**1. Introduction**

Football is the most popular sport in the worldwide sector and FIFA simulation game is created the by EA sports. The given datasets contain the player attributes for career mode from 2015 to 2020 FIFA game. This report main objectives are clustering the football players based on the skills, focuses on data analysis, model comparison and challenges faced during the analysis.

**2. Data Analysis**

**2.1 Dataset Overview**

The dataset includes attributes such as player name, age, height, overall rating, potential, preferred foot, work rate, position, and various skill ratings. The main aim is to analyse player statistics, distributions analysis, and performance metrics.

**2.2 Top 10 countries with most players:**

To analyse the number of football players per country, and we rank the top 10 countries contributing to professional footballers.

**2.3 Overall rating vs Age of players:**

The distribution analysis of overall rating with different age groups which can help to analyse the performance of the players who is underrated. This can lead the insights into peak performance years.

**2.4 Offensive players vs Salary:**

The comparison between the offensive players such as Right wingers, left wingers and strikers and salary trends to check which type of football player getting paid highest.

**3. Clustering of Football Players**

Using clustering techniques like K-Means or hierarchical clustering, players are grouped based on their attributes such as dribbling, finishing, sprint speed, passing, and physicality etc. This helps in identifying similar player archetypes and evaluating their playing styles.

**4. Model Comparison Report**

**4.1. Models Evaluated**

Using the machine learning model to group the players into clusters. The model which can be include:

* **K-Means Clustering** for skill-based segmentation
* **Elbow technique** to verify the number of clusters for predicting the silhouette score.
* **Silhouette score** to analyse the distance between the group of clusters from -1 to +1.

**5. Challenges Faced**

**5.1. Data Preprocessing Issues**

* Handling missing values in attributes like team position, pace, dribbling, defending, physic, shooting, release clause euro.
* Standardizing units of height, weight, overall, and potential.

**5.2. Feature Selection**

* Avoiding bias from overall ratings while clustering.
* Identifying the most relevant attributes for performance analysis.

**5.3. Model Selection and Training**

* Choosing an optimal number of clusters for K-Means.
* Avoiding overfitting while training predictive models.

**5.4. Interpretation of Results**

* Visualizing key insights through plots and ranking systems.
* Making logical conclusions from statistical trends.
* The silhouette score for 2 clusters in KMeans Algorithm is 0.78.
* The silhouette score for 3 clusters in KMeans Algorithm is 0.63.
* The silhouette score for 4 clusters in KMeans Algorithm is 0.54.
* The silhouette score for 5 clusters in KMeans Algorithm is 0.62.

**6. Conclusion**

Using the FIFA20 player dataset we have clustered the players based on the skills, performance and model comparisons. The insights drawn from this report contribute to better understanding player attributes and trends in football simulations.