

Mémoire de Fin d'Etudes

Pour l'Obtention du Diplôme de Master en Informatique

Présenté par :

GHEZ ABDELKADER

Domaine : Mathématiques & Informatique

Spécialité Réseaux et systèmes d'information.

Session 1 2023

THEME

Conception et Mise en œuvre d'un Système Intelligent et Distribué d'Aide au
Pronostic Sportif

Encadré par : Pr. HAMDADOU Djamila

Co-encadreur : Dr. SFIAT Sid Ahmed

Jury

Président : Pr. KADDOUR Mejdj

Examineur : Pr. HAFFAF Hafid

Code Master : 05/2023

Résumé.

L'Intelligence Artificielle est utilisée en amont d'un match de football afin de réaliser un pronostic. Elle offre donc la possibilité à des parieurs débutants ou experts de faire des pronostics sportifs. Même si la prédiction parfaite semble toujours impossible, l'algorithme prédictif reste un avantage certain face aux humains. Il analyse et compare des statistiques, probabilités et autres facteurs en amont d'un match. Dans cette optique, plusieurs plateformes en ligne proposent de s'appuyer sur l'Intelligence Artificielle pour établir des pronostics.

Ce projet consiste à exploiter les techniques de l'Intelligence Artificielle pour faire des prédictions intelligentes sur des contestes sportifs en se basant sur l'historique des contestants. Le système d'aide à la décision proposé fournit des pronostics sous forme de probabilité de gain, perte, ou match nul entre les équipes dans les différents matchs.

L'utilisateur du système a l'option d'utiliser ces prédictions ou prédire le résultat en faisant appel à ses connaissances personnelles. À la fin, une synthèse est présentée à l'utilisateur afin de faire une comparaison entre son choix et celui présenté par le système intelligent. Ce système prend en considération tous les aspects socio-économiques tout en évaluant son impact direct sur la population jeune ayant un intérêt commun qui est le football local et international.

Mots clés : Modèles Prédictifs, Pronostic sportif, Aide à la décision, Apprentissage Profond, Intelligence Artificielle, Application Mobile, Réseaux de Neurones.

Abstract

Artificial Intelligence is used prior to a football match to make predictions, providing an opportunity for both beginner and expert bettors to make sports forecasts. Although perfect prediction seems impossible, predictive algorithms still have a clear advantage over humans. They analyze and compare statistics, probabilities, and other factors before a match. In this regard, several online platforms offer the use of Artificial Intelligence to establish predictions.

This project aims to leverage Artificial Intelligence techniques to make intelligent predictions in sports contexts based on historical contestant data. The proposed decision support system provides predictions in the form of probabilities for winning, losing, or drawing between teams in various matches.

Users of the system have the option to rely on these predictions or make their own predictions based on personal knowledge. At the end, a summary is presented to the user for comparing their choice with that presented by the intelligent system. The system takes into consideration socio-economic aspects while assessing its direct impact on the young population with a shared interest in local and international football.

Keywords: Predictive Models, Sports Prediction, Decision Support, Deep Learning, Artificial Intelligence, Mobile Application, Neural Networks.

الملخص

يستخدم الذكاء الاصطناعي قبل مباراة كرة القدم للتنبؤ. لذلك فهو يوفر إمكانية للمراهنين المبتدئين أو الخبراء لعمل تنبؤات رياضية. حتى لو بدا التنبؤ المثالي دائماً مستحيلاً، تظل الخوارزمية التنبؤية ميزة أكيدة على البشر. يحلل ويقارن الإحصائيات والاحتمالات والعوامل الأخرى قبل المباراة. مع وضع ذلك في الاعتبار، تقدم العديد من المنصات عبر الإنترنت الاعتماد على الذكاء الاصطناعي لإنشاء التنبؤات.

يتمثل هذا المشروع في استغلال تقنيات الذكاء الاصطناعي لعمل تنبؤات ذكية في المسابقات الرياضية بناءً على تاريخ المتسابقين. يوفر نظام دعم القرار المقترح تنبؤات في شكل احتمالية الفوز أو الخسارة أو التعادل بين الفرق في المباريات المختلفة.

مستخدم النظام لديه خيار استخدام هذه التنبؤات أو التنبؤ بالنتيجة باستخدام المعرفة الشخصية. في النهاية يتم تقديم ملخص للمستخدم لإجراء مقارنة بين اختياره وتلك التي يقدمها النظام الذكي. يأخذ هذا النظام في الاعتبار جميع الجوانب الاجتماعية والاقتصادية أثناء تقييم تأثيره المباشر على الشباب ذوي المصلحة المشتركة وهي كرة القدم المحلية والدولية.

الكلمات المفتاحية: النماذج التنبؤية، الإنذار الرياضي، دعم القرار، التعلم العميق، الذكاء الاصطناعي، تطبيقات الهاتف المحمول، الشبكات العصبية.

REMERCIEMENT

Tous mes remerciements vont à ALLAH, le tout puissant et miséricordieux, pour m'avoir donné la patience et la force de persévérer, ainsi que la volonté pour dépasser toutes les difficultés auxquelles je suis confronté.

Je remercie très chaleureusement, mes très chers parents qui se sacrifient chaque jour pour me voir réussir dans ma vie.

Je remercie, également chaleureusement, mon encadrante Professeur HAMDADOU Djamilia, pour sa disponibilité, son attention constante à mon travail, et surtout pour ses conseils judicieux, qui ont contribué à guider mes réflexions. Je suis reconnaissant pour le temps considérable qu'elle m'a accordé et j'apprécie, considérablement, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. Son énergie et sa confiance ont été les éléments moteurs pour moi. J'ai pris un grand plaisir à travailler avec elle.

Je remercie sincèrement, mon Co-encadrant Dr SFIAT Sid Ahmed, pour ses conseils et sa précieuse aide tout au long de ce travail.

Je tiens à exprimer ma sincère reconnaissance et mon respect aux membres du jury qui m'ont honoré en évaluant ce travail.

J'exprime, consciencieusement, ma gratitude au corps professoral et administratif de notre université pour la qualité et la richesse de l'enseignement qu'ils m'ont aimablement dispensé pendant mes études. Il a constitué une base essentielle en me donnant une formation efficace qui a été un soutien colossal pour mon projet de fin d'études.

Afin de n'oublier personne, mes vifs remerciements s'adressent à tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce travail.



DÉDICACE

Je tiens à dédier cet humble travail comme preuve de respect, de gratitude et de reconnaissance.

À mes très chers parents,

Pour vos sacrifices infinis, vos prières, vos soutiens et pour tout l'amour inconditionnel que vous m'avez montré. Aucun mot ne pourra jamais décrire à quel point je suis reconnaissant pour tout ce que vous avez fait pour moi. J'espère que vous trouverez dans ce travail le témoignage de ma reconnaissance éternelle.

Que le bon Dieu procure mes parents une bonne santé et une longue vie.

*À mon cher frère **Mohamed Saad** et mes adorables sœurs*

Je les remercie pour leurs amours inconditionnels, leurs soutiens continuels, leurs fois en mes capacités, ainsi que pour les encouragements sincères qu'ils n'ont jamais manqué de m'offrir tout au long de mon parcours et pendant les moments dont j'avais le plus besoin.

*À mon cher petit neveu **Wassim***

Je dédie ce travail à mon petit adorable avec tous mes meilleurs vœux de bonheur, de santé et de réussite. Puisse Dieu le protéger, éclairer son chemin.

À toute ma famille

Que nulle dédicace ne puisse exprimer mon respect et ma considération pour votre soutien et encouragements. Je vous dédie ce travail en reconnaissance de l'amour que vous m'offrez quotidiennement et votre bonté exceptionnelle. Que Dieu le Tout Puissant vous garde et vous procure santé et bonheur.

Table des matières

Résumé.....	
Abstract.....	
الملخص.....	
REMERCIEMENT	
DÉDICACE.....	I
TABLE DES LUSTRATIONS	VII
LISTE DES TABLEAUX.....	IX
INTRODUCTION GÉNÉRALE.....	1
CHAPITRE1 SYNTHÈSE DE L'ÉTAT DE L'ART	4
1. Introduction.....	5
2. Le pronostic sportif	5
2.1. Limites et défis de l'informatique dans le football.....	6
3. L'Aide à la décision.....	7
3.1 Définition.....	7
3.2 L'Aide à la décision et le Pronostic sportif.....	8
4. L'Intelligence Artificielle (IA)	9
4.1. Définitions de l'Intelligence Artificielle.....	9
4.2. Historique de l'intelligence artificielle.....	9
4.3. Les approches de l'IA.....	10
4.4. Les défis de l'IA	10
4.5. Pourquoi des machines intelligentes ?	11
4.6. Le raisonnement dans l'IA	11
4.7. Les techniques de l'IA.....	12
4.8. Domaines d'applications de l'Intelligence Artificielle	13
4.9. L'apprentissage automatique (Machine Learning).....	13
4.10. Les types d'apprentissage	14
4.10.1. L'apprentissage supervisé.....	14
4.10.2. L'apprentissage non supervisé	15
4.10.3. L'apprentissage semi supervisé.....	15

4.10.4.	L'apprentissage par renforcement	15
4.11.	Les techniques de l'apprentissage automatique	16
4.11.1.	K plus proche voisins (K-PPV)	16
4.11.2.	Support Vector Machine (SVM)	17
4.11.3.	Les arbres de décision (Decision Tree)	17
4.12.	La classification	18
4.13.	Les types de classification	19
4.13.1.	Classification binaire	19
4.13.2.	Classification multi- classe	19
4.14.	L'Apprentissage profond	20
4.14.1.	Fonctionnement de l'apprentissage profond.....	20
4.14.2.	Architectures d'apprentissage profond.....	21
4.14.2.1.	Les réseaux Long Short Memory (LSTM)	21
4.14.2.2.	Les réseaux convolutifs (CNN).....	22
5.	Travaux connexes	23
6.	Conclusion.....	25
CHAPITRE2 CONCEPTION ET MODÉLISATION DU SYSTÈME PROPOSÉ		26
1.	Introduction.....	27
2.	Objectifs visés.....	27
3.	Étude de l'existant	27
3.1	Predict395	27
3.1.1.	Le site web Predict395.....	27
3.2	1xBet	29
3.2.1.	Le site web 1xBet	29
3.2.2.	Les limites de 1xBet.....	30
4.	Architecture générale du système proposé	30
4.1.	Module de structuration.....	32
4.2.	Module d'exploitation	32
4.3.	Module de concrétisation	32
5.	À propos des données utilisées	33
5.1.	Description des données récoltées	33
5.2.	Description des attributs	33
5.3.	Prétraitement et nettoyage des données	34
5.4.	Encodage des données.....	34

5.5.	Augmentation des données	35
5.6.	Sélection de variables (attributs) significatifs	36
5.7.	Approche basée sur la classification ou la régression	38
5.8.	Les méthodes d'apprentissage utilisées	40
5.8.1.	XGBoost	40
5.8.2.	Régression logistique	41
5.8.3.	Classification naïve bayésienne	41
5.8.4.	Arbres de décision	41
5.8.5.	Forêts aléatoires	41
5.8.6.	Machines à vecteurs de support (SVM).....	42
5.9.	Sélection et comparaison des modèles de prédiction.....	43
5.9.1.	Matrice de confusion.....	43
4.9.1.1.	Matrice de confusion avant le match « Before match ».....	44
49.1.2.	Matrice de confusion mi-temps de match « half match »	45
5.9.2.	Précision.....	46
5.9.3.	Rappel	46
5.9.4.	F1-score	46
5.9.5.	Exactitude (Accuracy).....	47
5.9.6.	Précision équilibrée (Balanced accuracy)	47
6.	Choix du modèle	51
7.	Modélisation UML du système d'aide à la décision.....	51
7.1.	Diagramme de Cas d'utilisation (Use Case).....	51
7.1.1.	Cas d'utilisation « User ».....	52
7.1.2.	Cas d'utilisation « Admin ».....	53
7.2.	Diagramme de séquences.....	55
8.	Conclusion.....	59
CHAPITRE3 MISE EN OEUVRE		61
1.	Introduction.....	62
2.	Outils de développement	62
2.1.	BackEnd	62
2.1.1.	PHP.....	62
2.1.2.	Symfony	62
2.1.3.	MySQL	63
2.2.	Frontend.....	63

2.2.1.	JavaScript.....	63
2.2.2.	CSS.....	63
2.2.3.	HTML.....	64
2.2.4.	Bootstrap.....	64
2.2.5.	Twig.....	64
2.2.6.	Flutter.....	64
2.2.7.	Dart.....	65
2.3.	Machine Learning.....	65
2.3.1	BeutifulSoup.....	65
2.3.2	SK Learn.....	65
2.3.3	Pandas.....	65
2.3.4	Numpy.....	65
2.3.5	Matplotlib.....	66
3.	Environnement de développement.....	66
3.1	Visual Studio Code.....	66
3.2	Jupyter NoteBook.....	66
4.	Outils matériels.....	66
5.	Fonctionnalités du site web proposé et scénario d'exécution.....	67
5.1.	Page d'authentification admin.....	67
5.2.	Page tableau de bord.....	67
5.3.	Page Statistique.....	68
5.4.	Page Prédiction.....	70
5.5.	Page questions.....	72
5.6.	Pages Post et Statut.....	72
5.7.	Page Adversaire.....	73
5.8.	Page Compétition.....	74
5.9.	Page matchs.....	75
5.10.	Page support messages.....	77
5.11.	Page users.....	78
5.12.	Page admin panel.....	78
6.	Fonctionnalités de l'application proposée « Pariball ».....	79
6.1.	Lancement de l'application.....	79
6.2.	Page d'accueil.....	80
6.3.	Page authentification et Profil utilisateur.....	81

6.4. Page match	83
6.5. Page bettin	83
6.6. Page pack	84
6.7. Page zone des fans.....	85
7. Conclusion.....	86
CONCLUSION GÉNÉRALE.....	87
Bibliographie	90

TABLE DES LUSTRATIONS

Figure 1: Les différentes techniques de l'IA.....	12
Figure 2:les techniques de l'apprentissage automatique.....	14
Figure 3:Exemple d'un agent utilisant l'apprentissage par renforcement.....	16
Figure 4:La tâche de la classification en utilisant l'algorithme K plus proche voisins	17
Figure 5:La classification en utilisant les arbres de décision.....	18
Figure 6: Le principe de la classification.	18
Figure 7:La classification dans l'apprentissage profond	20
Figure 8:Exemple de fonctionnement de Deep Learning.....	21
Figure 9:Fonctionnement d'un réseau LSTM.....	22
Figure 10:Fonctionnement d'un réseau convolutif (CNN)	23
Figure 11:Le site web Predict395.....	28
Figure 12: Le site web 1xBet.....	29
Figure 13:Architecture globale du système proposé.....	31
Figure 14:La matrice de corrélation	36
Figure 15:La forêt aléatoire	41
Figure 16:SVM.....	42
Figure 17:Matrice de confusion des 6 modèles utilisés avant le match.....	43
Figure 18:Matrice de confusion des 6 modèles utilisés mi-temps de match	45
Figure 19:Statistiques du modèle logistic regression avant le match	47
Figure 20: Statistiques du modèle Gaussain Naive Bayes Classifieur avant le match ..	47
Figure 21: Statistiques du modèle Gaussain Naive Bayes Classifieur avant le match ..	48
Figure 22 : Statistiques du modèle XGBClassifier avant le match	48
Figure 23: Statistiques du modèle logistic regression mi-temps	49
Figure 24: Statistiques du modèle Gaussain Naive Bayes Classifieur mi-temps	49
Figure 25: Statistiques du modèle Random Forest mi-temps	50
Figure 26:Statistiques du modèle XGBClassifier mi-temps	50
Figure 27: Diagramme cas d'utilisation User	52
Figure 28:Diagramme cas d'utilisation Admin	54
Figure 29: Diagramme de séquence authentification application.....	56
Figure 30:Diagramme de séquence des différentes opérations dans l'application mobile.....	58
Figure 31: Diagramme de séquence de la partie site web et la gestion de l'admin.....	59
Figure 32: Diagramme de classes de notre système	60

Figure 33: Page d'authentification	69
Figure 34: Onglet tableau de bord.....	70
Figure 35:Page statistique Before match	71
Figure 36: Matrice de confusion	72
Figure 37: Onglet prédiction automatique.....	73
Figure 38: Onglet prédiction manuelle.....	74
Figure 39: Page questions.....	75
Figure 40: Page posts.....	75
Figure 41: Création d'un post	76
Figure 42: Liste des status	76
Figure 43: Page adversaire	77
Figure 44: Création d'un nouveau club.....	77
Figure 45: Page compétition.....	78
Figure 46: Création d'une nouvelle compétition	78
Figure 47: Page matchs	79
Figure 48: Création d'un nouveau match	80
Figure 49: Page support messages	81
Figure 50: Page users.....	81
Figure 51: Admin pannel.....	82
Figure 52: Lancement de Pariball.....	83
Figure 53: Page d'accueil d'application.....	84
Figure 54: Page authentification.....	85
Figure 55: Profil utilisateur	85
Figure 56: Page match.....	86
Figure 57: Prédiction de match.....	87
Figure 58: Page pack	88
Figure 59: Page fans area	89

LISTE DES TABLEAUX

Tableau 1 : Travaux connexes.....	24
Tableau 2: Description des attributs	34
Tableau 3: Les données augmentées.....	36
Tableau 4:Description du diagramme cas d'utilisation User.....	53
Tableau 5: Description du diagramme cas d'utilisation Admin	55

INTRODUCTION GÉNÉRALE

Introduction Générale

Contexte de l'étude

Le football est l'un des sports les plus populaires et pratiqués au monde, il a connu de nombreuses évolutions au fil des ans. Parmi ces évolutions, le pronostic sportif est devenu une partie importante de l'univers du football, avec des millions de fans cherchant à prédire les résultats des matchs. Cela implique l'analyse de statistiques, de performances et d'autres facteurs afin d'anticiper le résultat d'un match. Cependant, cette analyse peut être complexe et nécessite des compétences avancées en analyse de données et en apprentissage automatique.

Problématique

L'analyse des données sportives peut être fastidieuse et complexe, mais elle offre de nombreuses possibilités, telles que l'évaluation des performances des joueurs et des équipes, la prédiction des résultats de matchs, l'optimisation des stratégies de jeu, etc. Les données sportives peuvent provenir de différentes sources et sont de plus en plus utilisées dans le sport professionnel, l'entraînement, la gestion d'équipe et les paris sportifs.

De plus, l'analyse des données sportives permet également de découvrir des tendances et des schémas de jeu, d'identifier les forces et les faiblesses des adversaires, et d'améliorer la prise de décision stratégique tant sur le terrain que dans la gestion des équipes. Elle contribue ainsi à une approche plus factuelle et basée sur des données pour améliorer les performances sportives.

Comment exploiter au mieux ces données pour fournir des prédictions intelligentes et utiles aux amateurs de football ?

Contribution

Ce projet de fin d'étude vise à exploiter les techniques de l'Intelligence Artificielle pour élaborer des prédictions intelligentes sur des contextes sportifs, en se basant sur l'historique des participants. Les données nécessaires sont disponibles gratuitement sous forme d'API sur internet. Le système proposé doit être capable de fournir des pronostics en termes de probabilité de gain, de perte ou de match nul entre les équipes. Les utilisateurs ont la possibilité d'utiliser ces prédictions ou de faire appel à leurs

propres connaissances. Une synthèse comparative entre les choix de l'utilisateur et ceux du système sera également présentée.

L'objectif est de prendre en compte les aspects socio-économiques de la société et d'évaluer l'impact sur la population jeune qui s'intéresse au football local et international, afin de fournir des pronostics sportifs adaptés à leurs besoins et à leurs intérêts spécifiques. Cette population jeune cherche à être informée de manière fiable et à prendre des décisions éclairées lorsqu'elle s'engage dans des activités de pronostic sportif. En fournissant des prédictions intelligentes basées sur l'analyse des données sportives, le système a contribué à leur expérience de paris sportifs en leur fournissant des informations précieuses pour maximiser leurs chances de succès et améliorer leur engagement dans le monde du football.

Organisation du mémoire

Ce mémoire s'articule autour de trois chapitres outre l'introduction générale, la conclusion et les perspectives qui se résument comme suit :

Le chapitre 1 : Synthèse de l'état de l'art

Étale un état de l'art sur Le pronostic sportif, il portera, également, sur les systèmes d'aide à la décision (SAD). Cet état de l'art permet aussi de maîtriser les techniques les plus importantes de l'IA, et mettre l'accent sur les travaux connexes les plus pertinents parmi les écrits récemment et réalisés dans le même contexte que notre étude.

Le chapitre 2 : Conception et Modélisation

Décrit les architectures globale et fonctionnelle ainsi que tous les détails se rapportant à la conception et la modélisation du système intelligent proposé. Nous abordons, dans ce même chapitre, la démarche décisionnelle adoptée au problème des prédictions à l'aide de l'Intelligence Artificielle et l'Informatique Décisionnelle dans le pronostic sportif ainsi que les méthodes d'IA utilisées tout en justifiant les différents choix retenus. Enfin, nous présentons les différents résultats d'expérimentation issus des deux approches prédictifs suggérées.

Le chapitre 3 : Implémentation et Mise en œuvre du système proposé

Est consacré au déroulement et au déploiement du système ainsi que la présentation des principales interfaces de notre système intelligent avec tous les environnements logiciels et le matériel de développement utilisés ayant servi à sa mise en œuvre.

Enfin, nous clôturons ce mémoire par une conclusion générale où nous récapitulons les apports de la présente étude et des perspectives relatives à l'extension de notre travail.

CHAPITRE 1 SYNTHÈSE DE
L'ÉTAT DE L'ART

1. Introduction

Au cours des dernières années, le football est devenu le sport le plus populaire au monde, il attire, généralement, des millions de personnes pour le regarder au niveau international. Le football moderne se présente comme un système dynamique, où il existe une relation de compétition entre les équipes et la coopération des membres de la même équipe. Cette modalité est caractérisée comme un système complexe, dans lequel l'organisation se déroule à plusieurs niveaux selon une analyse tactique.

Le pronostic sportif est devenu un domaine d'application incontournable de l'Intelligence Artificielle, qui s'est développé ces dernières années dans différents secteurs. Dans les paris hippiques ou de football, l'IA utilise des prédictions mathématiques pour permettre aux parieurs de miser intelligemment sur des courses ou des matchs préalablement recherchés.

Dans ce premier chapitre, nous aborderons l'historique et l'évolution de l'informatique dans le football, en mettant en relief les limites et les défis rencontrés. Nous discuterons, également, l'aide à la décision et ses applications pratiques. Ensuite, nous explorerons l'Intelligence Artificielle en profondeur, en abordant ses différentes approches, son importance, son raisonnement, ses techniques, ses domaines d'application et l'apprentissage automatique (Machine Learning) ainsi que l'apprentissage Profond (Deep Learning).

2. Le pronostic sportif

Les origines du football remontent à l'Antiquité, où de nombreux peuples jouaient à des jeux de balle avec des règles et des objectifs différents. Cependant, le football moderne a émergé en Angleterre au XIXe siècle, où les écoles ont commencé à jouer à des jeux de balle en utilisant des règles similaires à celles du football d'aujourd'hui.

L'impact de l'informatique dans le football est énorme et continue de se développer rapidement. Les innovations technologiques ont amélioré la gestion de l'équipe, les performances des joueurs, la prise de décision, les interactions avec les fans et les téléspectateurs, et même la sécurité des joueurs. Les technologies telles que la vidéo-assistance à l'arbitrage, la technologie de la ligne de but, l'analyse avancée des performances des joueurs et des équipes, la reconnaissance faciale et la réalité virtuelle sont autant d'exemples de la façon dont l'informatique a amélioré le sport. L'utilisation de l'informatique dans le football devrait continuer de se développer, ouvrant ainsi de

nouvelles opportunités pour améliorer les performances des joueurs, la sécurité des spectateurs, l'expérience des fans et la gestion des équipes. [1]

2.1.Limites et défis de l'informatique dans le football

Malgré les avantages de l'informatique dans le football, il y a aussi des limites et des défis à considérer [2] à savoir :

- **Coût élevé** : l'installation et la maintenance des systèmes informatiques sophistiqués peuvent être coûteuses, ce qui peut limiter l'accès à cette technologie pour les équipes moins riches.
- **Fiabilité** : comme toute technologie, les systèmes informatiques peuvent tomber en panne ou être sujets à des erreurs, ce qui peut entraîner des décisions erronées sur le terrain.
- **Contrôle de données** : les données collectées sur les performances des joueurs et des équipes sont précieuses, mais leur utilisation et leur stockage doivent être soigneusement contrôlés pour éviter les violations de la vie privée et les erreurs potentielles.
- **Éthique** : l'utilisation de la technologie dans le football peut soulever des questions éthiques, comme le recours à des outils de surveillance pour suivre les performances des joueurs ou la diffusion de données en temps réel sur les médias sociaux.
- **Impact sur l'expérience du spectateur** : l'utilisation excessive de la technologie peut perturber l'expérience du spectateur, notamment en prolongeant la durée des matchs ou en réduisant l'excitation des fans.
- **Déséquilibre compétitif** : les équipes qui ont accès à des systèmes informatiques sophistiqués peuvent avoir un avantage sur les équipes moins équipées, ce qui peut créer un déséquilibre compétitif dans le sport.

En conclusion, l'informatique a apporté de nombreux avantages au football, mais il y a aussi des limites et des défis à considérer. Pour maximiser les avantages de la technologie, il est important de comprendre ces défis et de travailler ensemble pour trouver des solutions équitables et efficaces pour les équipes, les joueurs et les fans.

Dans le domaine sportif, l'aide à la décision peut utiliser les pronostics sportifs comme l'une des sources d'information pour guider les choix et les décisions stratégiques.

3. L'Aide à la décision

L'aide à la décision, vise à fournir des informations et des recommandations pour aider les individus ou les organisations à prendre des décisions optimales

3.1 Définition

L'aide à la décision est un domaine interdisciplinaire qui implique des techniques et des outils de divers domaines tels que l'informatique, les mathématiques, la statistique, la psychologie, la théorie de la décision, l'ingénierie des systèmes, l'analyse des opérations et la gestion.

Les systèmes d'aide à la décision sont utilisés dans de nombreux domaines tels que les affaires, la médecine, l'ingénierie, les sciences sociales, la planification urbaine et l'environnement. Dans le domaine du sport, l'aide à la décision est de plus en plus utilisée pour améliorer les performances des équipes et des joueurs.

Les systèmes d'aide à la décision pour le sport peuvent utiliser une variété de techniques d'analyse de données et de modélisation. Ces dernières peuvent inclure l'analyse de variance, la régression, la classification et le clustering. Les modèles peuvent être basés sur des techniques telles que les réseaux de neurones, les arbres de décision, la logique floue, les algorithmes génétiques et la programmation linéaire.

L'aide à la décision peut être utilisée pour une variété de tâches dans le sport, telles que l'analyse de la performance des joueurs et des équipes, la planification stratégique, la gestion des ressources, la prévision des résultats et l'analyse de la concurrence.

Dans les sections suivantes, nous détaillerons ces applications et les techniques utilisées.

Analyse de la performance des joueurs et des équipes : Elle consiste à collecter des données sur les performances des joueurs et des équipes et à les analyser pour identifier les forces et les faiblesses.

Planification stratégique : Les techniques d'aide à la décision peuvent être utilisées pour identifier les domaines où l'équipe doit se concentrer, pour évaluer les risques et les avantages des différentes options de stratégie et pour prévoir les résultats possibles.

Gestion des ressources : consiste à allouer les ressources disponibles de manière à maximiser l'efficacité de l'équipe. Les ressources peuvent inclure les budgets, le temps, les compétences des joueurs etc.

Prévision des résultats : Les techniques d'analyse de données peuvent être utilisées pour identifier les facteurs qui affectent le résultat des matchs, tels que le niveau de forme des joueurs et le style de jeu

Analyse de la concurrence : Les techniques d'analyse de données peuvent être utilisées pour collecter et analyser les données sur les performances des adversaires.

3.2 L'Aide à la décision et le Pronostic sportif

L'aide à la décision a de nombreuses applications pratiques dans le football. Voici quelques exemples de la façon dont elle peut être utilisée [3] :

- **Évaluation des joueurs :** Les équipes de football peuvent utiliser des outils d'analyse de données pour évaluer les performances des joueurs. Les entraîneurs peuvent utiliser ces statistiques pour évaluer les performances des joueurs et déterminer ceux qui ont le plus grand impact sur le terrain.
- **Analyse de l'adversaire :** Les équipes de football peuvent utiliser des outils d'analyse de données pour étudier les performances de l'équipe adverse. Ils peuvent analyser les statistiques des matchs précédents, les tactiques utilisées et les schémas de jeu.
- **Planification des entraînements :** Les équipes de football peuvent utiliser des outils d'analyse de données pour planifier les entraînements. Les entraîneurs peuvent utiliser les données sur les performances des joueurs pour identifier les domaines à améliorer.
- **Sélection des joueurs pour l'équipe :** Les équipes de football peuvent utiliser des outils d'analyse de données pour sélectionner les joueurs pour l'équipe. Les entraîneurs peuvent utiliser les statistiques des performances des joueurs pour évaluer leur impact sur le terrain et leur contribution à l'équipe.
- **Gestion des ressources :** Les équipes de football peuvent utiliser des outils d'analyse de données pour gérer leurs ressources. Les entraîneurs peuvent utiliser les données pour surveiller l'utilisation des ressources telles que les joueurs et les équipements.

L'intelligence artificielle (IA) a révolutionné l'aide à la décision dans le domaine sportif en permettant l'analyse de grandes quantités de données sportives et en fournissant des recommandations personnalisées pour prendre des décisions éclairées. Dans les sections suivantes nous allons détailler les différents concepts liés à l'IA.

4. L'Intelligence Artificielle (IA)

Identifier cette activité en pleine évolution ou encore définir précisément l'Intelligence Artificielle (IA) n'est pas aisé. Le but des premiers chercheurs était en fait de concevoir des machines intelligentes capables d'imiter les humains dans les tâches intellectuelles les plus complexes. Ayant vite compris les limites de ces ambitions, leurs successeurs ont poursuivi leur travail en s'attachant à trouver de nouvelles technologies informatiques (soft ou hard), des langages objets, ou des machines LISP dédiées à la programmation symbolique [4]

4.1. Définitions de l'Intelligence Artificielle

L'IA est une branche de l'informatique dont le but est de créer des systèmes intégrant de grandes quantités de connaissances et de processus, appelés systèmes intelligents. Ainsi, l'IA conduit à cet ensemble d'explications, où [4]

- C'est l'étude de techniques pour résoudre des problèmes exponentiellement difficiles en temps polynomial en exploitant les connaissances sur le domaine du problème.

Programmer les ordinateurs pour qu'ils exécutent les tâches les mieux présentées ou complétées par des humains [5]

- Utiliser des machines pour réaliser des tâches qui seraient considérées comme intelligentes si elles étaient faites par des humains [4]. Les machines sont considérées comme se comportant intelligemment, tout comme les humains.
- Automatiser les activités associées à l'esprit humain, telles que la prise de décision, la résolution de problèmes ou l'apprentissage [6].

L'IA est la partie de l'informatique qui se concentre sur l'automatisation des comportements intelligents [6].

4.2. Historique de l'intelligence artificielle

L'intelligence artificielle (IA) a une longue histoire qui remonte aux années 1950. L'IA est née de l'ambition de créer des machines capables de simuler l'intelligence humaine. Voici un bref résumé de l'évolution de l'IA au fil du temps : [4]:

Années 1950 : Les fondements de l'IA sont posés avec les travaux d'Alan Turing et de John McCarthy, qui introduit le terme "intelligence artificielle". Les premiers programmes d'IA sont créés pour résoudre des problèmes de logique et de mathématiques.

| CHAPITRE 1

Synthèse de l'état de l'art

Années 1960-1970 : L'optimisme entoure l'IA. Des progrès sont réalisés dans des domaines tels que la résolution de problèmes, le traitement du langage naturel et les systèmes experts.

Années 1980-1990 : L'IA symbolique est prédominante, avec l'utilisation de représentations symboliques et de règles logiques pour modéliser l'intelligence. Les systèmes experts sont largement utilisés dans des domaines spécifiques.

Années 1990-2000 : L'IA connaît un déclin, avec des attentes non satisfaites et un manque d'intérêt public et d'investissements.

Années 2000-2010 : L'apprentissage automatique, en particulier les réseaux neuronaux profonds, connaît une renaissance de l'IA. Les progrès dans la puissance de calcul, les jeux de données massifs et les algorithmes d'apprentissage permettent des avancées significatives dans des domaines tels que la vision par ordinateur, la reconnaissance vocale et la traduction automatique.

Années 2010-2022 : L'IA devient omniprésente dans notre vie quotidienne. Les assistants virtuels, les systèmes de recommandation, les véhicules autonomes et les systèmes de reconnaissance d'images sont largement utilisés. L'IA éthique et responsable, l'IA explicative et l'IA quantique émergent comme des domaines importants de recherche et d'exploration.

4.3. Les approches de l'IA

Premièrement, l'IA est considérée comme le seul moyen de déterminer le fonctionnement des humains en référence aux sciences cognitives. Mais pour agir intelligemment, peu importe comment une personne travaille. L'IA utilise l'ordinateur comme métaphore du cerveau. C'est l'approche dominante depuis 60 ans.

L'IA doit être forte lorsqu'il s'agit de machines dotées d'une intelligence humaine, et faible lorsqu'il s'agit de machines qui se comportent intelligemment mais moins que les humains.

4.4. Les défis de l'IA

L'IA est née de problèmes dans divers domaines. Ces problèmes sont [4]:

- Connaissance d'algorithmes pouvant ne pas être implémentés : problème d'explosion combinatoire.
- Traitement audio : Pas de solution algorithmique connue.
- Heuristique : En raison du savoir-faire, les experts de certains domaines s'appuient sur des connaissances tacites.

- L'intelligence artificielle s'est avérée importante pour tenter de résoudre les problèmes susmentionnés, nous aidant à prendre des décisions car elle gère :
 - L'acquisition et l'utilisation des connaissances,
 - Transmission d'informations,
 - Coopération.

4.5. Pourquoi des machines intelligentes ?

Une machine est dite intelligente si elle imite le comportement humain dans tous les domaines et modélise la fonction humaine. C'est pourquoi, nous construisons des machines intelligentes qui offrent de nouvelles possibilités et utilisent des logiciels relativement peu coûteux à développer et faciles à modifier pour mieux résoudre les problèmes [4]

4.6. Le raisonnement dans l'IA

Il existe plusieurs catégories générales de raisonnement qui peuvent être appliquées dans n'importe quel domaine [4]

- **Raisonnement déductif** : manifester des connaissances tacites,
- **Pensée abductive** : Rechercher les causes probables des symptômes,
- **Pensée inductive** : construire des lois, des concepts à partir d'exemples,
- **Pensée Analogique** : concilier des situations particulières
- **Justification hypothétique ou canonique** : Comblent les lacunes d'information avec des hypothèses explicites ou implicites.

Il existe d'autres catégories de raisonnements spécifiques adaptés à des applications très spécifiques cités dans ce qui suit [4]:

- **Arguments temporaires** : Gérer le temps pour planifier ou simuler des événements obsolètes, tels que la conduite d'ateliers de production.
- **Pensée spatiale et géométrique** : Par exemple, pour la robotique
- **Pensée causale** : Propager une cause pour trouver son effet ou pour revenir de l'observation d'un effet à une cause possible
- **Raisonnement qualitatif** : décrit le comportement d'un système physique en chiffres symboliques.

4.7. Les techniques de l'IA

Il existe les arguments scientifiques et philosophiques suivants sur l'intelligence [4]:

- Comment fonctionne le cerveau humain ?
- Peut-on extraire l'intelligence humaine dans des machines ?
- Les machines peuvent-elles vraiment être intelligentes ?

Dans tous les domaines d'activité, les technologies d'IA tendent à élargir l'amplitude de mouvement des machines, leur offrant la capacité de voir, d'entendre, de raisonner, de parler, d'agir, etc. Compréhension du langage, apprentissage, raisonnement, résolution de problèmes, etc. Nous essayons simplement de faire en sorte que les ordinateurs se comportent intelligemment grâce à des technologies telles que : les Systèmes experts, la logique floue, les algorithmes génétiques, les réseaux de neurones, etc. Il y a deux principes généraux [4]

1- L'étude des processus de la pensée humaine.

2- La représentation mécanique de ces procédés : ordinateur, robot.

Par conséquent, diverses techniques d'IA peuvent imiter la pensée humaine et partager des points communs qui est : l'aide à la décision.

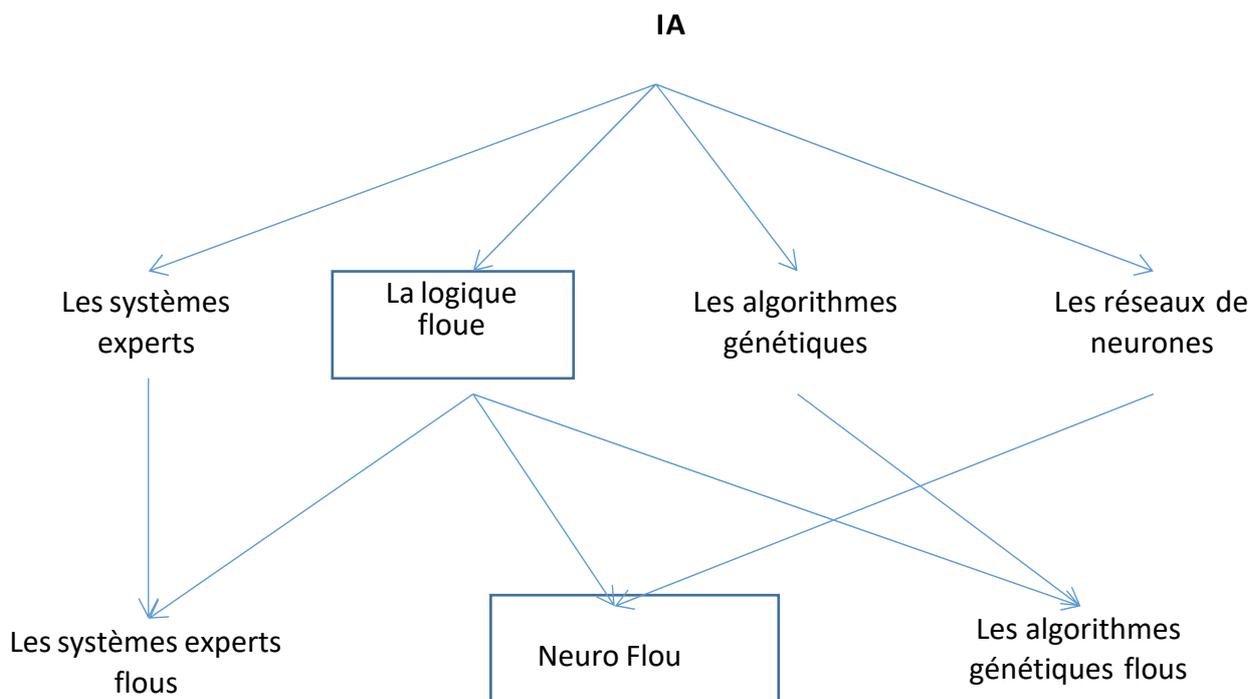


Figure 1: Les différentes techniques de l'IA

4.8. Domaines d'applications de l'Intelligence Artificielle

L'intelligence artificielle s'est développée parallèlement au calcul numérique, sans lui faire concurrence. Elle traite donc de domaines d'application considérés comme difficiles ou impossibles à résoudre numériquement. Entre autres, nous pouvons citer : L'algèbre informatique, la Vision : texte, image, la Robotique : plan, les Machines autonomes : percevoir, interpréter, décider, agir, le Langage naturel : traduire, comprendre, intégrer, la démonstration de théorème, les Jeux, la Représentation des connaissances dans divers domaines (droit, médecine, automatisme), etc.

Exemple : La robotique continue d'utiliser des algorithmes automatiques classiques pour le contrôle de bas niveau des mécanismes. L'intelligence artificielle intervient lorsque le robot a besoin d'interpréter l'évolution de l'environnement. Ce n'est pas le même geste répété qui peut fixer des objectifs différents (par exemple, replanifier la trajectoire si un "obstacle" survient [7])

4.9. L'apprentissage automatique (Machine Learning)

D'après L. Samuel (IBM, 1959) [7] « *L'apprentissage automatique (machine Learning) est le champ de l'IA qui permet à une machine (au sens large) d'apprendre. C'est-à-dire, d'améliorer progressivement ses performances sur une tâche spécifique en se basant sur des données, le tout sans être explicitement programmé pour résoudre cette tâche* ».

En particulier, l'apprentissage automatique est le concept selon lequel une machine explore la construction et l'étude d'algorithmes qui peuvent apprendre et faire des prédictions sur différents ensembles de données. En fonction du résultat de l'apprentissage automatique, nous pouvons distinguer différentes tâches de l'apprentissage automatique tels que : la classification, le clustering, la régression, etc. la figure suivante représente ces différentes tâches [8] :

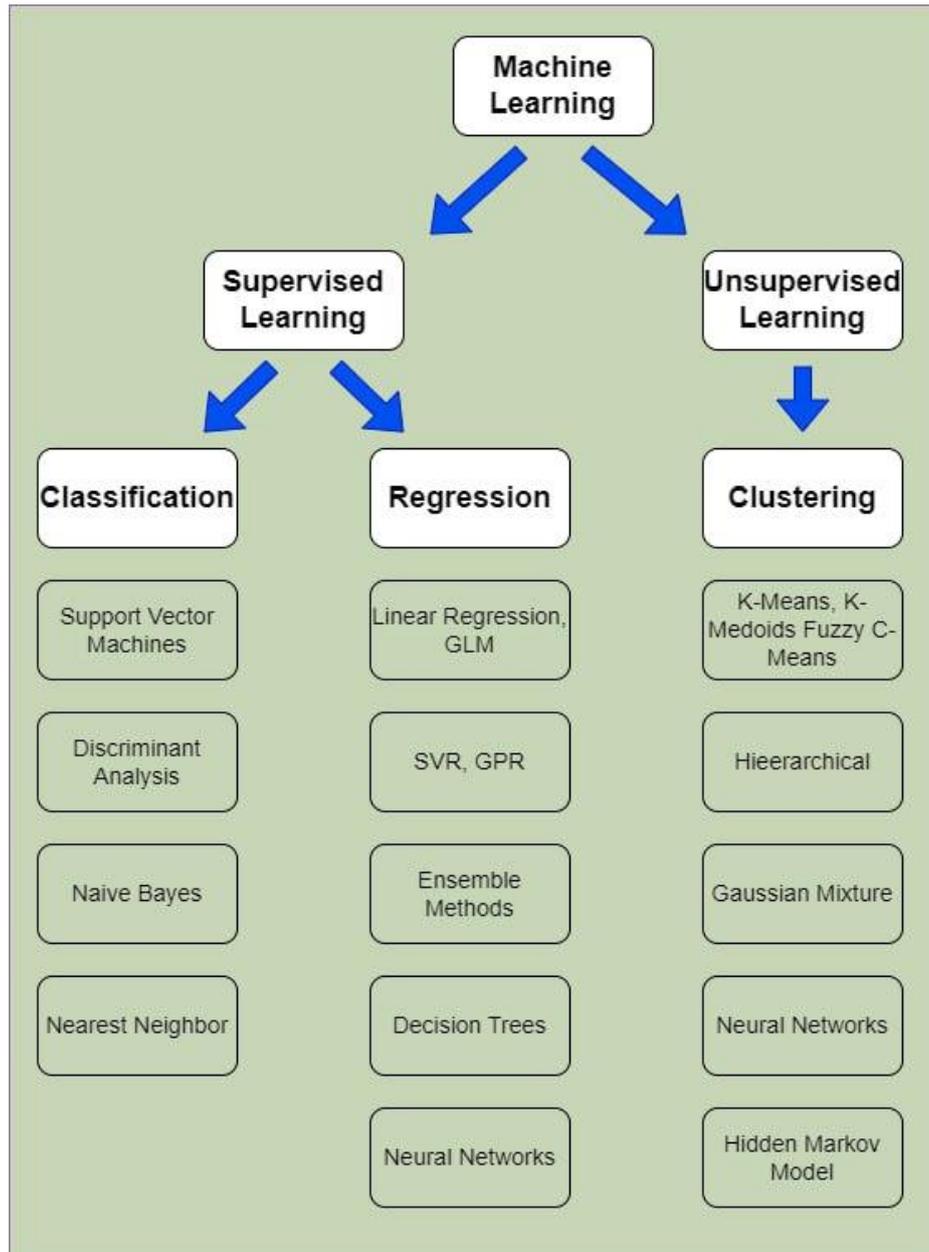


Figure 2:les techniques de l'apprentissage automatique

4.10. Les types d'apprentissage

Le problème de la classification peut être résolu en exploitant les approches décrites dans ce qui suit :

4.10.1. L'apprentissage supervisé

C'est une approche dans laquelle le programme informatique apprend les données qui lui sont transmises par l'utilisateur, puis utilise cet apprentissage pour classer une nouvelle observation. Cet ensemble de données peut simplement être bi-classe (comme identifier si la personne est un homme ou une femme ou si le courrier est du

spam ou non) ou il peut être multi-classes (reconnaissance vocale, reconnaissance de l'écriture manuscrite, identification biométrique, classification des documents, etc.) [8]

4.10.2. L'apprentissage non supervisé

C'est une technique d'apprentissage, où le système ne dispose que d'exemples, et que le nombre de classes et leur nature n'ont pas été prédéterminés. L'objectif de l'apprentissage non supervisé ou clustering est de découvrir des modèles cachés dans un ensemble de données pour classer les données brutes. L'apprentissage non supervisé est utilisé pour la détection d'anomalies, Exemples : les achats frauduleux de cartes de crédit, les systèmes de recommandation qui conseillent sur les produits à acheter, etc. [8]

4.10.3. L'apprentissage semi supervisé

Est une approche qui se situe entre l'apprentissage supervisé (utilise des données étiquetées) et l'apprentissage non-supervisé (utilise des données non-étiquetées). Il permet d'améliorer, significativement, la qualité de l'apprentissage avec une combinaison entre les données (étiquetées et non étiquetées).

4.10.4. L'apprentissage par renforcement

Il consiste à apprendre par interaction avec l'environnement dans lequel il doit atteindre un certain but à suivre dans une situation donnée. Par exemple : conduire un véhicule qui connaît le chemin dans un Labyrinthe et, en observant le résultat de certaines actions avec feedback sous forme de « récompenses » et de « punitions » pendant qu'il navigue dans l'espace du problème [8]

Voici un exemple sur l'utilisation de l'apprentissage par renforcement illustrée par la figure 3 :

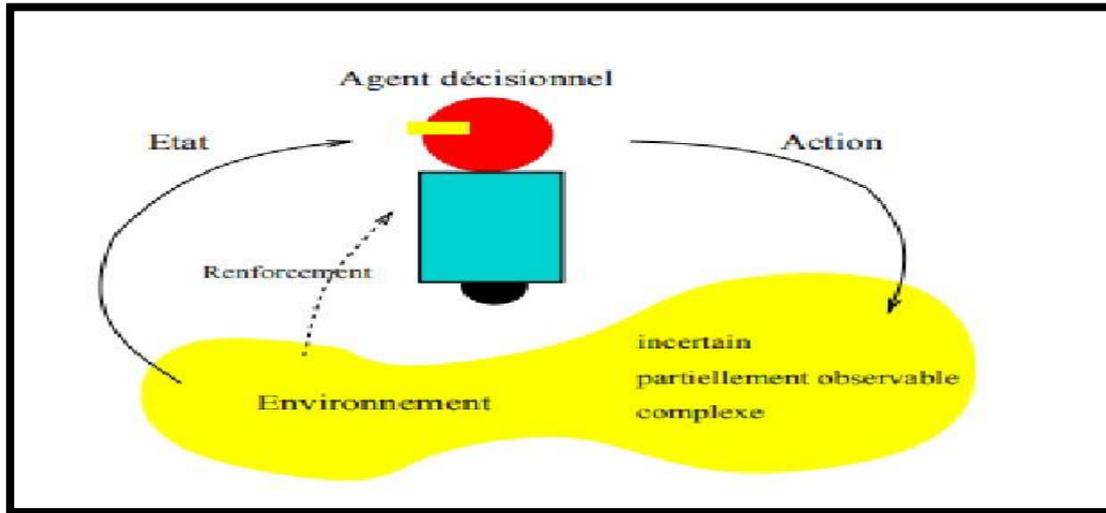


Figure 3: Exemple d'un agent utilisant l'apprentissage par renforcement.

4.11. Les techniques de l'apprentissage automatique

Les techniques d'apprentissage automatique sont appliquées dans un large éventail d'applications afin de résoudre un certain nombre de problèmes. Parmi les techniques les plus utilisées, nous détaillons dans les sections suivantes les plus populaires [8] :

4.11.1. K plus proche voisins (K-PPV)

L'algorithme K plus proche voisins (K-PPV) est un algorithme supervisé non paramétrique pour la classification. Cet algorithme est, largement, utilisé pour plusieurs tâches d'intelligence en classification, segmentation d'images et vidéo. Le principe de l'algorithme est de regrouper les échantillons selon leur voisinage, c'est-à-dire que chaque point est affecté à la classe la plus représentée parmi ses k voisins les plus proches. Cet algorithme est basé sur les deux éléments principaux suivants : [8]

- K : le nombre de cas les plus proches à utiliser.
- Une métrique pour mesurer le plus proche voisin(v).

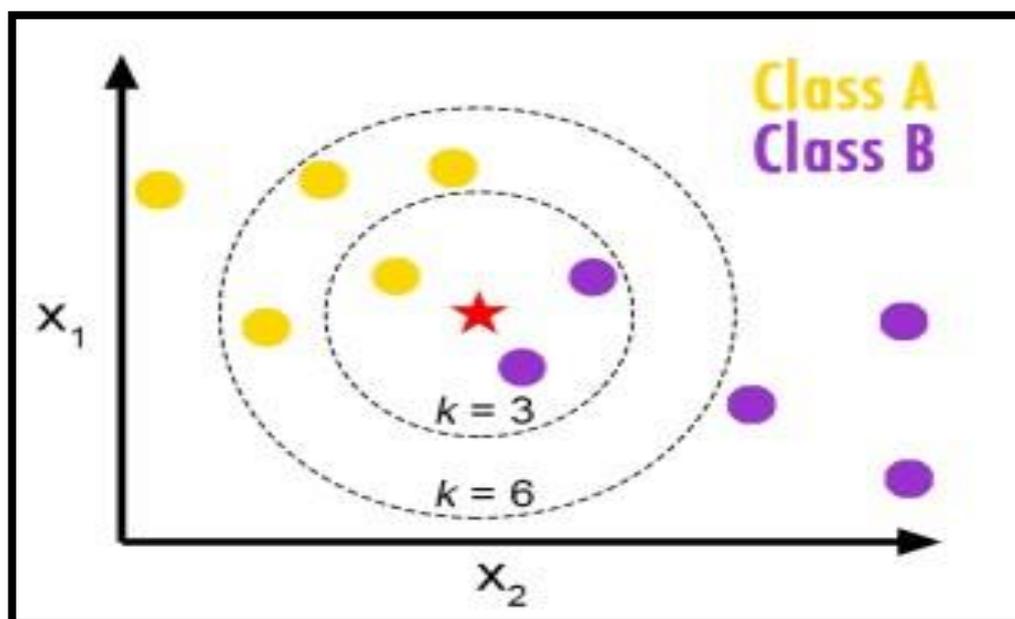


Figure 4: La tâche de la classification en utilisant l'algorithme K plus proche voisins

4.11.2. Support Vector Machine (SVM)

La technique « Support Vector Machine (SVM) » est développée par Vapnik en 1995 [9], c'est l'une des méthodes les plus populaires dans la famille des approches supervisées. Cette technique est destinée à résoudre des problèmes de classification ou de régression. Son but est de séparer les données d'apprentissage et maximiser la distance entre deux classes à partir d'un classificateur [10]. Cette technique est appliquée dans divers domaines, à savoir la reconnaissance des caractères, la reconnaissance des images et le diagnostic médical, etc.

4.11.3. Les arbres de décision (Decision Tree)

Les arbres de décision sont des modèles de ML supervisés, pouvant être utilisés pour la classification que pour la régression. Leur principe repose sur la construction d'un arbre de taille limitée. La racine constitue le point de départ de l'arbre et représente l'ensemble des données d'apprentissage. Les ensembles de données sont segmentées en plusieurs groupes possibles situés aux extrémités des branches (les « feuilles » de l'arbre), en fonction d'une variable discriminante (un des attributs).

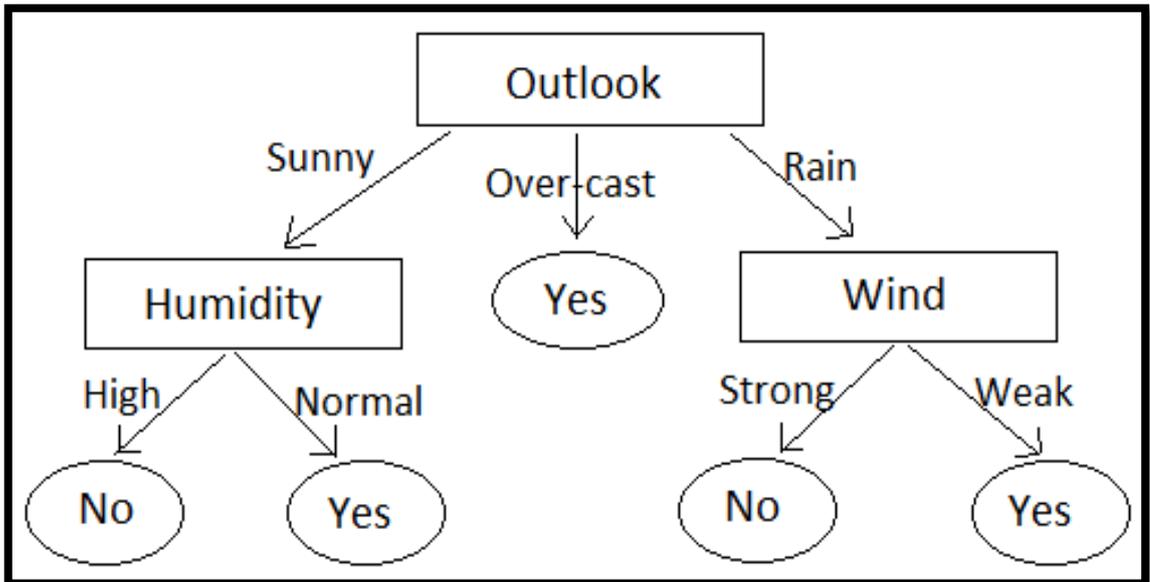


Figure 5: La classification en utilisant les arbres de décision

4.12. La classification

La classification consiste à déterminer une catégorie des nouvelles observations à partir des données existantes (training data). La classification est appliquée dans plusieurs domaines tel que : le diagnostic médical, le marketing ciblé, etc. La classification est une méthode mathématique d'analyse de données, elle est appliquée sur des données numériques (points, tableaux, images, sons, vidéo etc.), pour faciliter l'étude d'une population d'effectif important. [11]

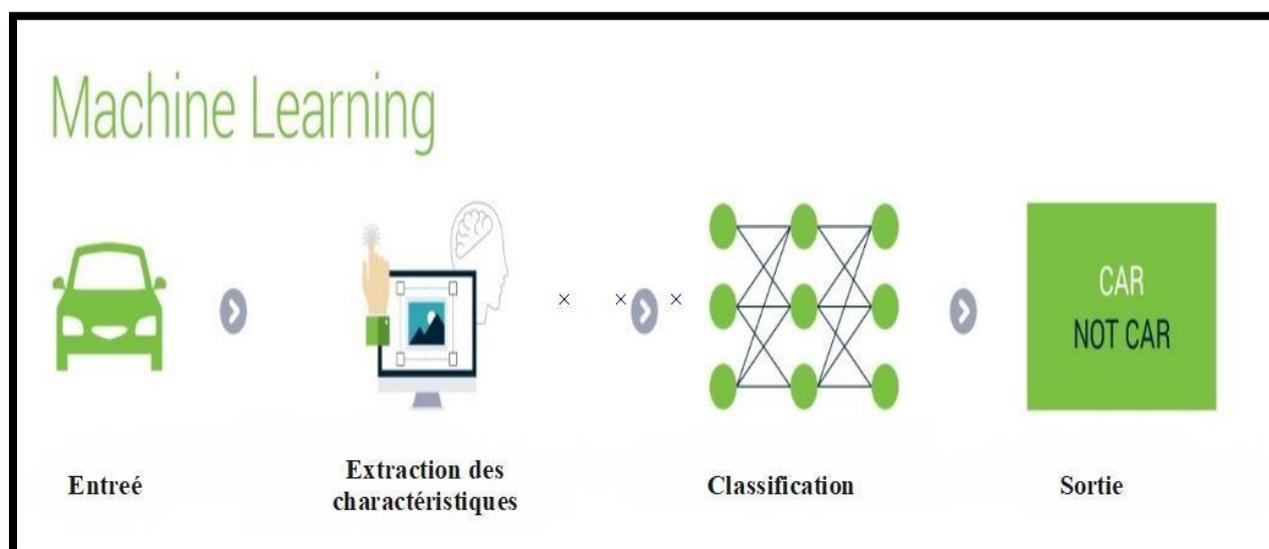


Figure 6: Le principe de la classification.

Comme son nom l'indique, la classification est obtenue en classant les objets dans des sous catégories à partir d'un groupe général, mais la machine peut-elle distinguer les personnes ou les animaux à l'aide des images, ou peut-elle distinguer et classer les textes en fonction de leur contenu.

La machine peut effectuer toutes les classifications à l'aide des données et les algorithmes appropriés, et donner des résultats précis en peu de temps. Toutefois, dans l'apprentissage automatique, la classification peut donner naissance à un problème d'identification d'une nouvelle observation à quel groupe de catégories (sous-populations) appartient, sur la base d'un ensemble de données de formation contenant des observations et dont les catégories sont connues. [8]

4.13. Les types de classification

4.13.1. Classification binaire

Les données entrantes sont classées dans l'une des deux catégories possibles. Par exemple pour le diagnostic d'une maladie, nous avons deux classes : malade {oui ou non}. [12]

4.13.2. Classification multi- classe

Les données sont divisées en plusieurs catégories possibles (supérieurs à 2). Ce type est très utile pour la grande masse de données. Exemple : la détection d'émotions selon le visage {Heureux, fatigué, triste, etc.}. [12]

4.14. L'Apprentissage profond

L'apprentissage profond est un type d'apprentissage automatique qui vise à former et à enseigner l'ordinateur à exécuter des fonctions humaines tels que la distinction d'objets visuels et l'identification du son et de l'image. Au lieu d'organiser les données, le deep Learning définit ses propres paramètres de base qui permettent à la machine d'apprendre de manière indépendante, et l'intérêt actuel pour le deep Learning est en partie dû à l'enthousiasme pour l'intelligence artificielle. Les techniques d'apprentissage approfondi ont amélioré la capacité de classer, de reconnaître, etc. [8]

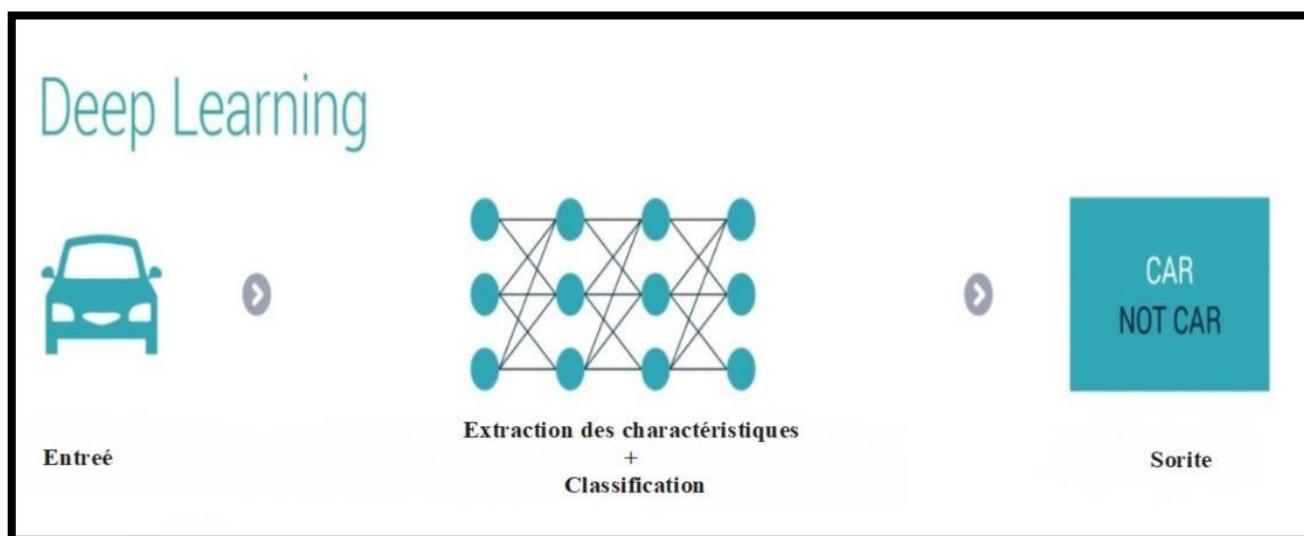


Figure 7: La classification dans l'apprentissage profond

Le deep Learning est une sous-catégorie du machine Learning, il concerne l'utilisation des réseaux de neurones pour améliorer des choses telles que la reconnaissance vocale, la vision par ordinateur et le traitement du langage naturel (tous des domaines particulièrement complexes pour les chercheurs en IA).

4.14.1. Fonctionnement de l'apprentissage profond

Le deep Learning est basé sur un réseau neuronal artificiel inspiré du cerveau humain. Ce réseau est composé de plusieurs « couches » de neurones, chacune recevant et interprétant des informations de la couche précédente. Par exemple, le système apprendra à déterminer s'il y a un visage sur une photo avant de découvrir quelle personne.

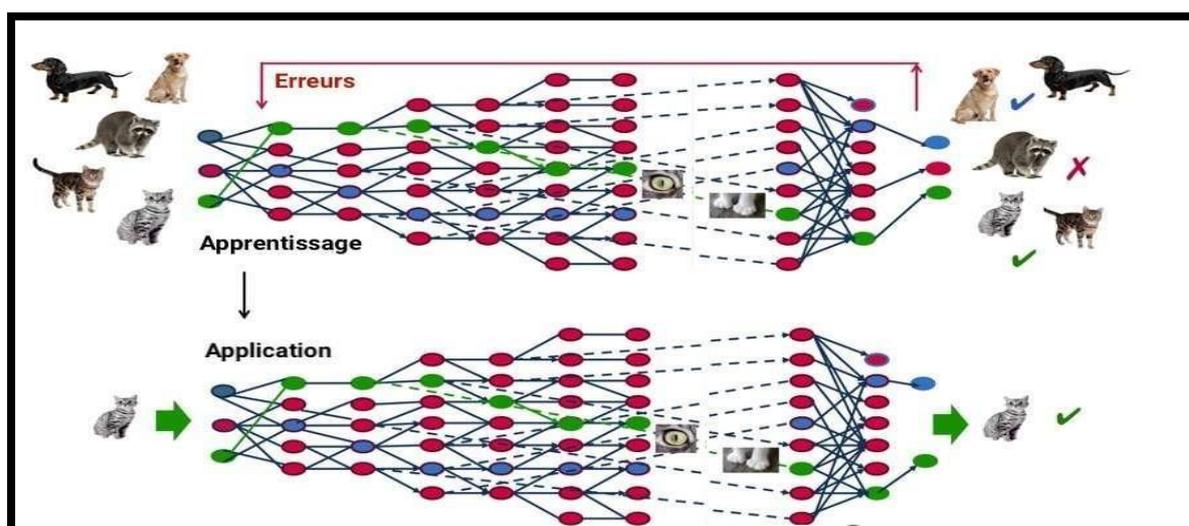


Figure 8:Exemple de fonctionnement de Deep Learning

Le Deep Learning découvre une structure complexe dans de grands ensembles de données en utilisant l'algorithme de propagation inverse pour indiquer comment une machine doit modifier ses paramètres internes qui sont utilisés pour calculer la représentation dans chaque couche à partir de la représentation dans la couche précédente. [13]

4.14.2. Architectures d'apprentissage profond

Il existe plusieurs cadres d'apprentissage en profondeur largement utilisés, tels que le réseau neuronal profond, le réseau neuronal convolutionnel (CNN), le réseau neuronal récurrent (RNN), et le LSTM décrits dans ce qui suit [8] :

4.14.2.1. Les réseaux Long Short Memory (LSTM)

Les réseaux de mémoire à court terme sont un type de réseau neuronal récurrent (RNN) introduit par Hochreiter et Schmidhuber en 1997, une technique capable d'apprendre et de se souvenir sur de longues séquences de données d'entrée.

Les réseaux LSTM contiennent leurs informations dans une mémoire (qui est très similaire à la mémoire d'un ordinateur). Cette mémoire peut être considérée comme une cellule fermée, où fermée signifie que la cellule décide de stocker ou de supprimer des informations de sa mémoire. Dans un LSTM vous avez trois portes :

- Porte d'entrée (input gate) : pour rôle d'extraire l'information de la donnée courante

| CHAPITRE 1

Synthèse de l'état de l'art

- Porte d'oubli (forget gate): Cette porte décide de quelle information doit être conservée ou jetée
- Porte de sortie (output gate) : décider de quel sera le prochain état caché, qui contient des informations sur les entrées précédentes du réseau et sert aux prédictions. Les portes d'un LSTM sont analogiques, modélisées par une fonction qui est généralement un sigmoïde. Ce sigmoïde (signifie qu'ils vont de 0 à 1) est appliqué à la somme pondérée des entrées (en bleu), des sorties (en vert) et de la cellule (en orange), par des poids. Vous pouvez voir une illustration d'un LSTM avec ses trois portes ci-dessous : [14]

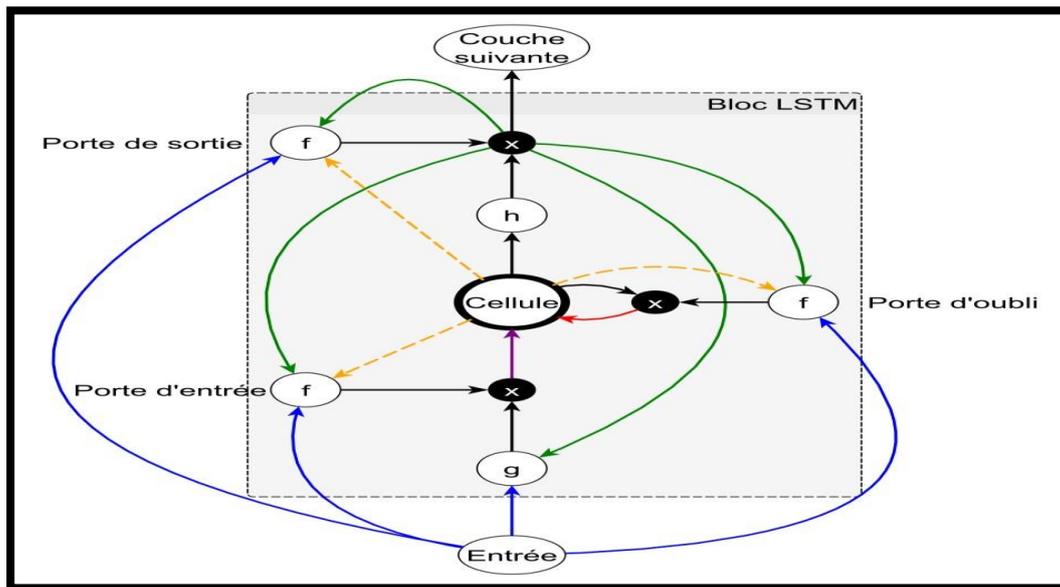


Figure 9: Fonctionnement d'un réseau LSTM

4.14.2.2. Les réseaux convolutifs (CNN)

Un CNN (réseau neuronal convolutionnel) est très similaire aux réseaux neuronaux de base. Il est composé de neurones avec des poids et des biais [5]. Il se compose de plusieurs couches permettant de reconnaître des objets sur une image :

- La couche convolution : qui tire d'une image ses caractéristiques principales,
- La couche ReLU : qui crée des modèles non linéaires et augmente la vitesse de formations du modèle,
- La couche Pooling : qui permet de réduire la taille de l'image et donc le nombre d'informations pour ne garder que le plus important
- La couche Entièrement connectée (fully-connected) : permet de classifier l'image en entrée d'un réseau (représentée sous forme d'un vecteur), pour indiquer la probabilité de chaque neurone dans une image [15].

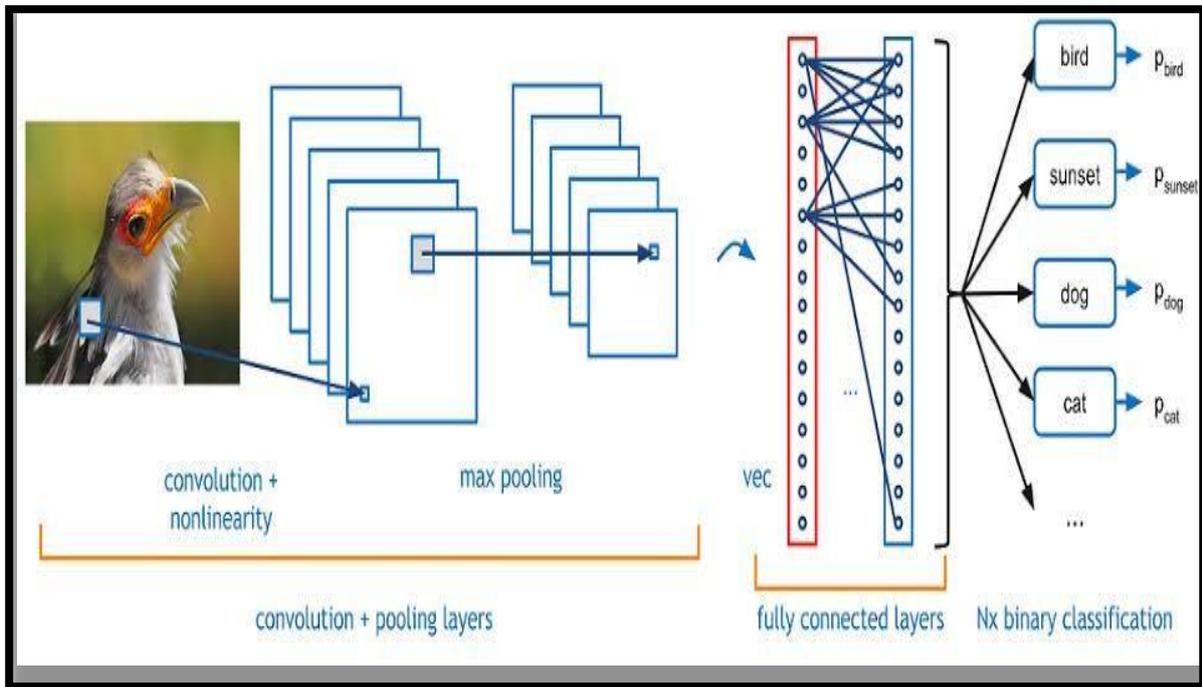


Figure 10: Fonctionnement d'un réseau convolutif (CNN)

5. Travaux connexes

Le tableau suivant synthétise les travaux les plus pertinents dans la littérature qui ont abordé l'utilisation de l'intelligence artificielle pour aider les équipes de football à améliorer leurs performances en utilisant des techniques telles que l'analyse de données, la reconnaissance d'image, la simulation de match, l'apprentissage par renforcement, le traitement du langage naturel et la vision par ordinateur,

| CHAPITRE 1

Synthèse de l'état de l'art

Référence du travail	Techniques	Description	Domaine d'utilisation	Date
Lefigaro[46]	Analyse des données	Utilisation de l'apprentissage automatique pour analyser les données de performance des joueurs et des équipes afin de prendre des décisions plus informées.	Analyse de données de match	2019
francetvinfo [47]	Reconnaissance d'image	Utilisation de la reconnaissance d'image pour identifier les joueurs lors de la surveillance des matchs en direct.	Surveillance de match en direct	2021
Futura sciences [48]	Simulation	Utilisation de l'IA pour simuler des matchs afin d'aider les entraîneurs à planifier leur stratégie.	Simulation de match	2022
strategie.gouv [49]	Apprentissage par renforcement	Utilisation de l'apprentissage par renforcement pour entraîner des agents d'IA à jouer au football et à prendre des décisions sur le terrain.	Simulation de match	2018
larevueia [50]	Traitement du langage naturel	Utilisation du traitement du langage naturel pour analyser les commentaires et les critiques de matchs pour aider les entraîneurs à prendre des décisions plus informées.	Analyse de commentaires et de critiques de matchs	2021-2022
economie.gouv [51]	Vision par ordinateur	Utilisation de la vision par ordinateur pour suivre les mouvements des joueurs sur le terrain et collecter des données sur leur performance.	Analyse de données de match	2023

Tableau 1 : Travaux connexes

6. Conclusion

Dans ce chapitre, nous avons parcouru l'histoire de l'informatique dans le football, découvert l'aide à la décision et présenté les fondements de l'Intelligence Artificielle, ainsi que ses différentes approches et domaines d'application.

Nous avons, également, exploré les techniques de l'apprentissage automatique et les différents types d'apprentissages tels que l'apprentissage supervisé et l'apprentissage non supervisé. Nous avons constaté que l'utilisation de l'IA dans le domaine du pronostic sportif a le potentiel de révolutionner l'industrie et de fournir des informations précieuses aux entraîneurs, aux joueurs, aux parieurs et aux fans de football.

Le chapitre suivant sera consacré à la conception et la modélisation de notre système intelligent d'aide à la décision proposé pour la prédiction des résultats de match.

CHAPITRE 2 CONCEPTION ET
MODÉLISATION DU SYSTÈME
PROPOSÉ

1. Introduction

Dans le chapitre précédent, nous avons exploré différents concepts liés à l'intelligence artificielle, en nous appuyant sur des travaux connexes issus de la littérature, qui partagent des similitudes avec notre projet.

Dans ce chapitre, nous commencerons par présenter les principaux objectifs de notre système, ainsi que les données et les méthodes utilisées par nos modèles. Ensuite, nous aborderons l'utilisation des concepts de l'UML (Unified Modeling Language) pour la conception et la modélisation de notre système intelligent.

2. Objectifs visés

L'objectif principal visé dans la présente étude est la conception et l'élaboration d'un système intelligent d'aide à la décision dans le football.

Ce système offre aux jeunes, un moyen pour se cultiver tout en jouant, il leurs permet de prédire les équipes gagnantes et les motiver avec des cadeaux intéressants selon les scores qu'ils peuvent accumuler. Cette tâche de prédiction est assurée par l'utilisation des techniques d'apprentissage automatique.

3. Étude de l'existant

Dans cette section, nous abordons quelques outils de prédiction existant dans le domaine du pronostic sportive et l'aide à la décision

3.1 Predict395

Predict395 est une filiale de TRUSTSCRIN LTD, une société enregistrée en Angleterre et au Pays de Galles sous le numéro d'entreprise 14193888. TRUSTSCRIN LTD fournit une analyse approfondie des données à l'aide de l'intelligence artificielle et de l'apprentissage automatique.

Predict395 promeut et soutient les conseils de paris responsables. Avec le système d'IA expert et robuste de Predict395, votre prédiction est bien analysée. [16]

3.1.1. Le site web Predict395

Predict395 est un site Web qui fournit des pronostics et des statistiques de football gratuits. Le site Web a été fondé en 2015 par Goddy Ezeh, un passionné de football nigérian. Predict395 offre une variété de fonctionnalités, notamment : [16]

- Prédiction gratuites de football pour toutes les ligues et compétitions majeures ;
- Statistiques détaillées sur les matchs passés et en cours ;
- Transfert des nouvelles et des analyses ;
- Astuces et conseils pour les paris.

Predict395 dispose d'une équipe d'analystes expérimentés qui utilisent une variété de données et de statistiques pour générer leurs prédictions. Le site Web dispose, également, d'un forum communautaire où les utilisateurs peuvent discuter de football et partager leurs pronostics.

Il a été félicité pour ses prévisions précises et ses statistiques complètes. Cependant, certains utilisateurs ont critiqué le site Web pour son manque de transparence et sa dépendance au marketing d'affiliation.

La figure suivante illustre l'interface principale de Predict395.



Figure 11:Le site web Predict395

Predict395 présente les limites suivantes [16]:

- Manque de transparence ;
- Dépendance au marketing d'affiliation ;
- Certains utilisateurs ont signalé des prédictions inexactes ;

Dans l'ensemble, Predict395 est une ressource précieuse pour les fans de football. Le site Web fournit des pronostics et des statistiques gratuits qui peuvent être utiles pour parier et simplement suivre le match. Cependant, les utilisateurs doivent être conscients des limites du site Web et doivent utiliser ses prédictions avec prudence.

3.2 1xBet

1xBet est une société de jeu en ligne fondée en 2007 et enregistrée à Chypre. Il est autorisé par la licence Curaçao eGaming et opère dans plus de 100 pays. 1xBet propose une grande variété de produits de jeu, notamment des paris sportifs, des jeux de casino et des paris en direct. La société est connue pour ses cotes élevées, ses bonus généreux et son large éventail de marchés de paris. [17]

3.2.1. Le site web 1xBet

1xBet est un site web de pronostics sportifs qui propose une plateforme en ligne permettant aux utilisateurs de parier sur une large gamme d'événements sportifs. Le site offre des prix et des cotes pour de nombreux sports populaires tels que le football, le tennis, le basketball, le hockey sur glace, et bien d'autres. Les utilisateurs peuvent placer des paris sur les résultats des matchs, les scores, les performances individuelles des joueurs, les statistiques et d'autres aspects liés aux événements sportifs. 1xBet offre également des fonctionnalités telles que le streaming en direct des matchs, les résultats en direct, les analyses statistiques et des promotions pour les parieurs [17]

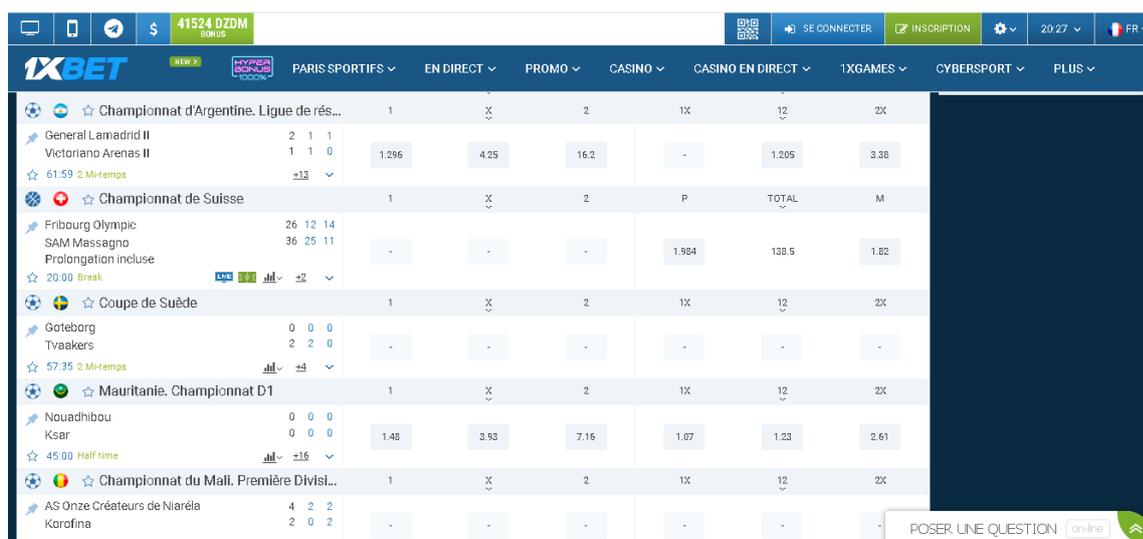


Figure 12: Le site web 1xBet

3.2.2. Les limites de 1xBet

Cet outil présente les limites suivantes :

- **Légalité et réglementation** : La légalité des activités de jeu en ligne peut varier selon les pays et les juridictions. Certains pays peuvent avoir des lois strictes régissant les jeux d'argent en ligne, ce qui peut rendre 1xBet inaccessibles ou illégaux dans ces régions. Il est donc important de vérifier la réglementation locale avant de s'engager sur le site.
- **Problèmes de fiabilité** : Certains utilisateurs ont signalé des problèmes de fiabilité avec 1xBet, notamment des retards de paiement, des problèmes de retrait

D'après ce que nous avons vu dans les deux sites précédents, nous allons proposer un système qui englobe les deux principes des deux sites, la prédiction des matchs et en plus d'utiliser le concept de gaming (de jouer en utilisant l'application mobile). Nous allons détailler et décrire le système proposé dans les sections suivantes.

4. Architecture générale du système proposé

L'architecture illustrée par la figure suivante représente de manière symbolique et schématique les différents composants d'un ou plusieurs systèmes informatiques, ainsi que leurs relations et interactions. Contrairement aux spécifications de l'analyse fonctionnelle, les modèles architecturaux produits lors de la phase de conception ne sont pas ce que le système informatique doit réaliser, mais plutôt comment il doit être conçu pour répondre aux spécifications. L'analyse se concentre sur le "quoi faire", tandis que l'architecture se concentre sur le "comment le faire" [NET 21].

La figure suivante présente l'architecture globale du système proposé, suivie par des détails sur chacun des sous-systèmes qui le composent.

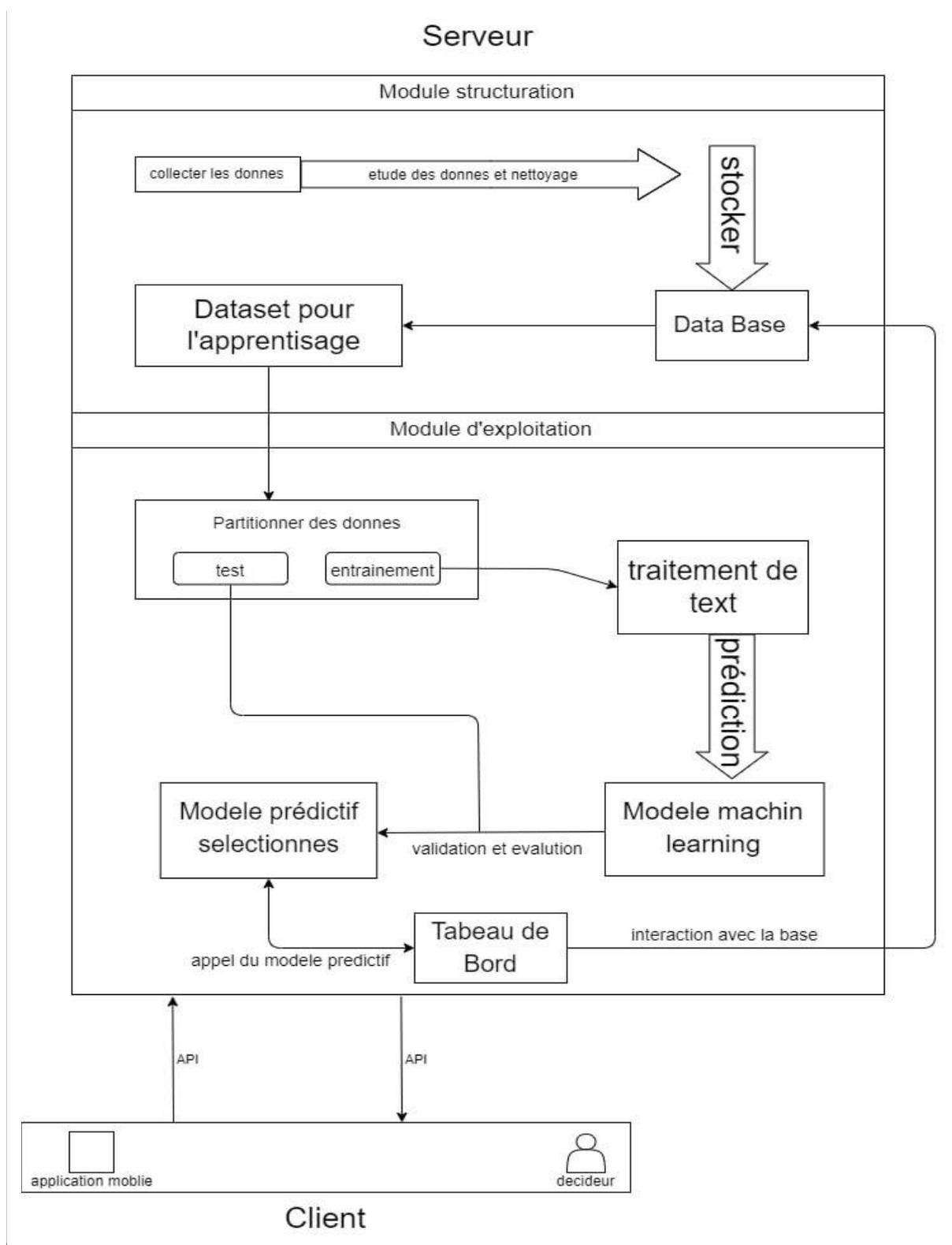


Figure 13:Architecture globale du système proposé

La figure précédente illustre que le système proposé est composé de trois modules principaux, chacun étant dédié à un traitement spécifique. Les détails de ces modules sont décrits ci-dessous :

4.1.Module de structuration

La structuration des données est la première étape intégrée dans l'architecture du système proposé, cette phase est primordiale pour faire convenablement le traitement nécessaire pour les prochaines étapes.

Dans cette phase, nous réalisons une analyse approfondie des données recueillies, suivie d'une étape de nettoyage. Le nettoyage des données nous permet de préparer la base de données qui sera utilisée pour l'apprentissage machine.

4.2.Module d'exploitation

La phase d'exploitation représente le noyau du fonctionnement du système. Cette phase est décomposée en deux grandes étapes à savoir :

1) **Le traitement de texte et apprentissage automatique** : il s'agit de la première étape de la phase d'exploitation. Durant cette étape, nous procédons au traitement des données et le partitionnement, pour ensuite exploiter les méthodes de *machine learning* ainsi que la validation du modèle choisi, afin de générer un modèle prédictif capable de prédire les résultats de matchs.

2) **L'élaboration de Interface IHM** : dans cette étape, nous allons créer un tableau de bord informatique qui prendra la forme d'un site web. Les utilisateurs ont la possibilité de créer un profil en fonction de leur cas d'utilisation, de consulter la base de données, de publier et de prédire les résultats d'un match en utilisant le modèle prédictif développé.

4.3.Module de concrétisation

Les utilisateurs de l'application mobile pourront prendre part des résultats de la prédiction des matchs en se basant sur les résultats prédits et affichés sur le tableau de bord et cela à travers des API.

Concernant la description de notre dataset et le composant d'apprentissage automatique, plus de détails seront donnés dans les sections suivantes.

5. À propos des données utilisées

Dans cette section, nous allons décrire avec détails les données exploitées par notre système.

5.1. Description des données récoltées

Les données utilisées sont disponibles dans des sites spécifiques ou à travers des API, cependant ces API sont payantes, pour remédier à cela nous avons utilisé la méthode scraping qui est manipulée avec Python, cette méthode permet de récupérer les données souhaitées des pages web gratuitement grâce à la bibliothèque BeautifulSoup.

Après la récupération des données, nous les avons classées dans un fichier CSV, ce fichier contient exactement cinq ligues : "Premier-League ", "La-Liga", "Ligue-1", "Bundesliga", "Serie-A" et toutes leurs années commençant par l'année 2005 jusqu'à l'année 2022.

5.2. Description des attributs

Le dataset contient les attributs décrits par le tableau suivant :

Variable	Description
'Comp'	Ligues ("Premier-League ", "La-Liga", "Ligue-1", "Bundesliga", "Serie-A")
'Season'	année de ligues
'Date'	Match Date (dd/mm/yy)
'HomeTeam'	Home Team (équipe à domicile)
'AwayTeam'	Away Team (Équipe extérieure)
'FTHG'	Full Time Home Team Goals (Objectifs de l'équipe à domicile à temps plein)
'FTAG'	Full-Time Away Team Goals (Objectifs de l'équipe à l'extérieur à temps plein)
'HTHG'	- Half Time Home Team Goals (Buts de l'équipe à domicile à la mi-temps)
'HTAG'	Half-Time Away Team Goals (Buts de l'équipe à l'extérieur à la mi-temps)
'FTR'	Full-Time Result (H=Home Win, D=Draw, A=Away Win) (Résultat à temps plein)
'HS'	Home Team Shots (Tirs de l'équipe à domicile)
'AS'	Away Team Shots (Tirs de l'équipe à l'extérieur)

'HST'	Home Team Shots on Target (Tirs cadrés de l'équipe à domicile)
'AST'	Away Team Shots on Target (Tirs cadrés de l'équipe à l'extérieur)
'HR'	Home Red Card (carte rouge pour l'équipe à domicile)
'AR'	Away Red Card (carte rouge pour l'équipe à l'extérieur)

Tableau 2: Description des attributs

5.3. Prétraitement et nettoyage des données

Le prétraitement des données est une phase très importante car elle permet de transformer des données brutes en un format compréhensible par les modèles d'apprentissage automatique.

De plus, les tâches de prétraitement et de nettoyage permettent de modifier ou d'enrichir notre base de données, créer de nouvelles variables et fournir des informations supplémentaires aux algorithmes d'apprentissage automatique.

Par la suite, effectuer la structuration du dataset sous forme Excel (fichier CSV) afin de le préparer à la phase d'apprentissage en faisant l'entraînement du modèle.

Étant donnée la qualité des données nécessaires, il était impératif pour nous de procéder comme suit :

- Comprendre et maîtriser les données récoltées ;
- Trier les données selon la catégorie ;
- Supprimer les redondances et les informations incomplètes ;
- Effectuer la structuration du dataset sous Excel

5.4. Encodage des données

Cette étape revêt une grande importance dans le processus d'apprentissage automatique, généralement après l'étape de prétraitement et de nettoyage des données. Au cours de cette dernière étape, nous utilisons des méthodes pour encoder les classes en valeurs numériques, afin de pouvoir appliquer des modèles de classification prédictive par la suite.

Nous avons choisi la méthode d'encodage *Label Encoding* pour traiter la variable "Team". Cette méthode consiste à attribuer une valeur numérique à chaque classe de la variable catégorielle et les classées par ordre alphabétique.

5.5. Augmentation des données

L'augmentation des données « Data Augmentation ou upsampling » est une technique souvent utilisée dans l'apprentissage machine, cette technique permet d'augmenter le nombre de données dans le dataset. Notre objectif consiste à réaliser deux modèles de prédiction : Le premier avant le match et le deuxième après la fin de la première mi-temps du match (45 min après le début du match).

Pour pouvoir mettre cela en œuvre, nous avons besoin d'autres données complémentaires aux données dont nous disposons, pour cela nous avons programmé des fonctions sur Python qui nous ont fourni les données suivantes :

Variable	Description
HTGS	le nombre de buts marqués par l'équipe durant tous les matchs de l'année.
ATGS	le nombre de buts marqués par l'équipe adverse durant tous les matchs de l'année.
HTGC	le nombre de buts marqués contre l'équipe durant tous les matchs de l'année
ATGC	le nombre de buts marqués contre l'équipe adverse durant tous les matchs de l'année
HTP	le nombre de points accumulés par l'équipe
ATP	le nombre de points accumulés par l'équipe adverse.
HTFormPtsStr	Résultats des cinq matchs choisis de l'historique (W : win, D : draw, L : Lost, M : pas en course) EX : LDWLD, Cela nous aidera à faire des prévisions pour évaluer le niveau de l'équipe et voir comment elle évolue.
HM1...5	HTFormPtsStr : Nous avons divisé ce dernier en 5 catégories comme suit : HM1 : L, HM5 : D, AM1...5 : L'adversaire.
ATFormPtsStr	L'adversaire.
HTFormPts	(win : 3pts, D : 1pts, L : 0pts) : La différences des points accumulés dans les cinq matchs, exemple :0+1+3+0+1
ATFormPts	l'adversaire.
MW	le nombre de matchs (1...38).
HTGD	différence : (HTGS – HTGC) /MW la différence entre les buts marqués par et contre l'équipe

ATGD	l'adversaire.
DiffPts	La différence des points accumulés par l'équipe et par son adversaire (HTP – ATP) /MW
DiffFormPts	La différence de points entre l'équipe et son adversaire dans les cinq derniers match (HTFormPts – ATFormPts) /MW

Tableau 3: Les données augmentées

5.6.Sélection de variables (attributs) significatifs

Pour pouvoir manipuler les bonnes données nécessaires, nous devons utiliser des méthodes telle que la matrice de corrélation pour dataset :

| CHAPITRE 2

CONCEPTION ET MODÉLISATION DU SYSTÈME PROPOSÉ

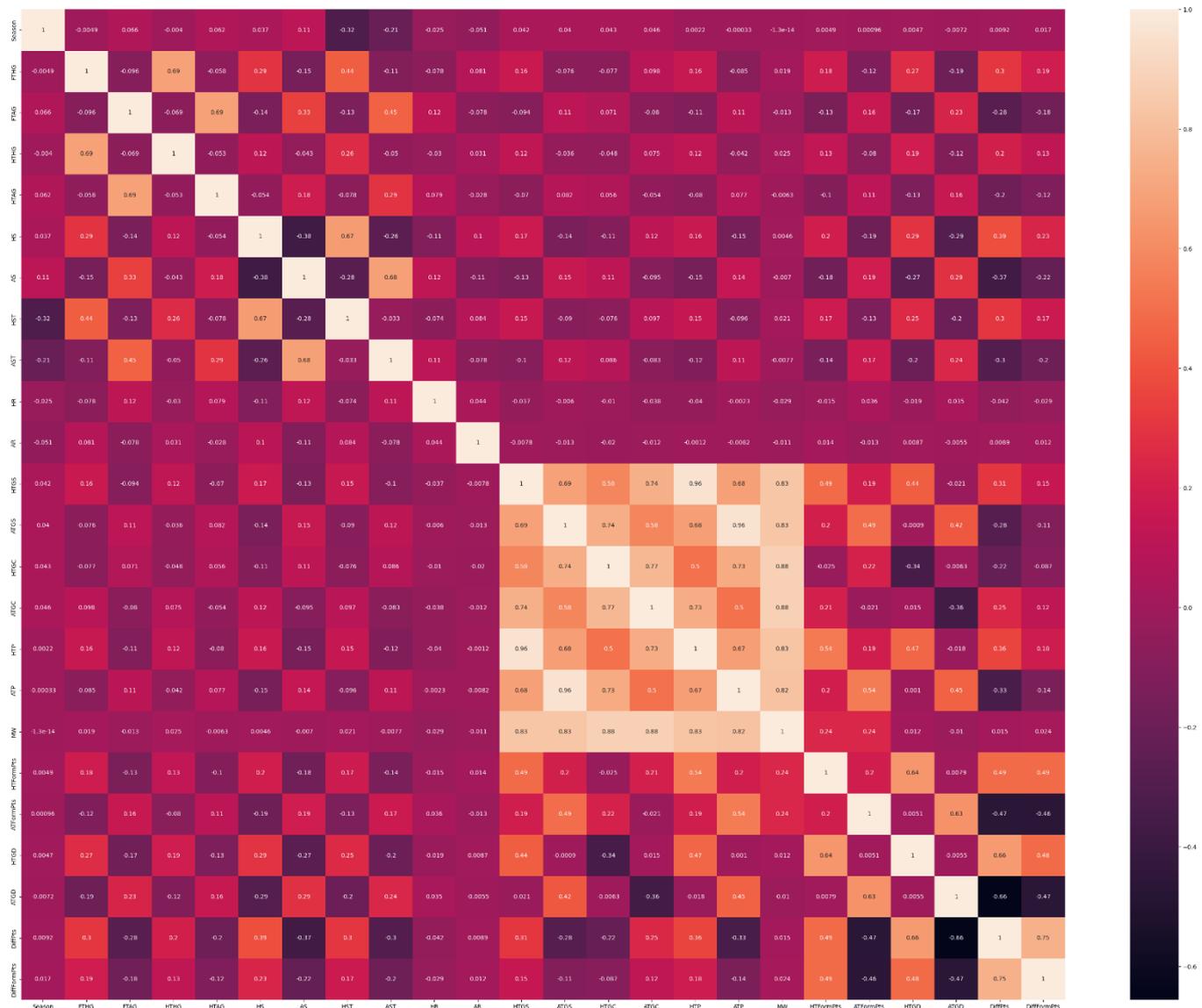


Figure 14: La matrice de corrélation

- **Corrélation positive** : Si une augmentation de la caractéristique « A » entraîne une augmentation de la caractéristique « B », alors elles sont positivement corrélées. Une valeur 1 signifie une corrélation positive parfaite.
- **Corrélation négative** : Si une augmentation de la caractéristique « A » entraîne une diminution de la caractéristique « B », alors elles sont négativement corrélées. Une valeur -1 signifie une corrélation négative parfaite.

Deux caractéristiques sont hautement ou parfaitement corrélées, de sorte que l'augmentation de l'une entraîne une augmentation de l'autre. Cela signifie que les deux entités contiennent des informations très similaires et qu'il y a très peu ou pas de

variance dans les informations. Ceci est connu sous le nom de Multi-Colinéarité car les deux contiennent presque les mêmes informations.

Alors nous devons utiliser les deux car l'un d'eux est redondant. Lors de la création ou de la formation de modèles, nous devrions essayer d'éliminer les fonctionnalités redondantes car cela réduit le temps de formation et de nombreux avantages similaires.

Après tout cela, nous avons généré les variables suivantes:

- 'HomeTeam', 'AwayTeam', 'FTR', 'HTP', 'ATP', 'HM1', 'HM2', 'HM3', 'AM1', 'AM2', 'AM3', 'HTGD', 'ATGD', 'DiffFormPts', pour le modèle « avant le match ».
- 'HomeTeam', 'AwayTeam', 'FTR', 'HTP', 'ATP', 'HTHG', 'HTAG', 'HS', 'AS', 'HST', 'AST', 'HR', 'AR', 'HM1', 'HM2', 'HM3', 'AM1', 'AM2', 'AM3', 'HTGD', 'ATGD', 'DiffFormPts', pour le modèle « mi-temps match ».

5.7.Approche basée sur la classification ou la régression

Cette approche consiste à prédire une variable Y à partir d'une autres variables X.

Nous avons réalisé les deux modèles suivants :

- 'HomeTeam', 'AwayTeam', 'HTP', 'ATP', 'HM1', 'HM2', 'HM3', 'AM1', 'AM2', 'AM3', 'HTGD', 'ATGD', 'DiffFormPts', nous les mettons dans la variable X.

Par la suite, nous prédirons la variable Y qui contient 'FTR' : (H : home win, NH : Home no win).

- Concernant le deuxième model nous rajoutons à la variable ces données :

'HTHG', 'HTAG', 'HS', 'AS', 'HST', 'AST', 'HR', 'AR'.

Avant d'entamer l'entraînement et la génération du modèle de prédiction, nous avons effectué un encodage sur ces variables car elles sont de type qualitatif.

L'algorithme suivant décrit en détails les étapes de prétraitement des données.

Algorithme : Prétraitement des données

Entrées : 'HomeTeam', 'AwayTeam', 'HTP', 'ATP', 'HTHG', 'HTAG', 'HS', 'AS', 'HST', 'AST', 'HR', 'AR', 'HM1', 'HM2', 'HM3', 'AM1', 'AM2', 'AM3', 'HTGD', 'ATGD', 'DiffFormPts' ; *'Voir tableau 3'*

Sorties : FTR : Full-Time Result (H=Home Win, NH=Away Win) (Résultat à temps plein)

Début

1. Charger les données
2. Sélectionner les variables pour chaque modèle
 - Pour le modèle "avant le match":
 - 'HomeTeam', 'AwayTeam', 'HTP', 'ATP', 'HM1', 'HM2', 'HM3', 'AM1', 'AM2', 'AM3', 'HTGD', 'ATGD', 'DiffFormPts'
 - Pour le modèle "mi-temps match":
 - 'HomeTeam', 'AwayTeam', 'HTP', 'ATP', 'HTHG', 'HTAG', 'HS', 'AS', 'HST', 'AST', 'HR', 'AR', 'HM1', 'HM2', 'HM3', 'AM1', 'AM2', 'AM3', 'HTGD', 'ATGD', 'DiffFormPts'
3. Effectuer l'encodage des variables qualitatives
 - Identifier les variables qualitatives dans chaque modèle
 - Appliquer l'encodage approprié en fonction du type de variable :
 - Pour les variables catégorielles nominales (sans ordre), utiliser le codage one-hot (ou one-hot-dummy)
 - Pour les variables catégorielles ordinales (avec ordre), utiliser le codage ordinal (assigner des valeurs numériques en fonction de l'ordre)
4. Séparer les données en ensembles d'entraînement et de test
 - Diviser les données en X (variables indépendantes) et Y (variable dépendante à prédire)
 - Séparer les ensembles X et Y en ensembles d'entraînement (pour entraîner le modèle) et d'évaluation (pour évaluer les performances du modèle)
5. Appliquer une normalisation ou une standardisation (si nécessaire)
 - Si les variables dans les ensembles d'entraînement ont des échelles différentes, il est recommandé d'appliquer une normalisation (mise à l'échelle des valeurs dans une plage spécifique, comme [0, 1]) ou une standardisation (mise à l'échelle des valeurs pour avoir une moyenne de 0 et un écart type de 1).
6. Entraîner les modèles de prédiction

- Pour le modèle "avant le match":
 - Utiliser les données d'entraînement pour entraîner un algorithme de classification ou de régression en utilisant X comme variables indépendantes et Y ('FTR') comme variable dépendante.
 - Pour le modèle "mi-temps match":
 - Utiliser les données d'entraînement pour entraîner un algorithme de classification ou de régression en utilisant X comme variables indépendantes et Y ('FTR') comme variable dépendante.
7. Évaluer les performances des modèles
- Utiliser les données d'évaluation pour évaluer les performances des modèles en comparant les prédictions avec les valeurs réelles de Y ('FTR').
 - Utiliser des mesures de performance appropriées en fonction du type de modèle (classification ou régression), telles que la précision, le rappel, la F-mesure, l'erreur quadratique moyenne (RMSE), etc.
8. Utiliser les modèles entraînés pour effectuer des prédictions
- Appliquer les modèles entraînés sur de nouvelles données pour prédire les valeurs de Y ('FTR') en utilisant les variables X correspondantes.

Fin

5.8. Les méthodes d'apprentissage utilisées

Dans la section précédente, nous avons décrit l'étape de préparation et de structuration des données pour pouvoir exploiter les méthodes d'apprentissage automatique. Dans ce qui suit, nous allons conduire une étude comparative entre certaines méthodes pour choisir la plus performante.

5.8.1. XGBoost

XGBoost (eXtreme Gradient Boosting) est une implémentation open source optimisée de l'algorithme d'arbres de boosting de gradient. En effet, Le boosting de gradient est un algorithme d'apprentissage supervisé. Son principe de base est de combiner les résultats d'un ensemble de modèles plus simples et plus fiables afin de fournir une meilleure prédiction [18]

Cet algorithme fonctionne de manière séquentielle. Cette technique est lente mais va améliorer l'algorithme par une capitalisation par rapport aux exécutions précédentes. Il commence par construire un premier modèle qu'il va évaluer.

À partir de cette première évaluation, chaque individu va être alors pondéré en fonction de la performance de la prédiction, etc. [18] XGBoost inclut un grand nombre d'hyper paramètres qui peuvent être modifiés et réglés à des fins d'amélioration [18]

5.8.2. Régression logistique

L'algorithme de régression logistique mesure la relation entre la variable dépendante catégorielle et une ou plusieurs variables indépendantes en donnant une estimation à la probabilité d'occurrence d'un événement à travers l'usage de sa fonction logistique.

5.8.3. Classification naïve bayésienne

Classification naïve bayésienne (en anglais naive bayes) est un classifieur probabiliste linéaire simple basé sur le théorème de Bayes en supposant que les attributs qui décrivent les objets de l'ensemble d'apprentissage sont indépendants [19]

La classification naïve bayésienne dispose d'un ensemble d'apprentissage « A » connue et chaque objet de cet ensemble est étiqueté par sa classe C_k . Son but crucial est de chercher à classer un nouvel objet « X_{new} » non encore étiqueté. Pour cela le classifieur choisit la classe C_k qui a la plus grande probabilité.

5.8.4. Arbres de décision

Les arbres de décision sont des algorithmes d'apprentissage automatique supervisé qui se représente comme leurs noms l'indique sous forme d'arbre. Dans l'arbre, la racine est la question principale posée, les noeuds internes représentent un test, et la feuille est le résultat du test (décision). Autrement dit, les arbres de décisions peuvent être sous forme d'une hiérarchie de question « si/autre » [20]

Il existe plusieurs algorithmes d'arbre de décision tels que : ID3, CART, C4.5 et C5.0 pour résoudre les problèmes de classification et de régression.

5.8.5. Forêts aléatoires

Les forêts aléatoires (en anglais « Random Forest ») sont des algorithmes d'apprentissage automatique supervisé utilisé pour la classification et la régression. Ils sont aussi connus par la technique d'apprentissage ensembliste (« bagging ou bootstrap aggregating »). Pour décrire succinctement cette technique, l'algorithme commence à créer plusieurs arbres de décisions ensuite il entraîne chacun de ces

arbres sur une portion aléatoire d'une collection de données, enfin il regroupe les résultats de chaque arbre afin d'effectuer la prédiction [20]. L'avantage principale des forêts aléatoires est qu'ils permettent de réduire le sur-ajustement par la moyenne des résultats des arbres.

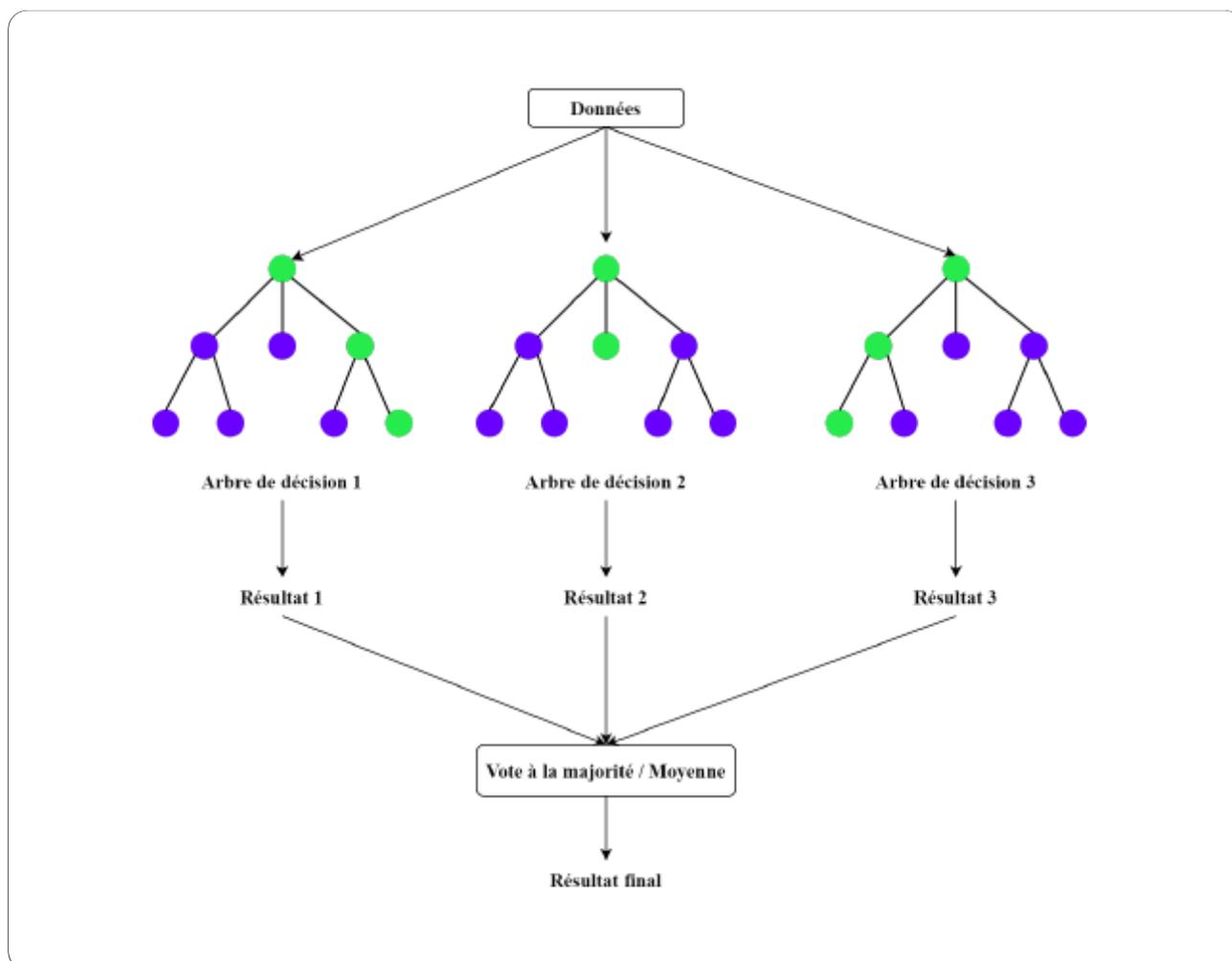


Figure 15: La forêt aléatoire

5.8.6. Machines à vecteurs de support (SVM)

Machines à vecteur de support (SVM) (en anglais « Support Vector Machine ») sont un type d'apprentissage automatique supervisé et notamment utilisé pour la classification binaire ou multi classe. De plus les SVM sont capable de résoudre les problèmes linéairement séparables ainsi que non séparables. En d'autres termes, elles permettent essentiellement de trouver un hyper plan séparateur en maximisant la marge entre les classes [33].

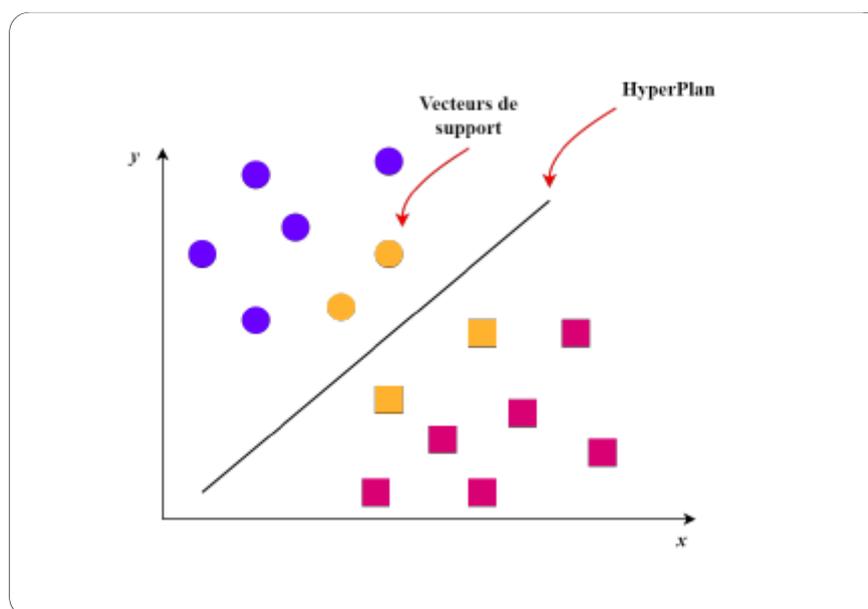


Figure 16:SVM

5.9.Sélection et comparaison des modèles de prédiction

Étant confrontés à une variété de modèles, il est essentiel de trouver un moyen de les comparer et de sélectionner le meilleur parmi eux. Pour ce faire, nous avons employé un ensemble de mesures pour évaluer les performances de chaque modèle. Dans les prochaines sections, nous présentons une description détaillée des métriques utilisées.

5.9.1. Matrice de confusion

Une matrice de confusion est un résumé des prédictions pour un problème de classification. Les prédictions correctes et incorrectes sont mises en évidence et ventilées par classe. Par conséquent, les résultats sont comparés avec la valeur réelle.

La précision est une mesure permettant d'évaluer les modèles de classification. De manière informelle, la précision est le rapport des prédictions correctes du modèle sur l'ensemble des prédictions. La précision à la définition suivante :

Exactitude = Nombre de prédiction correcte/Nombre totale de prédiction.

[17]

Pour chaque modèle utilisé, nous avons calculé sa matrice de confusion ainsi que son exactitude (Accuracy ou taux de bonnes prédictions).

4.9.1.1. Matrice de confusion avant le match « Before match »

Les résultats des modèles appliqués sur le dataset avant le match « Before match » sont représentés sur la figure 17 qui montre la matrice de confusion pour chaque modèle utilisé :

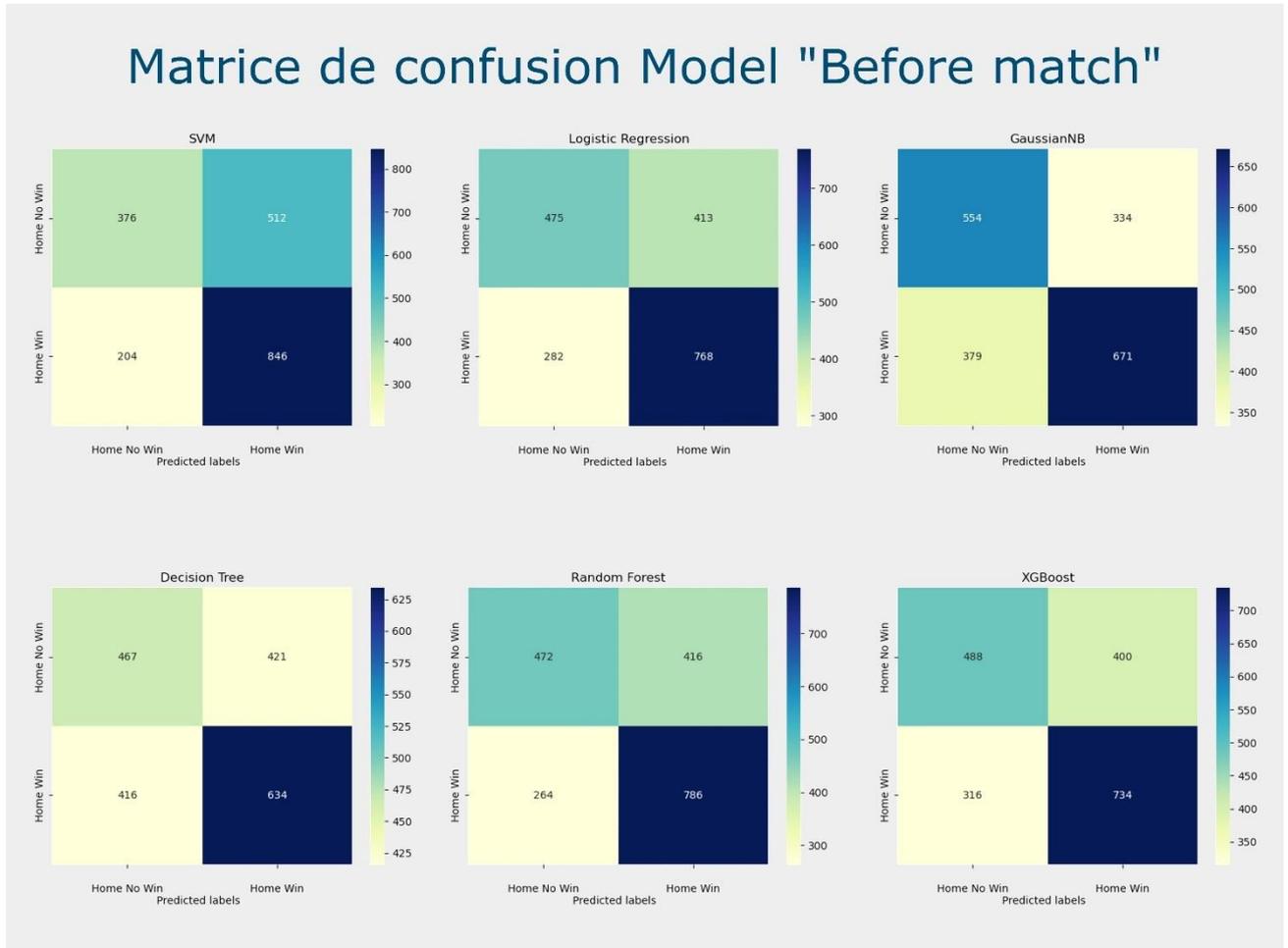


Figure 17: Matrice de confusion des 6 modèles utilisés avant le match

La matrice de confusion compare les prédictions du modèle avec les valeurs réelles pour chaque classe. Voici une explication de la matrice de confusion avant le match pour chaque modèle :

SVM		Logistic Regression		GaussianNB	
Vrai positif	Faux positif	Vrai positif	Faux positif	Vrai positif	Faux positif
376	512	475	413	554	334
Faux négatif	Vrai négatif	Faux négatif	Vrai négatif	Faux négatif	Vrai négatif
204	846	282	768	379	671

Decision tree		Random Forest		XGBoost	
Vrai positif 467	Faux positif 421	Vrai positif 472	Faux positif 416	Vrai positif 488	Faux positif 400
Faux négatif 416	Vrai négatif 634	Faux négatif 264	Vrai négatif 786	Faux négatif 316	Vrai négatif 734

La matrice de confusion montre la répartition des prédictions correctes et incorrectes pour chaque classe.

49.1.2. Matrice de confusion mi-temps de match « half match »

Les résultats des modèles appliqués sur le dataset mi-temps « half match » sont représentés par la figure 18 qui montre la matrice de confusion pour chaque modèle utilisé :

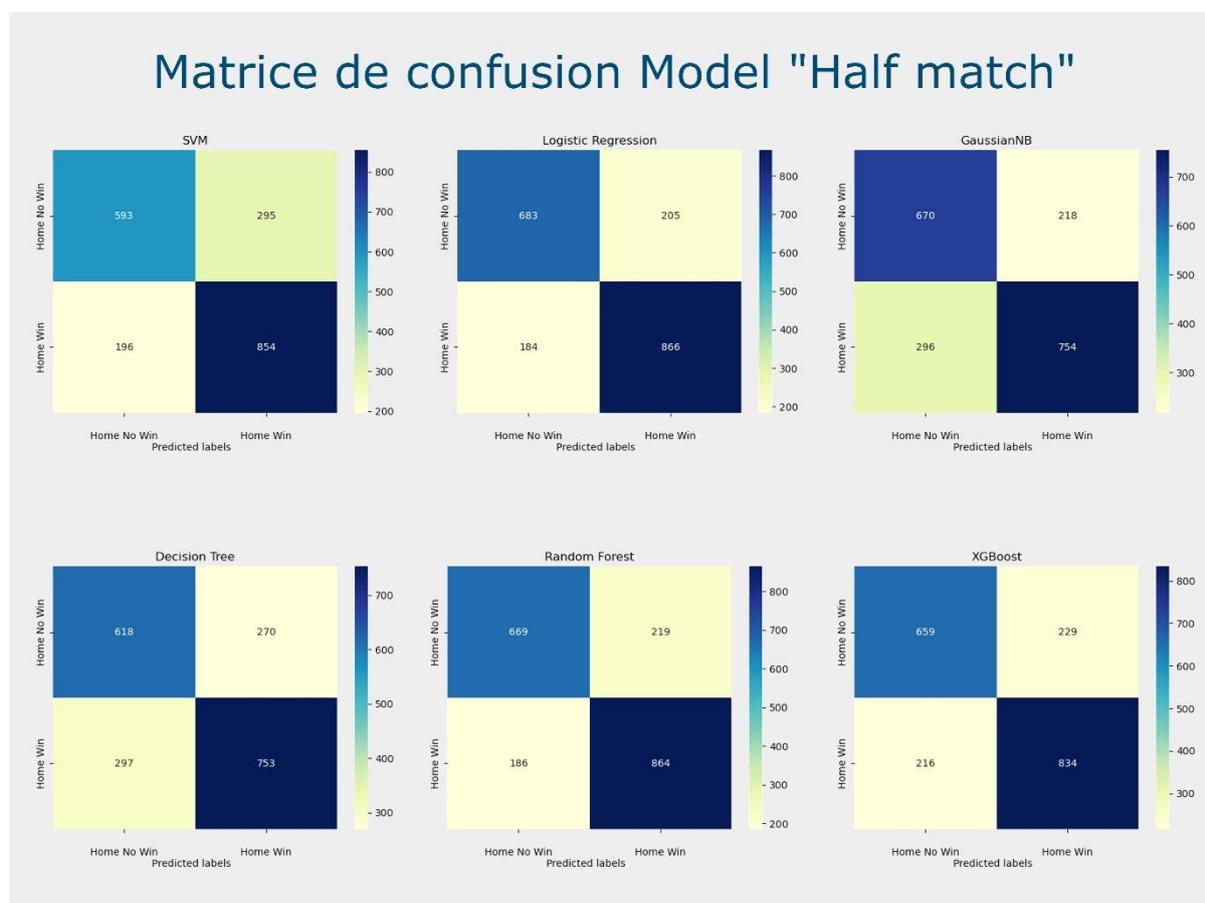


Figure 18: Matrice de confusion des 6 modèles utilisés mi-temps de match

| CHAPITRE 2

CONCEPTION ET MODÉLISATION DU SYSTÈME PROPOSÉ

Voici une explication des résultats de la matrice de confusion pour chaque modèle :

SVM		Logistic Regression		GaussianNB	
Vrai positif 593	Faux positif 295	Vrai positif 683	Faux positif 205	Vrai positif 670	Faux positif 218
Faux négatif 196	Vrai négatif 854	Faux négatif 184	Vrai négatif 866	Faux négatif 296	Vrai négatif 754

Decision tree		Random Forest		XGBoost	
Vrai positif 618	Faux positif 270	Vrai positif 669	Faux positif 219	Vrai positif 659	Faux positif 229
Faux négatif 297	Vrai négatif 753	Faux négatif 186	Vrai négatif 864	Faux négatif 216	Vrai négatif 834

Ces chiffres indiquent comment chaque modèle a effectué ses prédictions. Par exemple, dans le cas du modèle SVM, il a correctement prédit 593 vrais positifs, mais a également prédit à tort 295 faux positifs. De même, il a omis de prédire correctement 196 faux négatifs et a correctement prédit 854 vrais négatifs.

5.9.2. Précision

La précision est la fréquence à laquelle le nombre de prédictions de classe positive appartient réellement à la classe positive [18]

$$\text{Précision} = \text{TP} / \text{TP} + \text{FP}$$

5.9.3. Rappel

Le rappel, parfois appelé « sensibilité », est une métrique dont la quantité est le nombre de prédictions positives correctement identifiées parmi toutes les prédictions positives qui auraient pu être faites [18]].

$$\text{Rappel} = \text{TP} / \text{TP} + \text{FN}$$

5.9.4. F1-score

C'est la moyenne harmonique qui combine à la fois la précision et le rappel en un seul score et qui est défini par la formule suivante [18]:

$$\text{F1Score} = 2 \times \text{Rappel} \times \text{Précision} / \text{Rappel} + \text{Précision}$$

5.9.5. Exactitude (Accuracy)

C'est le taux de la bonne classification, il est donné par la formule suivante [18]:

$$\text{Exactitude} = \frac{TP+TN}{TP+TN +FP+FN}$$

5.9.6. Précision équilibrée (Balanced accuracy)

Balanced accuracy est utilisée dans la classification binaire et multi-classe. C'est la moyenne entre la sensibilité et la spécificité, elle est utilisée quand il y a des données déséquilibrées, (c'est-à-dire lorsque l'une des classes cibles apparaît beaucoup plus que l'autre). Sa formule est comme suit [18]:

$$\text{Précision équilibrée} = \frac{\text{rappel} + \text{spécificité}}{2}$$

D'après les résultats des figures 17,18, nous avons obtenu les statistiques des différents modèles de prédictions cités précédemment pour chaque méthode d'apprentissage, dans le cas avant le match dans les figures 19, 20, 21, 22, et ensuite pour le cas mi-temps les résultats sont représentées par les figures 23, 24, 25, 26 :

Logistic Regression					
	precision	recall	f1-score	support	
H	0.79	0.77	0.78	888	
NH	0.81	0.82	0.82	1050	
accuracy			0.80	1938	
macro avg	0.80	0.80	0.80	1938	
weighted avg	0.80	0.80	0.80	1938	
SVM					
	precision	recall	f1-score	support	
H	0.75	0.67	0.71	888	
NH	0.74	0.81	0.78	1050	
accuracy			0.75	1938	
macro avg	0.75	0.74	0.74	1938	
weighted avg	0.75	0.75	0.74	1938	

Figure 19:statistiques du modèle logistic regression avant le match

Gaussain Naive Bayes Classifier					
	precision	recall	f1-score	support	
H	0.69	0.75	0.72	888	
NH	0.78	0.72	0.75	1050	
accuracy			0.73	1938	
macro avg	0.73	0.74	0.73	1938	
weighted avg	0.74	0.73	0.74	1938	
Decision Tree Classifier					
	precision	recall	f1-score	support	
H	0.68	0.70	0.69	888	
NH	0.74	0.72	0.73	1050	
accuracy			0.71	1938	
macro avg	0.71	0.71	0.71	1938	
weighted avg	0.71	0.71	0.71	1938	

Figure 20: Statistiques du modèle Gaussain Naive Bayes Classifier avant le match

RandomForest					
	precision	recall	f1-score	support	
H	0.64	0.53	0.58	888	
NH	0.65	0.75	0.70	1050	
accuracy			0.65	1938	
macro avg	0.65	0.64	0.64	1938	
weighted avg	0.65	0.65	0.64	1938	
XGBoost					
	precision	recall	f1-score	support	
0	0.61	0.55	0.58	888	
1	0.65	0.70	0.67	1050	
accuracy			0.63	1938	
macro avg	0.63	0.62	0.62	1938	
weighted avg	0.63	0.63	0.63	1938	

Figure 21: Statistiques du modèle Gaussain Naive Bayes Classifier avant le match

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=0.8, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=0.4, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=0.1, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=3, max_leaves=None,
              min_child_weight=3, missing=nan, monotone_constraints=None,
              n_estimators=40, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, ...)
F1 score and accuracy score for training set: 0.7316 , 0.6802.
F1 score and accuracy score for test set: 0.7067 , 0.6502.
```

Figure 22 : Statistiques du modèle XGBClassifier avant le match

D’après les figures précédentes(20, 21, 22, 23) nous remarquons que l’ensemble test contient 70% du dataset.

Logistic Regression				
	precision	recall	f1-score	support
H	0.79	0.77	0.78	888
NH	0.81	0.82	0.82	1050
accuracy			0.80	1938
macro avg	0.80	0.80	0.80	1938
weighted avg	0.80	0.80	0.80	1938
SVM				
	precision	recall	f1-score	support
H	0.75	0.67	0.71	888
NH	0.74	0.81	0.78	1050
accuracy			0.75	1938
macro avg	0.75	0.74	0.74	1938
weighted avg	0.75	0.75	0.74	1938

Figure 23: statistiques du modèle logistic regression mi-temps

Gaussain Naive Bayes Classifier					
	precision	recall	f1-score	support	
H	0.59	0.62	0.61	888	
NH	0.67	0.64	0.65	1050	
accuracy			0.63	1938	
macro avg	0.63	0.63	0.63	1938	
weighted avg	0.63	0.63	0.63	1938	
Decision Tree Classifier					
	precision	recall	f1-score	support	
H	0.53	0.53	0.53	888	
NH	0.60	0.60	0.60	1050	
accuracy			0.57	1938	
macro avg	0.56	0.56	0.56	1938	
weighted avg	0.57	0.57	0.57	1938	

Figure 24: statistiques du modèle Gaussain Naive Bayes Classifier mi-temps

RandomForest					
	precision	recall	f1-score	support	
H	0.78	0.75	0.77	888	
NH	0.80	0.82	0.81	1050	
accuracy			0.79	1938	
macro avg	0.79	0.79	0.79	1938	
weighted avg	0.79	0.79	0.79	1938	
XGBoost					
	precision	recall	f1-score	support	
0	0.75	0.74	0.75	888	
1	0.78	0.79	0.79	1050	
accuracy			0.77	1938	
macro avg	0.77	0.77	0.77	1938	
weighted avg	0.77	0.77	0.77	1938	

Figure 25: Statistiques du modèle Random Forest mi-temps

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=0.8, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=0.4, gpu_id=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=0.1, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=3, max_leaves=None,
              min_child_weight=3, missing=nan, monotone_constraints=None,
              n_estimators=40, n_jobs=None, num_parallel_tree=None,
              predictor=None, random_state=None, ...)
F1 score and accuracy score for training set: 0.8133 , 0.7917.
F1 score and accuracy score for test set: 0.8163 , 0.7972.
```

Figure 26: Statistiques du modèle XGBClassifier mi-temps

D'après les figures précédentes (23, 24, 25, 26) nous remarquons que l'ensemble test contient 81% du dataset.

6. Choix du modèle

Après avoir appliqué différents modèles de classification, d'étude et de visualisation des résultats obtenus pour chaque modèle, et en spécification les modèles « Exactitude » et « précision », nous avons choisi le modèle « *régression logistique* ».

7. Modélisation UML du système d'aide à la décision

Nous avons adopté la modélisation UML pour une description détaillée de notre système décisionnel proposé. Cette approche offre une représentation simplifiée et planifiée des différentes étapes du système en réalisant un scénario d'exécution.

Dans les sections qui suivent, nous présentons les trois diagrammes UML les plus pertinents :

- Le diagramme de classes,
- Les diagrammes de cas d'utilisation,
- Les diagrammes de séquence.

7.1. Diagramme de Cas d'utilisation (Use Case)

Le diagramme de cas d'utilisation est utilisé pour décrire de manière schématique les interactions fonctionnelles entre les acteurs et le système. Son objectif principal est de

définir l'application de point de vue de l'utilisateur. Nous identifions deux acteurs qui interagissent avec le système :

- User : il doit s'authentifier afin de consulter les statistiques, publier poster ou statuts, choisir l'équipe qui va gagner et consulter le match afin de faire une prédiction sur le résultat de ce dernier.
- Admin : il a le droit d'accéder à tous les modules du système et exécuter les fonctionnalités de (consulter, modifier, ajouter vérifier, prédire...).

7.1.1. Cas d'utilisation « User »

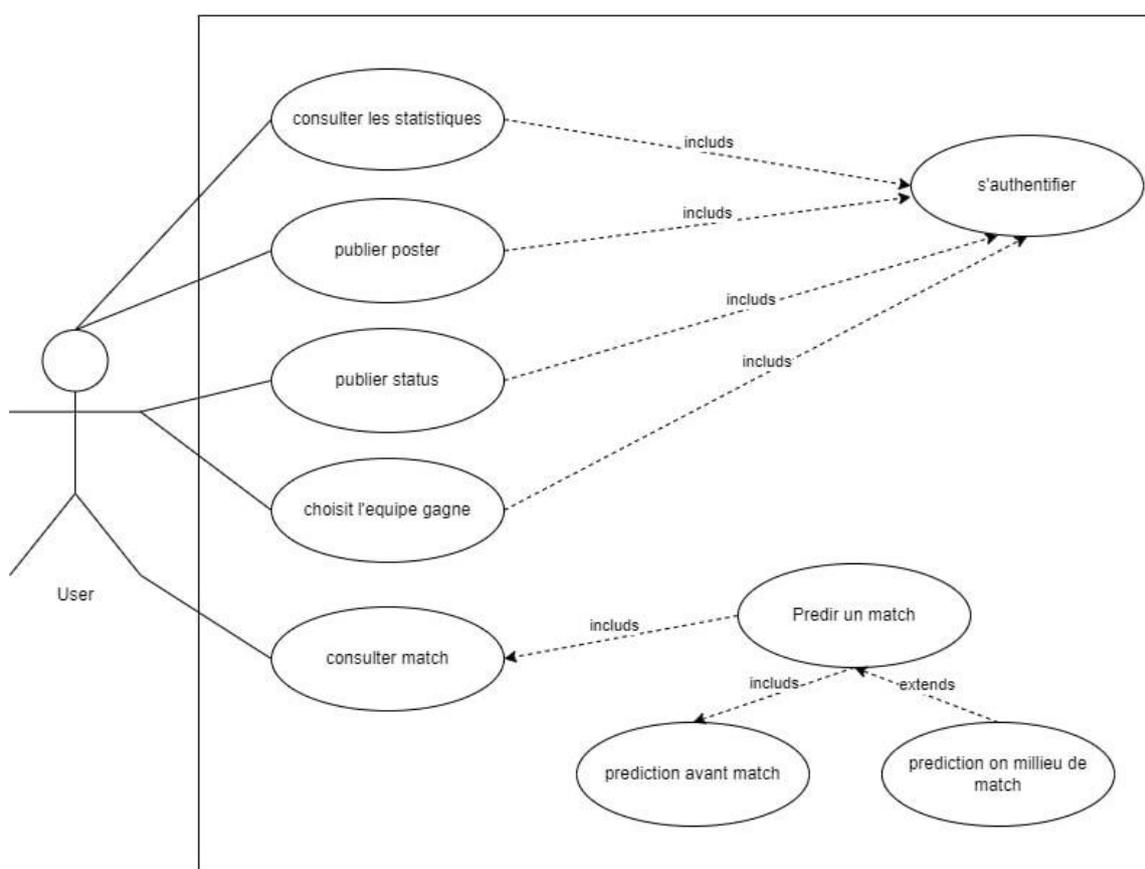


Figure 27: diagramme cas d'utilisation User

Cas d'utilisation	Description
S'authentifier	S'authentifier à l'application mobile pour accéder au profil de l'utilisateur après avoir fait un enregistrement.
Consulter statistiques Publier poster Publier statuts Choisir équipe gagnante	Après l'authentification, l'utilisateur peut accéder à ses fonctionnalités selon son profil
Consulter match	L'utilisateur peut consulter un match sans authentification, ce cas d'utilisation permet de prédire le résultat de ce match
Predir un match	Ce cas d'utilisation représente l'objectif principal de notre système, il permet d'effectuer une prédiction sur le résultat d'un match
Prédiction avant match	Prédiction avant le début de match
Prédiction en milieu de match	Prédiction durant le match ou après le premier mi-temps

Tableau 4:Description du diagramme cas d'utilisation User

7.1.2. Cas d'utilisation « Admin »

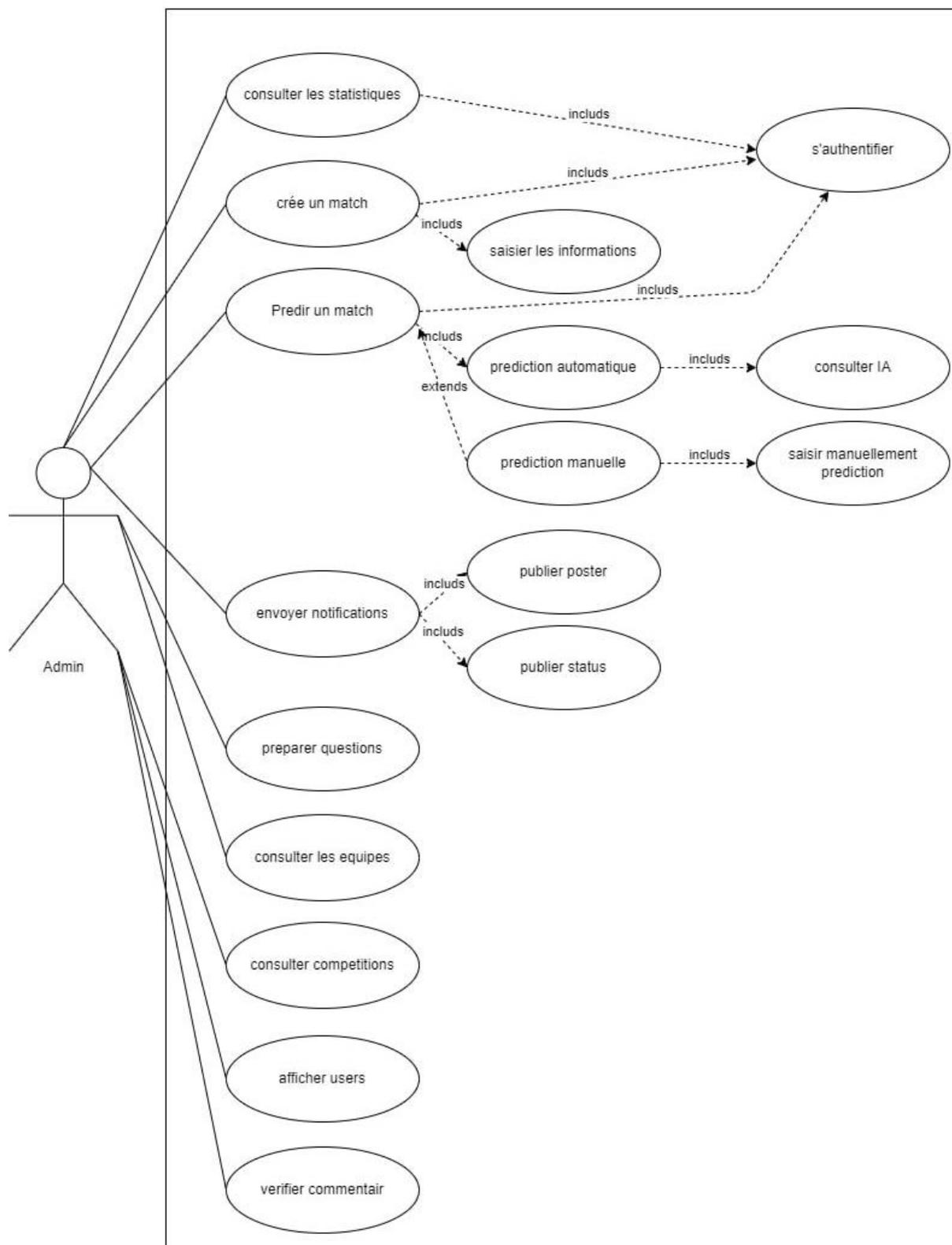


Figure 28:Diagramme cas d'utilisation Admin

Cas d'utilisation	Description
S'authentifier	L'admin effectue l'authentification pour accéder au profil d'administrateur.
Consulter les statistiques	Après l'authentification, l'admin peut consulter les statistiques sur le site web
Créé un match	L'admin peut créer un match pour les utilisateurs qui veulent faire une prédiction.
Saisir les informations	Après la création du match, l'admin saisie les informations de ce match.
Prédire un match	Pour prédire un match, il faut s'authentifier tout d'abord, ensuite il existe deux méthodes pour réaliser ça, manuelle ou automatique .
Prédiction automatique	La prédiction automatique est générée à l'aide des techniques de l'IA.
Prédiction manuelle	Cette prédiction est générée manuellement d'après le choix de l'utilisateur.
Envoyer notifications	L'admin envoie des notifications soit pour publier un status ou bien publier un poster.
Préparer questions	Après l'envoi de notification , l'admin prépare un questionnaire
Consulter les équipes, consulter compétitions, afficher users, vérifier commentaire	De plus l'admin peut gérer ces opérations.

Tableau 5: Description du diagramme cas d'utilisation Admin

7.2. Diagramme de séquences

Le diagramme de séquence permet de décrire les interactions et l'ordre chronologique des messages échangés entre les différentes entités du système. Dans

ce qui suit, nous allons décrire le diagramme de séquence adapté au module de prédiction de notre système et l'application mobile.

Nous aborderons trois scénarios, dont le premier concerne l'authentification des utilisateurs dans l'application.

Scénario 1 : Ce scénario représente la première interaction entre l'application et l'utilisateur pour que ce dernier puisse accéder à son compte.

La valeur « ref » signifie qu'un utilisateur a été invité à utiliser l'application par un autre utilisateur qui l'utilise déjà par une référence qui doit être saisie avant de se connecter par son compte « Facebook » ou bien « Google », ce dernier gagne des points à travers cette opération. Si cette valeur est égale à nul alors le nouvel utilisateur doit se connecter directement par un de ses compte.

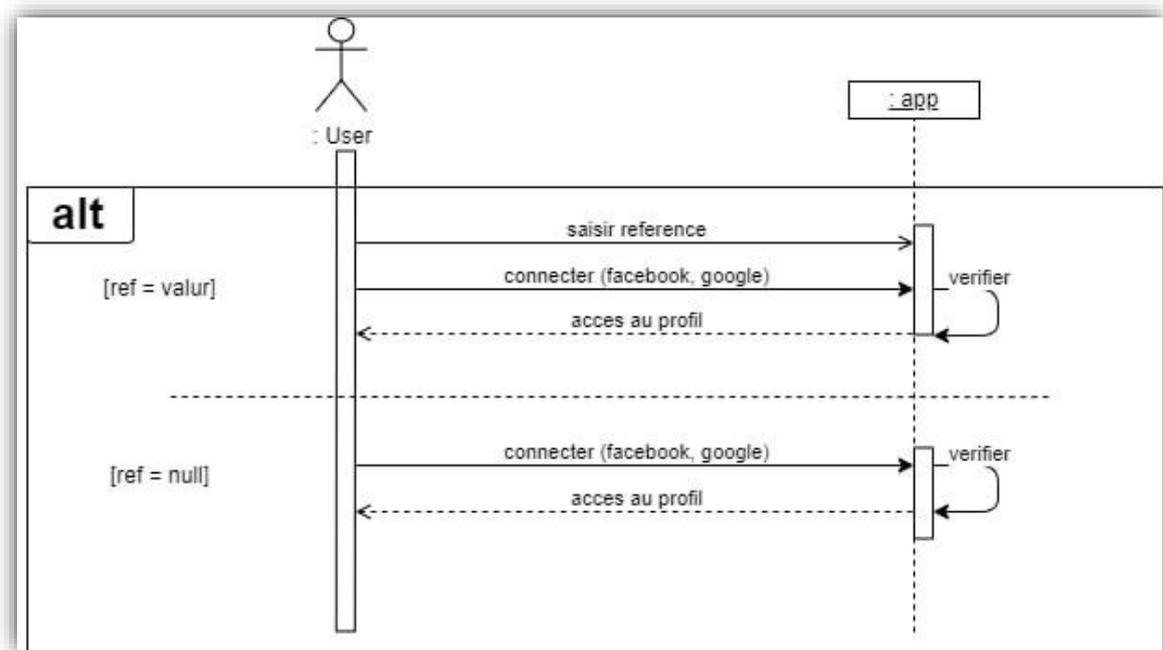


Figure 29: diagramme de séquence authentification application

Scénario 2 : Ce scénario représente la deuxième interaction de l'utilisateur durant l'utilisation de l'application après l'authentification.

- Si l'utilisateur n'est pas authentifié, il peut seulement consulter les matchs disponibles.

| CHAPITRE 2

CONCEPTION ET MODÉLISATION DU SYSTÈME PROPOSÉ

- Sinon il peut consulter les matchs et de plus de faire toutes les autres opérations existes dans l'application, ainsi la prédiction de résultat de match, selon la Figure30 :

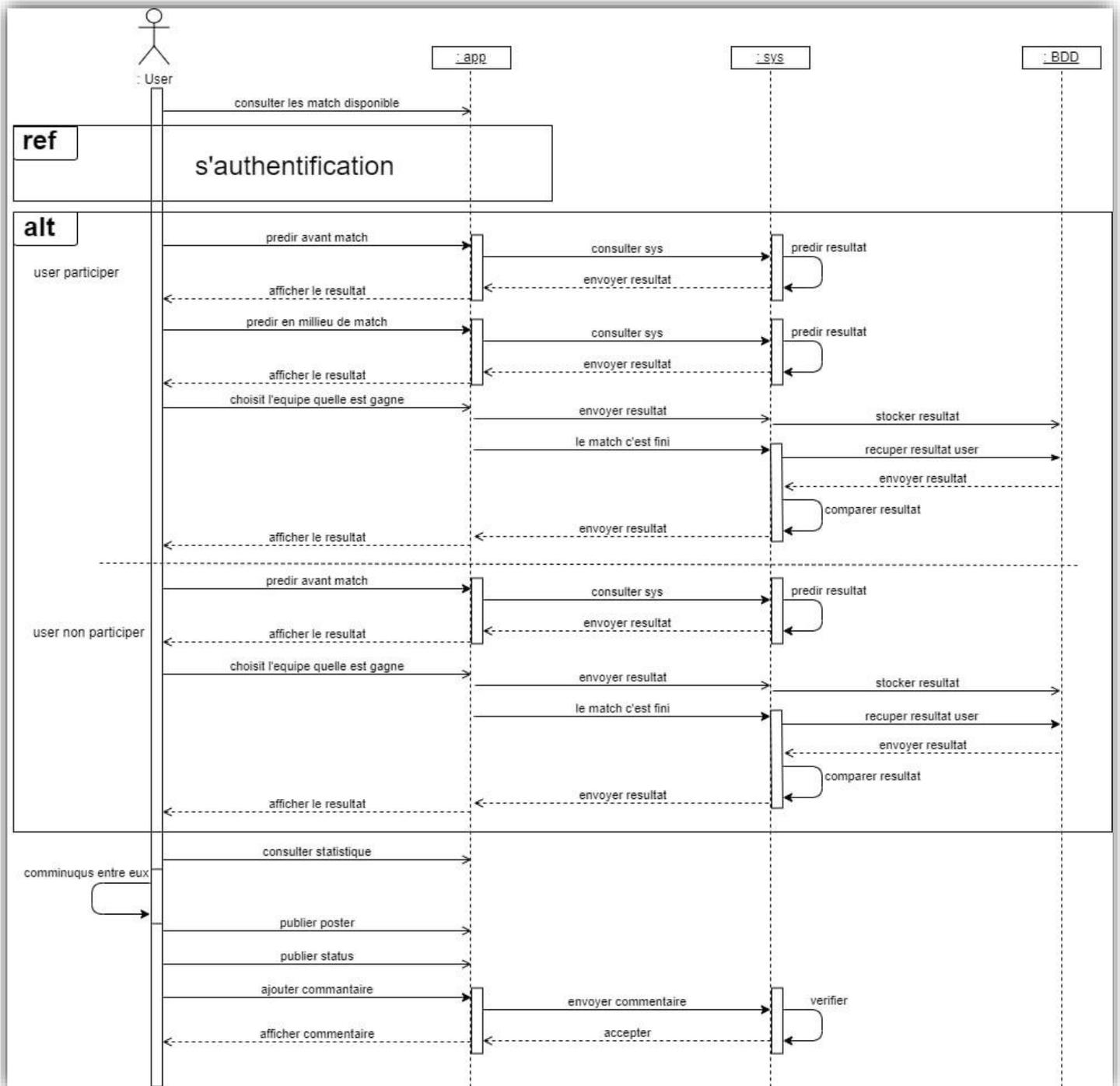


Figure 30: Diagramme de séquence des différentes opérations dans l'application mobile

Ce dernier scenario représente les différentes opérations qui gère l'admin sur le site web et comment se passe le traitement du modèle de prédiction et son exécution.

| CHAPITRE 2

CONCEPTION ET MODÉLISATION DU SYSTÈME PROPOSÉ

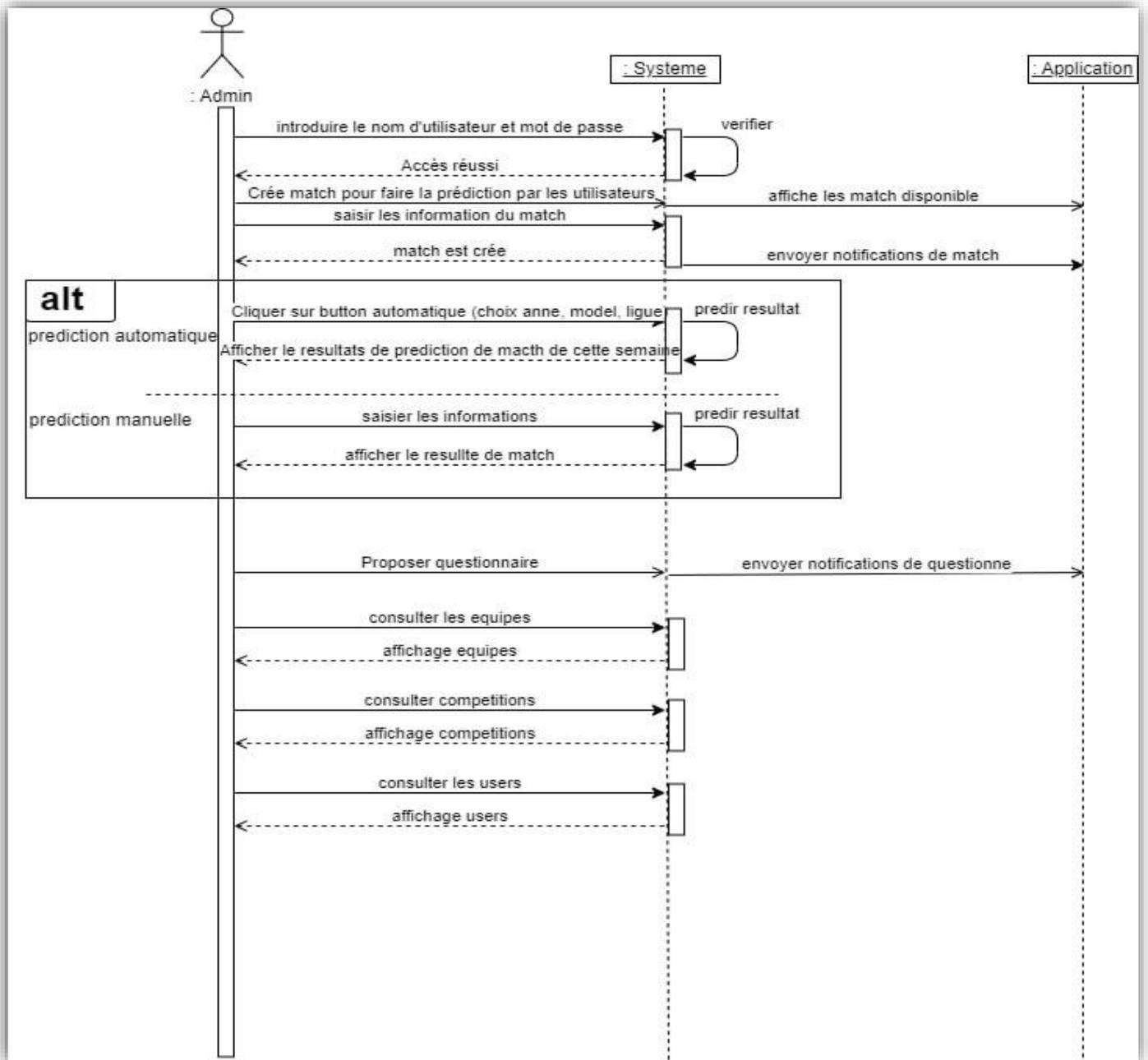


Figure 31: Diagramme de séquence de la partie site web et la gestion de l'admin

Diagramme de classes

Le diagramme de classe est un diagramme statique, permettant de décrire les opérations et les attributs des classes et des interfaces du système ainsi que les différentes relations entre celles-ci.

| CHAPITRE 2

CONCEPTION ET MODÉLISATION DU SYSTÈME PROPOSÉ

La Figure 32, représente le diagramme de classes de notre système d'aide à la décision intelligent.

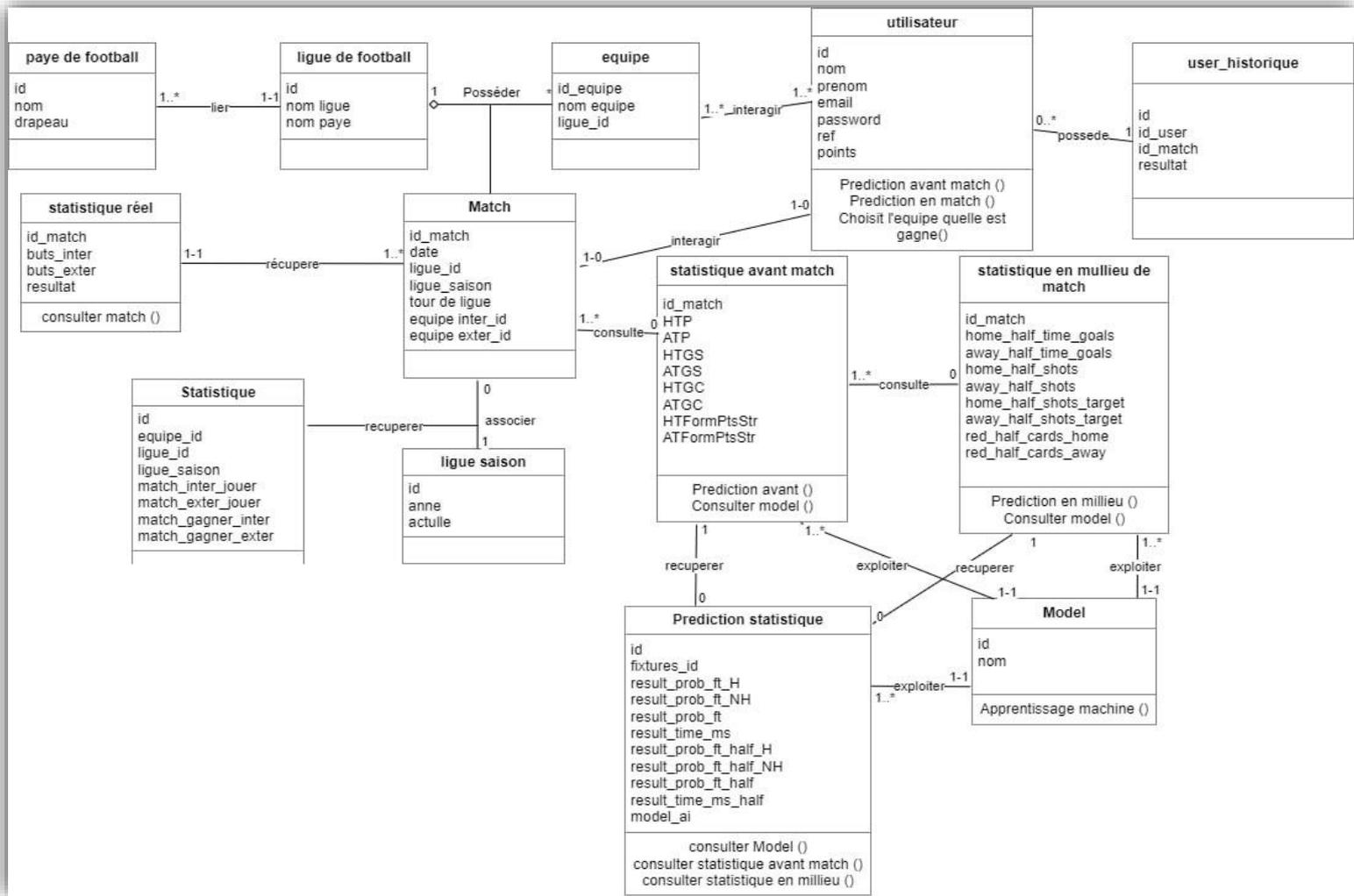


Figure 32: Diagramme de classes de notre système

8. Conclusion

Dans ce chapitre, nous avons présenté en détail notre système d'aide à la décision proposé. Nous avons présenté notre modèle de conception en commençant par une revue des objectifs, passant par l'architecture globale du système, ensuite l'étude des données et des modèles d'apprentissage automatique utilisés, puis nous avons modéliser notre outil d'aide à la décision en exploitant différents diagrammes UML.

| CHAPITRE 2

CONCEPTION ET MODÉLISATION DU SYSTÈME PROPOSÉ

Nous avons détaillé et justifié notre choix de modèle d'apprentissage machine en comparant les résultats obtenus par application de chaque modèle.

Ces étapes nous permettent de se préparer à démarrer la phase suivante, la phase de mise en œuvre et d'implémentation, qui est couverte dans le chapitre suivant.

CHAPITRE 3 MISE EN OEUVRE

1. Introduction

Après avoir élaboré la conception ainsi que la modélisation de notre système, nous abordons dans ce chapitre le dernier volet de ce projet, dont l'objectif est d'exposer la phase d'implémentation et de mise en œuvre. Nous allons, également, présenter les outils exploités pour le développement de notre système site web et de l'application « Pariball ».

2. Outils de développement

La mise en œuvre du système proposé a été réalisée à l'aide de différentes technologies décrites dans ce qui suit :

2.1.BackEnd

2.1.1. PHP

PHP (Hypertext Preprocessor) est un langage de programmation de script côté serveur utilisé pour le développement web. Il a été créé en 1994 par Rasmus Lerdorf et a depuis lors été largement utilisé pour créer des applications web dynamiques et interactives. PHP est un langage open-source, ce qui signifie que le code source est gratuitement disponible et modifiable pour tous les utilisateurs.

Le langage PHP est largement utilisé dans le développement web en raison de sa facilité d'utilisation et de sa polyvalence. Il permet aux développeurs de créer des pages web dynamiques et interactives en générant du contenu HTML à la volée. PHP est également compatible avec de nombreux systèmes de gestion de base de données tels que MySQL, ce qui facilite le stockage et la récupération de données pour les applications web. [22]

2.1.2. Symfony

Symfony est un framework open-source écrit en PHP, utilisé pour le développement d'applications web. Il a été créé en 2005 par Fabien Potencier et a depuis lors été largement utilisé pour développer des applications web complexes et évolutives. [23]

Symfony offre un ensemble de fonctionnalités et de composants prêts à l'emploi pour le développement web, tels que la gestion des formulaires, la sécurité, la gestion des sessions, la validation des données, l'ORM (Object-Relational Mapping), etc. Il est également compatible avec de nombreux systèmes de gestion de base de données tels que MySQL, PostgreSQL, Oracle, etc. [24]

2.1.3. MySQL

MySQL est un système de gestion de base de données (SGBD) open-source, utilisé pour stocker et récupérer des données pour les applications web. Il a été créé en 1995 par Michael Widenius et David Axmark et est maintenant maintenu par Oracle Corporation. [25]

MySQL est un SGBD très populaire, largement utilisé dans le développement web en raison de sa facilité d'utilisation, de sa vitesse et de sa fiabilité. Il est compatible avec de nombreux langages de programmation tels que PHP, Java, Python, etc., ce qui facilite l'intégration avec les applications web [26].

2.2. Frontend

2.2.1. JavaScript

JavaScript est un langage de programmation de script côté client utilisé pour le développement web. Il a été créé en 1995 par Brendan Eich. [27]

JavaScript est un langage de programmation polyvalent qui peut être utilisé pour créer des applications web, des extensions de navigateur, des applications mobiles, etc. Il est souvent utilisé en conjonction avec HTML et CSS pour créer des pages web interactives et dynamiques [28].

JavaScript offre de nombreuses fonctionnalités telles que la manipulation de DOM (Document Object Model), les événements, les animations, les transitions, les effets, etc. Il prend également en charge les bibliothèques et les frameworks tels que jQuery, Angular, React, Vue, etc. pour faciliter le développement web. [29]

2.2.2. CSS

CSS (Cascading Style Sheets) est un langage de feuille de style utilisé pour décrire la présentation d'un document écrit en HTML ou en XML. Il a été créé en 1996 par Håkon Wium Lie et Bert Bos et est maintenant largement utilisé pour concevoir et styliser des pages web. CSS permet aux développeurs de créer des designs visuellement attrayants pour les pages web en contrôlant la présentation, le formatage et la mise en page des éléments HTML. Il permet également de séparer la présentation de la structure HTML, ce qui facilite la maintenance et la mise à jour des pages web. [30]

CSS offre de nombreuses fonctionnalités telles que les sélecteurs, les attributs, les propriétés, les valeurs, etc. Il est également compatible avec les frameworks et les bibliothèques tels que Bootstrap, Materialize, Foundation, etc. [31].

2.2.3. HTML

HTML (Hypertext Markup Language) est un langage de balisage utilisé pour créer des pages web. Il a été créé en 1990 par Tim Berners-Lee et est maintenant largement utilisé pour créer des pages web statiques et dynamiques. [32]

HTML permet aux développeurs de structurer le contenu d'un document en utilisant des balises et des attributs. Les balises HTML définissent les éléments tels que les titres, les paragraphes, les listes, les images, les liens, etc. tandis que les attributs fournissent des informations supplémentaires sur ces éléments. [32]

HTML est souvent utilisé en conjonction avec CSS et JavaScript pour créer des pages web interactives et dynamiques. Il est également compatible avec les frameworks et les bibliothèques tels que Bootstrap, Foundation, jQuery, etc.

2.2.4. Bootstrap

Bootstrap est un framework open-source de développement web créé par Twitter. Il est basé sur HTML, CSS et JavaScript et est utilisé pour créer des sites web réactifs et mobiles. Bootstrap a été créé pour simplifier le processus de développement web en fournissant des éléments d'interface utilisateur prédéfinis tels que les boutons, les formulaires, les tables, les grilles, etc., et des fonctionnalités JavaScript telles que les modals, les onglets, les carrousels, etc. [33]

Bootstrap est facile à utiliser et à personnaliser, ce qui en fait un choix populaire pour les développeurs web débutants et expérimentés. Il est également compatible avec les navigateurs les plus populaires et les versions mobiles, ce qui permet de créer des sites web réactifs [33].

2.2.5. Twig

Twig est un moteur de template open-source pour PHP. Il a été créé par Fabien Potencier, le créateur de Symfony, et est maintenant largement utilisé pour générer des vues HTML dynamiques pour les applications web. [34]

Twig permet aux développeurs de créer des templates HTML réutilisables et de les remplir avec des données dynamiques en utilisant des variables, des boucles, des conditions, etc. Il offre également des fonctionnalités avancées telles que l'héritage, l'inclusion, la surcharge, les filtres, etc. [35] Twig est souvent utilisé en conjonction avec des frameworks PHP tels que Symfony, Laravel, Yii, etc.

2.2.6. Flutter

Flutter est un cadre open source développé et pris en charge par Google. Les développeurs front-end et full-stack utilisent Flutter pour créer l'interface utilisateur (UI) d'une application pour plusieurs plateformes avec une seule base de code. [36]

2.2.7. Dart

Dart est un langage de programmation, qui a été = développé par Google. Le langage Dart ressemble aux langages de programmation orientés objet établis, dont Swift, C# ou Java font partie, qui sont soumis à des paradigmes de programmation déterminés. Les règles de combinaison de signes établis, c'est-à-dire la syntaxe, ressemblent au langage de programmation C. Cette « parenté » facilite grandement la prise en main de sorte que l'on peut y accéder sans avoir à surmonter trop d'obstacles. [37]

2.3. Machine Learning

2.3.1 BeautifulSoup

BeautifulSoup est une bibliothèque Python open-source utilisée pour extraire des données à partir de fichiers HTML et XML. Elle a été créée par Leonard Richardson et est maintenant largement utilisée pour l'analyse de données web. [38]

BeautifulSoup permet aux développeurs de parcourir et de manipuler le contenu HTML et XML à l'aide de méthodes Python telles que `find()`, `find_all()`, `select()`, etc. Elle offre également des fonctionnalités avancées telles que la manipulation d'attributs, la gestion des erreurs, la navigation dans les arbres DOM, etc.

2.3.2 SK Learn

Est l'une des bibliothèques ML les plus connues. Elle est très utile pour les algorithmes de classification, de régression ou de cluster tels que les forêts d'arbres décisionnels, etc. [39]

2.3.3 Pandas

Est une bibliothèque Python open source conçue pour la manipulation et l'analyse de données, elle propose des structures de données et des opérations pour manipuler des tables numériques. Elle est aujourd'hui utilisée par un grand nombre de scientifiques et d'analystes [40]

2.3.4 Numpy

NumPy est une bibliothèque pour le langage de programmation Python conçue pour manipuler des matrices ou des tableaux multidimensionnels et des fonctions mathématiques qui opèrent sur ces tableaux. Plus précisément, cette bibliothèque de logiciels gratuite et open-source offre une variété de fonctions, permettant notamment de créer des tableaux directement à partir de fichiers ou à l'inverse d'enregistrer des tableaux dans des fichiers, ainsi que de manipuler des vecteurs, des matrices et des polynômes.

2.3.5 Matplotlib

Matplotlib est une bibliothèque open-source pour Python utilisée pour la visualisation de données. Elle a été créée en 2003 par John D. Hunter et est maintenant largement utilisée pour la création de graphiques et de visualisations en Python [41].

Matplotlib offre des fonctionnalités pour la création de graphiques en 2D et en 3D, la manipulation de couleurs, les sous-tracés, les annotations, etc. Elle prend en charge une variété de formats de sortie tels que PNG, PDF, SVG, etc.

Matplotlib est souvent utilisée en conjonction avec d'autres bibliothèques Python telles que NumPy, Pandas, Seaborn, etc. pour faciliter la visualisation des données.

3. Environnement de développement

3.1 Visual Studio Code

Visual Studio Code (VS Code) est un éditeur de code source open-source développé par Microsoft. Il a été lancé en 2015 et est maintenant largement utilisé pour le développement de logiciels. [42]. VS Code offre des fonctionnalités telles que l'auto-complétion, la coloration syntaxique, le débogage, la gestion de code source, etc. Elle prend en charge une variété de langages de programmation tels que Python, JavaScript, C++, etc.

3.2 Jupyter Notebook

Jupyter Notebook est une application web open-source utilisée pour la création et le partage de documents interactifs contenant du code, des visualisations et du texte. Elle a été créée en 2014 par Fernando Pérez et est maintenant largement utilisée dans les domaines de la science des données, de l'apprentissage machine, etc. [43] Jupyter Notebook est souvent utilisé en conjonction avec d'autres bibliothèques Python telles que NumPy, Pandas, Matplotlib, etc. pour faciliter l'analyse de données et la visualisation des résultats.

4. Outils matériels

Le développement de notre site web et application mobile a été effectués sur une machine avec les configurations suivantes :

Machine de Marque	HP EliteBook 15
Système	Windows 10
Microprocesseur	Intel (R) Core (TM) i5-6300U CPU @ 2.40Ghz 2.50 Ghz
Mémoire Vive	16,00 Go

Type du système	Systeme d'exploitation 64 bits, processeur x64
------------------------	--

5. Fonctionnalités du site web proposé et scénario d'exécution

Dans cette section , nous allons présenter les interfaces les plus importantes de notre système.

5.1. Page d'authentification admin

La figure suivante présente la page d'authentification pour l'administrateur du site web afin d'accéder au site par un nom d'utilisateur et un mot de passe.

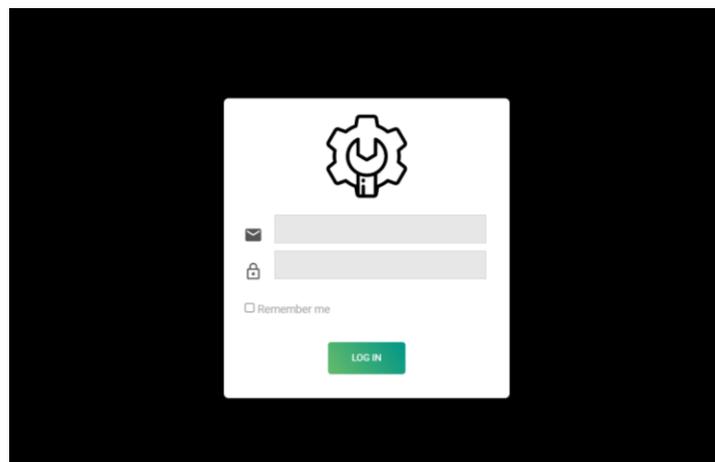


Figure 33: Page d'authentification admin

5.2. Page tableau de bord

La figure 34 représente le tableau de bord du décideur (l'administrateur du site) où nous pouvons y retrouver dans cet espace, le menu principal qui contient plusieurs onglets chacune avec des service et options spécifiques, le tableau de bord contient des informations numériques concernant le contenu de notre site, comme numéros d'utilisateurs, matches, joueurs...

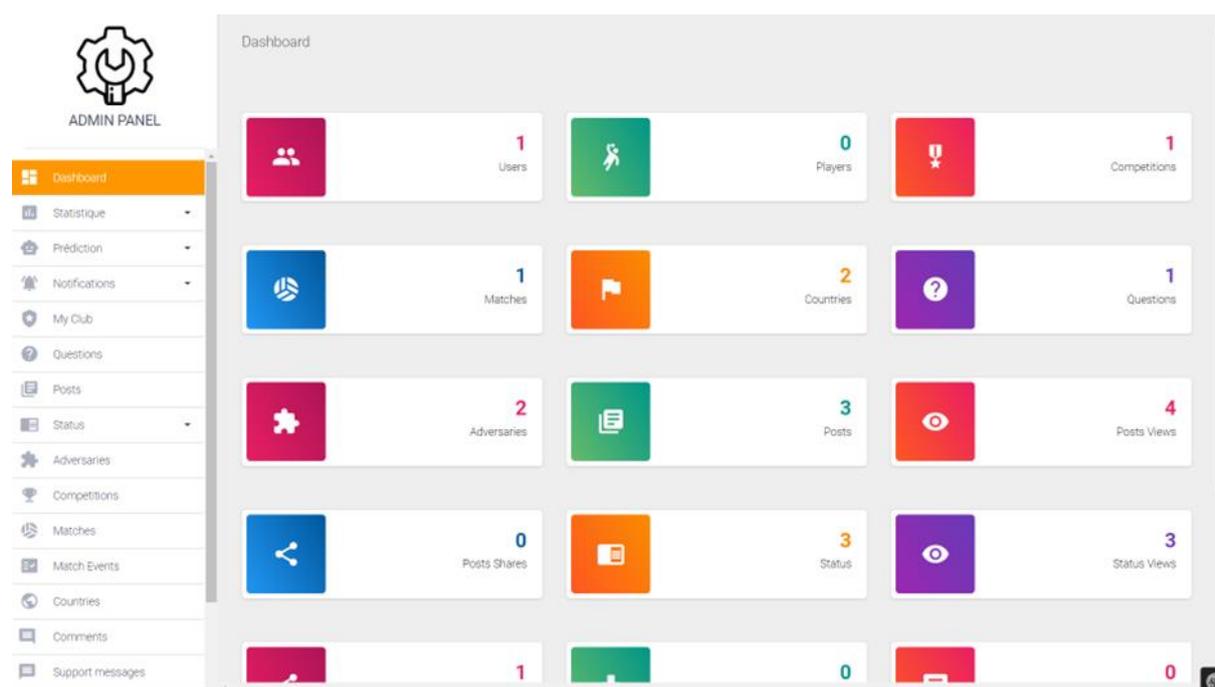


Figure 34: Onglet tableau de bord

Nous pouvons naviguer vers d'autres onglets du site web, l'onglet suivant est « Statistiques ».

5.3. Page Statistique

La figure 35 représente l'onglet Statistique qui contient les statistiques des modèles d'apprentissage appliqués sur les données de note site avant le match (Before match) et mi-temps (Half match) les résultats sont représentés par les diagrammes suivants :

- Diagramme circulaire : Il représente les proportions relatives au dataset des deux variables entraînement et test.
- Diagramme à barres : représente la moyenne de la validation croisée, F1-score en fonction de 6 modèles d'apprentissage utilisés
- Histogramme : représente le résultat des métriques, les modèles de prédiction en fonction de 6 méthodes d'apprentissage utilisées

CHAPITRE 3

Mise en œuvre

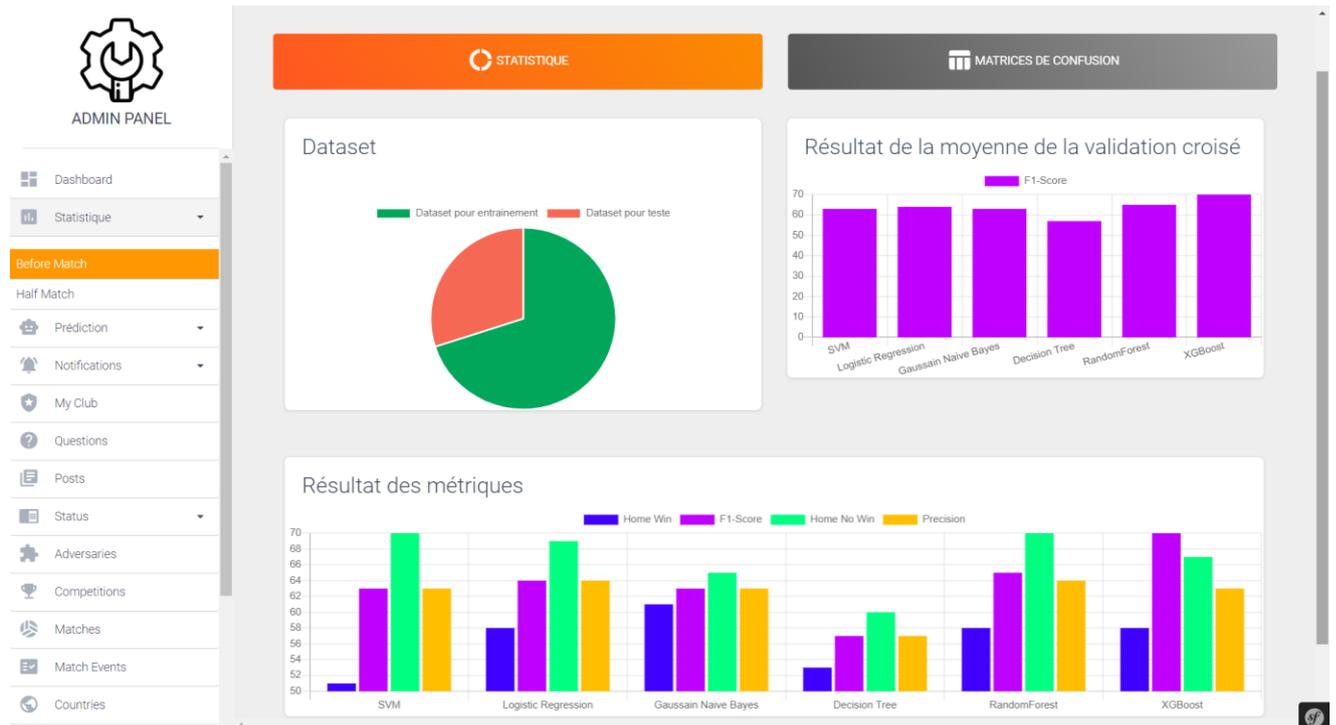


Figure 35 : Page statistique Before match

La page contient la matrice de confusion illustrée par la figure 36.

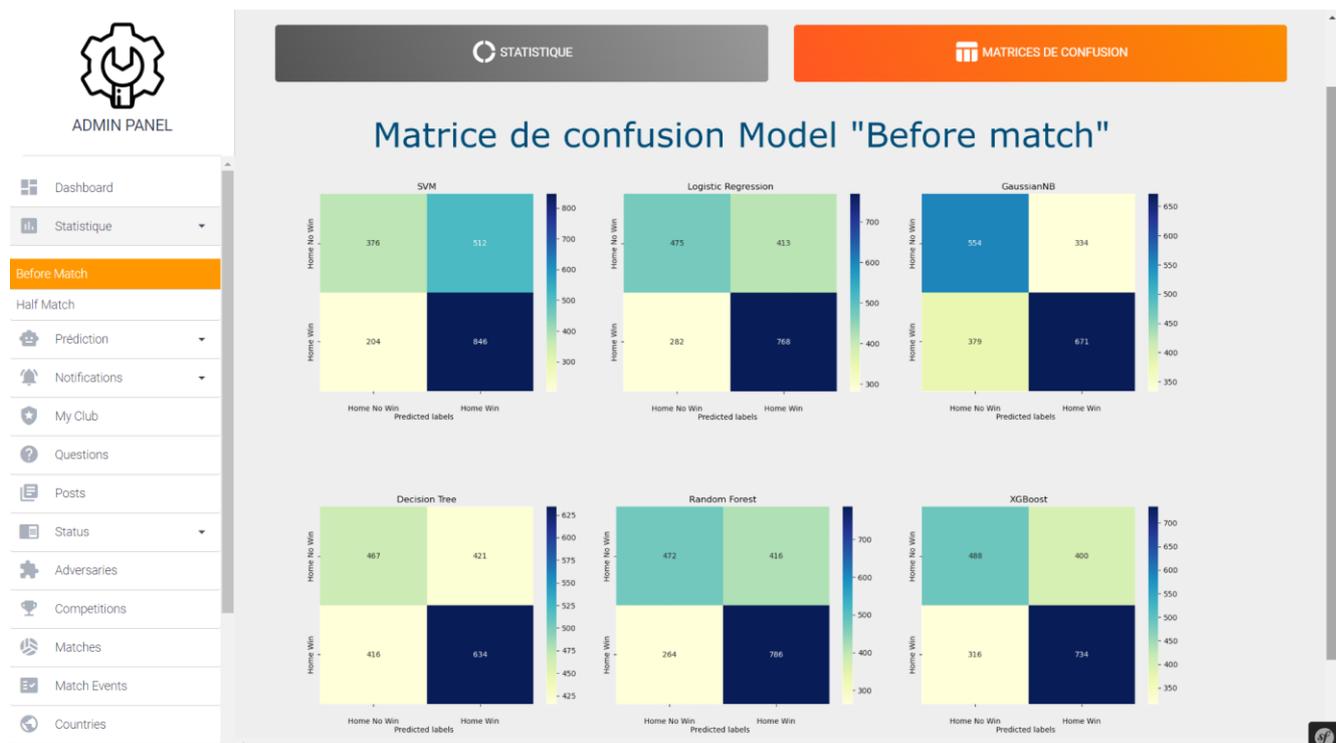


Figure 36 : Matrice de confusion

5.4. Page Prédiction

L'onglet Prédiction est divisé en deux onglets :

- **Onglet Automatique** : l'admin choisit une League, les matchs de cette semaine seront affichés ainsi que les résultats de chaque match durant toute la semaine, ces données sont récupérées à travers des API-football.

L'admin peut choisir également le modèle et la saison (par exemple 2022). Le résultat de la prédiction est obtenu sous forme de pourcentage et sera stocké dans la base de données. Quand le match prédit est terminé, le système vérifie et compare le résultat réel avec celui qui existe dans la base de données, si la prédiction est juste il sera affiché true sinon false.

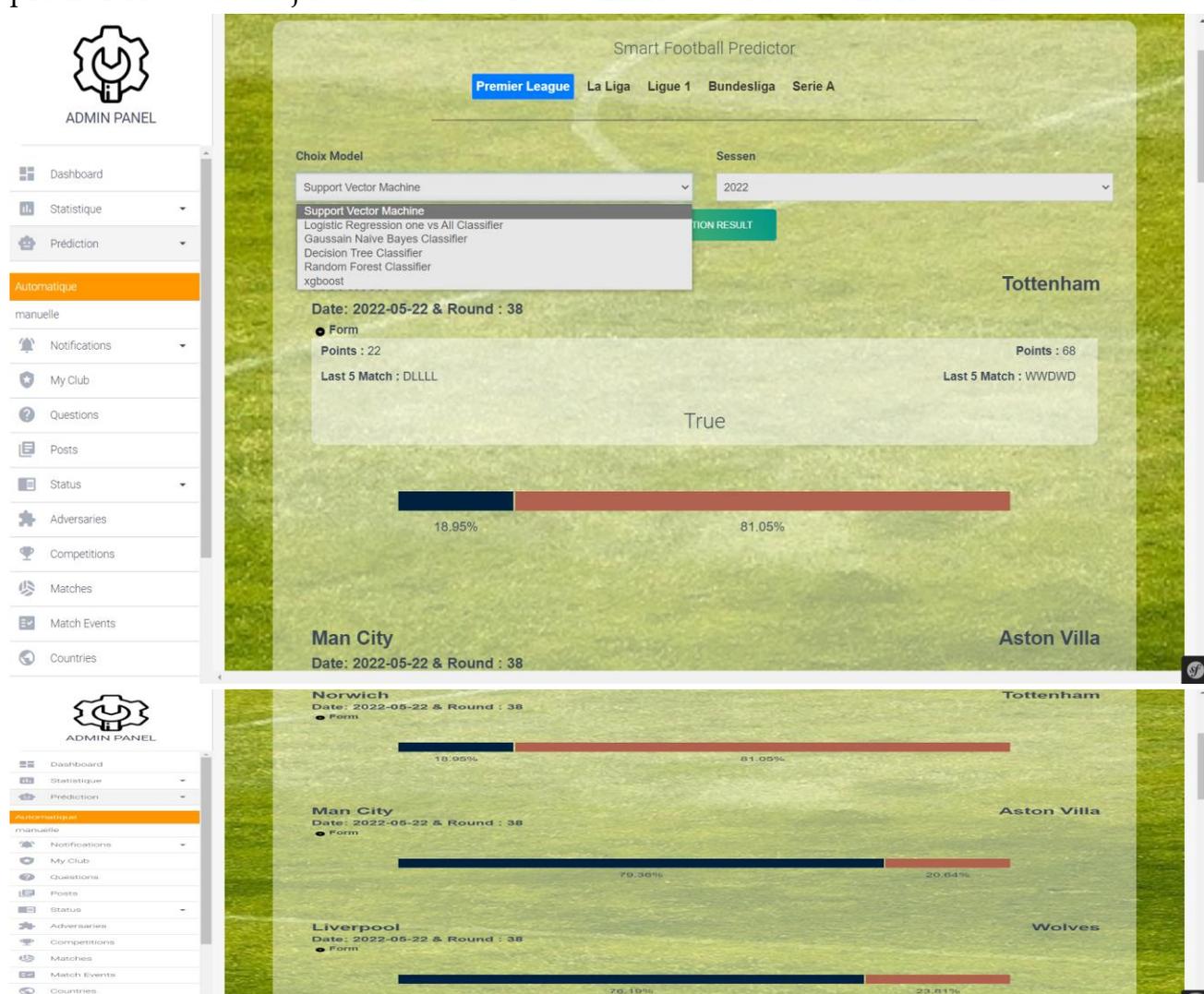


Figure 37: Onglet prédiction automatique

| CHAPITRE 3

Mise en œuvre

- **Onglet Manuelle** : l'admin peut choisir de faire la prédiction avant le match (before match) ou en mi-temps (half match), il choisit ainsi les équipes et introduire ses informations ; et finalement appliquer le modèle de prédiction, le résultat sera affiché comme le montre la figure suivante.

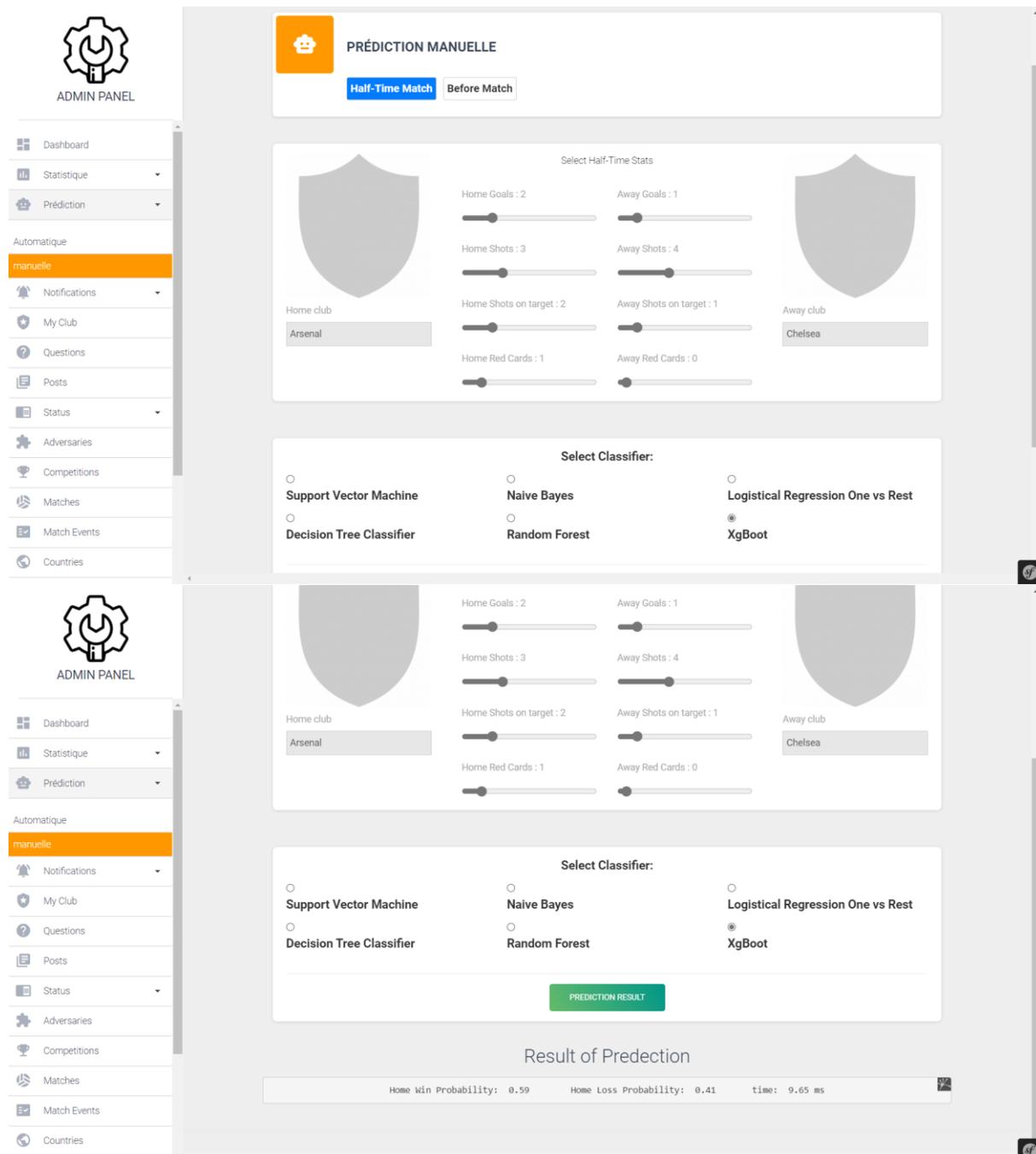


Figure 38 : Onglet prédiction manuelle

| CHAPITRE 3

Mise en œuvre

5.5. Page questions

L'admin prépare des questions aux utilisateurs d'application, ces questions seront affichées comme suit :

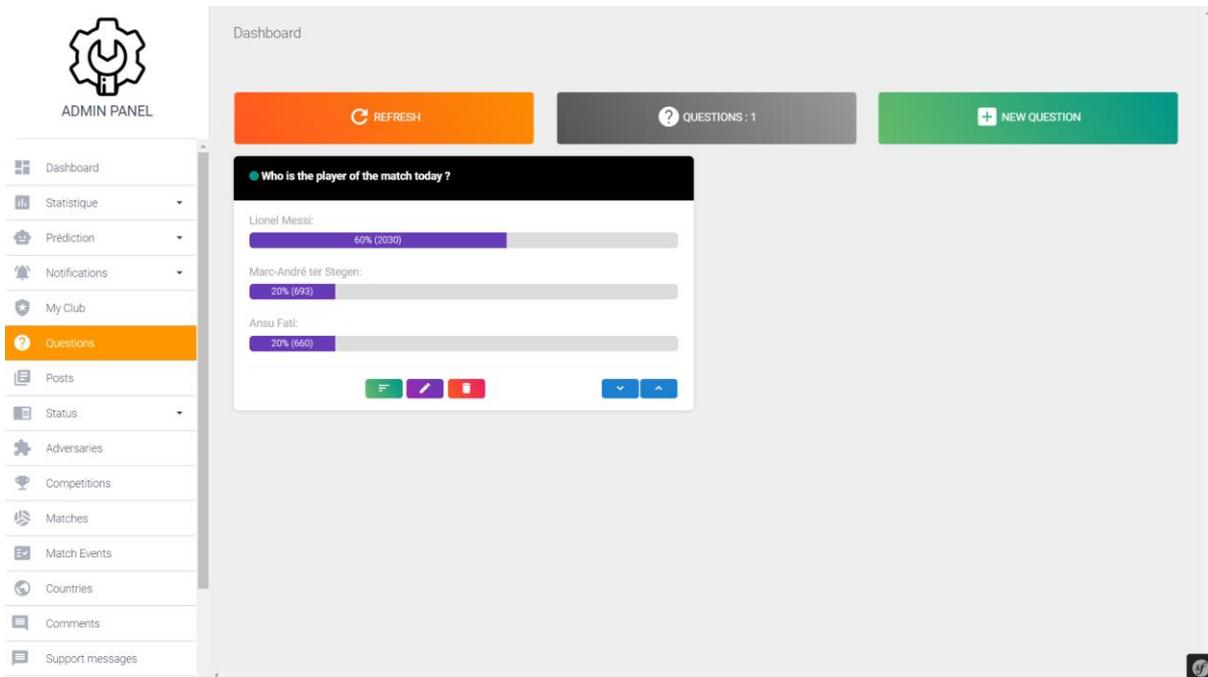


Figure 39 : Page questions

5.6. Pages Post et Statut

L'admin crée des postes et statuts comme illustré par la figure 40 :

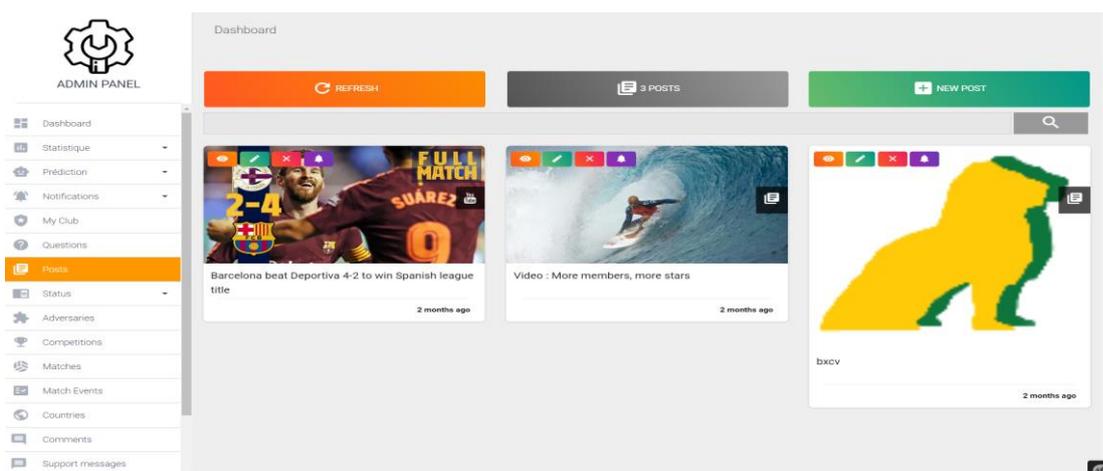


Figure 40 : Page Posts

| CHAPITRE 3

Mise en œuvre

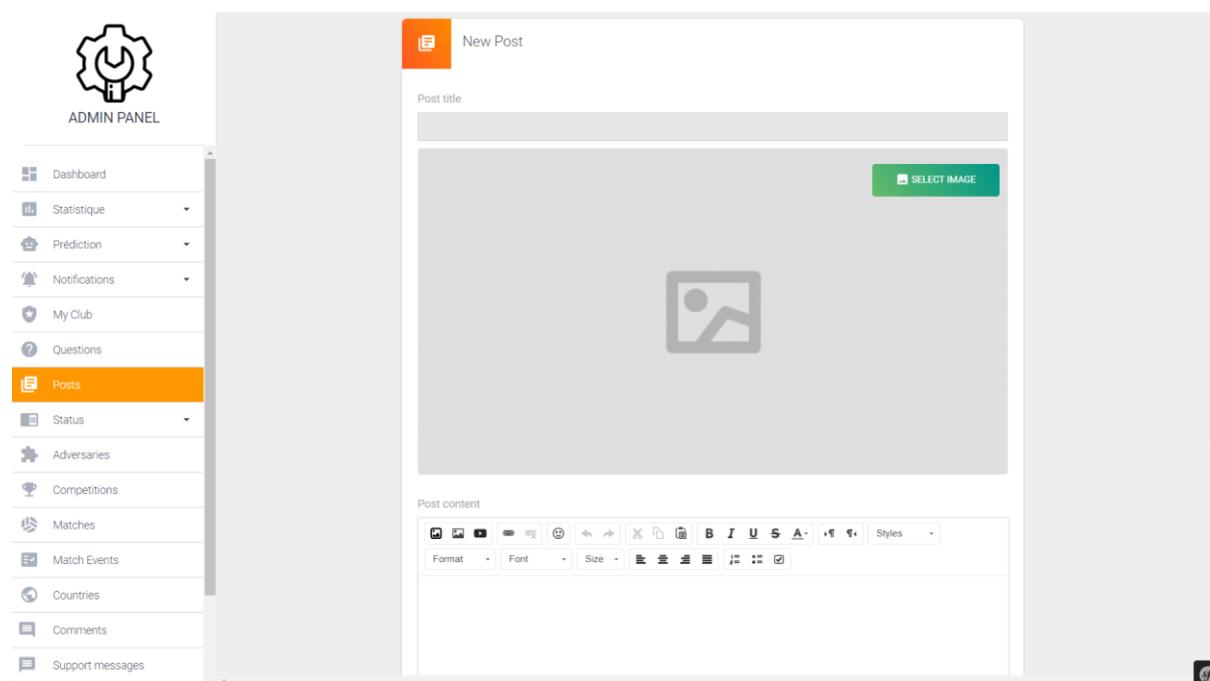


Figure 41: Création d'un Post

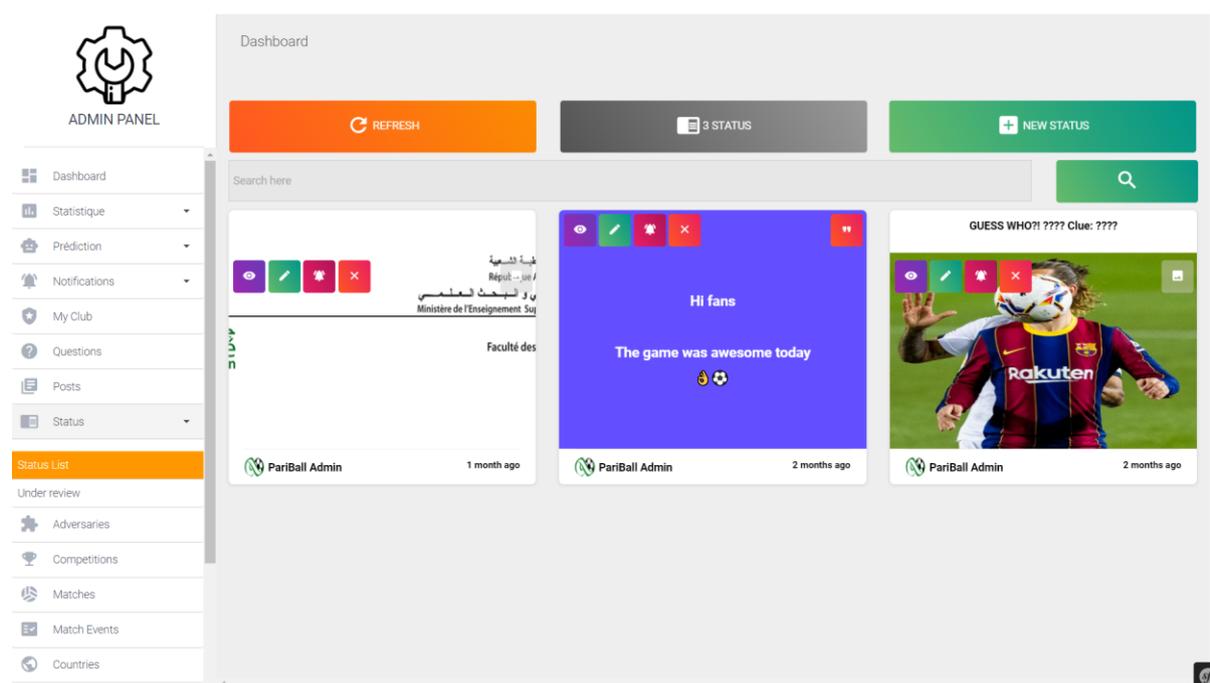


Figure 42 : Liste des statuts

5.7. Page Adversaire

Cette page contient la liste des clubs créés par l'admin ainsi que ses adversaires :

| CHAPITRE 3

Mise en œuvre

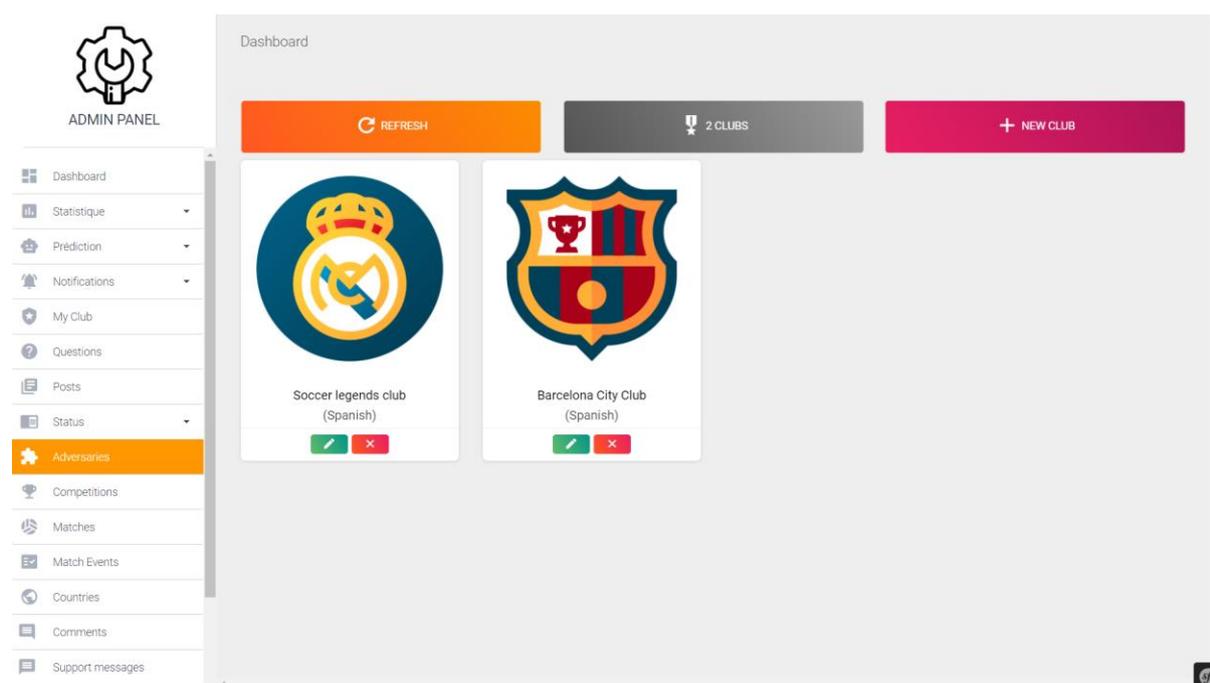


Figure 43 : Page adversaire

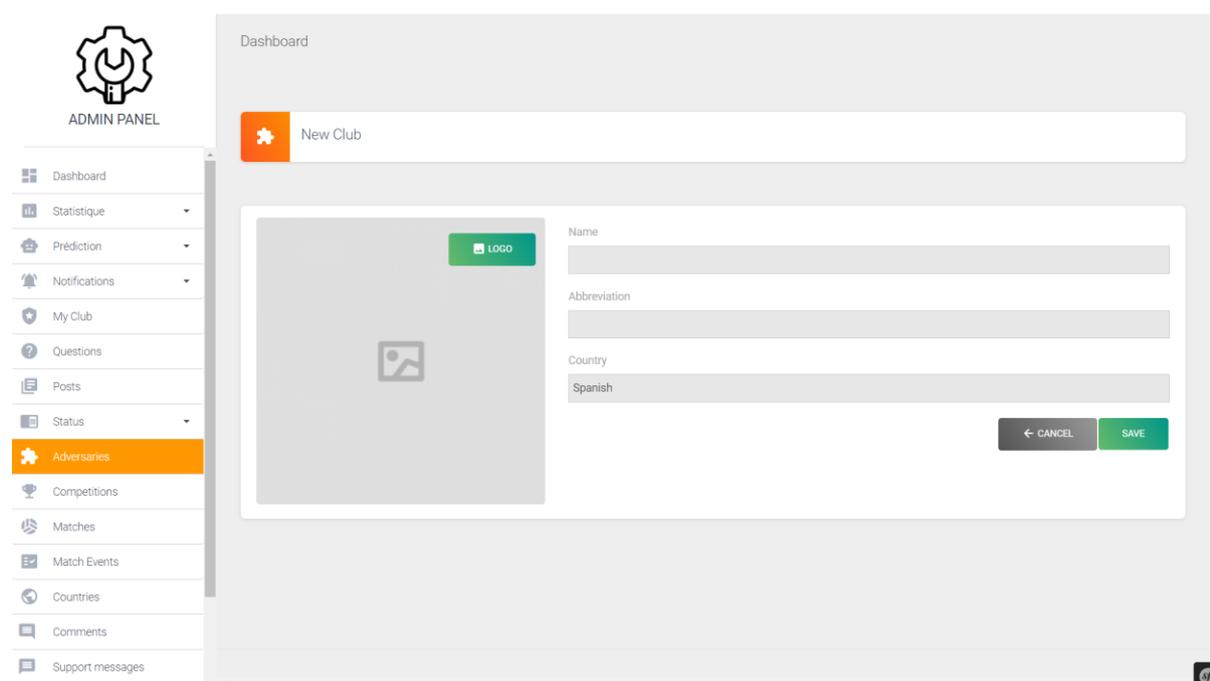


Figure 44 : Création d'un nouveau club

5.8. Page Compétition

Cette page contient les compétitions créées par l'admin :

| CHAPITRE 3

Mise en œuvre

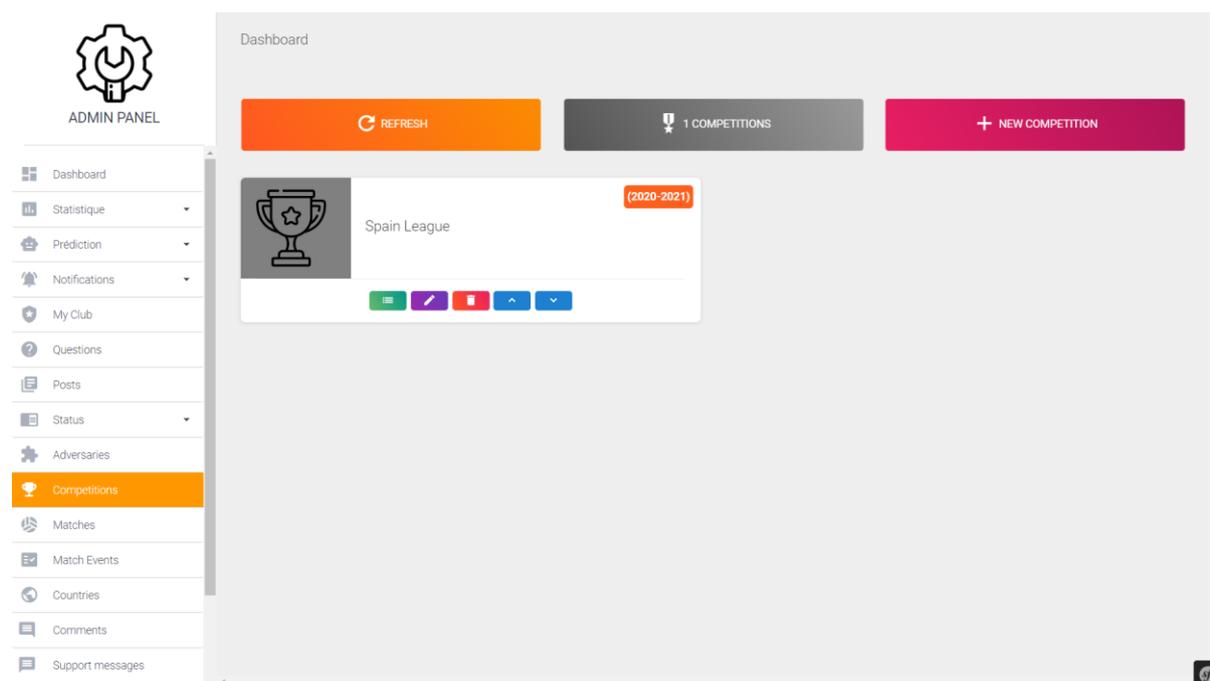


Figure 45: Page compétition

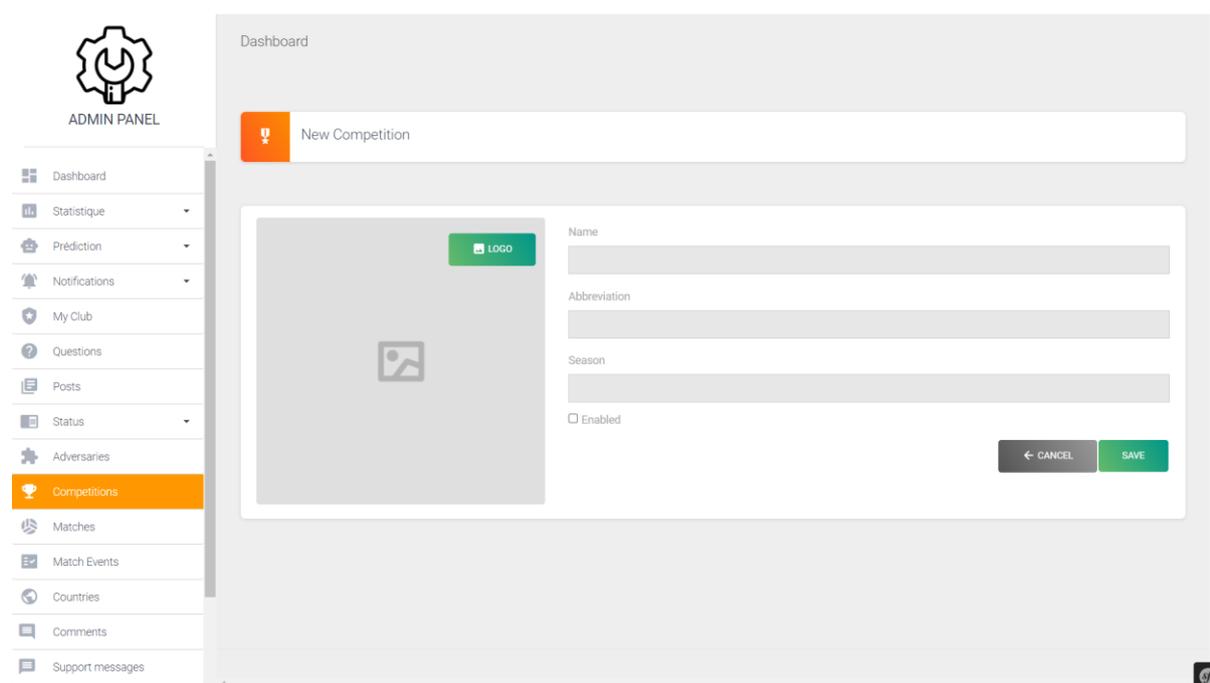


Figure 46 : Création d'une nouvelle compétition

5.9. Page matches

Cette page contient la liste des matchs créés par l'admin pour les utilisateurs d'application qui veulent jouer un match, en prédire le résultat et gagner des points :

| CHAPITRE 3

Mise en œuvre

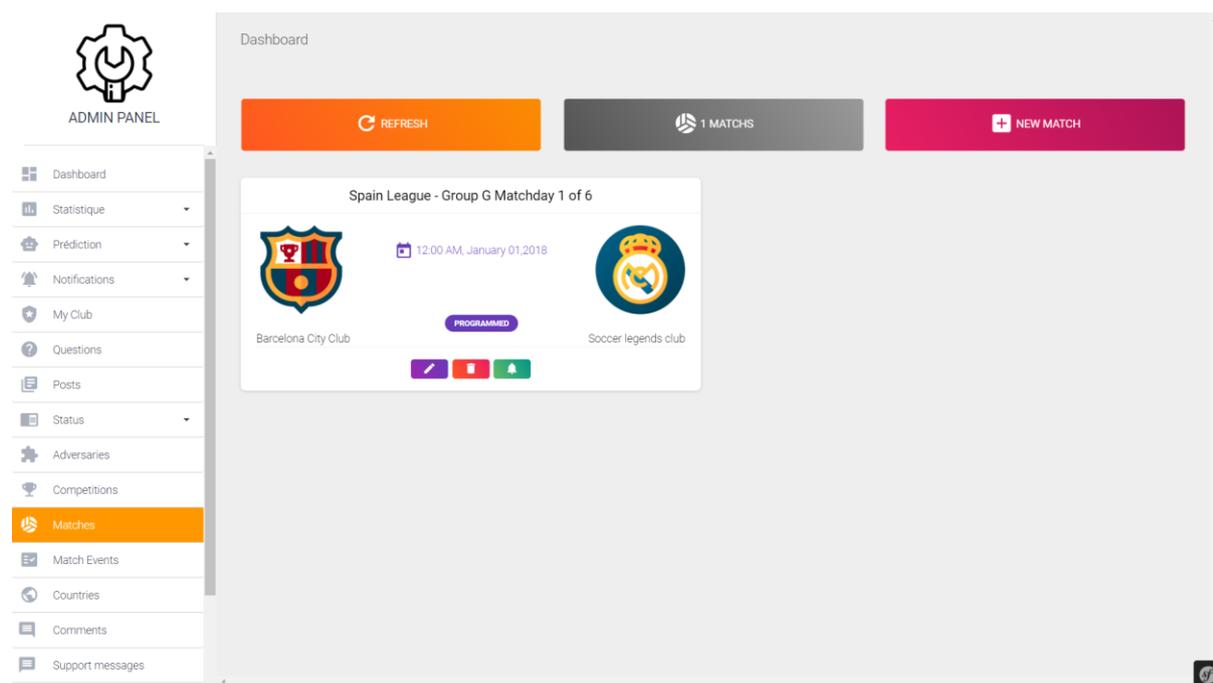


Figure 47 : Page matches

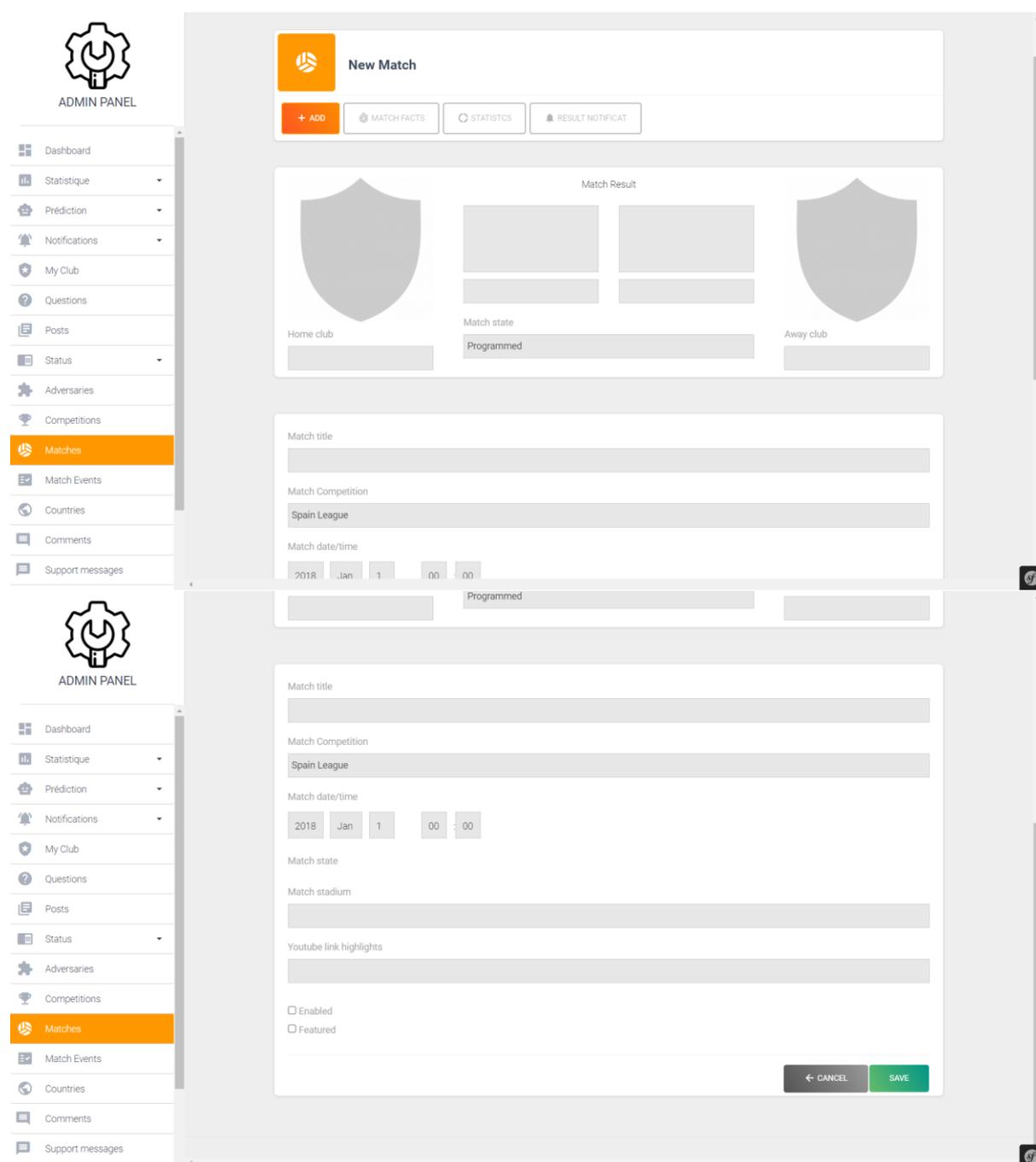


Figure 48 : création d'un nouveau match

5.10. Page support messages

Cette page contient les messages envoyés par les utilisateurs d'application pour signaler un problème dans l'application :

| CHAPITRE 3

Mise en œuvre

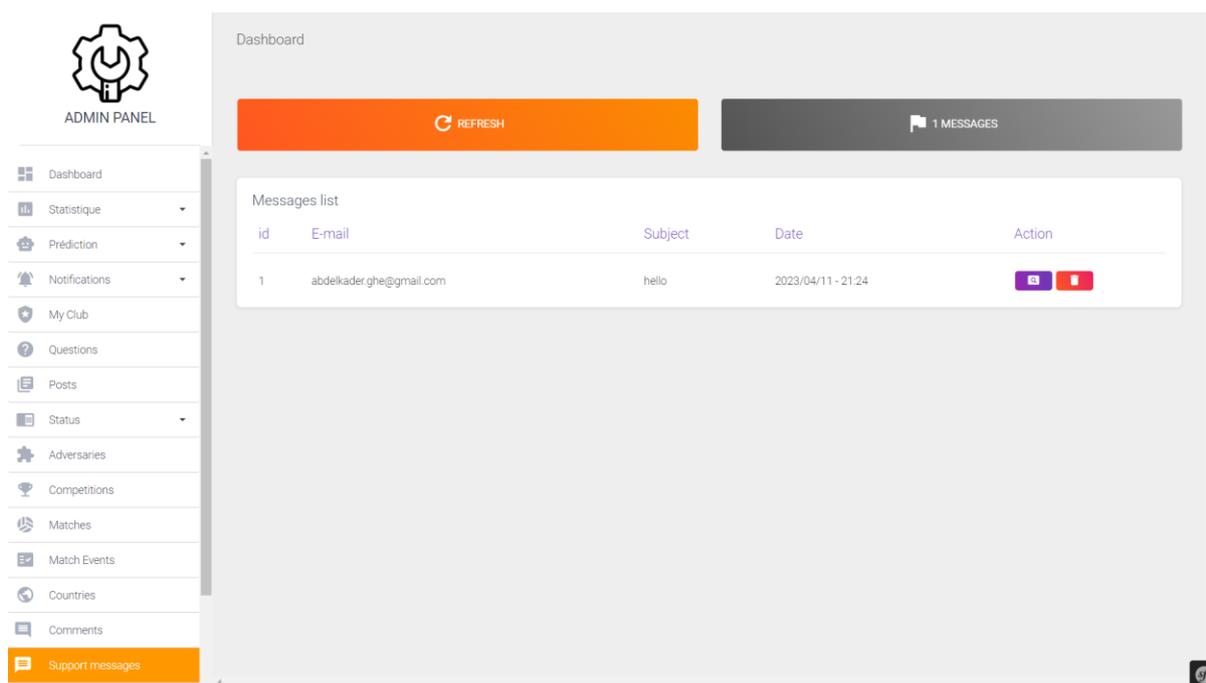


Figure 49 : Page support messages

5.11. Page users

Cette page contient la liste des utilisateurs d'application .

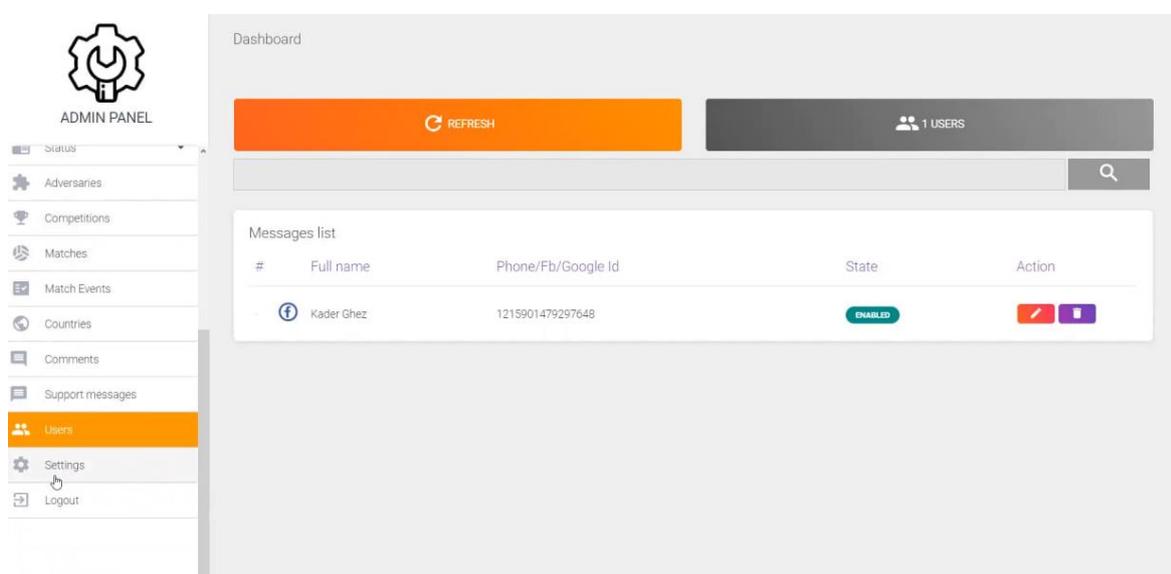


Figure 50 : Page users

5.12. Page admin panel

Cette page contient les paramètres du site :

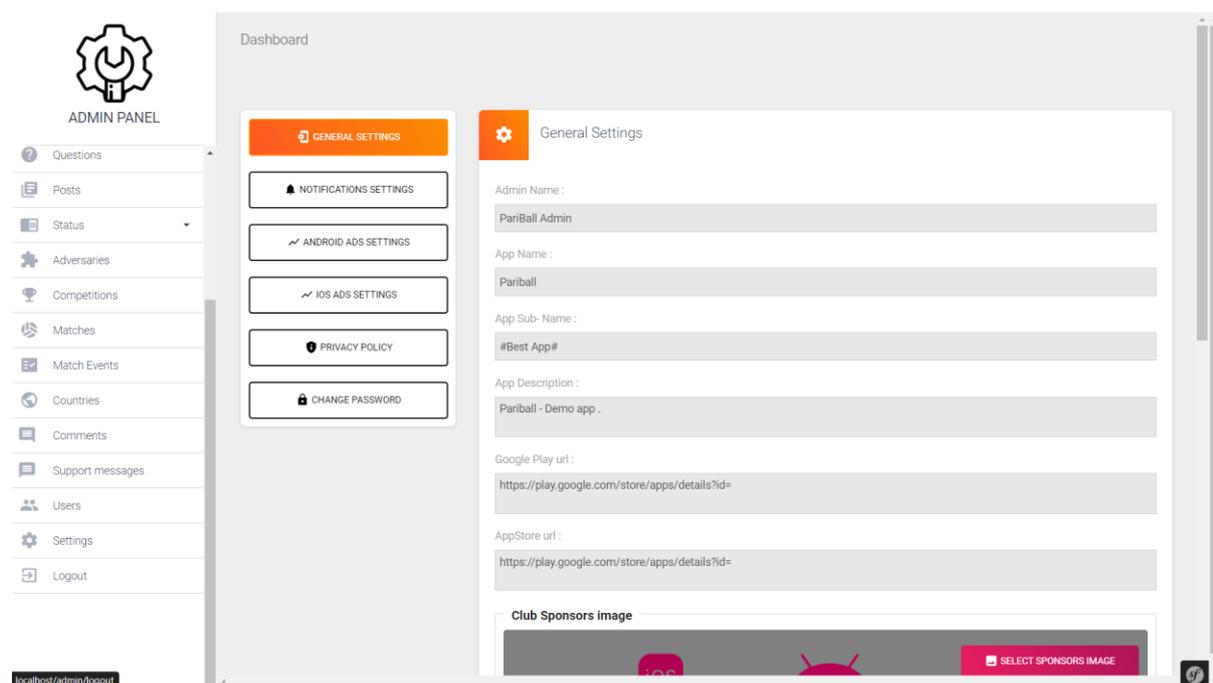


Figure 51: Admin panel

6. Fonctionnalités de l'application proposée « Pariball »

Dans cette partie, nous allons présenter les interfaces les plus importantes de notre application mobile.

6.1. Lancement de l'application

Lors de lancement d'application la première page affiche son nom qui est « Pariball » et logo :

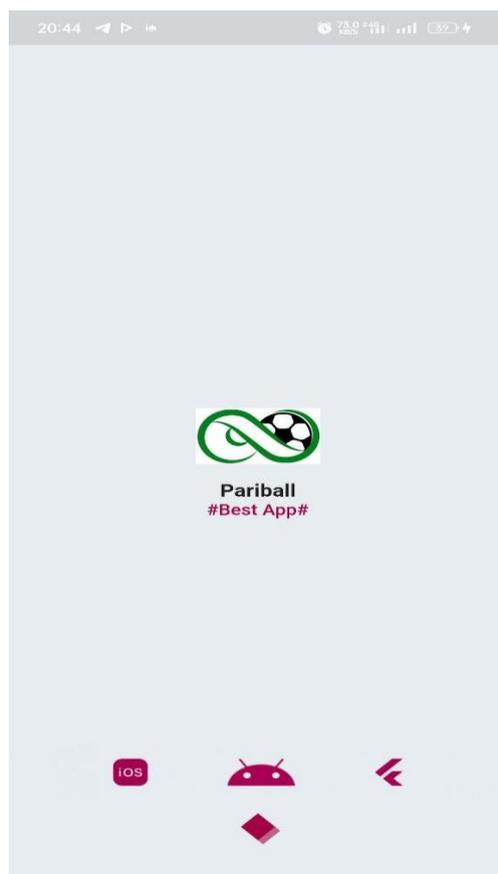


Figure 52: Lancement de Pariball

6.2. Page d'accueil

La page d'accueil associée à l'application contient les questions, postes et statuts créés par l'admin, et le bouton connexion :

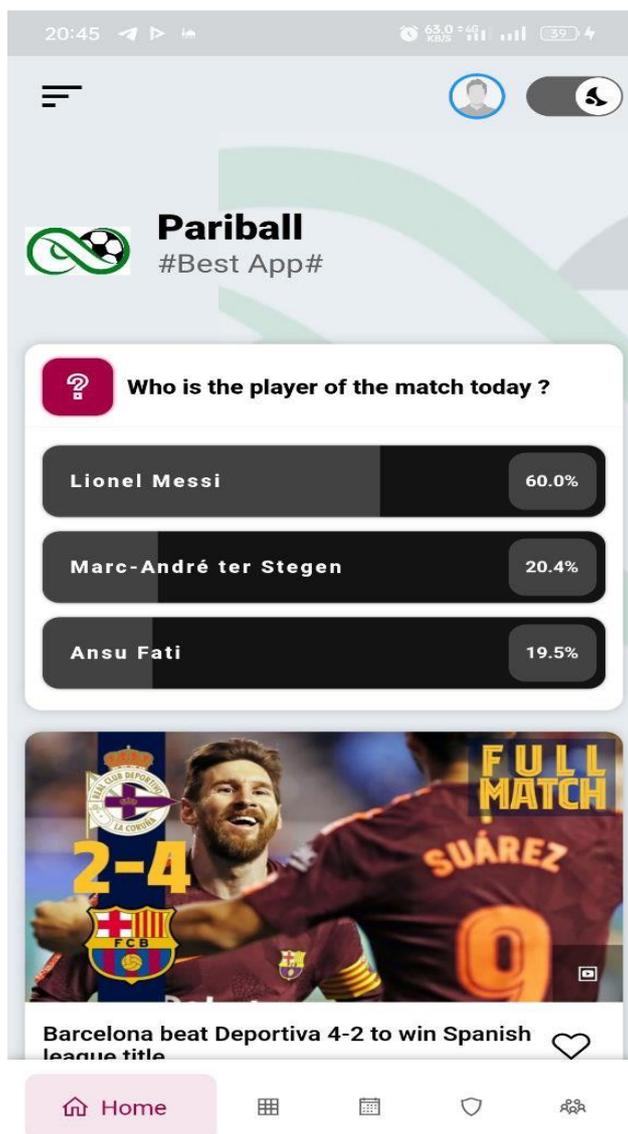


Figure 53: Page d'accueil associée à l'application

6.3. Page authentification et Profil utilisateur

Cette page permet aux utilisateurs de s'authentifier afin d'accéder à leurs espaces de travail selon leurs profils, l'authentification se fait soit par le compte Facebook ou le compte google d'utilisateur ou bien par la référence obtenue d'un ami(autre utilisateur d'application). Cette référence permet à l'utilisateur qui a partagé son référence de gagner des points :

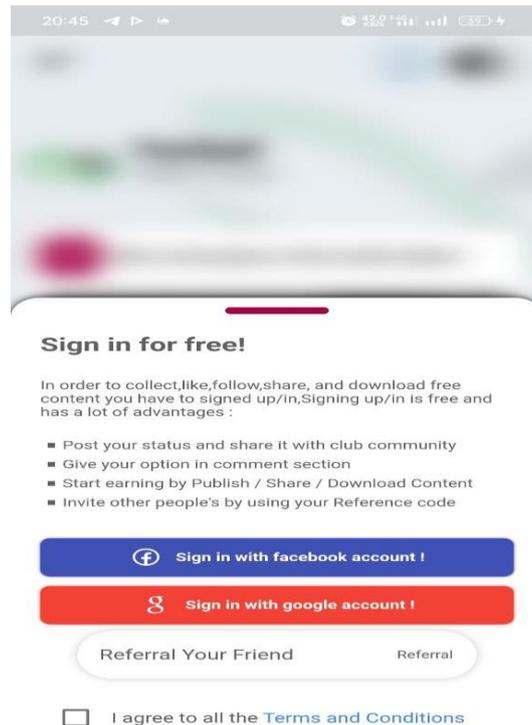


Figure 54 : Page authentication

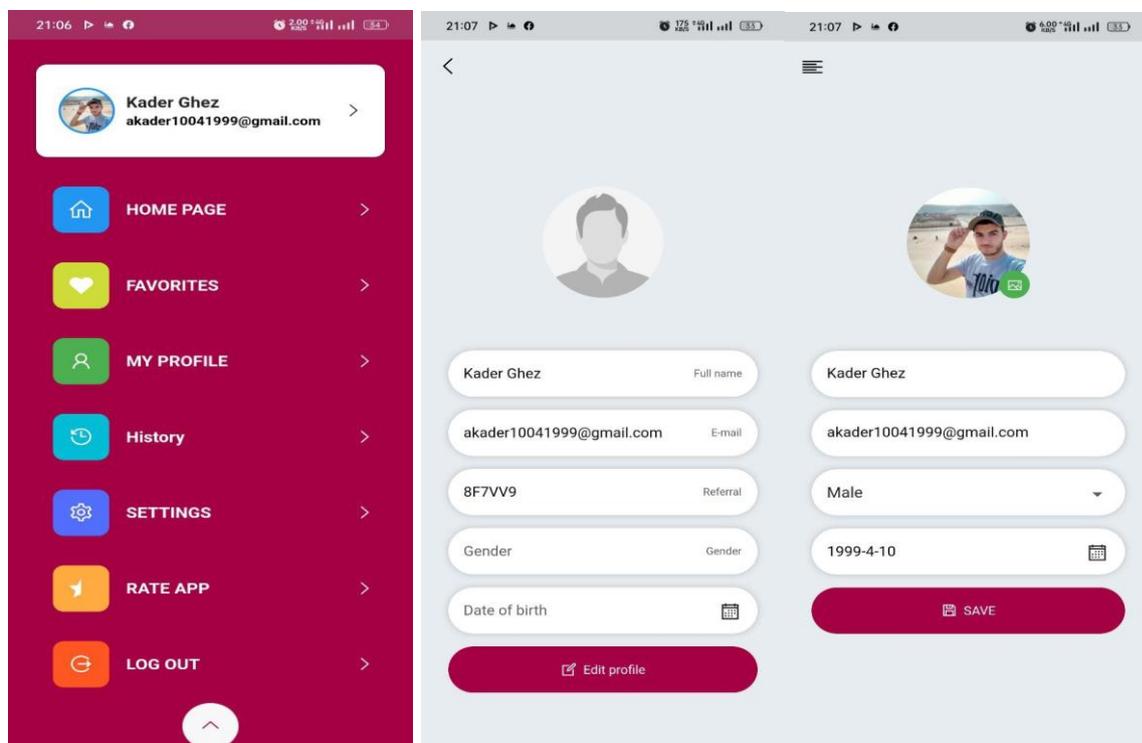


Figure 55 : Profil utilisateur

| CHAPITRE 3

Mise en œuvre

6.4. Page match

Contient la liste de tous les derniers matchs joués ainsi que ses informations et résultats obtenu à travers les API-football.

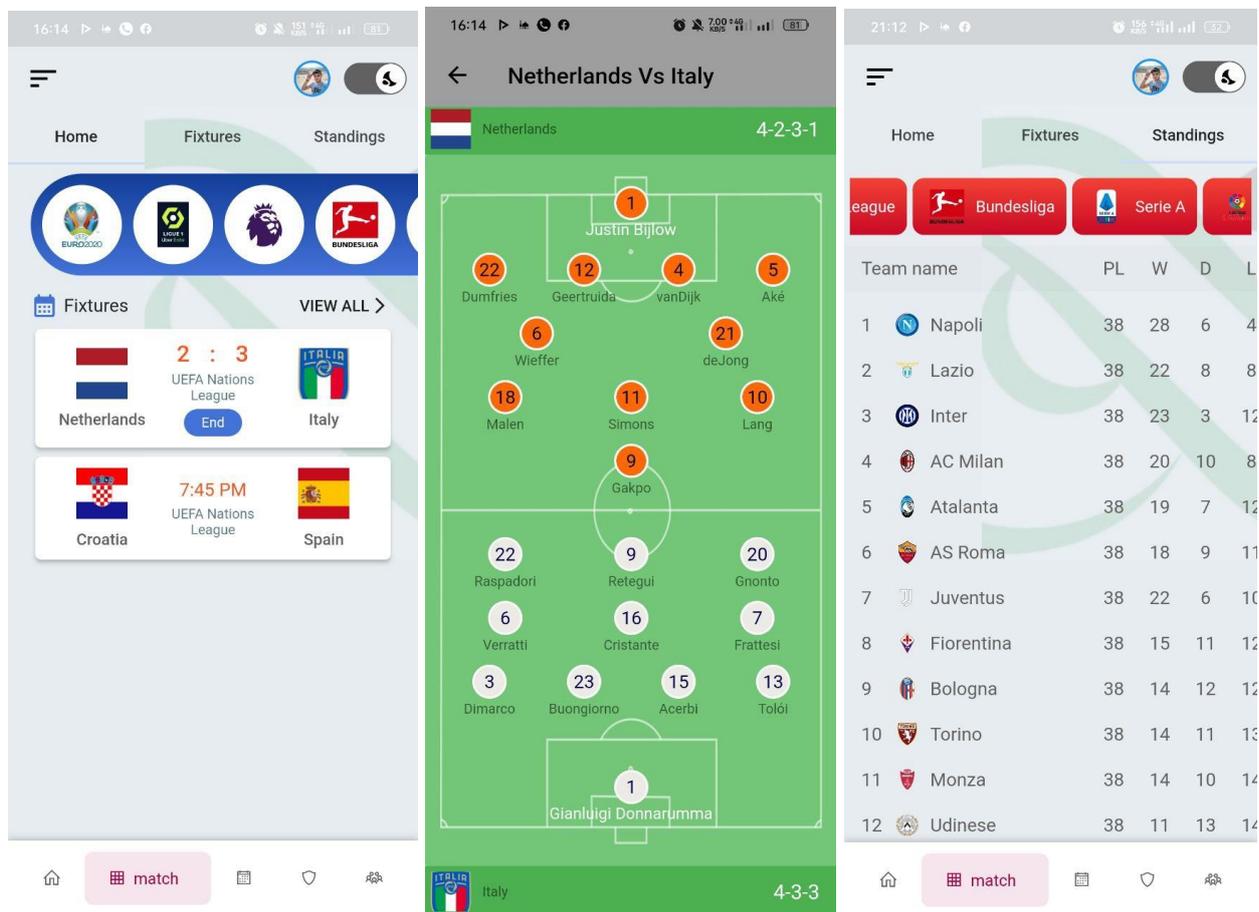


Figure 56 : Page match

6.5. Page bettin

Cette page contient les matchs créés par l'admin pour les utilisateurs afin de prédire des résultats, elle nécessite au début 10 points au minimum pour jouer, si le résultat est vrai ceci permet de gagner des points:

| CHAPITRE 3

Mise en œuvre

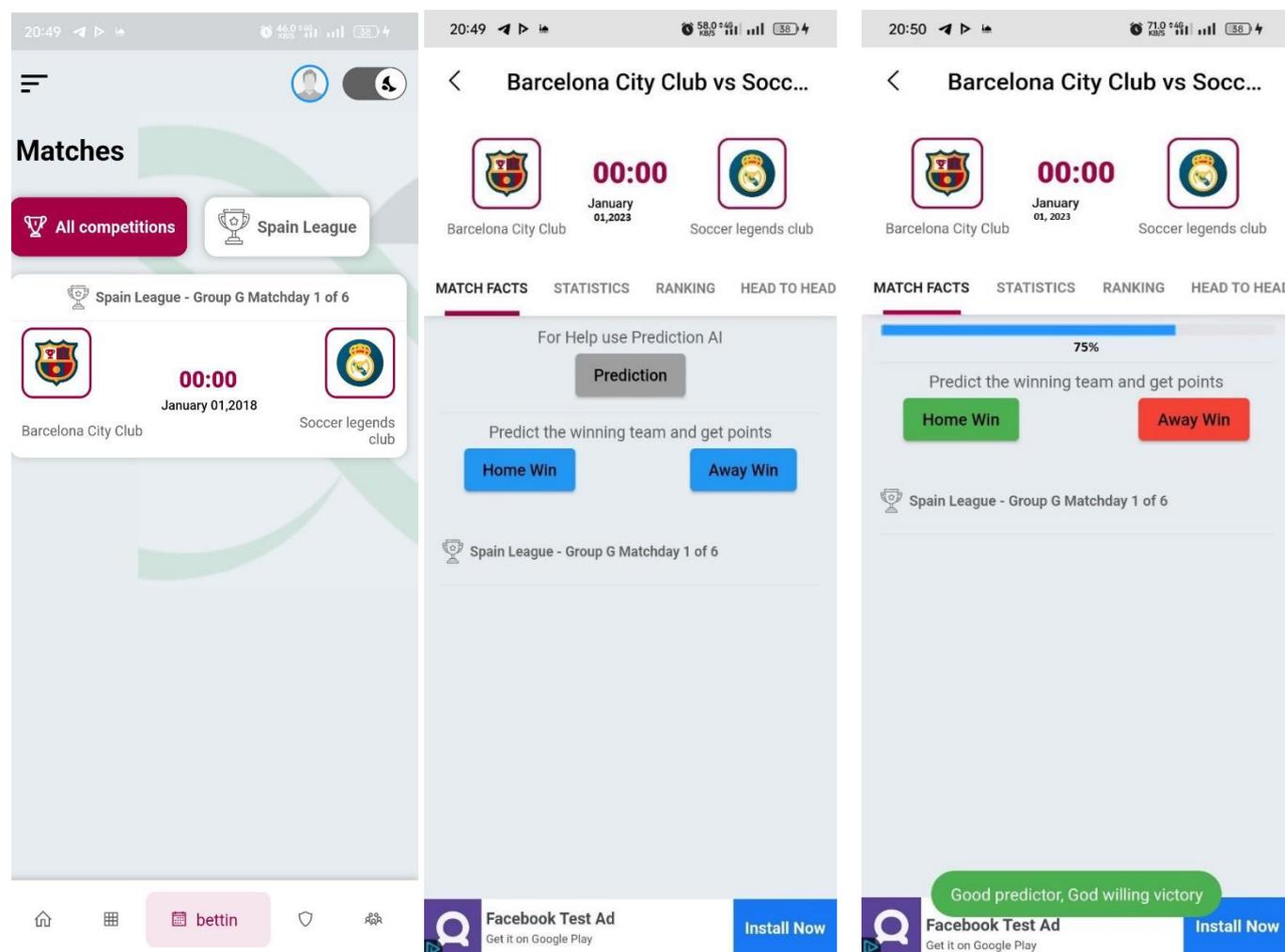


Figure 57 : Prédiction de match

6.6. Page pack

Cette page contient les packages suivants :

- Diamants : convertir les points en diamants et gagner des prix
- Subscriptions : inscrire pour augmenter le taux de prédiction
- Invertisment : obtenir des points à travers la reference d'utilisateur

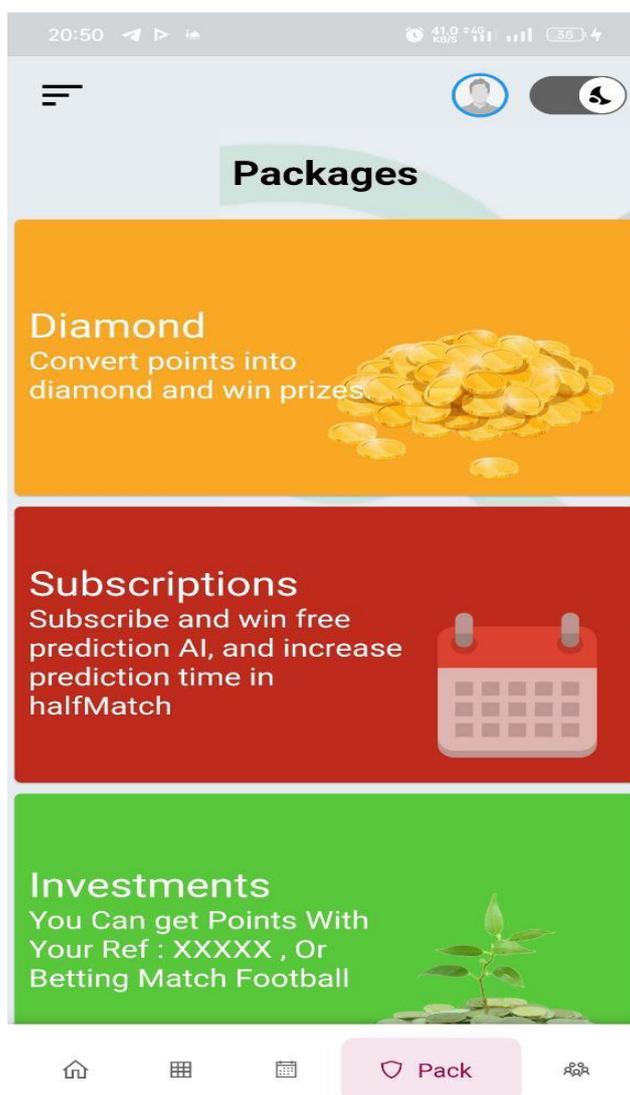


Figure 58 : page pack

6.7. Page zone des fans

Cette page contient les statuts, publications et commentaires des fans, c'est la partie réseau social de l'application :

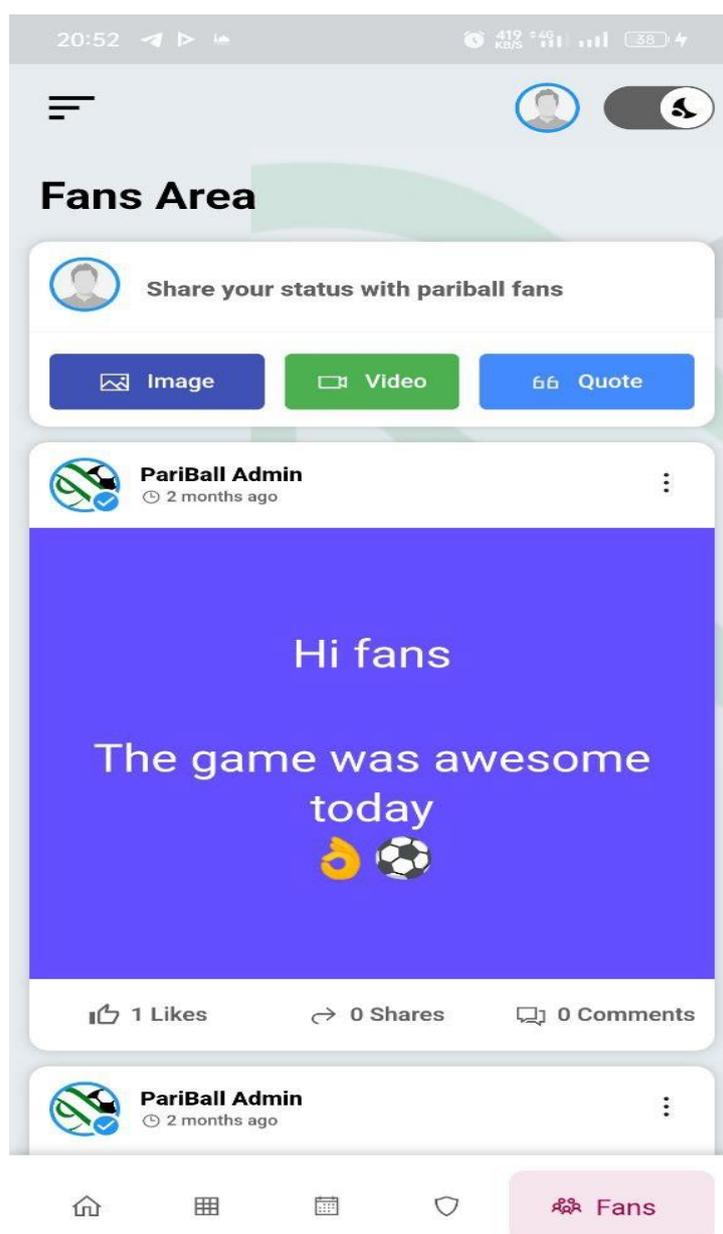


Figure 59: Page fans area

7. Conclusion

Dans ce chapitre nous avons présenté l'environnement de développement ainsi que les différents outils utilisés, ensuite nous avons conduit des scénarios illustratifs afin de décrire certaines fonctionnalités de notre site web et application mobile « Pariball ».

CONCLUSION GÉNÉRALE

Conclusion Générale

Au fil des années, le pronostic sportif, en particulier dans le domaine du football, a connu une popularité croissante parmi les fans qui cherchent à anticiper les résultats des matchs. Les progrès technologiques et l'émergence de l'Intelligence Artificielle (IA) ont ouvert de nouvelles perspectives pour améliorer la précision et la pertinence des prédictions sportives. L'utilisation de l'aide à la décision basée sur l'IA dans le pronostic sportif a ainsi gagné une importance, offrant aux amateurs de football des outils puissants afin d'analyser les données, évaluer les performances et prendre des décisions éclairées.

Dans ce contexte, la présente étude s'est concentrée sur l'implémentation et la mise en œuvre d'un système d'aide à la décision dédié au pronostic sportif, en se basant sur le football. Notre objectif était d'exploiter les techniques de l'Intelligence Artificielle pour fournir des prédictions intelligentes et utiles aux amateurs de football, en se basant sur l'historique des participants. Cette approche permet d'évaluer les performances des équipes, d'estimer les probabilités de victoire, de défaite ou de match nul, et d'optimiser les stratégies de jeu.

L'implémentation de notre système d'aide à la décision reposait sur des méthodes d'IA notamment l'apprentissage automatique. En utilisant des données sportives disponibles gratuitement sous forme d'API sur Internet, nous avons pu entraîner notre système à fournir des prédictions fiables et précises. Les utilisateurs ont la possibilité d'utiliser ces prédictions comme guide pour leurs propres pronostics ou de les combiner avec leurs connaissances personnelles pour prendre des décisions éclairées.

Notre approche prend également en compte les aspects socio-économiques, en particulier l'impact sur la population jeune intéressée par le football local et international. En fournissant des prédictions intelligentes et des informations précieuses, notre système vise à maximiser les chances de succès des utilisateurs dans leurs activités de pronostic sportif, tout en diminuant leur engagement dans le monde du football.

En conclusion, ce projet a montré les avantages de l'utilisation de l'Intelligence Artificielle et de l'aide à la décision dans le pronostic sportif, en mettant l'accent sur le football. En exploitant les techniques d'IA, il est possible d'analyser les données sportives de manière approfondie, d'anticiper les résultats des matchs et d'optimiser les stratégies de jeu.

Cependant, il existe encore des perspectives d'extension et d'amélioration, notamment:

- L'exploration de nouvelles techniques d'IA ;
- L'intégration davantage de sources de données ;
- L'élargissement de la portée du système à d'autres sports.

Ces avancées continueront à renforcer la pertinence et la précision des prédictions sportives basées sur l'IA,

Bibliographie

- [1] «l'impact de la technologie sur l'évolution du football,» FutballLab, [En ligne]. Available: <https://www.futbollab.com/fr/actualites/shift-de-paradigme-impact-de-la-technologie-sur-levolution-du-football>. [Accès le 23 05 2023].
- [2] «quand l'informatique juge les performances,» ZDNet, [En ligne]. Available: <https://www.zdnet.fr/actualites/football-quand-l-informatique-juge-les-performances-39351878.htm>. [Accès le 05 05 2023].
- [3] A. L. J. Paulus, «Outil d'aide à la décision et de prévention dans les sports collectifs. Exemple de centre d'entraînement de rugby,» *Journal de Traumatologie du Sport*, vol. 35, n° 13, pp. 134-142, oct 2018.
- [4] M. Benbrahim, «Introduction à l'Intelligence Artificielle,» GI, 2020.
- [5] M. Minsky, ««Steps toward Artificial Intelligence,»» MIT, 1968.
- [6] s. d. F. Torre, ««l'Intelligence Artificielle,»» université de Lille, [En ligne]. Available: <https://fabien-torre.fr/Enseignement/Cours/Intelligence-Artificielle/>. [Accès le 16 06 2023].
- [7] A. Samuel, ««Pioneer in Machine Learning,»» *Ibm Journal of Research and Development*, 1992.
- [8] S.Adnene, ««Apprentissage profond pour la classification d'actions dans les vidéos de football,»» Tébessa.
- [9] S. W. K. K. L. L. Z. Bio-Inspired Credit Risk Analysis: Computational Intelligence with Support Vector Machine By Lean Yu.
- [10] M.Belalia, «Classification des images selon la sémantique,» mémoire master , université de Mostaganem, 2016.
- [11] K. Rima, «difference-machine-learning-deep-learning,» 2014.
- [12] G. K. Cyril Bossard, «La prise de décision des joueurs de sports collectifs,» *Movement & Sport Sciences*, vol. 2, n° 173, pp. 3-22, 2011.
- [13] N. &. Locascio, «Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. " O'Reilly Media, Inc.",.» 2017.
- [14] B.Stuner, «Cohorte de réseaux de neurones récurrents pour la reconnaissance de l'écriture,» *Doctoral dissertation, Normandie*, 2018.
- [15] A.Angel, «Towards Distortion-Predictable Embedding of Neural Networks,» 2015.
- [16] «Predict395,» [En ligne]. Available: <https://predict395.com/>.
- [17] «1xBet,» [En ligne]. Available: <https://1xbet.com/fr>.

| Références

- [18] M. KAWTHER, «UN SYSTEME INTELLIGENT D'AIDE A LA DECISION POUR LA PREVENTION DES RUPTURES D'APPROVISIONNEMENT EN MEDICAMENTS,» Mémoire de Fin d'Etudes Pour l'Obtention du Diplôme de Master en Informatique, Université Oran1 Ahmed Ben Bella, 2022.
- [19] O. D. AIT TAHAR SABRINA, «Classification d'images A l'aide des réseaux de neurones convolutifs,» 2017.
- [20] S.SADI, «Implémentation et évaluation d'un modèle d'apprentissage automatique pour l'estimation de la valeur marchande de propriétés,» immobilières, 2020.
- [21] Z. Belkacem, «VERS UN SYSTEME INTELLIGENT D'AIDE A LA DECISION DEDIE A LA PHARMACOVIGILANCE,» Mémoire de Fin d'Etudes pour l'obtention du Diplôme de Master en informatique , 2022.
- [22] «"PHP for the Web: Visual QuickStart Guide" de Larry Ullman,» [En ligne].
- [23] «"Symfony 5: The Fast Track" de Fabien Potencier,» [En ligne].
- [24] s. o. d. Symfony. [En ligne]. Available: <https://symfony.com/>.
- [25] s. o. d. MySQL. [En ligne]. Available: <https://www.mysql.com/>.
- [26] «Learning MySQL: Get a Handle on Your Data de Seyed Mohammad Mehdi Moosavi,» [En ligne].
- [27] «"JavaScript: The Definitive Guide" de David Flanagan,» [En ligne].
- [28] «Site officiel de JavaScript,» [En ligne]. Available: <https://developer.mozilla.org/fr/docs/Web/JavaScript>.
- [29] «JavaScript and JQuery: Interactive Front-End Web Development de Jon Duckett,» [En ligne].
- [30] «Site officiel de CSS,» [En ligne]. Available: <https://www.w3.org/Style/CSS/>.
- [31] «CSS: The Definitive Guide de Eric Meyer,» [En ligne].
- [32] S. o. d. HTML. [En ligne]. Available: <https://www.w3.org/html/>.
- [33] «Site officiel de Bootstrap,» [En ligne]. Available: <https://getbootstrap.com/>.
- [34] «Site officiel de Twig,» [En ligne]. Available: <https://twig.symfony.com/>.
- [35] «"Twig for Template Designers" de Armin Ronacher,» [En ligne].
- [36] «Flutter,» [En ligne]. Available: <https://aws.amazon.com/fr/what-is/flutter/>.
- [37] Dart. [En ligne]. Available: <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/le-langage-de-programmation-dart/>.
- [38] «Site officiel de BeautifulSoup,» [En ligne]. Available: <https://www.crummy.com/software/BeautifulSoup/>.
- [39] [En ligne]. Available: <https://towardsdatascience.com/scikit-learn-tensorflow->

| Références

- [40] D.ABDELBAQI, «Failure prediction in cloud environment using deep learning,» Thèse.
- [41] «Site officiel de Matplotlib,» [En ligne]. Available: <https://matplotlib.org/>.
- [42] «Site officiel de Visual Studio Code,» [En ligne]. Available: <https://code.visualstudio.com/>.
- [43] «Site officiel de Jupyter Notebook,» [En ligne]. Available: <https://jupyter.org/>.
- [44] [En ligne]. Available: <https://www.php.net/>.
- [45] «MySQL Explained: Your Step-by-Step Guide de Andrew Comeau et Stephen Burge,» [En ligne].
- [46] <https://www.lefigaro.fr/secteur/high-tech/comment-l-intelligence-artificielle-s-immisce-dans-le-football-professionnel-20210512>. [En ligne].
- [47] https://www.francetvinfo.fr/sports/foot/l-intelligence-artificielle-au-service-du-foot-on-collecte-deux-millions-de-points-de-donnees-par-match_4680221.html. [En ligne].
- [48] [En ligne]. Available: <https://www.futura-sciences.com/tech/actualites/technologie-coupe-monde-football-2022-intelligence-artificielle-service-detection-hors-jeu-95230/>.
- [49] [En ligne]. Available: <https://www.strategie.gouv.fr/point-de-vue/impacts-de-lintelligence-artificielle-lavenir-travail>.
- [50] [En ligne]. Available: <https://larevueia.fr/lintelligence-artificielle-au-service-du-football/>.
- [51] [En ligne]. Available: <https://www.economie.gouv.fr/entreprises/intelligence-artificielle>.