July 2020

ALL ABOUT CRICKET

By

Vidya Menon

METHODOLOGY

- **ROSEMED Methodology** - This is the one of the most straightforward of the Data Science processes.
- **Research** - Find out about various models and see which model works best for our data.
- **Obtain** - Understanding stakeholder requirements, gathering information on the problem, and finally, sourcing data that we think will be necessary for solving this problem.
- **Scrub** - Focus on preprocessing our data. Important steps such as identifying and removing null values, dealing with outliers, normalizing data, and feature engineering/feature selection are handled around this stage.
- **Explore** - We create visualizations to really get a feel of the dataset. We focus on things such as understanding the distribution of different columns, checking for multicollinearity, and other tasks like that.
- **Model** - It consists of building and tuning models using all the tools we have in our data science toolbox.
- **Evaluate** - Interpret the results of models and communicate results to stakeholders.
- **Deploy** - Deploying the Model.

CHALLENGE

To create a Dashboard that helps us to understand :

- Which Teams have had more Wins/Losses
- Strike Rate v Runs of Batsmen
- Wickets v Economy for Bowlers
- Which players have been doing well in the recent time looking at the Trend Analysis Report.

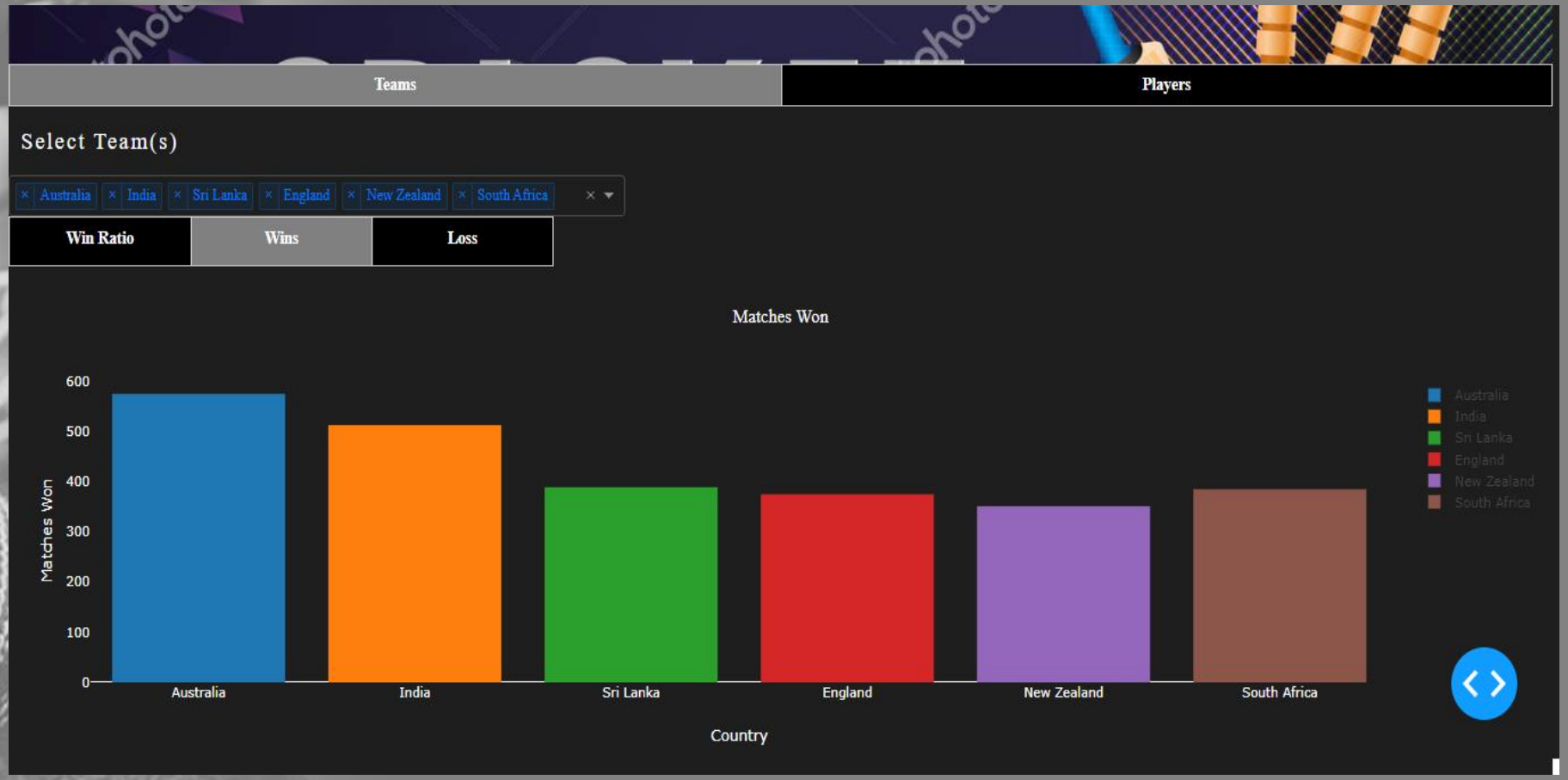
Also, with the help of Machine Learning:

- Predict the winner of World Cup.

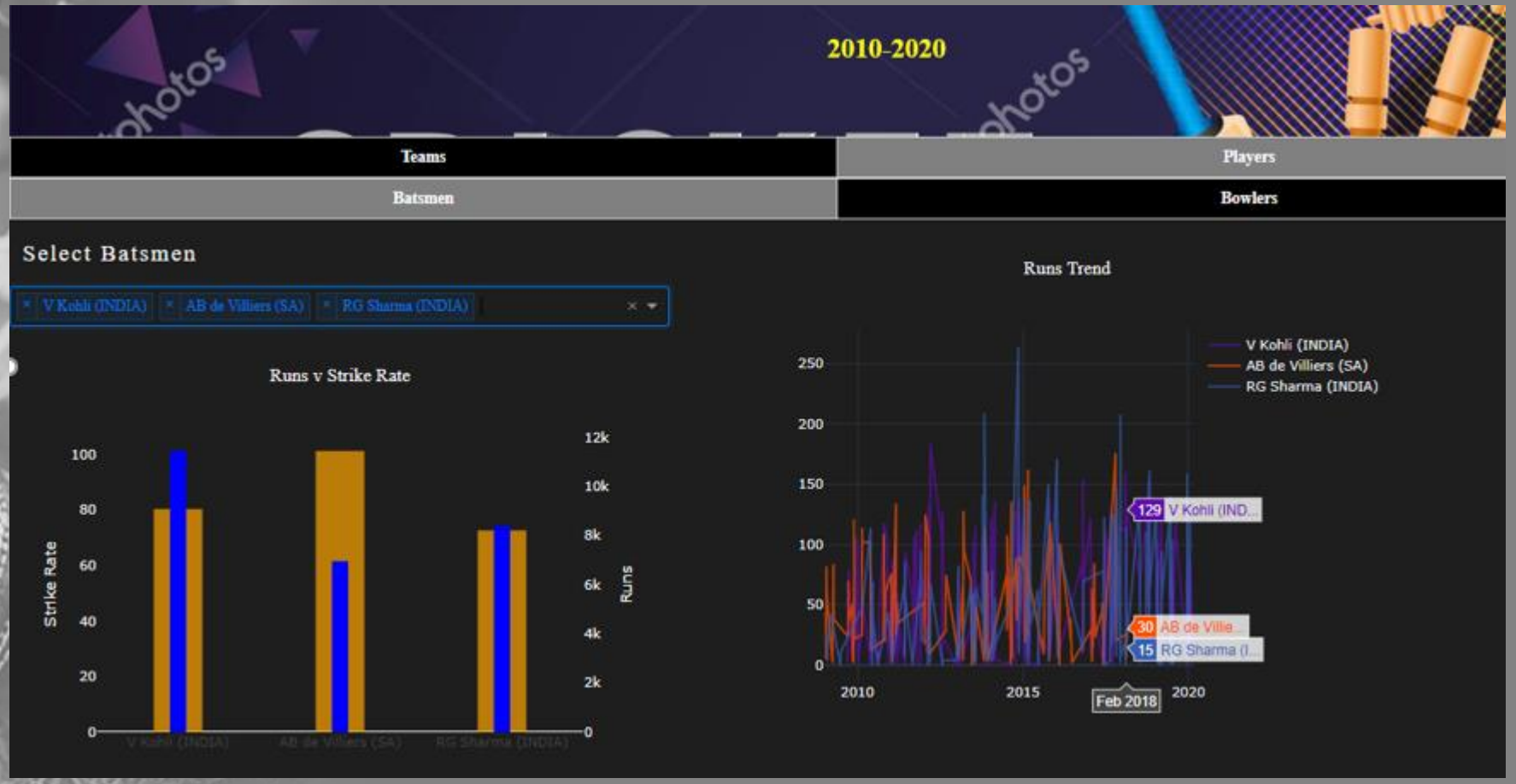
DASHBOARD LAYOUT

An interactive analytics dashboard which contains data from 2010 - 2020. One can analyze runs scored, strike rates by player, and other insights.

TEAM STATS


Teams
Players
Select Team(s)
Australia
India
Sri Lanka
England
New Zealand
South Africa
Win Ratio
Wins
Loss
Matches Won
600
500
400
300
200
100
0
Matches Won
Australia
India
Sri Lanka
England
New Zealand
South Africa
Country

BATSMAN STATS



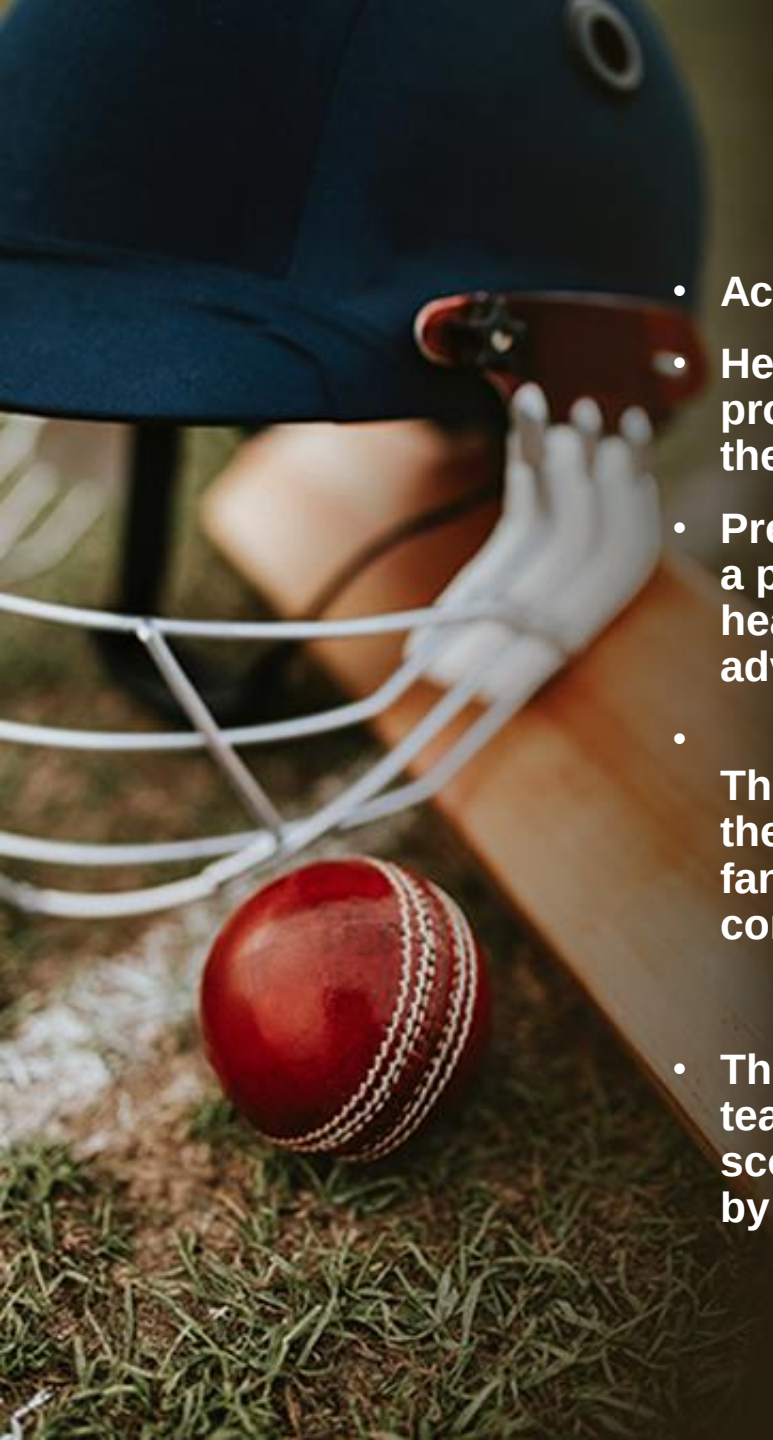
BOWLER STATS

WORLD CUP Prediction

- Using Random Forest-
  Training set accuracy : 68%
  Test set accuracy : 65%

- Top 4 teams :
  England, India, New Zealand, South Africa

- Winner : England ☺



CONCLUSION

- According to our model, England is likely to win the World Cup.
- Here, we can see how Machine Learning could be used to calculate the winning probabilities of a team based on the current rankings and the results of the matches they have played.
- Predicting the future sounds like magic whether it be detecting in advance the intent of a potential customer to purchase a product or figuring out where the price of a stock is headed. If we can reliably predict the future of something, then we own a massive advantage. Machine learning has only served to amplify this magic and mystery.
- The main objective of Sports prediction is to improve team performance and enhance the chances of winning the game. The value of a win takes on different forms like the fans filling the stadium seats, television contracts, fan store merchandise, parking, concessions, sponsorships, enrollment and retention.
- The Dashboard that has been created gives us an insight into the performances of the team and the players. For batsmen, it helps us to analyze their strike rate and the runs scored. Also, we can see their performance as a trend. The intent is similar for bowlers by checking their Economy and the wickets taken.

RECOMMENDATIONS/ FUTURE WORK

- We can use Machine Learning to predict a team's performance.
- The same approach can be used in predicting the outcomes in various fields.
- A Dashboard can be used to keep a track of various Projects in organizations. They can also be used to see the progress of employees, the kind of blockers they are facing. These dashboards are flexible and can be adapted to any environment based on the kind of data we have.
- Future work will be to include the data set of matches held before the year 2000 to see if our scores improve. Also, would like to include the current form of players and consider the grounds that the teams play on along with the team who wins the toss.

Any Questions ☺ ???