**Dataset Details**  
This dataset provides ball-by-ball coverage of:

* **1188 ODI matches** (data/odi.csv)
* **1474 T20 matches** (data/t20.csv)
* **617 IPL matches** (data/ipl.csv)

Each dataset contains the following columns:

* **mid**: Unique identifier for each match
* **date**: Match date
* **venue**: Stadium where the match was played
* **bat\_team**: Name of the batting team
* **bowl\_team**: Name of the bowling team
* **batsman**: Batsman facing the ball
* **bowler**: Bowler delivering the ball
* **runs**: Cumulative team runs at that ball
* **wickets**: Total wickets fallen at that point
* **overs**: Total overs bowled at that stage
* **runs\_last\_5**: Runs scored in the last 5 overs
* **wickets\_last\_5**: Wickets lost in the last 5 overs
* **striker**: Higher runs scored between striker and non-striker
* **non-striker**: Lower runs scored between striker and non-striker
* **total**: Total runs scored by the team after the first innings

**Prediction Models and Accuracy**  
Two regression models were utilized:

1. **Linear Regression** (linear\_regression.py)
2. **Random Forest Regression** (random\_forest\_regression.py)

**Features Used**: [runs, wickets, overs, striker, non-striker]  
**Label**: [total]

**Accuracy Metrics**:

* **R² Score**
* **Custom Accuracy**: Accuracy based on the difference between predicted and actual scores. Predictions within a specified threshold are considered accurate:
  + **T20**: Threshold = 10
  + **ODI**: Threshold = 20

| **Model** | **ODI Accuracy ([R², Custom])** | **T20 Accuracy ([R², Custom])** | **IPL Accuracy ([R², Custom])** |
| --- | --- | --- | --- |
| Linear Regression | [52, 43] | [52, 44] | [50, 44] |
| Random Forest Regression | [79, 77] | [64, 59] | [67, 65] |