

Mémoire du stage d'initiation

Spécialité : Génie Intelligence Artificielle et Génie Informatique

*** * Asmaa MOUSSAFIR * ***

*** * Imane ELMOUDEN * ***

Titre :

**Etude et modélisation de l'orientation dans le monde
d'intelligence artificielle**

Organisme d'Accueil : *ENSAM Casablanca*

Table des matières

Remerciements	4
Contexte	5
Introduction générale	6
Chapitre 1	7
1. Informations générales	7
2. Filières assurées par l'école	7
3. Avantages de la formation.....	7
Chapitre 2	11
1. Introduction	11
2. L'apprentissage supervisé.....	11
i. Background.....	11
ii. Types d'apprentissage supervisé.....	11
iii. Algorithmes d'apprentissage automatique supervisé	11
iv. Résumé	11
3. L'apprentissage non supervisé	11
i. Background.....	11
ii. Algorithmes d'apprentissage automatique non supervisé.....	11
iii. Résumé	11
4. Les langages pour le Machine Learning	12
i. Background.....	12
ii. Les langages les plus utilisés	12
iii. Résumé	12
Chapitre 3	25
1. Introduction	25
2. Les difficultés d'orientation pour les étudiants	25
3. Rôle de l'intelligence artificielle dans l'orientation.....	25
4. Conclusion.....	25
Chapitre 4	29
1. Introduction	29
2. Objectif.....	29

3. Encadrement du modèle	29
4. Données utilisées	29
5. Algorithmes à comparer	29
i. LightGbm	29
ii. Decision Trees (Arbres de Décision)	29
iii. Naive Bayes	29
iv. Neural Networks (Réseaux de Neurones).....	29
v. Random Forest Algorithme (RF)	29
6. Comparaison	29
▪ Explication des résultats	29
Le diagramme de Gantt	37
Conclusion générale	37
Webographie	38
Résumé.....	39

Remerciements

En espérant que ce travail soit de qualité et plait à tous les membres, nos remerciements s'adressent tout d'abord, à tous les membres du Jury qui nous font honneurs d'avoir accepté de juger ce modeste travail.

Ce projet est le fruit d'un travail mené au sein de l'École Nationale Supérieure d'Arts et Métiers de Casablanca dirigé par le directeur Ahmed MOUCHTACHI à qui nous voudrions exprimer toute notre gratitude pour sa confiance et pour tous les moyens techniques et scientifiques qu'il a mis à notre disposition.

Nous présentons nos sincères remerciements à Monsieur le docteur BAHNASS Ayoub, l'encadrant de projet, d'avoir accepté notre encadrement et pour son aide précieuse, son accompagnement professionnel, et l'intérêt qu'il nous a accordé durant la période du projet que ce soit en présentiel ou à distance.

Nos remerciements s'adressent également à tout le personnel de l'ENSAM de Casablanca pour leur soutien et pour leur générosité quant à l'offre de l'information.

Que tous ceux et celles qui ont contribué de près ou de loin à l'accomplissement de ce travail trouvent ici l'expression de nos remerciements les plus chaleureux.

Contexte

La digitalisation est un processus qui vise à convertir des objets, des outils ou des processus en code informatique afin de les remplacer et d'améliorer leur efficacité. Cette transformation digitale (numérique) s'est apparue dans de nombreuses alternatives, par exemple le courrier a été remplacé par le courrier électronique, les salons professionnels ont été remplacés par des forums en ligne et les magasins ont été remplacés par des sites de commerce électronique. Mais... avez-vous pensé à la digitalisation de l'éducation ?

La digitalisation (numérisation) de l'éducation est une tendance puissante en termes de réforme et de modernisation dans l'environnement éducatif mondial, en outre, plusieurs universités ont pris l'initiative d'abandonner l'époque des papiers et de s'évoluer vers la flexibilité et la performance, on parle ici de l'aire de l'intelligence artificielle.

L'intelligence artificielle a pris une évolution à long terme en tout ce qui concerne l'éducation en général, quel que soit son utilisation dont on cite par exemple les évaluations, la correction des examens, les cours en ligne et éventuellement la partie d'orientation de l'étudiant.

Qui dit l'orientation de l'étudiant, dit une orientation d'avenir dont on peut bien jeter la lumière sur chaque petite mauvaise vision de parcours plutôt qu'une mauvaise orientation de l'étudiant et qui peut lui générer à tout moment un recule jusqu'à une capitulation(céder) causée par le dialogue insuffisant, infructueux, ou souvent non instauré entre les élèves, les parents, les enseignants et les conseillers d'éducation et d'orientation pédagogique. Alors qu'une bonne orientation se base sur une masse complète d'informations, d'une bonne étude des compétences possédées et aussi nécessaires pour accomplir le bon parcours, car toute fois, le choix de l'orientation est une décision capitale qui doit être mûrement réfléchie et préparée par l'étudiant. Aujourd'hui, et spécialement grâce à l'IA on peut passer au niveau digital de l'orientation à travers un nombre important de données stockées, étudiées et précisément réutilisées pour offrir à chaque étudiant une aide technologique lui facilite cette tâche.

Outre que l'IA reste un peu limitée en ce qui concerne ce qu'on ne peut pas le représenter sous la forme de données. L'orientation en fin de compte est un choix ou plutôt une décision qui ne tient pas compte en disant juste les compétences, mais aussi plutôt des sentiments basculent entre un courage, un stress, une peur ou un état personnel etc. et que l'IA ne peut pas les prendre en considération.

Introduction générale

L'orientation est devenue un élément clé pour garantir la réussite de l'étudiant. Dans le cas de ENSAM CASABLANCA l'orientation s'avère indispensable pour les étudiants ayant un passage depuis les classes préparatoires vers le cycle d'ingénieurs.

La méthode actuelle se base sur le classement par note, alors que les notes parfois ne sont pas un bon critère pour affecter un étudiant à une filière, en considérant qu'il faut voir les notes de spécialité, l'intérêt de l'étudiant ainsi que ses capacités, d'où l'intelligence artificielle vient d'être proposé pour soulever ce genre de problèmes.

Dans ce stage de fin d'études, nous allons proposer une nouvelle méthode qui permet de faire une prédiction de la filière recommandée pour l'étudiant.

Notre méthode a comme entrée les notes des années préparatoires de l'étudiant afin de générer une filière recommandée en sortie et grâce à l'intelligence artificielle l'étudiant aura vers la fin une recommandation de la filière la plus recommandée

Le rapport de stage est organisé comme suit :

- Dans le premier chapitre, on va parler de l'état de l'art à travers laquelle nous allons présenter l'ensemble des algorithmes
- Dans le 2ème chapitre, on va traiter l'orientation et la vision de l'IA.
- Le 3ème chapitre sera concerné à la modélisation

Et enfin on va conclure.

Chapitre 1

Présentation de l'organisme d'accueil

Dans ce chapitre :

1. Informations générales
2. Filières assurées par l'école
3. Avantages de la formation

1. Informations générales

L'École Nationale Supérieure d'Arts et Métiers de Casablanca relevant de l'Université Hassan II de Casablanca, est une école d'ingénieurs étatique Créée en 2010 dans la plus grande agglomération industrielle du Royaume pour répondre aux besoins en ingénieurs. Elle assure une formation d'ingénieurs pluri-compétents, humanistes, innovants et dotés d'un esprit entrepreneurial, opérants dans différents secteurs industriels privés et publics à l'échelle national et international.

Afin d'être à la hauteur des attentes de ses élèves ingénieurs, et convaincue que leur formation doit combiner : Qualité technique, managériale et relationnelle ; l'ENSAM de Casablanca vise le développement d'une offre très large de formations. Partant de ce principe, et en incitant et motivant ses étudiants à organiser différents événements (Sociales, professionnels, culturels et sportifs), elle permet à ses futurs lauréats d'acquérir :

- Les compétences techniques et professionnelles de l'ingénieur.
- L'épanouissement de la personnalité.
- L'apprentissage de la responsabilité.
- L'adaptation à un milieu professionnel en constante évolution.

2. Filières assurées par l'école

La formation à l'ENSAM– Casablanca dure cinq ans : deux années préparatoires et trois années de formations spécialisées en l'une des filières :

- **GÉNIE MÉCANIQUE**
- **GÉNIE INDUSTRIEL**
- **GENIE ELECTROMECHANIQUE**
- **MANAGEMENT DES SYSTEMES ELECTRIQUES INTELLIGENTS**
- **INTELLIGENCE ARTIFICIELLE ET GENIE INFORMATIQUE**
- **MASTER BIG DATA ET INTERNET DES OBJETS**

En choisissant d'intégrer l'ENSAM-Casablanca, vous optez pour une formation d'ingénieurs de référence. L'accès en première année de l'ENSAM est réservé aux élèves ayant obtenu un baccalauréat scientifique (Sciences Mathématiques A ou B, Sciences Expérimentales) ou un baccalauréat technique.

L'accès en troisième et quatrième année de l'ENSAM est réservé aux élèves ayant réussi le concours national communs et titulaires des diplômes de DUT, DEUG, DEUST, BTS, DTS, licence ou tout diplôme reconnu équivalent.

3. Avantages de la formation

❖ Une pédagogie attrayante pour un métier d'avenir

80% de la formation à l'ENSAM-Casablanca est basée sur les sciences et techniques. Elle donne accès à de nombreux secteurs industriels porteurs (automobile, aéronautique, mécatronique, énergie...). 20% de la formation est centrée sur la pratique et à la maîtrise de l'ingénierie de projet, aux techniques de communication, au travail en équipe et à la gestion de l'entreprise. Ces disciplines apportent aux étudiants-ingénieurs les compétences managériales indispensables à leur futur métier et leur permet d'être immédiatement opérationnels dans le monde des entreprises.

❖ Une formation axée sur l'entreprise pour une insertion réussie

L'intégration dans l'entreprise, par le biais des stages de découverte, d'ingénieur et le Projet de fin d'étude, fait partie intégrante de la formation de l'ingénieur de l'ENSAM. L'enseignement couvre des domaines étendus en étroite relation avec les secteurs industriels porteurs et les enjeux sociétaux prioritaires.

La formation délivrée permet aussi aux jeunes ingénieurs de s'orienter vers des fonctions types bureaux d'études, Industrie automobile, aéronautique, navale et ferroviaire ...

❖ Un partenariat large pour plus d'ouverture

Les partenariats pédagogiques constituent un axe incontournable du développement de l'école.

Au niveau national, l'école fait partie du réseau national des grandes écoles d'ingénieurs.

A l'international, l'ENSAM a développé des liens forts avec de grandes écoles d'ingénieur en France comme l'UTC de Belfort Montbéliard, l'ENSAM ParisTech, SIGMA et l'ENSEM... Certes, dans une ère de mondialisation des efforts sont toujours déployés pour une grande ouverture sur l'international.

Les relations industrielles de l'ENSAM ne cessent de s'accroître dans un tissu industriel casablancais assez développé et assez varié.

Un cadre de travail privilégié pour une vie étudiante active et citoyenne

Pour rendre l'école plus dynamique et vivante, l'association des élèves, à travers ses clubs, propose aux étudiants une multitude d'activités allant du domaine artistique au domaine technique.

Elle assure l'animation et la gestion de la vie étudiante et fédère plus de 15 clubs et associations qui ont pour objectif le développement personnel des élèves ingénieurs de l'ENSAM-Casablanca.

Chapitre 2

Etat de l'art du Machine Learning

Dans ce chapitre :

1. Introduction
2. L'apprentissage supervisé
 - i. Background
 - ii. Types d'apprentissage supervisé
 - iii. Algorithmes d'apprentissage automatique supervisé
 - Régression linéaire
 - Régression logistique
 - Naïve Bayes
 - SVM (Support Vector Machine)
 - Arbre de décision
 - iv. Résumé
3. L'apprentissage non supervisé
 - i. Background
 - ii. Algorithmes d'apprentissage automatique non supervisé
 - K-means
 - Réseaux de Neurone
 - iii. Résumé

4. Les langages pour le Machine Learning

- i. Background
- ii. Les langages les plus utilisés
 - Python
 - C++
 - Le Langage R
 - JAVA
- iii. Résumé

1. Introduction

L'intelligence artificielle (IA) est une vaste branche de l'informatique qui implique la construction de machines intelligentes capables d'exécuter des tâches qui nécessitent normalement une intelligence humaine.

Les machines dotées d'une intelligence artificielle mémorisent des comportements. Ce travail de mémorisation leur permet de résoudre les problèmes plus tard et d'agir correctement face à cette situation. L'utilisation des bases de données et des algorithmes est nécessaire pour la réalisation de cet apprentissage.

Cette tâche complexe peut aider la machine à mesurer l'importance du problème, à analyser les solutions possibles et les situations antérieures similaires afin de bien agir. En fait, il s'agit d'un système statistique complexe et puissant qui peut amener les machines à prendre des décisions ou à avoir des comportements attendus.

Les domaines d'application et les usages potentiels de l'intelligence artificielle sont de plus en plus diversifiés : compréhension du langage naturel, reconnaissance visuelle, robotique, systèmes autonomes et le Machine Learning (Apprentissage Automatique).

De plus en plus de systèmes informatiques adoptent l'IA et l'apprentissage automatique. L'apprentissage automatique fait partie de l'intelligence artificielle qui se concentre sur l'étude des algorithmes informatiques et mathématiques et des ensembles de données pour prendre des décisions sans écrire de code manuel.

L'apprentissage automatique est un excellent exemple de technologie qui est récemment venue du milieu universitaire et des études théoriques aux applications pratiques et au soutien des activités commerciales au quotidien.

Le principe de base de l'apprentissage automatique est de créer des algorithmes capables de recevoir des données d'entrée et d'utiliser une analyse statistique pour prédire une sortie tout en les mettant à jour à mesure que de nouvelles données deviennent disponibles.

Les algorithmes du Machine Learning peuvent être classés selon le mode d'apprentissage qu'ils emploient : apprentissage supervisé, non supervisé ou semi-supervisé.

2. Apprentissage supervisé (supervised learning):

i. Background

L'apprentissage supervisé est une tâche d'apprentissage automatique qui apprend des fonctions prédictives à partir d'exemples annotés. L'apprentissage supervisé est basé sur la vérité. En d'autres termes, nous avons une connaissance préalable de la valeur de sortie de l'échantillon. Par conséquent, le but de l'apprentissage supervisé est d'apprendre une fonction qui se rapproche le mieux de la relation entre l'entrée et la sortie observables dans les données, à partir d'un échantillon de données et du résultat souhaité.

L'apprentissage supervisé se compose de variables d'entrée (X) et de variables de sortie (Y). Vous utilisez un algorithme pour apprendre la fonction de mapping entrée-sortie. $Y = f(X)$

Le but est de bien comprendre la fonction de mapping afin que, lorsque vous avez de nouvelles données d'entrée (x), vous pouvez prédire la variable de sortie (Y) des données.

Il s'est appelé apprentissage supervisé car le processus de l'algorithme obtenu dans l'ensemble d'apprentissage peut être considéré comme l'enseignant du processus d'apprentissage supervisé. Nous connaissons la bonne réponse, l'algorithme prédira itérativement les données d'apprentissage et les corrigera par l'enseignant. Lorsque l'algorithme atteint un niveau de performance acceptable, l'apprentissage s'arrête.

L'apprentissage supervisé est généralement réalisé dans le contexte de la classification et de la régression.

La classification est le processus de prédiction d'étiquettes ou de catégories de classes discrètes c'est à dire la sortie est une catégorie telle que « jaune », « vert ».

La régression est le processus de prédiction des valeurs continues c'est à dire la sortie est une valeur réelle.

ii. Algorithmes d'apprentissage automatique supervisé

- **Régression Linéaire**

La régression linéaire est un algorithme d'apprentissage automatique basé sur l'apprentissage supervisé. Il effectue une tâche de régression. La régression modélise une valeur de prédiction cible basée sur des variables indépendantes. Il est principalement utilisé pour découvrir la relation entre les variables et les prédictions. Différents modèles de régression dépendent du type de relation entre les variables dépendantes et indépendantes envisagées et du nombre de variables indépendantes utilisées.

La régression linéaire effectue la tâche de prédire une valeur de variable dépendante (y) basée sur une variable indépendante donnée (x). Ainsi, cette technique de régression découvre une relation linéaire entre x (entrée) et y (sortie). Par conséquent, le nom est régression linéaire.

Lors de la formation du modèle, on nous donne :

X : données d'entraînement d'entrée

Y : étiquettes aux données

Lors de la formation du modèle, il ajuste la meilleure ligne pour prédire la valeur de y pour une valeur donnée de x . Le modèle obtient la meilleure droite d'ajustement de régression en trouvant les meilleures valeurs θ_1 et θ_2 .

- **Régression Logistique**

La Régression logistique est un algorithme d'apprentissage automatique pour les problèmes de classification. C'est un algorithme d'analyse prédictive basé sur le concept de probabilité, utilisé pour prédire la probabilité d'une variable cible.

La régression logistique indique généralement là où la frontière entre les classes existe, aussi indique que les probabilités de classe dépendent de la distance par rapport à la frontière, dans une approche spécifique.

Généralement, la régression logistique signifie une régression logistique binaire ayant des variables cibles binaires, c'est à dire la nature de la variable cible ou dépendante est la dichotomie, ce qui signifie qu'il n'y aura que deux classes possibles, les données sont codées

comme 1 (indiquant le succès / oui) ou 0 (indiquant l'échec / non) mais elle peut être utilisée pour décrire les données et pour expliquer la relation une variable binaire dépendante et une ou plusieurs variables indépendantes nominales, ordinales, d'intervalle ou de rapport.

Nous pouvons appeler une régression logistique un modèle de régression linéaire, mais la régression logistique utilise une fonction de coût plus complexe, cette fonction de coût peut être définie comme la « fonction sigmoïde » ou également connue sous le nom de « fonction logistique » au lieu d'une fonction linéaire.

L'hypothèse de régression logistique tend à limiter la fonction de coût entre 0 et 1. Par conséquent, les fonctions linéaires ne parviennent pas à la représenter car elle peut avoir une valeur supérieure à 1 ou inférieure à 0, ce qui n'est pas possible selon l'hypothèse de régression logistique.

- **Naïve Bayes**

Le classificateur Naïve Bayes est un classificateur probabiliste simple basé sur le théorème de Bayes avec des hypothèses d'indépendance fortes et naïves, il prédit les probabilités d'appartenance pour chaque classe, telles que la probabilité qu'un enregistrement appartient à une classe particulière.

Pour comprendre le classificateur naïf de Bayes, nous devons comprendre le théorème de Bayes.

Théorème de Bayes nommé d'après le révérend Thomas Bayes. Cela fonctionne sur la probabilité conditionnelle. La probabilité conditionnelle est la probabilité que quelque chose se produise, étant donné que quelque chose d'autre s'est déjà produit. En utilisant la probabilité conditionnelle, nous pouvons calculer la probabilité d'un événement en utilisant ses connaissances antérieures.

La classe avec la probabilité la plus élevée est considérée comme la classe la plus probable. Ceci est également connu sous le nom de Maximum A Posteriori (MAP).

- **Support Vector Machine**

Le SVM est un algorithme d'apprentissage automatique supervisé qui peut être utilisé à la fois pour les problèmes de classification ou de régression. Cependant, il est principalement utilisé dans les problèmes de classification. Dans l'algorithme SVM, nous traçons chaque élément de données comme un point dans un espace à n dimensions (où n est le nombre d'entités dont vous disposez), la valeur de chaque entité étant la valeur d'une coordonnée particulière. Ensuite, nous effectuons la classification en trouvant l'hyper-plan qui différencie les deux classes.

Pour séparer les deux classes de points de données, il existe de nombreux hyperplans possibles qui pourraient être choisis. Notre objectif est de trouver un plan qui a la marge maximale, c'est-à-dire la distance maximale entre les points de données des deux classes. L'optimisation de la distance de marge fournit quelques améliorations afin que les futurs points de données puissent être classés avec plus de confiance.

- **Arbre de Décision**

L'arbre de décision est une technique d'apprentissage supervisé qui peut être utilisée à la fois pour les problèmes de classification et de régression, mais elle est principalement préférée pour résoudre les problèmes de classification.

Il s'agit d'un classificateur à structure arborescente, où chaque nœud représente une caractéristique (attribut), chaque lien (branche) représente une décision (règle) et chaque feuille représente un résultat (valeur catégorielle ou continue).

L'objectif de l'utilisation d'un arbre de décision est de créer un modèle d'apprentissage qui peut être utilisé pour prédire la classe ou la valeur de la variable cible en apprenant des règles de décision simples déduites de données antérieures (données d'apprentissage).

Dans les arbres de décision, pour prédire une étiquette de classe pour un enregistrement, nous partons de la racine de l'arbre. Nous comparons les valeurs de l'attribut racine avec l'attribut de l'enregistrement. Sur la base de la comparaison, nous suivons la branche correspondant à cette valeur et passons au nœud suivant.

iii. Résumé

L'apprentissage supervisé est une méthode populaire pour entraîner les algorithmes car les développeurs et les scientifiques peuvent garder un contrôle complet. Bien que les résultats de l'utilisation d'autres méthodes d'apprentissage soient généralement vagues, l'apprentissage supervisé peut vous donner une idée précoce de ce qui se passera pendant le processus d'apprentissage. Cependant, il est tout aussi important que les personnes chargées de l'apprentissage travaillent également dur.

3. Apprentissage non-supervisé (unsupervised learning) :

i. Background

L'apprentissage non supervisé signifie n'avoir que des données d'entrée (X) et aucune variable de sortie correspondante. Le but de l'apprentissage non supervisé est de modéliser la structure de base ou la distribution dans les données pour en savoir plus sur les données.

C'est ce qu'on appelle l'apprentissage non supervisé car, contrairement à l'apprentissage supervisé mentionné ci-dessus, il n'a pas de réponses correctes et pas d'enseignant. Ces algorithmes quittent leurs propres mécanismes pour découvrir et présenter des structures de données intéressantes.

L'apprentissage non supervisé comprend deux types d'algorithmes : les algorithmes de regroupement et les algorithmes d'association.

Le clustering implique le regroupement des points de données en fonction de leurs similitudes, tandis que l'association implique la découverte de relations entre les attributs de ces points de données.

ii. Algorithmes d'apprentissage automatique non supervisé

- **K-means**

Le clustering K-means est l'un des algorithmes d'apprentissage automatique non supervisé les plus simples et les plus populaires.

Un cluster fait référence à une collection de points de données agrégés en raison de certaines similitudes.

Vous allez définir un nombre cible k , qui fait référence au nombre de centres de gravité dont vous avez besoin dans l'ensemble de données. Un centroïde est l'emplacement imaginaire ou réel représentant le centre de l'amas.

Chaque point de données est alloué à chacun des clusters en réduisant la somme des carrés dans le cluster.

En d'autres termes, l'algorithme K-means identifie k nombre de centroïdes, puis alloue chaque point de données au cluster le plus proche, tout en gardant les centroïdes aussi petits que possible.

Le « means » dans le K-means se réfère à la moyenne des données ; c'est-à-dire trouver le centre de gravité.

- **Neural networks**

Les réseaux de neurones sont largement utilisés dans l'apprentissage non supervisé pour apprendre une meilleure représentation des données d'entrée.

Par exemple, étant donné un ensemble de documents texte, NN peut apprendre un mappage d'un document à un vecteur à valeur réelle de telle sorte que les vecteurs résultants soient similaires pour les documents avec un contenu similaire, c'est-à-dire en préservant la distance. Cela peut être réalisé en utilisant, par exemple, des codeurs automatiques - un modèle qui est formé pour reconstruire le vecteur d'origine à partir d'une représentation plus petite (activations de couche cachée) avec une erreur de reconstruction (distance à la fonction ID) comme fonction de coût.

Ce processus ne vous fournira pas de clusters, mais il créera une représentation significative qui peut être utilisée pour le cluster. Par exemple, vous pouvez exécuter un algorithme de clustering lors de l'activation de la couche masquée. Cluster : il existe de nombreuses architectures NN différentes spécialement conçues pour les clusters.

La plus connue est probablement les cartes autoorganisées. Un SOM est un NN qui a un ensemble de neurones connectés pour former une grille topologique (généralement rectangulaire).

Lorsqu'un modèle est présenté à un SOM, le neurone avec le vecteur de poids le plus proche est considéré comme gagnant et ses poids sont adaptés au modèle, ainsi que les poids de son voisinage. De cette manière, un SOM trouve naturellement des grappes de données. Un algorithme quelque peu apparenté est la croissance du gaz neuronal (il n'est pas limité à un nombre prédéfini de neurones).

Une autre approche est la théorie de la résonance adaptative avec deux couches : « champ de comparaison » et « champ de reconnaissance ». Le champ de reconnaissance détermine également la meilleure correspondance (neurone) avec le vecteur transféré à partir du champ de comparaison et ont également des connexions inhibitrices latérales. Les détails de mise en œuvre et les équations exactes peuvent facilement être trouvés en recherchant sur Google les noms de ces modèles, donc je ne les mettrai pas ici.

iii. Résumé

L'apprentissage automatique n'est pas seulement synonyme de progrès technologique, il permet également aux gens de se sentir détendus et de faciliter leur vie dans de nombreux domaines de la vie quotidienne. Il enrichit la vie quotidienne, l'économie et la recherche. Contrairement aux deux autres méthodes d'apprentissage (apprentissage supervisé et apprentissage par renforcement), les développeurs ne participent pas à l'entraînement à proprement parler, ce qui non seulement fait gagner du temps, mais présente également un autre avantage : l'apprentissage non supervisé permet aux gens d'identifier des modèles que les humains n'utiliseront pas.

Enfin, il faut souligner qu'il existe d'autres types d'apprentissages, comme l'apprentissage semi-supervisé, qui comprend un « hybride » entre l'apprentissage supervisé et non supervisé.

4. Les langages pour Machine Learning

i. Background

Presque tous les langages de programmation peuvent être utilisés pour écrire des applications basées sur le ML. Cependant, écrire chaque algorithme à partir de zéro est un processus qui prend du temps. Le langage de programmation le mieux adapté est celui qui est fourni avec des bibliothèques prédéfinies et qui prend en charge de manière avancée la science des données et les modèles de données

Selon les experts du secteur il n'y a pas de meilleur langage pour l'apprentissage automatique, chacun étant bien là où il convient le mieux. Cependant, il existe certainement des langages de programmation plus appropriés pour les tâches d'apprentissage automatique que d'autres.

De nombreux ingénieurs en apprentissage automatique choisissent un langage d'apprentissage automatique en fonction du type de problème commercial sur lequel ils travaillent.

Par exemple, la plupart des ingénieurs en apprentissage automatique préfèrent utiliser Python pour les problèmes de PNL tout en préférant également utiliser R ou Python pour les tâches d'analyse des sentiments, et certains sont susceptibles d'utiliser Java pour d'autres applications d'apprentissage automatique telles que la sécurité et la détection des menaces.

Les ingénieurs logiciels ayant une formation en développement Java et faisant la transition vers l'apprentissage automatique continuent parfois à utiliser Java comme langage de programmation dans les rôles de travail d'apprentissage automatique.

ii. Les langages les plus utilisés

- **Python**

Python est considéré comme le premier langage pour ML. La raison réside dans sa syntaxe simple. Python, créé par Guido van Rossum en 1991, est un langage de programmation open source, de haut niveau et polyvalent. Python est un langage de programmation dynamique qui

prend en charge les paradigmes de développement orientés objet, impératifs, fonctionnels et procéduraux.

Python est une recommandation pour la plupart des scientifiques des données et des développeurs ML pour une meilleure analyse et un meilleur déploiement des modèles.

De plus, il est indépendant de la plate-forme, ce qui signifie qu'il peut fonctionner sur n'importe quel système d'exploitation. Il est open-source, disponible pour tous les utilisateurs.

Python a de nombreuses bibliothèques qui facilitent les tâches telles que Teano, Keras et scikit-learn qui sont disponibles pour l'apprentissage automatique, l'apprentissage en profondeur, l'intelligence artificielle, la PNL, etc. Par exemple : Numpy est une bibliothèque qui aide à résoudre de nombreux calculs et Pybrain est à utiliser apprentissage automatique en python.

- C ++

Le C ++ fait partie des langages de programmation les plus anciens et les plus populaires.

C ++ est totalement basé sur le concept oops et prend en charge les fonctionnalités des langages de programmation bas et haut niveau, permettant ainsi à un utilisateur d'avoir un meilleur contrôle, augmentant ainsi son efficacité dans le contexte ML.

De plus, il peut exécuter du code à une vitesse beaucoup plus élevée. Cependant, le seul défaut est qu'un gros effort est nécessaire pour écrire du code complexe pour développer de nouvelles applications en ML.

De nombreuses bibliothèques puissantes telles que TensorFlow et Torch sont implémentées en langage de programmation C ++, donc l'apprentissage automatique et C ++ sont vraiment une excellente combinaison.

Fonctionnalités de C ++ :

Indépendant de la machine ou portable.

Langage de programmation de niveau intermédiaire.

Langage de programmation structuré.

Gestion de la mémoire.

Des pointeurs.

- **Le langage R**

Le langage R est un langage de programmation informatique dynamique, basé sur un tableau, orienté objet, impératif, fonctionnel, procédural et réflexif. Il a été développé en 1993.

R est un langage graphique utilisé pour le calcul statistique, l'analyse et les visualisations dans l'apprentissage automatique. C'est un langage graphique parfait pour explorer les données statistiques via un graphique largement utilisé par les professionnels des données de Facebook, Google, etc.

C'est un langage de programmation très populaire parmi les statisticiens, et il est également appliqué à des tâches d'apprentissage automatique telles que la régression, la classification et la formation d'arbres de décision.

R est hautement préférable dans la bio-ingénierie, la bio-informatique et les statistiques biomédicales. Les packages utilisés pour l'apprentissage automatique sont : RODBC, Gmodels, Class et TM. Il convient à des projets ponctuels tels que des rapports, des documents de recherche, qui incluent des artefacts.

- **Java**

Java est le langage de programmation le plus populaire au monde. Java a été développé par James Gosling chez Sun Microsystems qui a ensuite été racheté par Oracle. Il y a 9 millions de développeurs Java dans le monde.

Java est un autre langage de haut niveau très similaire au C ++. Java est utilisé pour implémenter une variété d'algorithmes utilisés dans le ML. C'est un langage très sûr. Il parvient également à surmonter la complexité globale de l'algorithme.

Java fournit de nombreux bons environnements tels que Weka, Knime, RapidMiner, Elka, qui effectuaient des tâches d'apprentissage automatique à l'aide d'interface utilisateur graphique.

iii. Résumé

L'apprentissage automatique est un domaine en croissance et le plus rapide de l'informatique. De nombreux langages de programmation prennent en charge les Framework et bibliothèques d'apprentissage automatique.

En guise de conclusion, le meilleur langage pour l'apprentissage automatique dépend du domaine sur lequel il va être appliqué.

5. Conclusion :

L'apprentissage automatique est l'un des domaines à la croissance la plus rapide qui a connu une croissance exponentielle dans le monde technique.

Les problèmes d'apprentissage automatique peuvent être distingués selon deux critères :

- Les données disponibles sont-elles annotées ou non ? Si c'est le cas, vous êtes confronté à un problème d'apprentissage supervisé. Sinon, vous serez obligé d'utiliser des algorithmes d'apprentissage non supervisés.
- Quel type de résultat souhaitez-vous prédire ? S'il s'agit d'un nombre (par exemple, le coût par clic d'une annonce), c'est un problème de régression. S'il s'agit d'une valeur discrète, d'une catégorie (comme le type d'animal présent sur la photo), alors c'est un problème de classification.

Il n'y a pas de meilleur langage pour l'apprentissage automatique, cela dépend de ce que vous voulez construire. Pour travailler dans ce domaine, il vous suffit d'apprendre un seul langage de programmation particulier en fonction de votre propre confort, des exigences du projet et de vos préférences. Explorez simplement certaines de ces langues les plus utilisées et choisissez l'un de vos choix, vous n'avez pas besoin de suivre les recommandations de qui que ce soit.

Chapitre 3

Orientation des étudiants

Dans ce chapitre :

1. Introduction
2. Les difficultés d'orientation pour les étudiants
3. Rôle de l'intelligence artificielle dans l'orientation
4. Conclusion

1. Introduction

L'orientation des étudiants dans les institutions publiques et le choix de leur cheminement académique ou de leur spécialisation appropriée sont importants pour que les étudiants poursuivent facilement leurs études dans leur carrière scolaire.

L'orientation de l'étudiant est un processus important et difficile car, prendre la décision concernant l'être humain est très compliqué, mais c'est un processus qui dépend principalement sur les points de l'apprenant en premier lieu et le désir au deuxième degré car si l'étudiant aime une spécialité, mais qu'il n'a pas les capacités suffisantes pour lui, il ne pourra pas suivre le rythme du programme d'études de cette spécialité.

Le choix de la filière est un moment crucial, voire décisif, dans la vie de tout étudiant. En effet, il ne s'agit pas seulement de choisir une filière, mais le plus important est de développer un plan d'étude et professionnel réfléchi et cohérent. En revanche, les principales difficultés rencontrées par ces futurs étudiants ne feront jamais avancer leurs tâches dans ce sens.

2. Les difficultés d'orientation pour les étudiants

Les raisons de ces obstacles sont le manque d'orientation et d'encadrement du système éducatif marocain au niveau scolaire, ainsi que le manque d'informations fiables et à jour sur la valeur et les perspectives des filières universitaires, des diplômes dans le domaine professionnel. Les difficultés rencontrées par les étudiants conduiront à une mauvaise orientation, elles ne pourront pas les aider à construire des projets personnels et professionnels basés sur leur enthousiasme, leur motivation et leur personnalité.

De plus, le dialogue insuffisant, ou souvent infructueux entre les étudiants, les parents, les enseignants, les conseillers d'orientation pédagogique et les administrateurs scolaires n'aident pas ces étudiants à trouver des réponses éclairées à leurs questions et hésitations.

Dans d'autres cas plus courants, cela dépend des résultats obtenus et de la position des établissements d'enseignement supérieur, en particulier des établissements d'enseignement supérieur réglementés, qui déterminent le choix de filière. Pire encore, les choix de carrière sont déterminés par l'opposition des parents à leur propre carrière.

Le choix de filière qui est affecté ou imposé par contrainte ne fera que déstabiliser et mettre plus de pression sur les étudiants. Lorsque vous savez que trouver votre propre chemin est un défi, il est raisonnable de penser que des choix et des risques inappropriés peuvent avoir un impact négatif sur votre carrière et votre vie personnelle.

De plus il faut mieux se connaître et s'informer pour mieux s'orienter, de ce point de vue, le choix de la filière représente une décision cruciale : l'étudiant doit considérer un certain nombre de paramètres, à savoir ses propres capacités, centres d'intérêt et forces, et réfléchir et se préparer avec sagesse. Cet exercice de réflexion doit être effectué tout au long du cursus afin d'entrer dans les études requises et dans le domaine professionnel le plus approprié.

3. Rôle de l'intelligence artificielle dans l'orientation

L'avenir de l'enseignement supérieur est désormais étroitement lié aux technologies de l'intelligence artificielle. Des applications innovantes continueront d'être développées et explorées, davantage de programmes et de cours incluront l'intelligence artificielle et des sujets connexes, et les programmes d'apprentissage existants permettront aux étudiants de maîtriser les compétences nécessaires pour s'adapter à l'évolution de ces technologies.

Prenons un exemple de deux étudiants des classes préparatoires à l'ENSAM Casablanca pour bien comprendre la situation :

Si on considère un étudiant « A » avec comme moyenne une note de '15' et un étudiant « B » avec une moyenne une note de '13'.

L'étudiant « A » a comme notes d'éléments :

- '12' en « langage C »
- '11' en « algorithmique »

L'étudiant « B » a comme notes d'éléments :

- '15' en « langage C »
- '17' en « algorithmique »

Les deux étudiants choisissent comme premier choix la filière 'IA&GI'. Cependant avec l'approche de l'ENSAM qui donne la priorité aux étudiants qui ont un bon classement, l'étudiant « A » va pouvoir facilement intégrer la filière 'IA&GI' même s'il a des notes inférieures que l'étudiant « B » en ce qui concerne les modules de la spécialité, en revanche ce dernier n'aura pas de chance d'intégrer la filière.

En conséquent, nous sommes devant une situation mal posée ni par rapport au premier étudiant, ni le deuxième car le résultat sera un étudiant qui va rencontrer plusieurs problèmes au niveau de ces études à cause de ses compétences non suffisantes pour le parcours de filière, et un autre qui n'aura pas l'opportunité de rejoindre l'IA&GI.

Bien que là qu'on peut constater l'utilité de l'intelligence artificielle qui va nous permettre de surmonter ces problèmes.

4. Conclusion

De multiples difficultés dans l'orientation des étudiants sont assez remarquables année après une autre. L'intelligence artificielle est toujours là comme étant la meilleure solution pour aider chaque étudiant à atteindre son objectif. Dans ce sens, une étude est bien faite pour acquérir le but. C'est ce que nous traiterons dans le chapitre suivant.

Chapitre 4

Modélisation de la solution

Dans ce chapitre :

1. Introduction
2. Objectif
3. Encadrement du modèle
4. Données utilisées
5. Algorithmes à comparer
 - i. LightGbm
 - ii. Decision Trees (Arbres de Décision)
 - iii. Naive Bayes
 - iv. Neural Networks (Réseaux de Neurones)
 - v. Random Forest Algorithme (RF)
6. Comparaison
 - Explication des résultats

1. Introduction

Apprentissage automatique se trouve au carrefour de nombreux autres domaines : intelligence artificielle, statistiques, sciences cognitives, théorie des probabilités, qui fait référence à des algorithmes informatiques qui permettent à une machine de s'évoluer grâce à un processus d'apprentissage, en se basant sur une sorte d'observations ou des données tel que des exemples ou des expériences pour remplir des tâches qu'il est difficile de les remplir par des moyens algorithmiques plus Classiques comme par exemple faire des prédictions précises ou d'extraire et d'exploiter automatiquement l'information présentée dans un jeu de données .

Nous faisons apprendre la machine afin qu'elle résolve des problèmes de même situation déjà connues et traitées par cette machine dans le passé en termes de donner une prédiction vers l'exacte réponse.

2. Objectif

L'objectif de ce travail est de créer une « Application » interactive avec l'utilisateur (étudiant), en se basant sur ses informations où l'app crée une future vision sur le choix souhaité, en donnant une prédiction nageant entre scores élevés et bas, dont l'utilisateur peut conclure si le choix est convenable ou non.

Notre tâche de modélisation a pour but de comparer de nombreux algorithmes, à travers une projection de chacun d'eux sur notre problématique à étudier. Cette comparaison va nous amener à classer nos algorithmes en algos convenables ou inconvenables, en se basant sur un ensemble d'arguments afin de trouver le meilleur algorithme entre eux, celui donnant une meilleure prédiction avec le minimum d'erreurs.

3. Encadrement du modèle

La variable à prédire est catégorique (Filière Recommandée), ce qui rend notre problème, un problème de Classification.

Le fait d'avoir un nombre de classes supérieur à deux (>2), alors la classification est multi-classe.

Après une longue recherche sur les algorithmes les plus puissants en termes de la classification multi-classe, nous tenons compte sur les suivants :

- Decision Trees
- Naive Bayesian (NB) Networks
- Neural Networks
- Random Forest Algorithm (RF)
- LightGbm

Le modèle d'apprentissage que nous travaillons sur, est basé sur la classification tant que notre variable à prédire est catégorique, et avec la contrainte qu'il existe un nombre de catégorie > 2 , alors que le travail consiste bien une classification multi-classe.

4. Données utilisées

En gros, notre modèle nécessite un nombre k de données (entrée + sortie)

Données nécessaires :

- Notes des étudiants (Lauréats) durant les API
- Notes des étudiants (Lauréats) durant le cycle d'ingénieur
- Filière choisie (étudiée) au cycle
- La variable à prédire est la « Filière recommandée »

5. Algorithmes à comparer

i. LightGbm

LightGBM est une implémentation d'arbres de décision par gradient boosté (GBDT) qui est une méthode d'ensemble qui combine les arbres de décision de manière sérielle, c'est-à-dire plutôt que d'utiliser un seul modèle, nous en utilisons **plusieurs** que nous **agrégeons** ensuite pour obtenir un seul résultat

Les arbres de décision sont combinés de manière à ce que chaque nouvel apprenant ajuste les résidus de l'arbre précédent afin que le modèle s'améliore, c'est comme s'il commence par construire un premier modèle qu'il va évaluer. Le modèle final regroupe les résultats de chaque étape et un apprenant fort est atteint.

Que les règles d'algorithme sont basées sur la syntaxe suivante :

- Si condition1 et condition2 et condition3 Alors résultat

Une raison mal posée est que le GBM léger est sensible au surajustement et peut facilement surajuster de petites données. Il n'y a pas de seuil sur le nombre de lignes, mais les expériences suggèrent de l'utiliser uniquement pour les données de plus de 10000 lignes.

ii. Decision Trees (Arbres de Décision)

L'arbre de décision (DT) est À travers une arborescence qui classe les instances en fonction des instances Sur la valeur caractéristique. Chaque nœud de l'arbre de décision Représente l'entité dans l'instance à classer, Chaque nœud de représentant de succursale peut penser. Les instances sont triées à partir de la racine Nœuds et trie-les en fonction de leurs valeurs caractéristiques. Apprentissage des arbres de décision pour l'exploration de données et Apprentissage automatique, utilisez comme Le modèle prédictif mappe les observations sur L'élément de la conclusion sur la valeur cible de l'élément.

Des noms plus descriptifs pour ces modèles d'arbres sont Arbre de classification ou décision d'arbre de régression la prise de décision Les classificateurs d'arbres sont généralement élagués après utilisation Techniques d'évaluation de la performance décisionnelle Arbres car ils sont élagués à l'aide du jeu de validation. Vous pouvez supprimer n'importe quel nœud et lui attribuer le plus de nœuds Institution de formation générale Trié.

iii. Naive Bayes

Le classifieur Naive Bayes est un algorithme populaire, simple et puissant pour la classification.

Pour comprendre ce classifieur, on va essayer de comprendre le théorème de Bayes que le classifieur s'est basé.

Le théorème de Bayes fonctionne sur la probabilité conditionnelle, autrement dit, quelle est la probabilité qu'un événement se produise sachant qu'un autre événement s'est déjà produit. C'est à dire qu'on peut calculer la probabilité d'un événement en utilisant ses connaissances antérieures.

iv. Neural Networks (Réseaux de Neurones)

Neural Un réseau capable d'exécuter efficacement plusieurs tâches (NN) Tâches de régression et / ou de classification, Bien que généralement chaque réseau ne peut exécuter qu'Un genre. Dans la plupart des instances, Le réseau aura une seule variable de sortie, bien que En cas de problèmes de classification dans plusieurs états, Cela peut correspondre à plusieurs unités de sortie (L'étape de post-traitement est responsable de la cartographie De l'unité de sortie à la variable de sortie). Le réseau neuronal (ANN) dépend de trois Aspects de base, fonctions d'entrée et d'activation Unité, architecture réseau et Chaque entrée est connectée.

Depuis les deux premiers L'aspect est fixe, le comportement du réseau neuronal artificiel est Défini par la valeur actuelle du poids. Ce Le poids du filet à former initialement est réglé sur Valeur aléatoire, puis une instance de l'ensemble d'apprentissage Exposition répétée à Internet. Valeurs L'entrée de l'instance est placée sur l'unité d'entrée, et Comparez le bénéfice net à la valeur attendue La sortie de cette instance. Et puis tous les poids Le filet est légèrement ajusté dans la direction Harmoniser la valeur de sortie du réseau La valeur de la sortie souhaitée. Il y a beaucoup de Algorithmes qui peuvent entraîner le réseau.

v. Random Forest Algorithme (RF)

RF est un package d'arbre de décision qui peut limiter aléatoirement les fonctions utilisées par chaque département, Travaillez en créant plusieurs arbres de décision pendant le processus de formation. Il n'y a pas de corrélation entre les deux L'arbre de décision dans RF, une fois le numéro de l'arbre généré, la classe de décision finale est basée sur le vote Lorsqu'un nouvel échantillon arrive, chaque arbre de décision RF déterminera quelle catégorie L'échantillon appartient à.

La RF est considérée comme l'un des meilleurs algorithmes qui ne sont pas sensibles à la multi colinéarité. Le résultat est relativement fiable pour les données perdues et les données déséquilibrées. De plus, il pourrait Prédire les effets de milliers de variables explicatives.

6. Comparaison

L'étude s'est passée en plusieurs étapes :

- ➔ La collecte et le tri des données nécessaires
- ➔ La division des données en deux sous-groupes de données (Entraînement et test)

- ➔ Création du modèle d'apprentissage adapté à chaque algorithme
- ➔ Tester la performance de chaque modèle

Les mesures résultantes de performance sont comme suite :

Algorithme	Précision	Temps d'Exécution (s)
Decision Tree	0.7900	2.0758
Naive Bayes	0.6828	1.8071
Neural Networks	0.7914	118.7610
Light Gbm	0.7947	1.8071
Random Forest	0.8105	0.8949

▪ Explication des résultats :

La prédiction de 'Decision tree' a donné 79% des résultats sont correctement prédis or 21% sont incorrectement prédis, avec un temps d'exécution de presque 2.0758 secondes.

Pour 'Naive Bayes', 68% correctement prédis et 32% mal prédis, avec un temps d'exécution de 1,80 seconde.

'Neural Networks', 79% correctement prédis vs 21% mal prédits. Le score de prédiction est bon mais si on constate le temps d'exécution est assez le plus grand par rapport aux autres algorithmes, ce qui diminue la performance de cet algorithme.

'Light Gbm', avec 79% des données bien prédites et 21% mal prédites, ainsi qu'un temps d'exécution court, il s'agit de 1,8071 seconde. NB que cet algorithme plus la taille des données s'augmente plus sa performance augmente.

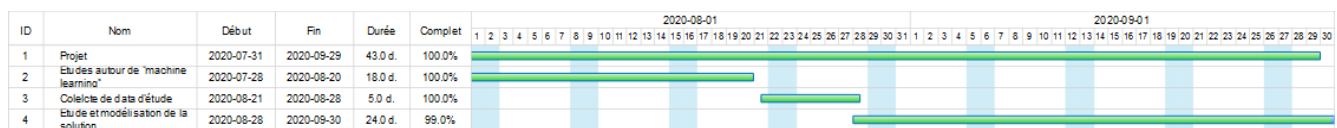
La valeur supérieure reste à l'algorithme 'Random Forest' avec une précision de '0.8105' qu'on peut l'expliquer que 81% des résultats sont correctement prédis, outre que 19% des résultats sont incorrectement prédis.

Comme on peut constater à partir des résultats de mesures que l'étude a permis de bien définir les performances de chaque algorithme. Les mesures concernant les précisions sont globalement d'une bonne valeur et proches les unes aux autres, et s'est aussi concernant le temps d'exécution sauf le NN (Neural Networks) a pris une longue durée durant son fonctionnement.

Le diagramme de Gantt

Le diagramme de Gantt est un outil de planification et de gestion de projet, qui peut être utilisé pour visualiser diverses tâches qui composent un projet au fil du temps et représenter graphiquement son avancement

Il va nous servir pour présenter le plan de notre travail avec une organisation meilleure, dans le temps accordé à chaque tâche exécutée comme le montre la figure :



Conclusion générale

A l'issue de ce stage de fin d'études on a pu toucher deux aspects très importants, il s'agit de génie informatique et de l'intelligence artificielle à travers lesquels on a eu toutes les compétences acquises durant notre parcours universitaire jusqu'à la deuxième année.

En effet, la problématique abordée sera pleinement utile pour notre établissement l'ENSAM CASABLANCA. On a pu développer un modèle qui permet de générer une recommandation de filière pour l'étudiant

Les résultats obtenus ont pu montrer l'efficacité de l'intelligence artificielle pour solver ce genre de problèmes bien entendu chaque modèle a des limites et des perspectives.

En ce qui concerne notre contribution, on envisage l'améliorer en intégrant de nouvelles idées.

Webographie

- <http://ensam-casa.ma/>
- <https://www.9rayti.com/article/difficultes-orientation-eleves-marocains>

Résumé

Que pensez-vous de l'orientation ?

L'orientation comme étant une importante étape dans la vie de l'étudiant, ayant des difficultés de choix et de décision, voilà que finalement une nouvelle idée a eu le jour.

L'intelligence artificielle a eu sa vision sur l'orientation pour faciliter le processus à chaque étudiant. En rassemblant toute idée, toute information et toute tendance de cette technologie, on vous présente ce document enrichi par une étude complète du sujet.

What do you think of the orientation?

Orientation as an important stage in the life of the student, having difficulties of choice and decision, finally a new idea was born.

Artificial intelligence had its vision on the orientation to make the process easier for every student. By bringing together any idea, any information and any trend in this technology, we present to you this document enriched by a complete study of the subject.