Football Player Position Prediction using FIFA Data

Project Overview

This project aims to predict the on-field playing position (e.g., CM, RW, CB) of football players based on their in-game stats
from FIFA 21 and FIFA 22 datasets. Instead of training and testing on the same dataset, FIFA 21 is used for training while
FIFA 22 is used for evaluation to simulate real-world generalization.

Pain Points Addressed

1. Inaccurate model generalization when trained/tested on the same dataset.
2. Difficulty predicting position from partial player names (e.g., 'Saka' instead of 'B. Saka').
3. Low accuracy due to inclusion of irrelevant or missing data such as goalkeepers or substitutes.
4. Complex features with NaNs that needed cleansing before model training.

Solution Summary

- Cleaned both FIFA 21 and FIFA 22 datasets by removing rows with missing or irrelevant values.
- Limited the model to outfield players only (excluding goalkeepers, substitutes, reserves).
- Selected numeric performance-related features for model training.
- Used a Random Forest Classifier trained on FIFA 21 to predict positions in FIFA 22.
- Implemented a partial name matching function to allow flexible player lookups.

Technical Approach

- Language: Python

- Libraries: pandas, scikit-learn, numpy, fpdf

- Model: RandomForestClassifier (100 trees, random_state=42)

- Feature Set: 30+ player attributes including passing, shooting, stamina, and vision.

- Evaluation: Accuracy score and classification report from FIFA 22 predictions.