

# AI BASED FEATURE SELECTION MODEL FOR SOCCER SPORTS MANAGEMENT

Karun Madan\*

Dr. Kavita Taneja\*

karanmadan9@gmail.com

Department of Computer Science and Applications, Panjab University, Sector 14, Chandigarh, India

Dr. Harmunish Taneja\*\*

Department of Computer Science and IT, DAV College, Sector 10, Chandigarh, India

## ABSTRACT

Due to swift development of data mining as well as machine-learning technology and the flare-up of big sports data mining expansion challenges, sports data mining cannot merely use data statistical methods such as how to club machine learning and data mining technology for efficient mining and analysis of the sports data, to supply useful advice for the public physical exercise, and this is an vital need to study. It is a kind of effective sports data mining work through feature selection algorithm. Around the tricky problems existing in the study of the sports effect, given the drawback of existing data sets and conventional research methods, this paper begins from data mining algorithm, construct the sports effect evaluation database, based on the feature selection scheme, using elastic system network algorithm, random forest algorithm, and the impact of sports on the outcome of physical gauges. The evaluation algorithm presents machine learning techniques and the feature selection algorithm to guide sports effect evaluation research. When studying this evaluation problem of the sports effect, according to created sports effect evaluation database, elastic system algorithm is appended to regularize, realize and optimize the feature selection. When selecting features of different sports skills using the information gains marked to rank the significance of characteristics, which can systematically and accurately provide the influence degree of the sports on diverse physical indicators, bring the physical fitness research little more scientific, and can uncover the effect of the sports as much as possible. Experimental results demonstrate that the selected features as well as ground-truth both have good accuracy and good evaluation as match up to the baseline method.

**Keywords:** Sports Management, Feature Selection, Data Mining, Accuracy, Machine-Learning.

## 1. INTRODUCTION

There is an accelerated accumulation trend in the data from many industries due to the Internet's quick development. In the era of big data, the proliferation of data in the fields of education,

healthcare, science, and finance has fueled the advancement of data mining and related technologies. It is clear that big data is crucial to the process of a country's growth, and people from all walks of life are focusing on the creation and

study of big data-related technologies in order to meet the demands of the modern world.

One significant component of big data resources is sports data. Sports efficacy and their effects on the human body can be efficiently understood through the mining and analysis of sports data. The emergence of big data sports has presented challenges for the development of sports data mining, given the rapid advancements in machine learning and data mining technologies. Current sports data mining techniques primarily concentrate on the creation and extraction of useful fundamental sports data features, while also employing statistical techniques to examine sports data. Sports data mining, however, cannot merely rely on data statistical techniques, such as combining machine learning and data mining technology for sports data mining and analysis, due to the rapid advancement of these technologies.

One significant area of big data analysis application is sports data mining. This sports data mining study employs an effective feature selection technique. The feature set that works best for system optimization is chosen for feature selection. [1]

The procedure contributes to the accuracy of the classification, these features can increase the system's classification effect, and feature selection can enhance model learning performance—a critical stage in pattern recognition [2].

Based on this context, this research presents an algorithm combining random forest and elastic network to choose the features of sports big data and study and assess various sports effects. This will enable the application of sports big data in sports impact evaluation. There are four sections throughout the entire text. The research background and necessity are introduced in Section 1, the paper's structure and random forest are explained in

Section 2, the theory and modeling process of the elastic network and random forest are explained in Section 3, and the significance of various sports is examined in Section 4, along with particular recommendations for sports training.

## 2. LITERATURE SURVEY

The statistical mean, standard deviation and basic correlation coefficient approaches are only a few of the research techniques that are currently being used in several domestic and international studies. The simple correlation coefficient approach is used to examine simply the influence relationship between the two indicators, but not the influence of other factors. When there is a correlation between the index characteristics, the research findings are not all-inclusive. At the same time, the physical fitness data is mainly the national physical fitness monitoring bulletin, investigation report, etc. The majority of the time and cross-sectional data are used to analyze the status of sports influence, but they are unable to demonstrate individual differences or provide statistically significant correlations.

Seldom are effective feature selection and data mining techniques employed, and even fewer are employed to examine how sports affect physical index data. Yu and colleagues used the ID3 algorithm to decision trees in pertinent domestic and international studies to analyze test data related to human grip and muscle strength. They were able to identify the root nodes of various test index parameters and derive the indicators that can be used to objectively assess human muscle strength [3]. Liu suggested utilizing artificial swarms to optimize the classifier in his optimized random forest technique. With comparatively high classification accuracy, the model is able to recognize patterns in human movements [4].

Additionally, some research studies the impact of physical fitness using sports data using statistical techniques.

In 2005, Jkman examined 11,407 adult data from the Shanghai National Physical Fitness Monitoring Database, focusing on individuals aged 20 to 39. Using data mining technology and relevant regulations, he processed and analyzed 21 significant physical indicators, as well as the function and quality of the research subjects, to determine the relationship between the physical indicators. The magnitude of handgrip strength is correlated with vital capacity, which influences handgrip strength, performance on the longitudinal jump, and other metrics; handgrip strength and body balance ability and body obesity degree are also correlated [5].

Ma and colleagues conducted research on the elements influencing university students' physical health and made the proposal that a conducive atmosphere for physical activity has a significant impact on students' ability to "exercise, enhancing students." Family support influences the growth of physical level, and physical awareness influences the enhancement of physical fitness. [6]

Feng et al. examined and analyzed the questionnaire survey results using two years' worth of physical test data from college students in this province [7]. Mei et al. performed a one-way variance analysis and descriptive analysis of four physical fitness indicators in Hebei Province: grip strength, sitting forward, one foot flexion, and responsiveness. They did this by mathematical statistics, literature study, and comparative research [8].

### 3. METHODOLOGY

Data mining technology has advanced quickly in the sports industry, and it has been extensively researched to include new theories into sports data mining technology. One common sports data mining problem is examining the impact of sports. The primary goal of data mining is to extract potentially valuable information from a vast amount of real-world application data through algorithmic mining. The use of data mining technology in the sports industry is known as sports data mining. Using picture, discrete, and video data, sports data mining technology is mostly employed in physical education instruction, sports training observation, and sports information administration. No appropriate public data set is available. Thus, the question is how to create representative sports data and then carry out an efficient and quick assessment of the sports effect.

The research team organized multiple objects for a period of time to train in wrestling, competitive foot, skills, and the four types of modern school sports. After more than 40 representative body forms, body functions, and physical quality indicators, and physical indicators changed as characteristic, the team observed object training. This allowed them to study the sports effect evaluation method. Initially, the research team separated sports into five groups: no sports, modern school sports, skills, competition, and wrestling. The purpose of the establishment of no sports is to represent how different sports affect the physical indications, not to create a unique sports exam.

The 785 pupils served as test subjects and were split into five groups based on five different sports categories. The research team looked at each group's physical indications prior to physical training, and the results were noted as P. Under the

supervision of designated people, each group will carry out the necessary training for three months during the physical training. Three times a week, for seven minutes each, you will prepare, exercise for thirty minutes, and then relax for three more minutes after work. At the conclusion of the last training session, the index data are recorded as Pi. The group used an electronic acoustic metronome, an electronic meter, a spirometer, a grip meter, and a reaction tester device to measure the team members' height, weight index, and sitting height.

Preprocessing of the data is necessary since there are uncertainties that may prevent us from obtaining the data we need, such as missing or duplicate values. After reviewing the literature, we discovered that data preprocessing, which is typically associated with the activity itself, does not have a single, standardized method. Additionally, different datasets typically require various attribute preprocessing methods. The following are typical data preprocessing procedures: data normalization, attribute coding, processing missing values, and removal of unique attributes. The preprocessing that we focused on was as follows in multiple sections. In order to eliminate unique attributes, one must first remove ID attributes from the data set, such as the "name" attribute in the data that was acquired, which is unable to characterize the sample's distribution pattern.

#### 4. RESULTS AND ANALYSIS

The sports database's SED, which describes the form's creation and usage in Section 3, provided the data for the experiment. A 4:1 ratio between the training and test sets was established in the database. In this research, the impact of four different sports on bodily markers is primarily examined. Sports data is obtained as positive, sports data is not obtained as negative, and the two

forms of data are compared. Four groups of trials comprise the experiments in this review: the first group consists of the data variance indexes.

The elastic network approach was used in conjunction with regularization theory to rank the attributes of the four sports effect kinds. It was discovered that the impacts of the four sports types varied on certain indexes. The ranking of the degree of angular force movement influence is shown in Figure 1, where the standing long jump, dorsal muscle strength, average grip strength, selection response time, and cardiac power index have the most influence. Wrestling really works the strength primarily, and the findings of the experiment support this theory. The random forest technique is another option.

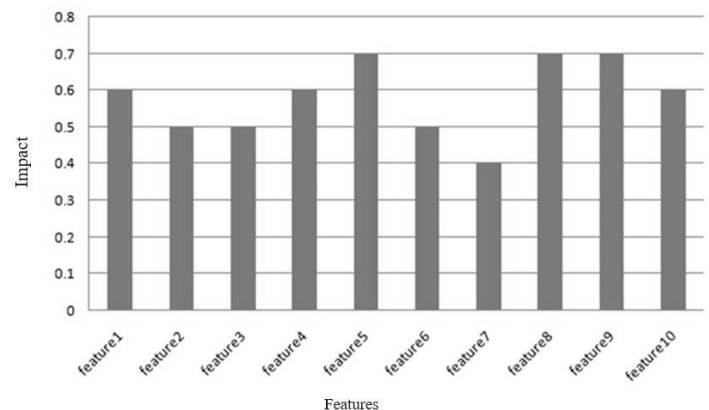


Figure 1: Uses the elastic network to analyze the influence degree of angular force movement on the body indicators.

Using simulation analysis and SPSS statistical analysis software, the effectiveness of the design model in implementing the sports skill assessment was confirmed. The sports skill ability evaluation is carried out in accordance with the simulation parameter settings mentioned above. Using big data analysis, the sports skill ability evaluation's big data mining findings are also considerable. The total movement ability declined as the number of

iterations increased and stayed below the benchmark line.

## 5. CONCLUSION

Due to the proliferation of sports big data, sport data can now be evaluated by relevant scholars in addition to being data mined and analyzed by pre-existing data sets using conventional methods. This paper begins with an analysis of the data mining algorithm, taking into account the constraints of the available data sets and conventional research techniques, and integrating the quickly evolving feature selection algorithm. In order to examine the impact of sports and direct sports, this paper suggests the random forest and elastic support vector network algorithms. The subjective evaluation and cognition of sports experts are in line with the experimental results. The two algorithm categorization findings are more accurate when compared to the baseline technique. In this way, we may select the appropriate exercise program based on the skills we wish to work on and the exercise outcomes, and we can also assist school physical education teachers in guiding their students in physical education classes. Simultaneously, the impact of the activity can be assessed using our sports exercise effect evaluation approach. The majority of young people should always be concerned about their physical health. They should understand the effects that all activities have on the body and know where they need to make adjustments, build new things, and stick to active, healthy sports while selecting the right ones.

## 6. REFERENCES

1. Li J., K. Cheng, S. Wang et al., (2016) "Feature selection: a data perspective," *ACM Computing Surveys*, vol. 50, no. 6.
2. Xu Y. U., W. Qian-Long, and X. U. Ling-Wei, (2019) "An efficient recommendation algorithm based on effective feature subset extraction," *Computer system application*, vol. 28, no. 7, pp. 162–168.
3. Yu. D., Y. Zhong, and Y. Yu, (2010) "Application based on data mining technology in the data analysis of human muscle strength. takes the study of human handgrip strength and muscle strength test data as an example," *Sport Science*, vol. 2, p. 6.
4. Liu Y., *Research on Human Movement Mode Recognition Based on Random Forest Algorithm*, (2018), Beijing University of Posts and Telecommunications, Beijing, China.
5. Jkman S. W., (2010) "Correlation analysis of body morphology, function, and quality indicators," *Journal of Zhoukou Normal University*, vol. 27, no. 5, p. 3.
6. Ma Q., L. Bai, and W. Liu, (2013) "Analysis on the physical condition and influencing factors of students in Liaoning university of science and technology," *Journal of Liaoning University of Science and Technology*, vol. 36, no. 6, p. 7.
7. Feng T. M., (2015) "Analysis on the physical health status and main problems of college students in Jiangxi province," *Contemporary Sports Technology*, vol. 35, p. 2.
8. Mei J., Y. He, and Y. Liu, (2017) "Analysis of physical fitness index in Hebei province," in *Proceedings of the 2017 Conference of Sports Physiology Committee of Chinese Physiology Association*, Rio de Janeiro, Brazil.