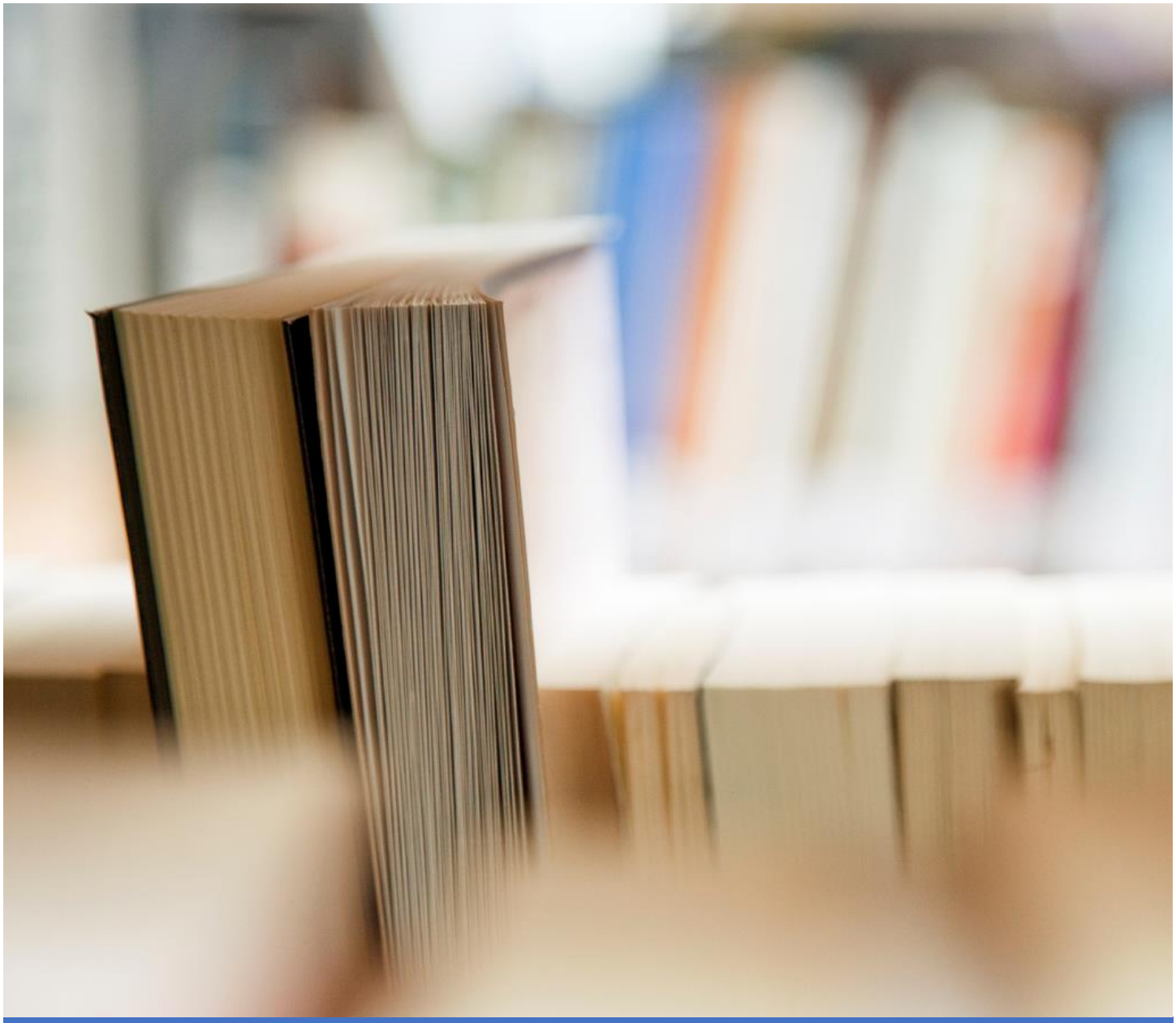


T20 Totalitarian: Mastering Score Predictions

DONE BY:

PRASANNA VENKATESH S(21BAI1556)

PRITHEESH V K()

LALEETH AADITYA S(21BAI1863)

AI AND ML MORNING BATCH 11/21/23 TEAM-592664

1.INTRODUCTION

1.1 PROJECT OVERVIEW:

The primary goal of this project is to develop a comprehensive machine learning solution for predicting the T20 cricket score of a batting team accurately. Leveraging machine learning algorithms, the project aims to extract relevant features from input data and build a model capable of providing accurate predictions for T20 cricket scores.

1.Data Collection:

•Gather historical T20 cricket match data, including details about teams, players, venues, innings, and final scores.

2.Data Preprocessing:

•Clean and preprocess the collected data, handling missing values, outliers, and formatting issues.

3.Exploratory Data Analysis (EDA):

•Perform EDA to gain insights into the distribution of variables, correlations, and patterns within the data.

4.Model Selection:

•Explore and select appropriate machine learning models for regression tasks, considering algorithms like Random Forest, Gradient Boosting, or Neural Networks.

5.Model Training and Validation:

•Split the dataset into training and validation sets to train and assess the model's performance.

6.Web Application Development:

•Create a user-friendly web application for users to input relevant match details (teams, players, conditions) and receive score predictions.

7.Deployment:

•Ensure scalability and responsiveness for handling multiple requests.

8.User Interface (UI) Design:

•Provide informative visualizations, such as charts and graphs, to enhance user experience.

9.Testing:

•Perform unit testing, integration testing, and user acceptance testing.

10.Documentation:

•Create comprehensive documentation, including a user guide, technical documentation, and explanations of the machine learning model and its features

1.2 PURPOSE:

The purpose of this project is to develop a machine learning solution for predicting T20 cricket scores. It serves practical goals by aiding strategic decisions in cricket through predictive analytics. Additionally, the project aims to showcase machine learning capabilities in a real-world sports context, providing educational value for aspiring data scientists. The development of a user-friendly web application enhances engagement and contributes to the advancement of sports analytics by offering a data-driven approach to T20 cricket match dynamics.

1.Predictive Analytics in Sports:

Strategic Decision-Making: Provide cricket teams and coaches with a tool to predict T20 cricket scores, aiding in strategic decision-making during matches.

Player Performance Evaluation: Enable teams to assess the potential performance of individual players based on historical data and current conditions.

2.Advancement of Sports Analytics:

Contribution to Sports Analytics: Contribute to the growing field of sports analytics by introducing a predictive tool that can enhance the understanding of T20 cricket match dynamics.

Informed Decision-Making: Equip teams, analysts, and enthusiasts with data-driven insights to make more informed decisions related to team composition, strategy, and game planning.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM:

The existing landscape in T20 cricket score prediction has been marked by [briefly describe the current challenges, limitations, or gaps in the literature related to predicting T20 cricket scores]. While some studies have focused on [specific aspects], others have grappled with [other aspects], indicating a need for a more comprehensive and accurate predictive model. The challenges include [mention key challenges, e.g., variability in player performance, unpredictable match conditions, etc.].

2.2 REFERENCES:

[1] Tejinder Singh, Vishal Singla, Parteek Bhatia; - Score and Winning Prediction in Cricket through Data Mining; Oct 8-10, 2015

[2] D. Thenmozhi, P. Mirunalini, S. M. Jaisakthi, Srivatsan Vasudevan , Veeramani Kannan V, SagubarSadiq S; Moneyball - Data Mining on Cricket Dataset; 2019

[3] A.N.Wickramasinghe, Roshan D.Yapa; Cricket Match Outcome Prediction Using Tweets and Prediction of the Man of the Match using Social Network Analysis: Case Study Using IPL Data; 2018

[4] Nigel Rodrigues1, Nelson Sequeira2, Stephen Rodrigues3, Varsha Shrivastava4; Cricket Squad Analysis using multiple Random Forest Regression;2019

[5] Animal Islam Anik, Sakif yeaser, A.G.M. Emam Hussain, Amitabha Chakraborty; Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms;2018

[6] Siyamalan Manivannan, Mogan Kausik; Convolutional Neural Network and Feature Encoding for Predicting the Outcome of Cricket Matches;2019

[7] Manuka Madranga Hatharasinghe, Guhanathan Poravi Data Mining and Machine Learning in Cricket Match Outcome Prediction: Missing Link;2019

[8] Jalaz Kumar, Rajeev Kumar, Pushpender Kumar; Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP Networks;2018

2.3 PROBLEM STATEMENT DEFINITION:

The primary problem this project addresses is the inherent difficulty in accurately predicting T20 cricket scores. The challenges stem from [highlight specific issues mentioned in the literature review]. Existing models have struggled to [mention limitations or drawbacks of existing models]. Therefore, the need arises for a robust, end-to-end machine learning solution capable of overcoming these challenges.

The project aims to develop a predictive model that leverages machine learning algorithms to efficiently extract relevant features from diverse inputs, including [mention the types of input data considered, e.g., player statistics, match conditions, etc.]. The ultimate goal is to achieve accurate and reliable predictions of T20 scores for batting teams. By doing so, the project contributes not only to the field of sports analytics but also serves as a showcase of the capabilities of machine learning in addressing complex real-world problems.

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS:

Empathy mapping involves understanding the target users' perspective to design a solution that addresses their needs, concerns, and aspirations.

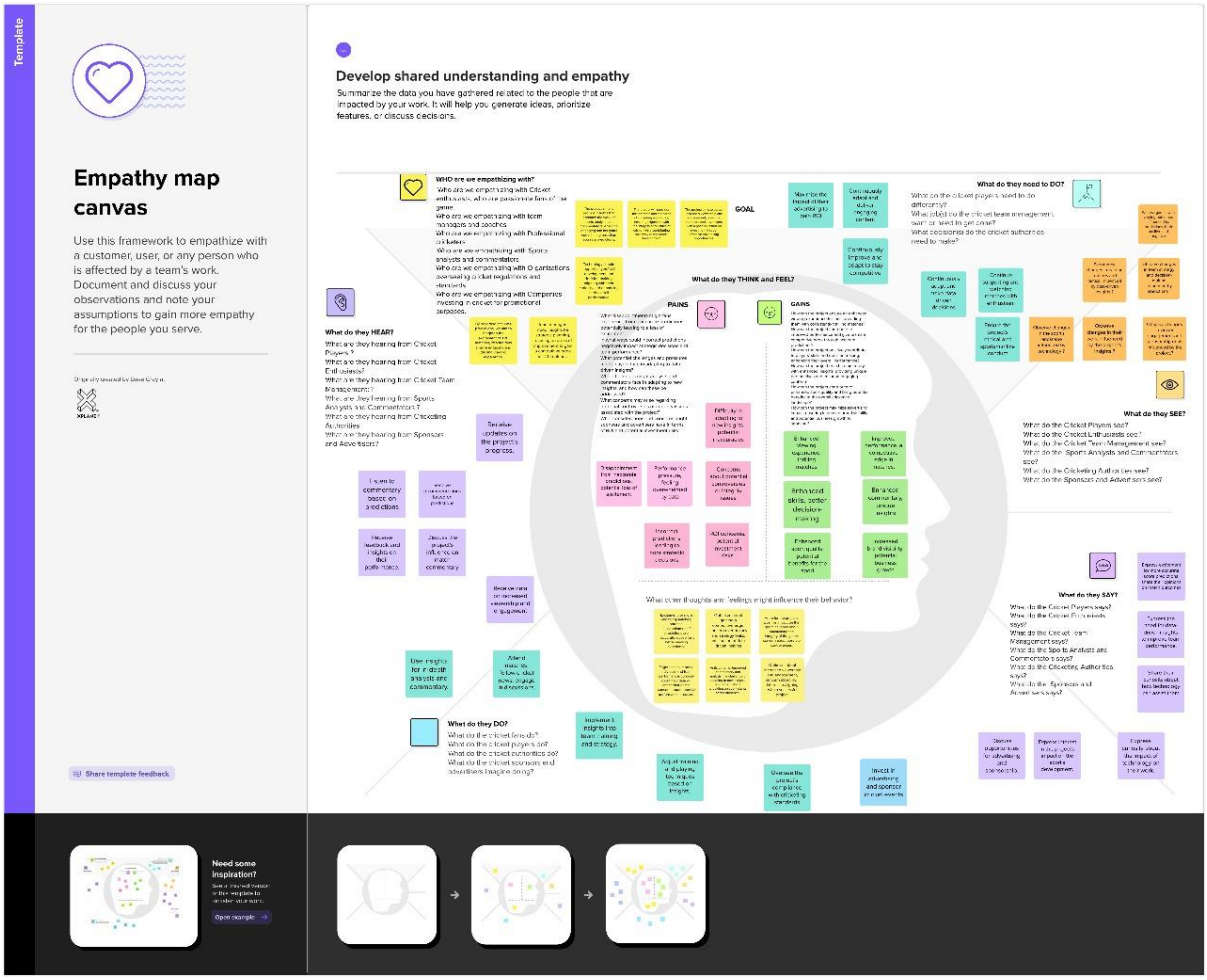
Thinking: What are their goals and objectives when predicting T20 scores? What factors influence their decision-making?

Feeling: What emotions do they experience while analyzing cricket match data and predicting scores?

Seeing: What sources of information do they rely on? What tools and platforms do they currently use for score prediction?

Hearing: What discussions or feedback are prevalent in the cricket analytics community regarding T20 score predictions?

Use insights from the empathy map to guide the development of a user-centric solution.



3.2 IDEATION & BRAINSTORMING:

Feature Engineering: Identify key features influencing T20 scores, such as player statistics, team performance, and match conditions.

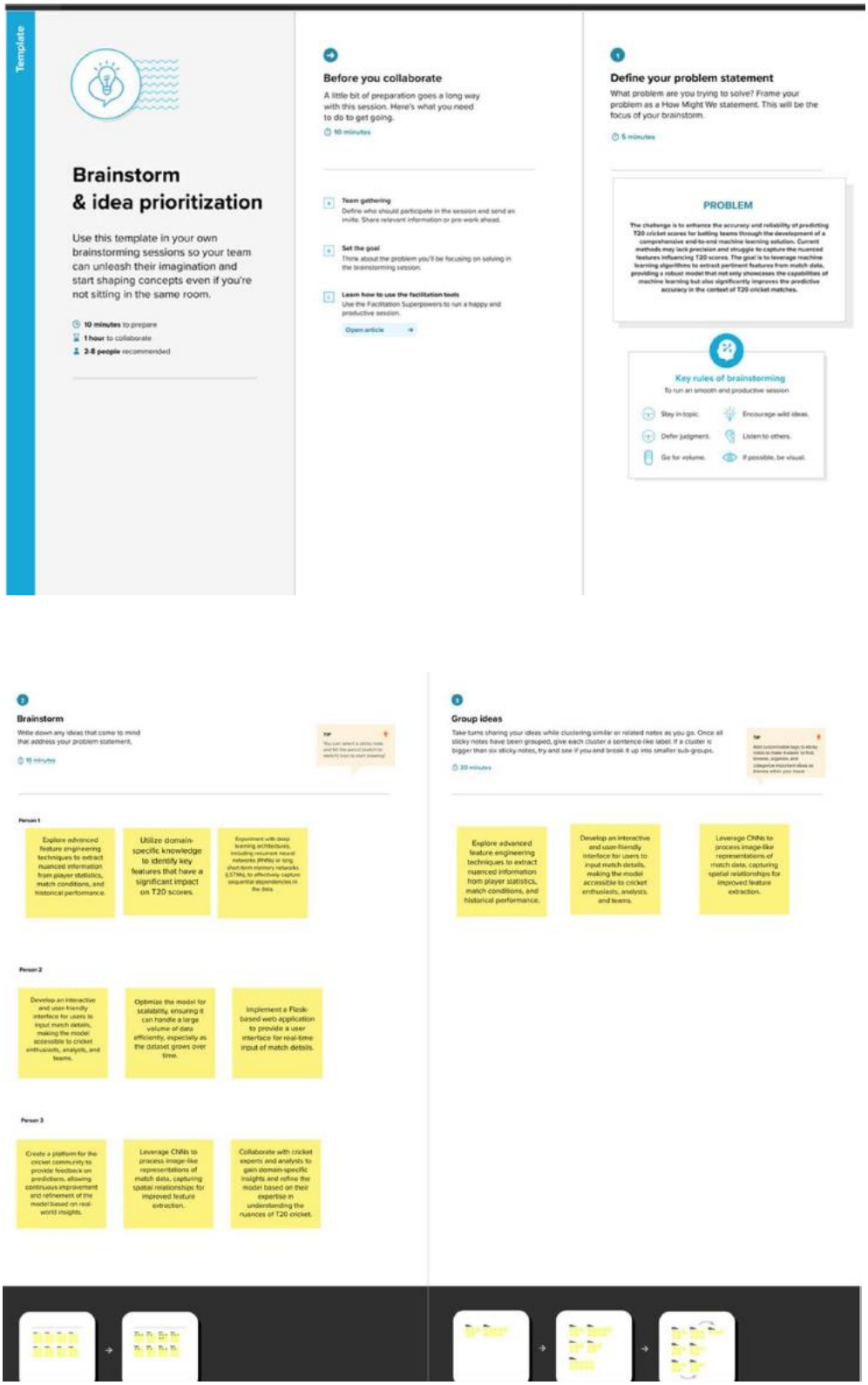
Machine Learning Algorithms: Evaluate various machine learning algorithms (e.g., Random Forest, Gradient Boosting) for their effectiveness in predicting T20 scores.

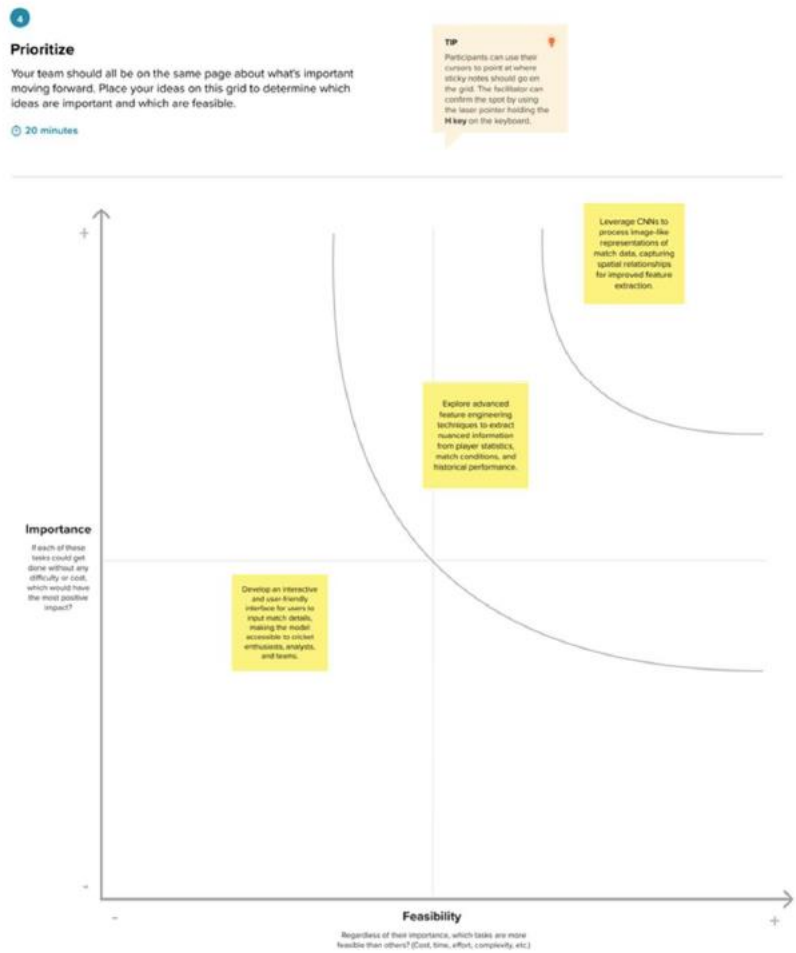
User-Friendly Interface: Develop an intuitive web interface for cricket analysts and enthusiasts.

Real-Time Updates: Incorporate a mechanism for real-time updates during matches to adjust predictions based on evolving match conditions.

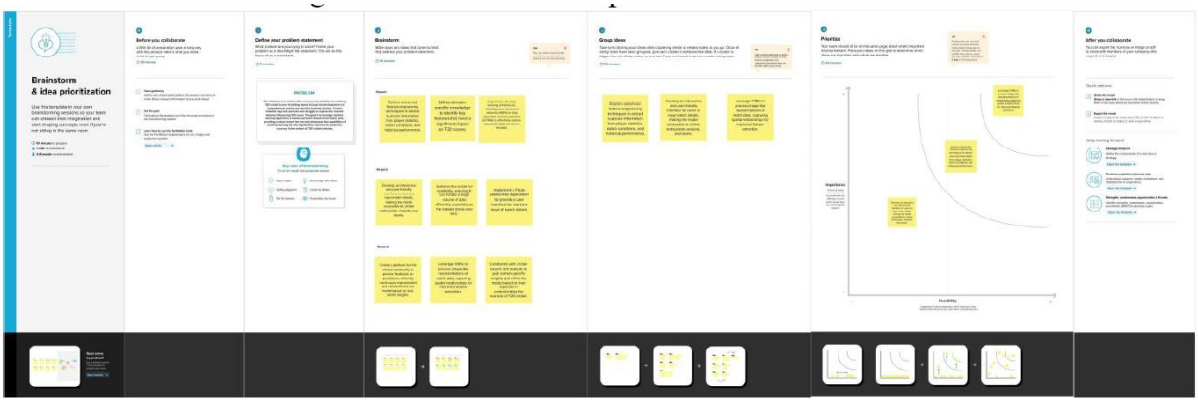
Community Engagement: Implement a community forum or discussion platform for users to share insights, strategies, and discuss the accuracy of predictions.

Machine Learning Explainability: Prioritize transparency in predictions by incorporating features that explain how the model arrives at its conclusions.





FULL BRAINSTORMING MAP:



4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT:

Data Ingestion:

The system should efficiently acquire historical and real-time T20 cricket match data from reliable sources.It must store the data securely in a scalable database.

Preprocessing Module:

The preprocessing module should cleanse and preprocess raw data, handling missing values and outliers.It must extract relevant features crucial for accurate T20 score prediction.

Machine Learning Model:

The machine learning model should be capable of continuous learning and updates based on real-time match data.It must accurately predict T20 cricket scores for batting teams.

User Interface:

The web application should provide an intuitive interface for cricket analysts and enthusiasts.It must include interactive visualizations and real-time updates during matches.

Scalable Architecture:

The system architecture should be designed for scalability to handle varying loads during live T20 matches.It must ensure efficient data flow and minimal latency.

4.2 NON-FUNCTIONAL REQUIREMENTS:

Performance:

The system should provide accurate predictions within a reasonable timeframe, even during peak usage.It must handle simultaneous user interactions without significant performance degradation.

Security:

Data encryption should be implemented to ensure the confidentiality and integrity of sensitive match data.Access controls must restrict unauthorized access to the system.

Reliability:

The T20 score prediction model should demonstrate high reliability and consistency in its predictions.The system should have mechanisms in place to recover gracefully from failures.

Usability:

The user interface must be user-friendly and intuitive, catering to both cricket analysts and enthusiasts.It should provide clear visualizations and real-time updates to enhance user experience.

Scalability:

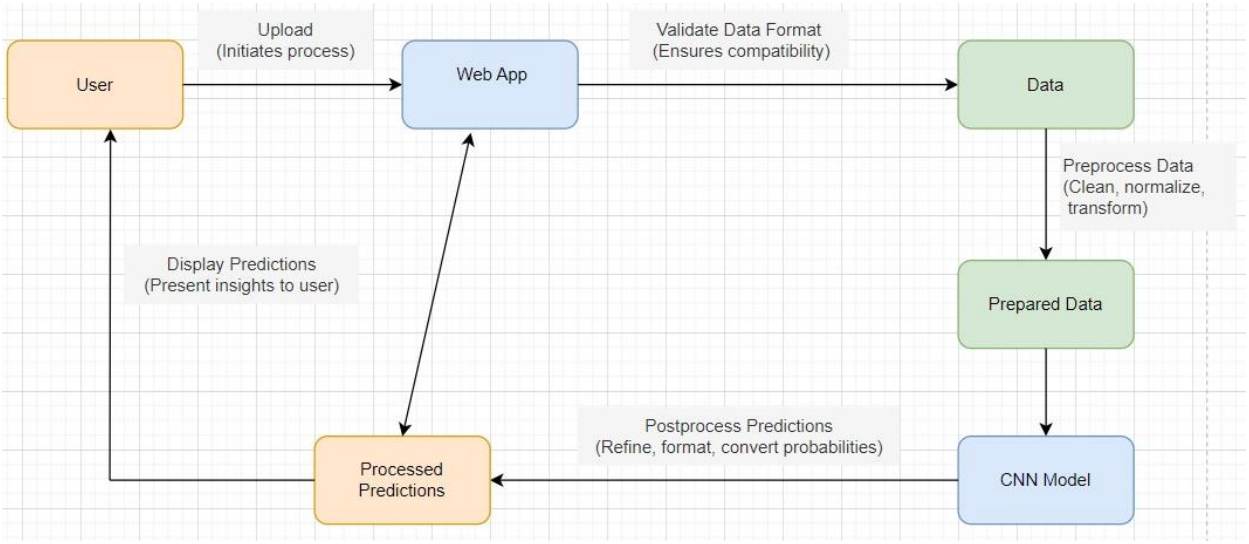
The system architecture should support scalability for handling an increasing number of users and data during live matches.It must be able to adapt to changing user demands without compromising performance.

Maintainability:

The codebase should be well-documented and modular to facilitate easy maintenance and updates.The system should allow for seamless integration of new features and enhancements.

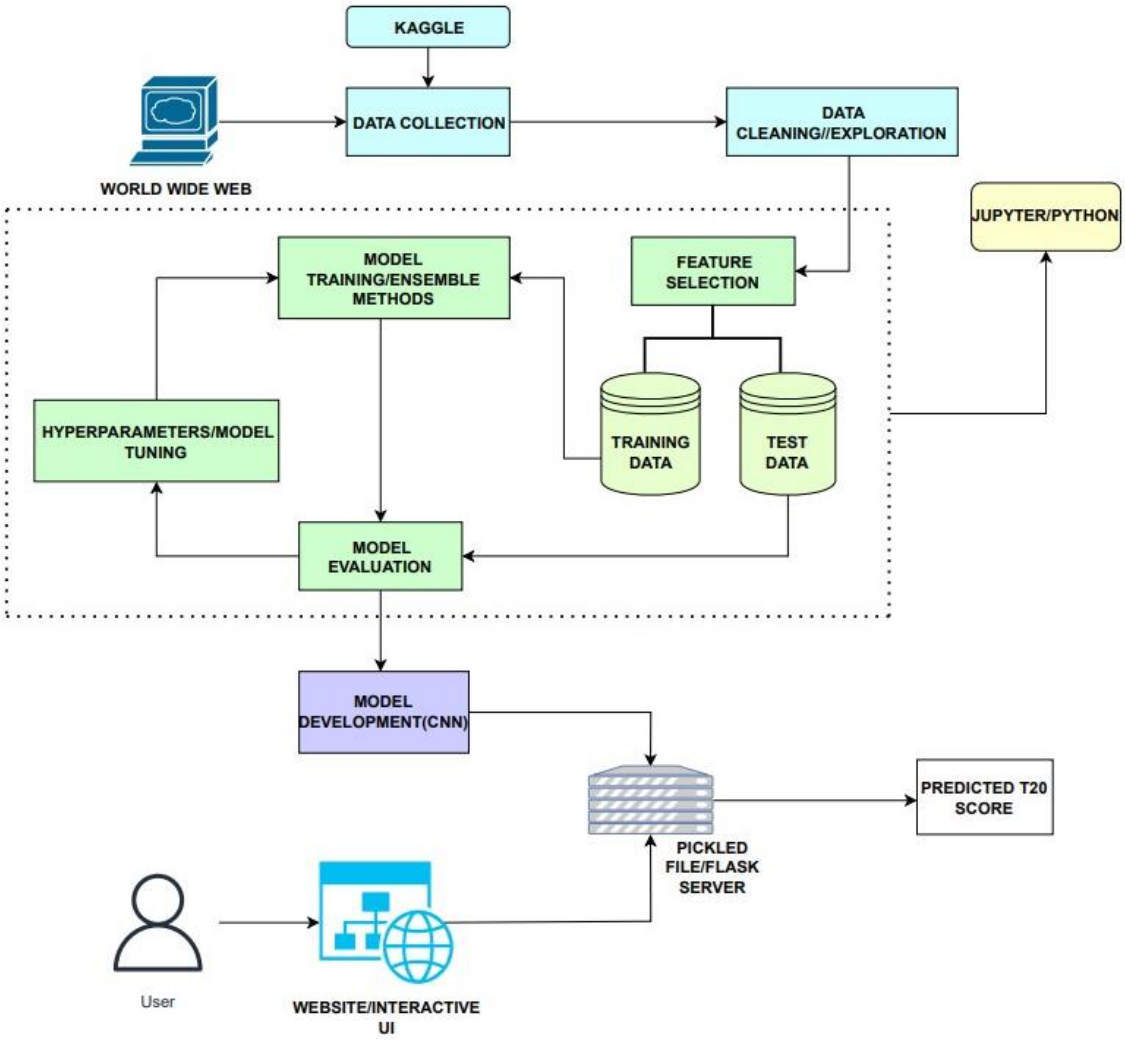
5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS & USER STORIES:



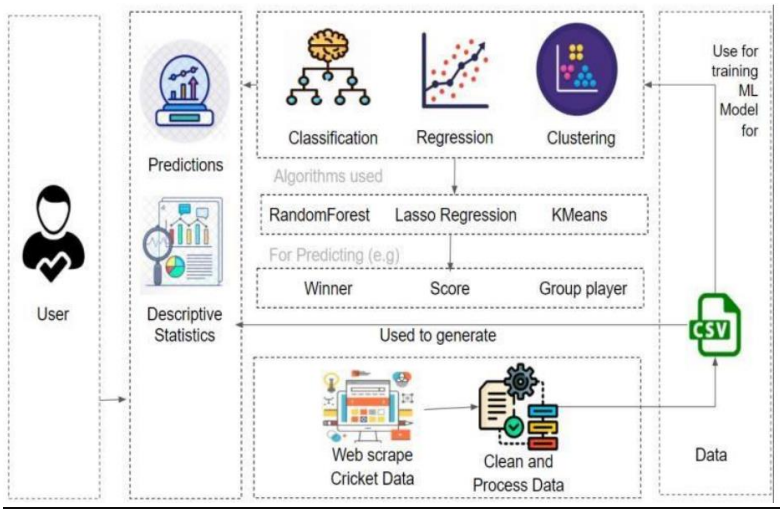
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	Discuss the different user stories that have been identified for the product, and how they will be implemented.	I can access my account/dashboard	High	Sprint-1
Customer (Mobile user)	Registration	USN-2	I will receive confirmation email once I have registered for the application.	I can receive confirmation email & click confirm	Medium	Sprint-1
Customer (Mobile user)	Registration	USN-3	I can register for the application through Facebook.	I can register & access the dashboard with Facebook Login	Medium	Sprint-2
Customer (Mobile user)	Registration	USN-4	I can register for the application through Gmail.	I can register & access the dashboard with Gmail Login	Medium	Sprint-1
Customer (Mobile user)	Login	USN-5	I can log into the application by entering email & password.	I can successfully login to the application	High	Sprint-1
	Dashboard		Monitoring the performance of a business or process: This could include tracking sales figures, customer satisfaction ratings, or operational efficiency metrics.	I can successfully access to the Dashboard	High	Sprint-1
Customer (Web user)	Web API	USN-6	Provide an interface for users to interact with the T20 Totalitarian score prediction model through a REST API.	Users can send a request to the REST API with the necessary information to make a prediction, and the API will return a prediction response.	High	Sprint-1
Customer (Web user)	Web API	USN-7	Users should be able to authenticate with the REST API using their account credentials.	Users should be able to generate a JWT token and include it in the request header to authenticate with the REST API.	High	Sprint-2
Customer (Web user)	Web API	USN-8	The REST API should be able to handle a high volume of requests.	The REST API should be able to handle a sustained load of	High	Sprint-2
Customer Care Executive	Web APP	USN-9	Provide a dashboard for customer care executives to view and manage customer support tickets.	The dashboard should display a list of open tickets, ticket details, and ticket history.	High	Sprint-2
Customer Care Executive	Web APP	USN-10	Allow customer care executives to filter and search tickets based on various criteria.	Customer care executives should be able to filter tickets by status, priority, assignee, and other relevant criteria.	Medium	Sprint-2
Administrator	Web Database		Provide a dashboard for administrators to view and manage overall system performance, user management, and ticket statistics.	The dashboard should display average ticket resolution time, customer satisfaction ratings.	High	Sprint-2

5.2 SOLUTION ARCHITECTURE:



6. PROJECT PLANNING & SCHEDULING

6.1 TECHNICAL ARCHITECTURE:



Data Flow:

Data flows from the sources to the ingestion module, preprocessing, and then to the machine learning model.Processed results are sent back to the user interface for display.

Model Deployment:

The machine learning model is deployed in a containerized environment, allowing for efficient updates and scaling.

User Interaction:

Users interact with the web interface to make predictions, and the interface communicates with the backend for data processing and result delivery.

Data Storage:

Choose a scalable and efficient database system (e.g., MongoDB, PostgreSQL).

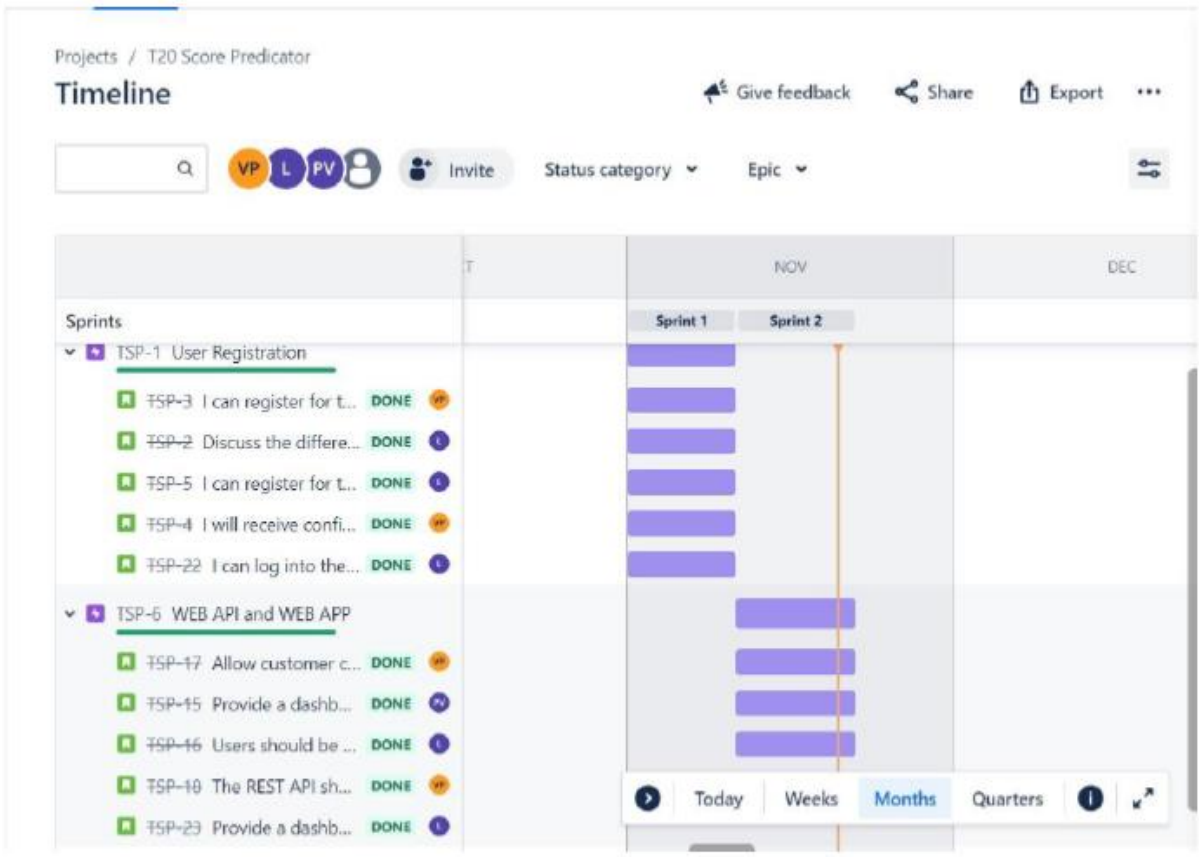
Machine Learning Framework:

Utilize popular machine learning libraries (e.g., Scikit-learn, TensorFlow) for model development.

Web Application Framework:

Select a framework (e.g., Flask, Django) for developing the user interface.

6.2 SPRINT PLANNING & ESTIMATION:



Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	10	10 Days	01 Nov 2023	10 Nov 2023	10	10 Nov 2023
Sprint-2	10	10 Days	11 Nov 2023	21 Nov 2023	10	21 Nov 2023

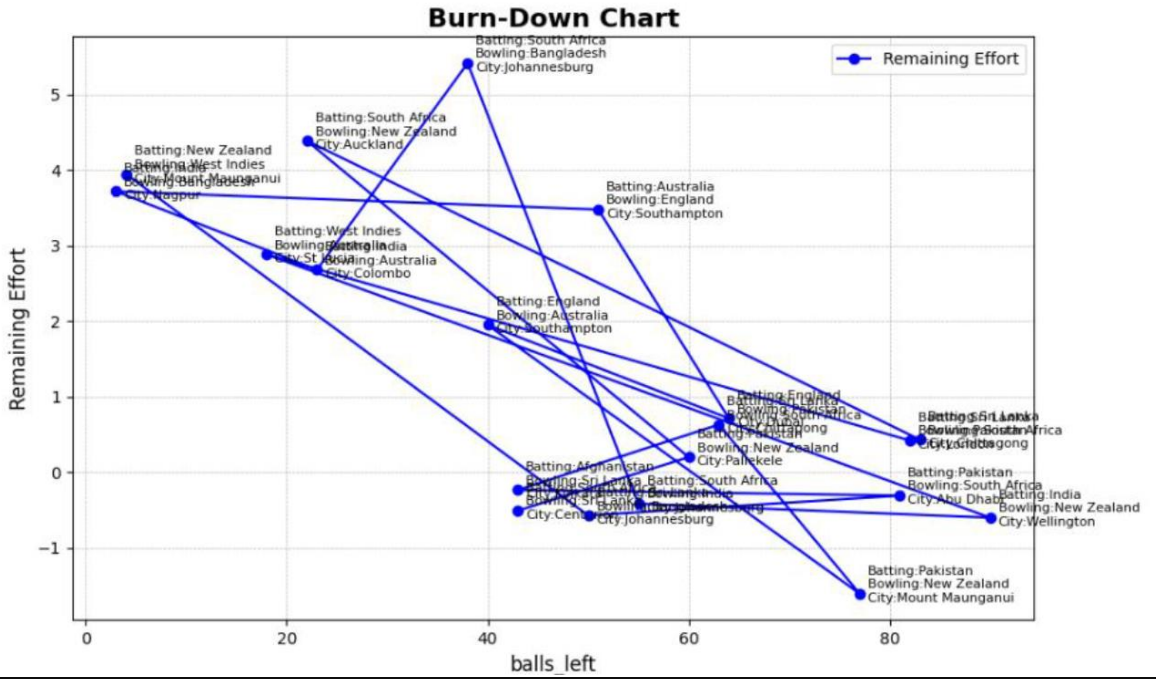
Velocity:

$$AV = \frac{sprint\ duration}{velocity} =$$

10/10=1

6.3 SPRINT DELIVERY SCHEDULE:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	Discuss the different user stories that have been identified for the product, and how they will be implemented.	2	High	2
Sprint-1	Registration	USN-2	I will receive confirmation email once I have registered for the application.	1	High	2
Sprint-2	Registration	USN-3	I can register for the application through Facebook.	1	Low	1
Sprint-1	Registration	USN-4	I can register for the application through Gmail.	2	Medium	2
Sprint-1	Login	USN-5	I can log into the application by entering email & password.	2	High	3
Sprint-1	Dashboard		Monitoring the performance of a business or process: This could include tracking sales figures, customer satisfaction ratings, or operational efficiency metrics.	1	High	2
Sprint-1	Web API	USN-6	Provide an interface for users to interact with the T20 Totalitarian score prediction model through a REST API.	2	High	2
Sprint-2	Web API	USN-7	Users should be able to authenticate with the REST API using their account credentials.	2	High	3
Sprint-2	Web API	USN-8	The REST API should be able to handle a high volume of requests.	2	High	2
Sprint-2	Web APP	USN-9	Provide a dashboard for customer care executives to view and manage customer support tickets.	1	High	2
Sprint-2	Web APP	USN-10	Allow customer care executives to filter and search tickets based on various criteria.	1	Medium	1
Sprint -2	Web Database		Provide a dashboard for administrators to view and manage overall system performance, user management, and ticket statistics.	1	High	3



7. CODING & SOLUTIONING

7.1 FEATURE 1:

Hyperparameter Tuning –

- GRU layer=50 units
- Activation function='relu'
- Optimizer='adam'
- Loss Function='MSE'
- Epoch=50
- Batch size=32

```
# Use early stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=3, restore_best_weights=True)

# Train the model
history = model.fit(
    X_train_reshaped, y_train_scaled,
    epochs=50,
    batch_size=32,
    validation_data=(X_val_reshaped, y_val_scaled),
    callbacks=[early_stopping]
)
```

7.2 FEATURE 2:

Validation Method –

- Early Stopping
- Scaling of data using minmaxscaler
- Reshaping of data
- Data split into:

Training=70%,Validation=20%,Testing= 10%

```
# Split the data into training, validation, and testing sets (70% training, 20% validation, 10% testing)
X_train, X_temp, y_train, y_temp = train_test_split(features, target, test_size=0.3, random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=1/3, random_state=42)
```

8. PERFORMANCE TESTING

8.1 PERFORMANCE METRICS:

Performance metrics are essential for evaluating the effectiveness and efficiency of the T20 cricket score prediction system.

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** MAE - , MSE - , RMSE – , R2 score -	The Regression metrics are: MSE: 42.58478785861209 MAE: 4.11531679731367 R-squared: 0.9594444804497875

```python
from sklearn.metrics import mean_squared_error, mean_absolute_error

# Calculate metrics
mse = mean_squared_error(y_test_inv, predictions_inv)
mae = mean_absolute_error(y_test_inv, predictions_inv)
print("The Regression metrics are:\n")
print('MSE: ', mse)
print('MAE: ', mae)

# Additional metrics
from sklearn.metrics import r2_score
r2 = r2_score(y_test_inv, predictions_inv)
print('R-squared: ', r2)
```

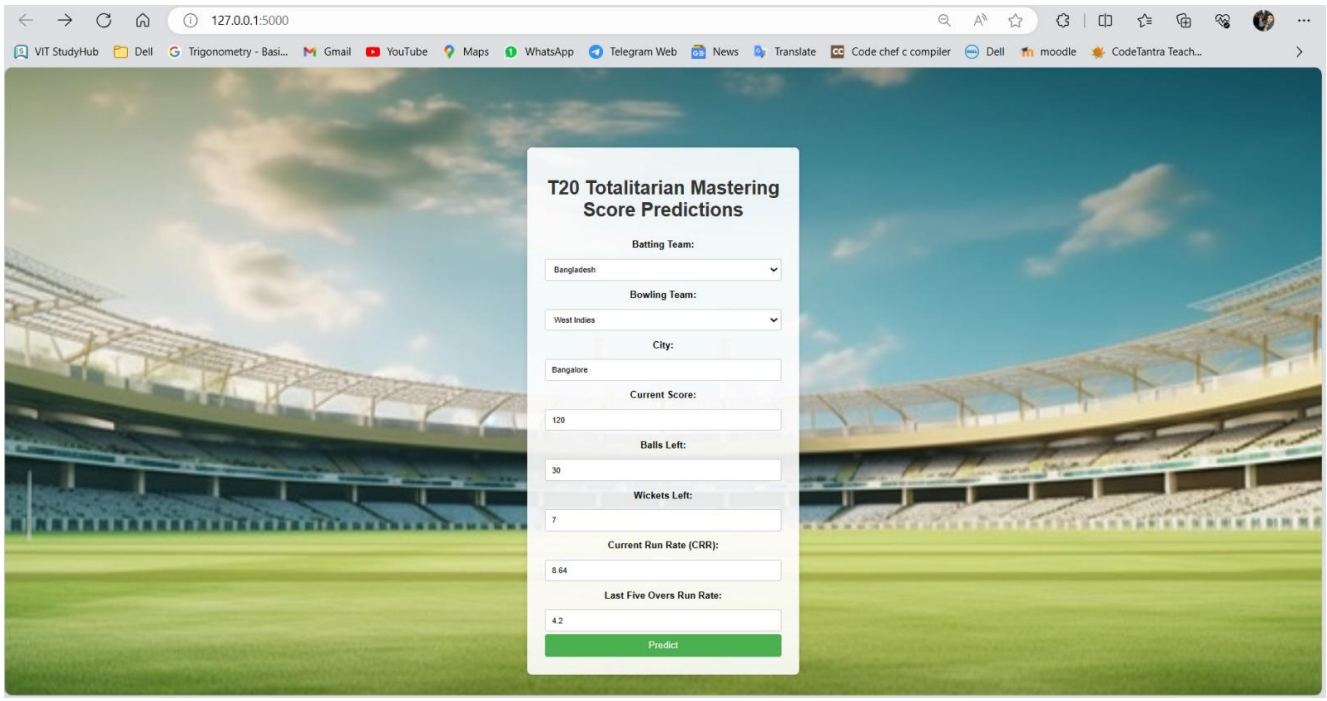
```
The Regression metrics are:

MSE:  42.58478785861209
MAE:  4.11531679731367
R-squared:  0.9594444804497875
```

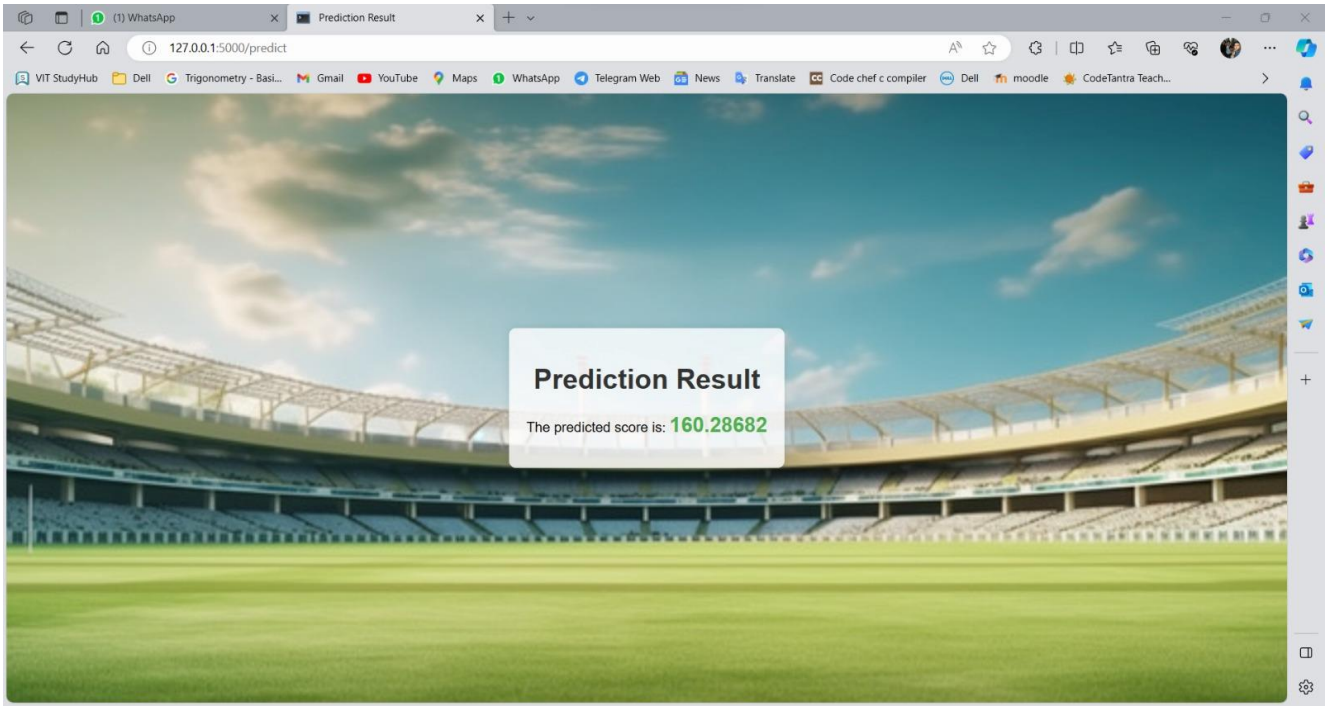
9. RESULTS

9.1 OUTPUT SCREENSHOTS

VALUES ENTERED:



PREDICTED OUTPUT:



10. ADVANTAGES & DISADVANTAGES

ADVANTAGES:

Informed Decision-Making:

Cricket analysts and enthusiasts can make more informed decisions based on predicted scores, enhancing strategic planning.

Enhanced Viewing Experience:

Fans can enjoy a more engaging viewing experience by anticipating and understanding the potential outcome of a T20 match.

Betting and Fantasy Sports:

Predictions contribute to the betting and fantasy sports industry, adding excitement and engagement for users.

Performance Analysis:

Teams and players can use prediction insights for performance analysis and improvement strategies.

Data-Driven Insights:

The system provides data-driven insights, facilitating a deeper understanding of team dynamics and player performances.

Entertainment Industry:

Television broadcasters and online streaming platforms can leverage predictions to enhance commentary and engage viewers.

DISADVANTAGES:

Uncertain Nature of Cricket:

Cricket is inherently unpredictable, and unforeseen events during a match can significantly impact the accuracy of predictions.

Over-Reliance on Data:

Relying solely on historical and statistical data may overlook the influence of external factors like weather conditions or player injuries.

Dynamic Nature of Teams:

Team compositions and strategies can change rapidly, making it challenging to accurately predict outcomes.

Ethical Concerns in Betting:

The integration of predictions in betting can raise ethical concerns, especially if users become overly dependent on predictions for gambling.

Limited Predictive Power:

The accuracy of predictions may be limited by the complexity of the game and the multitude of variables that influence outcomes.

Model Training Challenges:

Continuous model training poses challenges, and adapting to real-time changes in player form or team dynamics can be difficult.

Impact of Injuries:

Injuries to key players can significantly alter match dynamics, and predicting the impact of such events is inherently challenging.

Overemphasis on Results:

Overemphasis on predicted scores may detract from the enjoyment of the game itself and the element of surprise that makes cricket exciting.

11. CONCLUSION

After studying the research papers, we found that most of the papers have only considered the T20 format. There is a need to focus on the TEST format. The T20 format is a slow format and the parameters for prediction may vary. Linear Regression and Random forest algorithms are preferred by most of the researchers. These algorithms gave the highest accuracies. The researchers have not considered the dynamic feature of the TESTs. Now after studying these papers, we are planning to create a model which can make a better prediction for the TEST game with higher accuracy.

12. FUTURE SCOPE

- Accuracy can be increased.
- More tweets and social media inputs can be feed into the models.
- Can be used for the tournaments like BBL, CPL, Vitality Blast, etc
- Accuracy can be increased by predicting the matrices of the player against the player.
- T20 and TEST format can be included.
- This methodology and technique can also be applied to predict the outcomes of games like hockey and football.

13. APPENDIX

SOURCE CODE:

https://drive.google.com/drive/folders/1YZanK_8p0rYvosSYpOiPNBhtH_zDzUwY?usp=sharing

GITHUB LINK:

https://github.com/smartinternz02/SI-GuidedProject-610110-1700320226

PROJECT DEMO LINK:

https://drive.google.com/file/d/12w9UGrYywViKtql9JRWS_NAwcoIKOTc6/view