

RAPPORT DE PROJET DE Beast League

Application FOOTBALL.

Élaboré par :

BOUFTINI MOHAMMED

EL BADRI ACHRAF

ER-RAMACH BOUCHRA

MAHFOUDH YASSINE

Année universitaire : 2023-2024

Table des matières

I. Introduction.....	3
1. Aperçu du Projet.....	3
a) Objectifs.....	3
II. Collecte et Stockage des Données	3
1. Sources de Données	3
2. Méthode de Collecte	4
3. Stockage	4
III. Traitement des Données	4
1. Prétraitement	4
a) Gestion des Valeurs Manquantes :	4
b) Normalisation des Données :	4
c) Création de Caractéristiques :	4
2. Stockage des Données Traitée	5
IV. Modèle AI.....	5
1. Sélection du Modèle.....	5
2. Entraînement et Évaluation:	5
a) Séparation des données:	5
b) Entraînement du modèle de régression:.....	5
c) Prédiction des points:.....	5
d) Entraînement du modèle de classification:.....	6
3. Résultats	6
a) Random Forest Regressor:	6
b) Random Forest Classifier:.....	6
V. Tableau de Bord.....	6
1. Choix de l'Outil	6
2. Fonctionnalités du Tableau de Bord.....	6
a) Comparaison des Lignes :	6
b) Performance Historique	6
c) Insights du Modèle :	6
VI. Conclusion.....	7

I. Introduction

Pour développer une approche basée sur les données afin d'identifier la meilleure ligue de football européenne des 3 à 5 dernières années. L'objectif est d'enrichir un nouveau forfait de streaming en incluant la ligue de football la plus attrayante. Ce projet implique la mise en place d'un pipeline ETL (Extract, Transform, Load) pour traiter les statistiques des ligues, avec éventuellement l'application de modèles AI/ML pour des analyses plus approfondies, et la création d'un tableau de bord pour la visualisation.

1. Aperçu du Projet

a) Objectifs

- Collecte de Données : Extraire des données des sites officiels des ligues de football européennes.
- Stockage de Données : Stocker les données brutes et traitées dans des formats structurés.
- Traitement de Données : Nettoyer et transformer les données pour l'analyse.
- Modèle AI : Construire et déployer un modèle AI pour analyser les performances des ligues.
- Tableau de Bord : Créer un tableau de bord visuel pour présenter les résultats.

II. Collecte et Stockage des Données

1. Sources de Données

Les données seront collectées à partir des sites officiels des principales ligues de football européennes, tels que :

- Premier League anglaise (<https://www.premierleague.com/>)
- La Liga (<https://www.laliga.com/>)
- Serie A (<https://www.legaseriea.it/>)
- Bundesliga (<https://www.bundesliga.com/>)
- Ligue 1 (<https://www.ligue1.com/>)

2. Méthode de Collecte

Les données seront extraites à l'aide de techniques de web scraping et d'appels API lorsque disponibles. Les outils et bibliothèques suivants seront utilisés :

- Python : Pour le script et l'automatisation.
- BeautifulSoup : Pour le web scraping.
- Requests : Pour gérer les requêtes HTTP.
- Pandas : Pour la manipulation et le stockage des données.

3. Stockage

Les données brutes seront stockées dans une base de données SQL avec des tables correspondant à chaque ligue. Les tables clés incluront :

Résultats des Matches : ID du match, date, équipe à domicile, équipe à l'extérieur, score, etc.

Classements des Ligues : Saison, équipe, points, victoires, défaites, nuls, etc.

Statistiques des Équipes : Équipe, saison, buts marqués, buts encaissés, etc.

III. Traitement des Données

1. Prétraitement

Les étapes de nettoyage et de prétraitement des données incluront :

- a) Gestion des Valeurs Manquantes : Imputation ou suppression des données manquantes.
- b) Normalisation des Données : Assurer la cohérence des formats de données.
- c) Création de Caractéristiques : Création de nouvelles caractéristiques telles que la différence de buts, les points par match, etc.

2. Stockage des Données Traitées

Les données traitées seront stockées dans des tables séparées au sein de la base de données SQL pour maintenir l'intégrité des données et faciliter l'accès. Exemples de tables :

- Résultats_Matches_Traitée
- Classements_Ligues_Traitée
- Statistiques_Équipes_Traitée

IV. Modèle AI

1. Sélection du Modèle

Pour prédire les points futurs et le classement des équipes de football, nous avons utilisé deux modèles de forêts aléatoires (Random Forest) : un pour la régression (prédiction de la meilleure league) et un pour la classification (prédiction des positions). Ces modèles sont implémentés à l'aide de la bibliothèque scikit-learn.

2. Entraînement et Évaluation:

a) Séparation des données:

Les données de l'année 2023 sont utilisées comme ensemble de test, tandis que les autres années sont utilisées pour l'entraînement.

Les caractéristiques et la cible sont séparées pour les ensembles d'entraînement et de test.

b) Entraînement du modèle de régression:

Le modèle Random Forest Regressor est entraîné sur les données d'entraînement.

La performance sur l'ensemble d'entraînement est évaluée à l'aide du score R^2 .

c) Prédiction des points:

Les points des équipes pour l'année 2023 sont prédits à l'aide du modèle de régression.

La meilleure ligue pour l'année 2023 est identifiée en fonction des points prédits.

d) Entraînement du modèle de classification:

Le modèle Random Forest Classifier est entraîné sur les données de la meilleure ligue prédite.

La position des équipes pour l'année 2023 dans cette ligue est prédite.

3. Résultats

a) Random Forest Regressor:

Score d'entraînement : Le modèle a montré une bonne performance sur l'ensemble d'entraînement, suggérant qu'il a bien capturé les relations entre les caractéristiques et les points des équipes.

Meilleure ligue prédite : La ligue avec les points prédits les plus élevés pour l'année 2023.

b) Random Forest Classifier:

Prédiction des positions : Le modèle a prédit les positions des équipes dans la meilleure ligue pour l'année 2023, fournissant une vue d'ensemble des performances futures des équipes dans cette ligue.

V. Tableau de Bord

1. Choix de l'Outil

Pour le tableau de bord, Tableau ou Power BI sera utilisé pour créer des visualisations interactives. Le choix dépendra de la facilité d'intégration avec les sources de données et de la capacité à créer des tableaux de bord dynamiques et perspicaces.

2. Fonctionnalités du Tableau de Bord

- a) Comparaison des Ligues : Comparaison visuelle des statistiques clés entre les ligues.
- b) Performance Historique : Tendances des classements des ligues et des performances des équipes au fil des ans.
- c) Insights du Modèle : Visualisation des résultats du modèle AI et des classements des ligues.

VI. Conclusion

Ce projet vise à fournir une analyse complète des ligues de football européennes, aidant le diffuseur à prendre des décisions basées sur les données pour leur nouveau forfait de streaming. La combinaison des processus ETL, des modèles AI et des tableaux de bord interactifs garantira des résultats perspicaces et exploitables.