International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Horse Race Results Prediction Using Machine Learning Algorithms With Feature Selection

[1]Meenakshi Gupta, [2]Latika Singh

Submitted: 17/07/2023 **Revised**: 08/09/2023 **Accepted**: 25/09/2023

Abstract— People's interest in horse racing has skyrocketed along with its rapid expansion. Some experts and academics have studied the best practices for managing decisions and making predictions in horse racing. In the areas of categorization and prediction, applying the machine learning (ML) paradigms has demonstrated hopeful results. Betting on sports is big business, making accurate predictions in this field increasingly important. In addition, club executives are looking for classification models to better comprehend the game and develop winning plans. Research has shown that Machine Learning algorithms offer a good answer to the categorization and prediction problem in horse racing, where traditional prediction algorithms have failed. In this study, we present several ML approaches for predicting the outcome of horse races, including K-nearest-neighbor (KNN), Linear-regression (LR), Randomforest(RF), Gaussian NaiveBayes (NB), ADA Boost (BAG), along with Bagging. These models take into account several aspects of the games, including past match outcomes, player and horse statistics, information about the competition, and more. The results of the massive-scale studies showed that the RF approach produced more accurate predictions than any of the other models tested. We believe that researchers who delve into this field in the future will find our work both enlightening and useful.

Keywords— *Machine Learning, Random forest, Horse race results prediction, feature selection*

1. Introduction

Predictions in sports are typically viewed as a one-class problem (win, loss, or draw). The topic of predicting a numerical value, such as the size of a victory margin, has also been studied by a number of scholars. Predicting the outcome of a sporting event requires collecting a vast number of variables, such as past performance, match outcomes, and player information, to let various parties assess their chances of success. Bookies, fan-audiences, and punters are all concerned in guessing the results of a racing ahead of the start owing to the monetary stakes involved in making a prediction about which team is most likely to prevail [1]. After a match prediction has been made, there is still the issue of whether or not to bet on a particular horse given the odds offered by the bookmaker [2]. Sport management are also trying to develop effective ways for sizing up a possible opponent before a match. Therefore, the difficulty of forecasting sports outcomes has long piqued the interest of various parties, including the media. More and more sports-related data is being collected and stored digitally, making it possible to create sophisticated models and prediction systems to anticipate the outcomes of games [3].

Nowadays, people from all walks of life enjoy going to outdoor horse races. This sector encompasses a wide range of activities, from sports to entertainment to gambling to commerce. Every year there are 3.2M races, and the total-prize pool is worth 360 billion yuan. More than 70 countries and areas throughout the world are actively cultivating horse racing right now. Horse racing has flourished and become a major industry in many countries [4]. Horse racing serves as a catalyst for other industries, a backbone for social services and nonprofits, and a major source of revenue for the government. It facilitates social interaction and serves as a major source of amusement for most people. People today often use their membership in equestrian organizations and their own of horses as a means of displaying their individuality and social standing [5].

As people's standard of living rises dramatically, so do efforts to improve the horse racing industry. The horse racing lottery, as a subset of the wider horse racing industry, has attracted widespread interest and financial backing from the general public. The horse racing lottery has attracted a growing number of casual fans. In the United Kingdom, for instance, horse racing now generates over a billion pounds annually and is the country's second-largest sports sector, behind only football [6]. About 130 million people worldwide guess the results of horse races each year as a result of the growing popularity of the industry. People started compiling general laws across the several sorts of horse racing, looking for ways to increase the accuracy of their predictions. More academics and the general public are interested in learning how to construct a management model for making decisions in horse racing and how to use scientific methods to address issues with forecasting [7].

Artificial intelligence (AI) is an emerging field that combines the theory, methods, techniques, models, systems, and applications of human intelligence with the fast advancements in computer science. Machine learning (ML)

[1]*SET, Sushant University*
[2]*SET, Sushant University*
**meenakshi78gupta@gmail.com*

is just one subfield of this larger area of study, which has implications for medicine, teaching, farming, space exploration, renewable energy, manufacturing, and even sports [8]. One of the most common applications of ML is classification, which entails making predictions about a target variable using data that has not been seen before. Classification is a technique for making predictions about a target variable by first creating a classification approach from a training-data and applying it to the class corresponding to the test-data [9] to make predictions. As the class variable is considered during model construction in the data-processing phase, hence the ML process is called supervised learning. Loan approval, medical diagnosis, and email screening are just a few examples of common uses for classification. To fully comprehend a sport, its strategy, and how to make judgments while minimizing uncertainty, it is necessary to collect and catalog a comprehensive collection of relevant metrics [10].

Sports are a complex system in which many independent variables interact. When compared to a single agent, the learning space for multiple agents exponentially grows with the number of agents, making training more challenging. As a result, researchers can utilize approach of machine learning (ML) so that models can be designed and systems can be simulated to make predictions [11]. Researchers are attempting to use various features, such as a team's past performance, the outcomes of previous matches, and information gathered about players and equipment, because traditional approaches to sports prediction consider the task as a classification issue with a single class to be anticipated. This means that a correlation must be established between machine learning prediction outputs and the strategy employed by the sport team manager in order to ensure that both are accurate [12].

In this study, we suggest a method for predicting the outcomes of horse races using machine learning. The model is built with the help of the various feature selection techniques, which take these variables as input. KNN, LR, ADA Boost, Bagging, and RF are the five machine learning algorithms used to generate prediction models using these custom-designed features. Finally, the analysis of the proposed models is being done in a variety of ways. The structure of this study is as follows: The rationale and scope of this research are presented in Section 2. The proposed machine learning approaches employed in this study are presented in Section3. Results and discussion are presented in Section4. Conclusions and further research are summarized in Section5.

2. Related Works

The ability of humans to train and direct horses at high speeds is put to the test in horse racing. It is a major competition in the equestrian sports world. It takes many different shapes, but at its core, it's a race against the clock. Here we summarize the numerous sports prediction models developed by academics. In the early days of employing ANN models for NFL prediction, one of the first researches was undertaken by Purucker [13]. Five features were used to analyze the data, which included total yards, rushing yards, marginal turn-over, occupied time and odds of betting-line from the first eight rounds of the competition. Clustering-based unsupervised algorithms were utilized to categorize teams into strong and weak performers. Then, a neural network (ANN) equipped with backpropagation (BP) was used. When compared to the domain experts' 72% accuracy, Purucker only managed 61%. It was determined that the BP algorithm provided the best results. The small number of characteristics employed is one of the studies flaws.

Researchers Edelmann-Nusser et al. [14] looked into modeling an Olympic-caliber female swimmer's performance in the final lap of the 200-meter back-stroke during 2000 Summer-Games, Sydney. Information was gathered from the swimmer's training in the final four weeks before the Olympics, as well as from the results of 19 races in the 200-meter. The network employed was an ML Perceptron consisting of input of 10 neurons, with 2 hidden ones and 1 output neuron. The outcomes demonstrate the MLP's efficacy, with a prediction error of only 0.05 s. Results from the MLP were compared to those from linear regression, which were shown to be less reliable. This research, along with Karetnikov et al. [15], emphasizes the potential of machine learning approaches for use by sports professionals including staff and analysts, not just to predict results or outocmes but also for determining which elements should be prioritized when designing training programs.

According to Zhang and Liu [16], condition of the horses has a direct impact on the success of the competition as the horse represent a "unique match between two life forms". According to studies conducted by Izzo et al., [17] equestrian contests require highly trained emergency medical personnel due to the high number of injuries sustained by competitors. Based on their findings, Borowski et al. [18] recommend establishing a professional talent training system. Competitions in equestrian sports have been shown to promote and drive associated consumption, according to research by Ngyyen JK [19]. Padalino et al. [20] investigated Hong Kong's gaming sector, what factors decide and foster the growth of and sports culture by analyzing the peculiarities of horse racing, the foundation of huge number of sporting activities in Hong Kong, and the mechanisms behind their management. To help with the informatization of events, Hoseini and Amani presented a B/S model-based equestrian event management system [21].

When comparing three different neural networks, the authors of [22] found that a feed-forward network yielded the highest accuracy (74.39). The authors [23] predict outcomes using Naive Bayes and get an accuracy of 67%. They employed multivariate linear regression to predict the outcome dispersion, but were only 10% accurate. Predictions of this nature are often tricky, therefore this outcome can hardly be considered terrible. The author of [24] attained an accuracy of 70.0% utilizing these ML algorithms and these datasets. With the first feature set, the Support-Vector-Machine achieved 70.0% exactness, while the best relative feature set for logistic regression only reached 69.7%. In [25], the author reports highest accuracy of 68.4% when using a multilayer perceptron, 68.0% when using linear regression, and 66.8% while using the maximum likelihood classifier. The authors of [26], after employing multiple ML algorithms, reach the conclusion that a team's winning record in previous games is an important factor in determining game outcomes. Using the random forest strategy, they were able to improve accuracy to 65.2%.

In [27], the author proposes a model built on matrix factorization, which, with just one season of training, achieves a state-of-the-art 71.0% accuracy. The best accuracy was found in, when a maximum entropy model was presented and achieved. Predictions of game outcomes were made using a variety of classification and regression ML techniques in, with the authors achieving a 65.5% accuracy using Gaussian discriminant analysis. Using an SVM prediction model and a feature selection approach, the authors of [28] were able to increase prediction accuracy to 85.2%. An ML model using stacked Bayesian regressions was presented and optimized by the authors of [29], who report an optimal accuracy of 85.3%. The authors of [30] employed four separate feature sets to get an accuracy of 88.1% at its best.

Features based on a team's concert and standing within the observed league were used in most of the aforementioned research. Transferring practices from one league to another and prescribing a uniform structure that would satisfy all standards for a watched sport is not simple tasks. Therefore, the amount of data utilized and the competitiveness of the league analyzed greatly affect the presented prediction results. The specifics of feature-selection (FS) and extraction procedures also varied widely, depending on the study methodology employed and the features of the game that were deemed most important for determining the outcomes of the games in question. However, this step is fundamental to the entire ML process and significantly affects the final outcome of the predictions. In this research, we will demonstrate how choosing the proper features illustrated by data acquired from the ideal time frames affects the prediction outcomes and highlight the significance of doing so.

3. Proposed Methodology

We contend that the best results may be obtained from any given data set by using a well-organized investigational approach for predicting the sport results accurately. This section presents a smart architecture for predicting the outcomes of horse races, including potential ML approach steps along with details of the data utilized in such predictions within the proposed frameworks. The conceptual diagram for the suggested model is presented in Figure 1.

A. Data Preparation and Feature Selection

A live web site of a well -known Horse Racing Company in India was used to obtain the dataset. The data was in .csv format and included 14,750 rows with 56 attributes. Finding the optimal feature set for modeling the phenomena of interest is the purpose of feature selection approaches in ML. In ML, feature selection strategies can be broken down into two categories: supervised and unsupervised. We have implemented the information-gain along with Chi Square filtering algorithms to select features [31]. Information-gain measures how much entropy is lost when a dataset is transformed. For feature selection, it's useful to compare target variable to each candidate variable's Information gain. Chi-square test is used in case of categorical values in the dataset. Chi-square score for every feature in relation to the target is computed, and the high score features are chosen. Dataset features should be independently sampled and categorical along with independent sampling with a known limit specified in order for the chi-squared test to be applied correctly to examine the association between them and the target variable [32].

Wrappers need a way to find the best subset of features, evaluate it by training and testing a classifier with just that subset, and then choose the best one. When fitting a machine learning algorithm to a dataset, its output will inform the feature selection process. Predictive accuracy is typically higher when using wrapper techniques rather than filter methods. The following are the three categories. With Forward Feature Selection, we first compare the features we have so far with the features we want to use. Then, we pick a second variable that improves the model's performance in tandem with the initial one. This procedure is repeated until the target is reached. Contrast this with the Forward Feature Selection (FFS) approach, in which the opposite features are chosen [33]. In this case, we begin by constructing a model using all of the data at our disposal. Then, we take the best-valued evaluation measure variable from the model. This procedure is repeated until the target is reached. The Sequential Feature Selection (SFS) method encompasses both of the aforementioned approaches. Recursive feature elimination (RFE) works by choosing features after considering progressively small groups of features by assigning weights to features (external estimator such as linear model coefficients). An initial collection of features is used to train the estimator, and the relevance of every selected feature is then determined using a coefficient-attribute. Next, we narrow down the final feature-set by eliminating ones we deem unnecessary. These steps are iterated recursively with reduced attribute-set till the target of feature-set for selection is achieved.

B. Machine Learning Models

Algorithms developed using machine learning may analyze data, make predictions, and optimize themselves based on past results. Machine learning allows for a wide variety of algorithmic approaches to problems.

Linear regression

One of the most widely used and accessible ML methods for predictive analysis is linear regression. Predictive analysis provides a definition for prediction, and linear regression forecasts interval values [34]. It illustrates the linear connection between the two variables and the way in which the dependent variable (y) varies as a function of the independent variable (x). Regression analysis seeks to find the line that provides the best fit amid dependent and independent variables. Here is the formula for the regression line:

$y= a_0+ ax+ b$

Here, y= target variable, x= independent variable $a_0 =$ Intercept of line.

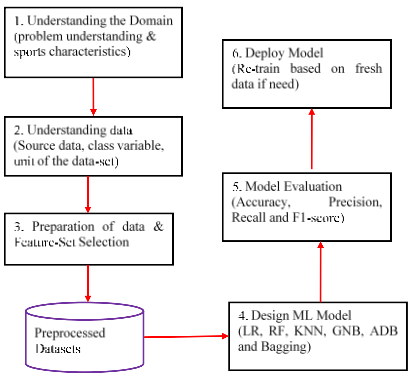

Fig. 1. Proposed model architecture

K-Nearest Neighbour (KNN)

The supervised learning approach known as K-Nearest Neighbour can be used for both classification and regression problems. This method is able to produce results because it makes assumptions about the existing data and the new data point. The new information is classified into groups with the most similar characteristics based on the similarities found [35]. This approach saves all available datasets and K-neighbours is used to classify every new sample, sometimes referred to as the "lazy learner algorithm." Any distance function can be used to determine which existing class best fits the characteristics of the new example.

Gaussian Naive Bayes (GNB)

This is a probabilistic method and Gaussian distribution based classification strategy used in Machine Learning (ML). Each parameter (also known as features or predictors) in Gaussian Naive Bayes is treated as though it were capable of predicting the output variable on its own. When the continuous variables associated with each feature are assumed to follow a Gaussian distribution, Gaussian Naive Bayes is utilized. Another name for the Gaussian distribution is the Normal distribution. With continuous data, we can use this formula to calculate the likelihood of probabilities [35].

Random Forest

When it comes to machine learning, supervised learning method of random forest is the choice for tasks involving both classification and regression problems. In this method of ensemble learning, results from numerous classifiers are used together to make more accurate predictions. To enhance the model's forecast precision, it includes numerous decision trees for different slices of the input dataset and calculates an average. Between 64 and 128 trees are ideal for a random forest. The more trees there are, the more precise the algorithm becomes. Each tree votes on how best to categorize a new dataset or item, and the algorithm then makes a prediction based on the results with the most votes [36].

ADA Boost Algorithm

Including trees in a model is how the additive model is described. In order to avoid altering preexisting trees in the model, we should only add a single tree at a time. To further favor the gradient descent approach, we can employ trees to mitigate loss. In recent times, the gradient descent technique has been utilized to find the optimal values for a neural network's weights and regression equation coefficients. Error or loss is first calculated, then the weight parameter is employed to attain the best possible result. However, these parameters are no longer favored by most ML experts, who instead use weak learner sub-models or decision trees. In which case, we need to incorporate a tree into the model to enhance its accuracy and efficiency. A final prediction is reached by adding the forecast from the newly added tree to the predictions from the previously established sequence of trees. This procedure is repeated until the amount of loss is tolerable, at which point further enhancement is unnecessary [37].

Bagging

Bagging (or bootstrap aggregation) is an example of an ensemble learning approach used to reduce the impact of noise in otherwise difficult datasets. To select the same data points several times, bagging includes randomly selecting a subset of a training set with replacement. The idea behind "bagging" is that by working together, a group of "average" students can surpass an individual "exceptional" student. It helps lower the overall amount of variation, which keeps models from being overfit. The interpretability of a model suffers when bagging is used [38].

4. Results & Discussions

In this section, we compare accuracy for each of the suggested machine learning models across three distinct datasets: the original dataset, the FS info gain dataset, and the FS-CHI2 dataset. It was crucial to address the class imbalance issue that could cause biased results, as there is a clear correlation between the amount of components in each class and the reliability of the results. To deal with the issue of skewed horse-racing data, a preprocessing method known as SMOTE (Synthetic-Minority-Oversampling-Technique) was implemented. Oversampling techniques like SMOTE, which employs the k nearest nodes algorithmically to construct the false data, were used, with k set to 5. Setting the minority class as set M in the k-NN algorithm allows the SMOTE algorithm to generate new samples. Each e > M data point had its k nearest neighbors calculated. Next, the vector between e and one of these k neighbors (ek) was calculated and multiplied by a random number between 0 and 1. After adding this value to e, we obtained the fabricated data point (e'). Using interpolation between nearby positive cases, fresh synthetic data are generated, and a new minority class set M1 is formed that is equivalent to the majority class. New samples were created using the method described in the following formula (Equation 1). Random numbers between 0 and 1 are represented by rnd (0, 1). Model accuracy determined by equation 2. Table 1 and Figure 2 display the accuracy values. According to the data, the RF algorithm has a higher success rate (93.1%) than the other models.

$$e' = e + rnd\,(0, 1) * |e - e_k| \qquad \text{Eq. (1)}$$

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}} \qquad \text{Eq. (2)}$$

Table 1. Accuracy values of ML Models

Accuracy	Bagging	ADB	K-NN	GNB	RF	LR
Before applying FS	73.7	84.6	73.6	59.1	93.1	54.6
After applying Information Gain	75.3	84.8	74.7	54.2	92.2	55.5
After applying CHI2	69.8	70.2	70.1	66.1	71.3	68.1

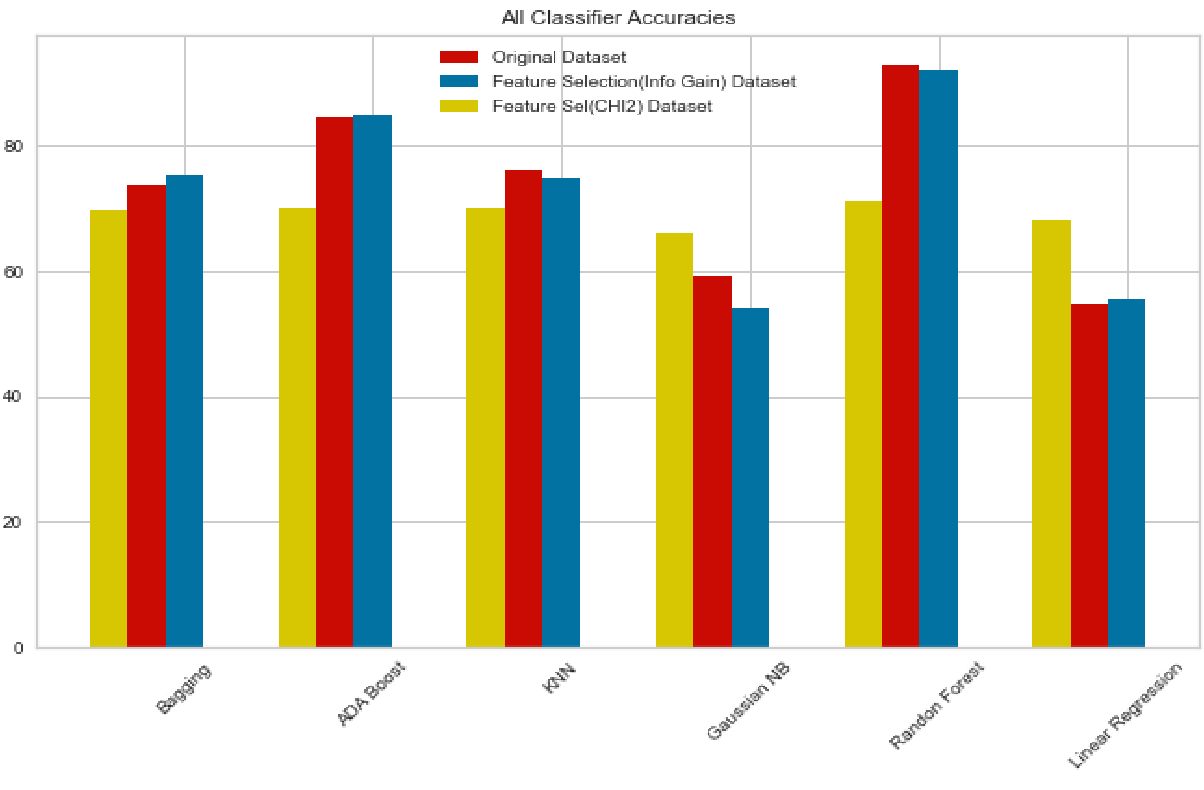

Fig. 2. Accuracy of the proposed ML Models in different levels of dataset

Accuracy and ROC are examined for RF models run with various forms of data feature selection. Table 2 and Figure 3 show the accuracies and ROC values of the RF model using various FS methods. The Information Gain FS method consistently outperforms the alternatives.

Table 2. Accuracy and ROC of RF with different FS approaches

	Information Gain	CHI2	SFS	BFS	RFE
Accuracy	92.2	71.3	75.3	76.0	90.0
ROC	92.2	71.3	84.0	84.6	90.0

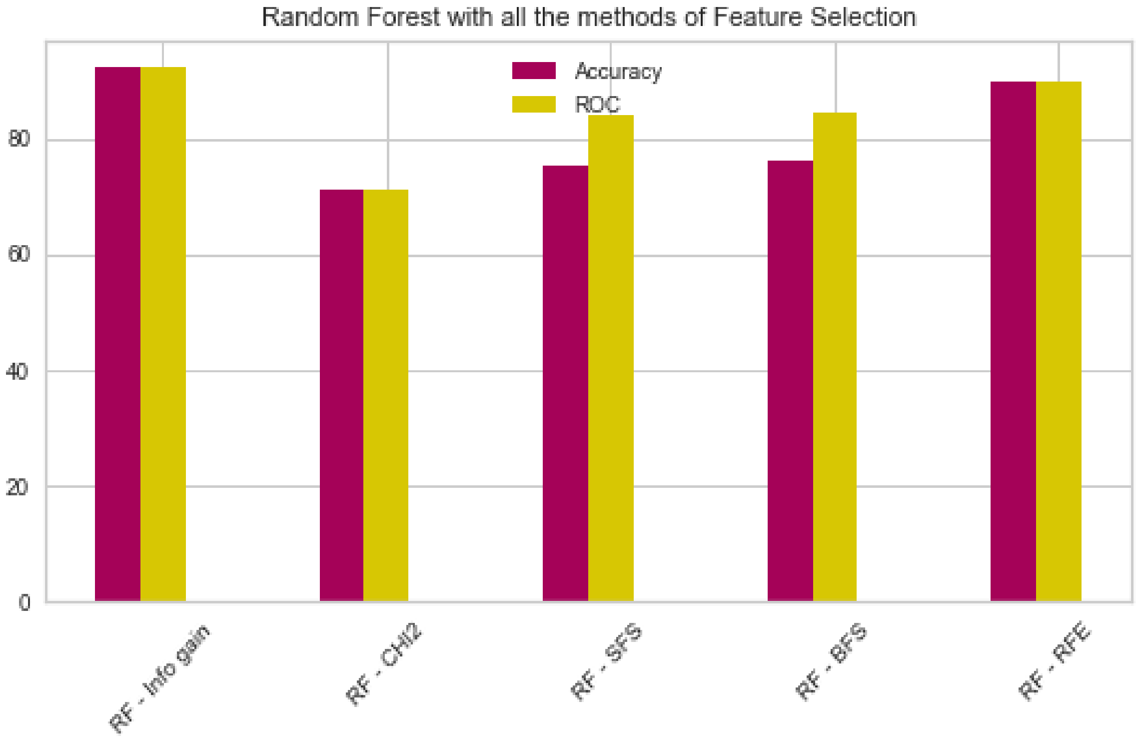

Fig. 3. Accuracy and ROC of the RF Model in different FS approaches

5. Conclusions And Future Works

The number and complexity of information used to predict the outcomes of horse races have expanded significantly alongside the expansion of digital technologies. Large-scale data and dimensionality disasters have become an issue during model training as data and feature sizes have grown. Many researchers and industry professionals have been interested in ML-based algorithms that can effectively process nonlinear data. This work explores a big dataset and presents ML algorithms to construct a horse racing results forecast using several FS methods. As part of this research, we put out five different ML models, including the GNB, LR, RF, KNN, ADAboost, and bagging. Empirical data demonstrate that, relative to competing models, RF provides the highest levels of precision. This approach provides a solid quantitative foundation for sound training management decisions and training programs, the ability to scientifically compare methods for enhancing horse racing performance, and the ability to scientifically predict the public's interest in and support of horse racing. The system is currently functional, but it has certain flaws, its functions aren't flawless, and a lot more data is needed to verify the accuracy of the final prediction. The next stage is to utilize deep learning algorithms to enhance the current system's predictive abilities.

References

[1] Bunker, R. and Susnjak, T., 2022. The application of machine learning techniques for predicting match results in team sport: A review. *Journal of Artificial Intelligence Research*, *73*, pp.1285-1322.

[2] Borowski, P. and Chlebus, M., 2021. *Machine learning in the prediction of flat horse racing results in Poland*. University of Warsaw, Faculty of Economic Sciences.

[3] Muthappa, K.A., Nisha, A.S.A., Shastri, R., Avasthi, V. and Reddy, P.C.S., 2023. Design of high-speed,

low-power non-volatile master slave flip flop (NVMSFF) for memory registers designs. *Applied Nanoscience*, pp.1-10.

[4] Chen, J.C., Dunn, A., Hood, K.K., Driessen, A. and Batch, A., 2019. Off to the races: A comparison of machine learning and alternative data for predicting economic indicators. *Big data for 21st century economic statistics*.

[5] Reddy, P.C.S., Suryanarayana, G. and Yadala, S., 2022, November. Data analytics in farming: rice price prediction in Andhra Pradesh. In *2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)* (pp. 1-5). IEEE.

[6] Gonçalves, R., Ribeiro, V.M., Pereira, F.L. and Rocha, A.P., 2019. Deep learning in exchange markets. *Information Economics and Policy*, *47*, pp.38-51.

[7] Min, M., Lee, J.J., Park, H. and Lee, K., 2021. Detecting anomalous transactions via an iot based application: A machine learning approach for horse racing betting. *Sensors*, *21*(6), p.2039.

[8] Chevallier, J., Zhu, B. and Zhang, L., 2021. Forecasting Inflection points: Hybrid methods with multiscale machine learning algorithms. *Computational Economics*, *57*, pp.537-575.

[9] Singhal, A., Varshney, S., Mohanaprakash, T.A., Jayavadivel, R., Deepti, K., Reddy, P.C.S. and Mulat, M.B., 2022. Minimization of latency using multitask scheduling in industrial autonomous systems. *Wireless Communications and Mobile Computing*, *2022*, pp.1-10.

[10] Lotfi, S., 2021. Machine Learning for sport results prediction using algorithms. *International Journal of Information Technology and Applied Sciences (IJITAS)*, *3*(3), pp.148-155.

[11] Kumar, K., Pande, S.V., Kumar, T.C., Saini, P., Chaturvedi, A., Reddy, P.C.S. and Shah, K.B., 2023. Intelligent Controller Design and Fault Prediction Using Machine Learning Model. *International Transactions on Electrical Energy Systems*, *2023*.

[12] Shanmugaraja, P., Bhardwaj, M., Mehbodniya, A., VALI, S. and Reddy, P.C.S., 2023. An Efficient Clustered M-path Sinkhole Attack Detection (MSAD) Algorithm for Wireless Sensor Networks. *Adhoc & Sensor Wireless Networks*, *55*.

[13] Bunker, R.P. and Thabtah, F., 2019. A machine learning framework for sport result prediction. *Applied computing and informatics*, *15*(1), pp.27-33.

[14] Edelman, D., 2007. Adapting support vector machine methods for horserace odds prediction. *Annals of Operations Research*, *151*(1), pp.325-336.

[15] Karetnikov, A., 2019. Application of data-driven analytics on sport data from a professional bicycle racing team. *Eindhoven University of Technology, The Netherlands*.

[16] Zhang, S., 2022. Optimal Model of Horse Racing Competition Decision Management Based on Association Rules and Neural Network. *Scientific Programming*, *2022*, pp.1-10.

[17] Izzo, C., 2022. *On Explainable Deep Learning for Macroeconomic Forecasting and Finance* (Doctoral dissertation, UCL (University College London)).

[18] Borowski, Piotr, and Marcin Chlebus. *Machine learning in the prediction of flat horse racing results in Poland*. University of Warsaw, Faculty of Economic Sciences, 2021.

[19] Nguyen, J.K., Karg, A., Valadkhani, A. and McDonald, H., 2022. Predicting individual event attendance with machine learning: a 'step-forward'approach. *Applied Economics*, *54*(27), pp.3138-3153.

[20] Baragli, P., Padalino, B. and Telatin, A., 2015. The role of associative and non-associative learning in the training of horses and implications for the welfare (a review). *Annali dell'Istituto superiore di sanita*, *51*, pp.40-51.

[21] D'Urso, P., De Giovanni, L. and Swartz, T., 2023. Big data and data science in sport. *Annals of Operations Research*, pp.1-7.

[22] Lotfi, S., 2021. Machine Learning for sport results prediction using algorithms. *International Journal of Information Technology and Applied Sciences (IJITAS)*, *3*(3), pp.148-155.

[23] Mouloodi, S., Rahmanpanah, H., Martin, C., Gohari, S., Burvill, C. and Davies, H.M., 2022. The use of deep learning algorithms to predict mechanical strain from linear acceleration and angular rates of motion recorded from a horse hoof during exercise. *International Journal of Mechanical Sciences*, *216*, p.106972.

[24] Schumaker, R.P., 2013. Machine learning the harness track: Crowdsourcing and varying race history. *Decision Support Systems*, *54*(3), pp.1370-1379.

[25] Baboota, R. and Kaur, H., 2019. Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, *35*(2), pp.741-755.

[26] Selvaraj, P., 2017. *Predicting The Outcome Of The Horse Race Using Data Mining Technique* (Doctoral dissertation, Dublin, National College of Ireland).

[27] Pudaruth, S., Nemdharry, N., Ramma, T.R.H. and Mungroo, R., 2017, February. Impact of the variation of horse racing odds on the outcome of horse races at the Champ de Mars. In *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-4). IEEE.

[28] Chiang, S. and Denes, G., 2023. Supervised Learning for Table Tennis Match Prediction. *arXiv preprint arXiv:2303.16776*.

[29] Mazidi, A., Golsorkhtabaramiri, M. and Etminan, N., 2020. Sport Result Prediction Using Classification Methods. *Journal of Applied Dynamic Systems and Control*, *3*(2), pp.39-48.

[30] Cliff, D., Hawkins, J., Keen, J. and Lau-Soto, R., 2021. Implementing the BBE Agent-Based Model of a Sports-Betting Exchange. *arXiv preprint arXiv:2108.02419*.

[31] Kholkine, L., Servotte, T., De Leeuw, A.W., De Schepper, T., Hellinckx, P., Verdonck, T. and Latré, S., 2021. A learn-to-rank approach for predicting road cycling race outcomes. *Frontiers in sports and active living*, *3*, p.714107.

[32] Rastegari, H., 2013. A review of data mining techniques for result prediction in sports. *Advances in Computer Science*, *2*.

[33] Štrumbelj, E. and Vračar, P., 2012. Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting*, *28*(2), pp.532-542.

[34] Haghighat, M., Rastegari, H., Nourafza, N., Branch, N. and Esfahan, I., 2013. A review of data mining techniques for result prediction in sports. *Advances in Computer Science: an International Journal*, *2*(5), pp.7-12.

[35] Esme, E. and Kiran, M.S., 2018. Prediction of football match outcomes based on bookmaker odds by using k-nearest neighbor algorithm. *International Journal of Machine Learning and Computing*, *8*(1), pp.26-32.

[36] Baboota, R. and Kaur, H., 2019. Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, *35*(2), pp.741-755.

[37] Niinikoski, E., 2020. Using Random forests to predict total career performance of young Finnish trotters with early racing performance and other variables.

[38] Schultz, J., 2018. *Women's Sports: What Everyone Needs to Know®*. Oxford University Press.

[39] Ahire, P. G. ., & Patil, P. D. . (2023). Context-Aware Clustering and the Optimized Whale Optimization Algorithm: An Effective Predictive Model for the Smart Grid. International Journal on Recent and Innovation Trends in Computing and Communication, 11(1), 62–76. https://doi.org/10.17762/ijritcc.v11i1.5987

[40] Davis, W., Wilson, D., López, A., Gonzalez, L., & González, F. Automated Assessment and Feedback Systems in Engineering Education: A Machine Learning Approach. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/102

[41] Rohokale, M.S., Dhabliya, D., Sathish, T., Vijayan, V., Senthilkumar, N. A novel two-step co-precipitation approach of CuS/NiMn2O4 heterostructured nanocatalyst for enhanced visible light driven photocatalytic activity via efficient photo-induced charge separation properties (2021) Physica B: Condensed Matter, 610, art. no. 412902, .