

# **Rapport de projet fin d'année**

**5<sup>ème</sup> année**

**Ingénierie Informatique et Réseaux**

---

## **PROJET DE FIN D'ANNEE : ETUDE DE FAISABILITE D'UN PROJET -PROJECT ADVISOR-**

---

**Réalisé par :**

Asmae MOUBARRIZ

Achraf AKRACHE

**Encadré par :**

Docteur Salma Chrit

ANNEE UNIVERSITAIRE: 2024-2025



# Table des matières

Table des figures .....	5
Remerciement .....	7
Introduction .....	8
Chapitre I. Présentation du projet.....	9
A. Problématique et contexte du projet .....	9
1. Contexte du projet.....	9
2. Problématique du projet .....	9
B. Objectifs du projet .....	10
Chapitre II. Spécification Fonctionnelle.....	10
A. Système de prédiction .....	10
Module 1 : Gestion des entrepreneurs .....	10
1. Prédiction du profil entrepreneur.....	10
Module 2 : Gestion des projets .....	11
2. Prédiction du profit d'un projet .....	11
3. Prédiction du classement parmi le top 500 .....	11
B. Système de gestion.....	12
Module 1 : Gestion des entrepreneurs .....	12
1. Créer un entrepreneur .....	12
2. S'authentifier .....	12
Module 2 : Gestion des projets .....	12
3. Editer un projet.....	12
4. Supprimer un projet.....	12
Chapitre III. Spécification non-fonctionnelle .....	13
A. Architecture micro-services .....	13
B. Evolutivité et Adaptabilité.....	13
C. Maintenabilité .....	13
D. Gestion des charges .....	14

E. Précision et fiabilité des modèles de prédiction.....	14
Chapitre IV. Spécification Technique des modèles de prédiction .....	15
A. Jeux de données .....	15
B. Méthodes de pré-traitement utilisées .....	16
1. Transformation .....	16
2. Feature Engineering.....	18
C. Modèles d'entraînement .....	19
1. Random Forest Classifier.....	19
2. KNeighborsClassifier .....	20
D. Critères de validation .....	21
1. Accuracy Score .....	21
2. Classification Report .....	22
3. Cross Validation .....	23
Chapitre V. Spécification Technique du système .....	24
A. Choix de technologie .....	24
Spring Boot .....	24
Flask.....	24
Thymeleaf .....	25
PostgreSQL .....	25
Job lib .....	26
Scikit-learn .....	26
B. Outils de collaboration .....	26
GitHub.....	26
Chapitre VI. Conception .....	27
A. Acteurs .....	27
B. Diagramme de cas d'utilisation .....	28
1. Gestion des entrepreneurs.....	29
2. Gestion des projets .....	30
Chapitre VII. Implémentation .....	31
A. Architecture du Système Project Advisor.....	31

Flask API .....	31
Spring boot API.....	32
Base de données PostgreSQL.....	32
B. Module 1 : Gestion des entrepreneurs .....	33
1. Processus de détection .....	33
2. Composants clés du modèle .....	34
3. Mapping du système.....	34
C. Module 2 : Gestion des Projets .....	35
1. Processus de détection .....	35
2. Composants clés du système .....	37
3. Mapping du système.....	38
Annexes (UIs) .....	39
Conclusion .....	42
Les jeux de données .....	43
Bibliographie.....	43

## Table des figures

Tableau 1 Dimensions des jeux de données .....	15
Tableau 2 Colonnes renommées .....	16
Tableau 3 Valeurs des colonnes transformés.....	17
Tableau 4 Colonnes encodées .....	18
Tableau 5 Accuracy Scores .....	22
Tableau 6 Classification Reports .....	23
Tableau 7 Validation croisée.....	24
Tableau 8 Composants clés pour la prédiction du profil.....	34
Tableau 9 Mapping de la prédiction du profil entrepreneur .....	34
Tableau 10 Composants clés pour la prédiction de profit.....	37
Tableau 11 Composants clés pour la prédiction du classement (top 500) .....	37

Tableau 12 Mapping de la prédiction du profit .....	38
Tableau 13 Mapping de la prédiction du classement (Top 500) .....	38
Figure 1 Représentation visuelle du Random Forest Classifier .....	20
Figure 2 KNeighbors Classifier Exemple .....	21
Figure 3 Diagramme des acteurs .....	27
Figure 4 Diagramme de cas d'utilisation générale .....	28
Figure 5 Diagramme de cas d'utilisation pour un visiteur.....	29
Figure 6 Diagramme de cas d'utilisation pour un entrepreneur .....	30
Figure 7 Diagramme d'architecture.....	31
Figure 8 Processus de détection de profil.....	33
Figure 9 Processus de détection de profit .....	35
Figure 10 Processus de détection de classement (Top500) .....	36
Figure 11 Formulaire de connexion .....	39
Figure 12 Formulaire de prédiction de profil .....	39
Figure 13 Formulaire de prédiction de profil (suite) .....	40
Figure 14 Formulaire de création de compte .....	40
Figure 15 Dashboard de gestion et étude de faisabilité des projets.....	41
Figure 16 Affichage du détail d'un projet .....	41

## Remerciement

Nous tenons à exprimer notre profonde gratitude à notre encadrante, Salma Chrit, pour son aide précieux, sa disponibilité et sa motivation constante qui ont grandement contribué au développement de ce projet. Son expertise et ses conseils avisés nous ont permis de surmonter de nombreux défis techniques et méthodologies.

Le développement d'une application « Project Advisor » s'est avéré particulièrement difficile en raison du manque de ressources, et son soutien ainsi que ses réponses à nos questions ont été d'une aide inestimable. Sa volonté d'approfondir les concepts liés au Machine Learning a été une source d'inspiration pour nous.

Nous souhaitons également exprimer notre gratitude envers notre école l'Ecole Marocaine des Sciences d'Ingénieur, qui nous a offert le cadre et les ressources nécessaires pour mener à bien ce projet.

Enfin, nous remercions tout particulièrement notre binôme, avec qui nous avons partagé cette aventure. La collaboration, l'échange d'idées et la complémentarité de nos compétences ont été un appui pour mener à bien ce projet.

Ce projet a renforcé notre esprit d'équipe et notre capacité à relever ensemble des défis complexes.

Merci à toutes les personnes qui, de près ou de loin, ont contribué à faire de ce projet une expérience enrichissante, formatrice et amusante.

## Introduction

Ce projet a été initié à la suite d'une demande de notre école dans le cadre du projet de fin d'année (PFA). Cependant, au-delà de l'obligation académique, il a également été motivé par le désir d'améliorer nos compétences techniques et de mettre en pratique les concepts appris en cours. En effet, l'un des objectifs principaux était de renforcer nos compétences en gestion des données, en couvrant tout le cycle de vie des données, de la collecte à l'analyse. De plus le projet nous a permis d'implémenter concrètement les notions vues en Data Mining et en Deep Learning, en utilisant des algorithmes avancés pour l'analyse des données et en appliquant des modèles de prédiction pour fournir des recommandations pertinentes.

En outre, l'application Project Advisor a été conçue pour évaluer la faisabilité des projets en fournissant des prédictions basées sur des données fiables et concrètes. Elle se positionne ainsi comme un outil d'aide à la décision permettant aux entrepreneurs d'avoir une vision plus claire sur leur projet et de prendre des décisions stratégiques et éclairées. En analysant divers critères de faisabilité tels que le calcul de profit, la pertinence du profil investisseur ainsi que la pertinence du projet. Project Advisor aide à minimiser les risques et à optimiser l'allocation des ressources, tout en augmentant les chances de réussite des projets.

Enfin, le projet a été pensé pour être évolutif et adaptable aux besoins du marché. Project Advisor peut être amélioré en ajoutant d'autres critères de faisabilité afin de répondre aux exigences spécifiques de différents secteurs d'activité, tels que la finance, l'industrie ou les services. L'ambition à long terme est de faire de cette application une plateforme complète d'aide à la décision, capable de s'adapter à un large éventail de projets et de secteurs.



# Chapitre I. Présentation du projet

## A. Problématique et contexte du projet

### 1. Contexte du projet

Notre projet est une application nommée Project Advisor qui vise à aider les startups et les projets en cours d'étude ainsi que les fournisseurs, les entrepreneurs et les décideurs à évaluer la faisabilité de leurs projets.

Elle se base sur une base de données robuste, qui a été collecter, prétraiter et analyser pour extraire les informations cruciales nécessaires à la prise de décision à travers l'entraînement des modèles de prédiction.

### 2. Problématique du projet

Lors de la planification et de la mise en œuvre de projets, de nombreuses entreprises font face à des défis pour évaluer la faisabilité des projets. Cela peut être dû à un manque de données fiables, à des analyses subjectives ou à des processus décisionnels inefficaces. Ces obstacles peuvent entraîner des pertes financières, un gaspillage de ressources et l'échec du projet.

#### **Solution :**

Notre application Project Advisor utilise des modèles robustes entraînée depuis une base de données prétraitée et normalisée pour obtenir les informations critiques nécessaires à l'évaluation de la faisabilité des projets.

En s'appuyant sur les résultats prédits par ces modèles l'application fournit des recommandations objectives et basées sur des données pour aider les décideurs à prendre des décisions éclairées et stratégiques.

## B. Objectifs du projet

L'objectif principal de l'application Project Advisor est de fournir une évaluation objective et précise de la faisabilité des projets en utilisant des données fiables et des modèles de prédiction avancés. Cela permet aux décideurs de minimiser les risques, d'optimiser l'allocation des ressources et d'augmenter les chances de réussite du projet.

## Chapitre II. Spécification Fonctionnelle

Pour garantir l'efficacité du développement et obtenir un résultat optimal, nous avons adopté une méthodologie de travail basée sur la modularisation. Cela consiste à diviser le fonctionnement de notre application en modules distincts.

### A. Système de prédiction

#### Module 1 : Gestion des entrepreneurs

##### 1. Prédiction du profil entrepreneur

Cette prédiction aide un individu visiteur à savoir son potentiel d'être un bon entrepreneur. Cela à travers une prédiction qui détermine si cet individu possède le profil entrepreneur ou non.

Le modèle utilisé dans cette prédiction a été baser sur des critères concrète comme l'âge, le degré de persévérance, le degré de confiance en soi, la présence d'un trouble mentale chez l'individu, et d'autre critères que nous allons détailler davantage tout au long du reste du rapport.

## Module 2 : Gestion des projets

### 2. Prédiction du profit d'un projet

Cette prédiction évalue le degré de rentabilité d'un projet, les résultats de cette prédiction varie entre un profit Très bon jusqu'à un profit Très faible. Ces résultats sont obtenus en utilisant les montants investit par l'entrepreneur dans les principales secteurs et départements que chaque entreprise peu importe sa taille doit avoir. Je cite le département de recherche et développement qui garantit un long cycle de vie au projet entreprise, le département de Marketing qui touche directement à la réputation de l'entreprise, et le département d'administration qui garantit le bon fonctionnement de l'entreprise, en surveillant le respect de la hiérarchie administrative.

### 3. Prédiction du classement parmi le top 500

Cette fonctionnalité prédit si le projet figurera parmi les 500 meilleurs dans la région ou il nécessite plus de travail et de paramétrages pour y parvenir.

Cette prédiction comme toute les autres elle se basent sur des critères réelle, concrète, et convaincante sur le projet. Je cite en guise d'exemple le nombre de relations que le projet aura, la présence ou l'absence d'un investisseur providentiels (Business Angel), le nombre de participants dans ce projet, ainsi que d'autres critères qui seront détaillés tout au long du reste du rapport.

## B. Système de gestion

### Module 1 : Gestion des entrepreneurs

#### 1. Créer un entrepreneur

La création de l'objet entrepreneur est géré par un système de gestion puissant qui communique avec une base de données d'une manière sécurisée. Le but de cette création est de rassembler le maximum d'information sur l'entrepreneur afin de faciliter la tâche pour le système de prédiction par la suite.

#### 2. S'authentifier

Cette fonctionnalité permet à un visiteur de se connecter au système, et changer son droit d'accès de visiteur à un entrepreneur. Après l'authentification, le visiteur maintenant entrepreneur, accède à un Dashboard administratif qui fournit des fonctionnalités d'édition de projet, de suppression de projet, de consultation et prédictions des indicateurs d'étude de faisabilité offerte par le système de prédiction.

### Module 2 : Gestion des projets

#### 3. Editer un projet

L'édition du projet est offerte par le système à seul les entrepreneurs inscrits et authentifiés. L'édition est présentée sous forme de formulaire vide en cas de création de projet, ou remplie en cas de modification de projet déjà existant.

#### 4. Supprimer un projet

La fonctionnalité de suppression du projet, de même, elle est offerte par le système à seul les entrepreneurs inscrits et authentifiés via le Dashboard de gestion de projet.

## Chapitre III. Spécification non-fonctionnelle

### A. Architecture micro-services

Le système est conçu selon une approche modulaire où chaque fonctionnalité est gérée par un service distinct. Cette séparation permet d'assurer une meilleure organisation du code et de faciliter les mises à jour ou l'ajout de nouvelles fonctionnalités sans affecter l'ensemble de l'application. Par exemple, le moteur de prédiction fonctionne indépendamment du système de gestion des entrepreneurs et des projets, garantissant une exécution fluide et efficace des différentes tâches de prédiction et de stockage.

### B. Evolutivité et Adaptabilité

L'architecture du système est conçue pour permettre une intégration fluide de nouvelles fonctionnalités sans perturber les services existants. Grâce à l'indépendance du moteur de prédiction, il est possible d'ajouter de nouveaux critères d'évaluation ou d'améliorer les modèles sans impacter la gestion des entrepreneurs et des projets. De plus, les prédictions sont fournies sous forme de données JSON standardisées, garantissant une compatibilité avec n'importe quel Framework ou technologie backend. Cette approche permet ainsi de modifier ou remplacer les services de gestion en fonction des besoins sans besoin d'une refonte complète du système de prédiction, assurant ainsi une évolutivité et une adaptabilité satisfaisantes.

### C. Maintenabilité

La conception du système permet une maintenance efficace grâce à une séparation claire des différentes couches applicatives. Les composants sont documentés afin de faciliter leur compréhension et leur modification.

## D. Gestion des charges

Le système est conçu pour assurer une exécution rapide des différentes opérations, je cite la création des comptes entrepreneur ou des projets et l'analyse des projet et leur faisabilité. Grâce à des mécanismes de traitement asynchrones et à une gestion optimisée des ressources, les prédictions et la gestion des données s'effectuent avec un temps de réponse réduit. Cette approche garantit une expérience fluide et réactive pour les utilisateurs.

## E. Précision et fiabilité des modèles de prédiction

Les modèles de prédiction mis en place garantir un haut niveau de précision afin d'assurer des résultats fiables et pertinents. Pour cela, ils sont entraînés sur des ensembles de données représentatifs et compatibles aux tendances entrepreneuriales et économiques. Des techniques de validation croisées et de tests sur des jeux de données diversifiés permettent d'évaluer leur performance et d'éviter les biais.

## Chapitre IV. Spécification Technique des modèles de prédiction

### A. Jeux de données

Nos trois jeux de données ont été sélectionnés à partir de la plateforme Kaggle. Notre choix de fournisseur de Dataset nous a offert une diversification, une bonne qualité de données et une compatibilité aux différentes problématiques de classification liées à l'entrepreneuriat et à la rentabilité des projets.

Ces données ont été soigneusement choisies afin de garantir une représentativité adéquate pour l'entraînement et l'évaluation de nos modèles de prédiction. Chaque Dataset a été prétraité et nettoyé pour assurer une meilleure cohérence et fiabilité des résultats.

MODELE DE PREDICTION	DIMENSION	
	Initiale	Après prétraitement
PROFIL	(219,17)	(219, 16)
PROFIT	(50,5)	(50,5)
TOP500	(923,49)	(923, 8)

*Tableau 1 Dimensions des jeux de données*

## B. Méthodes de pré-traitement utilisées

### 1. Transformation

#### Indices des colonnes

Pour résoudre les lacunes des structures des données de nos jeu de données. Nous avons utilisé une stratégie de transformation et de traduction des colonnes qui consiste à **renommer les indices de toutes les colonnes** pour faciliter l'utilisation et la compréhension des variables du fichier csv. Le tableau ci-dessous montre les différentes colonnes qui ont été renommé durant ce processus.

MODELE DE PREDICTION	COLONNES TRANSFORMEES
<b>PROFIL</b>	Education, Projet_Individuel, Genre , EnVille , Influence , DegPerseverance, DegInitiative , DegCompetition , DegAutonomie, DegSante, DegConfiance, DegBesoinReussite, TraitsCles, TroubleMental
<b>PROFIT</b>	R&D, Marketing, Region
<b>TOP 500</b>	Ville, Nombre_relations, Capitale_fonds, top500 Categorie, Tours_financement, Investisseurs_providentiels , Nombre_participants

*Tableau 2 Colonnes renommées*



## Valeurs des variables

Pour assurer une meilleure utilisabilité du modèle à travers des formulaires, nous avons opté pour la conversion des colonnes numériques en colonnes qualitatives. Cette transformation permet de regrouper les valeurs numériques en classes cohérentes et compatibles avec la logique des données. Ainsi, au lieu d'utiliser des valeurs continues ou discrètes difficiles à interpréter pour l'utilisateur, nous avons défini des catégories significatives qui facilitent la saisie et l'analyse des informations. Cette approche améliore l'accessibilité du modèle tout en conservant la pertinence des données pour les prédictions.

MODELE	COLONNES	VALEURS
PROFIL	Education	Sciences de l'ingénierie, Autres, Sciences économiques, études commerciales, commerce et droit, Art, musique ou design...
	Traits clés	Passion, Vision, Resilience, Positivity, Work Ethic.
PROFIT	Région	Rabat, Casablanca, Tanger
TOP 500	Catégorie	music, entreprise, web, software, games_video, network_hosting, finance, mobile, education, public_relations, security...
	Villes	Casablanca, Rabat, Marrakech, Fez, Tanger, Agadir, Mohammedia, Beni Mellal, Khenifra, Kenitra, Tetouan ...

Tableau 3 Valeurs des colonnes transformés

## Nulles et doublons

Pour garantir la fiabilité des données, nous avons effectué un nettoyage en supprimant les valeurs nulles et les doublons. Les valeurs nulles peuvent fausser les prédictions, c'est pourquoi nous avons appliqué la méthode `dropna()`. De même, les doublons ont été supprimés à l'aide de la fonction `drop_duplicates()`.

## 2. Feature Engineering

### Label Encoding

Le label Encoding c'est d'attribuer un nombre entier unique à chaque catégorie d'une variable catégorielle. Autrement le Label Encoding est une technique utilisée pour convertir des colonnes catégorielles en colonnes numériques afin qu'elles puissent être ajustées par des modèles de Machine Learning qui ne prennent en charge que les données numériques. Il s'agit d'une étape de prétraitement importante car ces encodeurs sont stockés dans des fichiers `.pkl` et seront invoqué par la suite dans la réutilisation du modèle lors de la prédiction des résultats.

MODELE DE PREDICTION	COLONNES ENCODEES
PROFIL	Education, Enville , Genre , Influence , TraitsCles , TroubleMental, Projet_Individuel
PROFIT	Region
TOP 500	Categorie, investisseurs_providentiels , Ville

*Tableau 4 Colonnes encodées*

## C. Modèles d'entraînement

### 1. Random Forest Classifier

#### a) Fonctionnement du Random Forest

Random forest est un algorithme d'apprentissage supervisé qui est utilisé à la fois pour la classification et la régression. Cependant, il est principalement utilisé pour les problèmes de classification. Comme nous le savons, une forêt est composée d'arbres et plus il y a d'arbres, plus la forêt est robuste. De même, un algorithme de forêt aléatoire crée des arbres de décision sur des échantillons de données, puis obtient la prédiction de chacun d'eux et sélectionne finalement la meilleure solution.

En général, plus il y a d'arbres dans une forêt, plus elle paraît robuste. De la même manière, dans le classificateur de forêt aléatoire, plus le nombre d'arbres dans la forêt est élevé, plus les résultats sont précis.

Le concept fondamental de la forêt aléatoire est simple mais puissant : Un grand nombre de modèles relativement non corrélés (arbres) fonctionnant en comité surpasseront n'importe lequel des modèles constitutifs individuels. « La clé est la faible corrélation entre les modèles » les prédictions (et donc les erreurs) faites par les arbres individuels doivent avoir de faibles corrélations entre elles.

#### b) Système de vote

Pour effectuer la prédiction à l'aide de l'algorithme de forêt aléatoire formé, nous devons faire passer les caractéristiques de test par les règles de chaque arbre créé aléatoirement. Supposons que nous ayons formé 100 arbres de décision aléatoires à partir de la forêt aléatoire.

Chaque arbre aléatoire prédit des résultats différentes pour la même fonction de test. Ensuite, en considérant chaque résultat prédit, les votes seront calculés.

Supposons que les 100 arbres de décision aléatoires prédisent 3 résultats uniques x, y, z, les votes x tout seul n'ont aucune valeur, mais sur 100 arbres de décision aléatoires, combien d'arbres prédisent x. De même pour les 2 autres résultats (y, z). Si x obtient le plus de votes, la forêt aléatoire finale renvoie x comme cible prédite ou résultat prédit. Ce concept de vote est connu sous le nom de vote majoritaire.

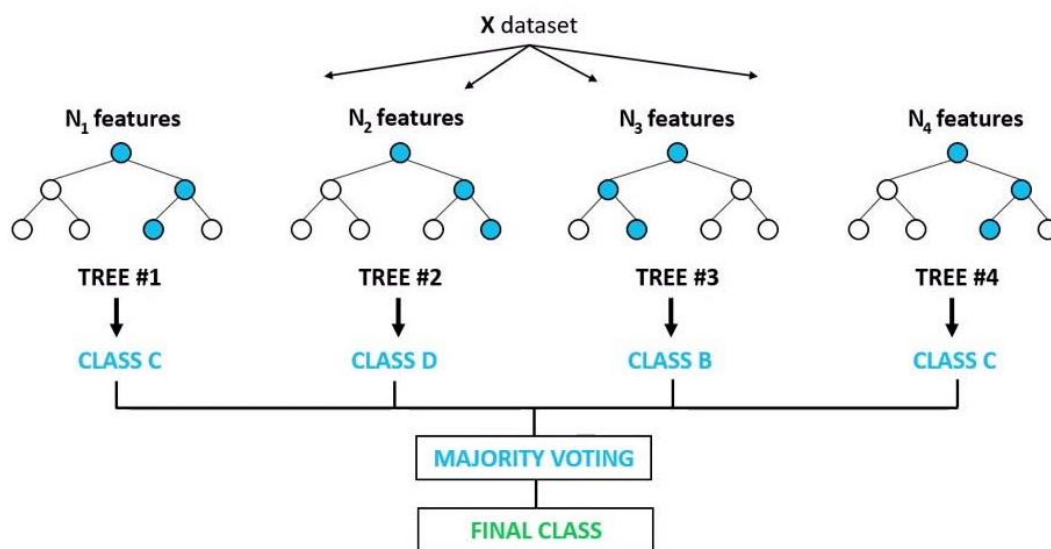


Figure 1 Représentation visuelle du Random Forest Classifier

## 2. KNeighborsClassifier

KNN, ou k-Nearest Neighbors, c'est comme avoir un ami très intelligent qui vous aide à prendre des décisions en fonction de ce que font vos voisins.

KNN se distingue par sa polyvalence et les hypothèses minimales qu'il formule sur les données. Contrairement à de nombreux modèles d'apprentissage automatique qui nécessitent des configurations et des hypothèses complexes sur la distribution des données, KNN travaille directement avec les données qui lui sont présentées, sans faire d'hypothèses sur la distribution sous-jacente. Cette caractéristique le rend adaptable à diverses tâches.

Dans KNN, lorsque nous obtenons un nouveau point de données et que nous devons décider à quel groupe il appartient, nous examinons ce point de données pour trouver les « k » voisins les plus proches. Si la plupart de ces voisins appartiennent à un groupe particulier, nous disons que le nouveau point de données appartient également à ce groupe. Le « k » dans KNN n'est qu'une façon de dire combien de voisins prendre en compte. Si « k » est 3, vous examinez les 3 voisins les plus proches pour prendre votre décision.

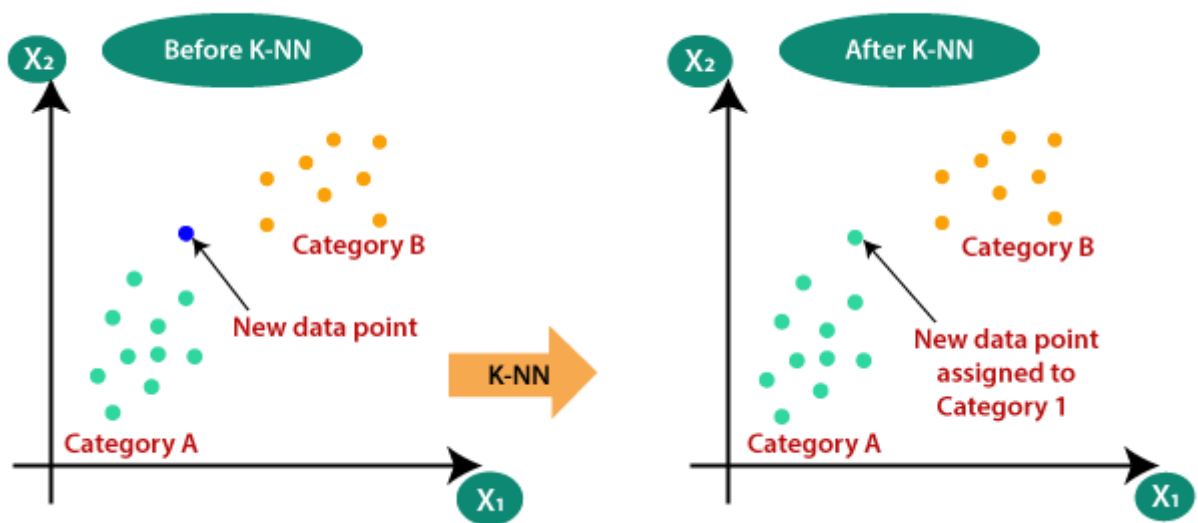


Figure 2 KNeighbors Classifier Exemple

## D. Critères de validation

### 1. Accuracy Score

L'Accuracy Score est une métrique essentielle pour évaluer la performance d'un modèle de classification. Il représente le pourcentage de prédictions correctes effectuées par le modèle par rapport au total des observations.

Une Accuracy élevée indique un bon niveau de performance, mais cette métrique peut être trompeuse lorsque les classes sont déséquilibrées. Dans ce cas, un modèle peut obtenir une haute précision en favorisant la classe majoritaire sans

bien prédire les classes minoritaires. C'est pourquoi nous avons opté à l'accompagner par le rapport de classification et la validation Croisée.

Le tableau ci-dessous montre les accuracy Scores des différents modèles entraînés par notre binôme pour prédire les trois critères de faisabilité : le profit, le profil entrepreneur et la présence parmi les Top 500.

MODELE DE PREDICTION	ACCURACY SCORE
PROFIL	0,59
PROFIT	0,90
TOP 500	0,81

*Tableau 5 Accuracy Scores*

## 2. Classification Report

Le Classification Report fournit une analyse plus détaillée pour chaque classe de la variables prédit, ceci à travers des indicateurs clés :

- ✓ La précision (*precision*) : qui mesure la proportion de prédictions correctes parmi celles effectuées pour une classe donnée
- ✓ Le rappel (*recall*) : qui évalue la capacité du modèle à identifier correctement les instances d'une classe
- ✓ Le score F1 : qui est une moyenne harmonique entre précision et rappel.

Ce rapport permet d'identifier les forces et faiblesses du modèle, notamment en cas de déséquilibre des classes.

	CLASSES	PRECISION	RECALL	F1-SCORE	SUPPORT
PROFIL	0	0.61	0.85	0.71	26
	1	0.50	0.22	0.32	18
PROFIT	Bon	0.80	1	0.89	4
	Faible	1	1	1	1
	Moyen	1	1	1	3
	Très bon	1	0,5	0,67	2
TOP 500	0	0.44	0.34	0.39	32
	1	0.87	0.97	0.89	152

Tableau 6 Classification Reports

### 3. Cross Validation

La Cross Validation est une technique statistique permettant d'évaluer la robustesse et la capacité de généralisation d'un modèle. Elle consiste à diviser l'ensemble des données en plusieurs sous-ensembles appelés *folds*, puis à entraîner et tester le modèle sur différentes partitions.

La validation croisée réduit le risque de surapprentissage (*overfitting*) en s'assurant que le modèle ne se contente pas de mémoriser un ensemble de données spécifique, mais qu'il est capable de bien généraliser sur de nouvelles observations.

MODELE DE PREDICTION	CROSS VALIDATION(MEAN)
PROFIL	0,56
PROFIT	0,77
TOP 500	0,79

*Tableau 7 Validation croisée*

## Chapitre V. Spécification Technique du système

### A. Choix de technologie

#### Spring Boot

Spring Boot est un Framework Java conçu pour simplifier le développement d'applications Spring. Il propose une configuration automatique, un serveur embarqué Tomcat et une gestion simplifiée des dépendances, permettant de créer des applications autonomes et prêtes pour la production avec un minimum de configuration. Il est souvent utilisé pour les micro-services et les applications d'entreprise.

- ⇒ Nous avons utilisé Spring boot pour créer le système de gestion.  
L'application Spring boot est chargé de créer les entrepreneurs et de gérer les projets de l'entrepreneur.

#### Flask

Flask est un Framework web léger en Python utilisé pour développer des applications web et des API. Il suit une approche minimaliste, laissant aux développeurs la liberté d'ajouter uniquement les composants nécessaires. Flask



est apprécié pour sa simplicité, sa flexibilité et son intégration facile avec des modèles de machine Learning, des bases de données et des services RESTful.

- ⇒ Nous avons utilisé cette technologie pour créer une api Flask qui expose des Endpoints pour les fonctions de prédiction. L'application Flask est chargée de fournir les valeurs des variables, ainsi de retourner les résultats de prédiction sous format JSON.

## Thymeleaf

Thymeleaf est un moteur de template moderne basé sur Java, principalement utilisé avec Spring Boot pour générer des pages web dynamiques. Il permet d'intégrer du code Java directement dans du HTML, facilitant la création d'interfaces web interactives. Thymeleaf prend en charge les formulaires, les boucles, les conditions et l'internationalisation, ce qui en fait un choix populaire pour les applications MVC côté serveur.

- ⇒ Thymeleaf est utilisée dans l'API Spring, fusionnée avec d'autres langages web comme le CSS et JS, pour fournir au visiteur ou entrepreneur une interface utilisateur conviviale, attrayante, fonctionnelle et facile à utiliser, garantissant ainsi une expérience utilisateur (UX) élevée.

## PostgreSQL

PostgreSQL est un système de gestion de bases de données relationnelles (SGBDR) open-source reconnu pour sa robustesse, son extensibilité et sa conformité aux normes SQL. Il offre des fonctionnalités avancées telles que le support des données semi-structurées avec JSONB, la recherche en texte intégral et les transactions ACID, le rendant adapté aux applications de grande envergure et à haute performance.

## Job lib

Job lib est une bibliothèque Python optimisée pour la gestion efficace de la sérialisation et du calcul parallèle. Elle est couramment utilisée pour enregistrer et charger des modèles de machine Learning (comme ceux de scikit-learn) sous forme de fichiers .pkl, ce qui permet d'éviter de recalculer des résultats coûteux.

⇒ Job lib a garantie une portabilité au différents modèles entraînée et stockée dans des fichiers .pkl. De plus cette bibliothèque a été utiliser pour stocker les différents encodeurs utilisés lors du prétraitement des données et pour récupérer ces encodeurs lors du processus de prédiction.

## Scikit-learn

Scikit-learn est une bibliothèque Python incontournable pour l'apprentissage automatique. Elle fournit un ensemble d'outils efficaces pour la classification, la régression, le clustering et la réduction de dimension. Basée sur NumPy, SciPy et Matplotlib.

⇒ Elle est également compatible avec Job lib, permettant la sauvegarde et le chargement rapide des modèles entraînés.

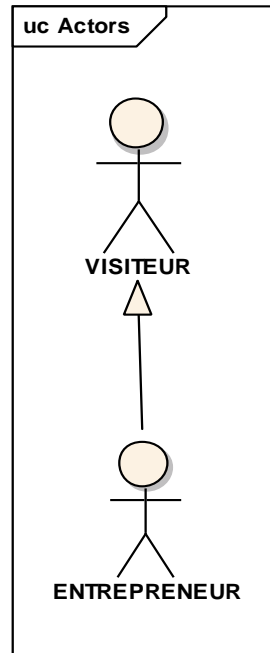
## B. Outils de collaboration

### GitHub

GitHub est une plateforme de gestion de code source basée sur Git, permet de collaborer et versionner un projet. Cet outil a facilité le travail et la collaboration entre notre binôme, grâce à ces fonctionnalités de versions comme les branches et les fonctionnalités de documentation comme le fichier readme qui sert d'une présentation générale sur notre projet publié dans le répertoire GitHub.

## Chapitre VI. Conception

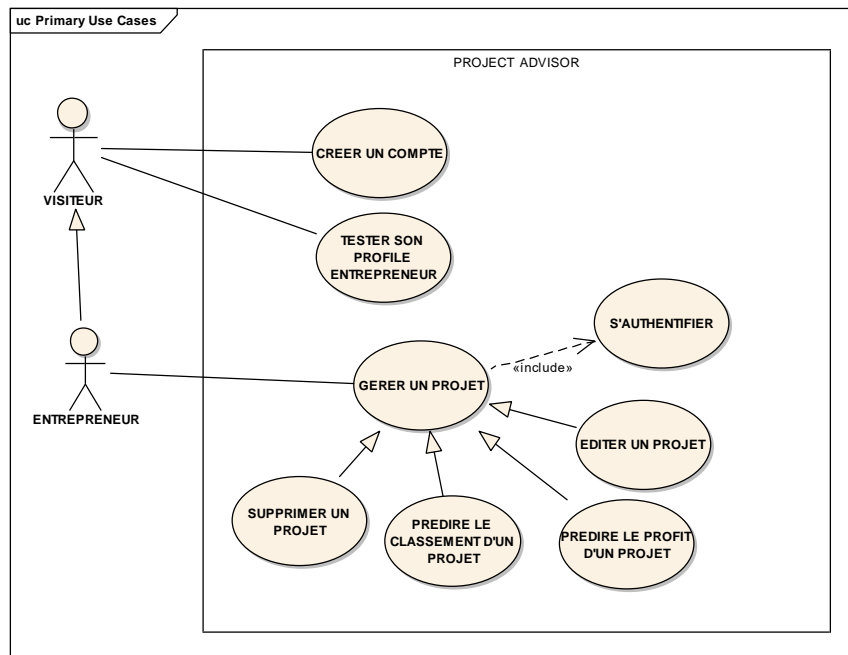
### A. Acteurs



*Figure 3 Diagramme des acteurs*

Le diagramme des acteurs montre la relation entre les différents acteurs du système. Le Visiteur représente toute personne accédant au système sans authentification et un Entrepreneur, qui hérite des fonctionnalités du visiteur tout en ayant des fonctionnalités supplémentaires.

## B. Diagramme de cas d'utilisation



*Figure 4 Diagramme de cas d'utilisation générale*

Ce diagramme représente une idée générale sur les différentes fonctionnalités offerte par notre système d'«étude de faisabilité» . Les fonctionnalités comprennent :

- ✓ Une gestion de projet pour les entrepreneurs.
- ✓ Un test de profil pour les visiteurs ainsi que les entrepreneurs.
- ✓ La possibilité de création d'un compte entrepreneur pour gérer les projets et les étudier selon les critères de faisabilité offerte.

## 1. Gestion des entrepreneurs

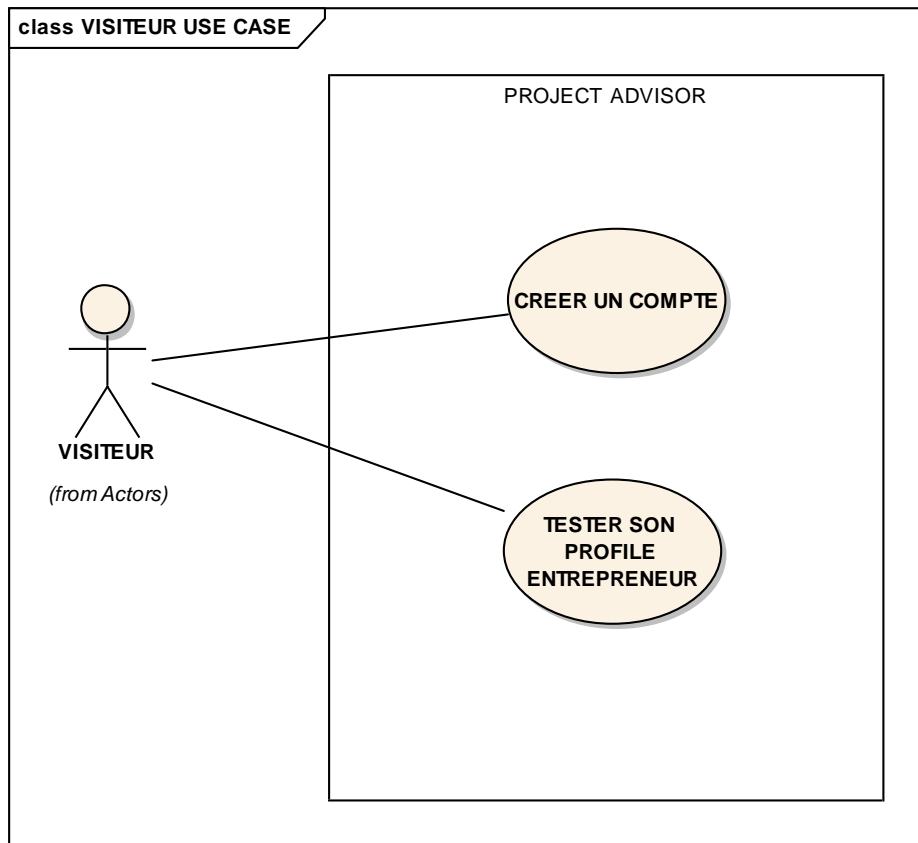


Figure 5 Diagramme de cas d'utilisation pour un visiteur

Ce diagramme montre les fonctionnalités accessibles via un privilège visiteur, ainsi qu'un privilège entrepreneur puis que l'acteur entrepreneur hérite tout fonctionnalité accessible à un visiteur.

- ✓ Créer un compte : La création d'un compte entrepreneur pour avoir l'accès fonctionnalité de gestion de projet et d'étude de faisabilité de ce projet
- ✓ Tester son profil entrepreneur : un formulaire dynamique permettant au visiteur de tester s'il a un profil entrepreneur pour un certain type de projet choisit depuis le formulaire même , ou bien il doit encore y travailler sur soie même ou sur la thématique du projet pour adopter un profil entrepreneur.

## 2. Gestion des projets

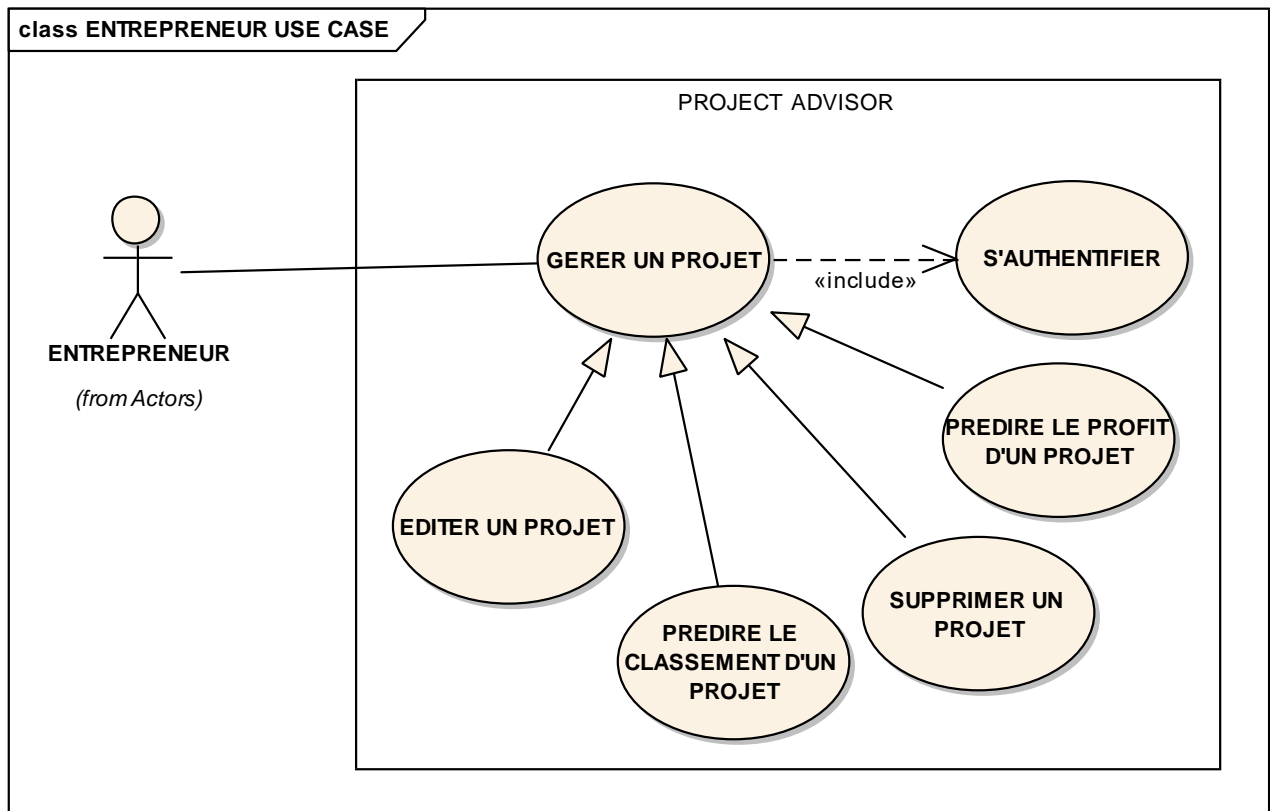


Figure 6 Diagramme de cas d'utilisation pour un entrepreneur

Ce diagramme montre les fonctionnalités accessibles via un privilège entrepreneur, ces fonctionnalités tous se résume dans la gestion de projet.

- ✓ Editer un projet : La création d'un projet ou la modification d'un projet sont faite à partir d'un formulaire qui gère l'objet projet.
- ✓ Prédire le classement d'un projet à travers son existence parmi les tops 500.
- ✓ Prédire le profit d'un projet à partir des montant investit dans ses départements.
- ✓ Suppression d'un projet.
- ✓ Authentification : cette fonctionnalité permet au visiteur de changer son privilège à un entrepreneur et accéder aux fonctionnalités de gestion de projet.

## Chapitre VII. Implémentation

### A. Architecture du Système Project Advisor

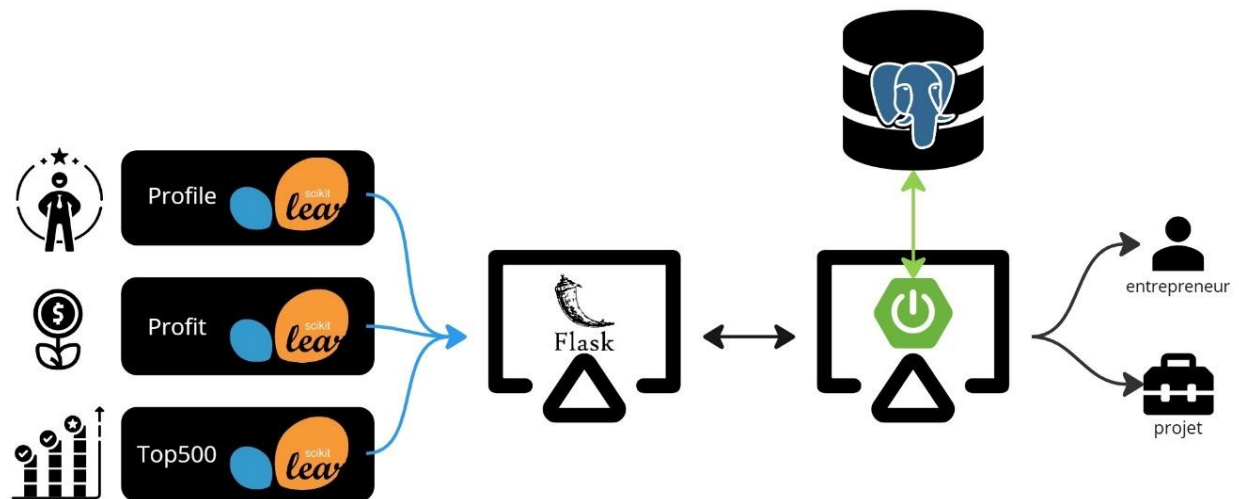


Figure 7 Diagramme d'architecture

Ce diagramme représente les différents composants du système Project Advisor , qui sont :

- ✍ Un modèle de prédiction de profil
- ✍ Un modèle de prédiction de profit
- ✍ Un modèle de prédiction du classement Top 500
- ✍ Une API Flask
- ✍ Une application Spring Boot
- ✍ Une base de données PostgreSQL

#### Flask API

- ✓ Cette API charge les modèles de prédiction stocker dans des fichiers d'extension .pkl par l'outil job lib , ainsi que les encodeurs .
- ✓ Après avoir chargé ces modèles , ils les utilisent pour prédire le profit , le profil entrepreneur ou le classement parmi le Top 500 , Selon la requête

HTTP de type POST reçu depuis l'API Spring , puis renvoie le résultat dans la réponse sous format JSON .

- ✓ L'API Flask fournit aussi les différentes classes ou valeurs des colonnes quantitatifs via des requêtes HTTP de type GET .

## Spring boot API

- ✓ Cette API gère les données des entrepreneurs et des projets .
- ✓ Il stocke les informations dans une base de données , et utilise le JPA comme ORM pour mapper et récupérer les informations depuis cette base de données .
- ✓ Il utilise la classe RestTemplate pour communiquer avec l'API Flask . Elle s'occupe d'envoyer les données concernant les projets ou les entrepreneurs sous format JSON avec une requête HTTP de type POST et récupère les réponses de prédiction renvoyer par l'API Flask .
- ✓ Lors de la construction des formulaires , l'Application Spring envoie des requêtes HTTP de type GET pour récupérer les différentes classes pour chaque critère qualitatif , et les injectes , avec le Thymeleaf , dans les pages HTML pour garantir une compatibilité entre les données utilisées pour entrainer le modèle et les données fournit par l'utilisateur finale .

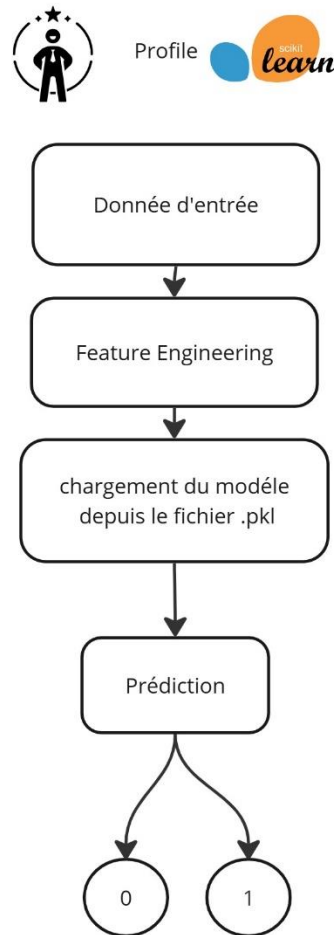
## Base de données PostgreSQL

- ✓ Le SGBD PostgreSQL s'occupe de stocker les informations sur les entrepreneurs et les projets .
- ✓ Il interagit avec le Spring boot pour stocker et récupérer ces informations nécessaires pour le bon fonctionnement du système.



## B. Module 1 : Gestion des entrepreneurs

### 1. Processus de détection



*Figure 8 Processus de détection de profil*

- ✓ Les catégories textuelles (Genre, TraitsCles, Education...) sont transformées via les Label Encoder pré-entraînés.
- ✓ Les données numériques (Age, Deg\* scores...) sont conservées telles quelles.
- ✓ Le modèle RandomForestClassifier utilise les 15 variables transformées pour prédire.

La sortie est une classe binaire (0 ou 1) correspondant à la cible 'y'.

## 2. Composants clés du modèle

<b>Fichier du modèle</b>	model/profile/profile_model.pkl
<b>Encodeurs sauvegardés</b>	model/profile/*.pkl
<b>Valeurs des colonnes</b>	model/profile/*_valeurs.pkl
<b>Fichier de jeux de données</b>	data/profile_data.csv
<b>Type de Modèle</b>	Classification avec RandomForestClassifier
<b>Pipeline</b>	Pandas + scikit-learn
<b>Hyper paramètre</b>	<div> ▼ RandomForestClassifier ⓘ ? </div> <div> RandomForestClassifier(random_state=42) </div>

Tableau 8 Composants clés pour la prédiction du profil

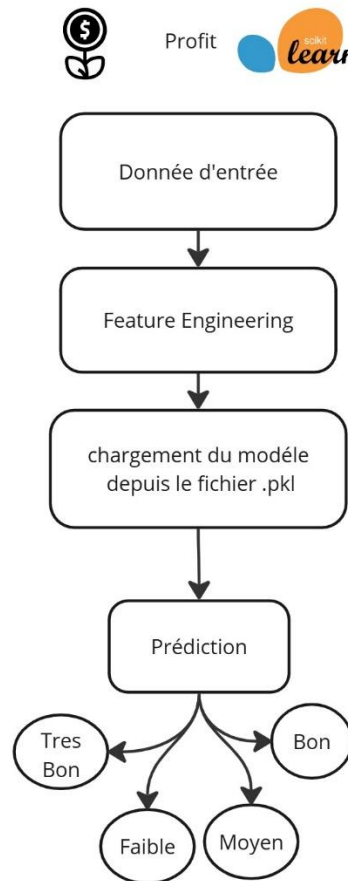
## 3. Mapping du système

Api	Type de requêtes	Mapping
Flask	GET	/profile
	POST	/profile
Spring	GET	/profile
	POST	/profile

Tableau 9 Mapping de la prédiction du profil entrepreneur

## C. Module 2 : Gestion des Projets

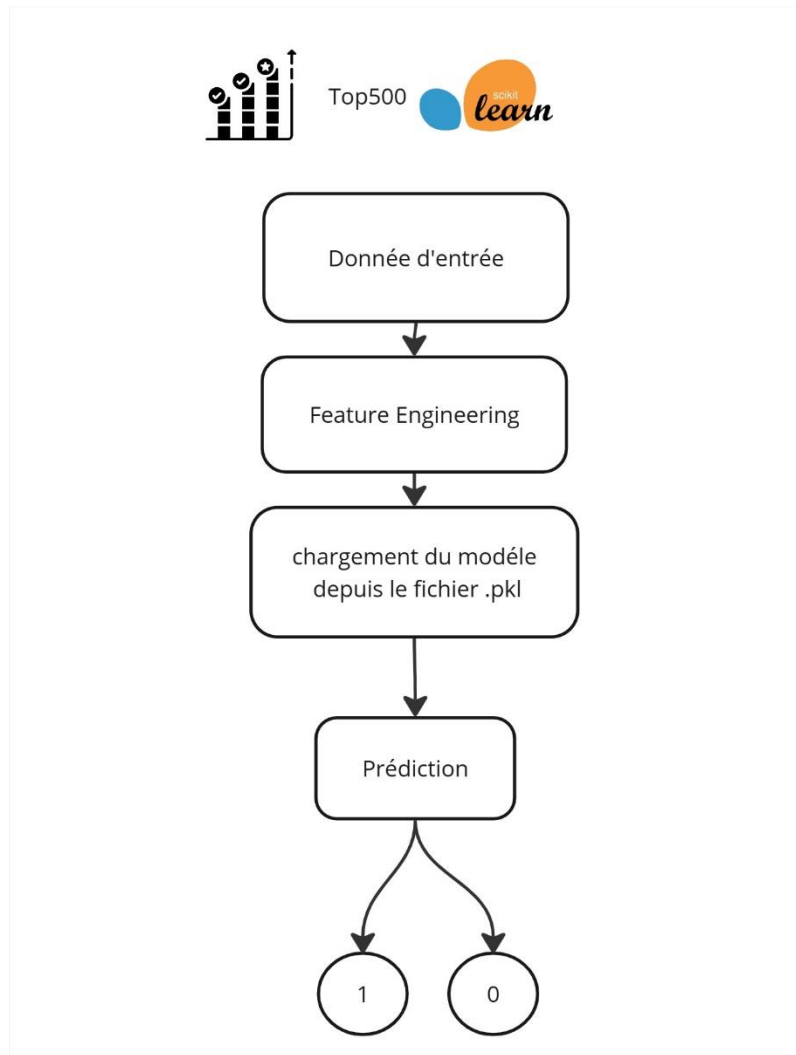
### 1. Processus de détection



*Figure 9 Processus de détection de profit*

- ✓ Les catégories textuelles (Region) sont transformées via les Label Encoder pré-entraînés.
- ✓ Les données numériques (R&D, Marketing, Administration ) sont conservées telles quelles.
- ✓ Le modèle RandomForestClassifier utilise les 4 variables transformées pour prédire.

La sortie est multi-classe (Très Bon , Faible , Moyen , Bon ) correspondant à la cible 'profit'.



*Figure 10 Processus de détection de classement (Top500)*

- ✓ Les catégories textuelles ( Ville, Catégorie ,Investisseurs providentiels(business Angel) ) sont transformées via les Label Encoder pré-entraînés.
- ✓ Les données numériques (nombre de relations ,fond des capitaux) sont conservées telles quelles.
- ✓ Le modèle KNeighborsClassifier utilise les 7 variables transformées pour prédire.

La sortie est une classe binaire (0 ou 1) correspondant à la cible 'top500'.

## 2. Composants clés du système

<b>Fichier du modèle</b>	model/profit/profit_model.pkl
<b>Encodeurs sauvegardés</b>	model/profit/*.pkl
<b>Valeurs des colonnes</b>	model/profit/*_valeurs.pkl
<b>Fichier de jeux de données</b>	data/profit_data.csv
<b>Type de Modèle</b>	Classification avec RandomForestClassifier
<b>Pipeline</b>	Pandas + scikit-learn
<b>Hyper paramètre</b>	<div> ▼ RandomForestClassifier ⓘ ? </div> <div> RandomForestClassifier(n_estimators=20) </div>

*Tableau 10 Composants clés pour la prédiction de profit*

<b>Fichier du modèle</b>	model/top500/top500_model.pkl
<b>Encodeurs sauvegardés</b>	model/top500/*.pkl
<b>Valeurs des colonnes</b>	model/ttop500/*_valeurs.pkl
<b>Fichier de jeux de données</b>	data/top500_data.csv
<b>Type de Modèle</b>	Classification avec KNeighborsClassifier
<b>Pipeline</b>	Pandas + scikit-learn
<b>Hyper paramètre</b>	<div> ▼ KNeighborsClassifier ⓘ ? </div> <div> KNeighborsClassifier(n_neighbors=3) </div>

*Tableau 11 Composants clés pour la prédiction du classement (top 500)*

### 3. Mapping du système

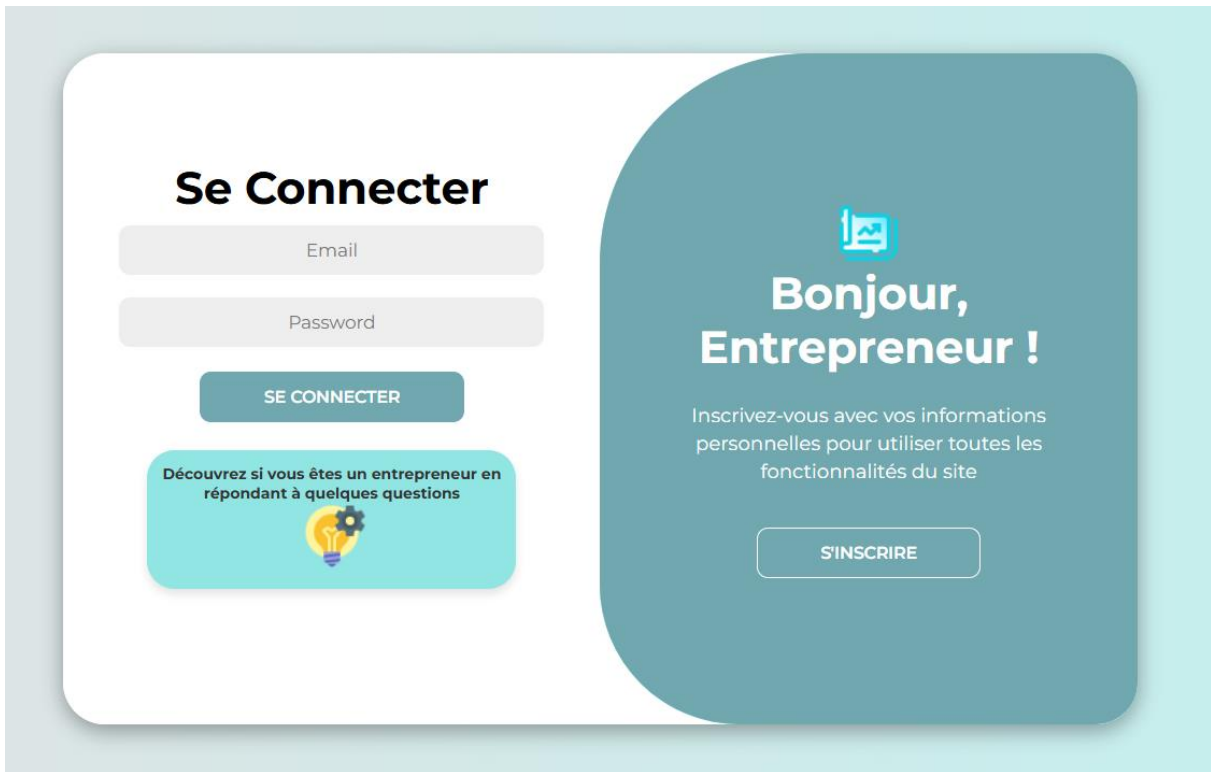
Api	Type de requêtes	Mapping
Flask	GET	/profit
	POST	/profit
Spring	GET	/profit
	POST	/profit

*Tableau 12 Mapping de la prédiction du profit*

Api	Type de requêtes	Mapping
Flask	GET	/top500
	POST	/top500
Spring	GET	/top500
	POST	/top500

*Tableau 13 Mapping de la prédiction du classement (Top 500)*

## Annexes (UIs)




The image shows a user interface for a login and registration section. On the left, there is a white rounded rectangle containing a login form. The form has a title "Se Connecter" in bold black text. Below the title are two input fields: "Email" and "Password". Below these fields is a blue button labeled "SE CONNECTER". Below the button is a light blue rounded rectangle with the text "Découvrez si vous êtes un entrepreneur en répondant à quelques questions" and a small icon of a lightbulb with a gear inside. On the right, there is a large teal rounded rectangle. At the top of this rectangle is a small icon of a laptop with a bar chart. Below the icon is the text "Bonjour, Entrepreneur !" in bold white text. Below this text is a smaller line of white text: "Inscrivez-vous avec vos informations personnelles pour utiliser toutes les fonctionnalités du site". At the bottom of the teal rectangle is a white button labeled "S'INSCRIRE".

Figure 11 Formulaire de connexion



The image shows a user interface for a profile prediction form. On the left, there is a white rounded rectangle containing a form with several questions. The first question is "Age" with a text input field. The second question is "Choisissez Votre Education" with a dropdown menu showing "Sciences de l'ingénierie". The third question is "Choisissez Votre Trait de Personnalité" with a dropdown menu showing "Passion". Below these are three sets of radio button questions: "Est-ce que c'est un Projet Individuel ?" (Oui/Non), "Est-ce que votre Projet est dans la même Ville où vous habitez ?" (Oui/Non), and "Est-ce que quelqu'un vous a influencé pour créer votre projet ?" (Oui/Non). The last question is "Est-ce que vous avez un trouble mental ?" (Oui/Non). On the right, there is a large teal rounded rectangle. At the top of this rectangle is a small icon of a lightbulb with a gear inside. Below the icon is the text "Bonjour, Entrepreneur !" in bold white text. Below this text is a smaller line of white text: "Tester Votre Potentiel autant qu'Entrepreneur". Below this text is a white button labeled "SUIVANT". At the bottom of the teal rectangle is a white button labeled "SORTIR DE LA PRÉDICTION DU PROFILE ENTREPRENEUR".

Figure 12 Formulaire de prédiction de profil



## Vous y êtes Presque

Entrez les informations manquantes

RETOUR

Quel est votre degré de persévérance ?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Quel est votre niveau d'initiative ?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Quel est votre niveau de compétition ?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Quel est votre degré d'autonomie ?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

À quel point avez-vous un besoin de réussite ?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Quel est votre niveau de confiance en vous ?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Quel est votre niveau de santé actuel ?  
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

PRÉDIR

Figure 13 Formulaire de prédiction de profil (suite)



## Bienvenue à nouveau !

Entrez vos informations personnelles pour utiliser toutes les fonctionnalités du site

SE CONNECTER

## Entrepreneur

FirstName

LastName

Age

☐ Homme ☐ Femme

Email

Password

S'INSCRIRE

Figure 14 Formulaire de création de compte





## Conclusion

Le projet Project Advisor représente une avancée significative dans l'évaluation de la faisabilité des projets pour les startups, les entrepreneurs et les décideurs.

En s'appuyant sur une base de données robuste et des modèles de prédiction bien entraînés, l'application offre une solution innovante pour surmonter les défis liés à la planification, la mise en œuvre et la faisabilité des projets. Les modules de gestion des entrepreneurs et des projets, couplés à une architecture micro-services, garantissent une maintenabilité et une adaptabilité.

Les résultats obtenus grâce aux modèles de prédiction et de classification tels que le RandomForestClassifier et le KNeighborsClassifier démontrent une grande précision, comme en témoignent les scores d'accuracy et les rapports de classification.

Ces modèles, entraînés sur des jeux de données prétraités, permettent de fournir des prédictions objectives et basées sur des données réelles, aidant ainsi les utilisateurs à prendre des décisions éclairées sur la faisabilité du projet.

Enfin, le choix des technologies modernes comme Spring Boot, Flask, et Scikit-learn, ainsi que l'utilisation d'outils de collaboration, a permis d'implémenter un système performant tout en collaborant entre binôme. Les diagrammes de cas d'utilisation et l'architecture du système illustrent clairement la manière dont les différents acteurs interagissent avec l'application pour atteindre l'objectif du projet.

En somme, Project Advisor se positionne comme un outil intéressant pour les entrepreneurs, en leur offrant une plateforme intuitive et efficace pour évaluer la faisabilité de leurs projets tout en minimisant les risques et maximisant les chances de succès.

## Les jeux de données

- ✓ <https://www.kaggle.com/code/vishwakumar5201/entrepreneurial-competency-in-university-students/input>
- ✓ <https://www.kaggle.com/code/codebreaker619/predicting-profits-with-multiple-linear-regression/input>
- ✓ <https://www.kaggle.com/code/suraj520/logistic-reg-to-adaboost-implementations/input>

## Bibliographie

- <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
- <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- <https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6>
- <https://medium.com/@sahin.samia/demystifying-k-neighbors-classifier-knn-theory-and-python-implementation-from-scratch-f5e76d6f2d48>
- <https://www.free-work.com/fr/tech-it/blog/actualites-informatiques/impact-de-la-data-preparation-en-machine-learning>
- <https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/>
- <https://spring.io/projects/spring-data-jpa>
- <https://flask.palletsprojects.com/en/stable/>