****

**RAPPORT DE PROJET DE Beat 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](#_Toc167367810)

[1. Aperçu du Projet 3](#_Toc167367811)

[a) Objectifs 3](#_Toc167367812)

[II. Collecte et Stockage des Données 3](#_Toc167367813)

[1. Sources de Données 3](#_Toc167367814)

[2. Méthode de Collecte 4](#_Toc167367815)

[3. Stockage 4](#_Toc167367816)

[III. Traitement des Données 4](#_Toc167367817)

[1. Prétraitement 4](#_Toc167367818)

[a) Gestion des Valeurs Manquantes : 4](#_Toc167367819)

[b) Normalisation des Données : 4](#_Toc167367820)

[c) Création de Caractéristiques : 4](#_Toc167367821)

[2. Stockage des Données Traitée 5](#_Toc167367822)

[IV. Modèle AI 5](#_Toc167367823)

[1. Sélection du Modèle 5](#_Toc167367824)

[a) Modèles de Régression : 5](#_Toc167367825)

[b) Modèles de Clustering : 5](#_Toc167367826)

[2. Construction du Modèle 5](#_Toc167367827)

[a) Division des Données : 5](#_Toc167367828)

[b) Entraînement du Modèle 5](#_Toc167367829)

[c) Évaluation du Modèle : 5](#_Toc167367830)

[d) Ajustement des Hyperparamètres : 5](#_Toc167367831)

[3. Résultats 6](#_Toc167367832)

[V. Tableau de Bord 6](#_Toc167367833)

[1. Choix de l'Outil 6](#_Toc167367834)

[2. Fonctionnalités du Tableau de Bord 6](#_Toc167367835)

[a) Comparaison des Ligues : 6](#_Toc167367836)

[b) Performance Historique 6](#_Toc167367837)

[c) Insights du Modèle : 6](#_Toc167367838)

[VI. Conclusion 6](#_Toc167367839)

# 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.

## Aperçu du Projet

### 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.

# Collecte et Stockage des Données

## 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/)

## 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.

## 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 Matchs : 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.

# Traitement des Données

## Prétraitement

Les étapes de nettoyage et de prétraitement des données incluront :

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

## Stockage des Données Traitée

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\_Matchs\_Traitée
* Classements\_Ligues\_Traitée
* Statistiques\_Équipes\_Traitée

# Modèle AI

## Sélection du Modèle

Pour ce projet, un modèle prédictif sera construit pour classer les ligues en fonction de leur nature compétitive et des performances des équipes. Les modèles possibles incluent :

1. Modèles de Régression : Pour prédire les classements des ligues basés sur les performances passées.
2. Modèles de Clustering : Pour regrouper les ligues similaires en fonction des caractéristiques statistiques.

## Construction du Modèle

1. Division des Données : Diviser les données en ensembles d'entraînement et de test.
2. Entraînement du Modèle : Entraîner le(s) modèle(s) choisi(s) sur les données historiques.
3. Évaluation du Modèle : Évaluer les performances du modèle à l'aide de métriques telles que RMSE, précision, etc.
4. Ajustement des Hyperparamètres : Optimiser les paramètres du modèle pour de meilleures performances.

## Résultats

Le modèle fournira un classement des ligues, indiquant la meilleure ligue des 3 à 5 dernières années. Les résultats seront expliqués dans la documentation, mettant en évidence la signification statistique et les métriques de performance.

# Tableau de Bord

## 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.

## Fonctionnalités du Tableau de Bord

1. Comparaison des Ligues : Comparaison visuelle des statistiques clés entre les ligues.
2. Performance Historique : Tendances des classements des ligues et des performances des équipes au fil des ans.
3. Insights du Modèle : Visualisation des résultats du modèle AI et des classements des ligues.

# 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.