Bangladesh Army University of Science and Technology (BAUST)

Department of Computer Science and Engineering

Course Title: Machine Learning Sessional

Course Code: CSE 4140

Project Title: Prediction of IPL Win Probability Using Machine Learning Techniques

Presented by:

Md. Tanvirul Islam

220201002

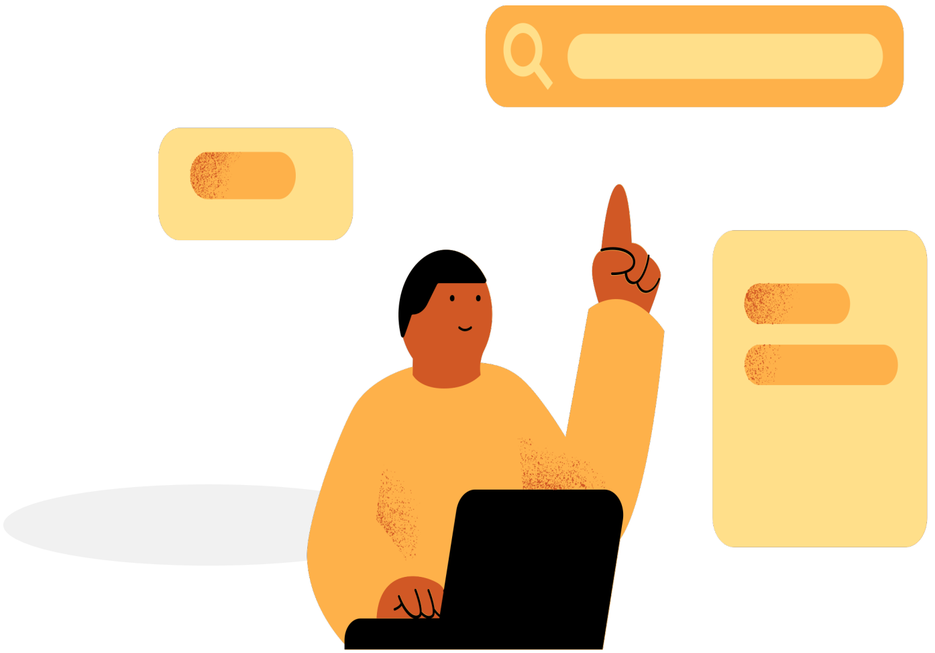
Presente for:

Engr. Rohul Amin, Lecturer

Nadim Reza, Lecturer

25 Nov 2025

Contents



Introduction

- Cricket is a highly dynamic sport where the match situation changes every ball.
- Predicting win probability requires understanding runs, wickets, overs, and match context.
- Machine learning can help analyse match situations in real-time.
- This project predicts IPL win probability during the second innings using ball-by-ball data.
- Logistic Regression, SVM, and TabNet models were implemented and compared.

Objectives

- Build a machine learning model to predict real-time IPL win probability.
- Use ball-by-ball second innings data for mid-match prediction.
- Implement safe, leakage-free feature engineering.
- Evaluate multiple models using accuracy, F1-score, recall, ROC-AUC.
- Deploy the model with a user-friendly Streamlit web interface.

Related Work (Paper 1)

In [1], A. T. Maigu. in 2024, proposed a “IPL Win Probability Predictor using Machine Learning Techniques.”

Findings:

- Used historical IPL match and player data.
- **Methods:** Logistic Regression, Gradient Boosting.
- **Strengths:** Good baseline models, feature exploration.
- **Limitations:** No real-time prediction, train-test split unclear, risk of data leakage.

Related Work (Paper 2)

In [2], S. M. Zohaib, N. Sharma, R. Singh, and M. Sonia. in 2025, proposed a "IPL Win Probability Prediction System using Machine Learning Techniques."

Findings:

- **Dataset:** ball-by-ball IPL data for 16 seasons.
- **Methods:** Random Forest, Logistic Regression.
- Used 70–30 split.
- Focused mainly on end-of-match prediction.
- Limited mid-innings prediction.

Related Work (Paper 3)

In [3], A. Vishwanath, C. M. Gulzar, J. S. Reddy, E. Manjunath, and M. S. Sukumar. in 2025, proposed a "Analyzing and Estimating IPL Winner Prediction Using Machine Learning."

Findings:

- Combined historical and real-time match data.
- **Methods:** Logistic Regression, Gradient Boosting.
- Deployed with Streamlit and Heroku.
- **Limitations:** No proper leakage control, No advanced deep tabular models.

Related Work (Paper 4)

In [4], A. Tripathi, R. Islam, V. Khandor, and V. Murugan. in 2020, proposed a "Prediction of IPL Matches using Machine Learning while Tackling Ambiguity in Results."

Findings:

- Used team composition, toss, venue, player stats.
- **Models**: LR, SVM, RF, Gradient Boosting, RNN, HMM.
- Feature-rich system but computationally heavy.
- Not suitable for real-time mid-match prediction.

Common Gaps in Literature

- Limited focus on mid-innings real-time prediction.
- Complex models (RNN/HMM).
- Lack of deep tabular models like TabNet.
- Real-time deployment often missing.

Proposed System

- Uses only mid-match features from second innings.
- Leakage-free dataset split (match-based).
- **Features include:**
  - Current score
  - Balls left
  - Wickets left
  - CRR, RRR
- **Algorithms used:** Logistic Regression, SVM, TabNet.
- Deployed with Streamlit for real-time use.

Methodology Flow

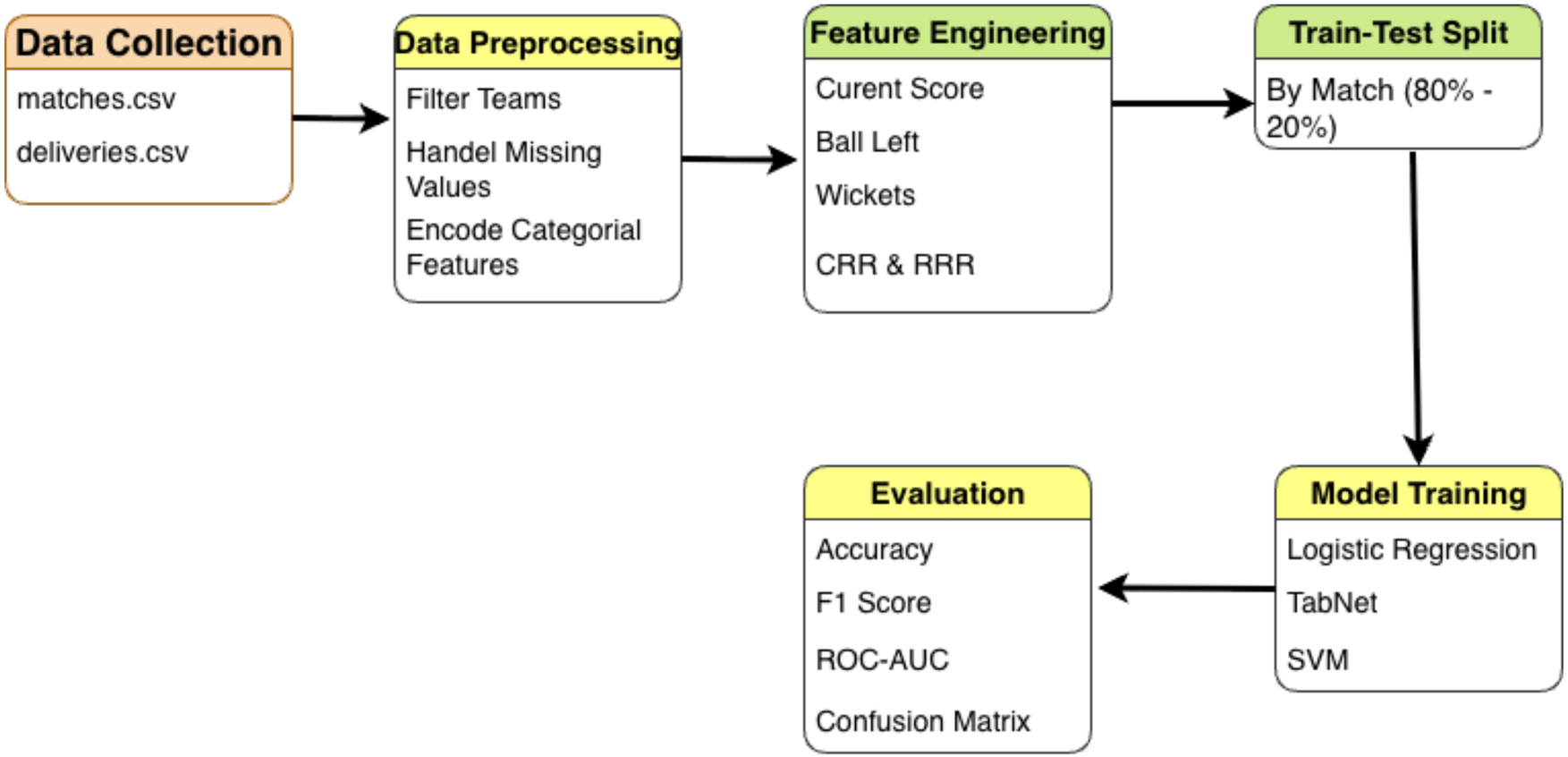


Fig. 1. Research Methodology

Feature Engineering

- **Runs Left:**

  runs_left = target - current_score

- **Balls Bowled:**

  balls_bowled = over * 6 + ball

- **Balls Left:**

  120 - balls_bowled

- **Current Run Rate (CRR):**

  (current_score * 6) / balls_bowled

- **Required Run Rate (RRR):**

  (runs_left * 6) / balls_left

- **Wickets Left:**

  10 - cumulative_dismissals

Dataset Split

- Splitting by rows causes leakage.
- I used match-wise split:
  - 80% matches → training
  - 20% matches → testing
- Ensures real-world scenario simulation.

Models Used

Logistic Regression:

- Simple, interpretable baseline.

Support Vector Machine (SVM):

- Non-linear boundary, improved accuracy.

TabNet:

- Deep learning tabular model.
- Uses sequential attention.
- Best performance.

Model Results Table

Table 1. Model Perfomance

Model	Accuracy	F1 Score	Recall	ROC-AUC
Logistic Regression	0.8145	0.8248	0.8324	0.8962
SVM (Support Vector Machine)	0.8262	0.8394	0.8680	0.9102
TabNet	0.8434	0.8519	0.8602	0.9317

TabNet achieves the best results.

ROC Curves

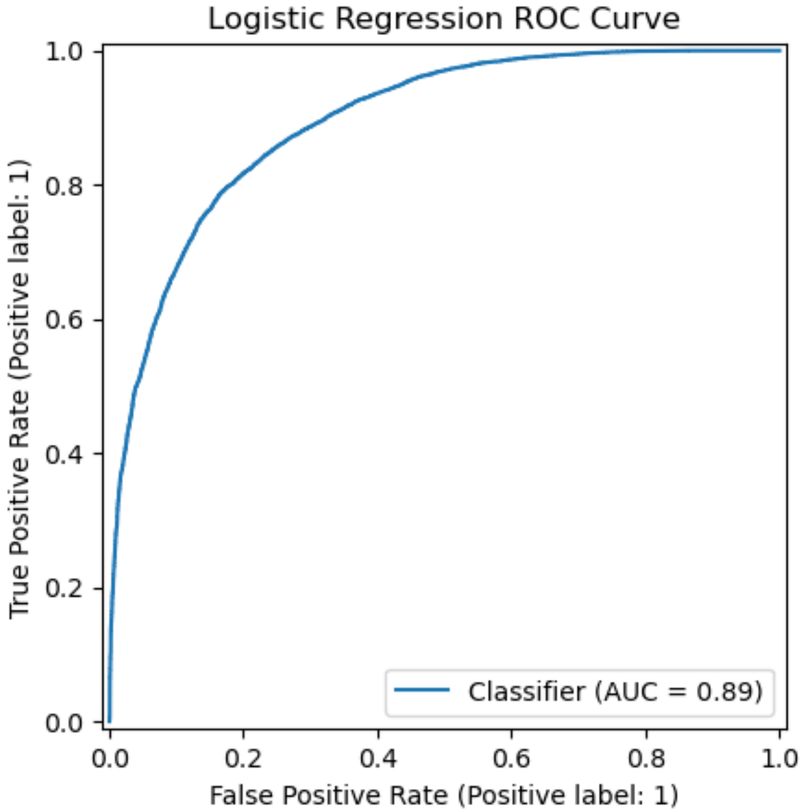


Fig. 2. LR ROC Curve

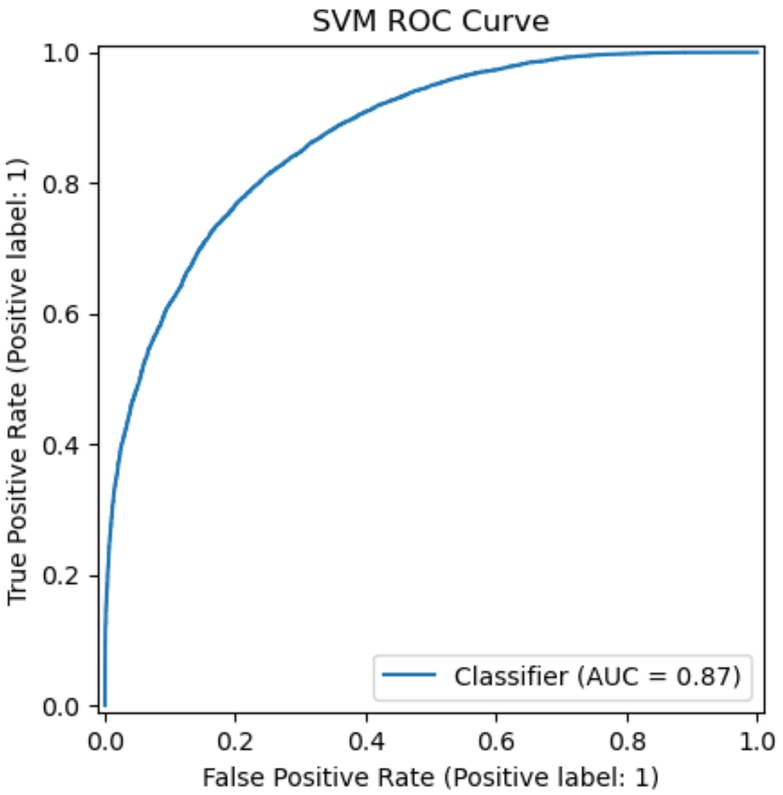


Fig. 3. SVM ROC Curve

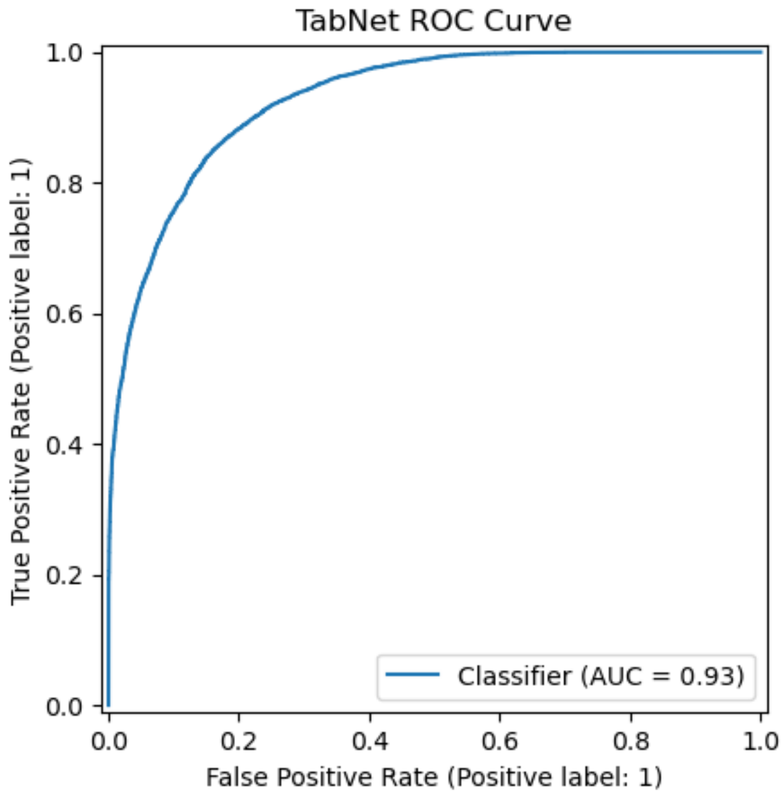


Fig. 4. TabNet ROC Curve

Confusion Matrices

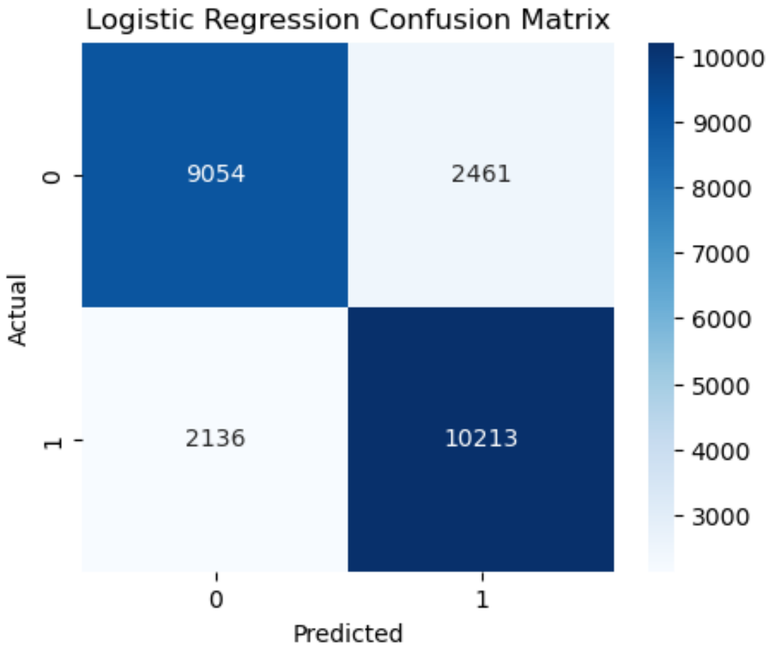


Fig. 5. LR Confusion Matix

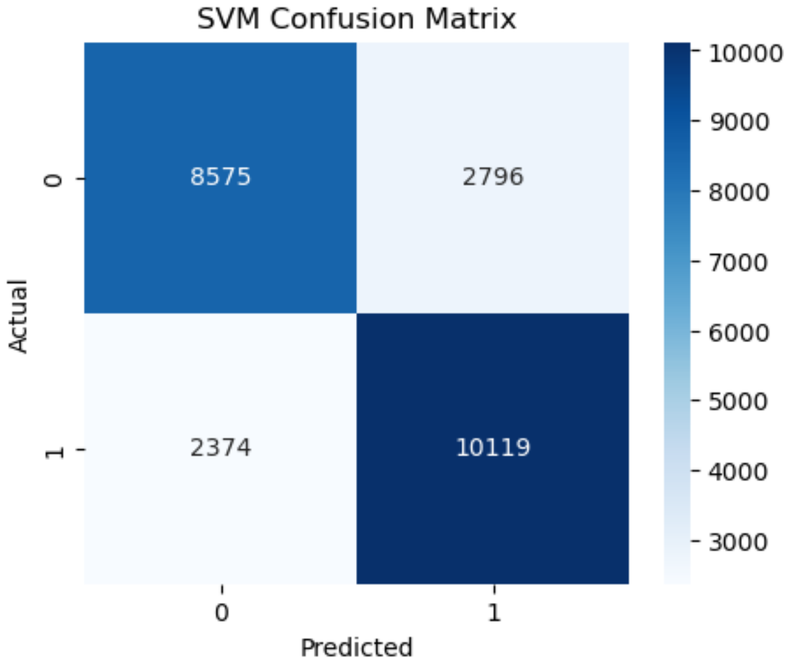


Fig. 6. SVM Confusion Matix

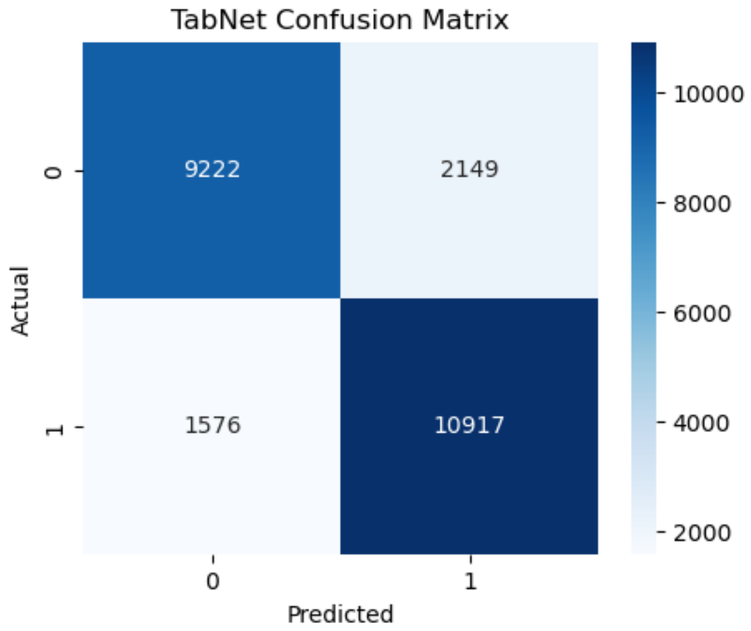


Fig. 7. TabNet Confusion Matix

Streamlit UI

Features of UI:

- Dropdowns for team and city
- Numeric inputs for overs, score, wickets
- Automatic computation of CRR, balls left
- Outputs win probability
- Easy for non-technical users

Fig. 8. Streamlit UI of the ML Project

Future Work

- Add player-level features (strike rate, economy).
- Use LSTM/Transformer for sequence modelling.
- Predict over-by-over win probability graph.
- Deploy end-to-end mobile or web app.

Conclusion

- Built a leakage-free IPL win prediction model.
- Feature engineering based on real match conditions.
- TabNet performed best across all metrics.
- Streamlit deployment enables real-time use.
- System useful for analysts, fans, and teams.

References

[1] A. T. Maigu, “IPL Win Probability Predictor using Machine Learning Techniques,” International Journal of Research Publication and Reviews, vol. 5, no. 9, pp. 3146–3151, Sep. 2024.

[2] S. M. Zohaib, N. Sharma, R. Singh, and M. Sonia, “IPL Win Probability Prediction System using Machine Learning Techniques,” International Journal for Research in Applied Science & Engineering Technology (IJRASET), Paper ID: IJRASET69906, Apr. 28, 2025, ISSN: 2321-9653.

[3] A. Vishwanath, C. M. Gulzar, J. S. Reddy, E. Manjunath, and M. S. Suku-mar, “Analyzing and Estimating IPL Winner Prediction Using Machine Learning,” International Journal of Scientific Research & Engineering Trends, vol. 11, no. 2, pp. –, Mar.-Apr. 2025, ISSN (Online): 2395-566X.

[4] A. Tripathi, R. Islam, V. Khandor, and V. Murugan, “Prediction of IPL Matches using Machine Learning while Tackling Ambiguity in Results,” Indian Journal of Science and Technology, vol. 13, no. 38, pp. 4013–4035, 2020.

Do you have any questions?

Thank You