Using Support Vector Regression Kernel Models for Cricket Performance Prediction in the Women's Premier League 2024

Ponnusamy Yoga Lakshmi¹ABCD, Swamynathan Sanjaykumar¹BCD, Maniazhagu Dharuman²CD and Aarthi Elangovan¹CD

¹SRM Institute of Science and Technology
²Central University of Tamil Nadu

Authors' Contribution: A – Study design; B – Data collection; C – Statistical analysis; D – Manuscript Preparation; E – Funds Collection

Corresponding Author: Swamynathan Sanjaykumar, E-mail: sanjayswaminathan007@gmail.com

Abstract

Background. The interest in women's premier league cricket has caused the need for advanced analytics to understand the multifaceted dynamics of the sport.

Study Purpose. This study aimed to contribute to sports analytics by assessing the efficacy of Support Vector Regression (SVR) kernel models in predicting the most valuable player. Such research methods as ANOVA, Bessel function, and Inverse MultiQuadratic kernel application have been deliberately chosen for their diverse mathematical approaches, aligning with the nuanced intricacies of women's premier league cricket.

Materials and methods. Player performance was analyzed by using the following study methods: ANOVA, Bessel function and Inverse MultiQuadratic kernel application. The data, sourced from espncricinfo.com and the International Cricket Council, includes essential metrics for five teams. Rigorous preprocessing techniques, such as imputation and outlier removal, enhance data reliability, ensuring robust predictive models.

Results. The application of the Inverse MultiQuadratic kernel exhibits exceptional predictive performance, surpassing ANOVA and Bessel function models. The kernel's radial basis function proves effective in capturing the intricate dynamics of women’s premier league cricket. The findings underscore the suitability of kernel method for predicting standout performers in the Women's Premier League 2024 season.

Conclusions. The study revealed the dynamic interplay between sports analytics and machine learning in women's premier league cricket. The application of the Inverse MultiQuadratic kernel stands out as the most effective model, providing key insights into player predictions. This emphasizes the continual integration of advanced analytical techniques to enhance our understanding of the evolving landscape of women's premier league cricket. As the sport gains prominence on the global stage, such analytical endeavors become imperative for strategic decision-making and sustained growth.

Keywords: women's cricket, support vector regression, machine learning, comparative analysis, performance prediction.

Introduction

The realm of women's premier league cricket has been experiencing unprecedented growth, attracting heightened attention from enthusiasts and stakeholders alike. In response to this surge, the significance of employing advanced predictive modeling techniques cannot be overstated (Kapadia et al., 2022). This research endeavors to explore the application of Support Vector Regression (SVR) with distinct kernel models, namely ANOVA, Bessel function, and Inverse MultiQuadratic, to predict the player's performance of women's premier league cricket in the upcoming 2024 season (Sumathi et al., 2023; Aburas et al., 2018). Cricket, as a sport, is influenced by a multitude of factors, ranging from individual player form to team dynamics and external conditions. Leveraging sophisticated analytics and machine learning methodologies holds the potential to unravel the intricate relationships within these factors, surpassing...
The Support Vector perspective to the analysis (Hudnurkar & Rayavarapu, 2022). Function, known for its oscillatory behavior, brings a unique capture nuanced variations in player’s performances. Bessel Inverse MultiQuadratic represent diverse mathematical selected kernels for this study, ANOVA, Bessel function, and most valuable player in women’s premier league cricket 2024 of distinct SVR kernel models concerning predicting the nuanced understanding of the strengths and weaknesses. Through our comparative analysis, we aim to provide a layer of sensitivity to the SVR model (Gu et al., 2023; Saikia, 2020). The Inverse MultiQuadratic kernel introduces a radial basis function, offering a unique perspective on the relationship between input features. By incorporating the inverse of the Euclidean distance between data points, this kernel underscores the significance of proximity in predictive modeling. In the domain of women’s premier league cricket, where team dynamics and interactions significantly impact performance, the kernel’s focus on proximity adds a layer of sensitivity to the SVR model (Gu et al., 2023; Wickramasinghe, 2014).

This research’s intentional focus on women’s premier league cricket underscores the importance of gender-inclusive sports research (Subburaj et al., 2023). As women’s cricket continues to ascend on the global stage, understanding and predicting team dynamics become imperative for various stakeholders (Passi and Pandey, 2018b). These include team management making strategic decisions, sponsors assessing investment opportunities, and fans engaging in the sport. Through our comparative analysis, we aim to provide a nuanced understanding of the strengths and weaknesses of distinct SVR kernel models concerning predicting the most valuable player in women’s premier league cricket 2024 (Bunker and Thabtah, 2019; Wickramasinghe, 2014). The selected kernels for this study, ANOVA, Bessel function, and Inverse MultiQuadratic, represent diverse mathematical approaches. ANOVA, emphasizing variance analysis, aims to capture nuanced variations in player’s performances. Bessel function, known for its oscillatory behavior, brings a unique perspective to the analysis (Hudnurkar & Rayavarapu, 2022). The Inverse MultiQuadratic kernel introduces a radial basis function that may enhance predictive capabilities. The objective in comparing these kernels is to identify the model that best aligns with the complex dynamics inherent in women’s premier league cricket (Van Eetvelde et al., 2021; Anam et al., 2021).

The purpose of the research. The Support Vector Regression (SVR) with ANOVA, Bessel function, and Inverse MultiQuadratic kernels to predict player performance in the Women’s Premier League Cricket 2024 season.

Materials and Methods

Study Participants

The dataset is meticulously curated from two reliable sources: espncricinfo.com and the official website of the International Cricket Council (ICC). The dataset metrics for five teams and a total of 90 players participating in the WPL, including the number of matches played, batting averages, bowling averages, catches taken, and the T20 status of the 18 players in each squad, as shown in Figure 1 (International Cricket Council; ESPNcricinfo; Women’s Premier League).

Study organization

This investigation aims to predict the most valuable player of the upcoming Women’s Premier League (WPL) 2024 season, focusing on essential performance metrics using three different SVR kernel models such as ANOVA, Bessel function, and Inverse MultiQuadratic. Data preprocessing to address missing values is fundamental to the reliability of predictive models. Robust imputation methods, such as mean or median imputation, are employed to fill in missing values. Additionally, data integrity checks are performed to identify and rectify any anomalies in the dataset. Identifying and removing outliers is critical for refining the dataset. Outliers can distort model training, and their removal enhances the robustness and generalization capabilities of the SVR models (Subburaj et al., 2023).

The data splitting adheres to a well-established principle, allocating 80% of the acquired dataset for training the Support Vector Regression (SVR) models, while reserving the remaining 20% for rigorous testing and evaluation. This meticulous division ensures a robust and comprehensive training phase, allowing the SVR models to glean intricate patterns and relationships from the majority of the data. Subsequently, the earmarked 20% functions as an independent dataset, remaining untouched during the training process, thereby facilitating an unbiased assessment of the models’ predictive capabilities (Kapadia et al., 2022; Men,
2022). The training process involved the application of three distinct kernels - ANOVA Kernel, Bessel Function Kernel and Inverse MultiQuadratic Kernel. Parameters were fine-tuned through methods like grid search, ensuring optimal model performance. Testing involved making predictions on the independent dataset, and the models’ effectiveness was evaluated using metrics such as Mean Squared Error and R-squared (Sanjaykumar et al., 2023). The ultimate aim is to contribute to the creation of reliable and precise predictive models capable of accurately forecasting the performance of women’s premier league cricket players in the anticipated 2024 season.

**Statistical analysis**

Mean Squared Error (MSE) calculates the average squared difference between predicted (Apredict) and actual (Aactual) values, while Root Mean Squared Error (RMSE) is its square root. Mean Absolute Error (MAE) measures the average absolute difference. R-Squared (R²) evaluates predictive accuracy by comparing actual and predicted values to the mean.

- Mean Squared Error (MSE) = \[ \frac{1}{n} \sum (A_{\text{predict}} - A_{\text{actual}})^2 \]
- Root Mean Squared Error (RMSE) = \[ \sqrt{\text{MSE}} \]
- Mean Absolute Error (MAE) = \[ \frac{1}{n} \sum |A_{\text{predict}} - A_{\text{actual}}| \]
- R-Squared (R²) = \[ 1 - \frac{\sum (A_{\text{actual}} - \text{mean})^2}{\sum (A_{\text{predict}} - \text{mean})^2} \]

In the given equation, Aactual signifies observed performance values, Apredict represents predicted performance, Amean stands for the mean of actual performance values, and x denotes the number of data points (Sanjaykumar et al., 2023; Rajendiran and Rethnaraj, 2023).

**Results**

The research study unfold the outcomes on predicting the most valuable player in the upcoming Women’s Premier League (WPL) 2024 season. The research focus lies on leveraging Support Vector Regression (SVR) with distinctive kernel models – ANOVA, Bessel function, and Inverse MultiQuadratic. This segment serves as an exclusive platform to present our analyses and engage in a comprehensive discussion, dissecting the performance of each SVR kernel model in forecasting women’s premier league cricket.

**Comparative analysis**

The Comparative Analysis section unfolds a detailed assessment of the diverse Support Vector Regression (SVR) kernel types ANOVA, Bessel function, and Inverse MultiQuadratic based on essential metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and R-Squared (R²) Table 1. These metrics serve as pivotal indicators in gauging the precision and efficacy of each kernel in predicting the most valuable player for the Women’s Premier League 2024 season. The ensuing discourse meticulously explores the distinctive characteristics of each kernel, presenting unique insights into their individual strengths and areas where refinement may enhance predictive capabilities.

The Support Vector Regression (SVR) kernel types ANOVA, Bessel function, and Inverse MultiQuadratic reveals the distinct performance metrics crucial for assessing their predictive capabilities in the context of the Women’s Premier League 2024 season. In analyzing the Mean Squared Error (MSE), the average squared disparity between predicted and actual values, it emerges as the most proficient

Table 1. Performance metrics of support vector regression with different kernels

<table>
<thead>
<tr>
<th>Support Vector Regression – Kernel Type</th>
<th>Mean Squared Error (MSE)</th>
<th>Root Mean Squared Error (RMSE)</th>
<th>Mean Absolute Error (MAE)</th>
<th>R Squared (R²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>20.1357</td>
<td>4.4872</td>
<td>8.0684</td>
<td>0.79386</td>
</tr>
<tr>
<td>Bessel function</td>
<td>18.2389</td>
<td>4.2707</td>
<td>7.1290</td>
<td>0.81563</td>
</tr>
<tr>
<td>Inverse MultiQuadratic</td>
<td>13.6294</td>
<td>3.6918</td>
<td>2.5639</td>
<td>0.86392</td>
</tr>
</tbody>
</table>

The predictions based on the Inverse MultiQuadratic kernel for the most valuable player in the upcoming Women's Premier League (WPL) season 2024 reveal intriguing insights into the potential standout performers. Hayley Matthews emerges as a front-runner with a substantial prediction percentage of 84.1%, indicating a high likelihood of her securing the coveted title. Following closely are players like Natalie Sciver-Brunt (83.9%), Marizanne Kapp (80.4%), and Meg Lanning (81.8%), showcasing strong predictions for their significant impact on the league. Notably, the kernel anticipates consistent performances from key players such as Ellyse Perry (78.9%), Deepthi Sharma (79.7%), and Smriti Mandhana (74.2%) (Figure 6).

Fig. 4. Accuracy analysis of support vector regression with different kernels

Performance Prediction Based on Inverse MultiQuadratic kernel.

The implementation of the Inverse MultiQuadratic kernel, which stands out for its radial basis function, displays an impressive predictive accuracy of 86.4% (Figure 4). This outperforms alternatives like ANOVA and the Bessel function, emphasizing its proficiency in identifying intricate patterns within the dataset, leading to heightened precision in performance predictions.

Fig. 5. Visualizing the performance prediction graph of the Inverse MultiQuadratic kernel model, the predicted values (donated in red) and actual value (represented in blue color) with slight deviation

Discussion

We looked at how well Support Vector Regression (SVR) kernel models ANOVA, Bessel function, and Inverse MultiQuadratic worked at predicting who would be the most valuable player in the 2024 Women's Premier League (WPL) season. Notably, the kernel exhibits superior predictive accuracy, boasting the lowest mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), alongside the highest R-squared (R²) (Table 1) (Kaur et al., 2021; Xu et al., 2023).

The predictions based on the Inverse MultiQuadratic kernel for the most valuable player in the Women's Premier League (WPL) season 2024 reveal intriguing insights into the potential standout performers. Hayley Matthews emerges as a front-runner with a substantial prediction percentage of 84.1%, indicating a high likelihood of her securing the coveted title. Following closely are players like Natalie Sciver-Brunt (83.9%), Marizanne Kapp (80.4%), and Meg Lanning (81.8%), showcasing strong predictions for their significant impact on the league. Notably, the kernel anticipates consistent performances from key players such as Ellyse Perry (78.9%), Deepthi Sharma (79.7%), and Smriti Mandhana (74.2%) (Figure 6).
based on this kernel, offer compelling insights. With an 84.1% prediction percentage, Hayley Matthews pulls ahead of Natalie Sciver-Brunt, Marizanne Kapp, and Meg Lanning as the leading contender. Additionally, consistent performances are anticipated from key players such as Ellyse Perry, Deepti Sharma, and Smriti Mandhana, providing valuable information for various stakeholders. The Inverse MultiQuadratic kernel proves to be a standout choice for predicting the most valuable player in the WPL 2024 season. Its efficacy in capturing the dynamics of women's premier league cricket positions it as a valuable tool for stakeholders, aiding in strategic decision-making and enhancing overall understanding of player performances in this evolving sporting landscape (Mandoli et al., 2021; Passi and Pandey, 2018c).

While this research study provides valuable insights, there are opportunities for future exploration and refinement of predictive modeling in women's premier league cricket (Awan et al., 2021). The inclusion of additional features, such as player fitness, team strategies, and external factors like weather conditions, could enhance predictive models. Exploring ensemble methods and hybrid models may further amplify predictive accuracy. Ongoing advancements in data analytics and machine learning techniques call for continuous adaptation and refinement of models. Future research could extend to other cricket leagues or sports, expanding the applicability of predictive analytics in women's sports (Kruglov and Khudolidi, 2022: Bullock et al., 2022). In essence, this research lays the groundwork for a dynamic interplay between sports analytics and machine learning in women's premier league cricket. As the sporting landscape evolves, predictive modeling becomes an indispensable tool, offering strategic insights and fostering a deeper understanding of the intricate dynamics influencing player performances (Aburas et al., 2018). Through ongoing research and refinement, the field of sports analytics is poised for continuous growth and contribution to the expanding domain of women's cricket (Simsek et al., 2021).

Conclusions

The Inverse MultiQuadratic kernel emerges as the top performer, displaying superior results across key metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R2). Its unique radial basis function and emphasis on proximity contribute to its adeptness in capturing nuanced patterns within the dataset. The predictions based on this kernel for the most valuable player in the upcoming Women's Premier League (WPL) season 2024 reveal intriguing insights into the potential standout performers. Hayley Matthews emerges as a front-runner with a substantial prediction percentage of 84.1%, indicating a high likelihood of her securing the coveted title.

Acknowledgement

We extend our sincere appreciation to the International Cricket Council, ESPNcricinfo, and the official website of the Women's Premier League for graciously providing indispensable data and resources essential for the successful completion of this study.

Conflict of interest

The authors declare that there is no conflict of interest.

References


Застосування ядерних моделей опорно-векторної регресії для прогнозування результатів гри в крикет у жіночій прем’єр-лізі 2024 року

Поннусамі Йога Лакшмі1ABCD, Свамінатан Санджайкумар1ABCD, Маніажагу Дхаруман2CD, Арті Елангован3CD

1SRM Інститут науки та технологій
2Центральний університет Таміл Наду

Авторський вклад: А – дизайн дослідження; В – збір даних; С – статаналіз; Д – підготовка рукопису; Е – збір коштів

Реферат. Стаття: 7 с., 1 табл., 6 рис., 29 джерел.

Історія питання. Інтерес до жіночої прем’єр-лізі з крикету викликав потребу в застосуванні розширеної аналітики для розуміння багатогранної динаміки даного виду спорту.

Мета дослідження. Метою дослідження є внесок у спортивну аналітику шляхом оцінки ефективності застосування ядерних моделей опорно-векторної регресії (ОВР) у прогнозуванні визначення найбільш результативного гравця. Такі
Методи дослідження, як дисперсійний аналіз (ANOVA), функція Бесселя та застосування звортінх мультиквадратичних ядер, були свідомо обрані через їхні різноманітні математичні підходи, що відповідають тонкістям гри в крикет у жіночій прем’єр-лізі.

Матеріали та методи. Результативність гравців було проаналізовано за допомогою наступних методів дослідження: дисперсійний аналіз (ANOVA), функції Бесселя та застосування звортінх мультиквадратичних ядер. Дані, отримані з сайту espncricinfo.com та Міжнародної ради крикету, включають основні показники для п’яти команд. Ретельні методи попереднього опрацювання, такі як імпутація та виключення відхилень, підвищують достовірність даних, забезпечуючи отримання надійних прогнозних моделей.

Результати. Застосування звортінх мультиквадратичного ядра демонструє виняткову прогнозну ефективність, перевершує моделі дисперсійного аналізу (ANOVA) та функції Бесселя. Радіально-базисна функція ядра ефективно відображає складну динаміку жіночої прем’єр-ліги з крикету. Отримані результати підкреслюють доцільність застосування ядрового методу з метою прогнозування найбільш результативних гравців у сезоні Жіночої прем’єр-ліги 2024 року.

Висновки. В результаті дослідження було встановлено динамічну взаємодію між спортивною аналітикою та машинним навчанням у жіночій прем’єр-лізі з крикету. Застосування звортінх мультиквадратичного ядра показало найбільшу ефективність, надаючи ключову інформацію щодо прогнозування результатів гравців. Це підкреслює постійну інтеграцію сучасних аналітичних методів для покращення нашого розуміння динаміки розвитку жіночої прем’єр-ліги з крикету. Оскільки цей вид спорту набуває все більшої популярності на світовій арені, такі аналітичні дослідження стають необхідними для прийняття стратегічних рішень та сталого розвитку.

Ключові слова: жіночий крикет, опорно-векторна регресія, машинне навчання, порівняльний аналіз, прогнозування результатів.

Information about the authors:

Lakshmi, Ponnusamy Yoga: yogalkp@srmist.edu.in; https://orcid.org/0000-0002-3461-8226; Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Tamil Nadu 603203, India.
Sanjaykumar, Swamynathan: sanjayswaminathan007@gmail.com; https://orcid.org/0000-0001-9945-2223; Department of Physical Education and Sports Sciences, Faculty of Science and Humanities, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Tamil Nadu 603203, India.
Dharuman, Maniazhagu: maniazhagu@cutn.ac.in; https://orcid.org/0000-0001-7937-8724; Department of Physical Education and Sports, Central University of Tamil Nadu, CUTN Bridge, Neelakudy, Tamil Nadu 610005, India.
Elangovan, Aarthi: aarthi.devpal@gmail.com; https://orcid.org/0000-0002-0324-4997; Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Tamil Nadu 603203, India.


Received: 06.01.2024. Accepted: 12.02.2024. Published: 29.02.2024

This work is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0).