Hindawi

Retraction

Retracted: Prediction of Sports Performance and Analysis of Influencing Factors Based on Machine Learning and Big Data Statistics

Journal of Sensors

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] P. Wang, J. Liu, and B. Liao, "Prediction of Sports Performance and Analysis of Influencing Factors Based on Machine Learning and Big Data Statistics," *Journal of Sensors*, vol. 2022, Article ID 3276576, 9 pages, 2022.

Research Article

Prediction of Sports Performance and Analysis of Influencing Factors Based on Machine Learning and Big Data Statistics

Panpan Wang,[1] Jiangbo Liu [ID],[2] and Benlu Liao[1]

[1]*Ministry of Sport, Hebei GEO University, Shijiazhuang, 050031 Hebei, China*
[2]*Art Sports Institute, Hebei Institute of Communications, Shijiazhuang, 050000 Heibei, China*

Correspondence should be addressed to Jiangbo Liu; liujb@hebic.edu.cn

Performance prediction is one of the important contents of sports future development strategy research. This paper conducts research on sports performance prediction and influencing factor analysis based on machine learning and big data statistics. The purpose of forecasting research is to formulate short-term, medium-term, and long-term sports development planning and policy services for decision-makers of sports development. Advantages and disadvantages make sports workers consciously revise and realize their own sports future plans. First, the analysis of the influencing factors of sports performance is proposed, including the setting and scoring of sports examination items, research on standards, training methods for physical examination items, and physical examination learning. Second, the application of big data statistics based on machine learning algorithms is studied. Based on machine optimization algorithms, it summarizes the naive Bayesian classification algorithm and big data statistical classification algorithm. The theoretical idea of rough set algorithm and finally through the comparative analysis experiment based on big data statistics of male and female sports test scores and the comparison experiment of the accuracy of three algorithms to predict the results, it is concluded that the endurance quality of boys is better than that of girls, which indicates the quality of sustainable development. There is a certain gender gap, there are significant differences in the skills of male and female students, and the machine-optimized algorithm can predict sports performance with the highest accuracy among the three algorithms.

1. Introduction

Machine getting to know methods, a household of statistical strategies originating in the area of synthetic intelligence, are viewed to preserve awesome promise for advances in the appreciation and prediction of ecological phenomena. These modeling methods are bendy adequate to manage complicated issues with a couple of interacting factors and regularly outperform standard methods, making them perfect for enhancing ecosystems [1]. One of the most leaping boundaries in AI lookup is in the area of laptop learning. This includes exploration to boost computational fashions of the mastering process. The fundamental cause of lookup in this location is to construct computer systems that can enhance their overall performance in the area and independently gather records [2]. One of the most active areas of artificial intelligence research is the field of machine learning. The

main goal of research in this area is to build computers that can improve their performance in the field and acquire information independently [3]. Boosting is a general method to improve the accuracy of any learning algorithm. It mainly focuses on the AdaBoost algorithm, including analyzing the AdaBoost training error and generalization error, strengthening the connection with game theory and linear programming, and improving the relationship with logistic regression [4]. Statistical machine learning methods further reveal qualitatively different trade-offs between small-scale and large-scale learning difficulties. Large-scale instances contain the computational complexity of the underlying optimization algorithm, and much less probable optimization algorithms (such as stochastic gradient descent) are proposed in big problems with shocking overall performance [5]. In contemporary implementations, computational fashions primarily based on first concepts of physics

can notably gain from the latest explosion of facts science. In fact, these two branches of utilized arithmetic can engage benignly to a massive extent they already do [6]. The improvement of huge statistics will turn out to be a principal disruptive innovation in the manufacturing of authentic statistics, bringing a sequence of opportunities, challenges, and dangers to the work of the National Statistical Institute (NSI) [7]. Big data is becoming more and more important as an additional data source for official statistics (OS), the contribution of machine learning lies in the overview of big data technologies related to the operating system, and IT is derived from the actual projects we carry out at Istat Problem description and classification [8]. The hidden-missing-artificial-utility algorithm handles transactions to hide sensitive itemsets through transaction deletions. It considers three facet results of hidden failures, lacking itemsets, and synthetic itemsets for evaluating whether or not processed transactions want to be deleted at some stage in cleanup [9]. This meta-analysis investigated two relationships in aggressive sports: country cognitive anxiousness and overall performance and kingdom self-belief and performance. Paired sample t-tests confirmed that the suggested impact measurement of self-confidence was once extensively large than the implied impact dimension of cognitive anxiety. The moderators of the cognitive anxiety-performance relationship had been gender and opposition standards [10]. In sports, the distinction between the domestic and away ratings is modeled as a Brownian movement procedure described at $t \in (0, 1)$ that obtains the domestic team's lead (or lag) l at time t with the domestic crew prevailing A easy relationship between possibilities [11]. In order to set up a scientific and truthful assessment approach for bodily schooling colleges, thinking about the distinction in the significance and problem of courses, the fuzzy analytic hierarchy technique is used to decide the weight coefficient of every course, and the gray relational evaluation is used to assemble a complete comparison system. Evaluate the model. Practice indicates that this approach is greater scientific than different techniques [12]. Construct a structural mannequin of university students' bodily schooling overall performance in one-of-a-kind educational year via structural equation evaluation and instructing data, and then describe the relationship between the models that have an impact on bodily training overall performance in educational years and quantify the relationship between university students' bodily training overall performance in one-of-a-kind educational years. Finally, some positive guidelines on university physique are put ahead [13]. Physical schooling can play a doubtlessly essential position in improving public fitness with the aid of fostering advantageous athletic attitudes and advertising health-related health programs, emphasizing the significance of perceived competence and intrinsic motivation in obligatory sports activities [14]. A cross-context mannequin proposes that adolescents' perceived self-reliant guide in bodily training influences their perceived causality, intentions, and bodily undertaking behaviors at some point of entertainment time. Perceived autonomic aid in bodily training without delay and not directly influences leisure-time bodily exercise thru motivational sequences involving

inside perceived causality, attitudes, perceived behavioral control, and intention [15].

2. Analysis of Factors Affecting Sports Performance

2.1. Research on the Target Layout and Scoring Criteria of Physical Education Degree. From the perspective of educational measurement, Ma Ying pointed out that the test difficulty of physical examination subjects should be around 0.5 and the discrimination degree should be above 0.3. Wang Quanhui suggested that there should be greater depth and difficulty in the design of test centers and the formulation of classification standards, so that it is easier to distinguish the level and admission ability of candidates. Yang Zongyou and Li Zhenzhong analyzed the results of sports candidates in Chongqing from 2011 to 2014. There is a strong correlation between the 100-meter run, the 5-meter three-way run, the medicine ball thrown in front, and the long-ball standing. Skip to Chongqing medical examination target. Change the original quality testing method and add special quality testing. For the validity and economy of the test, it is recommended to reset the physical condition of the exercise to more effectively measure the physical fitness of the examinee. Li Yuling analyzed the physical education achievements in Hubei Province from 2006 to 2008 and pointed out that there are significant differences between men and women in the total score and individual score. The overall score of boys is much lower than that of girls, and the scores of girls in individual destinations are relatively low. Wang Guanxiong believes that the weight of basic physical quality is higher than that of special qualities, which is not conducive to the development of sports ability of one specialization and multiple abilities. The method of determining the score only according to the arithmetic sequence of the next paragraph cannot objectively reflect the law of athletic ability to take the test.

2.2. Research on the Culture Method of Physical Examination Objects. Wu Bin pointed out in the article that grassroots coaches need to update their training concepts, not only focusing on special movements and strength training but also focusing on strengthening speed, endurance, flexibility and agility, and other physical training. In addition to the traditional training of waist, abdomen, and limbs, the coaches also pay attention to nuclear energy training and carry out targeted training according to the situation of different athletes. Jiang Jin pointed out in the article that the vertical triple jump is a compulsory item in some provinces and cities. From a technical standpoint, pauses are more complex and require more than just good strength, speed, coordination, and jumping abilities of the candidate. The pace of the applicant is also required. Standing triple jumps and triple jumps are key and manage the connection between the triple jumps. In addition to good physical fitness, applicants must also have normal technical movements in order to better display their technical level. In order to get a high score, candidates must strengthen their strength in standard practice; throwing medicine balls with both hands on the spot is

also a project with short training time and more complex technical movements. Zhu Rui pointed out in the article that running 800 meters requires stable attention and easy and natural movements and rhythms, and general physical fitness should be developed in the basic training stage, and continuous training should be used. Introduce interval training. The above researchers are all from the perspective of sports events. Most of the training strategies proposed for a certain event are theoretical, and there are few studies on the arrangement of individual stages.

2.3. Research on PE Exam Study. The cultivation of physical education candidates and the discussion on the cultural course study of physical education candidates provide theoretical support for the renewal and development of sports in our country and have far-reaching significance. In response to sports issues, many scholars express their opinions and strive to find a more scientific, more reasonable, and fairer way. My country has made great progress in formulating sports content and viewing standards. Although there are still some gaps in foreign countries, we can learn from their advanced experience, remove the rough and extract the essence, and continuously improve. With the continuous development of society and the development of school sports in our country, it is believed that some problems in sports will be gradually solved. In view of this situation, this paper uses statistical software and statistical data to analyze and test the test scores of Chongqing badminton super players and reveals the rules of the quality test and special test of Chongqing badminton super players. Get out of the shortcomings and put forward detailed and effective countermeasures, hoping to provide some reference for the sports department of Chongqing, and make sports in Chongqing more objective, scientific, fair, rational, and effective.

3. Research on the Application of Big Data Statistics Based on Machine Learning Algorithm

3.1. Fundamentals of Machine Learning. After building the model hypothesis space, machine learning should consider the criteria for learning or selecting the optimal model. The loss function usually measures the quality of the first prediction in the model, the hazard function measures the expected value predicted by the model given the input, and the difference between the predicted value and the actual value corresponding to the model output selected from the hypothesis space is stored as the loss function. The formula of the commonly used functions is as follows:

$$\Gamma(y, f(x; \theta)) = \frac{1}{2}(y - f(x; \theta))^2. \tag{1}$$

The lower the value of the loss function, the better the performance of the model. Each sample can be considered to be independently randomly drawn from a common space according to an unknown common distribution, and the quality of the model can be measured by a hazard function

or expected risk as follows:

$$\Gamma(y, f(x; \theta)) = -y^T \log f(x; \theta). \tag{2}$$

In general, the lower the expected risk, the better the performance of the model. If the training set contains N independent and random samples, the average loss of the model relative to the training set D is called empirical risk, and the formula is as follows:

$$R_D^{\mathrm{emp}}(\theta) = \frac{1}{N} \sum_{n=1}^{N} \Gamma(y^n, f(x^n; \theta)). \tag{3}$$

Therefore, a feasible learning criterion is to find a set of parameters 0 that minimize the empirical risk, namely,

$$\theta^* = \arg \max R_D^{\mathrm{emp}}(\theta). \tag{4}$$

3.2. Machine Optimization Algorithms. Because laptop studying frequently has many mannequin parameters and a massive quantity of academic data, it is not possible to use the secondary optimization technique with a giant quantity of calculation, and the principal secondary optimization technique is typically much less fine in education. In order to use some mature and environment friendly optimization techniques in convex optimization theory, the most often used optimization algorithm is the gradient calculation method, and the formulation is as follows:

$$\theta_{t+1} = \theta_t - \alpha \frac{\partial R(\theta)}{\partial \theta} = \theta_t - \alpha \frac{1}{N} \sum_{n=1}^{N} \frac{\partial \Gamma(y^n, f(x^n; \theta))}{\partial \theta}. \tag{5}$$

This is the most primitive form of gradient landing, which corresponds to the estimated risk to empirical risk approximation of samples independent of the actual data distribution, but when the training samples are large, each iteration is computationally intensive and the training process is slow. To solve this problem, stochastic gradient descent only computes the gradient of randomly selected samples and uses it to update the parameters updated at each iteration. For the tenth iteration, the update method is as follows:

$$\theta_{t+1} = \theta_t - \alpha \frac{\partial \Gamma(y^n, f(x^n; \theta))}{\partial \theta}. \tag{6}$$

Stochastic gradient descent is easy to implement, has fast convergence speed, and is widely used. So in the 10th iteration, a subset S of K samples is randomly selected, the gradient of the loss function is calculated and averaged for each sample in the subset, and the gradient is updated:

$$\theta_{t+1} = \theta_t - \alpha \frac{1}{K} \frac{\partial \Gamma(y^n, f(x^n; \theta))}{\partial \theta}. \tag{7}$$

The small gradient descent method has gradually become the main optimization algorithm for large-scale

machine learning because of its fast convergence speed and low computational cost.

3.3. Naive Bayesian Classification Algorithm. Naive Bayesian fashions have secure classification efficiency. It performs nicely on small-scale data, can cope with multiclassification tasks, and is appropriate for incremental training, specifically when the quantity of information exceeds the memory, it can be deincremental coaching in batches. It is much less touchy to lacking data, and the algorithm is highly simple and is regularly used in textual content classification. The naive Bayes classification algorithm is based on the development of probability and statistics and is a typical classification method. The main point is to use previous knowledge or experience to get the previous probability of an event and then use the Bayesian formula to calculate the post probability based on the previous probability, that is, the probability to which a particular thing belongs. The category corresponding to the highest probability value is determined as the category of the current case. Let h be a random event, $P(h)$ represents the condition of h, and if D represents the training sample, then $P(D)$ is the prior probability obtained by D. In this case, assuming h is true, the probability is that D can be detected, and the result obtained is called the conditional probability of D. However, in this calculation, the probability value of h is calculated based on D, the posterior probability of h. By adjusting the previous probability, the data can show the probability that h will be determined later.

$$P(h|D) = \frac{P(h|D) \times P(h)}{P(D)}. \tag{8}$$

Equation (8) is a complete Bayesian formula. Its naive nature is to assume that each condition is independent of each other, use a preprocessed and selected set of exercises, pour it into a naive Bayesian model, practice and learn the common probability distribution of X and Y, and finally accept the model for training, such that Y with posterior probability is obtained by condition X as input. Equation (9) can be used to calculate the final conditional probability; that is, if the attribute element t appears in a given document x, it can be classified as the conditional probability of class c.

$$P(c|x) = \frac{P(c)\prod_{i=1}^{n}P(t_i|c)}{\prod_{i=1}^{n}P(t_i)}. \tag{9}$$

In formula (9), $P(c)$ represents the probability of class c that can be obtained from the statistics of a complete data set, and $P(t/c)$ is a conditional probability, which represents the probability of adjective t in the text class c. The last n represents the number of all attribute items in document x. As mentioned above, $P(c)$ and $P(t/c)$ should be calculated using statistical methods so that maximum probability estimates can be used to find them. Using the maximum probability estimation, the preprobability $P(c)$ can be obtained from NN, where N represents the total number of documents of class c in the entire training sequence and the total

number of documents in the entire training, and the calculation of $P(t/c)$ is shown in formula (10):

$$P(t_i|c) = \frac{T_{ct}}{\sum T_{ct}}. \tag{10}$$

After the model is taught through a series of exercises and final parameters are determined, the model is predicted in a series of tests, the model's predicted class value is compared to the actual class value, and an evaluation performance evaluation index is calculated to evaluate the model. Typically, all or part of the three measurements accuracy, memory, and $F1$ are used to measure and evaluate the model. Accuracy represents the ratio of the number of sample logs to the total number of samples in the test set when testing the model with the labeled test set. Recovery represents the relationship between the number of samples that the model can correctly determine when tested on a labeled test series and the number of samples in that test series that actually belong to that class. The $F1$ value takes into account both the accuracy and the recovery speed, and the specific calculation takes the average between the two. In text classification, the defined categories are usually divided into positive examples P and negative examples N, and the formula for calculating the accuracy speed is given by Equation (10). The efficiency calculation formula is shown in formula (11), and the evaluation index mainly measures whether the model does not match the inspection data samples of other categories.

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{11}$$

The formula for calculating the $F1$ value is shown in formula (12), and the evaluation index represents a comprehensive measure of the classification effect of the model.

$$F = \frac{2 \times P \times R}{P + R}. \tag{12}$$

3.4. Rough Set Algorithm of Big Data Statistical Classification Algorithm. The basic idea of classical rough set theory is the granulation and approximate data analysis method based on equivalence relation. The core foundation of tough set principle and software is a pair of approximation operators derived from the approximation space, specifically the top approximation operator and the decrease approximation operator (also acknowledged as higher and decrease approximation sets). Currently, there are two predominant lookup techniques to outline approximate operators: optimistic techniques and axiomatic methods. The development technique takes the binary relation, division, coverage, nearby system, Boolean subalgebra, etc. on the universe as the primary factors to outline the tough approximation operator, thereby deriving the difficult set algebra gadget. The original research continued to follow the traditional idea of reducing complexity to simplify uncertainty and become deterministic. Therefore, the uncertainty problem can be expressed as a set of lower approximation sets

and roughly upper approximation sets, which can be precisely defined as follows:

$$\underline{R}X = \{x \in U \,|\, [x]_R \subseteq X\}, \tag{13}$$

$$\bar{R}X = \{x \in U \,|\, [x]_R \cap X \approx \Phi\}. \tag{14}$$

And since the accuracy is inversely proportional to the size of the boundary region, the model is evaluated in terms of approximate accuracy and approximate classification quality, which is determined by

$$\alpha_R(X) = \frac{|\underline{R}X|}{|\bar{R}X|}. \tag{15}$$

The approximate classification quality is defined by the following: $\alpha_R(X)$ represents the percentage of correct decisions in the decisions made by classifying data elements using knowledge, and $\gamma_R(X)$ represents the percentage of data elements that can be correctly classified into predetermined classes using knowledge.

$$\gamma_R(X) = \frac{|\sum_{i=1}^n \underline{R}X_i|}{|U|}. \tag{16}$$

Deterministic factor expressions define rules by

$$\mu(X_i, Y_i) = \frac{|Y_i \cap X_i|}{|X_i|}. \tag{17}$$

Discriminant matrix is another form of knowledge representation, and it can easily calculate the reduction and kernel of attributes, so Equation (13) is the set of all attributes that distinguish objects and sums, and the expression is as follows:

$$\Delta = \prod_{(x,y) \Xi U \times U} \sum a(x,y). \tag{18}$$

It is usually implemented by a wrapping method, that is, a rough set is used to select attributes, and finally, the approximate accuracy and approximate classification of the model are calculated.

The approximate accuracy of the mass model is expressed as

$$\alpha_{a_1, a_2, \cdots a_n}(V_{y1}) = \frac{|\underline{R}V_{y1}|}{|\bar{R}V_{y1}|}. \tag{19}$$

The approximate classification quality is expressed as

$$\gamma_{a_1, a_2, \cdots a_n}(V_{y1}) = \frac{|\underline{R}V_{y1}|}{|\bar{R}V_{y1}|}. \tag{20}$$

3.5. Classification Algorithm Based on Big Data Statistics. There is no saturation area, there is no gradient disappearance problem, there is no complicated exponential operation, the

calculation is simple, and the effectivity is improved; the genuine convergence velocity is faster, a whole lot quicker than Sigmoid/tanh; it is greater in line with the organic neural activation mechanism than Sigmoid. A random variable obeying a normal distribution is called the central limit theorem in probability theory. A random variable whose distribution function satisfies for any x:

$$\lim F_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \tag{21}$$

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}. \tag{22}$$

Rating models use customer reviews as an expression for ratings. Since the essence of classification is to establish a model to realize the mapping from the conditional attributes of data elements to the class identification attributes of data elements, the main idea of the scoring model based on statistics is

$$y = x_1 + x_2 + \cdots + x_n. \tag{23}$$

The statistical scoring model is

$$x_i' = \frac{Y - q \times (Y + N)}{\sqrt{q \times (1 - q) \times (Y + N)}}. \tag{24}$$

4. Prediction and Analysis of Sports Test Scores Based on Big Data Statistics

4.1. Selection of Experimental Subjects. This paper takes the physical examination results as the research object. In order to more intuitively reflect the changing trend of the physical examination performance indicators, boys and girls are required to run 100 m, 200 m, 400 m, and 4 × 100 m, respectively, and the physical examination results are statistically and predicted.

This paper uses the DPS data processing system to theoretically predict the average grades of men and women. As can be seen from Figure 1 and Tables 1 and 2, the fixed base ratio and chain ratio are based on the average of the test scores of male and female students in 2016. The participation rate of Luoyang sports test showed a trend of first decreasing and then increasing year by year. In 2017, the endurance quality of boys and girls deteriorated, and the decline was smaller. In terms of endurance quality, the increase in endurance quality of girls was slightly higher than that of boys, reflecting the gender gap in the increase in endurance quality.

4.2. Analysis of Different Factors of Physical Examination Results. As can be viewed from Table 3, the null speculation is rejected due to the fact the P value for this check is an awful lot much less than 0.05; i.e., the numbers are appreciably exceptional between the three years. Therefore, the outcomes of the evaluation right here have to be precisely equal to the preceding assessments of the ordinary model. Therefore, it suggests that there is a giant distinction in the
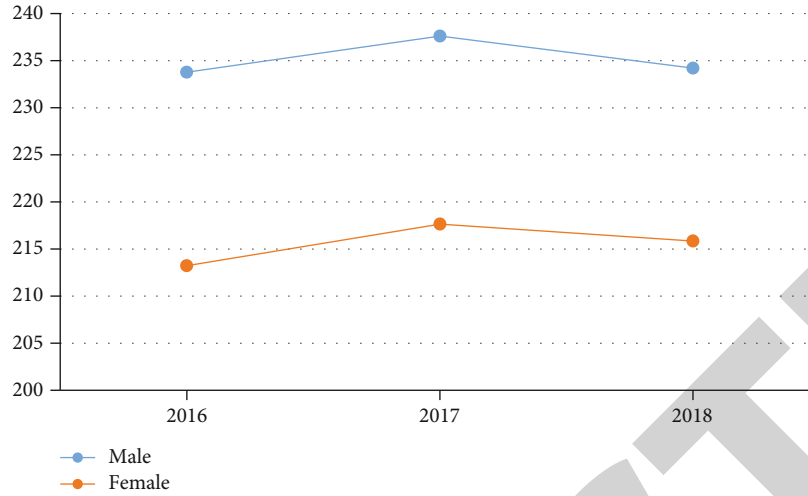
Figure 1: Dynamic graph of the average running of men and women.

Table 1: Men's running average performance prediction.

Project	Predictive model	Predictive value
100 meters	$x(k+1) = 2016.509081\mathrm{e}^{(0.005202\mathrm{k})} - 2005.929092$	10.67
200 meters	$x(k+1) = 8241.187274\mathrm{e}^{(0.002576\mathrm{k})} - 8220.837272$	21.42
400 meters	$x(k+1) = -75010.510001\mathrm{e}^{(-0.000631\mathrm{k})} + 75058.451$	47.31
4×100 meters	$x(k+1) = -4969.967165\mathrm{e}^{(-0.008248\mathrm{k})} + 5010.83717$	39.82

Table 2: Women's running average performance prediction.

Project	Predictive model	Predictive value
100 meters	$x(k+1) = -1246.701956\mathrm{e}^{(0.009661\mathrm{k})} + 1258.5219756$	11.65
200 meters	$x(k+1) = 0.171728\mathrm{e}^{(0.314068\mathrm{k})} - 0.087293$	24.17
400 meters	$x(k+1) = 23212.099807\mathrm{e}^{(0.002365\mathrm{k})} - 23157.749807$	55.35
4×100 meters	$x(k+1) = -0.345766\mathrm{e}^{(-0.846651\mathrm{k})} + 0.460778$	45.57

Table 3: One-way ANOVA of men's 1000 m running (s).

Source	Type III sum of squares	Df	Mean square
Calibration model	63214.631[a]	3	31773.421
Intercept	1765102360	2	1765102360
Error	43924953.04	34287	1452.053

Table 4: One-way analysis of variance for women's standing long jump (m).

Source	Type III sum of squares	Df	Mean square
Calibration model	14.587[a]	3	7.542
Intercept	102359.257	2	102359.257
Error	1489.627	31648	0.132

common values of the boys' 1000-meter run in 2016, 2017, and 2018, so more than one comparison of the common values is required. In statistics, df refers to the wide variety of impartial or freely altering records in the pattern when the parameters of the populace are estimated by using the statistic of the sample, which is known as the diploma of freedom of the statistic. In general, ranges of freedom are equal to the impartial variable minus the quantity of its derivatives. For example, the variance is described as the pattern minus suggest (a spinoff decided by using the sample) and for this reason has N-1 tiers of freedom for N random samples. The P price is a parameter used to decide the outcomes of a speculation test and can additionally be in contrast in accordance to one-of-a-kind distributions the use of the rejection location of the distribution. The P fee is the chance of a pattern remark or extra severe consequences when the null speculation is true. If the P fee is small, it has the capacity that the chance of the incidence of the null speculation is very small, and if it occurs, in accordance to the precept of small probability, we have a purpose to reject
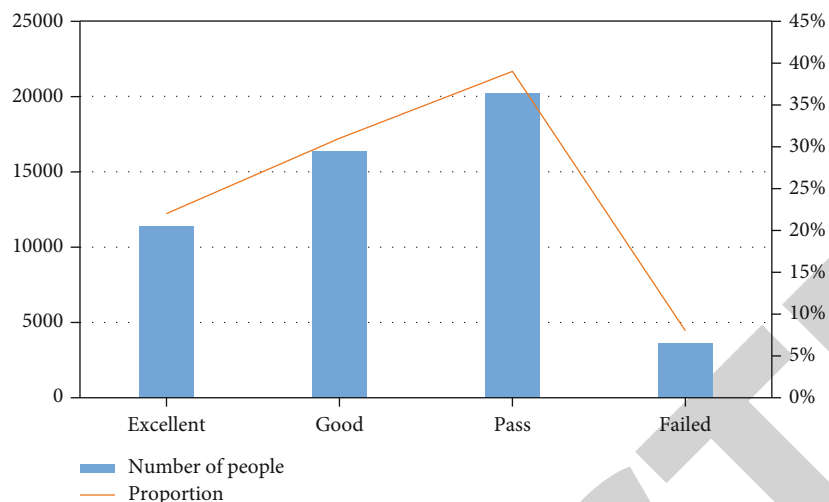
FIGURE 2: The excellent rate of the total score in the physical exam.
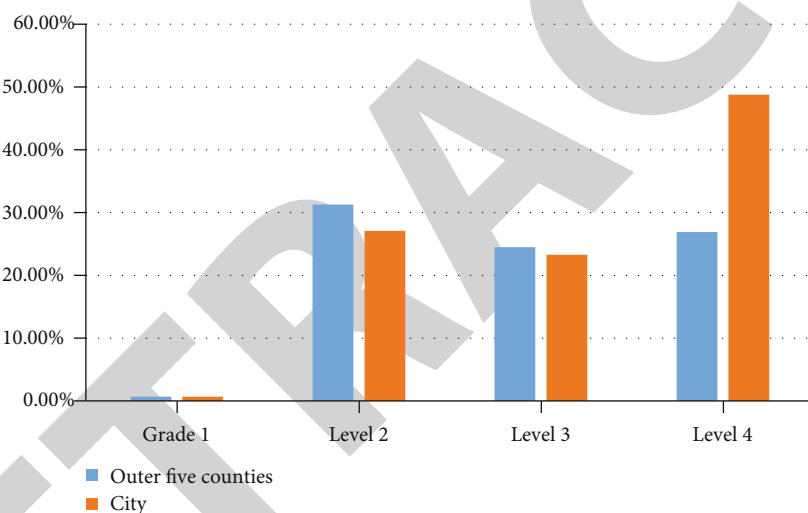

FIGURE 3: Comparison of grades and classifications of boys' sitting forward flexion test scores in the two regions.

TABLE 5: Prediction mannequin parameters of physical health.

Parameter name	Parameter value
The quantity of neurons in the enter layer	7
The wide variety of neurons in the output layer	2
The wide variety of neurons in the hidden layer	12
Threshold	0.004
Study efficiency	0.2
The most variety of iterations	1.0e20
Machine optimization algorithms	rprop+
Big data statistical algorithms	SSE
Naive Bayes algorithm	Logistic

the null hypothesis. The smaller the P value, the better the cause for us to reject the null hypothesis.

As shown in Table 4, the P value of this test is much less than 0.05, indicating that there is a significant difference in the means of the women's strength event (permanent long jump) in 2017, 2018, and 2019. Therefore, multiple comparisons of means are required.

4.3. Statistics on the Basic Situation of Physical Education Examinations. The results of the physical examination were divided into four grades: excellent, good, pass, and fail. It was concluded that the physical fitness examination reached the excellent level and accounted for 18.96% of the total number of people and 32.95% of the good level. Among the total number of students who passed the standard, 41.21% reflected that nearly 50% of the students in our city had unsatisfactory test scores, indicating that the overall physical condition of the students was not optimistic and polarized (see Figure 2).

For the boys who choose the prone body as an optional target, the comparison table of subjects in the two districts shows that the sports performance of boys in urban areas is higher than that of other boys. It shows that urban boys
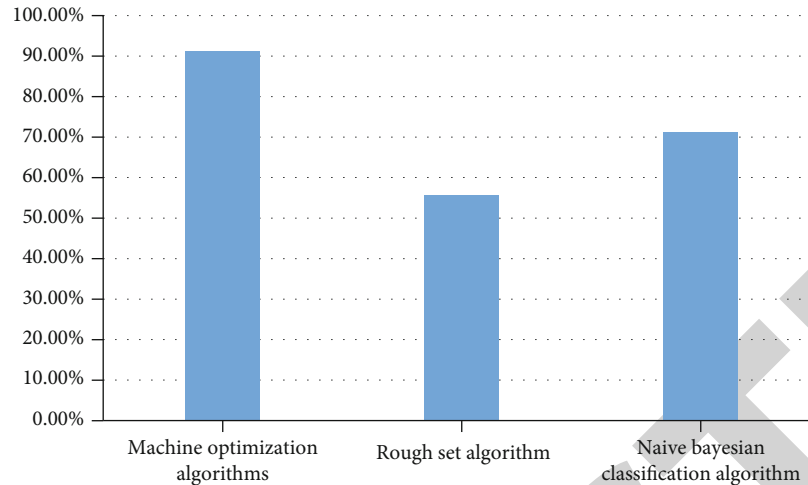
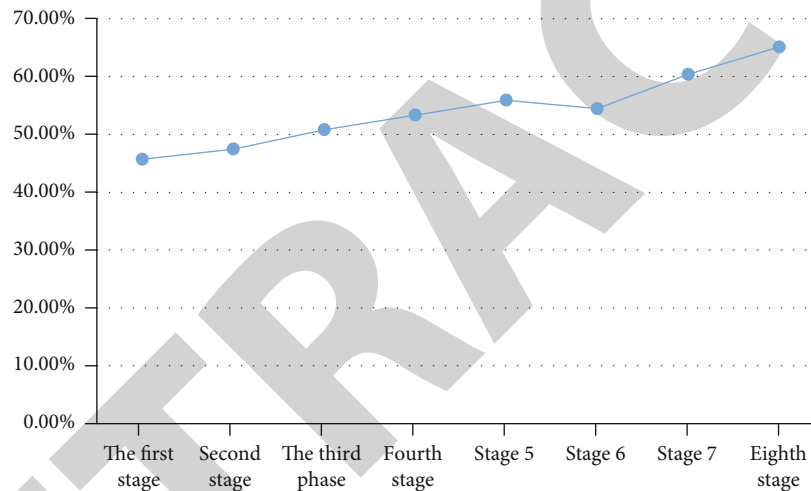FIGURE 4: Comparison of the prediction accuracy of the three algorithms.

FIGURE 5: Accuracy rate of stage performance prediction model.

have better sports performance. The overall physical quality is higher than that of boys in other provinces and cities. There is little difference in the subject classification between the two areas of sitting forward flexion. The ratio of the first level of sitting forward flexion in both areas is 0.5%, indicating that the boys who choose the sitting posture forward flexion have chosen the items they have. The excellence of the property is high, as shown in Figure 3.

4.4. Prediction Model of Physical Education Test Scores. The bodily health take a look at statistics is preprocessed and normalized, the preliminary information is modified with the aid of essential aspect evaluation method, and then, a BP neural community is used to generate a model. In this project, a neural network model is built using the neuralnet package, in which boys and girls have different scoring standards and methods, so the test data of boys and girls are separated, and models are generated separately for prediction. The model is continuously adjusted and optimized using 80% of the training set, and the final model parameters are shown in Table 5. The fundamental precept of the RPROP

algorithm is as follows: first, assign an preliminary cost to every weight change, set the weight exchange acceleration component and deceleration factor, and in the community feedforward iteration, when the nonstop error gradient signal does now not change, the acceleration approach is adopted to velocity up the education speed. When the signal of the nonstop error gradient changes, a deceleration approach is adopted to anticipate secure convergence. The community combines the cutting-edge error gradient image and the alternate step dimension to recognize BP. At the equal time, in order to avoid oscillation or underflow of community learning, the algorithm requires placing the top and decrease limits of the weight change; SSE is regularly used as the abbreviation of "sum of squares for error." In monadic regression, we use the weighted residual sum of squares of the basic factors to replace the method of screening eigenvalues only by residual sum of squares, which greatly improves the systematic error and the model has higher accuracy. The accuracy has been greatly improved; there is a law of development in nature: in the early stage of its development, the number or scale increases faster

and faster, and at a certain period, its growth rate gradually slows down, and the final number or scale no longer grows, thus stabilizing at the limit of number or size.

4.5. Evaluating the Performance of the Three Algorithms. For machine optimization algorithm and rough set algorithm, it is different from naive Bayesian algorithm, so the divided training set is divided into 7 : 3. For the three decision tree algorithms of growth set and pruning set, machine optimization, naive Bayes, and original set, respectively, build the performance prediction model of the final stage from the same training data set, and then, use the same test data to test the performance prediction of three different models. The performance setting evaluates and selects the algorithm with the best performance on this data set. As shown in Figure 4, it is a comparison chart of the prediction accuracy of the three algorithms.

The accuracy of the stage performance prediction model is shown in Figure 5. The figure shows that with the increase of the number of learning behavior characteristics in each stage, the accuracy of the stage performance prediction model generally shows an upward trend.

5. Conclusion

In the context of big data, this paper conducts dynamic analysis and research on the results of physical education surveys from 2016 to 2018,and researches on project settings, scoring standards, and the distinction of examination subjects to improve students' test scores and improve students' physique. The strategy is more scientific. Scientifically and reasonably formulate effective exercise methods for students to exercise, practice scientifically, teach students according to their aptitude, fully mobilize the enthusiasm and enthusiasm of students and teachers for sports, make students act consciously, love exercise, and teachers actively explore. Thus, a new way to improve the physical quality of students is realized. This paper also builds a phase performance prediction model based on the performance of the three algorithms. Experimental results show that a machine-optimized algorithm may have the highest accuracy of the three algorithms in predicting athletic performance. Given that there are many deficiencies in professional knowledge in the understanding of sports examination-related policies and the emerging information technology method of "big data," this paper only conducts dynamic analysis and research on sports examination results in the context of big data. Through in-depth mining of effective information in the data, it finds the existing problems and influencing factors and then finds out strategies to improve students' test scores and enhance students' physical health. There is no in-depth research on the setting of physical examination items, whether the scoring standards are reasonable, and whether the test items have a good degree of distinction between male and female students.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] K. Rieck, "Similarity measures for sequential data," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, no. 6, pp. 322–379, 2009.

[2] M. S. Chen, K. L. Wu, and P. S. Yu, "Optimizing index allocation for sequential data broadcasting in wireless mobile computing," *Knowledge & Data Engineering IEEE Transactions on*, vol. 15, no. 1, pp. 161–173, 2003.

[3] D. M. Hawkins and D. F. Merriam, "Optimal zonation of digitized sequential data," *Journal of the International Association for Mathematical Geology*, vol. 5, no. 4, pp. 389–395, 1973.

[4] A. J. Robinson, "Recognition of sequential data using finite state sequence models organized in a tree structure: US," *US5983180 A*, vol. 9, no. 2, pp. 34–37, 2004.

[5] X. He, Z. Zhang, and A. Kusiak, "Performance optimization of HVAC systems with computational intelligence algorithms," *Energy and Buildings*, vol. 81, pp. 371–380, 2014.

[6] T. Fukuda and T. Arakawa, "Computational intelligence in robotics and automation," *Computer Standards & Interfaces*, vol. 21, no. 2, pp. 17–23, 1997.

[7] L. Bottou, "Large-scale machine learning with stochastic gradient descent," *Physica-Verlag HD*, vol. 23, no. 5, pp. 23–78, 2010.

[8] A. Quarteroni, "The role of statistics in the era of big data: a computational scientist' perspective," *Statistics & Probability Letters*, vol. 136, no. 1, pp. 63–67, 2018.

[9] R. Kitchin, "Big data and official statistics: opportunities, challenges and risks," *Social Science Electronic Publishing*, vol. 1, no. 3, pp. 355–457, 2015.

[10] G. Barcaroli, S. D. Francisci, and O. M. Scannapi Ec, "Dealing with big data for official statistics: IT issues[C]," *Meeting on the Management of Statistical Information Systems (MSIS 2014)*, vol. 56, no. 34, pp. 1–18, 2014.

[11] C. Lury and A. Gross, "The downs and ups of the consumer price index in Argentina: from national statistics to big data[M]," *Springer Berlin Heidelberg*, vol. 4, no. 5, pp. 669–671, 2014.

[12] T. Woodman and L. Hardy, "The relative impact of cognitive anxiety and self-confidence upon sport performance: a meta-analysis," *Journal of Sports Sciences*, vol. 21, no. 6, pp. 443–457, 2003.

[13] S.,. S. Hal, "A Brownian motion model for the progress of sports scores," *Journal of the American Statistical Association*, vol. 46, no. 3, pp. 255–271, 2012.

[14] P. P. Gorman, R. J. Butler, M. J. Rauh, K. Kiesel, and P. J. Plisky, "Differences in dynamic balance scores in one sport versus multiple sport high school athletes," *International Journal of Sports Physical Therapy*, vol. 7, no. 2, pp. 148–153, 2012.

[15] Y. H. Zhang, "Comprehensive evaluation model and application of sports scores in college students," *Journal of Henan Normal University (Natural Science Edition)*, vol. 40, no. 4, pp. 29–592, 2012.