

. Asma Ghariani
. Ghassen Abidi
. Malek Fitouri
. Rayen Rekik
. Taher Moussa
. Paul Ulrich Koulagna

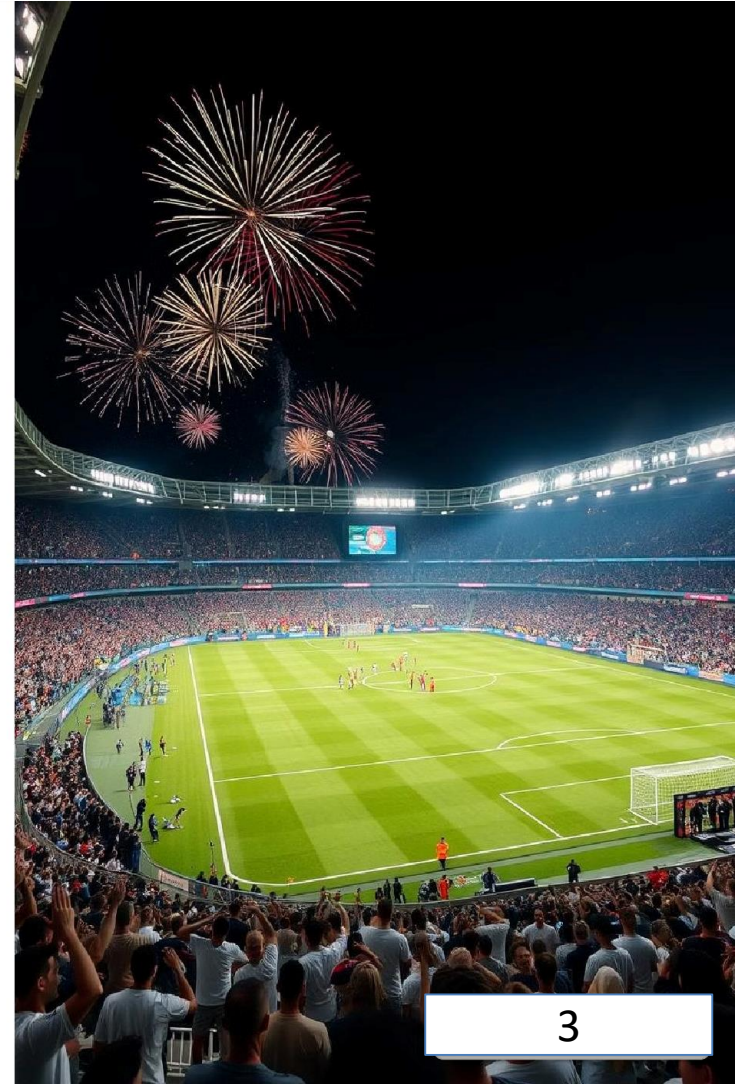


Sommaire

1	Introduction
2	Objectives du Projet
3	Méthodologie : CRISP-DM
4	Modélisation et Résultats
5	Déploiement et Visualisation
6	Conclusion et Recommandations

1.Introduction

Le football, un sport suivi par des milliards de fans à travers le monde, est aujourd'hui à la croisée des chemins entre passion et technologie. Grâce au **Machine Learning**, nous pouvons analyser en profondeur les statistiques des matchs pour prédire le **résultat des rencontres** et estimer le **nombre total de buts**. Ce projet exploite des données telles que les tirs cadrés, la possession ou encore les passes réussies pour offrir des analyses précises, bénéfiques pour les **clubs**, **analystes** et **fans**.



2.Objectifs du Projet

Le but du projet est de **prédire les résultats des matchs de football** (victoire, défaite ou match nul) et **d'estimer le nombre total de buts marqués** pendant un match, en utilisant des **modèles de Machine Learning** sur des données collectées.

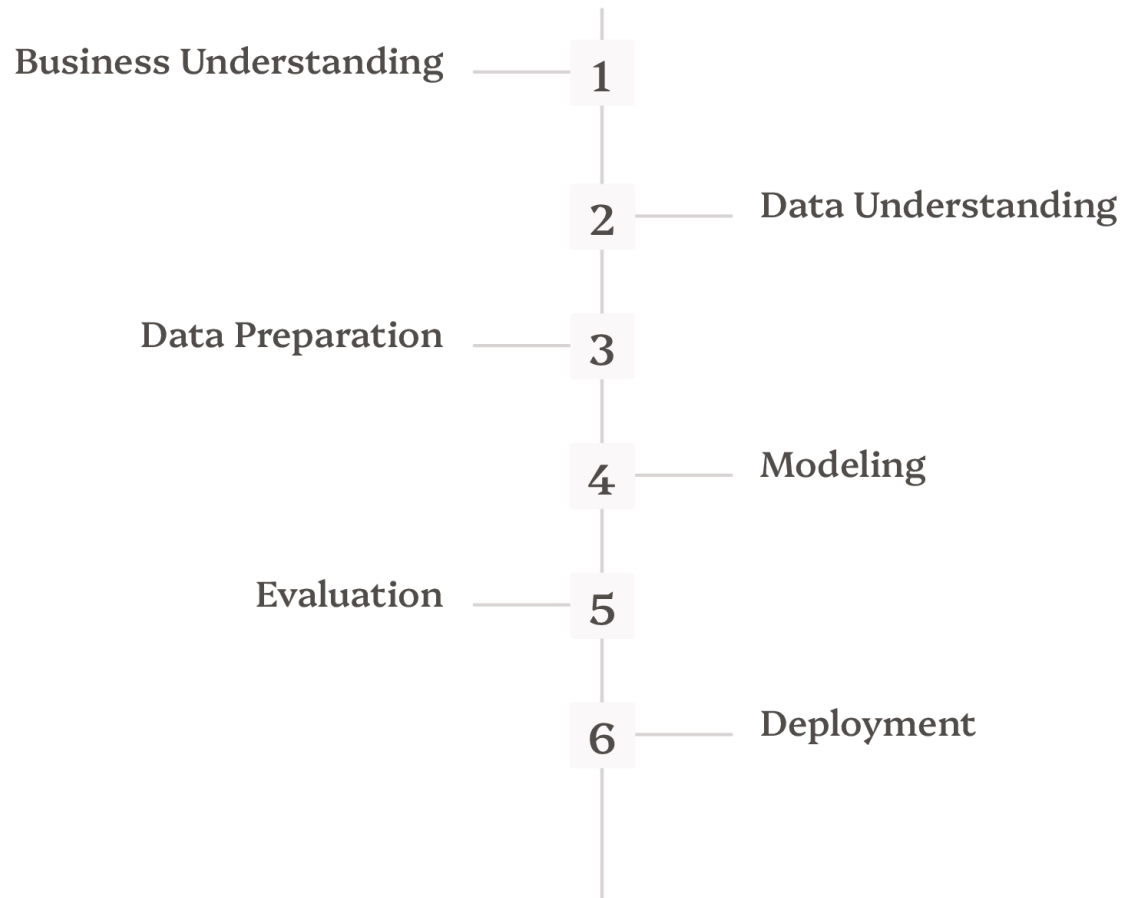
- **Prédiction du résultat du match (Classification) :**

Utiliser les caractéristiques comme les tirs cadrés, la possession, et les passes réussies pour classer les matchs en trois catégories : victoire, défaite ou match nul.

- **Estimation du nombre de buts marqués (Régression) :**

Estimer précisément le nombre de buts marqués dans un match, en tenant compte de statistiques comme les tirs, les corners, et les performances des équipes.

3.La Méthodologie CRISP-DM



Compréhension du Projet et des Données

Compréhension du Business(Business Understanding):

- **Objectif** : Prédire le **résultat des matchs** (victoire, défaite, match nul) et estimer le **nombre total de buts marqués** pendant un match.
- **Données** : Statistiques des matchs (tirs cadrés, possession, passes réussies, historique des résultats, etc.).

Compréhension des Données (Data Understanding):

- **Données collectées** : Statistiques des matchs, résultats passés des équipes.
- **Analyse exploratoire (EDA)** : Visualisation des variables comme les tirs cadrés, la possession, etc.

Préparation des Données (Data Preparation):

- **Nettoyage des données** : Suppression des valeurs manquantes et doublons.
- **Transformation des données** : Encodage des résultats, mise à l'échelle des variables numériques.
- **Séparation des données** : Jeu d'entraînement et jeu de test.

Modélisation, Évaluation et Déploiement

Modélisation (Modeling):

- **Classification** pour prédire le résultat : Random Forest, XGBoost, Logistic Regression.
- **Régression** pour estimer les buts : Random Forest Regressor, XGBoost Regressor.

Évaluation (Evaluation):

- **Métriques de performance** : Accuracy, F1-Score pour la classification, MSE et R^2 pour la régression.
- **Sélection du meilleur modèle** : Choisir le modèle le plus performant.

Déploiement (Deployment):

- **Visualisation des résultats** : Graphiques comparant les résultats réels et prévus.
- **Recommandations pour amélioration future** : Ajouter des données supplémentaires, tester des modèles plus complexes.

4.Modélisation

Classification :

1 Random Forest Classifier

Un ensemble d'arbres de décision pour une meilleure gestion des données bruyantes.

3 Logistic Regression

Modèle simple, utile pour des tâches de classification binaire (ici, adapté à une classification multiclasse avec ajustements).

2 XGBoost Classifier

Un algorithme basé sur le gradient boosting, souvent plus performant pour des problèmes complexes.

4 Gradient Boosting Classifier

Un autre algorithme basé sur le boosting, similaire à XGBoost mais avec une approche différente.

Regression :

MSE

La moyenne des erreurs au carré, une mesure de la précision du modèle.

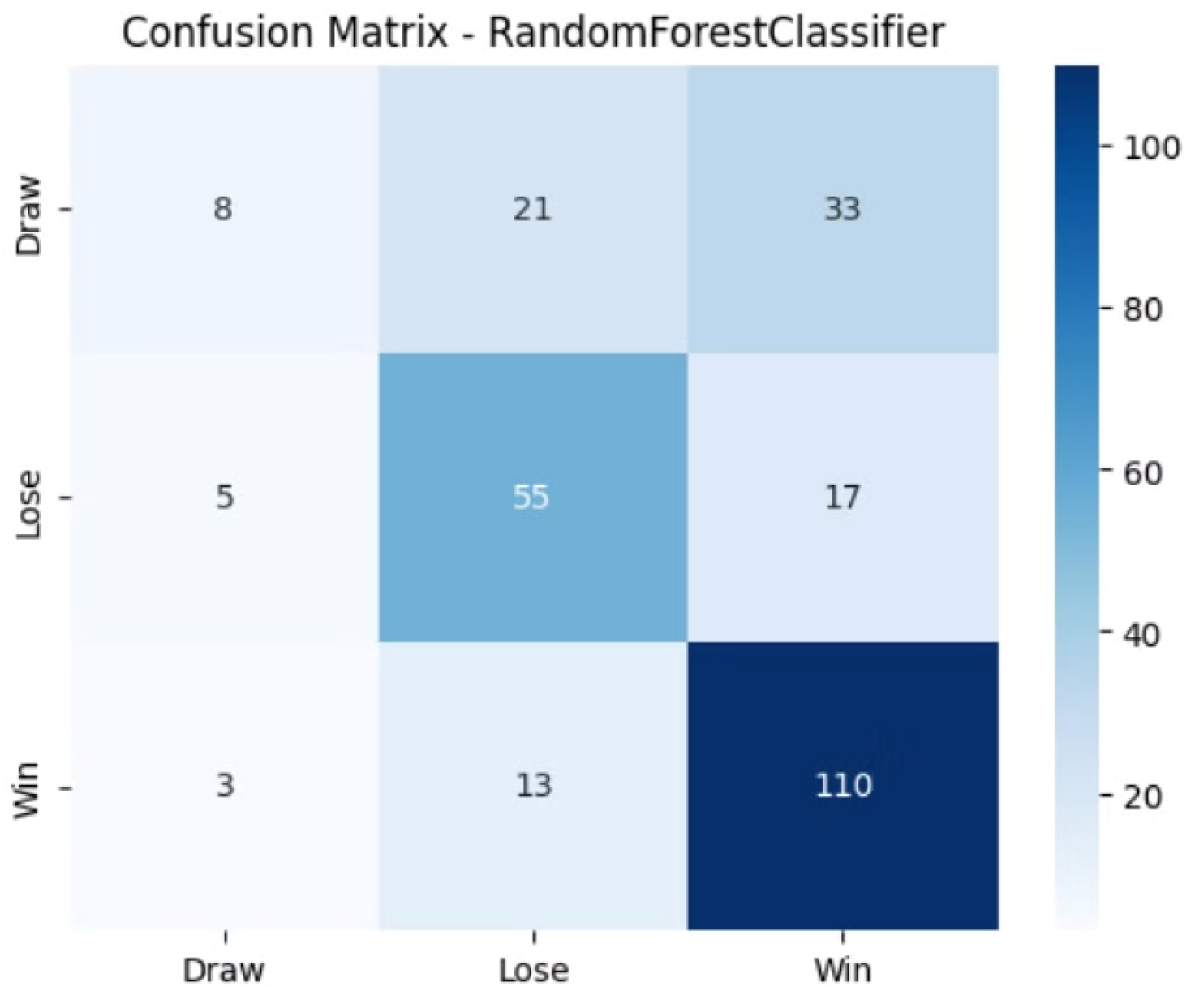
R^2

Indicateur qui montre la proportion de la variance expliquée par le modèle, plus R^2 est proche de 1, mieux c'est.

MAE

Moyenne des erreurs absolues, qui montre la distance moyenne entre les valeurs réelles et les prédictions.

Random Forest Classifier



Results for RandomForestClassifier - Classification:

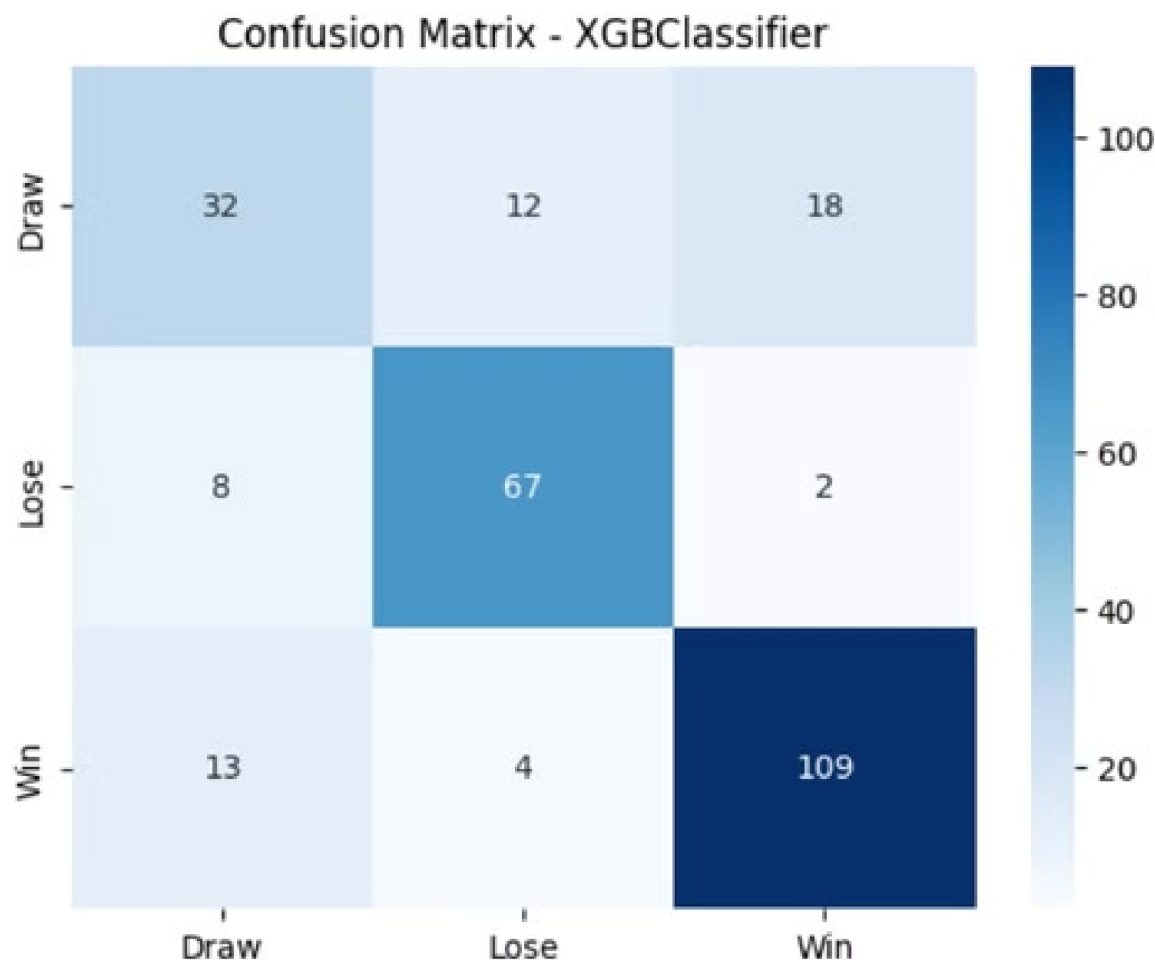
Accuracy: 0.6528301886792452

F1 Score: 0.606283479345539

Precision: 0.6234312062751749

Recall: 0.6528301886792452

XGBoost Classifier



Results for XGBClassifier - Classification:

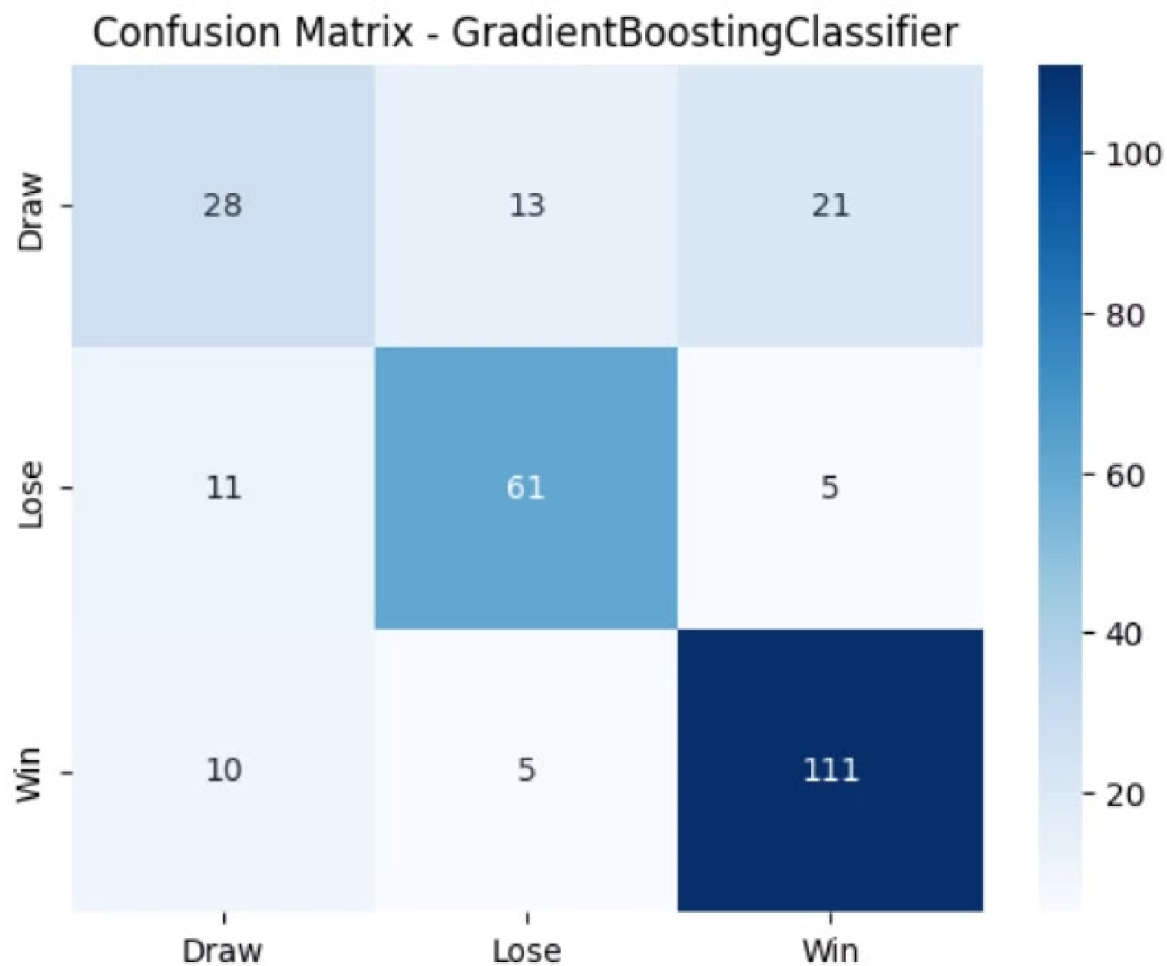
Accuracy: 0.7849056603773585

F1 Score: 0.7800358297543792

Precision: 0.777568698299037

Recall: 0.7849056603773585

Gradient Boosting Classifier:



Results for GradientBoostingClassifier - Classification:

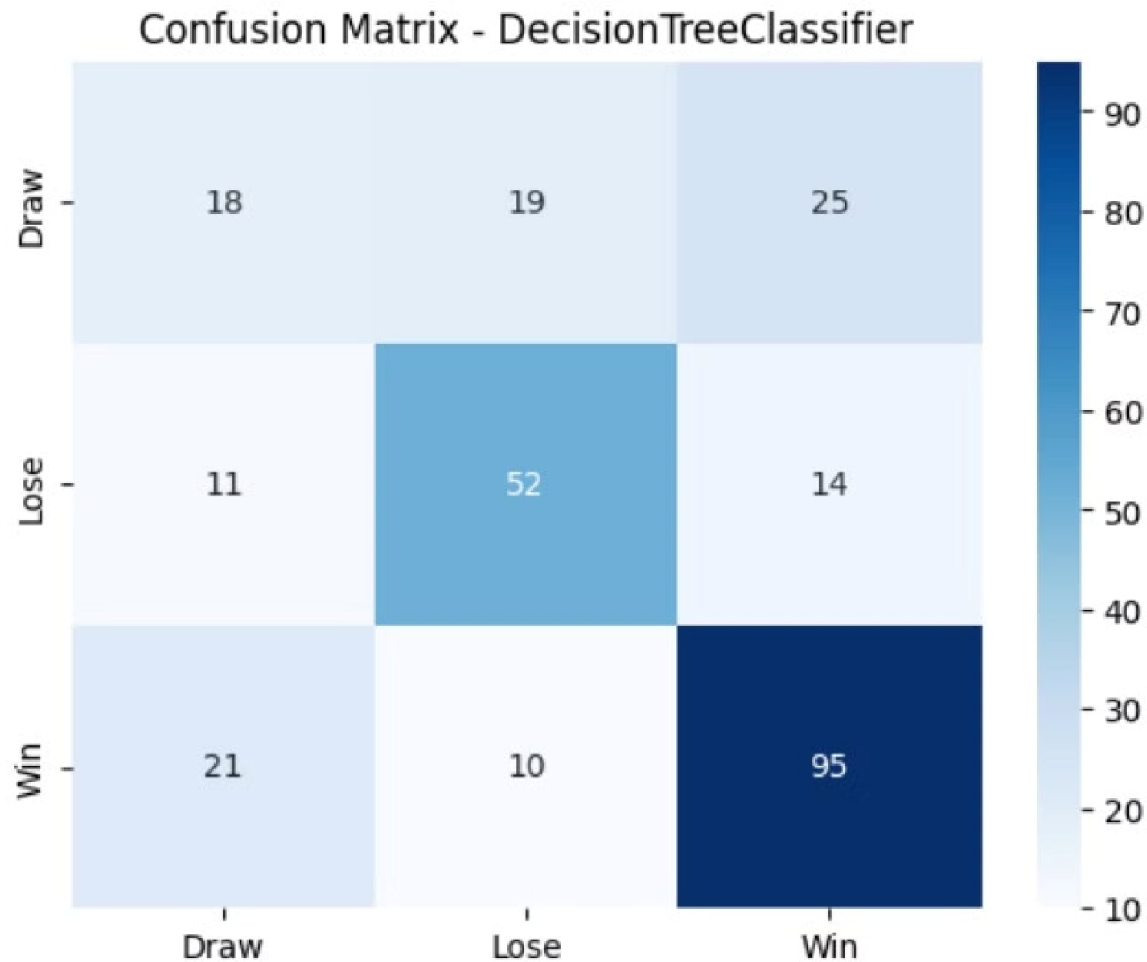
Accuracy: 0.7547169811320755

F1 Score: 0.7466212922486686

Precision: 0.743290033479166

Recall: 0.7547169811320755

Decision Tree Classifier:



Results for DecisionTreeClassifier - Classification:

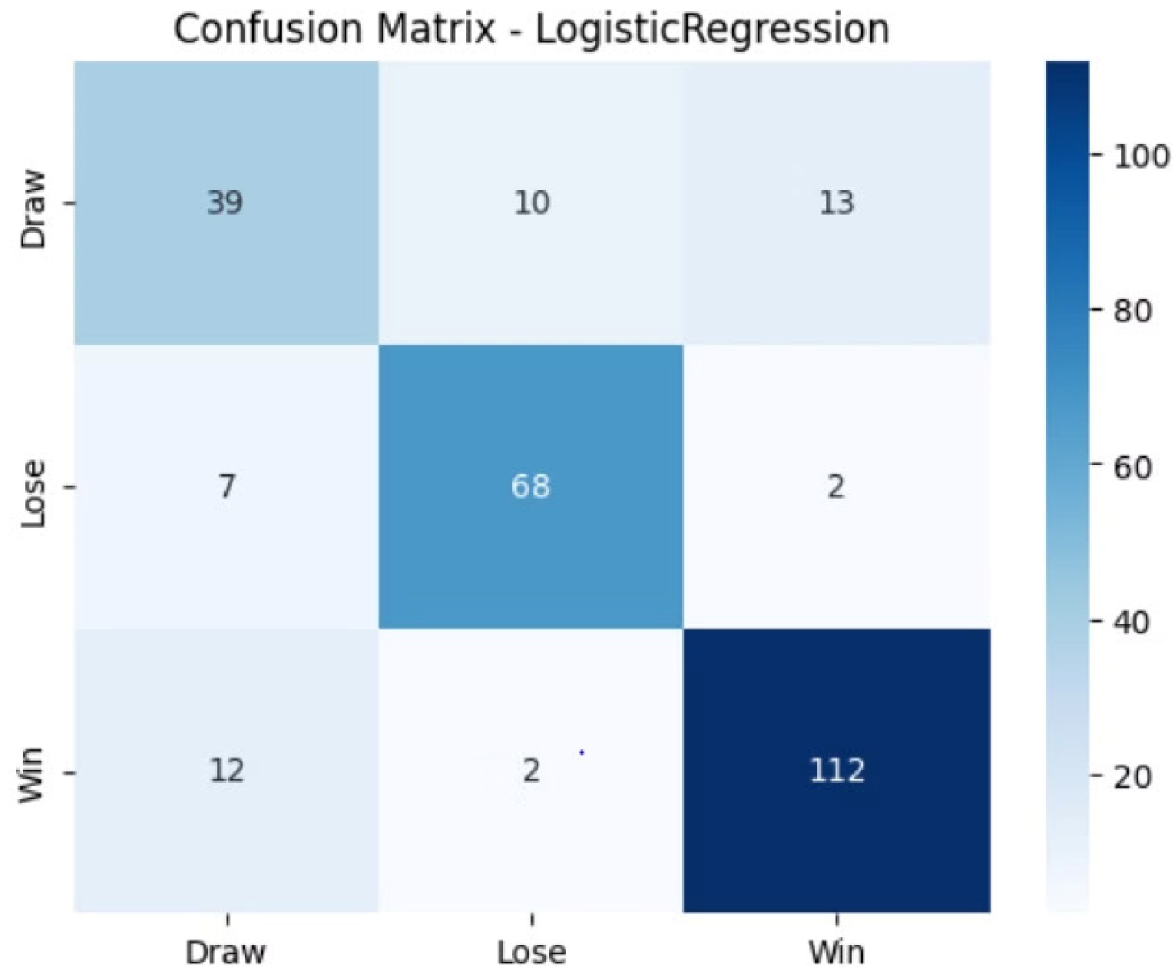
Accuracy: 0.6226415094339622

F1 Score: 0.6139209011671395

Precision: 0.6078507810354239

Recall: 0.6226415094339622

Logistic Regression :



Results for LogisticRegression - Classification:

Accuracy: 0.8264150943396227

F1 Score: 0.8247469780967532

Precision: 0.8236142090891859

Recall: 0.8264150943396227

5. Déploiement

Le modèle a été déployé pour prédire les résultats du match entre **Arsenal** et **Southampton** en utilisant les données des **5 derniers matchs** entre les deux équipes. Les prédictions ont montré que **Arsenal** avait une probabilité de victoire de **44.29%**, mais le match s'est terminé par un **match nul**, soulignant une erreur de prédiction. Le modèle a une **précision de 83%** globalement, mais l'optimisation est nécessaire pour mieux prédire les matchs nuls.



The screenshot shows a web interface titled "Match Predictor" on a blue background. It features two columns for team selection. The left column, labeled "Home Team:", shows the Arsenal logo and a dropdown menu currently set to "Arsenal". The right column, labeled "Away Team:", shows the Southampton logo and a dropdown menu currently set to "Southampton". Between the columns is a "VS" label. Below the dropdowns is a blue button labeled "Predict".

Capture 1 : Les 5 derniers matchs

Montre les **5 derniers matchs** entre **Arsenal** et **Southampton** utilisés pour entraîner le modèle.

```
Arsenal 1 - 1 Southampton
Southampton 1 - 3 Arsenal
Arsenal 3 - 0 Southampton
Southampton 1 - 0 Arsenal
Southampton 1 - 1 Arsenal
```

Capture 2 : Probabilités prédites

Illustre les **probabilités prédites** pour chaque classe de résultat (victoire, défaite, match nul).



This screenshot shows the same "Match Predictor" interface as before, but with a white box at the bottom displaying the "Prediction Result:". The text inside the box reads: "Home Team: Arsenal", "Away Team: Southampton", "Prediction: Home Win", "Home Win Probability: 44.29%", "Away Win Probability: 30.1%", and "Draw Probability: 25.6%".

Capture 3 : Résultat réel du match

Affiche le **résultat réel** du match, qui a été un **match nul**

21.04.23	+	PL		Arsenal	3
				Southampton	3

6.Conclusion

En conclusion, le modèle a montré une bonne précision de **83%** pour prédire les résultats des matchs, bien qu'il ait eu des difficultés à prédire les **matches nuls**. Le déploiement du modèle via une **interface interactive** permet des prédictions en temps réel. Avec des améliorations telles que l'ajout de données contextuelles et l'optimisation des modèles, la précision des prédictions pourrait encore être améliorée.