FIFA 19 PLAYER RATING AND WAGE ANALYSIS

Group 44: Lakshmi Narayanan Ramasamy, Mahima Regunathan, Srivathsan Gomadamramesh

THE UNIVERSITY OF UTAH

Problem & Motivation

- FIFA 19 is a soccer-based video game which features around 18,000 domestic and international players
- Players depicted in the game possess skillsets that are exactly similar to their real-life counterparts
- The game has an intuitive rating system that weighs each player's skillset according to the position they play
- Our motive is to identify the player attributes that contribute majorly to his rating in the game

Data

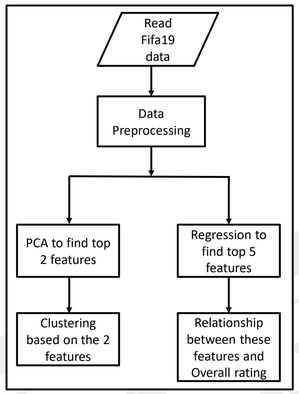
- The dataset chosen for this project is the FIFA 19 complete player dataset from https://www.kaggle.com/karangadiya/fifa19
- The dataset consists of 18,200 rows and 89 columns where each row represents a player and each column represents their attributes/skills

Data Preprocessing

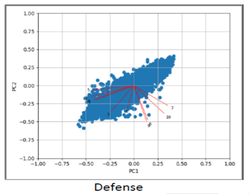
- Columns like Wage, Height and Weight had to be normalized to a scale of 100.
- Missing values filled with mean of the entire column.
- Columns irrelevant to our analysis were removed.
- We mined the dataset with 34 features and 27 positions and found the top 10 relevant features needed for each position.
- We then mapped the 27 positions to 4 broad categories like 'Attack', 'Defense', 'Midfield', 'Goalkeeping'.

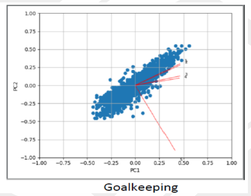
Methods

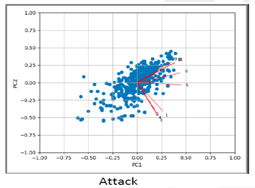
- The key idea behind this project is to find the factors that influence the ratings/wage of a player by applying techniques like **Principal Component Analysis (PCA), K-means ++ clustering** and **Linear Regression**.

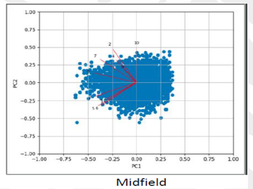


Dimensionality Reduction - PCA





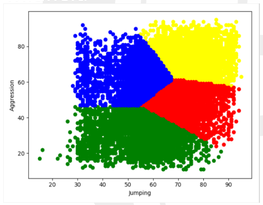




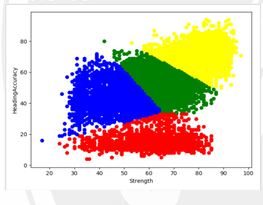
Attack: Dribbling, BallControl
Midfield: Jumping, Aggression
Defense: - Strength, HeadingAccuracy
Goalkeeping: GKReflexes, GKDiving

Clustering – K Means++

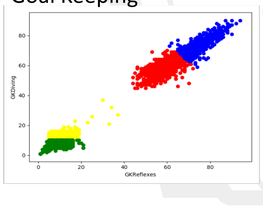
Midfield



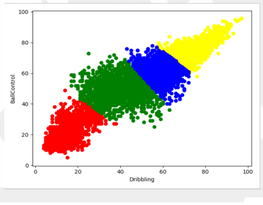
Defense



Goal Keeping

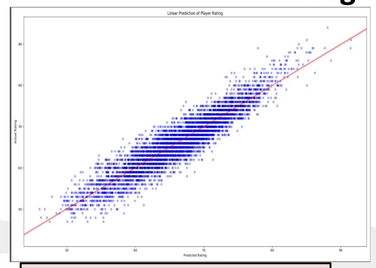


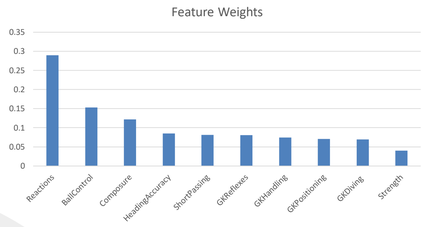
Attack



Attackers: Yellow
Midfielders: Green
Defenders: Red
Goal Keepers: Blue

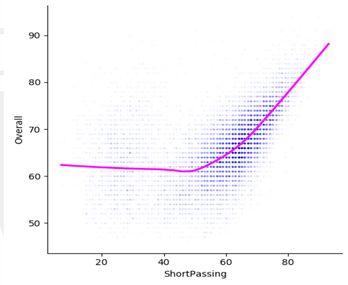
Regression

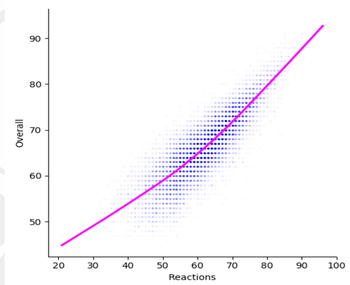


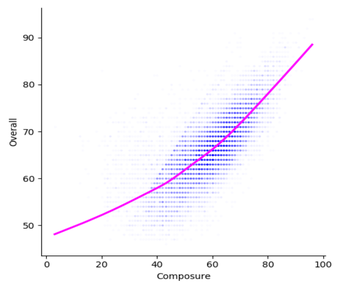


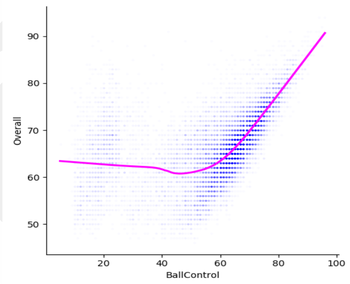
Accuracy of prediction: 86%

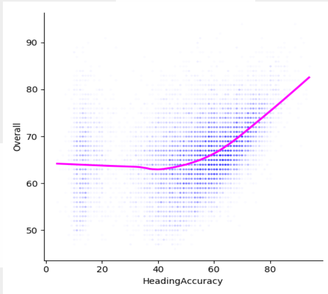
Results











- As the values of the features go above 60, the overall rating also increases linearly.
- This results in a higher rating for the player with higher values for these features.
- Thus, we can see that these features influence the rating the most.