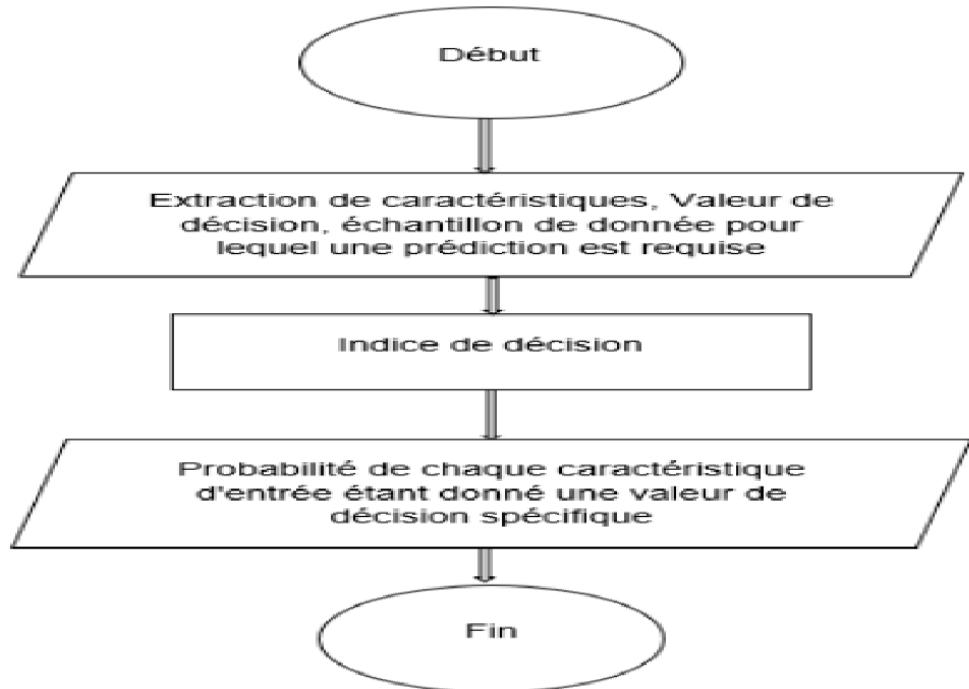


Figure 2-3: Organigramme de l'algorithme Naïve Bayes.[19]

2.1.2.2. L'Arbre de Décision

Un arbre de décision DT (Decision Tree) est un enchaînement hiérarchique de règles logiques ou de production construite de manière automatique à partir d'un ensemble de données. La construction de l'arbre de décision consiste à utiliser les descripteurs, pour subdiviser progressivement l'ensemble en sous-ensembles de plus en plus fins.

L'algorithme estime la probabilité qu'une observation se trouve dans le nœud t en utilisant la formule :

$$P(T) = \sum_{j \in T} w_j \quad (2.5)$$

Où :

w_j : Poids des observations,

T : Ensemble de tous les indices d'observation du nœud

n : Égale au nombre d'observations.

Les arbres de décision sont des algorithmes de classification de données reposant sur un apprentissage supervisé. Comme son nom l'indique, la structure de ces algorithmes ressemble à des arbres constitués de nœuds, de branches et de feuilles (voir figure 2.4). La construction de ces arbres est réalisée à l'aide d'une base de données brute (vecteur des caractéristiques et classes) et de lois qui permettent de déterminer les variables discriminantes pour la classification efficace des futures données. Chacun des nœuds constituant l'arbre représente une règle de classification préalablement déterminée de manière récursive. [20]

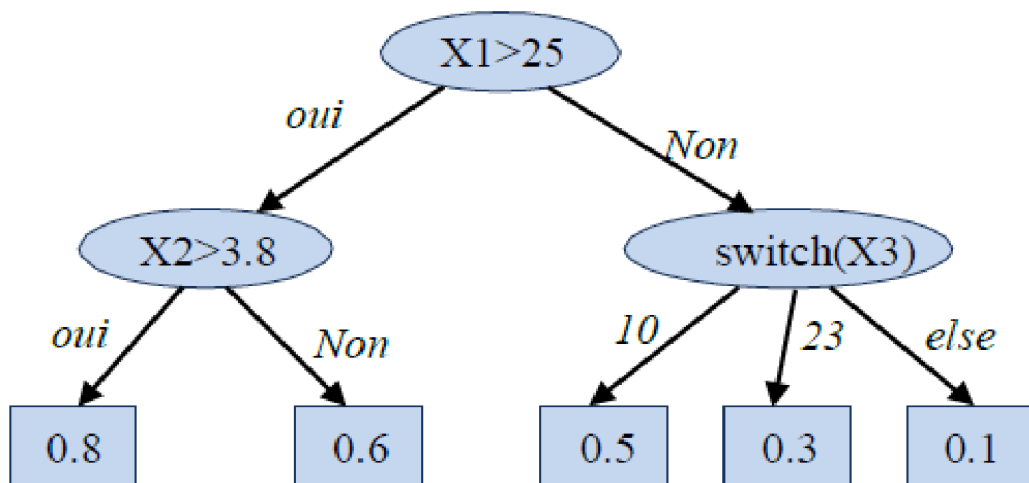


Figure 2-4: Exemple de l'arbre de décision

2.1.2.3. Machines à Vecteurs de Support (SVM)

Les machines à vecteurs supports (Support Vector Machines, SVM) sont des méthodes généralistes d'apprentissage et de discrimination. Ces méthodes ont montré leur efficacité dans de nombreuses applications. Les SVM ont été à l'origine des travaux de Vapnik [21] et ont été conçus pour la décision binaire. L'originalité principale de ces méthodes consiste à utiliser efficacement les exemples étiquetés afin de produire une fonction de décision qui maximise la marge entre deux classes données. L'efficacité repose sur la sélection des éléments d'apprentissage les plus représentatifs de la tâche de décision tout en maximisant la capacité de généralisation du modèle. Ces éléments correspondent aux exemples proches de la frontière de décision. Ils sont appelés les supports vecteurs et sont situés sur la marge. Pour garantir la bonne généralisation de la fonction de décision recherchée, le problème mathématique correspond à la maximisation de la distance des exemples annotés de la frontière (voir figure 2-5).

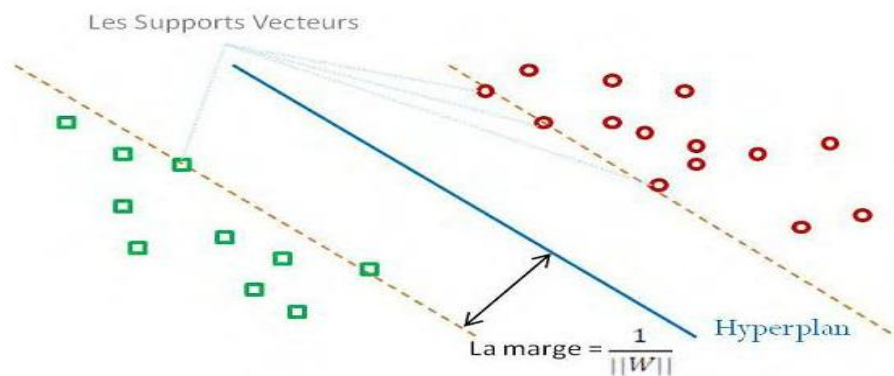


Figure 2-5: Méthode à noyau SVM. [22]

Cette méthode a montré son efficacité dans de nombreux domaines d'applications tels que le traitement d'image, la catégorisation de textes ou le diagnostics médicales et ce même sur des ensembles de données de très grandes dimensions. Cependant, malgré qu'un algorithme de SVM soit en général peu coûteux en temps, il faut compter que la recherche des paramètres permettant d'obtenir la performance maximale, peut requérir des phases de test assez longues [23].

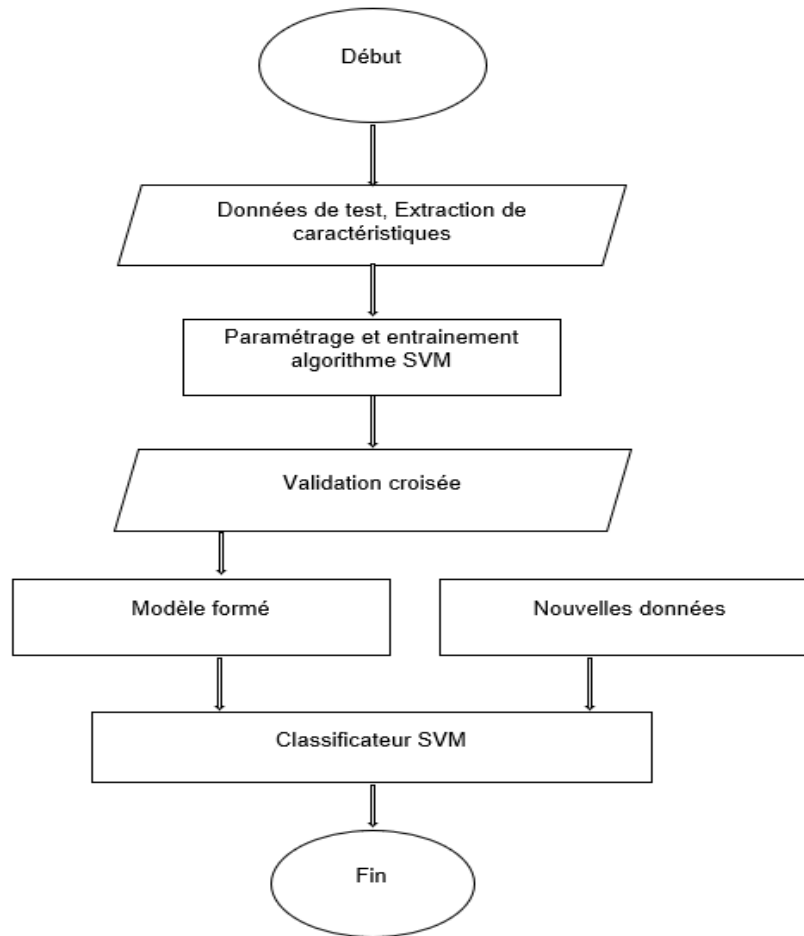


Figure 2-6: Organigramme de l'algorithme SVM.

2.2. Apprentissage Non Supervisé (Unsupervised Learning)

L'apprentissage non supervisé est le mieux adapté lorsque le problème nécessite une quantité massive de données non étiquetées. Par exemple, les applications de médias sociaux, telles que Twitter, Instagram, Snapchat, etc., contiennent toutes de grandes quantités de données non étiquetées. Comprendre la signification de ces données nécessite des algorithmes qui peuvent commencer à comprendre la signification en étant capables de classer les données en fonction des modèles ou des clusters qu'ils trouvent. Par exemple, dans le domaine de la santé, la collecte d'énormes quantités de données sur une maladie spécifique peut aider les praticiens à mieux comprendre les schémas des symptômes et à les relier aux résultats des patients. Il faudrait trop de temps pour étiqueter toutes les sources de données associées à une maladie comme le diabète. Par conséquent, une approche d'apprentissage non supervisé peut aider à déterminer les résultats plus rapidement qu'une approche d'apprentissage supervisé. [31]

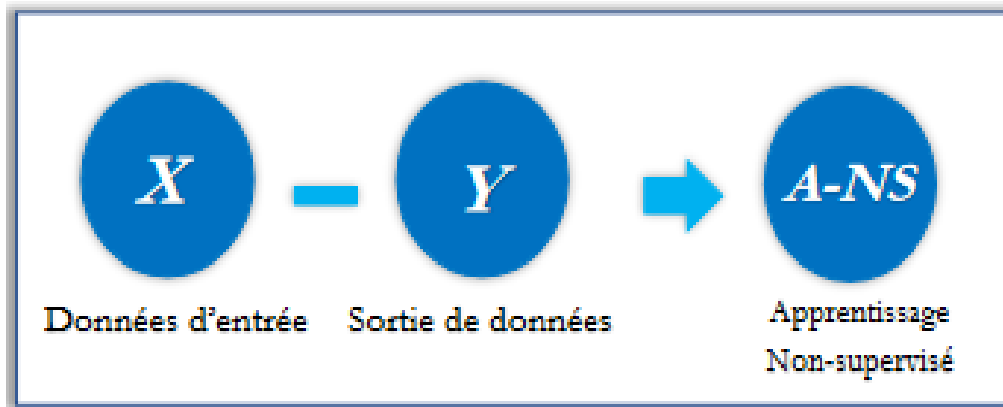


Figure 2-7: Apprentissage non supervisé [14]

- **Regroupement (Clustering)** : c'est une méthode d'analyse statistique utilisée pour organiser des données brutes en silos homogènes, à l'intérieur de chaque grappe, les données sont regroupées selon une caractéristique commune.
- **Réduction de la dimension** : l'objectif est de simplifier les données sans perdre trop d'informations, à titre d'exemple, fusionner plusieurs caractéristiques en un seul caractère.
- **Règle d'association** : Une règle d'association est une relation d'occurrence qui lie deux ou plusieurs items particuliers. L'objectif des règles d'associations, traditionnelles ou séquentielles, consiste à découvrir les associations fréquentes dans les données étudiées.

Les algorithmes les plus célèbres utilisés dans cette approche sont :

- **K-Moyenne** : est un algorithme de Regroupement (Clustering) il regroupe dans les même Cluster (Groupes) les données similaires (qui se ressemblent). Il utilise un raffinement itératif pour produire un résultat final.
- **PCA (Analyse des composants principaux)** : l'algorithme PCA consiste à transformer des variables liées entre elles, vers de nouvelles variables séparées les uns des autres. Ces nouvelles variables sont nommées les composantes principales, elles permettent au praticien de réduire le nombre de variables et de rendre l'information moins redondante.
- **Apriori** : l'algorithme Apriori s'utilise dans une base de données transactionnelle pour extraire des ensembles d'éléments fréquents, puis générer des règles d'association.

2.3. Apprentissage par Renforcement (Reinforcement Learning)

L'apprentissage par renforcement est une approche de l'intelligence artificielle qui met l'accent sur l'apprentissage du système à travers ses interactions avec l'environnement. Avec l'apprentissage par renforcement, le système adapte ses paramètres en fonction des réactions reçues de l'environnement, qui fournit ensuite un retour d'information sur les décisions prises. C'est un système qui apprend avec le renforcement. La recherche actuelle sur l'apprentissage avec renforcement est hautement interdisciplinaire et comprend des chercheurs spécialisés dans les algorithmes génétiques, les réseaux de neurones, la psychologie et les techniques de contrôle. [24]

Il existe plusieurs algorithmes dans l'apprentissage par renforcement, nous citerons deux algorithmes :

- **Q-learning** : Le Q-learning est une technique d'apprentissage par renforcement. Cette technique ne nécessite aucun modèle initial de l'environnement. La lettre 'Q' désigne la fonction qui mesure la qualité d'une action exécutée dans un état donné du système.
- **Algorithme génétique (Genetic algorithm)** : Les algorithmes génétiques appartiennent à la famille des algorithmes évolutionnistes. Leur but est d'obtenir une solution approchée à un problème d'optimisation. Les algorithmes génétiques utilisent la notion de sélection naturelle et l'appliquent à une population de solutions potentielles au problème donné.

2.4. Apprentissage en Profondeur (Deep Learning)

L'apprentissage en profondeur (Deep Learning) est un domaine de recherche sur le ML basé sur un type particulier de mécanisme d'apprentissage.

Il est caractérisé par l'effort de créer un modèle d'apprentissage à plusieurs niveaux, dans lequel les niveaux les plus profonds prennent en compte les résultats des niveaux précédents, les transformant et en faisant toujours plus d'abstraction. Cet aperçu des niveaux d'apprentissage est inspiré par la façon dont le cerveau traite l'information et apprend en réagissant aux stimuli externes. Chaque niveau d'apprentissage correspond, par hypothèse, à l'une des différentes zones qui composent le cortex cérébral. [24]

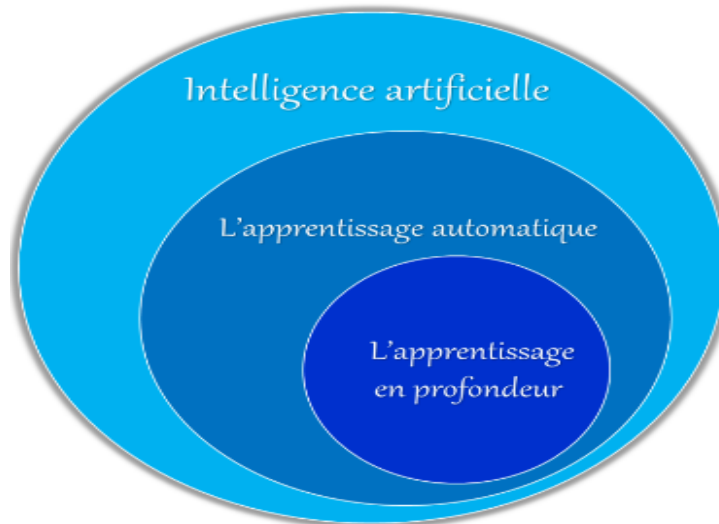


Figure 2-8: Les sous-branches de l'intelligence artificiel [13]

- **RNN :** Les réseaux de neurones récurrents sont adaptés pour des données d'entrée de taille variable. Ils conviennent en particulier pour l'analyse de séries temporelles. Ils sont utilisés en reconnaissance automatique de la parole ou de l'écriture manuscrite.
- **CNN :** Les réseaux de neurones convolutifs désignent une sous-catégorie de réseaux de neurones. Cependant, les CNN sont spécialement conçus pour traiter des images en entrée.

2.5. Apprentissage Ensembliste (Ensemble Learning)

Utiliser un modèle pour prédire un comportement, c'est un premier pas. Mais agréger des milliers de modèles ayant des avis divergents mais pouvant être chacun spécialisés sur des parties de la data donne le plus souvent de meilleurs résultats. Nous parlons alors de méthodes ensemblistes, dont les plus connues sont le Boosting et le Bagging.

- **Boosting :** Il s'agit ici de construire un système où les classifieurs sont utilisés les uns à la suite des autres, chacun réduisant l'ensemble de classes possibles pour un objet à classer.
- **Bagging :** L'organisation parallèle des experts d'un ensemble de classifieurs implique que chaque classifieur individuel produise une sortie simultanée. L'ensemble de ces sorties est ensuite fusionné de façon à produire une décision unique.

Conclusion

Dans ce chapitre, nous avons présenté le ML et certains de ses algorithmes. Nous avons d'abord commencé par décrire le machine Learning et ses différentes architectures proposées. Ensuite, nous sommes passés à une description détaillée des algorithmes d'apprentissage supervisé. Enfin, une description de certains travaux de prévision des incendies de forêt utilisant le ML est discutée.

Le prochain chapitre sera consacré à la présentation détaillée de notre approche proposée, qui a un rôle dans la prédiction des incendies de forêt, basée sur l'apprentissage automatique supervisé.

CHAPITRE 3

L'APPROCHE PROPOSEE

Introduction

Les incendies de forêt sont le type de catastrophe le plus grave et un problème qui menace toutes les forêts du monde. Au cours de ces dernières années, divers systèmes basés sur le ML ont été mis au point pour améliorer la classification et la prédiction du feu de forêt, et ils dépendent sur des données étiquetées.

Dans ce chapitre, nous allons présenter notre approche, qui consiste à prédire les incendies de forêt en utilisant l'apprentissage automatique supervisé avec la classification et la régression.

Dans ce qui suit, nous allons tout d'abord présenter les travaux antérieurs relatifs au sujet étudié, ensuite nous présenterons notre modèle proposé.

1. L'état de l'art

Les incendies de forêt causent d'énormes dégâts environnementaux et des souffrances humaines, ce qui a contribué à attirer l'attention du monde sur ce problème croissant.

Au cours des dernières décennies, les scientifiques ont fait un effort mondial majeur pour construire des systèmes d'alerte précoce basés sur des données météorologiques qui réduisent les risques forestiers.

Amparo et al. Ont utilisé un modèle de prévision des incendies de forêt dans le nord de l'Espagne basé sur un réseau de neurones dont la production est classée en quatre catégories de risques symboliques, obtenant une précision de 80 % [25].

Imas et al. Ont développé des modèles de classification pour l'occurrence des points chauds en utilisant l'arbre de décision dans la Province de Riau (Indonésie), le résultat a montré que l'arbre de décision avait la meilleure précision (69,59 %) par rapport aux autres techniques [26].

Illiadis et al, ont adopté la logique floue pour prédire les incendies de forêt en Grèce, la précision retrouvée était de 85,25 % [27].

Zahou et al. Ont utilisé la technique LDA pour prédire les incendies de forêt dans le Shanxi (Chine). Les résultats des tests ont montré que la méthode était généralement pratique avec une précision moyenne de 82 % dans la région d'essai [28]. (Voir Tableau 3.1)

Auteurs	Titre de l'article	Lieu d'étude	Approche utilisée	Précision
Amparo et al.	Un système intelligent de prévision et de lutte contre les incendies de forêt gestion en Galice	Nord de l'Espagne	Réseau de neurones	80 %
Imas et al.	Application of classification algorithmes d'exploration de données pour la prédiction de l'occurrence des points chauds dans la province de Riau en Indonésie	Province de Riau (Indonésie)	Arbre de décision	69,59 %
Illiadis et al.	Un système expert heuristique de guidage des feux de forêt en Grèce	forêts de Grèce	logique floue	85,25 %
Zahou et al.	Fisher Linear Discriminant Method for Forest Fire Risk Points on Transmission Line	Shanxi (Chine)	technique LDA	82 %

Table 3-1: L'état de l'art

2. Conservation des forêts de M'sila

- Décret de création : la conservation des forêts de la wilaya de M'sila a été instituée par la Résolution n°95/333 du 25 octobre 1995. Le domaine forestier domanial occupe une superficie : 388.792 ha
- Nombre de provinces : 10 (M'Sila , Hammam Dalaa, Ouled Derradj, Sidi Aissa, Ben Srour, Bou Saada, Magra, Medjedel, Djebel Messaad, Aïn El Hadjel)
- Siège social : Al-Muwailiha quartier 1000 logements coopératifs

2.1. Définition

La capacité juridique : une institution à caractère administratif Sous la tutelle du Ministère de l'Agriculture et du Développement Rural (Direction Générale des Forêts). Au niveau de l'Etat, le domaine d'Al-Ghabi s'étend sur une superficie estimée à : 388.792 hectares, soit 21.5 % de la superficie totale de M'sila.

Le siège de la conservation comprend quatre (04) départements et neuf (09) bureaux qui sont :

- Administration et moyens :
 - Bureau de gestion des ressources humaines.
 - Bureau de gestion budgétaire et moyens.
- Direction de la Gestion de Patrimoine, des Etudes et de la Programmation :
 - Le Bureau de Régulation et de Police Forestière.
 - Bureau des Etudes et des Programmes.
 - Bureau de l'inventaire, de la préparation et des produits forestiers.
- Service de la protection des végétaux et des animaux :
 - Bureau de la prévention et du contrôle des incendies et des maladies.
 - Office des Espèces Protégées, des Activités de Chasse et de Pêche.
- Le Département de l'Expansion de la Richesse et de la Protection des Terres :
 - Bureau de bonification des terres.
 - Bureau d'expansion des ressources forestières.

3. Lieu d'étude

M'sila est l'une des wilayas de l'intérieur composé de 15 arrondissements et 47 communes. Il est délimité au nord par les états de Sétif et Bordj Bou Arreridj, à l'ouest par les états de Bouira et Médéa, au sud par les états de Djelfa et Biskra, et à l'est par l'état de Batna, son climat est continental, et c'est un milieu entre la colline et le désert, et la majeure partie de l'état est plate, avec une hauteur de 200 à 300 mètres d'altitude.

La Wilaya de M'sila est dominée par un climat continental, qui se caractérise par des étés chauds et secs et des pluies modérément froides en hiver.

4. Méthodologie et préparation des données

Nous avons obtenu les données relatives aux différents incendies de M'sila à partir des extraits de rapports établis par la Conservation des forêts de la wilaya. Ces données s'étendaient sur une période de 2011 jusqu'à 2021 comprenant 69 incendies de forêt.

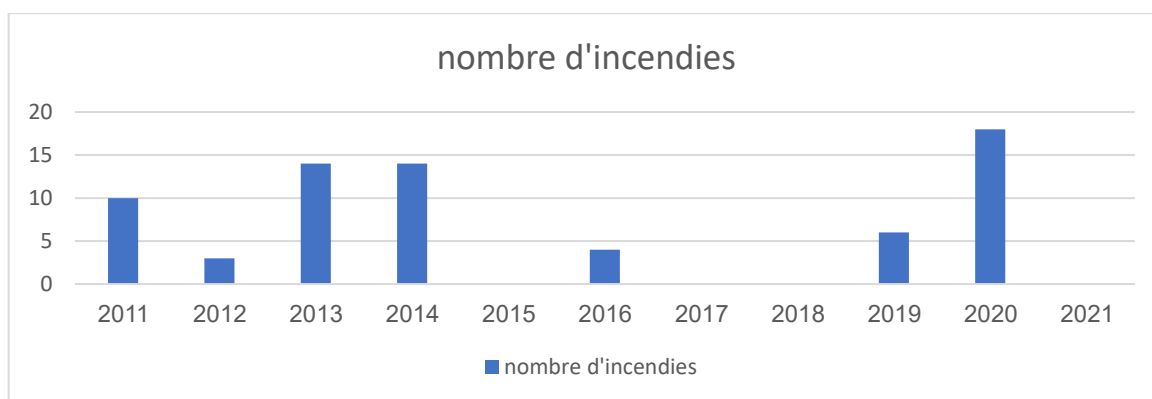


Figure 3-1: nombre d'incendies dans M'sila (2011-2021)

Nous avons enregistré les données dans des tableaux et intégrées dans un fichier Excel. Ces données peuvent inclure les variables suivantes : Daïra, Commune, Nom de forêt, Date, Heure, Surface, Type des plantes brûlées et Température.

En utilisant l'historique météo "Historique météo : archives météo pour le monde entier" [32], nous avons ajouté trois autres attributs qui sont : la vitesse du vent, l'humidité et les précipitations. Voir le tableau (3-2)

Variable	Explication
Daïra	Daïra où le feu s'est déclaré
Commune	La commune où le feu s'est déclaré
Nom de forêt	Le nom de la forêt où le feu s'est déclaré
Date	La date de l'incendie (JJ/MM/AAAA)
Heure	Le heure de l'incendie (HH:MM)
Surface	La zone brûlée (en Hectares)
Type	Type de plantes brûlées (maquis, broussailles, forêt, halfa, autres)
Température	La température de l'air dans la zone d'incendie mesurée par (c°)
Vents	Vitesse du vent dans la zone d'incendie exprimée par (Km/h)
Précipitations	La quantité de pluie dans la zone d'incendie mesurée par (mm)
Humidité	Le pourcentage d'humidité dans la zone d'incendie.

Table 3-2: Attributs de l'ensemble de données

	A	B	C	D	E	F	G	H	I	J	K
1	daira	commune	nome foret	date	heure	surface(H)	type	températ ure(c°)	vents(km /h)	précipitat ions(mm)	humidité (%)
2	hammam dala	hammam dala	dreat	09/06/11	14:53	7,00	foret	30	13	0	73
3	hammam dala	hammam dala	dreat	11/06/11	21:45	3,00	autres	25	10	0	77
4	hammam dala	hammam dala	dreat	06/07/11	21:45	8,00	alhalfa	35	9	0	45
5	magra	magra	righa dahra	07/07/11	14:30	2,00	maquis	37	21	0	29
6	hammam dala	hammam dala	dreat	18/07/11	17:00	350,00	maquis	38	23	0	34
7	hammam dala	hammam dala	dreat	18/07/11	17:00	125,00	roussaille	38	23	0	34
8	hammam dala	hammam dala	dreat	18/07/11	17:00	100,00	alhalfa	38	23	0	34
9	hammam dala	hammam dala	dreat	21/07/11	17:00	50,00	foret	39	26	0	42
10	hammam dala	hammam dala	dreat	21/07/11	17:00	20,00	maquis	39	26	0	42

Figure 3-2: Extrait de la Base de données en Excel

5. Les outils et détails d'implémentation

Pour évaluer et tester les performances de l'approche proposée, nous devons d'abord passer par l'étape de l'implémentation. Dans cette section, nous décrivons les différents outils et langages de programmation utilisés.

5.1. Langage de programmation

De nos jours, il existe plusieurs langages de programmation et chaque langage possède ses propres caractéristiques. Parmi ses langages, notre choix s'est focalisé sur Python.

5.1.1. Python

C'est l'un des langages de programmation les plus intéressants du moment, et il l'a inventé "Guido van Rossum", la première version de Python sortie en 1991. C'est un langage Programmation interprétée, multi-paradigme, multiplateforme. Encourage la programmation Impératif structuré, fonctionnel et orienté objet. Python est simple et puissant à la fois, il permet d'écrire des scripts très simples, grâce à Ses nombreuses bibliothèques nous permettent de travailler sur des projets plus ambitieux [29].

5.1.2. Pourquoi Python ?

Python est un langage stable et flexible qui fournit aux développeurs une variété d'outils. Quelle Faites-en un choix de premier ordre pour le ML, du développement, la mise en œuvre et l'entretien. Python aide les développeurs à développer des produits Confiance [30].

Python présente de nombreux avantages :

- Simple et cohérent : la simplicité de Python facilite la gestion des développeurs Algorithmes complexes pour le ML.
- Flexibilité : Le facteur de flexibilité réduit la possibilité d'erreur, il laisse Le programmeur a le contrôle total et s'y engage Confortable.
- Librairies et Framework : Algorithmes de Machine Learning Très complexe, mais Python est une solution de repli avec une large gamme Bibliothèques telles que Scikit-learn, Keras et Framework TensorFlow.
- Lisibilité : Python est facile à lire, les développeurs peuvent donc Il est facile de comprendre le code.
- Indépendance de la plate-forme : Python est un langage indépendant de la plate-forme. Plate-forme. Il est pris en charge par de nombreuses plates-formes, notamment Windows, Linux et MacOS.

5.2. Environnement de développement

5.2.1. Jupyter

Jupyter est une application web utilisée pour programmer dans plus de 40 langages de programmation, dont Python, Julia, Ruby, R, ou encore Scala. C'est un projet communautaire dont l'objectif est de développer des logiciels libres, des formats ouverts et des services pour l'informatique interactive.



Figure 3-3: Logo de l'IDE Jupyter

5.2.2. Sklearn

Scikit-learn (sklearn) est une bibliothèque Python gratuite pour le ML. Il est développé par de nombreux contributeurs², notamment universitaires, par des instituts d'enseignement supérieur et de recherche français comme Inria³. Il fournit de nombreuses librairies d'algorithmes dans son Framework à implémenter, clé en main.

Ces bibliothèques sont particulièrement utiles pour les data scientists. Il comprend spécifiquement des fonctions d'estimation de forêts aléatoires, de régression logistique, d'algorithmes de classification et de machines à vecteurs de support. Il est conçu pour se coordonner avec d'autres bibliothèques Python gratuites, notamment NumPy et SciPy.



Figure 3-4: Logo de bibliothèque Sklearn

5.2.3. Pandas

Pandas est une bibliothèque open source conçue principalement pour travailler avec des données relationnelles ou étiquetées de manière simple et intuitive. Il fournit diverses structures de données et opérations pour manipuler des données numériques et des séries chronologiques. Cette bibliothèque est construite au-dessus de la bibliothèque NumPy. Pandas est rapide et offre des performances et une productivité élevée pour les utilisateurs.

5.2.4. Matplotlib

Matplotlib est une bibliothèque multiplateforme de visualisation de données et de traçage graphique pour Python et son extension numérique NumPy. En tant que tel, il offre une alternative open source viable à MATLAB.

5.2.5. NumPy

NumPy est une bibliothèque Python utilisée pour travailler avec des tableaux. Elle possède également des fonctions pour travailler dans le domaine de l'algèbre linéaire, de la transformée de Fourier et des matrices.

5.2.6. Matrice de Confusion

Une matrice de confusion ou tableau de contingence sert à évaluer la qualité d'une classification. Elle est obtenue en comparant les données classées avec des données de référence qui doivent être différentes de celles ayant servi à réaliser la classification. Elle ne doit pas être confondue avec la table de performance qui sert à évaluer l'homogénéité thématique des zones-test servant à réaliser une classification assistée.

		Valeur de prédiction	
		Positive	Négative
Valeurs réelles	Positive	VP	FN
	Négative	FP	VN

Table 3-3: Exemple de Matrice de Confusion.

Les paramètres d'évaluation sont définis comme suit :

- Vrai positive VP : classe positive classée comme étant positive.
- Vrai négative VN : classe négative classée comme étant négative.
- Faux positive FP : classe négative considérée positive.
- Faux n'égative FN : classe positive considérée négative.

Nous avons utilisé les mesures de performances suivantes pour évaluer et montrer l'efficacité de notre modèle :

La justesse (Accuracy) : La justesse ou (Accuracy en Anglais) est l'un des critères permettant d'évaluer les modèles de classification. Elle désigne la proportion des prédictions correctes effectuées par le modèle, l'équation représente la formule de calcul de cette mesure:

$$ACC = (VP + VN) / (VP + VN + FP + FN)$$

La Précision : Permet de déterminer la proportion d'identifications positives qui ont été effectivement correctes :

$$Précision = VP / (VP + FP)$$

La Sensibilité : Elle mesure le pourcentage de valeurs positives correctement identifiés par le modèle :

$$Sensibilité = VP / (VP + FN)$$

La Spécificité : Elle mesure le pourcentage de valeurs négatives correctement identifiées par le modèle :

$$\text{Spécificité} = \text{VN} / (\text{VN} + \text{FN})$$

F1-mesure : Il mesure la similarité de deux échantillons, il calculé à partir de la Précision et la Sensibilité du modèle, l'équation :

$$\text{F1-mesure} = (\text{Précision} \times \text{Sensibilité}) / (\text{Précision} + \text{Sensibilité})$$

ROC : Une courbe ROC (receiver operating characteristic) est un graphique représentant les performances d'un modèle de classification pour tous les seuils de classification. Cette courbe trace le taux de vrais positifs (TVP **Sensibilité**) en fonction du taux de faux positifs (TFP **Spécificité**).

AUC : signifie "aire sous la courbe ROC". Cette valeur mesure l'intégralité de l'aire à deux dimensions situées sous l'ensemble de la courbe ROC (par calculs d'intégrales) de (0,0) à (1,1).

Pour dessiner le courbe ROC, nous utilisons les valeurs TVP (taux de vrais positifs) sur l'axe Y, et TFP (taux de faux positifs) sur l'axe X. Pour cela nous avons utilisé les équations suivantes :

$$\text{TVP} = \text{Sensibilité} = \text{Vrai_Positif} / (\text{Vrai_Positif} + \text{Faux_Négatif})$$

$$\text{TFP} = 1 - \text{Spécificité} = 1 - [\text{Vrai_Négatif} / (\text{Vrai_Négatif} + \text{Faux_Positif})]$$

6. L'apprentissage du modèle proposé (Classification)

6.1. Présentation des ensembles d'apprentissage utilisés

Nous avons défini les variables indépendantes "température, vitesse du vent, humidité et précipitations" (voir figure 3-5) et les variables dépendantes "type" et "classe", où nous avons préparé 3 ensembles d'apprentissage de données pour construire un modèle de prédiction selon les objectifs de prédiction tracés.

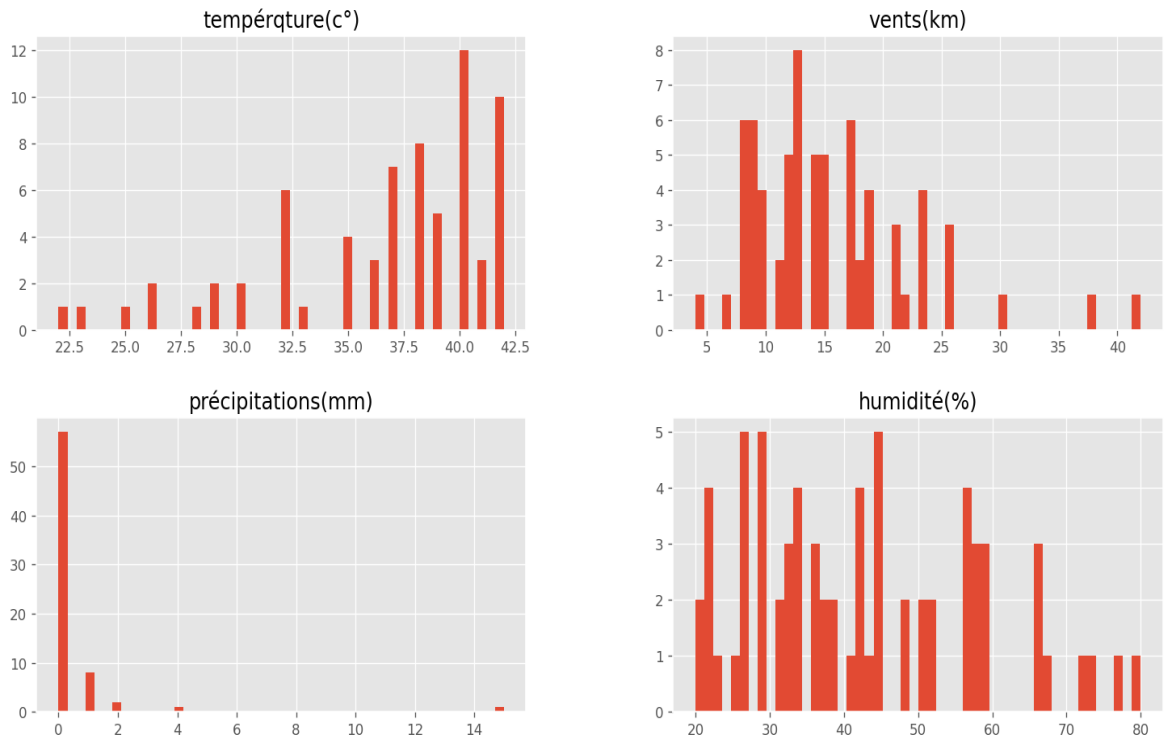


Figure 3-5 Les graphiques des variable indépendantes.

1er Ensemble d'apprentissage (BDD1) : Cet ensemble est composé de 04 attributs (Température, Vent, Précipitations et Humidité) ainsi que la colonne de classe qui indique si l'incendie peut se produire ou non. Cet ensemble est utilisé dans l'objectif de prévoir si un incendie peut se produire ou non selon certaines conditions météorologiques (voir Fig 3.6).

	température(c°)	vents(km)	précipitations(mm)	humidité(%)	class
0	30	13	0	73	oui
1	30	12	0	70	non
2	25	10	1	77	oui
3	28	12	0	55	non
4	30	10	0	56	oui

Figure 3-6: Extrait de l'ensemble d'apprentissage pour la prédiction d'incendie (BDD1).

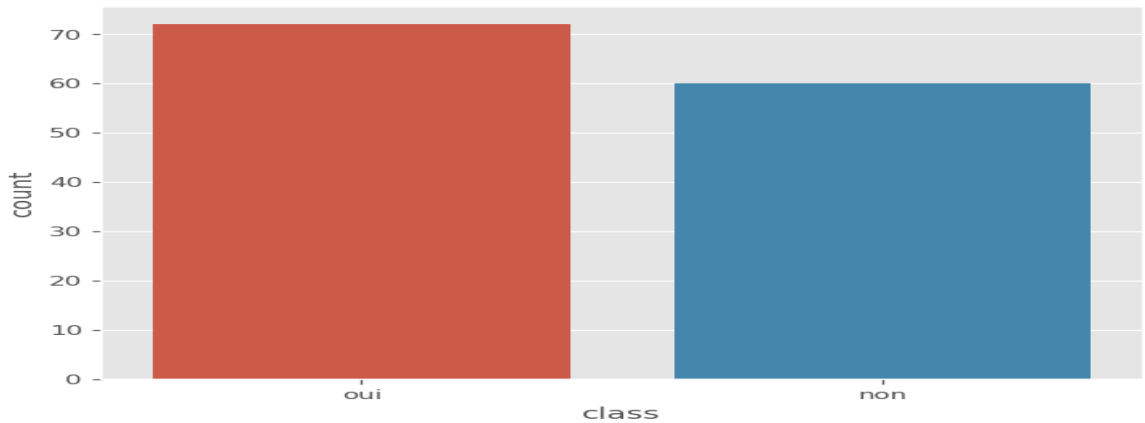


Figure 3-7: Graphique de la variable dépendante Incendie ou non

2^{ème} Ensemble d'apprentissage (BDD2) : Cet ensemble est composé de 04 attributs (Température, Vent, Précipitations et Humidité) ainsi que la colonne de classe qui indique le type de plantations ravagées par un incendie une fois il est produit. Cet ensemble est utilisé pour prédire si l'incendie est produit dans un maquis ou une plantation autre que le maquis. (Voir Fig. 3.8).

	tempérqture(c°)	vents(km)	précipitations(mm)	humidité(%)	type
0	30	13	0	73	autres
1	25	10	0	77	autres
2	35	9	0	45	autres
3	37	21	0	29	maquis
4	38	23	0	34	maquis

Figure 3-8: Extrait de l'ensemble d'apprentissage pour la prédiction d'incendie des maquis (BDD2)

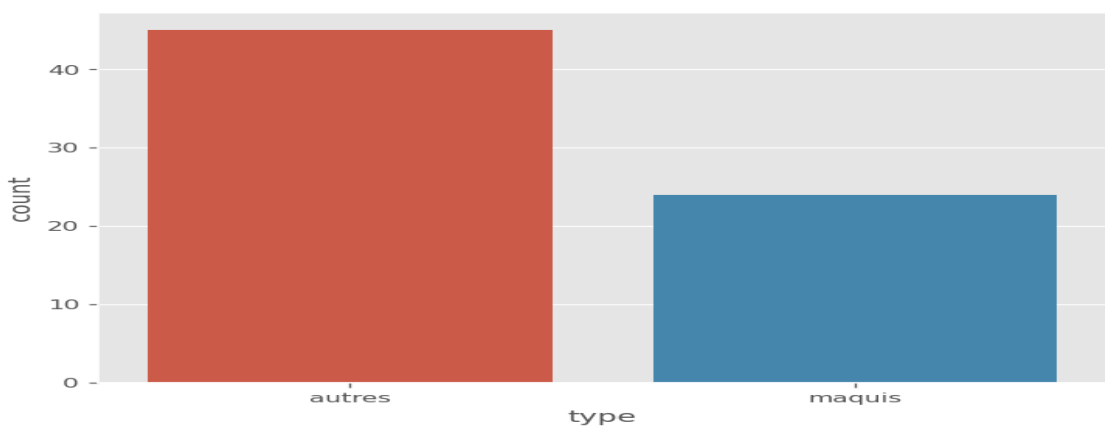


Figure 3-9: Graphique de la variable dépendante Maquis ou autres.

3éme Ensemble d'apprentissage (BDD3) : Cet ensemble est composé de 04 attributs (Température, Vent, Précipitations et Humidité) ainsi que la colonne de classe qui indique le type de plantations ravagées par un incendie une fois il est produit. Cet ensemble est utilisé pour prédire si l'incendie est produit dans une forêt ou une plantation autre que la forêt. (Voir Fig. 3.10).

	tempérqture(c°)	vents(km)	précipitations(mm)	humidité(%)	type
0	30	13	0	73	foret
1	25	10	0	77	autres
2	35	9	0	45	autres
3	37	21	0	29	autres
4	38	23	0	34	autres

Figure 3-10: Extrait de l'ensemble d'apprentissage pour la prédiction d'incendie des forêts (BDD3)

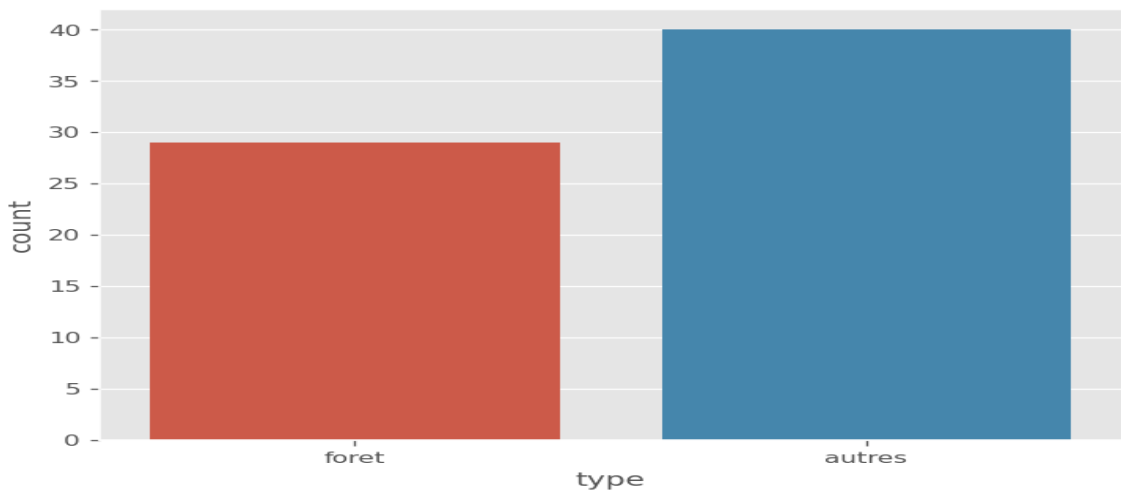


Figure 3-11: Graphique de la variable dépendante Forêt ou autres.

6.2. L'importation et le partitionnement de la base de données

Nous définissons d'abord la base de données "feux de forêt". Nous introduisons ensuite les variables indépendantes et les variables dépendante (Fig. 3.12) :

```
Entrée [4]: data=pd.read_csv('feux de forêt.csv')
X=data.drop(['variables dépendante '], axis=1)
y=data['variables dépendante ']
```

Figure 3-12: L'importation de l'ensemble d'apprentissage et la définition des variables.

Nous importons l'ensemble de données en utilisant le Framework Pandas. La trame de données est chargée dans la variable `data`. Ensuite nous définissons les variables indépendantes et les variables dépendantes. Nous avons partitionné la base de données en deux parties différentes, partie pour l'entraînement « training » 80% et une partie pour les tests 20% pour SVC et Arbre de Décision, et 30% testes, 70% entraînement pour Naïve Bayes par l'utilisation de la fonction « `train_test_split` » (Voir Fig. 3.13). Les modèles utilisés sont discutés dans les sections qui suivent.

```
Entrée [655]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
```

Figure 3-13: Code de Partitionnement de l'ensemble d'apprentissage.

Et pour analyser l'efficacité des modèles on utilise ce Code pour crée le graph de courbe ROC :

```
Entrée [146]: # plot ROC courbe
from sklearn.metrics import roc_curve
# construire la courbe ROC du modèle optimisé
fpr, tpr, thresholds = roc_curve(y_test, y_pred)
plt.figure(figsize=(6,4)) # créer une figure
plt.plot(fpr, tpr, linewidth=2) # afficher la courbe ROC
plt.plot([0,1], [0,1], 'k--' )
plt.rcParams['font.size'] = 12
# donner un titre aux axes et au graphique
plt.title('ROC courbe de Model Classification')
plt.xlabel('Taux de Faux Positifs (1 - Spécifité)')
plt.ylabel('Taux de Vrais Positifs (Sensibilité)')
plt.show() # afficher l'image
```

Figure 3-14: Lignes des codes pour obtenu la courbe de ROC.

6.3. Naïf Bayésien

Nous devons créer une instance de l'algorithme Naïve Bayes. Pour le créer, nous utilisons la classe GaussianNB du package sklearn. naive_bayes pour créer une instance de l'algorithme. Après nous utilisons la méthode d'ajustement pour ajuster l'algorithme sur les données d'apprentissage.

Après nous utilisons la classe confusion_matrix de sklearn. metrics pour calculer la matrice de confusion du modèle.

```
Entrée [10]: from sklearn.naive_bayes import GaussianNB
model_gnb = GaussianNB()
model_gnb.fit(X_train, y_train)
y_pred = model_gnb.predict(X_test)
from sklearn.metrics import confusion_matrix
cm1 = confusion_matrix(y_test, y_pred)
print('Confusion Matrix')
```

Figure 3-15: modèle Naïve Bayes et matrice de confusion.

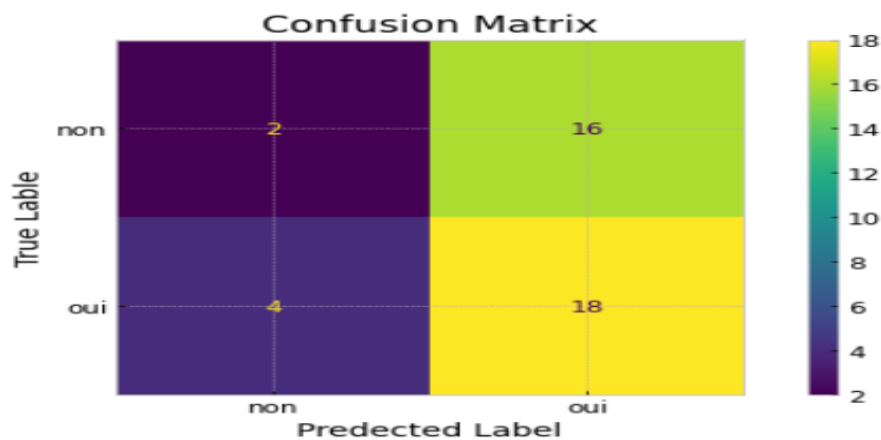


Figure 3-16: Les résultats de matrice de confusion BDD1.

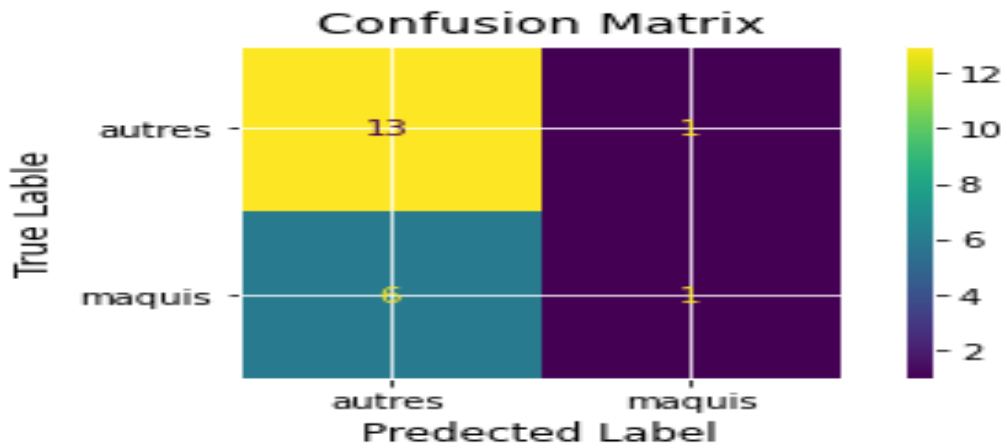


Figure 3-17: Les résultats de matrice pour BDD 2.

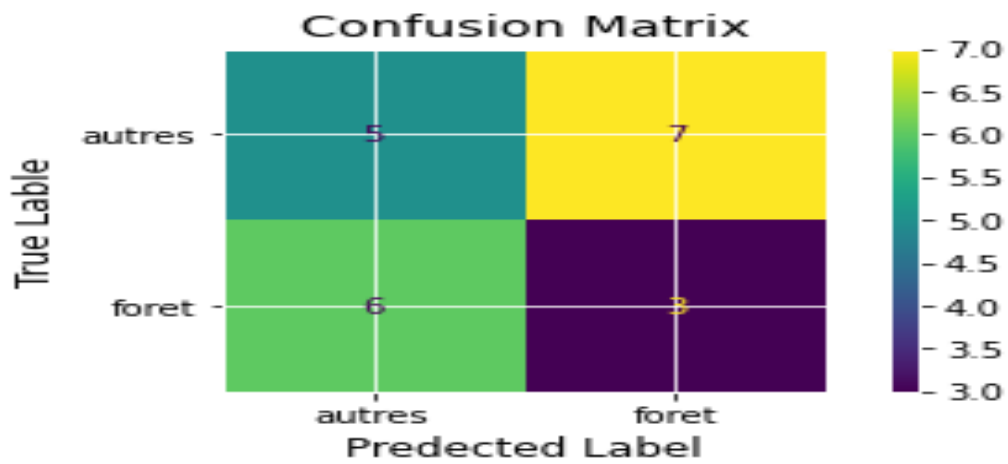


Figure 3-18: Les résultats de matrice pour BDD 3.

Par la matrice de confusion, les résultats indiqués dans le tableau ci-dessous ont été obtenus (voir tableau 3.4)

Attribue	Pour BDD1	Pour BDD 2	Pour BDD3
Précision	10%	92.85%	62.5%
Sensibilité	11.11%	92.85%	41.66%
Spécificité	81.81%	14.28%	33.33%
F1-Measure	10.53%	92.85%	50.0%
AUC	46.46%	7.15%	37.50%
Accuracy	50 %	66.66%	38.09%

Table 3-4: Résultats obtenus par le modèle Naïve bayes.

6.4. Support Vector Machine (SVC) :

Afin de gérer les classifications, Sklearn fournit une classe de classificateur de machines à vecteurs de support appelée SVC.

Nous devons créer un modèle de l'algorithme SVC sur les 3 datasets. Pour le créer, nous utilisons :

- Le noyau Kernel qui définit le type de fonction utilisé pour transformer l'ensemble de données de type 'rbf'.
- Le paramètre C qui définit la régularisation de l'erreur.
- Le coefficient du Kernel Gamma définit le degré d'ajustement du modèle aux données d'entraînement, ce qui vous permet d'éviter le surajustement et le sous-ajustement.
- Degree est le degré de fonction polynomiale.

```
Entrée [102]: from sklearn.svm import SVC #Pour SVM
              model = SVC( C=1.0, kernel='rbf', degree=3, gamma=0.01, )
              model.fit(X_train, y_train) #Exécution SVM
```

Figure 3-19: Le modèle de SVC.

Ici, nous utilisons dans le modèle SVC le module GridSearchCV afin de tester un certain nombre de combinaisons de paramètres qui peuvent optimiser les performances de notre modèle. Pour le réglage des hyperparamètres, nous avons 2 paramètres à prendre en compte (paramètre C, paramètre gamma) :

```
Entrée [83]: # choisir 6 valeurs pour C, entre 1e-2 et 1e3
              C_range = np.logspace(-2, 3, 6)

              # choisir 4 valeurs pour gamma, entre 1e-2 et 10
              gamma_range = np.logspace(-2, 1, 4)

              # grille de paramètres
              param_grid = {'C': C_range, 'gamma': gamma_range}

              # critère de sélection du meilleur modèle
              score = 'roc_auc'
              |
              # initialiser une recherche sur grille
              grid = model_selection.GridSearchCV(svm.SVC(kernel='rbf'),
                                                  param_grid,
                                                  cv=5,
                                                  scoring=score)

              grid.fit(X_train, y_train)
              print("The optimal parameters are {} with a score of {:.2f}"
                    .format(grid.best_params_, grid.best_score_))
```

Figure 3-20: le code de module GridSearchCV.

Grâce à cette fonction nous obtenir les résultats suivants :

- Pour le BDD1 : C=1000, gamma=0.01
- Pour le BDD2 : C=0.01, gamma=0.01
- Pour le BDD3 : C=1000, gamma=0.01

Pour obtenir la matrice de confusion nous avons utilisé la fonction «`confusion_matrix`» de la bibliothèque «`sklearn`» (Fig. 3.21) :

```
Entrée [110]: from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
cm1 = confusion_matrix(y_test, y_pred)
print(cm1)
print(classification_report(y_test, y_pred, target_names=["autres", "maquis"]))
```

Figure 3-21: La matrice de confusion.

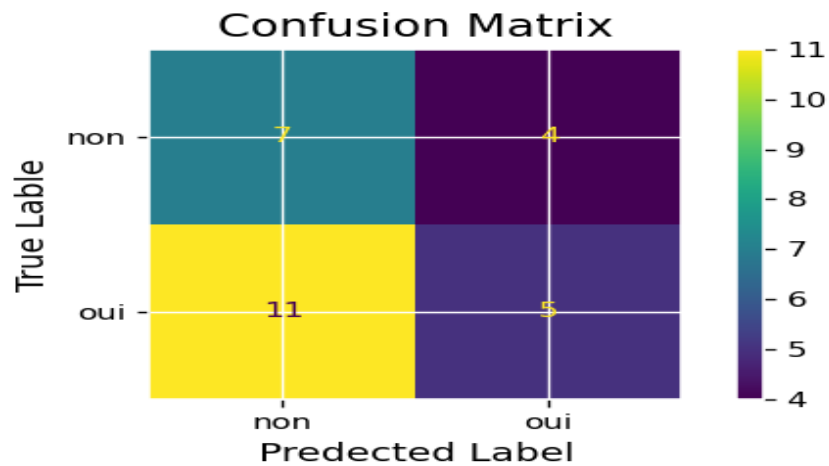


Figure 3-22: Les résultats de matrice de confusion SVC BDD1.

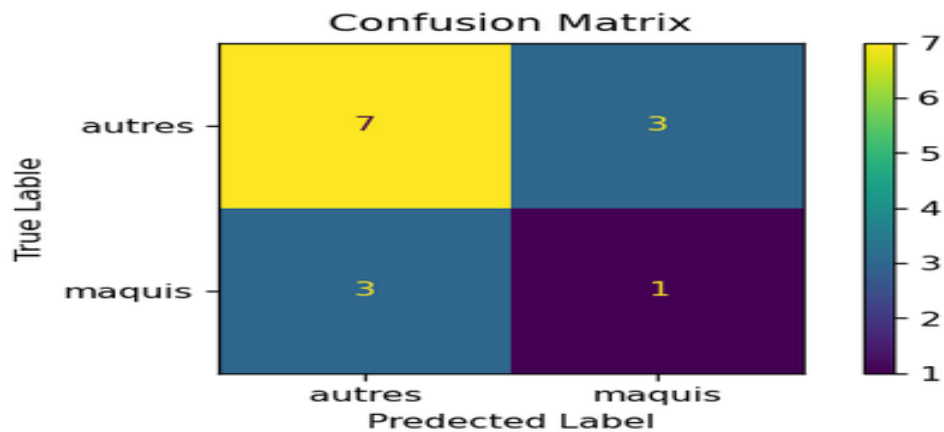


Figure 3-23: Les résultats de matrice pour BDD 2.

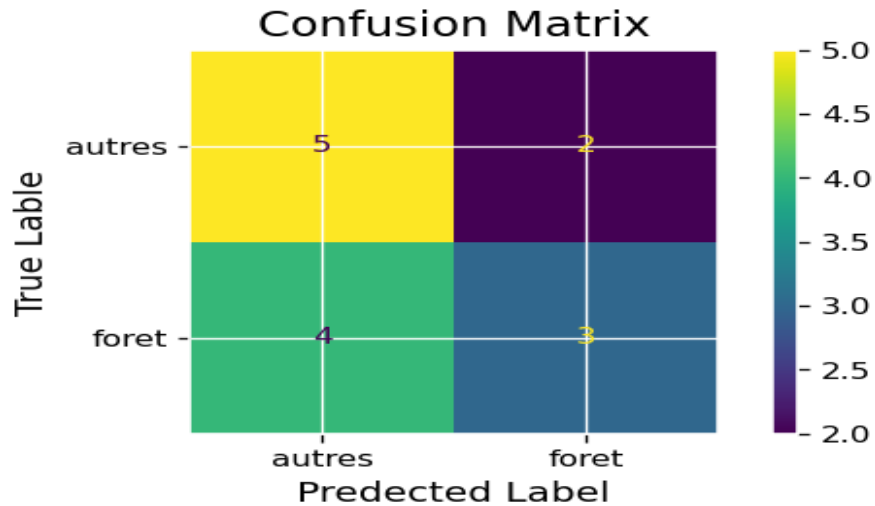


Figure 3-24: Les résultats de matrice pour BDD 3.

Grâce aux résultats obtenus par la matrice de confusion, nous pouvons maintenant calculer la justesse, la précision, la sensibilité, la spécificité et F1-Measure de notre modèle SCV sur les deux Datasets. Les résultats sont décrits dans le tableau ci-dessous (Voir Tableau 3.5)

Attribue	Pour BDD1	Pour BDD2	Pour BDD3
Précision	43.75 %	87.5 %	62.5%
Sensibilité	63.64%	70%	71.43%
Spécificité	56.26%	25%	42.86%
F1-Measure	51.85 %	77.77%	66.67%
AUC	59.34%	47.5%	57.14%
Accuracy	59.26%	57.14%	57.14%

Table 3-5: Les résultats obtenus en SVC.

6.5. Arbre de décision (Decision tree)

Le code utilisé pour créer l'arbre de décision est en Python et utilise la bibliothèque pandas pour les DataFrames et la bibliothèque scikit-learn pour le ML. Cette bibliothèque sklearn contient des méthodes d'arbre de décision pour créer des arbres de décision. La méthode principale utilisée pour la formation est le DecisionTreeClassifier(). Plusieurs hyperparamètres sont disponibles pour affiner chaque modèle. Nous utiliserons la fonction GridSearchCV pour rechercher et fixer la valeur appropriée de chaque hyperparamètre (les

méthodes Gini ou Entropy pour le type d'arbre et restreindre la profondeur de l'arbre résultant en définissant max_depth) (Voir Fig. 3.25).

```

Entrée [13]: Hyper_paramters={'criterion':['gini', 'entropy'],
                             'max_depth':[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]}

Tree_Gridsearch_paramters=GridSearchCV(DecisionTreeClassifier(random_state=4),
                                         Hyper_paramters,scoring='roc_auc',
                                         n_jobs=-1,cv=10,verbose=1)

Tree_crossvalidation=Tree_Gridsearch_paramters.fit(X_train,y_train)

print ("The best paramter combination is ")
print(Tree_crossvalidation.best_params_) #obtient le meilleur estimateur
Final_Model=Tree_crossvalidation.best_estimator_ #obtient le meilleur estimateur
print("The best AUC score was ")
print(Tree_crossvalidation.best_score_) #obtient le meilleur score pour le meilleur estimateur
importances=Tree_crossvalidation.best_estimator_.feature_importances_
importances
    
```

Figure 3-25: Le model de l'Arbre de Décision.

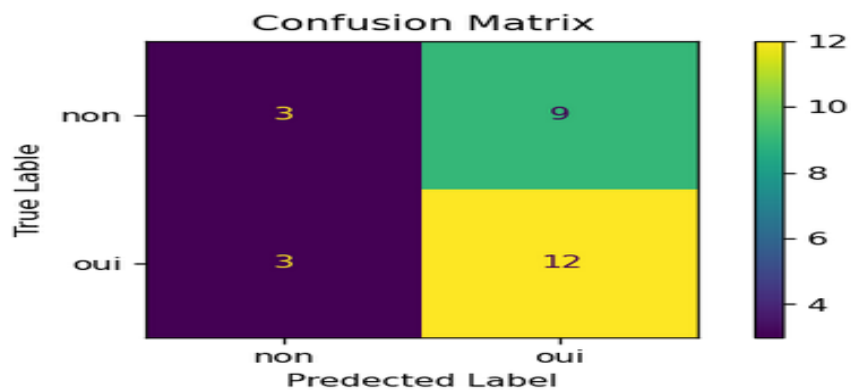


Figure 3-26: Les résultats de matrice de confusion Arbre Décision BDD1.

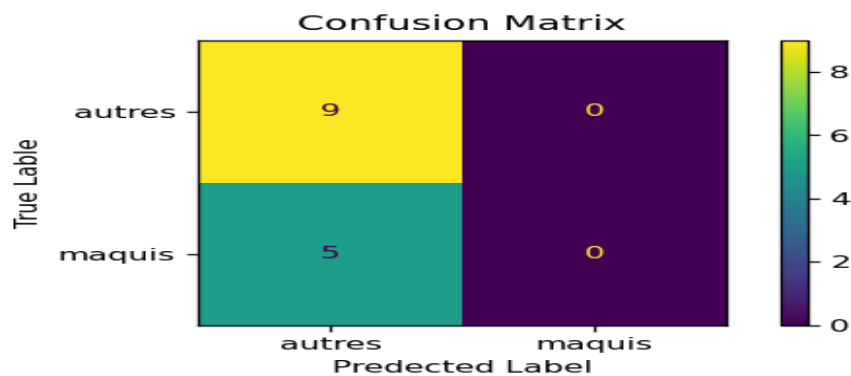


Figure 3-27: Les résultats de matrice de confusion Arbre Décision BDD2.

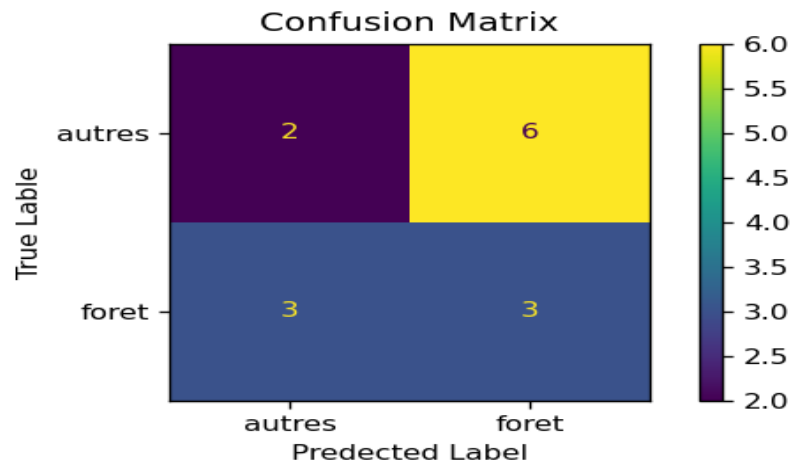


Figure 3-28: Les résultats de matrice de confusion Arbre Décision BDD3.

Pour obtenir une bonne représentation de notre résultat, nous utilisons cette méthode pour dessiner le graphe d'arbre (Voir Fig. 3.29) :

```
Entrée [20]: fig = plt.figure(figsize=(15,10))
             _ = tree.plot_tree(Final_Model, feature_names=X_train.columns.values,
                               class_names='class', filled=True)
```

Figure 3-29: un code pour dessiner l'Arbre de Décision.

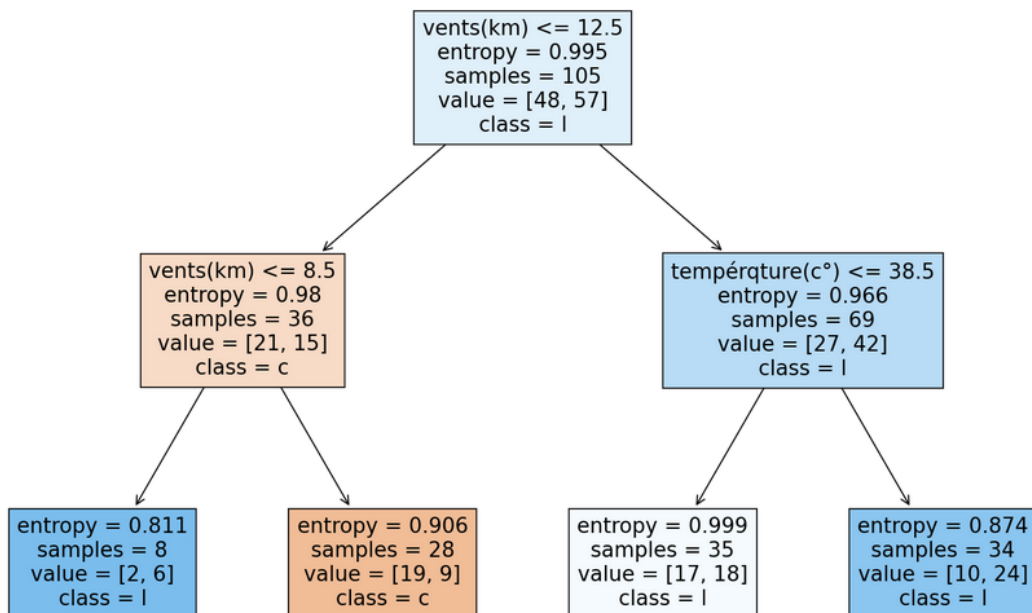


Figure 3-30: le graphe d'Arbre de Décision (BDD1).

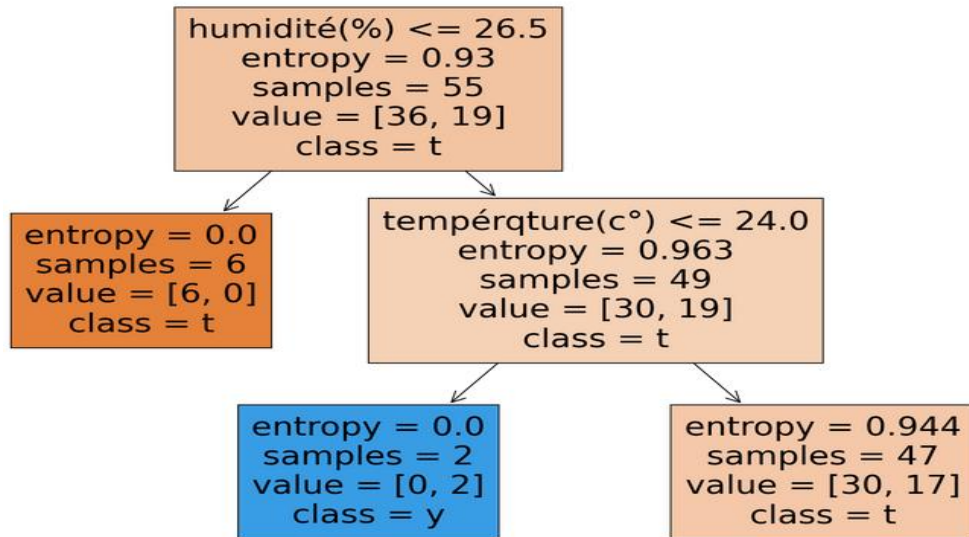


Figure 3-31: Le graphe d'Arbre de Décision (BDD2).

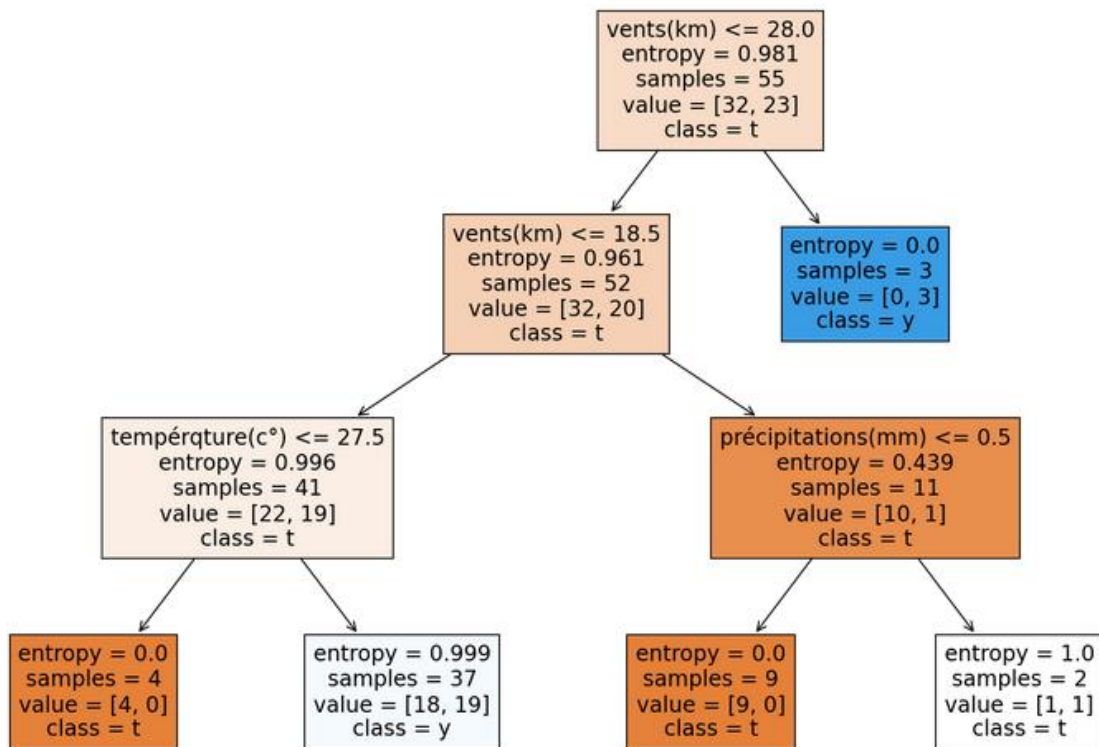


Figure 3-32: Le graphe d'Arbre de Décision (BDD3).

En fonction des résultats obtenus par la matrice de confusion nous pouvons maintenant calculer l'exactitude, la précision, la sensibilité, la spécificité et la mesure F1 de notre modèle (Voir Tableau 3.6) :

Attribue	Pour BDD1	Pour BDD2	Pour BDD3
Précision	20 %	100%	40%
Sensibilité	25%	100%	25%
Spécificité	80%	0%	50%
F1-Measure	22.22%	100%	30.77%
AUC	52.5%	50%	37.5%
Accuracy	55.56%	64.29%	35.71%

Table 3-6: Les résultats obtenus en Arbre de Décision.

Conclusion

Le feu joue un rôle vital dans la majorité des écosystèmes forestiers. Au cours des dernières décennies, de grands efforts ont été déployés pour créer des outils de détection efficaces qui peuvent aider les systèmes de gestion des incendies dans différentes villes.

Dans ce chapitre, nous avons brièvement en revue les différentes tendances et applications de l'exploration de données depuis sa création jusqu'à l'avenir.

CHAPITRE 4

RESULTATS ET DISCUSSION

Introduction :

Nous avons présenté l'implémentation d'Approche proposée et les résultats obtenus. Dans ce chapitre nous discutons des résultats de chaque expérience et les commentons. Dans la dernière partie, nous comparons les résultats de chaque Algorithmes d'apprentissage automatique utilisés entre eux.

1. Résultats d'évaluation des approches d'apprentissage automatique sur les trois classes de donnée :

Commençons, tout d'abord par comparer notre méthode de classification aux différentes approches existantes, en termes de Accuracy pour chaque Ensemble d'apprentissage, nous avons déroulé les algorithmes suivants :

SVC, Arbre de décision. Avec la taille du train est égale à 0.8 et la taille du tests et égale à 0.2 (qui est la division la plus fréquente).

Et pour le Naïf bayes une taille d'entrainement 0.7 et une taille de test égala à 0.3.

Ensuite, Nous avons calculé l'Accuracy des approches par rapport à notre datasets comme suivants :

1.1. Résultat de 1^{er} Ensemble d'apprentissage (BDD 1)

	Naïve bayes	SVC	Arbre de Décision
Accuracy	50%	59.26%	55.56%

Table 4-1: Pourcentage de Accuracy pour BDD1.

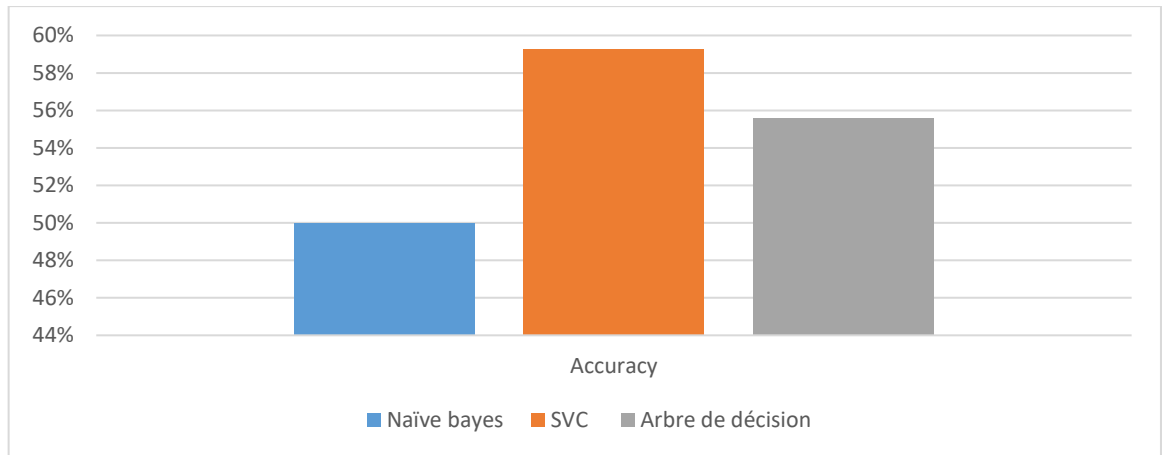


Figure 4-1: Graph de Accuracy pour BDD1.

Après avoir confronté ensemble d'apprentissage BDD 1 aux algorithmes, nous avons constaté que SVC ont donné un bon résultat par rapport à l'Arbre de décision et Naïve Bayes avec Accuracy 59.26%.

Le résultat de la courbe (ROC) :

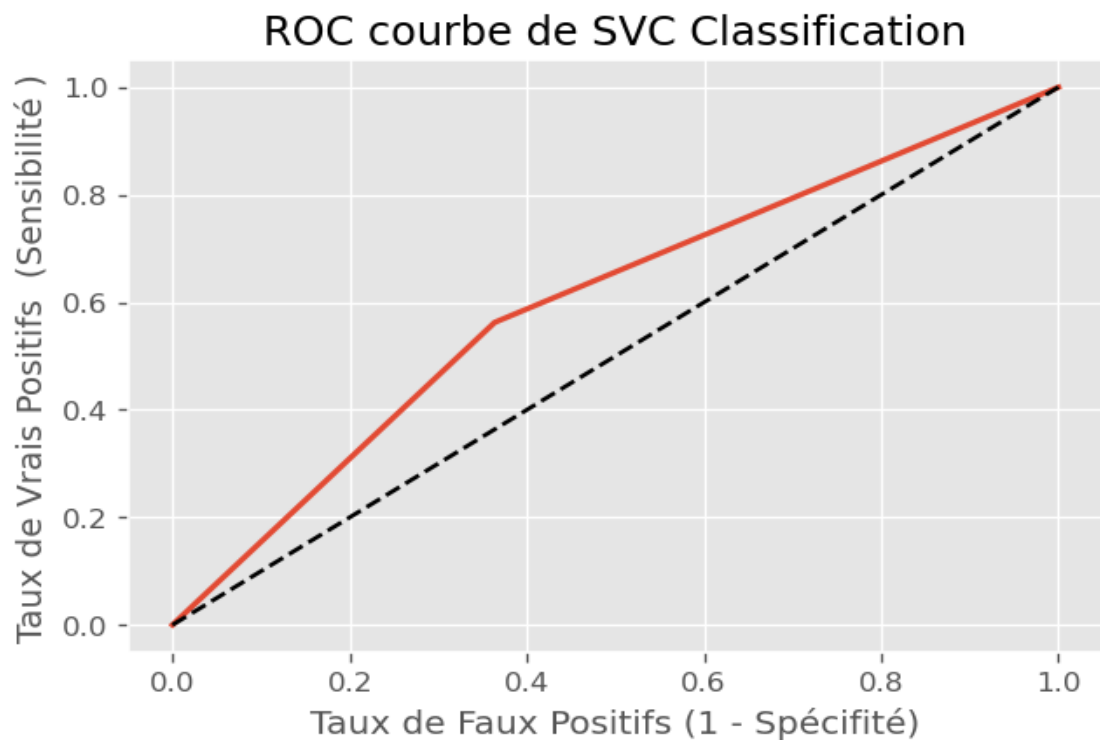


Figure 4-2: La courbe de caractéristique de fonctionnement du récepteur (ROC) de SVC pour BDD1.

1.2. Résultat de 2^{-ème} Ensemble d'apprentissage (BDD 2)

	Naïve bayes	SVC	Arbre de décision
Accuracy	66,66%	57,14%	64,29%

Table 4-2: Pourcentage de Accuracy pour BDD 2

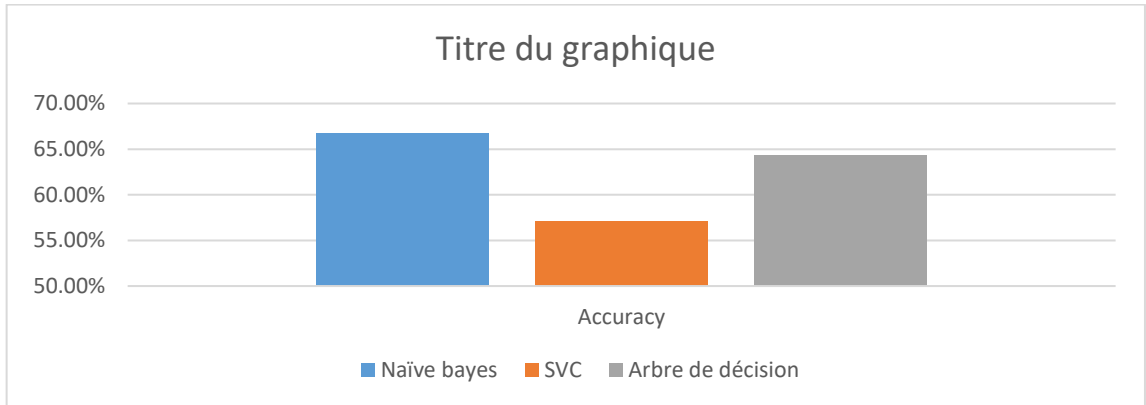


Figure 4-3: Graph de Accuracy pour BDD2

En comparant les résultats obtenus après on constate que le meilleur algorithme Naïf Bayes, qui a une précision de 66,66% lorsqu'il est appliqué à une base de données BDD2.

Le résultat de la courbe (ROC) :

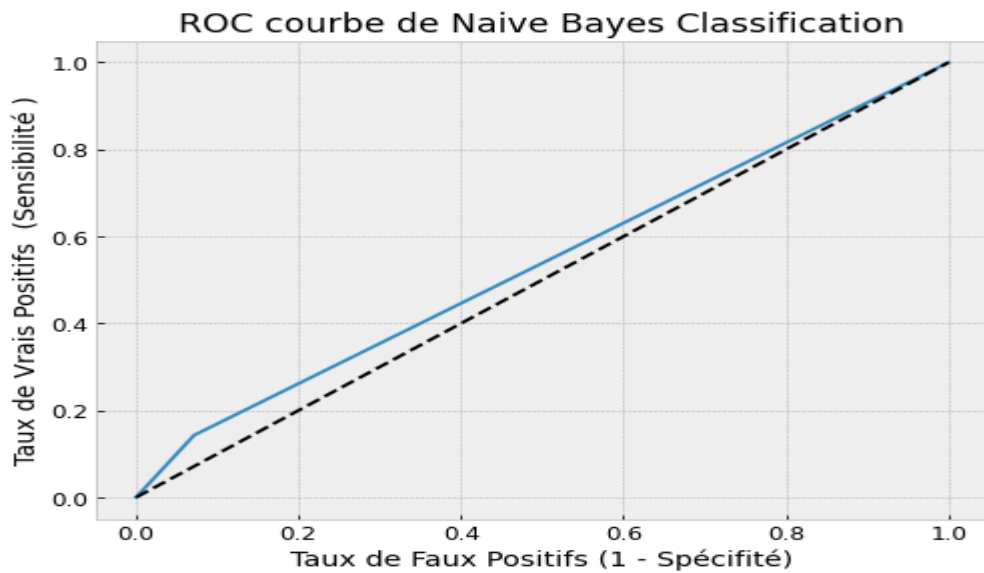


Figure 4-4 : La courbe de caractéristique de fonctionnement du récepteur (ROC) de Naive Bayes pour BDD2

1.3. Résultat de 3^{-ème} Ensemble d'apprentissage (BDD 3) :

	Naïve bayes	SVC	Arbre de décision
Accuracy	38,09%	57.14%	35,71%

Table 4-3: Pourcentage de Accuracy pour BDD 3.

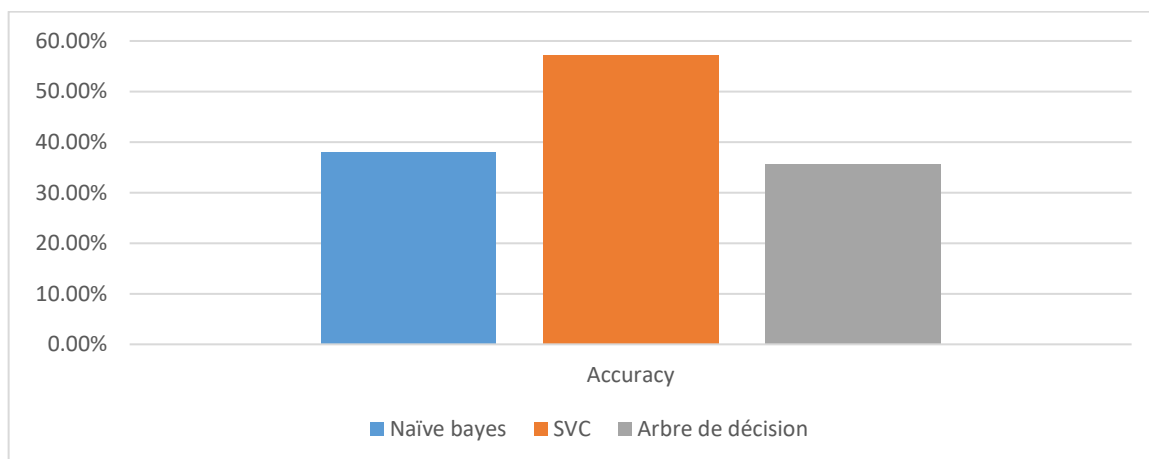


Figure 4-5: Graph de Accuracy pour BDD 3

Nous avons constaté que SVC ont donnée de bon résultat avec Accuracy 57.14 %.

Le résultat de la courbe (ROC) :

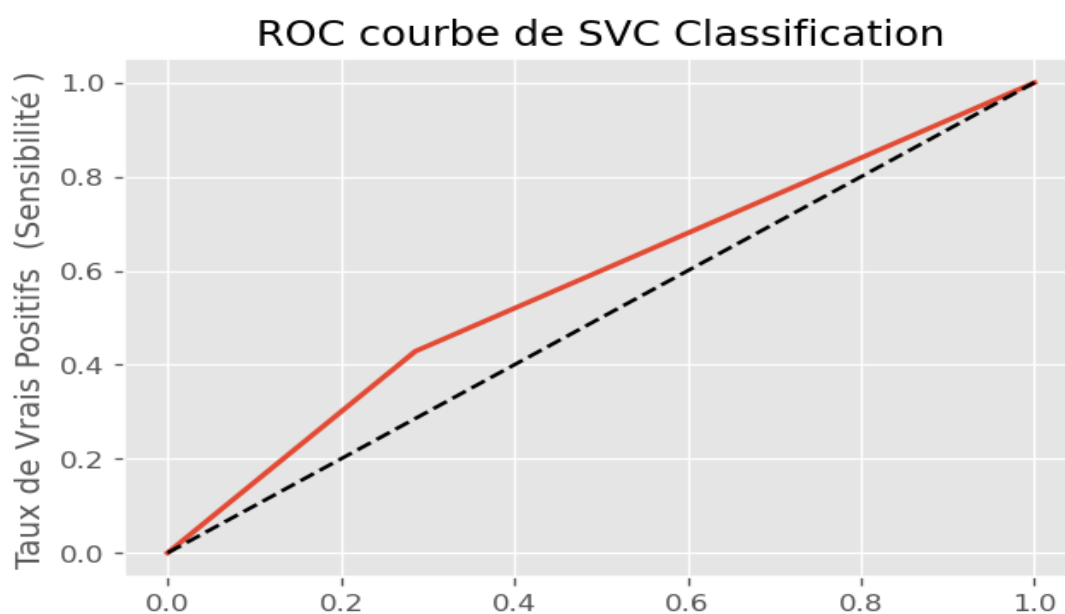


Figure 4-6 : La courbe de caractéristique de fonctionnement du récepteur (ROC) de SVC pour BDD3

1.4. Résultat de les 3 Ensembles d'apprentissage

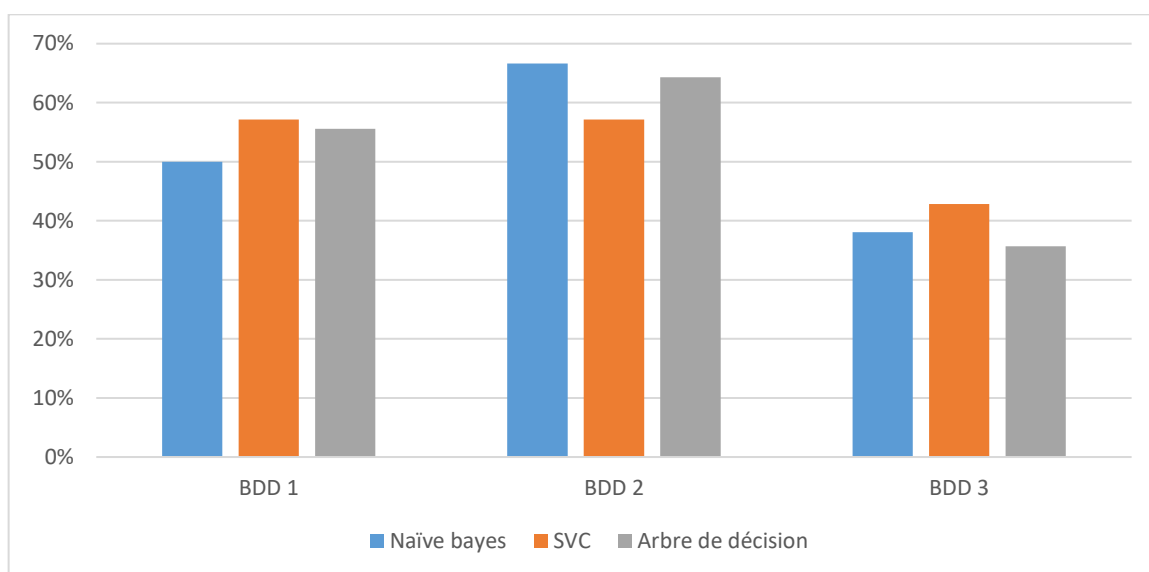


Figure 4-7 : Performance des algorithmes d'apprentissage sur les ensembles d'apprentissage (BDD1, BDD2, BDD3)

L'algorithme le plus précis dans notre approche proposée est SVC qui appliqué à trois ensembles d'apprentissage.

2. Limites de la recherche

Dans La Régression nous utilise deux model de régression Support Vector Regression (SVR) et La Régression linéaire sur la surface brulée mais à cause de la quantité limitée et insuffisante de données, le résultat obtenu était négatif. Donc les deux modèles sont éliminés.

La limitation rencontrée était l'impossibilité de collecter un grand nombre de bases de données (quantité insuffisante de données d'entraînement).

Nous n'avons utilisé que 69 lignes. Cette limite est due, d'une part, Diminution du nombre d'incendies dans M'sila.

Conclusion :

D'après les résultats de l'approche proposée, il a été constaté que les résultats de précision sont moyens, et cela n'est pas suffisant pour obtenir un modèle idéal pour prédire les incendies de forêt. Pour obtenir de meilleurs résultats, il est suggéré d'augmenter le nombre d'années qu'il faut pour la formation et les tests.

À l'avenir, nous suggérons de tester l'approche proposée en utilisant un environnement d'apprentissage en ligne dans le cadre du système de gestion des incendies. Cela nous

permettra d'obtenir après un certain temps un retour d'information précieux des gestionnaires d'incendie, concernant la justesse et l'acceptabilité de ces solutions alternatives.

Conclusion générale

Dans ce mémoire, nous avons abordé, dans le premier chapitre, une description du feu de forte la nécessité de la détection précoce de cette catastrophe pour réduire les nombres des Pertes matérielles et humaines. Ensuite, dans le deuxième chapitre, nous avons montré le rôle de l'informatique, en particulier l'intelligence artificielle, pour la recherche des solutions et des moyens pour la détection du feu de forte, dans le but de réduire le risque de cette catastrophe. Pour cela, nous avons détaillé les concepts du domaine de l'intelligence artificiel et le ML.

Notre approche et architecture proposées sont détaillées dans le troisième chapitre. Elles sont basées sur l'utilisation de trois types de algorithmes pré-entraînés :

- Machine à vecteurs de support (SVM)
- Classification naïve bayésienne.
- l'Arbre de décision

Notre choix de ces toi types est justifié par le fait qu'ils sont parmi des algorithmes les plus précis et les plus largement utilisés par rapport à d'autres algorithmes d'apprentissage supervisés.

Nous avons prouvé l'efficacité de nos modèles pour la prédiction du feu de forêt tel que le SVC était le model le plus efficace sur les trois datasets avec une prédiction passé le 50% malgré que la limite de donnée.

Perspectives

Malgré les difficultés que nous avons eu durant la réalisation de notre travail, tel que le manque des bases de données, le manque des sources d'informations, et la difficulté de comprendre le fonctionnement des algorithmes Machine à vecteurs de support (SVM), Arbre de Décision et Classification naïve bayésienne que nous avons vécu durant cette période. Nous avons réussi à réaliser et implémenter notre modèle qui a pour rôle pour prédire le feu, cela nous a permis de dégager plusieurs perspectives de travail :

- Modifier/Adapter notre architecture proposée, afin qu'elle puisse utiliser, aussi, une base de données composée de plusieurs variables plus précises.

Conclusion générale

- Remplacer algorithmes Machine à vecteurs de support (SVM), Classification naïve bayésienne et l'Arbre de décision par des autres algorithmes afin d'améliorer la précision des résultats obtenus.

- Appliquer le modèle proposé réellement sur le terrain Et son exploitation par la Direction du Gouvernorat des Forêts.

Bibliographie

- [1] FAO.2001. Protection des forêts contre l'incendie. Fiches techniques pour les pays du bassin méditerranéen. Partie 1,2,3,4,5,6,7,8. <https://www.fao.org>
- [2] Mauguen G., Lesne P., Petit J.M., Sallé B., Marc F., 2012 : Evaluation du risque méditerranéen. Écologie plantarum.
- [3] Maylan, Doméne. (2002). Les feux de forêt. MINISTÈRE DE L'ÉCOLOGIE ET DU DÉVELOPPEMENT DURABLE (MEDD). Dossier d'information. France.
- [4] FAO 2020 Protection des forêts contre l'incendie. Fiches techniques pour les pays du bassin méditerranéen. Partie 1,2,3,4,5,6,7,8. <https://www.fao.org>
- [5] Margeritt J., 1998. Modélisation et simulations numériques de la propagation de feux de forêts. Thèse, Doctorat. Inst. National polytechnique de lorraine. Nancy, France.
- [6] Plana, E. ; Font, M. ; Serra, M. ; Chauvin, S. ; Gladiné, J. 2016. Les incendies en forêt, guide pour les journalistes et les médias. Projet eFIRECOM. Edition CTFC.
- [7] Emergency Management Australia, La Prévention des Feux de Forêt en Australie (EMA)
- [8] San-Miguel-Ayanz, J., Durrant, T., Boca, R., Maianti, P., Libertá, G., Artés-Vivancos, T., Oom, D., Branco, A., de Rigo, D., Ferrari, D., Pfeiffer, H., Grecchi, R., Nuijten, D. 2022. Advance Report on Forest Fires in Europe, Middle East and North Africa 2021, EUR 31028 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-49633-5, doi:10.2760/039729, JRC128678.
- [9] CHOLLET Francois. 2017. Deep learning with python.
- [10] Innocent Mateyaunga, « Prédictive Maintenance Using Machine Learning », 2020, la thèse de master, sous la direction de Hadj Abdelkader, Faculté de technologie de l'université de Tlemcen, 61p..
- [11] Zeki Murat Çinar, Abubakar Abdussalam Nuhu, Qasim Zeeshan, Orhan Korhan, Mohammed Asmael and Babak Safaei, October 2020 , « Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0 » Sustainability.
- [12] Patrick Jahnke, June 19, 2015, «Machine Learning Approaches for Failure Type Detection and Predictive Maintenance».

- [13] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press .
- [14] Géron, A. (2017). *Hands-On_Machine_Learning_with_Scikit-learn_and_trensflow*. USA: O'Reilly Media.
- [15] Panagiotis Korvesis, le 21 novembre 2017, « Apprentissage Automatique pour la Maintenance Predictive dans le Domaine de l'Aviation », Thèse de doctorat, sous la direction de Michalis Vazirgiannis, Paris, de l'Université Paris-Saclay préparée à l'école Polytechnique, 139p.
- [16] Roger Bagra, Valentine Fontana, Wee-hyong Tok, 2014, « Predictive analytics with Microsoft azure machine learning», edition 1,188p.
- [17] Innocent Mateyaunga, 2020, « Prédictive Maintenance Using Machine Learning », la thèse de master, sous la direction de Hadj Abdelkader, Faculté de technologie de l'université de Tlemcen, 61p
- [18] NF EN 13306, Juin 2001, Norme européenne, Éditée et diffusée par l'Association Française de Normalisation (AFNOR).
- [19] Gavin Hackeling, Juin 2017, « Mastering Machine Learning with scikit-learn Second Edition ». Packt Publishing Ltd, Livery Place, UK.
- [20] Sreerama K Murthy, Simon Kasif, and Steven Salzberg. 1994. A system for induction of oblique decision trees. In *Journal of Arti_cial Intelligence Research* [55], pages 1_32.
- [21] Cortes, C., & Vapnik, V. 1995, "Support-vector networks", *Machine learning*,20(3), 273-297.
- [22] Christopher J.C.H. Watkins and Peter Dayan, May 1992, Technical note: Q-learning. *Machine Learning*, 8(3-4) :279–292.
- [23] Bassem Besbes. Int'égration de methodes de representation et de classification pour la d'etecion et la reconnaissance d'obstacles dans des scenes routieres. *Computer Science*. INSA de Rouen,2011. French. <NNT: 2011ISAM0007>. <tel 00633109>
- [24] ZACCONE Giancarlo, MD REZAUL Karim, MENSRAWY Ahmed. 2017. *Deep learning with tensorflow*.

Bibliographie

- [25] A. Betanzos, O. Romero, B. Berdiñas, E. Pereira, M. Inmaculada, E. Jimenez, J. Soto, T. Carballas, , 2003, « Un système intelligent de prévision et de lutte contre les incendies de forêt gestion en Galice », *Journal of Expert Systems with Applications*, pp.545–554.
- [26] I. Sitanggang, R. Yaakob, N. Mustapha, A. Ainuddin, , septembre 2012, « Application of classification algorithmes d'exploration de données pour la prédiction de l'occurrence des points chauds dans la province de Riau en Indonésie » *Journal of Theoretical and Applied Information Technology*, 43 No.2.
- [27] L. Iliadis, A. Papastavrou, P. Lefakis, , juillet 2002, « Un système expert heuristique de guidage des feux de forêt en Grèce » *J Environ Manage*, pp.327-36.
- [28] Z. Jing, M. Weiqing et Z. Ye, , 2015, « Fisher Linear Discriminant Method for Forest Fire Risk Points on Transmission Line », *International Journal of Smart Home*, 9, pp. 25-34.
- [29] python.doctor. (2019). Apprendre le langage de programmation python. Retrieved 04 29, 2020, from python.doctor: <https://python.doctor/>
- [30] Gupta, N. (2019, 10 29). Why is Python Used for Machine Learning? Retrieved 04 29, 2020, from Hacker Noon: <https://hackernoon.com/why-python-used-for-machine-learning-u13f922ug>
- [31] Judith Hurwitz, Daniel Kirsch . (2018). *Machine Learning For Dummies®*, IBM Limited Edition.
- [32] "Historique météo : archives météo pour le monde entier", 2022, <https://www.historique-meteo.net/afrique/algerie/msila> .

Résumé

Les incendies des forêts constituent l'un des principaux risques naturels des forêts dans le monde. Dans certains pays et certaines régions, comme par exemple en Méditerranée, il s'agit même du risque naturel le plus important en forêt.

En dépit des efforts considérables déployés dans des actions de prévention et d'extinction, les grands incendies de forêt restent des événements récurrents qui ont souvent des interactions avec les infrastructures et les habitations, concernant ainsi de larges franges de la population.

L'intelligence artificielle est largement utilisée dans la vie quotidienne et dans de nombreux domaines différents tels que l'industrie, l'automatisation et les systèmes experts, la médecine et la biologie, l'éducation et même les jeux vidéo.

L'apprentissage supervisé est l'une des branches les plus importantes et les plus exploitées de l'apprentissage automatique. Grâce à son succès, il a réalisé un saut quantique dans le domaine de l'intelligence artificielle. Par conséquent, il est devenu le domaine le plus important dans le monde de l'informatique. L'apprentissage supervisé est basé sur l'apprentissage et l'acquisition d'expériences. Il simule la façon dont les humains pensent et apprennent.

Machine à vecteurs de support (SVM), Classification naïve bayésienne et arbre de décision sont considérés comme l'un des premiers algorithmes d'apprentissage supervisé. Ils ont donné des résultats très efficaces, en particulier dans la classification et la régression.

Dans ce travail, nous nous sommes intéressés par la proposition d'une nouvelle approche basée sur une branche d'intelligence artificielle qui est « l'apprentissage supervisée », pour la prédiction du Feu de forêt. Cette approche est basée sur l'utilisation de deux types des algorithmes pré-entraînés : Machine à vecteurs de support (SVM), Classification naïve bayésienne et arbre de décision. Ces modèles peuvent être utilisés comme outils de prévention efficaces dans la gestion des incendies de forêt.

Mots clés : l'intelligence artificielle, Les incendies des forêts, l'apprentissage automatique supervisé, la prédiction.

Abstract

Forest fires are one of the main natural hazards to forests worldwide. In some countries and regions, such as the Mediterranean, it is even the most significant natural risk in the forest.

Despite the considerable efforts deployed in prevention and extinction actions, large forest fires remain recurring events that often interact with infrastructure and housing, thus affecting large sections of the population.

Artificial intelligence is widely used in everyday life and in many different fields such as industry, automation and expert systems, medicine and biology, education and even video games.

Supervised learning is one of the most important and exploited branches of machine learning. Thanks to his success, he achieved a quantum leap in the field of artificial intelligence. Therefore, it has become the most important field in the computer world. Supervised learning is based on learning and the acquisition of experiences. It simulates the way humans think and learn.

Support Vector Machine (SVM), Naive Bayesian Classification and decision tree is considered as one of the first supervised learning algorithms. They gave very efficient results, especially in classification and regression.

In this work, we are interested in the proposal of a new approach based on a branch of artificial intelligence which is "supervised learning", for the prediction of forest fire. This approach is based on the use of two types of pre-trained algorithms: Support Vector Machine (SVM), Naive Bayesian Classification and Decision Tree. These models can be used as effective prevention tools in forest fire management.

Keywords: artificial intelligence, forest fires, supervised machine learning, prediction.

ملخص:

حرائق الغابات هي واحدة من المخاطر الطبيعية الرئيسية للغابات في جميع أنحاء العالم. في بعض البلدان والمناطق، مثل البحر الأبيض المتوسط، يعتبر الخطر الطبيعي الأكثر أهمية في الغابة. على الرغم من الجهود الكبيرة المبذولة في إجراءات الوقاية والانقراض، لا تزال حرائق الغابات الكبيرة أحداثًا متكررة تتفاعل غالبًا مع البنية التحتية والإسكان، مما يؤثر على قطاعات كبيرة من البشر. يستخدم الذكاء الاصطناعي على نطاق واسع في الحياة اليومية وفي العديد من المجالات المختلفة مثل الصناعة والرقمنة والأنظمة الذكية والطب والبيولوجيا والتعليم وحتى ألعاب الفيديو.

يعد التعلم الخاضع للإشراف أحد أهم فروع التعلم الآلي وأكثرها استغلالًا. بفضل نجاحه حقق نقلة نوعية في مجال الذكاء الاصطناعي. لذلك، أصبح أهم مجال في عالم الكمبيوتر. يعتمد التعلم الخاضع للإشراف على التعلم واكتساب الخبرات. إنه يحاكي طريقة تفكير البشر وتعلمهم. تعتبر آلة المتجهات الداعمة (SVR)، تصنيف بايزي الساذج وشجرة القرار من أولى خوارزميات التعلم الخاضعة للإشراف. أعطوا نتائج فعالة للغاية، خاصة في التصنيف والانحدار.

في هذا العمل، نحن مهتمون باقتراح نهج جديد قائم على فرع من الذكاء الاصطناعي وهو "التعلم الخاضع للإشراف"، للتنبؤ بحرائق الغابات. يعتمد هذا النهج على استخدام نوعين من الخوارزميات المدربة مسبقًا: آلة المتجهات الداعمة (SVR) تصنيف بايزي الساذج وشجرة القرار. يمكن استخدام هذه النماذج كأدوات وقائية فعالة في إدارة حرائق الغابات.

الكلمات المفتاحية: الذكاء الاصطناعي، حرائق الغابات، التعلم الآلي الخاضع للإشراف، التنبؤ.