

INF 1521 : Projet tutoré

THÈME : Modèle de prédiction d'un résultat de match de football

Groupe N° 2

AZONYIBO Kossi Florentin

ATITSO Adjovi Sinyouli Anne

GANDONOU Koffi Patrick-Léon

Chargée : Mme AMOUZOU

SOMMAIRE

Introduction-----	2
Contexte-----	3
Problématique-----	3
Objectif-----	4
Proposition de solution-----	5
Choix de solution-----	7
Conception et Modélisation-----	8
RÉFÉRENCES-----	8

Introduction

Le football, sport le plus populaire au monde, suscite une passion sans précédent, rassemblant des millions de supporters et d'acteurs économiques autour d'un même engouement. Cet intérêt s'accompagne d'un besoin croissant de mieux comprendre et anticiper les résultats des matchs, que ce soit pour optimiser les stratégies des équipes, améliorer l'expérience des amateurs de football, ou orienter les investissements liés à ce domaine.

Dans ce cadre, le **Football Insight Predictor (FIP)**, un modèle innovant de prédiction des résultats de matchs, a été conçu pour répondre à ces attentes. S'appuyant sur l'analyse de données et les techniques avancées d'apprentissage automatique, le FIP vise à offrir des prédictions fiables et précises en exploitant une large variété de facteurs influençant les performances des équipes.

Ce document présente le cadre conceptuel, méthodologique et technique du développement de ce modèle. Il aborde également les défis inhérents à la prédiction des résultats sportifs et les solutions apportées par le FIP pour surmonter ces obstacles. En intégrant des approches modernes et des données pertinentes, notre objectif est de contribuer à la transformation des méthodes de prédiction dans le domaine sportif.

Contexte

Le football est l'un des sports les plus suivis et appréciés à travers le monde, générant un engouement considérable parmi les supporters et les investisseurs. Cet intérêt croissant s'accompagne d'une recherche constante de prédictions fiables pour orienter les décisions, qu'elles soient liées à des enjeux sportifs, financiers, ou stratégiques. Cependant, les méthodes traditionnelles de prédiction des résultats restent souvent limitées en termes de précision et d'exhaustivité.

Dans ce contexte, notre projet vise à améliorer la précision des modèles de prédiction en intégrant de nouveaux paramètres, issus d'une analyse approfondie des données disponibles. Cette approche innovante a pour objectif de fournir des insights plus pertinents et exploitables, tant pour les passionnés de football que pour les acteurs économiques du domaine.

Problématique

Le football, en tant que sport universellement suivi, génère un intérêt croissant pour la prédiction des résultats de matchs, que ce soit à des fins de divertissement, de stratégie ou

d'investissement. Cependant, les modèles traditionnels de prédiction se heurtent à plusieurs limitations :

1. La complexité du football, qui dépend d'une multitude de facteurs souvent imprévisibles, tels que les performances individuelles, les dynamiques collectives, ou les conditions extérieures.
2. L'exploitation incomplète ou inefficace des données disponibles, qui peuvent inclure des aspects tactiques, historiques, contextuels et statistiques.
3. La difficulté d'obtenir des prévisions suffisamment précises et exploitables pour répondre aux besoins des différents utilisateurs (supporters, analystes sportifs, investisseurs, etc.).

Dans ce contexte, comment concevoir un modèle de prédiction capable d'intégrer ces multiples paramètres de manière cohérente et d'améliorer significativement la précision des résultats prédits ?

Objectif

Objectif principale

L'objectif principal de ce projet est de concevoir un modèle de prédiction performant permettant de fournir des prévisions fiables et précises sur l'issue des matchs de football. Ce modèle vise à exploiter des données variées et pertinentes pour améliorer la qualité des prédictions, en tenant compte de facteurs clés influençant les performances des équipes.

Objectif spécifiques

Afin de concrétiser l'objectif principal du projet, plusieurs objectifs spécifiques ont été définis pour structurer les étapes clés et garantir une mise en œuvre efficace du modèle Football Insight Predictor (FIP).

- **Aide à la prise de décision pour les parieurs** : Les modèles de prédiction peuvent aider les parieurs à prendre des décisions plus éclairées en leur fournissant des probabilités de victoire, de défaite ou de match nul.
- **Analyse et stratégie pour les équipes** : Les entraîneurs et analystes sportifs peuvent utiliser ces prédictions pour développer des stratégies de jeu, en se basant sur les forces et faiblesses prévues des équipes adverses.

- **Engagement des fans** : Les supporters peuvent utiliser ces prédictions pour augmenter leur engagement et leur intérêt pour les matchs, en suivant les probabilités et les performances attendues.
- **Estimation des scores** : Fournir des prédictions précises sur le nombre de buts marqués par chaque équipe.
- **Exploration des tendances** : Les prédictions peuvent également aider à identifier des tendances à long terme, comme l'évolution des performances d'une équipe sur une saison ou plusieurs saisons.

Proposition de solution

Avant de voir les différentes options qui se présentent à nous pour réaliser nos objectifs, faisons un tour de ce qui se fait déjà.

Existant

Classements Elo [\[1\]](#)

Développés à l'origine pour les échecs, les classements Elo ont été adaptés à divers sports, y compris le football. Chaque équipe se voit attribuer une cote basée sur sa performance, et cette cote est ajustée après chaque match en fonction du résultat.

- . Ajustement dynamique
- . Différence entre les deux notes correspond à la probabilité de victoire

Simulations de Monte Carlo [\[2\]](#)

Les simulations de Monte Carlo impliquent de simuler un grand nombre de matchs pour estimer la probabilité de différents résultats. Ces simulations prennent en compte divers facteurs tels que :

- . Force de l'équipe
- . Variabilité aléatoire

Modèles bayésiens [\[3\]](#)

Les modèles bayésiens intègrent des connaissances ou des croyances antérieures sur les équipes et mettent à jour ces croyances au fur et à mesure que de nouvelles données deviennent disponibles. Cette approche est particulièrement utile dans le football car elle permet :

- . Flexibilité
- . Mise à jour continue

Différents Options

Solution 1

Modèle basé sur les algorithmes de Machine Learning classiques

Description :

Cette approche repose sur l'utilisation de modèles d'apprentissage automatique classiques, adaptés à la prédiction de catégories (victoire, défaite, nul).

1. Étapes :

- **Collecte des données** : Utilisation de bases de données fiables, incluant les résultats des matchs, les statistiques des équipes, et les conditions de jeu.
- **Prétraitement** : Nettoyage et transformation des données (encodage des variables catégoriques, normalisation des données numériques).
- **Création de caractéristiques** : Intégration de paramètres clés comme les performances récentes, les historiques des confrontations, les données des joueurs clés, ou encore l'impact des matchs à domicile/extérieur.
- **Entraînement du modèle** : Utilisation d'algorithmes tels que la régression logistique, les forêts aléatoires, ou les algorithmes de boosting (XGBoost, LightGBM).
- **Évaluation et optimisation** : Utilisation de métriques adaptées (accuracy, matrice de confusion) et optimisation des hyperparamètres.

2. Avantages :

- Mise en œuvre rapide et efficace.
- Facilité d'interprétation des résultats.
- Performance solide avec des données bien structurées.

3. Inconvénients :

- Sensibilité aux données bruitées ou incomplètes.

- Limitations face à des relations complexes ou non linéaires.

Solution 2

Modèle basé sur des réseaux de neurones artificiels (Deep Learning)

Description :

Cette solution utilise des techniques avancées de Deep Learning pour exploiter les relations complexes entre les données et améliorer la précision des prédictions.

1. Étapes :

- **Collecte et enrichissement des données** : En plus des données classiques, inclure des données plus complexes (données en temps réel, positions des joueurs, etc.).
- **Prétraitement avancé** : Gestion des valeurs manquantes, standardisation et augmentation des données (ex. oversampling pour classes déséquilibrées).
- **Conception du modèle** : Création d'un réseau neuronal adapté (Dense Neural Networks) ou d'un modèle récurrent (RNN/LSTM) pour analyser les séquences temporelles, comme les performances d'une équipe sur plusieurs matchs.
- **Entraînement et validation** : Utilisation de techniques avancées comme le dropout, la régularisation L2, et la validation croisée pour éviter le surapprentissage.
- **Évaluation** : Comparaison des résultats avec les méthodes classiques pour valider l'efficacité du modèle.

2. Avantages :

- Capacité à capturer des relations complexes entre les données.
- Meilleures performances sur des jeux de données volumineux et variés.
- Évolutivité pour inclure des données plus riches ou non structurées (ex. vidéos).

3. Inconvénients :

- Besoin de ressources importantes pour l'entraînement.
- Difficulté d'interprétation des résultats comparée aux modèles classiques.
- Dépendance à des jeux de données volumineux et bien annotés.

Choix de solution

Spécificité

Modèle basé sur les algorithmes de Machine Learning classiques

Pour le développement de notre modèle **Football Insight Predictor (FIP)**, nous avons opté pour une approche fondée sur les algorithmes classiques de Machine Learning. Ce choix est motivé par plusieurs raisons stratégiques et pratiques, en lien avec les objectifs et les ressources du projet.

1. Simplicité et rapidité de mise en œuvre :

Les modèles classiques tels que la régression logistique, les forêts aléatoires ou les algorithmes de boosting (XGBoost, LightGBM) permettent une mise en œuvre rapide et efficace. Leur structure relativement simple facilite la compréhension et l'interprétation des résultats, un atout essentiel pour identifier les facteurs clés influençant les prédictions.

2. Compatibilité avec les données disponibles :

Les données collectées pour ce projet (résultats historiques, statistiques des équipes, performances récentes) sont majoritairement structurées et bien adaptées aux algorithmes classiques. Ces derniers offrent de bonnes performances dans ce type de contexte sans nécessiter des volumes de données excessifs ou des prétraitements complexes.

3. Ressources limitées :

Contrairement aux approches basées sur le Deep Learning, les modèles classiques demandent moins de puissance de calcul et de temps d'entraînement, tout en produisant des résultats satisfaisants. Cela correspond mieux aux contraintes en termes de ressources matérielles et de temps disponible pour ce projet.

4. Évaluation et optimisation facilitées :

L'évaluation des performances et l'optimisation des modèles classiques sont relativement simples à mettre en œuvre grâce à des techniques éprouvées telles que la validation croisée et l'ajustement des hyperparamètres. Cela permet d'obtenir des résultats fiables tout en maîtrisant le risque de surapprentissage.

Implémentation

Les technologies utilisées dans notre projet Football Insight Predictor depuis le Front-End, Back-End jusqu'au déploiement:

Frontend :

- 1-HTML5
- 2-CSS3 avec Tailwind CSS
- 3-JavaScript (ES6+)

Backend :

- 1-Python 3
- 2-Flask (Framework web de Python)
- 3-Flask-CORS (Pour la gestion des CORS)
- 4-Requests (Pour les appels API externes)
- 5-Netlify (pour le déploiement)

Outils et Dépendances :

- 1-Node.js et NPM (Pour la gestion des dépendances frontend)
- 2-Git (Pour le contrôle de version)

Structure de Données :

- 3-JSON pour le stockage des données statiques
- 4-API REST (football-data.org) pour la communication client-serveur

Voici donc le Lien vers un premier livrable du projet: [Football Insight Predictor \(FIP\)](#)

L'interface d'accueil

Football Insight Predictor (FIP)

Prédiction de match

Championnat

Sélectionnez un championnat

Équipe 1

Sélectionnez une équipe

Équipe 2

Sélectionnez une équipe

Prédire le match

Matches à venir

Premier League

Manchester United VS Brighton

dimanche 12 janvier 2025 à 22:30

Premier League

Liverpool VS Arsenal

mercredi 8 janvier 2025 à 20:22

Champions League

Inter Milan VS Chelsea

vendredi 10 janvier 2025 à 07:28

Exemple de prédiction avec l'exemple du FC Barcelone VS R  al Madrid

Football Insight Predictor (FIP)

Pr  diction de match

Championnat

Ligue des Champions

  quipe 1

Barcelona

  quipe 2

Real Madrid

Pr  dire le match

R  sultats de la pr  diction

Barcelona

1

Real Madrid

2

Statistiques d  taill  es

Barcelona

Score: 1
Possession: 46%
Tirs: 12
Tirs cadr  s: 5
Cartons jaunes: 2
Cartons rouges: 1
Passes r  ussies: 439
Coups francs: 6
Corners: 6
Fautes: 11

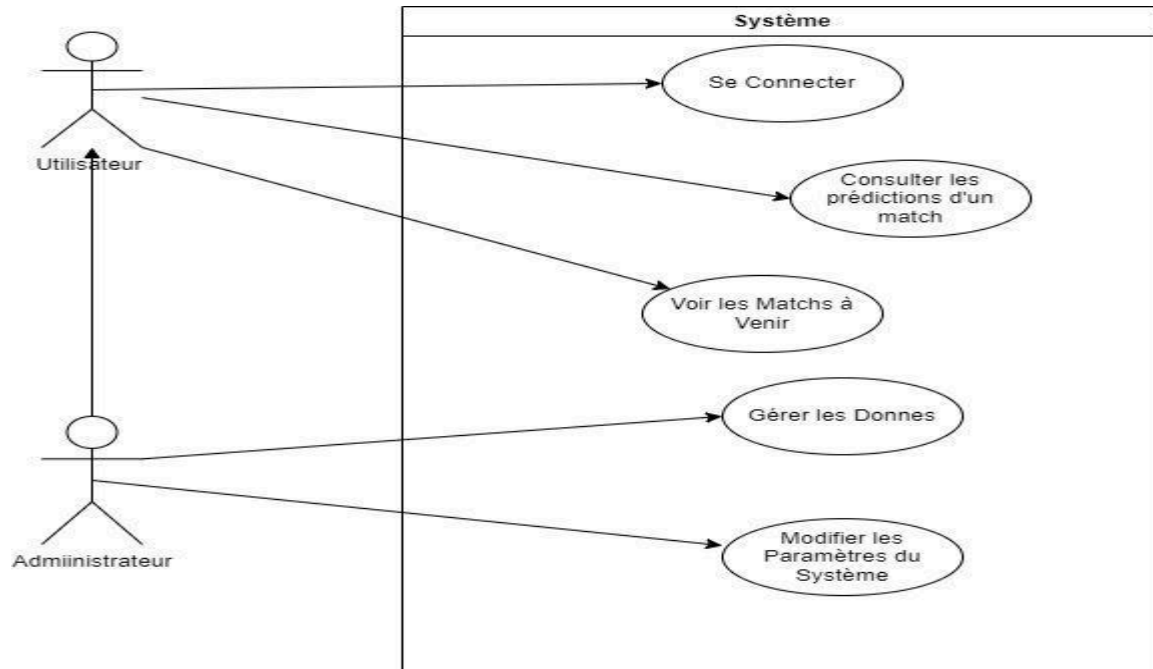
Real Madrid

Score: 2
Possession: 54%
Tirs: 17
Tirs cadr  s: 7
Cartons jaunes: 0
Cartons rouges: 1
Passes r  ussies: 600
Coups francs: 4
Corners: 6
Fautes: 8

Pr  cision de la pr  diction: 82.0%

Conception et Modélisation

Diagramme de Cas d'Utilisation



Description des Cas d'Utilisations

1. Se Connecter

Acteur principal : Utilisateur

Description :

L'utilisateur doit s'authentifier en fournissant un identifiant et un mot de passe pour accéder au système. Cela permet de sécuriser l'accès et de personnaliser l'expérience en fonction du type d'utilisateur (Administrateur ou Utilisateur standard).

Scénario de base :

1. L'utilisateur ouvre l'application.
2. Il saisit son identifiant et son mot de passe.
3. Le système vérifie les informations.
4. Si les informations sont valides, l'utilisateur est redirigé vers la page principale.
5. Sinon, un message d'erreur est affiché.

2. Consulter les prédictions d'un match

Acteur principal : Utilisateur

Description :

L'utilisateur peut consulter les prédictions pour un match spécifique en sélectionnant les équipes concernées. Les prédictions sont générées à partir des données statistiques des équipes.

Scénario de base :

1. L'utilisateur accède à la section "Prédictions".
2. Il sélectionne deux équipes à comparer.
3. Le système affiche les prédictions pour le match, basées sur les statistiques disponibles.

3. Voir les matchs à venir

Acteur principal : Utilisateur

Description :

L'utilisateur peut consulter la liste des matchs programmés dans un futur proche.

Scénario de base :

1. L'utilisateur accède à la section "Matchs à venir".
2. Le système affiche une liste des matchs avec leurs dates et les équipes participantes.
3. L'utilisateur peut cliquer sur un match pour obtenir plus de détails.

4. Gérer les données

Acteur principal : Administrateur

Description :

L'administrateur peut ajouter, modifier ou supprimer les données du système, telles que les équipes, les joueurs ou les statistiques des matchs.

Scénario de base :

1. L'administrateur se connecte au système.
2. Il accède à la section "Gestion des données".
3. Il choisit de créer, mettre à jour ou supprimer des données.
4. Le système enregistre les modifications.

5. Modifier les paramètres du système

Acteur principal : Administrateur

Description :

L'administrateur peut configurer ou modifier les paramètres du système, comme les seuils statistiques utilisés pour les prédictions ou les informations générales affichées.

Scénario de base :

1. L'administrateur se connecte au système.
2. Il accède à la section "Paramètres".
3. Il modifie les options souhaitées.
4. Le système sauvegarde les modifications.

Diagramme de classe

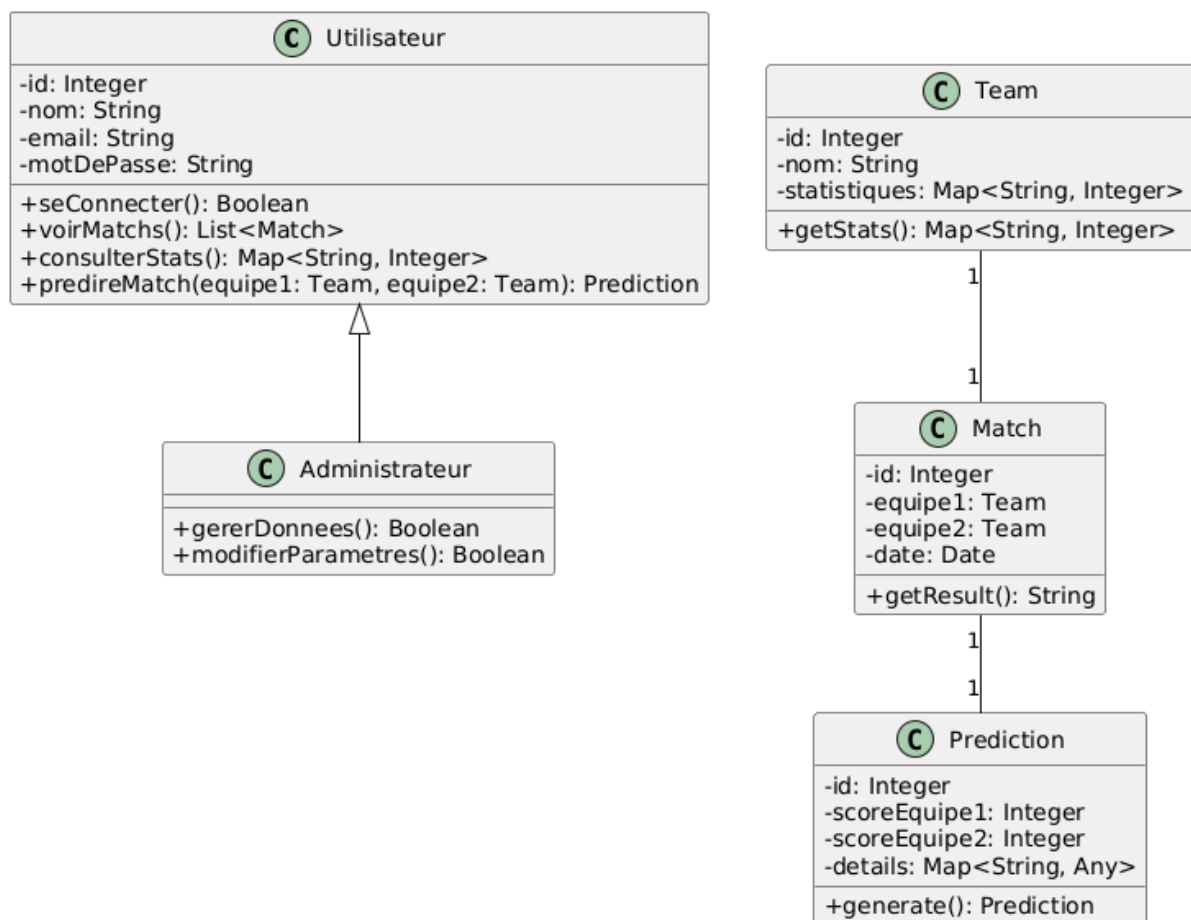
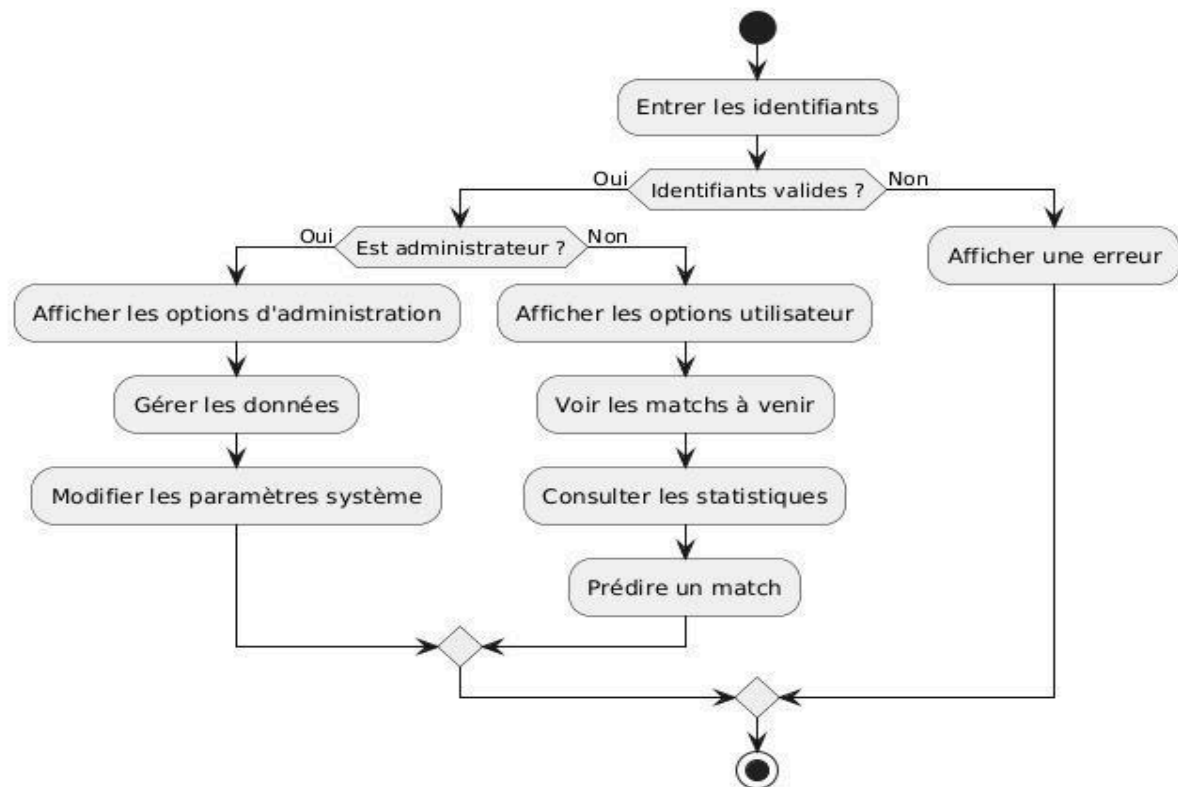


Diagramme d'Activité



Conclusion

Le développement du modèle **Football Insight Predictor (FIP)** marque une étape significative dans l'exploration des capacités de prédiction des résultats de matchs de football. En combinant des données variées et des techniques avancées de Machine Learning, ce projet vise à offrir une solution fiable et performante, adaptée aux besoins des passionnés, des analystes, et des acteurs économiques du domaine.

Grâce à une approche méthodique, incluant la collecte, le traitement et l'analyse des données, le modèle a été conçu pour maximiser la précision des prédictions tout en restant adaptable à des évolutions futures, telles que l'intégration de données supplémentaires ou l'utilisation de méthodes plus sophistiquées.

Ce projet met en lumière l'importance croissante des technologies d'analyse de données dans le secteur sportif et ouvre la voie à de nouvelles opportunités pour améliorer la compréhension et l'anticipation des performances des équipes.

RÉFÉRENCES

- [1] https://fr.wikipedia.org/wiki/Classement_Elo#Annexes
- [2]<https://medium.com/@markfootballdata/thats-not-how-it-should-have-ended-a30a9ff0a049>
- [3] football-data.org - ur src for machine readable football data

TABLE DES MATIÈRES

Introduction.....	3
Contexte.....	3
Problématique.....	3
Objectif.....	4
Objectif principale.....	4
Objectif spécifiques.....	4
Proposition de solution.....	5
Existant.....	5
Classements Elo.....	5
Simulations de Monte Carlo.....	5
Modèles bayésiens.....	6
Différents Options.....	6
Solution 1.....	6
Solution 2.....	7
Choix de solution.....	8
Spécificité.....	8
Implémentation.....	8
Frontend :.....	9
Backend :.....	9
Outils et Dépendances :.....	9
Structure de Données :.....	9
L'interface d'accueil.....	10
Exemple de prédiction avec l'exemple du FC Barcelone VS R��al Madrid.....	10
Conception et Mod��lisation.....	11
<u>Diagramme de Cas d'Utilisation.....</u>	<u>11</u>
<u>Description des Cas d'Utilisations.....</u>	<u>11</u>
Diagramme de classe.....	13
Diagramme d'Activit��.....	14
Conclusion.....	14
R��F��RENCES.....	15