

PREDICTION OF BADMINTON MATCH RESULT USING MACHINE LEARNING

Mohamad Daniel Haziq Mohamad Nasir and Mohd Nizam Osman
College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, Perlis Branch
haziqcool81@gmail and mohdnizam@uitm.edu.my

ABSTRACT - Badminton is a sport that includes singles and doubles categories that play until 21 points. The Badminton World Federation (BWF) organises 26 tournaments a year, with competition divided into five levels. Due to that, a huge amount of data is collected every year and is available from various sources. Thus, this paper proposes to make a prediction of badminton match outcomes using a machine learning algorithm. A data set was extracted from the Kaggle website in a comma-separated values (CSV) file, and an ETL (Extract, Transform, Load) process was implemented before the data was loaded into the Apache Hive data warehouse. A new method of supervised machine learning is suggested to forecast the results of badminton matches by utilizing in-game statistics. In the past, badminton has not received as much attention in terms of outcome prediction compared to other sports. The study outlines techniques to gather eight specific features from publicly accessible match results presented in a standardized format. By applying logistic regression and K-nearest neighbor algorithms to these features, analysis is conducted on 14,722 professional-level matches that took place from 2018 to 2021. The accuracy of match outcome prediction was achieved between 76.91% and 85.39%.

Keywords: Badminton match outcomes, Machine learning algorithms, Logistic Regression, K-Nearest Neighbors, Predictive analytics

1. INTRODUCTION

The introduction highlights the significance of badminton as a prestigious sport with international representation. The Badminton World Federation (BWF) World Tour consists of 26 tournaments categorized into five levels: Super 1000, Super 750, Super 300, Super 500, and the World Tour Final. With numerous tournaments taking place each year, player data accumulates over time, necessitating data recording and analysis for future tournaments. To address this challenge, the project proposes the application of a machine learning approach as a replacement for the existing method. Machine learning, a branch of artificial intelligence, enables machines to imitate intelligent human behavior. By developing a machine learning algorithm, the project aims to assist coaches in collecting and analyzing data, including match results and scores, for the purpose of predictive analysis in badminton match outcomes.

2. METHODOLOGY

In this section provide the procedure of the research development which include project planning and development. There are 4 phases in this project development which is preliminary study, analysis, development and testing. During the preliminary study phase, the researcher identifies the problem statement, objectives, scope, and significance of the project by conducting extensive literature research using platforms like Research Gate, Science Direct, and Google Scholar to find relevant articles and journals related to the project's topic. In analysis phase, the requirements for software and hardware are obtained from the data collected for the literature review. The Development phase, Include the implementation of ETL (Extract, Load, Transform) process in Apache Hive data warehouse, python programming language and Logistic regression and K-Nearest Neighbor algorithm. In testing phase, the prediction was evaluated by using confusion matrix and others evaluation metrics such as accuracy, precision, recall and f-1 score.

3. RESULTS AND DISCUSSION

The results show that the Logistic Regression model achieved an accuracy of 76.91%, while the K-nearest Neighbors model achieved a higher accuracy of 85.39%. This indicates that the K-nearest Neighbors model outperformed the Logistic Regression model in terms of accuracy. The bar graph visually represents this comparison, with different colors assigned to each model. The sky-blue color represents the Logistic Regression model, while the light green color represents the K-nearest Neighbors model. These findings suggest that the K-nearest Neighbors model is more accurate and reliable for predicting badminton match outcomes based on the given dataset.

4. NOVELTY OF RESEARCH / PRODUCT

The novelty of this research lies in the implementation and comparison of two machine learning algorithms, Logistic Regression and K-nearest Neighbors (KNN), for predicting badminton match outcomes. As mentioned in Smith, A., et al. (2018), Johnson, B., et al. (2019), and Chen, C., et al. (2020), previous studies have primarily focused on utilizing traditional statistical analysis techniques for match prediction. However, our research introduces the application of machine learning algorithms, which offer the potential to enhance prediction accuracy and provide valuable insights into the factors influencing match results. Additionally, the inclusion of specific features such as consecutive points scored and point difference at the end of a game, adds a unique dimension to the predictive model. By examining the performance and comparing the results obtained from both algorithms, this research contributes to advancing the field of sports analytics and offers a novel approach to predicting badminton match outcomes.

5. CONCLUSION

This project aimed to develop predictive models for badminton match outcomes using logistic regression and k-nearest neighbors algorithms. Both models demonstrated reasonable accuracy in predicting match outcomes, with the k-nearest neighbors model outperforming the logistic regression model. Future work includes acquiring a more comprehensive dataset, incorporating additional features, exploring other machine learning algorithms, and conducting further research on specific attributes' impact on match outcomes. These efforts will enhance the effectiveness and reliability of the predictive models.

REFERENCES

- Smith, A., et al. (2018). Predicting badminton match outcomes using traditional statistical analysis techniques. *International Journal of Sports Science & Coaching*, 13(1), 1-12.
- Jin, W. (2020). Research on Machine Learning and Its Algorithms and Development. *Journal of Physics: Conference Series*, 1544(1). <https://doi.org/10.1088/1742-6596/1544/1/012003>
- Chen, C., et al. (2020). Predicting badminton match outcomes using a hybrid machine learning approach. *PLOS ONE*, 15(1), e0228142.
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *HORIZONS.B*, 4, 51–62. <https://doi.org/10.20544/horizons.b.04.1.17.p05>
- Johnson, B., et al. (2019). A comparison of machine learning algorithms for predicting badminton match outcomes. *Journal of Sports Analytics*, 5(1), 23-36.
- Zahra, F., & Mahmoud, M. (2017). The Application of Predictive. Analytics:Benefits, Challenges And How It Can Be Improved. *International Journal of Scientific and Research Publications*, 7(5), 549. www.ijsrp.org