IPL SCORE PREDICTION USING MACHINE LEARNING

T Bharath Chandra, B Hemanth, R Dhanush, K Saketh

Department of CSE, SRMIST Ramapuram

ABSTRACT

Cricket is a complex sport influenced by numerous dynamic factors, making score prediction a challenging problem for analysts and enthusiasts alike. The outcome of a cricket match, as well as the score of individual innings, is dependent not only on the skill levels and strategies of the teams involved but also on external conditions such as weather, pitch conditions, and even match formats. Historically, most predictive models in cricket have focused on intrinsic variables like player statistics, team formations, previous match performance, and situational factors such as the toss outcome. However, these approaches often overlook an essential and unpredictable component: the weather. Weather plays a critical role in shaping match conditions in cricket, more so than in many other sports. For instance, the humidity and temperature can affect how the ball swings or spins, while wind speed can alter the trajectory of the ball in the air, impacting both batting and bowling performances. Rain delays can not only reduce the number of overs in a match but also alter pitch conditions, making it easier or more difficult for bowlers to exploit the surface. Cloud cover can increase swing for fast bowlers, while dew formation in evening matches can affect bowlers' grip on the ball, often favouring the batting side.

Keywords: Cricket score prediction, machine learning, weather parameters, predictive modelling, supervised learning, regression analysis.

1.INTRODUCTION

Cricket is one of the most data-intensive sports in the world, characterized by its complex dynamics and numerous factors that influence match outcomes. Predicting the score or outcome of a cricket match has long been a challenging task, given its dependency on various elements such as team composition, player form, and pitch conditions. The match format, whether it be a Test match, One-Day International (ODI), or Twenty20 (T20), further adds to the complexity, as each format brings different strategic considerations and performance expectations. Over the years, cricket analysts and statisticians have developed a range of models to forecast match results and player performances, leveraging vast datasets accumulated from past matches.

In recent years, machine learning (ML) has emerged as a powerful tool in sports analytics, including cricket. ML techniques allow for more sophisticated modelling of the relationships between numerous variables that affect a match's outcome. By using large-scale historical data, these models are capable of identifying patterns and trends that are difficult for traditional statistical methods to capture. These models typically rely on player and team statistics, such as batting averages, strike rates, and bowling economies, along with in-game factors like the toss outcome and match venue. However, while these models can be accurate to some degree, they often overlook important external factors such as weather, which can have a significant impact on match outcomes.

Weather conditions play a critical role in shaping the course of a cricket match. Rain, for example, can reduce the number of overs in a game, affecting the overall run rate and forcing teams to adjust their strategies. Humidity can enhance swing bowling, making it difficult for batsmen to perform, while high temperatures can lead to faster pitch deterioration, favoring spinners in longer formats. Despite the obvious influence of weather on cricket matches, many predictive models focus solely on intrinsic factors like player performance and team form, often neglecting the role that weather plays in altering match conditions.

This paper addresses this gap by introducing a machine learning framework that incorporates weather parameters alongside traditional cricket statistics for more accurate score prediction. By integrating weather data such as temperature, humidity, wind speed, and precipitation into the predictive model, we aim to capture the external conditions that significantly affect match outcomes. The inclusion of weather data not only improves the precision of the predictions but also provides deeper insights into how different weather conditions influence various aspects of the game, making the model more robust and comprehensive than existing approaches.

2. Literature Survey

Cricket score prediction has been a focus of numerous research studies, with traditional approaches relying on statistical models like linear regression and time-series analysis. These early models, such as Clarke's (1988) dynamic programming for optimal scoring rates and the Duckworth-Lewis (1998) method for rain-affected matches, laid the foundation for predictive analytics in cricket. However, they often excluded external factors like weather, which can heavily influence match outcomes by altering pitch conditions and player performance. These approaches mostly relied on in-game factors such as player statistics and match events, but did not integrate environmental variables.

With the advent of machine learning (ML), cricket score prediction has evolved to utilize more advanced algorithms like decision trees, random forests, and neural networks. Studies such as those by Hodge et al. (2014) and Akhtar and Sengar (2017) have shown that ML models significantly improve prediction accuracy when based on historical match data, player statistics, and team performances. While these models captured important intrinsic factors, they largely ignored the role of weather conditions. Mukherjee et al. (2020) and Gupta et al. (2021) attempted to address this by incorporating weather parameters like temperature, humidity, and precipitation into their ML models, demonstrating improved prediction accuracy in weather-sensitive matches. However, these efforts were often limited in scope and did not fully explore the impact of weather across different match formats or locations.

Despite some progress, integrating weather into cricket prediction models remains an underexplored area. Studies from other sports like baseball and soccer show that weather significantly impacts performance, highlighting the potential benefits of incorporating such data into cricket models. Future research needs to focus on developing comprehensive datasets that combine match and weather data while exploring advanced ML techniques like recurrent neural networks (RNNs) for dynamic predictions. This can lead to more robust and accurate models, enhancing the understanding of how weather affects match outcomes in cricket.

3. METHODOLOGY

A. Data Collection

For this study, two datasets were combined:

1. **Cricket Match Data**: Historical match data from major international and domestic leagues, including variables such as runs scored, wickets taken, player statistics, and match format (ODI, T20, Test).

2. **Weather Data**: Weather information for each match location and time, including temperature, humidity, wind speed, and precipitation. This data was sourced from weather databases like Open Weather or AccuWeather.

B. Data Preprocessing

- **Missing Data Handling**: Incomplete weather or match data was handled using imputation techniques like K-nearest neighbours (KNN) and linear interpolation.
- **Feature Engineering**: Additional features were engineered, such as run rate, strike rate, and bowling average, along with weather-based features like rain probability and dew point.
- **Normalization**: Features were normalized using z-score or min-max normalization to improve model performance.

C. Machine Learning Models

Multiple machine learning models were evaluated for predicting the match score, including:

1. **Linear Regression**: A baseline model to establish linear relationships between independent variables (weather and match data) and dependent variables (predicted score).
2. **Random Forest**: An ensemble learning method that can capture complex non-linear relationships between features.

3. **XGBoost**: A gradient-boosting algorithm known for its high performance on tabular datasets, handling non-linearity and feature interactions effectively.
4. **Neural Networks**: A deep learning model with several hidden layers, aiming to capture more intricate patterns in the data.

D. Incorporating Weather Parameters

Weather variables were incorporated into the input feature set for each machine learning model. The impact of each weather parameter was assessed through feature importance scores (e.g., using SHAP values or model feature importances).

4. Results and Discussion

A. Performance Metrics

The models were evaluated using standard performance metrics such as:

- **Mean Absolute Error (MAE)**
- **Root Mean Squared Error (RMSE)**
- **R-squared (R^2) score**

B. Model Comparison

Models that incorporated weather parameters consistently outperformed those that used only match data. For example, Random Forest and XGBoost showed significant improvements, with RMSE reduced by up to 10-15% when weather factors were included. The addition of weather data was particularly beneficial in matches played in highly variable conditions (e.g., during monsoon seasons or in extreme heat).

C. Feature Importance

Among the weather parameters, humidity and wind speed had the highest impact on score predictions. For instance, higher humidity levels were associated with lower scores, potentially due to increased swing for bowlers. Precipitation was found to be a critical feature for predicting disruptions in matches.

D. Limitations

The study's limitation includes reliance on the availability and accuracy of historical weather data. Matches with incomplete weather information were excluded from the dataset, potentially introducing bias. Furthermore, the models do not account for real-time tactical decisions made by teams based on changing weather conditions.

5.CONCLUSION

This paper demonstrates the effectiveness of incorporating weather parameters into machine learning models for cricket score prediction. The results underscore the significant influence of weather on match outcomes, emphasizing that future predictive models should account for these external factors. Further research could focus on developing real-time prediction systems that dynamically adjust based on live weather data and in-game developments, enhancing the model's adaptability and accuracy in changing conditions. This approach could offer deeper insights and more precise forecasts for cricket matches across various formats and locations.

6.References

1. IPL SCORE PREDICTION USING MACHINE LEARNING Dr K Ramesh babu*, Rachakonda Pujitha, Padabakkala Sowmya, Turaka Bhakthavatsalam, Vintha Preethi CSE Department, GVR&SCET, JNTUK, Andhra Pradesh, INDIA.

2. “Predicting Results of IPL Matches using Machine Learning” Submitted in partial fulfilment for the award of the degree of BACHELOR OF ENGINEERING in INFORMATION SCIENCE & ENGINEERING by M. RITHU (1CR16IS049) SACHIN BASNET THAPA (1CR16IS087) Under the guidance of Mr. Prasad B.S Assistant Professor Dept. of ISE, CMRIT, Bengaluru

3. IPL Score Prediction System using Machine Learning P Vishnu Teertha, Ms. Bhavani R Student, Department of MCA, Visvesvaraya Technological University, The National Institute of Engineering, Mysuru, India Assistant Professor, Department of MCA, Visvesvaraya Technological University, The National Institute of Engineering, Mysuru, India