Cricket Analytics
ML.Net

Praveen Raghuvanshi

@praveenraghuvan

AGENDA

01 Cricket

02 Dataset

03 Data Cleaning and Analysis

04 Analysis using ML.Net

05 Prediction using ML.net

06 Demo

INTRODUCTION!

- Cloud Architect @ HARMAN A SAMSUNG COMPANY
- Domain: Professional Audio, Video & Control
- Area of Expertise: Cloud, Distributed computing
- Area of Interest: AI/ML, Cloud and IoT
- Location: Bangalore, India
- Azure certified
- Member .NET foundation

History

- Invented in 1550, originated by England
- First International Match was played between Canada and USA in New York
- First ODI was played in 1971
- First Cricket World Cup in 1975
- First T20 Match was played in 2003

Formats

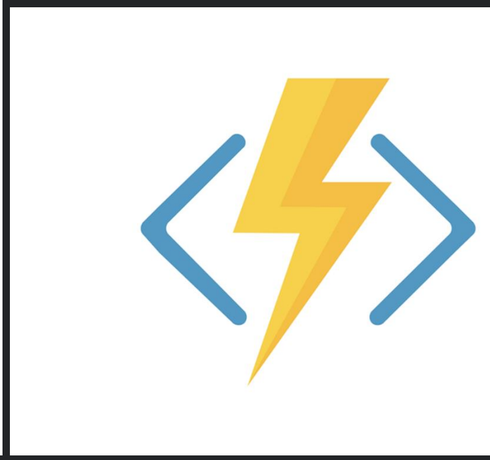
- ODI – One Day International
- Test Match
- T20
- County
- IPL

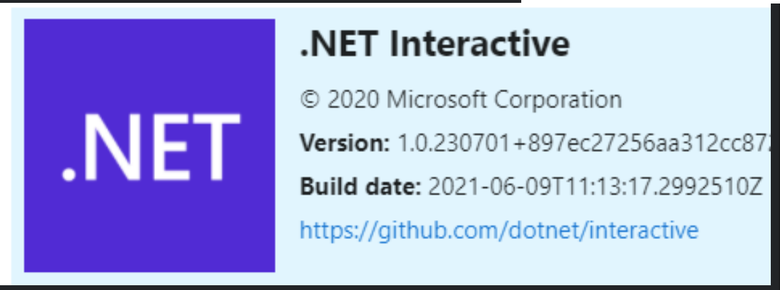
Statistics(2019-2020)

- Played by 104 Nations
- IPL Valuation: $6.7 billion
- IPL Viewers : 370 Million

PROBLEM STATEMENT

- Perform Analysis on cricket dataset
- Predict the team score till 6 overs





TOOLS AND FRAMEWORKS

Dataset

Source: Cricksheet.org
T20 Matches-Men

- Duration: 2017 – 2021
- Matches : 1010
- Teams: 56
- Columns: 22
- Records: 231 K
- Mix of number and strings

- match_id
- season
- start_date
- venue
- innings
- ball
- batting_team
- bowling_team
- striker
- non_striker
- bowler
- runs_off_bat
- extras
- wides
- noballs
- byes
- legbyes
- penalty
- wicket_type
- player_dismissed
- other_wicket_type
- other_player_dismissed

Data Cleaning

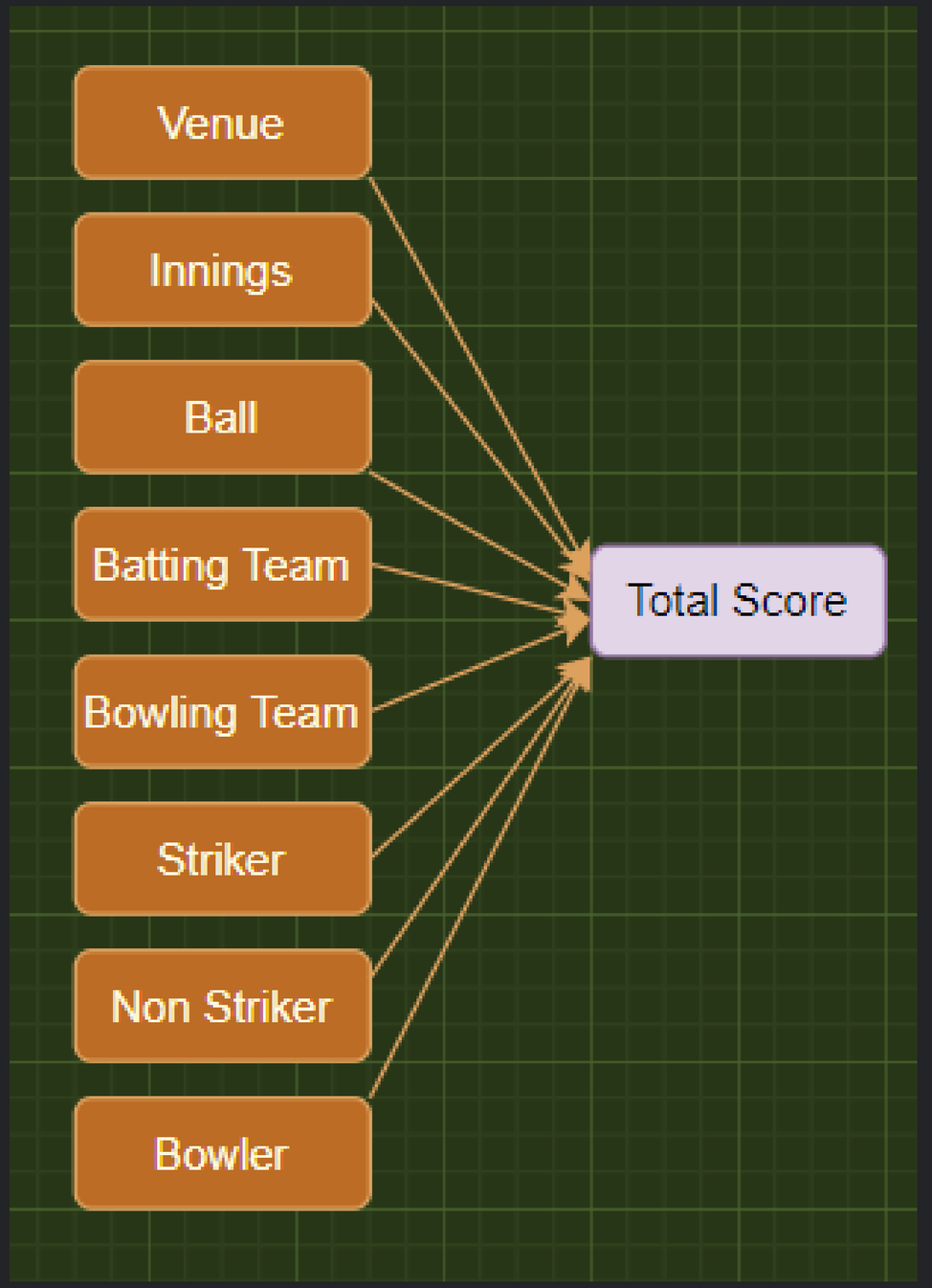
- Filter dataset to include records till 6 over. Low memory and fast execution
- Check for Null Values
- Aggregation : Score per ball → runs_off_bat + extras
- Cumulative sum : Total Score per ball
- Remove features/columns

- match_id
- season
- start_date
- venue
- innings
- ball
- batting_team
- bowling_team
- striker
- non_striker
- bowler
- runs_off_bat
- extras
- wides
- noballs
- byes
- legbyes
- penalty
- wicket_type
- player_dismissed
- other_wicket_type
- other_player_dismissed

Prediction

Regression Problem using ML.Net

- Define Classes : Match, MatchScorePrediction
- Load Dataset
- Split Dataset: Train/Test : 80/20
- One Hot Encoding
- Model Algorithm : FastTree
- Train → Evaluate → Predict the Model



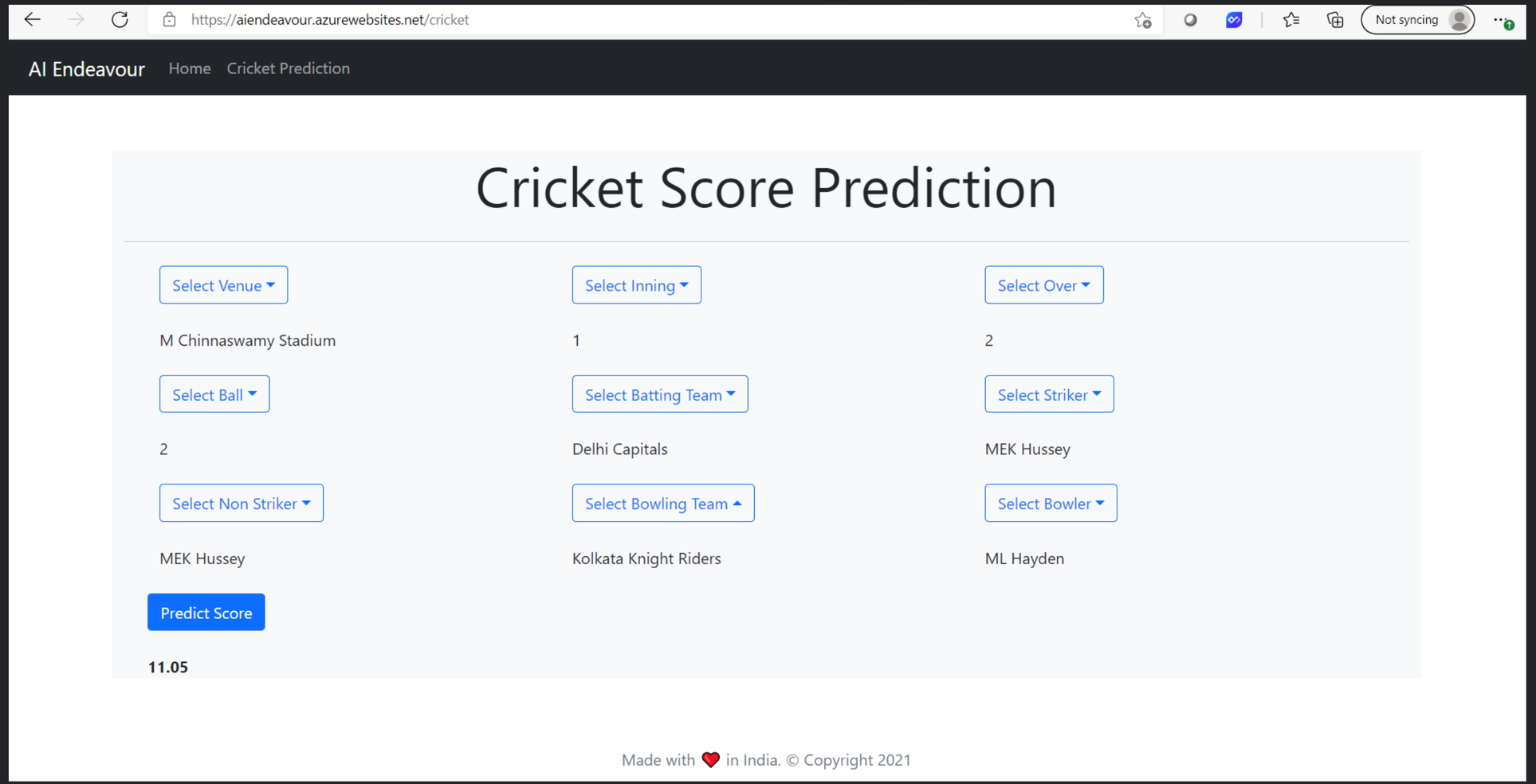
ML.Net – API(Jupyter Notebook)

```
************************************************
*       Model quality metrics evaluation
*------------------------------------------------
*       RSquared Score:      0.82
*       Root Mean Squared Error:      6.73



********** Predict...
Match Info:

Venue: Vidarbha Cricket Association Stadium_ Jamtha
Batting Team: India
Bowling Team: New Zealand
Inning: 1
Ball: 3.4
Striker: V Kohli
Non-Striker: Yuvraj Singh
Bowler: CJ Anderson
^^^^^^ Prediction:  24.15512
************************************************************************
Predicted score: 24.1551, actual score: 20
************************************************************************
```

ML.Net – Model Builder

Top 3 models explored						
	Trainer	RSquared	Absolute-loss	Squared-loss	RMS-loss	Duration	#Iteration
1	LightGbmRegression	0.9012	3.72	25.48	5.05	6.7	1
2	SdcaRegression	0.8234	5.09	45.56	6.75	4.0	2
3	FastTreeRegression	0.8213	5.08	46.09	6.79	4.7	3

ML.Net - WebApp



06 DEMO

IMPROVEMENTS

- Large Dataset
- Feature selection strategies
- Algorithm

RESOURCES

- Getting Started: https://dotnet.microsoft.com/apps/machinelearning-ai/ml-dotnet
- Slides and Source: https://github.com/praveenraghuvanshi/tech-sessions/tree/master/16062021-Global-AI-Community-2021

THANK YOU FOR WATCHING!

ANY QUESTIONS?

https://linktr.ee/praveenraghuvanshi

@praveenraghuvan
Github: praveenraghuvanshi