**Project Title:** Beyond Cricket

**Main Problem:**

To analyze Batsman performance and create predictive models using historical data to aid in decision-making for fantasy cricket enthusiasts.

**Objective:**

**Data and Initial Findings:**

For this Analysis, data is collected from the cricinfo website, which provided the variables impacting the performance of the player. Some of them are Innings played, Runs scored, Balls Faced and 6s and 4s hit. Preliminary assessments indicate a strong positive relationship between the balls faced and runs scored per game.

**Approach:**

In our endeavour to enhance decision-making in fantasy cricket, we harnessed machine learning methodologies, employing Linear Regression to discern the optimal picks based on opposition and ground dynamics. Additionally, leveraging the K-means Algorithm, we clustered batsmen data into four distinct groups, a determination made through the insightful Elbow Method. This clustering aided in identifying nuanced player categories, enhancing our ability to recommend diverse strategies.

Furthermore, our exploration delved into the intricate relationships between key performance indicators like batsmen runs and innings played. This comprehensive analysis involved constructing a predictive model trained on fundamental attributes such as innings, run-scored, balls-faced, opposition, and ground conditions. Rigorous validation processes and meticulous adjustments fortified the reliability of our forecasts, aligning with our primary objective of delivering dependable predictions.

Ensuring transparency and trust in our analyses,. Our paramount aim remains the generation of robust and reliable predictions, empowering fantasy cricket enthusiasts with actionable insights for strategic decision-making.

This multifaceted approach, augmented by advanced methodologies and rigorous validation, underpins our commitment to delivering trustworthy and insightful predictions, fostering an environment of informed choices in the dynamic realm of fantasy cricket.

**Recommendations:**

According to the analysis, the Linear Regression model has the maximum accuracy, precision, and recall for both classes. Although logistic regression is a descent method, it does not achieve the same level of accuracy as linear regression. K-Means clustering efficiently classifies players, assisting users in forecasting the optimal player for various scenarios. It is advisable to continue conducting research and optimizing the model to further improve its functioning and assure accurate predictions.

**Conclusions:**

With all the analysis, we found that Linear Regression is the best model with the best accuracy, good precision and recall for both classes. In addition to that, logistic regression has been descent but didn’t match the accuracy of Linear Regression. On the other hand,K-Means clustering has successfully divided the data into clusters of Different categories like Opener, Middle Order, All-rounder and finisher. This helps the User in predicting the best player when predicting. It can be concluded that Linear Regression model is the best option among all Supervised models, however additional research and optimization are suggested to improve the model’s functionality and guarantee reliable forecasts.

Team Members :

* Chilappagari Anantha Krishna
* ShivaKumar Dasari
* Sai Swetha Kancharlapalli