

IPL Score Prediction Using Linear Regression

Project Overview

This project predicts the **final IPL team score** during an ongoing match using a **Linear Regression model**. By analysing real-time match data such as runs, wickets, overs, and recent performance, the model forecasts the expected final total.

Objective

To build a machine learning model that estimates the **final score** of a batting team based on the match's current progress.

Dataset

- **Source:** Kaggle – IPL Ball-by-Ball Dataset
 - **Features Used:**
 - bat_team – Batting team name
 - bowl_team – Bowling team name
 - overs – Overs completed
 - runs – Current total runs
 - wickets – Wickets fallen
 - runs_last_5 – Runs scored in the last 5 overs
 - wickets_last_5 – Wickets lost in the last 5 overs
 - total – Final match score (target)
-

Data Preprocessing

1. Filtered data to include recent IPL seasons.
 2. Removed incomplete or invalid records.
 3. Applied **One-Hot Encoding** to team names.
 4. Split data into:
 - **Training set:** Matches before 2017
 - **Testing set:** Matches from 2017 onwards
-

Model Building

- **Algorithm Used:** Linear Regression
- **Why:** It's simple, interpretable, and performs well for numerical predictions with linear relationships.

```
from sklearn.linear_model import LinearRegression
```

```
model = LinearRegression()  
model.fit(X_train, y_train)  
y_pred = model.predict(X_test)
```

Evaluation Metrics

Metric	Value (approx.)
--------	-----------------

MAE (Mean Absolute Error) ~12.4

RMSE (Root Mean Square Error) ~16.2

 The model performs decently, with predictions close to actual final scores.

Prediction Example

```
final_score = predict_score(model, 'Sunrisers Hyderabad', 'Delhi Daredevils',  
    overs=12.4, runs=98, wickets=3, runs_in_prev_5=40, wickets_in_prev_5=2)  
print(f"Predicted Final Score: {final_score - 10} to {final_score + 5}")
```

Output Example:

Predicted Final Score: **165 to 180**

Results & Insights

- The Linear Regression model captures general scoring trends well.
 - Predictions vary slightly for aggressive or slow matches due to linear assumptions.
 - Increasing overs and runs proportionally raises the predicted total, aligning with match logic.
-

Tools & Technologies

- **Language:** Python
 - **Libraries:** pandas, numpy, scikit-learn, matplotlib
 - **Techniques:** Data Cleaning, One-Hot Encoding, Regression, Model Evaluation
-

Conclusion

The **Linear Regression model** provides a simple yet effective method to estimate final scores in IPL matches. It demonstrates how **machine learning can enhance sports analytics** by using in-match data to forecast outcomes.