Phase 2: Machine Learning Models for Football Match Prediction

Introduction

In this phase, we will explore different machine learning models to predict match outcomes in the Somali Football League. The goal is to use various models and compare their performance to select the best one. Below are the models chosen for this project and the reasons for selecting them.

Selected Models

1. Logistic Regression

- **What it does**: Predicts match outcomes (Win, Draw, Loss) using probability.
- **Why use it**: Simple, fast, and easy to interpret.
- **Best for**: Understanding key factors affecting match results.

2. Decision Tree

- **What it does**: Creates a tree-like structure to make decisions based on historical match data.
- **Why use it**: Easy to visualize, handles both numerical and categorical data.
- **Best for**: Explaining predictions in an easy-to-understand way.

3. Random Forest

- **What it does**: Uses multiple decision trees to improve accuracy and reduce overfitting.
- **Why use it**: More reliable than a single decision tree, good for complex patterns.
- **Best for**: Getting balanced accuracy while avoiding overfitting.

4. K-Nearest Neighbors (KNN)

- **What it does**: Compares a match with similar past matches to predict outcomes.
- **Why use it**: Simple, requires no training, and works well with structured data.
- **Best for**: When we have historical data with repeated patterns.

5. Support Vector Machine (SVM)

- **What it does**: Finds the best boundary to separate match results.
- **Why use it**: Works well for small datasets and handles complex relationships.
- **Best for**: Ensuring high accuracy in cases where data is well-structured.

6. XGBoost (Extreme Gradient Boosting)

- **What it does**: Uses multiple decision trees in an optimized way for better accuracy.
- **Why use it**: One of the most effective models for structured data.
- **Best for**: High-performance predictions in structured datasets.

Conclusion

These models provide a variety of approaches to predicting match outcomes. Some models, like Logistic Regression and KNN, are simpler and easier to implement, while others like XGBoost offer higher accuracy. We will train and evaluate all these models to determine the best-performing one for Somali football match prediction.