|  |  |  |  |
| --- | --- | --- | --- |
| Machine Learning Project Proposal Template | | | | |
| | Student name | ROLL NUMBER | BATCH | EMAIL ID | | --- | --- | --- | --- | | MUHAMMED SAFVAN PS | LLAI24259 | F | safvanlearnlogic@gmail.com | | | | | |

| Project Overview |
| --- |
|  |

* **Project Title:** (IPL-Analysis)
* **Project Description:**

| The problem statement | We have a dataset with ball-by-ball details of IPL matches, including batting, bowling, runs, extras, and wickets. The goal is to analyze this data to understand team performance, evaluate players, and predict match results |
| --- | --- |
| Objectives | * **Study how players perform in batting and bowling.** * **Find patterns in team performance to improve strategies.** * **Predict match results using past data.** |
| The real-world significance | * **Helps teams make better decisions using data.** * **Gives fans more insights into matches.** |
| References | * **IPL dataset from Kaggle.** * **ChatGpt.** |

* **Expected Outcome:**

| Accuracy or performance expectations | * Find the best players and teams. * Predict match results with 85-90% accuracy. * Give useful insights for better decisions. |
| --- | --- |
| Future Scope | * **Make predictions using live match data.** * **Add more features like player fitness and weather conditions.** |

* **Dataset Description:**

| Source of dataset | Kaggle IPL dataset: <https://www.kaggle.com/datasets/patrickb1912/ipl-complete-dataset-20082020> |
| --- | --- |
| Number of Instances and features | * **2,60,920 rows (instances)** * **17 columns (features)** |
| Description of features and target variable | * **match\_id – Unique ID for each match.** * **inning – The inning number (1st or 2nd).** * **batting\_team – The team currently batting.** * **bowling\_team – The team currently bowling.** * **over – The over number in the innings.** * **ball – The ball number in the over.** * **batter – The batsman facing the ball.** * **bowler – The bowler delivering the ball.** * **non\_striker – The batsman at the non-striker’s end.** * **batsman\_runs – Runs scored by the batsman.** * **extra\_runs – Extra runs due to no-ball, wide, etc.** * **total\_runs – Total runs scored in that ball (batsman + extras).** * **extras\_type – Type of extra run (wide, leg bye, etc.).** * **is\_wicket – 1 if a wicket fell, 0 otherwise.** * **player\_dismissed – Name of the player who got out.** * **dismissal\_type – How the batsman got out (bowled, caught, etc.).**   **Target Variable:**   * **total\_runs – Used to analyze team performance and predict scores.** |
| Data Type | * **Numbers and text (integer, float, string)** |
| Any preprocessing required | * **Handle missing values in some columns.** * **Convert text data into numerical form if needed.** |

* **Proposed Methodology:**

(Outline the steps you plan to follow, including model selection, preprocessing techniques, and evaluation metrics.)

| Data Collection:  Data Cleaning:  Data Exploration:  Feature Selection:  Model Selection:  Training & Testing:  Evaluation:  Visualization: | Get the IPL dataset from Kaggle.  Remove missing values and correct errors.  Find patterns in batting, bowling, and team performance.  Pick important factors like runs, wickets, and extras.  Use machine learning models like Decision Trees or Random Forest.  Train the model on past data and test its accuracy.  Check performance using accuracy, precision, and recall.  Show results using graphs and charts in Power BI Dashboard. |
| --- | --- |

| **APPROVAL STATUS** (To be filled by faculty) |
| --- |
|  |

| STATTUS | **☐ Approved ☐ Not Approved** |
| --- | --- |
| **Faculty Comments** |  |
| Faculty name | DATE | Signature |
|  |  |  |