MATCHUP MAYHEM: DATA MINING FOR FANTASY CRICKET RIVALRY (IPL)

Presenter Name
Het Vashi
Laxmikant Kabra
Vaibhav Lodhiya

AGENDA

Introduction: Fantasy Cricket

About Dataset

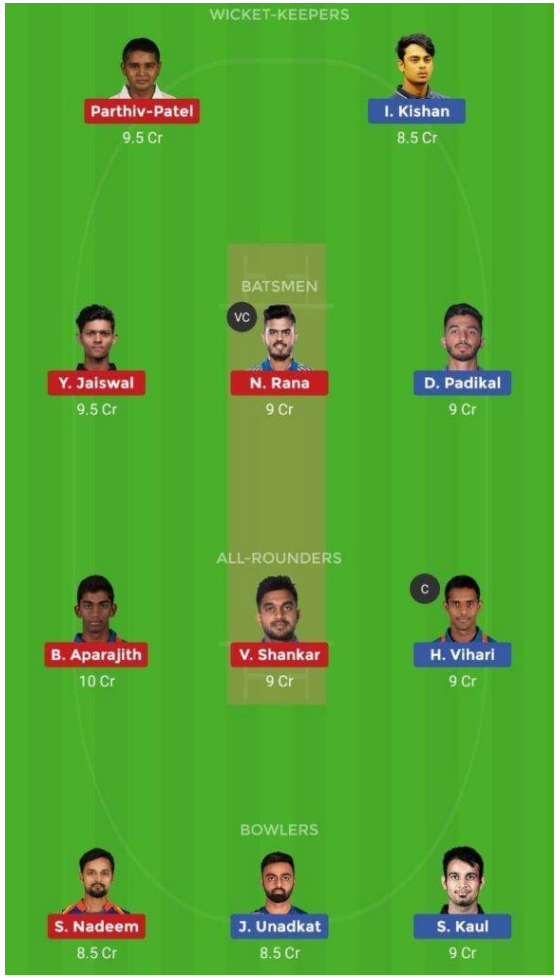
EDA for Dataset

Data Preprocessing

Methodology

Introduction

Fantasy cricket is a destination where participants create virtual teams of real cricket players and earn points based on the players' actual performance in live matches. Participants act as team managers, strategically selecting players to compete against others in leagues and contests. Indian Premier League(IPL) is a festival in itself and since recent introduction, fantasy cricket has become an innate component.



Motivation

Expanding participation in fantasy cricket poses a challenge to select teams on mere intuition.

In age of data, making informed decision gives an edge for winning.

This project aims to bridge this gap of intuition and information through help of data mining techniques.

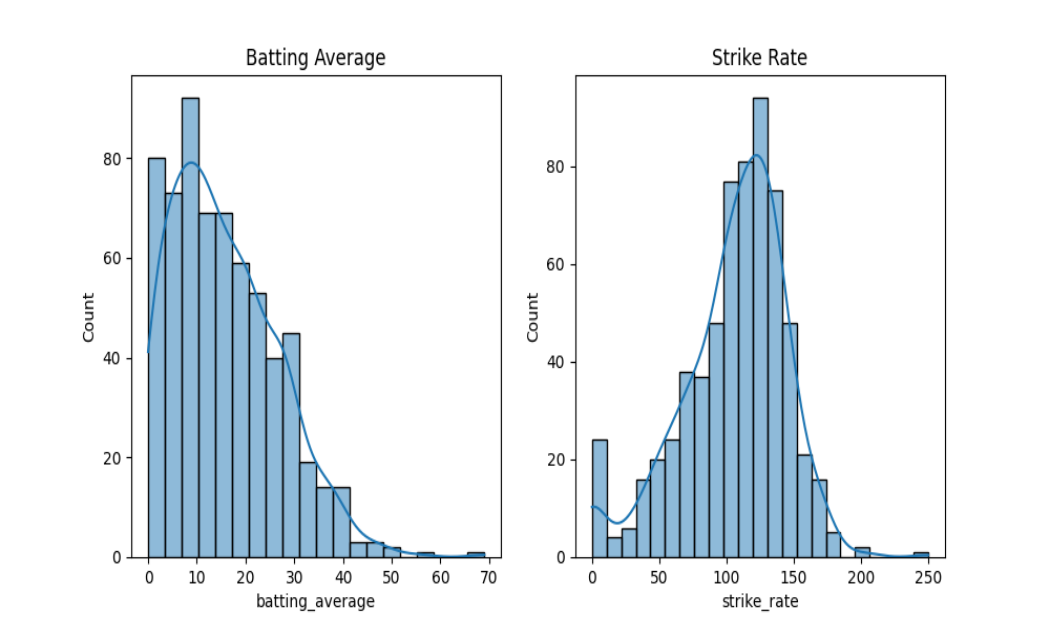
About Dataset

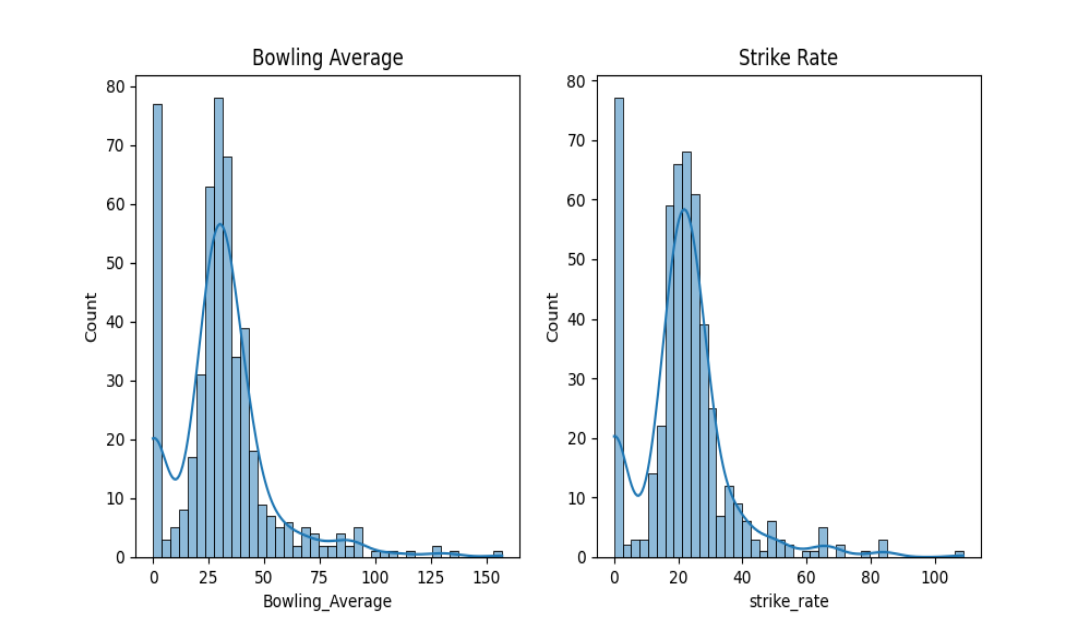
BALL BY BALL DATA

BATTING STATISTICS DATA

BOWLING STATISTICS DATA

EDA FOR DATASET





Data Preprocessing

- ball-by-ball data for each match of every season is used to create a custom dataset for each match.
- Batsman and bowler performances are separated into two data with set of different features.
- We perform standard data cleaning and processing are along with MinMaxScaler to scale the data.
- These set of features are used to create a customized score metric that quantifies player performance across different features into one.
- This weighted score will serve as label to the model.

Linear Regression

2 models

One to predict best set of batsmen

One to predict best set of bowlers

Openers and wicketkeepers included

Implementing a Linear Regression, we trained a model on that predicts player performance metric.

For predicting performance metric in upcoming match, we provide mean player performance in past n years.

The players with highest predicted score will be chosen in squad.

Fine tune weights used in metric calculation to get better results

Selection of best 11 Using Clustering and Weighted Scores

In this approach we are using batting_stats and bowling_stats dataset which gives information about each player from 2008 to 2022.

Applied KMeans clustering to both batsmen and bowlers separately by utilizing features like Batting Average, Strike Rate, Bowling Average, and Economy Rate etc. and identified distinct clusters representing different performance levels.

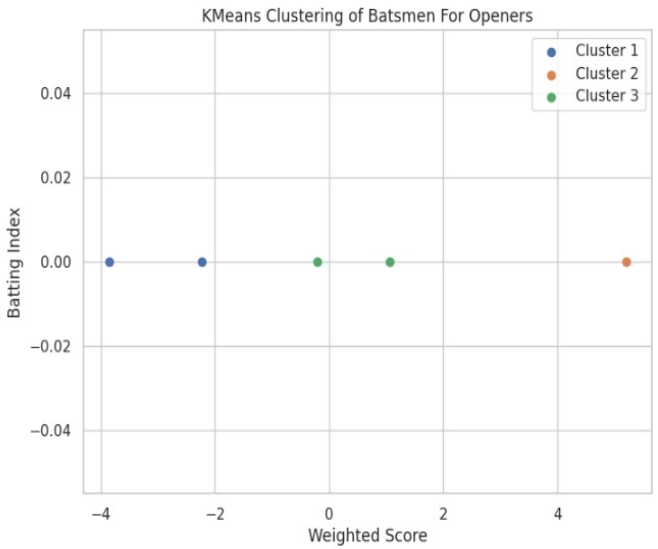
Introduced a Weighted Score system to assign significance to various performance metrics whereWeights were strategically assigned to features such as Batting Average, Strike Rate, Bowling Average, and Economy Rate .This comprehensive approach provides a nuanced evaluation of Cluster(player performance).

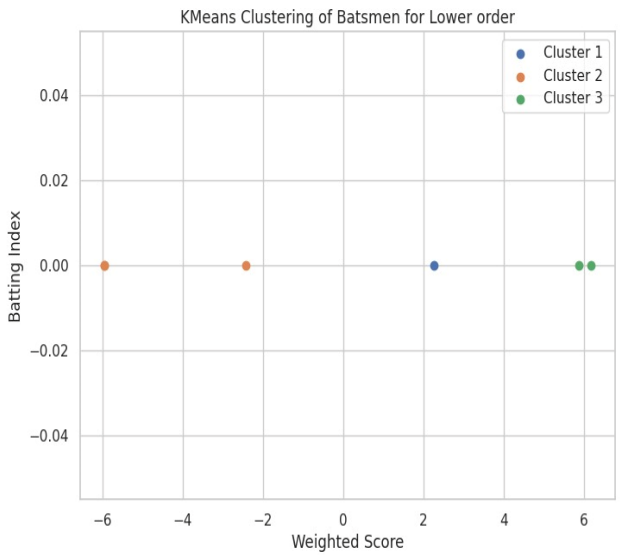
Contd...

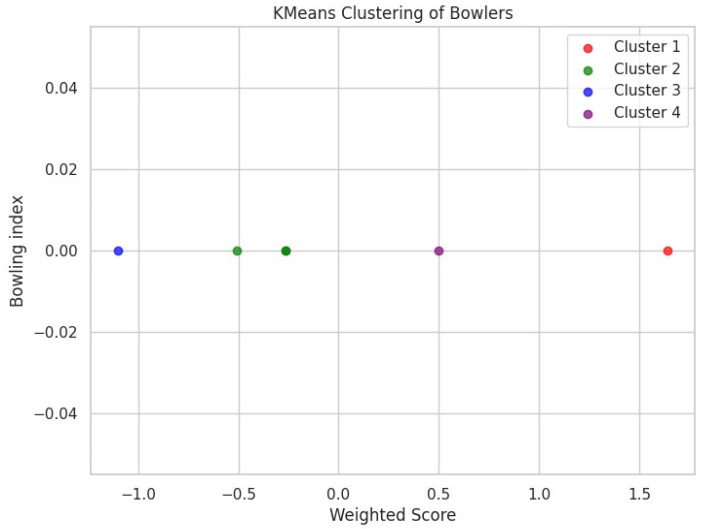
Combining Weighted Scores with KMeans not only refines our player evaluation process but also facilitates the creation of clusters that capture the nuanced strengths and weaknesses of each player. This dual approach enhances the precision of our dream team selection.

In a nutshell, using clustering techniques helps us pick the perfect eleven for an optimal fantasy gaming thrill!

Cluster


KMeans Clustering of Batsmen For Openers
Cluster 1
Cluster 2
Cluster 3
Batting Index
0.04
0.02
0.00
−0.02
−0.04
−4
−2
0
2
4
Weighted Score


KMeans Clustering of Batsmen for Lower order
Cluster 1
Cluster 2
Cluster 3
Batting Index
0.04
0.02
0.00
−0.02
−0.04
−6
−4
−2
0
2
4
6
Weighted Score


KMeans Clustering of Bowlers
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Bowling index
0.04
0.02
0.00
−0.02
−0.04
−1.0
−0.5
0.0
0.5
1.0
1.5
Weighted Score

MATCHUP MAYHEM

THANK YOU

Het Vashi
Laxmikant Kabra
Vaibhav Lodhiya