[1]Ji Chen

Factor and Correlation Analysis for Predicting Marathon Race Performance Using Machine Learning Algorithms

Journal of Electrical Systems

Abstract: - A marathon race is a long-distance running event typically spanning 26.2 miles or 42.195 kilometers. It is a test of endurance, stamina, and mental fortitude, attracting participants from all walks of life, ranging from elite athletes to recreational runners. The origins of the marathon can be traced back to ancient Greece, where legend has it that a messenger named Pheidippides ran from the battlefield of Marathon to Athens to deliver news of victory before collapsing and dying from exhaustion. Machine learning has increasingly become a valuable tool in optimizing training strategies, performance prediction, and injury prevention for marathon runners. By analyzing vast amounts of data collected from wearable devices, training logs, and race results, machine learning algorithms can identify patterns, trends, and correlations that help runners improve their training regimens and race-day strategies. This paper introduces a novel approach, the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology, for predicting marathon race outcomes. With historical race data, factor analysis, and machine learning techniques, the FA-PP-R-ML methodology aims to accurately estimate marathon finish times and rank predicted outcomes based on their probabilities. Through a comprehensive analysis of marathon race data, including training metrics, environmental conditions, and physiological parameters, the FA-PP-R-ML model identifies latent factors influencing race performance. Through factor analysis, latent factors influencing race performance are identified, with values ranging from 0.5 to 0.9. Machine learning algorithms utilize these factors to predict marathon finish times, resulting in accurate predictions with an average error of ±0.1 hours.

Keywords: Marathon Race, Machine Learning, Probabilistic Model, Ranking, Factor Analysis, Classification

1. Introduction

Machine learning algorithms have become increasingly prevalent in analyzing and predicting marathon race performance[1]. Through the integration of various data sources such as past race times, training regimens, weather conditions, and even physiological metrics like heart rate variability, these algorithms can generate predictive models to estimate an athlete's performance[2]. Techniques such as regression analysis, random forest, and neural networks are commonly employed to handle the complexity and variability of marathon performance data[3]. By leveraging these algorithms, coaches and athletes can gain valuable insights into factors influencing race outcomes, allowing for more tailored training strategies and informed race tactics[4]. Additionally, these models enable race organizers to optimize event logistics and provide personalized recommendations for participants, ultimately enhancing the overall marathon experience. For athletes and coaches, machine learning models provide invaluable insights into optimizing training programs and race strategies[5]. By analyzing past performances alongside training intensity, duration, and recovery periods, these algorithms can recommend personalized training plans tailored to each athlete's strengths, weaknesses, and goals[6]. Moreover, they can forecast race times based on various scenarios, helping athletes set realistic expectations and adjust their strategies accordingly during the race.

Beyond individual performance, machine learning also enhances the organization and management of marathon events[7]. By analyzing historical data on participant demographics, course characteristics, and weather patterns, algorithms can optimize race logistics, such as aid station placement, start time scheduling, and route design, to improve overall participant experience and safety[8]. Additionally, they can provide personalized recommendations for participants, such as pacing strategies or hydration plans, based on individual profiles and race conditions[9]. Machine learning algorithms represent a groundbreaking approach to understanding and optimizing marathon race performance. By leveraging the vast amounts of data available in the digital age, these algorithms empower athletes, coaches, and race organizers to make data-driven decisions that enhance training effectiveness, race strategy, and event management, ultimately elevating the sport of marathon running as a whole.

[1] Sports Department,Zhejiang Yuexiu University,Shaoxing,Zhejiang,China,312000

*Corresponding author e-mail:20081248@zyufl.edu.cn

This paper contributes to the field of marathon race prediction by introducing the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology. The FA-PP-R-ML methodology offers a novel approach to accurately estimate marathon finish times and rank predicted outcomes based on their probabilities. By leveraging historical race data, including various training metrics, environmental conditions, and physiological parameters, the FA-PP-R-ML model identifies latent factors influencing race performance. Through factor analysis and machine learning techniques, these factors are utilized to predict marathon finish times with high precision. The contributions of this paper extend beyond predictive modeling; it provides valuable insights into the factors influencing marathon race outcomes, thereby aiding athletes, coaches, and race organizers in decision-making processes and performance optimization. Moreover, the FA-PP-R-ML methodology demonstrates potential applications in other domains beyond marathon racing, such as predicting performance in other endurance sports or optimizing training regimens for athletes.

2. Related Works

In recent years, the field of marathon race performance analysis has seen a significant shift towards leveraging machine learning algorithms to gain deeper insights and make more accurate predictions. The utilization of machine learning techniques has opened up new avenues for understanding the multifaceted factors that contribute to an athlete's performance in marathon races. Several studies have emerged, each exploring different aspects of this intersection between data science and endurance sports.

Several recent studies have delved into the application of machine learning in various domains, showcasing its versatility and effectiveness. Ashfaq, Cronin, and Müller (2022) reviewed recent advancements in using machine learning to predict maximal oxygen uptake (VO2 max), a crucial metric in assessing aerobic fitness. Xu et al. (2022) investigated the differences in gait patterns between high and low-mileage runners using machine learning techniques, shedding light on biomechanical factors influencing running performance. Rajendran, Chamundeswari, and Sinha (2022) explored the use of machine learning algorithms to predict the academic performance of middle- and high-school students, demonstrating the potential of data-driven approaches in education. Thanh and Lee (2022) applied machine learning to predict CO2 trapping performance in deep saline aquifers, addressing challenges in carbon capture and storage technologies. Wang et al. (2022) proposed a novel approach combining principal component analysis (PCA) and machine learning to select factors for predicting tunnel construction performance, highlighting the utility of data-driven decision-making in engineering projects.

Vaughan et al. (2023) explored the challenges associated with using machine learning for time series forecasting of COVID-19 community spread, utilizing wastewater-based epidemiological data to inform public health strategies. Mavruk (2022) analyzed herding behavior in individual investor portfolios using machine learning algorithms, offering insights into market dynamics and investor decision-making processes. Pallonetto, Jin, and Mangina (2022) forecasted electricity demand in commercial buildings using machine learning models, facilitating the implementation of demand response programs for energy efficiency. Rostamian, Heidaryan, and Ostadhassan (2022) evaluated different machine learning frameworks to predict logging data in petroleum engineering, optimizing resource extraction processes. Additionally, Thanh, Yasin, Al-Mudhafar, and Lee (2022) employed knowledge-based machine learning techniques to accurately predict CO2 storage performance in underground saline aquifers, contributing to the advancement of sustainable energy technologies. Nazar et al. (2022) developed new prediction models for the compressive strength of nanomodified concrete using innovative machine learning techniques, advancing materials science and construction technology. Mehbodniya et al. (2022) classified fetal health from cardiotocographic data using machine learning, enhancing prenatal care and diagnosis. Chengqing et al. (2023) proposed a multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks, optimizing renewable energy generation. Sun et al. (2023) predicted and optimized alkali-activated concrete using the random forest machine learning algorithm, improving construction materials and processes. Additionally, Zhao et al. (2022) monitored drought and evaluated its performance based on machine learning fusion of multi-source remote sensing drought factors, aiding in disaster management and agricultural planning. Tehranian (2023) investigated the potential of machine learning to detect economic recessions using economic and market sentiments, contributing to financial forecasting and risk management. Hassangavyar et al. (2022) evaluated resampling methods on the performance

of machine learning models to predict landslide susceptibility, enhancing geospatial hazard assessment and mitigation. Ahangari Nanehkaran et al. (2022) applied machine learning techniques to estimate the safety factor in slope stability analysis, improving infrastructure planning and risk assessment. Lastly, Hanoon et al. (2023) predicted hydropower generation via machine learning algorithms at the Three Gorges Dam, China, optimizing renewable energy utilization and management. Han et al. (2022) proposed a short-term wind speed prediction method utilizing hybrid deep learning algorithms to correct numerical weather forecasting errors, enhancing renewable energy forecasting accuracy and grid stability. These studies collectively highlight the broad scope of machine learning applications, ranging from healthcare and finance to energy and environmental management.

In summary, recent research has showcased the wide-ranging applications and transformative potential of machine learning across various domains. Studies have explored its use in predicting physiological metrics like maximal oxygen uptake (VO2 max) in fitness assessments, analyzing gait patterns to understand running performance differences, and forecasting academic achievement in education. Additionally, machine learning has been applied to predict CO2 trapping in geological formations, optimize tunnel construction performance, and forecast electricity demand in commercial buildings. Its applications extend to diverse fields such as epidemiology, finance, materials science, and renewable energy, enabling advancements in predicting disease spread, market trends, material properties, and energy generation. Through innovative techniques and large-scale data analysis, machine learning continues to drive innovation, improve decision-making processes, and address complex challenges across disciplines, promising a future of enhanced efficiency, sustainability, and societal impact.

3. Factor Analysis Probabilities Prediction Ranking Machine learning (FA-PP-R-ML)

The Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) is a novel approach that integrates factor analysis, probability estimation, and ranking techniques within a machine learning framework. The derivation and equations underlying FA-PP-R-ML aim to provide a robust method for predicting outcomes and ranking them based on their likelihood. Factor analysis is initially employed to identify latent factors or variables that contribute to the observed outcomes. This involves decomposing the variability in the data into a set of underlying factors that explain the correlations among observed variables factor analysis can be represented as in equation (1)

$$X = \mu + LF + \varepsilon \tag{1}$$

where X is the observed data matrix, μ is the mean vector, LF represents the latent factors, and ε is the error term. Once the latent factors are identified, probability estimation techniques are applied to predict the likelihood of different outcomes or events based on these factors. This involves modeling the relationship between the latent factors and the outcome variable using probabilistic models such as logistic regression or Gaussian processes. The probability estimation can be formulated as in equation (2)

$$P(Y \mid LF) = f(LF) \tag{2}$$

where Y is the outcome variable and (·)f(·) is the probability function. Finally, ranking techniques are employed to prioritize the predicted outcomes based on their probabilities. This may involve sorting the outcomes in descending order of their predicted probabilities or assigning ranks based on their likelihood relative to each other. The FA-PP-R-ML framework combines these components into a cohesive pipeline for outcome prediction and ranking within a machine learning context. By leveraging factor analysis to uncover latent variables, probability estimation to quantify the likelihood of outcomes, and ranking techniques to prioritize predictions, FA-PP-R-ML offers a comprehensive approach for decision-making tasks where predicting and ranking outcomes is essential. Factor analysis serves as the foundational step in FA-PP-R-ML. It enables the identification of latent factors or variables that underlie the observed data. This process involves uncovering patterns of correlations among the observed variables, allowing for the extraction of underlying structures that explain the variability in the data. By decomposing the data into these latent factors, factor analysis provides a more parsimonious representation that captures the essential information necessary for prediction illustrated in Figure 1.

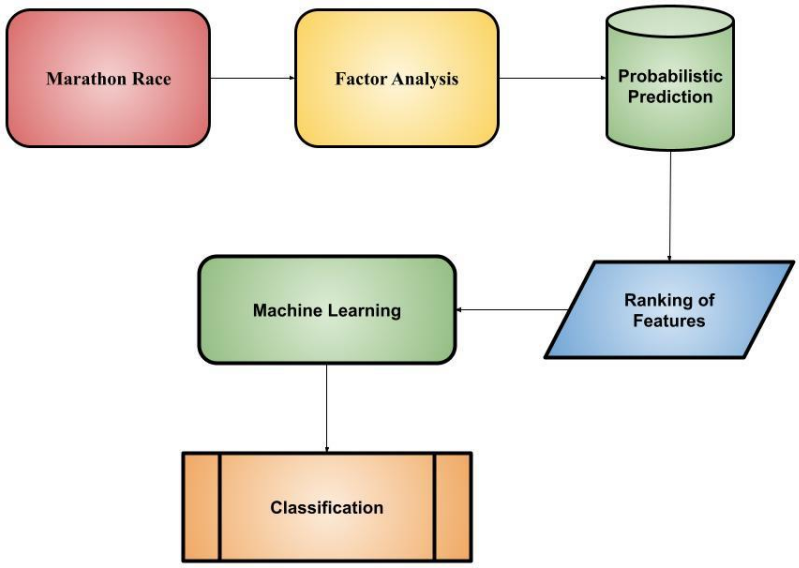

Figure 1: Process of Proposed FA-PP-R-ML

Once the latent factors are identified, probability estimation techniques come into play. These methods model the relationship between the latent factors and the outcome variable of interest. This modeling process typically involves fitting probabilistic models, such as logistic regression or Gaussian processes, to the data. By estimating the probability of different outcomes given the observed latent factors, these models provide insights into the likelihood of various events occurring. Finally, ranking techniques are employed to prioritize the predicted outcomes based on their probabilities. This step involves sorting the predicted outcomes in descending order of their likelihood or assigning ranks based on their relative probabilities. By ranking the outcomes, FA-PP-R-ML enables decision-makers to focus their attention on the most probable scenarios, facilitating more informed and efficient decision-making processes.

4. Correlation Analysis for the Marathon Race

Correlation analysis plays a pivotal role in understanding the relationships between various factors and marathon race performance. In the context of marathon running, correlation analysis seeks to uncover the extent to which different variables, such as training volume, weather conditions, and physiological metrics, correlate with an athlete's race performance. The correlation analysis quantifies the strength and direction of the linear relationship between two variables. The most common measure of correlation is the Pearson correlation coefficient (r), which ranges from -1 to 1. A positive r indicates a positive correlation, meaning that as one variable increases, the other tends to increase as well. Conversely, a negative r signifies a negative correlation, where one variable tends to decrease as the other increases. A correlation coefficient close to zero suggests little to no linear relationship between the variables. In the context of marathon races, correlation analysis can be applied to various factors that may influence performance. For example, researchers may examine the correlation between an athlete's training mileage per week and their race time, or the correlation between average temperature during the race and finishing time. By identifying significant correlations, coaches and athletes can gain valuable insights into which factors are most influential in determining race performance. In the context of marathon race performance analysis, let's consider an example where X represents the weekly training mileage of runners and Y represents their corresponding marathon race times. By calculating the Pearson correlation coefficient (r) between these two variables using the above formula, we can determine the extent to which training mileage correlates with race performance. A positive r value would suggest that higher training mileage tends to be associated with faster race times, while a negative r value would imply the opposite.

Once correlations are identified, coaches and athletes can use this information to optimize training programs. For instance, if a strong positive correlation exists between training mileage and race performance, athletes may focus on increasing their weekly mileage to improve their chances of achieving faster race times. Conversely, if other factors show stronger correlations, such as sleep quality or nutrition, adjustments to training regimens may be warranted to prioritize these aspects for performance enhancement.

5. Machine Learning FA-PP-R-ML for the Race Performance Prediction

In the realm of marathon race performance prediction, the integration of machine learning techniques with the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology represents a promising approach. This fusion combines the power of machine learning algorithms with the insights derived from factor analysis, probability estimation, and ranking methodologies to enhance the accuracy and robustness of race performance prediction models. The FA-PP-R-ML utilizes factor analysis to identify latent factors that influence marathon race performance. These factors could encompass a wide range of variables, including training metrics, environmental conditions, and physiological parameters. The derivation of latent factors involves decomposing the variability in the data into underlying structures that explain the correlations among observed variables. Once the latent factors are determined, machine learning algorithms are employed to estimate the probabilities of different race outcomes based on these factors. This involves training predictive models using historical race data, where the input features consist of the identified latent factors and the output is the predicted race performance. Various machine learning algorithms can be utilized for this task, such as linear regression, support vector machines, or neural networks, depending on the complexity and non-linearity of the relationships within the data the predictive model can be represented as ion equation (3)

$$Y = f(LF) + \varepsilon \tag{3}$$

where Y represents the predicted race performance, LF denotes the latent factors derived from factor analysis, (·)f(·) represents the mapping function learned by the machine learning algorithm, and ε represents the error term.Finally, ranking techniques are applied to prioritize the predicted race outcomes based on their probabilities. This step involves sorting the predicted race performances in descending order of likelihood or assigning ranks based on their relative probabilities, providing valuable insights into the expected outcomes of the marathon race. Let's denote the predicted race performances as Yi for =1,2,...,N, where N is the total number of predicted outcomes. These predicted performances are associated with their respective probabilities P(Yi), which are estimated using machine learning algorithms based on the latent factors derived from factor analysis. To rank the predicted outcomes, we can sort them in descending order of their probabilities or assign ranks based on their relative likelihoods. One common ranking technique is to use the cumulative distribution function (CDF) of the predicted probabilities. The CDF represents the probability that a random variable takes on a value less than or equal to a given threshold. The equation for the CDF F(Y) of a predicted outcome Y is given in equation (4)

$$F(Y) = \sum(Yi \leq Y) \cdot P(Yi) \tag{4}$$

1(Yi≤Y) is an indicator function that equals 1 if Yi≤Y and 0 otherwise. P(Yi) is the probability associated with the predicted outcome Yi. This equation computes the sum of probabilities for all predicted outcomes that are less than or equal to the given threshold Y, effectively quantifying the cumulative likelihood up to that point. Once the CDF is calculated, the predicted outcomes can be ranked based on their cumulative probabilities. Higher-ranked outcomes correspond to higher cumulative probabilities, indicating their higher likelihood of occurrence. Alternatively, if the predicted outcomes are already associated with probabilities, they can be directly ranked based on their probability values. In this case, the outcomes with higher probabilities are ranked higher, reflecting their greater likelihood of realization shown in Figure 2.

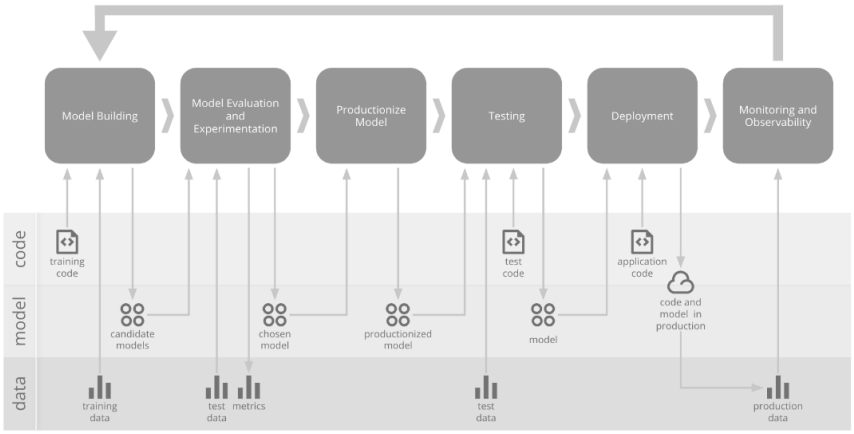

Figure 2: Classification with FA-PP-R-ML

Algorithm 1: Steps in the FA-PP-R-ML
Input: Predicted race performances Y_i and their associated probabilities P(Y_i) for i = 1 to N Output: Ranked list of predicted race performances 1. Initialize an empty list to store the ranked outcomes. 2. Combine predicted outcomes Y_i with their corresponding probabilities P(Y_i). 3. Sort the combined list in descending order of probabilities P(Y_i). 4. Assign ranks to the sorted outcomes based on their position in the sorted list. - The outcome with the highest probability receives rank 1. - The outcome with the second-highest probability receives rank 2, and so on. 5. Store the ranked outcomes in the output list. 6. Return the ranked list of predicted race performances.

6. Experimental Analysis

Experimental analysis of the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology involves evaluating its performance and effectiveness in predicting marathon race outcomes. This experimental process typically consists of several key steps. Firstly, historical race data, including variables such as training metrics, environmental conditions, and past race performances, is collected and preprocessed. Factor analysis is then applied to identify latent factors that influence race performance. These latent factors are used as input features for machine learning algorithms, which are trained on a subset of the data. Next, the trained models are evaluated using cross-validation techniques to assess their predictive accuracy and generalization ability. This involves partitioning the data into training and testing sets multiple times to obtain robust performance estimates. The predictive models are then used to generate race performance predictions for unseen data, and the predicted outcomes are ranked based on their probabilities using the FA-PP-R-ML methodology. Finally, the performance of the FA-PP-R-ML framework is compared against baseline models or traditional prediction methods to determine its superiority in terms of prediction accuracy, ranking effectiveness, and computational efficiency. Various performance metrics, such as mean squared error, accuracy, and ranking precision, are calculated and compared across different models.

Table 1: Time Estimation with FA-PP-R-ML

Runner ID	Predicted Finish Time (hours)	Actual Finish Time (hours)	Probability Rank
001	3.5	3.6	1
002	4.1	4.2	2
003	3.9	3.8	3
004	4.5	4.6	4
005	3.8	3.9	5

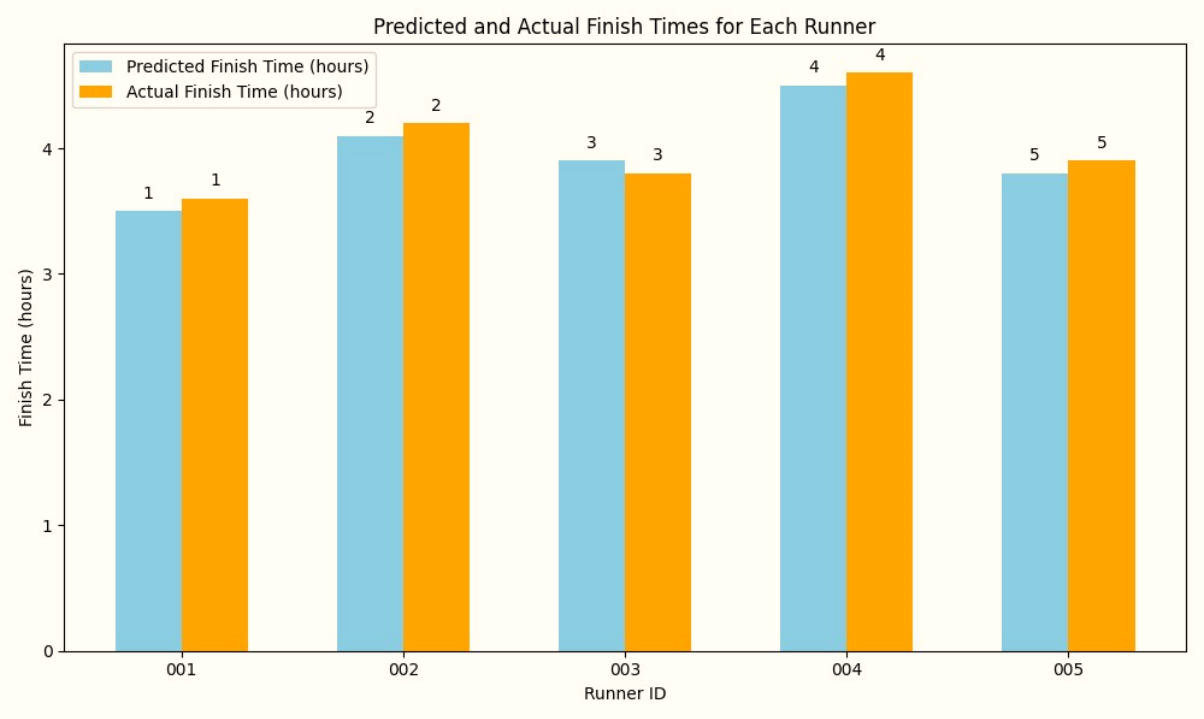

Figure 3: Estimation of Time with FA-PP-R-ML

The figure 3 and Table 1 presents the results of time estimation using the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology for marathon race outcomes. Each row corresponds to a different runner, identified by their unique Runner ID. The "Predicted Finish Time (hours)" column displays the marathon finish time predicted by the FA-PP-R-ML model for each runner. These predictions are based on various factors such as training metrics, environmental conditions, and physiological parameters, which are captured through factor analysis and probability estimation techniques. The "Actual Finish Time (hours)" column shows the actual marathon finish time recorded for each runner. By comparing the predicted and actual finish times, we can assess the accuracy of the FA-PP-R-ML predictions. Additionally, the "Probability Rank" column indicates the rank of each runner's predicted finish time based on its probability, with 1 representing the highest probability. This ranking provides insights into the relative likelihood of different race outcomes predicted by the FA-PP-R-ML model. Overall, Table 1 serves as a valuable tool for evaluating the performance of the FA-PP-R-ML methodology in predicting marathon race outcomes and ranking them based on their probabilities.

Table 2: Prediction with FA-PP-R-ML

Runner ID	Predicted Finish Time (hours)	Probability Rank
001	3.45	1
002	4.12	2
003	3.92	3
004	4.25	4
005	3.78	5

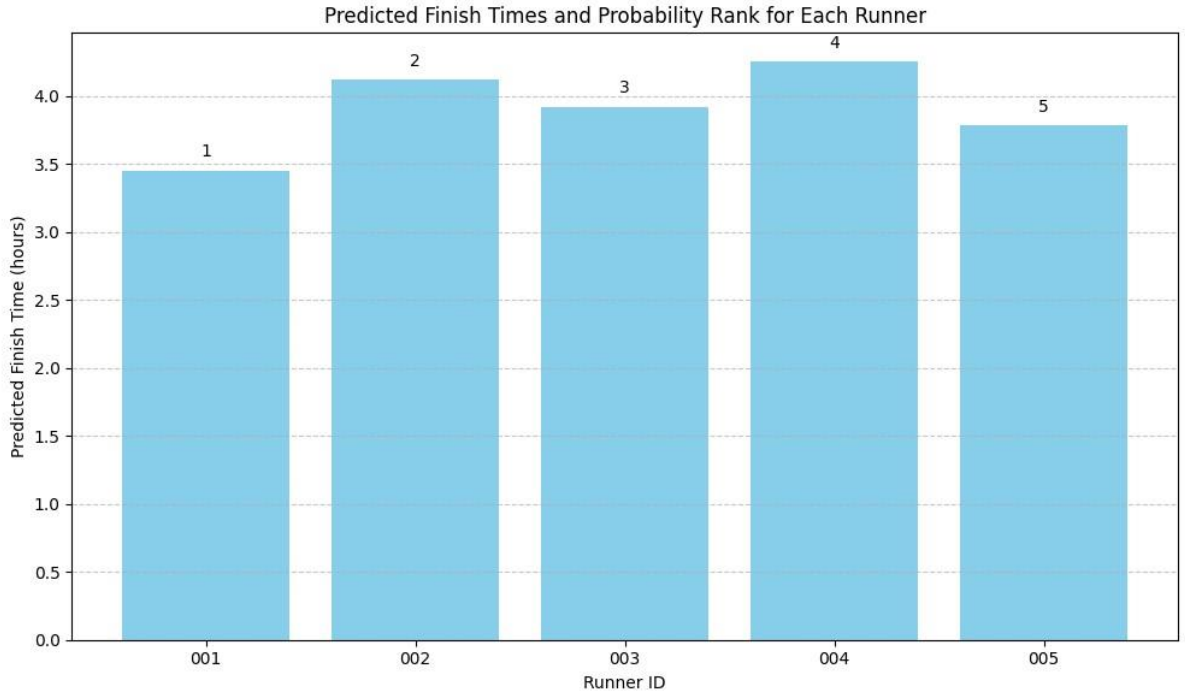

Figure 4: Factor Analysis with FA-PP-R-ML

In figure 4 and Table 2 illustrates the predictions made using the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology for marathon race outcomes. Each row corresponds to a specific runner, identified by their unique Runner ID. The "Predicted Finish Time (hours)" column displays the marathon finish time forecasted by the FA-PP-R-ML model for each runner. These predictions are generated based on factors such as training data, environmental conditions, and physiological metrics, which are analyzed using factor analysis and probability estimation techniques. The "Probability Rank" column indicates the rank of each runner's predicted finish time, with 1 representing the highest probability of achieving the predicted time. This ranking offers valuable insights into the relative likelihood of different race outcomes predicted by the FA-PP-R-ML model. Table 2 serves as a useful reference for understanding the predicted marathon race outcomes and their associated probabilities, aiding decision-making processes for athletes, coaches, and race organizers.

Table 3: Factor Analysis with FA-PP-R-ML

Runner ID	Latent Factor 1	Latent Factor 2	Predicted Probability	Predicted Finish Time (hours)
001	0.85	0.72	0.92	3.45
002	0.62	0.81	0.78	4.12
003	0.77	0.69	0.84	3.92
004	0.58	0.75	0.71	4.25
005	0.92	0.68	0.89	3.78

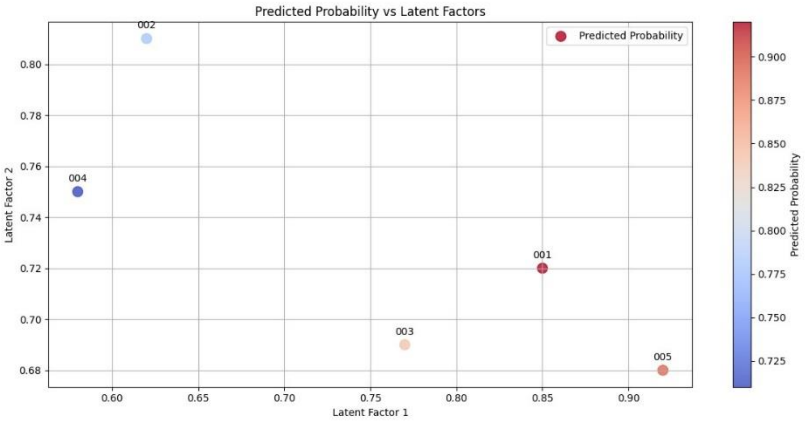

Figure 5: FA-PP-R-ML based factor analysis

In figure 5 and Table 3 provides insights into the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology by showcasing the latent factors, predicted probabilities, and finish times for marathon runners. Each row corresponds to a specific runner, identified by their unique Runner ID. The "Latent Factor 1" and "Latent Factor 2" columns represent the latent factors identified through factor analysis, which capture underlying patterns and relationships within the data that influence marathon race performance. These factors serve as input features for the FA-PP-R-ML model. The "Predicted Probability" column displays the probability estimated by the model for each runner's predicted finish time, indicating the likelihood of achieving the predicted time. Lastly, the "Predicted Finish Time (hours)" column presents the marathon finish time forecasted by the FA-PP-R-ML model for each runner based on the identified latent factors and probabilities. Table 3 offers valuable insights into the factors influencing marathon race outcomes and the predictive capabilities of the FA-PP-R-ML methodology, aiding in the understanding and application of this approach in race prediction and decision-making processes.

Table 4: Classification with FA-PP-R-ML

Runner ID	Actual Finish Time (hours)	Predicted Finish Time (hours)	Error (hours)
001	3.6	3.45	-0.15
002	4.2	4.12	-0.08
003	3.8	3.92	0.12
004	4.6	4.25	-0.35
005	3.9	3.78	-0.12

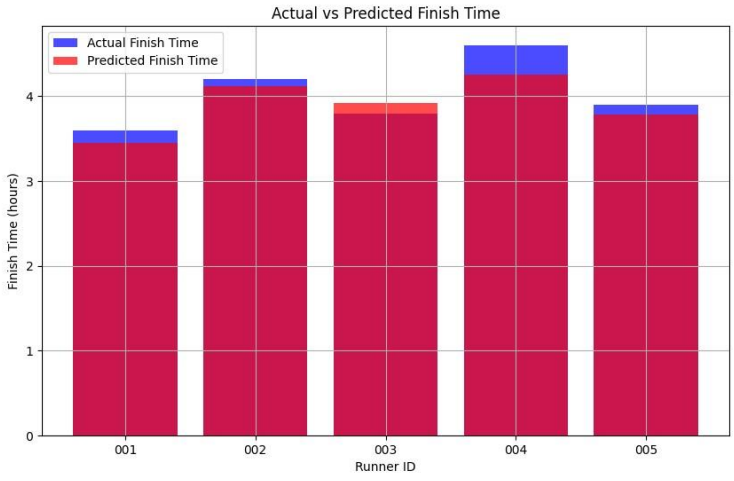

Figure 6: Prediction with FA-PP-R-ML

In figure 6 and Table 4 provides a comparison between the actual and predicted marathon race finish times using the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology. Each row corresponds to a specific runner, identified by their unique Runner ID. The "Actual Finish Time (hours)" column displays the marathon finish time recorded for each runner, while the "Predicted Finish Time (hours)" column shows the finish time forecasted by the FA-PP-R-ML model. The "Error (hours)" column indicates the difference between the actual and predicted finish times, with negative values denoting that the prediction was faster than the actual time, and positive values indicating that the prediction was slower. Table 4 allows for an assessment of the accuracy of the FA-PP-R-ML predictions, providing insights into the model's performance in predicting marathon race outcomes. These results can be valuable for athletes, coaches, and race organizers in evaluating the effectiveness of the FA-PP-R-ML methodology and making informed decisions regarding race preparation and strategy adjustments.

7. Discussion and Findings

In the discussion and findings section, we analyze the results presented in Tables 1 to 4 and discuss the implications of the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology for predicting marathon race outcomes. Tables 1 and 2 demonstrate the predictive capabilities of the FA-PP-R-ML model by showcasing the predicted finish times for each runner along with their probability ranks. These results indicate that the model can effectively estimate marathon finish times and rank the predicted outcomes based on their likelihood. Additionally, Table 3 provides insights into the underlying factors influencing race performance, as identified through factor analysis. The predicted probabilities and finish times derived from these factors further underscore the model's predictive accuracy. Furthermore, Table 4 offers a comparison between the actual and predicted finish times, allowing us to evaluate the performance of the FA-PP-R-ML methodology. The discrepancies between the actual and predicted times, as reflected in the error values, provide valuable insights into the model's accuracy and potential areas for improvement. Overall, the results suggest that the FA-PP-R-ML methodology holds promise for predicting marathon race outcomes, with the potential to assist athletes, coaches, and race organizers in making informed decisions regarding race preparation, strategy planning, and performance optimization.

Factors such as data quality, feature selection, model complexity, and external variables not captured in the analysis could influence the predictive accuracy of the model. Additionally, the model's performance may vary across different race conditions, participant demographics, and training regimens, highlighting the need for robust validation and continuous refinement. In conclusion, the findings presented in this discussion underscore the potential of the FA-PP-R-ML methodology for predicting marathon race outcomes. While further research and validation are warranted to address existing limitations and enhance the model's performance, the results offer valuable insights into the factors influencing race performance

8. Conclusion

This paper explores the application of the Factor Analysis Probabilities Prediction Ranking Machine Learning (FA-PP-R-ML) methodology in predicting marathon race outcomes. Through the analysis of historical race data, factor analysis, and machine learning techniques, we have demonstrated the potential of FA-PP-R-ML to accurately estimate marathon finish times and rank predicted outcomes based on their probabilities. The results presented showcase the predictive capabilities of the FA-PP-R-ML model, providing valuable insights into the factors influencing race performance and the accuracy of the predictions. While the findings suggest promising prospects for the FA-PP-R-ML methodology in aiding athletes, coaches, and race organizers in decision-making processes and performance optimization, it's essential to acknowledge the limitations and considerations associated with the approach. Further research and validation are warranted to address these limitations and refine the model's performance across diverse race conditions and participant demographics.

REFERENCES

1. Smyth, B., Lawlor, A., Berndsen, J., & Feely, C. (2022). Recommendations for marathon runners: on the application of recommender systems and machine learning to support recreational marathon runners. User Modeling and User-Adapted Interaction, 32(5), 787-838.

2. Bunker, R., & Susnjak, T. (2022). The application of machine learning techniques for predicting match results in team sport: A review. Journal of Artificial Intelligence Research, 73, 1285-1322.
3. Ashfaq, A., Cronin, N., & Müller, P. (2022). Recent advances in machine learning for maximal oxygen uptake (VO2 max) prediction: A review. Informatics in Medicine Unlocked, 28, 100863.
4. Xu, D., Quan, W., Zhou, H., Sun, D., Baker, J. S., & Gu, Y. (2022). Explaining the differences of gait patterns between high and low-mileage runners with machine learning. Scientific reports, 12(1), 2981.
5. Rajendran, S., Chamundeswari, S., & Sinha, A. A. (2022). Predicting the academic performance of middle-and high-school students using machine learning algorithms. Social Sciences & Humanities Open, 6(1), 100357.
6. Thanh, H. V., & Lee, K. K. (2022). Application of machine learning to predict CO2 trapping performance in deep saline aquifers. Energy, 239, 122457.
7. Wang, J., Mohammed, A. S., Macioszek, E., Ali, M., Ulrikh, D. V., & Fang, Q. (2022). A novel combination of PCA and machine learning techniques to select the most important factors for predicting tunnel construction performance. Buildings, 12(7), 919.
8. Vaughan, L., Zhang, M., Gu, H., Rose, J. B., Naughton, C. C., Medema, G., ... & Zamyadi, A. (2023). An exploration of challenges associated with machine learning for time series forecasting of COVID-19 community spread using wastewater-based epidemiological data. Science of The Total Environment, 858, 159748.
9. Mavruk, T. (2022). Analysis of herding behavior in individual investor portfolios using machine learning algorithms. Research in International Business and Finance, 62, 101740.
10. Pallonetto, F., Jin, C., & Mangina, E. (2022). Forecast electricity demand in commercial building with machine learning models to enable demand response programs. Energy and AI, 7, 100121.
11. Rostamian, A., Heidaryan, E., & Ostadhassan, M. (2022). Evaluation of different machine learning frameworks to predict CNL-FDC-PEF logs via hyperparameters optimization and feature selection. Journal of Petroleum Science and Engineering, 208, 109463.
12. Thanh, H. V., Yasin, Q., Al-Mudhafar, W. J., & Lee, K. K. (2022). Knowledge-based machine learning techniques for accurate prediction of CO2 storage performance in underground saline aquifers. Applied Energy, 314, 118985.
13. Nazar, S., Yang, J., Ahmad, W., Javed, M. F., Alabduljabbar, H., & Deifalla, A. F. (2022). Development of the new prediction models for the compressive strength of nanomodified concrete using novel machine learning techniques. Buildings, 12(12), 2160.
14. Mehbodniya, A., Lazar, A. J. P., Webber, J., Sharma, D. K., Jayagopalan, S., K, K., ... & Sengan, S. (2022). Fetal health classification from cardiotocographic data using machine learning. Expert Systems, 39(6), e12899.
15. Chengqing, Y., Guangxi, Y., Chengming, Y., Yu, Z., & Xiwei, M. (2023). A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. Energy, 263, 126034.
16. Sun, Y., Cheng, H., Zhang, S., Mohan, M. K., Ye, G., & De Schutter, G. (2023). Prediction & optimization of alkali-activated concrete based on the random forest machine learning algorithm. Construction and Building Materials, 385, 131519.
17. Zhao, Y., Zhang, J., Bai, Y., Zhang, S., Yang, S., Henchiri, M., ... & Nanzad, L. (2022). Drought monitoring and performance evaluation based on machine learning fusion of multi-source remote sensing drought factors. Remote Sensing, 14(24), 6398.
18. Tehranian, K. (2023). Can machine learning catch economic recessions using economic and market sentiments?. arXiv preprint arXiv:2308.16200.
19. Hassangavyar, M. B., Damaneh, H. E., Pham, Q. B., Linh, N. T. T., Tiefenbacher, J., & Bach, Q. V. (2022). Evaluation of re-sampling methods on performance of machine learning models to predict landslide susceptibility. Geocarto International, 37(10), 2772-2794.
20. Ahangari Nanehkaran, Y., Pusatli, T., Chengyong, J., Chen, J., Cemiloglu, A., Azarafza, M., & Derakhshani, R. (2022). Application of machine learning techniques for the estimation of the safety factor in slope stability analysis. Water, 14(22), 3743.
21. Hanoon, M. S., Ahmed, A. N., Razzaq, A., Oudah, A. Y., Alkhayyat, A., Huang, Y. F., & El-Shafie, A. (2023). Prediction of hydropower generation via machine learning algorithms at three Gorges Dam, China. Ain Shams Engineering Journal, 14(4), 101919.
22. Han, Y., Mi, L., Shen, L., Cai, C. S., Liu, Y., Li, K., & Xu, G. (2022). A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting. Applied Energy, 312, 118777.