Research Article

Optimization of Human Resource Management System Based on Data Mining Technology and Random Forest Algorithm

Dicao Tang

Huanghe Jiaotong University, Jiaozuo Henan 454950, China

Correspondence should be addressed to Dicao Tang; 19403651@masu.edu.cn

Received 23 June 2022; Revised 18 July 2022; Accepted 22 July 2022; Published 8 August 2022

Academic Editor: Kalidoss Rajakani

Copyright © 2022 Dicao Tang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Talent is the first resource, and the role of talents in an enterprise is self-evident. Retaining key talents is crucial to the development of an enterprise. However, traditional data statistics methods cannot meet the needs of the government or enterprises in talent introduction, training, and reserve. For this reason, data mining technology and random forest algorithm are used to discover the hidden laws in the human resource management system and provide the government or enterprises with information. Decision support has become an urgent concern of the government and enterprises. This paper mainly studies human resources analysis and forecasting methods based on data mining technology and random forest. First, the data mining technology, naive Bayesian algorithm, decision tree algorithm, and random forest algorithm used in the research are introduced, and finally, the paper concludes this research.

1. Introduction

The human resource management system, as the name suggests, is developed and designed from the perspective of human resource management. All human resource-related content, such as recruitment, approval, salary, and other modules, is collected in the system through data collection and analysis [1]. It facilitates the daily work of human resource managers. The main reason why enterprises adopt human resource management system is that they expect to use human resources to the best economic benefits through human resource management system. Also, due to the advent of knowledge economy, the concept of so-called human capital has been formed, and the importance of human capital is no less than land, plant, equipment, and capital, or even beyond. In addition, people are the carrier of knowledge. In order to effectively use knowledge, to give full play to knowledge, proper human resource management is needed to give full play to the best utility of human resources.

The human resource management system has gone through three stages from its birth to the present:

(1) The development of human resource management system began in the late 1960s, which is almost the same as the development process of computer technology [2]. However, due to the immature development of computer technology at that time, the human resource management system in the initial stage can only realize very basic computing functions

(2) In the 1970s, with the computer in the database and application [3], great achievements have been made in the development of tools, and the development of human resource management system has also received sufficient technical support. However, the second-generation human resource management system is mainly researched and developed by computer technicians and cannot achieve both technical and practical improvement, and its practical value is not high
2. Related Concepts and Theoretical Basis

2.1. Human Resource Decision-Making System. With the advent of knowledge economy, developing countries are facing the dual challenges of completing industrialization and accepting informatization and intellectualization. From now on, enterprises in developing countries, especially large and medium-sized enterprises, must further strengthen their awareness of scientific and technological innovation, improve the capital and product structure, rebuild enterprise organizations, adjust marketing strategies, and change management methods. However, in the final analysis, the challenge brought by the knowledge economy is the competition of talents. Strengthening enterprise human resource management is of great significance to promote the modernization of enterprises. Human resource is the main strategic resource of an enterprise in business decision-making and human resource [16]. As shown in Figure 2, human resources is the strategy of the enterprise [17].

2.2. Overview of Data Mining Technology Theory. Data mining refers to the process of obtaining a new understanding of the original things by using mathematical methods such as statistics from the large amount of data with complex structure that has existed and is still being generated. Compared with the long-developed computer science and technology, data mining is a relatively new scientific field, and Chinese and foreign scholars have given corresponding definitions from different perspectives [18]. In recent years, data mining technology has developed rapidly, and many data mining algorithms based on different purposes have appeared. According to their application characteristics, they can be divided into the following five types of tasks.

(1) Classification. “Classification” is to classify the data according to certain rules. For example, the data in the commercial database can be used to mine the effective information by the classification algorithm, and all customers can be classified according to the criteria such as unified preference or annual contribution estimate. Currently, neural network algorithms, naïve Bayesian algorithms, and K-nearest neighbor (KNN) algorithms can all complete the classification task [19].

(2) Clustering. The clustering task is mainly to use clustering algorithms, such as K-means (K-means) algorithm, maximum expectation (expectation-maximum, EM) algorithm, and K-center point (K-medoids) algorithm, to the highly similar data objects. Clustered together into a cluster, and there are obvious differences between different clusters [20].

(3) Prediction. “Prediction” refers to making scientific judgments about the future development of a business based on underlying information from data. For example, the banking system learns and revises the customer credit data existing in the database through data mining algorithms to obtain a credit evaluation model [21]. When a customer has a loan...
demand, the bank can use this model to predict the customer’s future loan repayment ability. Reference information to avoid loss of funds.

(4) Association. “Association” refers to determining whether there is a strong association between transactions by exploring the intrinsic relationship between data. The Apriori algorithm is a commonly used association rule algorithm in it. It can help people to explore the relationship rules that exist between entities or attributes and analyze the degree of association. It has been widely used in the field of commercial networks.

(5) Investigation. The main purpose of reconnaissance tasks is to search for anomalies, patterns, and outlier data and to provide supportive explanations for decision-making, which is not common in general data mining tasks.

Data mining is divided into guided data mining and unsupervised data mining. Guided data mining is to use available data to build a model, which is the description of a specific attribute. Unsupervised data mining is to find a relationship among all attributes. Specifically, classification, investigation, and prediction belong to guided data mining; association rules and clustering belong to unsupervised data mining.

2.3. Random Forest Algorithm. Decision trees are a common class of machine learning methods that are based on tree structure to make decisions with people. Classes have very similar solutions when facing problems. The decision tree is easy to understand and realize. People do not need users to know a lot of background knowledge in the learning process. This is also its feature that it can directly reflect the data. As long as it is explained, it has the ability to understand the meaning expressed by the decision tree. The preparation of data is often simple or unnecessary, and it can deal with data and conventional attributes at the same time and can make feasible and effective results for large data sources.
in a relatively short time. It is easy to evaluate the model through static testing and can measure the reliability of the model. Given an observed model, it is easy to deduce the corresponding logical expression according to the generated decision tree. In real diagnosis and treatment, doctors’ thinking is often based on it. Some variables make judgments of right and wrong, which is also consistent with the idea of the decision tree method. Data used for real data relationships are data distributed by factor blocks and can be used when data relationships are nonlinear. Data does not need to make any assumptions. The decision tree method groups the features of the sample and the knot where there are multiple features in the sample. The structure resembles the shape of a tree. The algorithm for grouping is that the branch node of the decision tree contains all samples. To belong to the same category, the higher the purity of the sample grouping under the node, the better. When making predictions, in the sample group with the same characteristics, the dependent variable in the sample group is used as the predicted result. The dependent variable is categorized. For variables, the majority vote method is used to obtain the predicted results. When the dependent variable is a numerical type, the sample group is used. The mean of the dependent variable in was used as the prediction result. The decision tree has a very good feature, namely, that it does not require itself. The type of variable can be handled very well when the independent variable is a categorical variable, unlike the neural network model that requires that all independent variables be numerical types. It is generated if all samples and all variables in the data build a model with the decision tree algorithm A tree, which is an individual learner. Building multiple individual learners and combining them through a strategy are an integrated learning method. For individual learners, determine the integrated learning method of individual learners, use grouped samples in the individual decision tree, and use similar methods to predict the dependent variables. The ensemble learning method has more results than a single individual learner that is stable and accurate. Integrated learning methods of decision tree include bagging method, random forest method, and promotion method. Random forest method is a special integrated learning method composed of multiple decision trees. Its characteristic is the self-service sampling method to build multiple sample sampling groups. The k is a specified parameter, usually using the square of the total number of attributes. In building random forests, the algorithm excluded most of the independent variables from the model. This method can avoid some independent variables with strong influence on the results being selected by the majority of the tree model. If most of the results of the tree model will have high correlation, averaging the highly related multiple results with large variance and small deviation, which will affect the accuracy of prediction. Random forest maximizes the difference between individual learners by double randomization of samples and attributes and often achieves good generalization error. The random forest algorithm refers to ensemble learning, solves the shortcomings of a single decision tree model, and integrates multiple decision tree models into a forest to predict the final result. The classification ability of a single decision tree is limited, but by randomly combining a large number of decision trees, a single sample data will produce the most likely classification result by superimposing the classification results of all decision trees. The schematic diagram of it is shown in Figure 3.

3. Related Technologies

3.1. Application Research of Data Mining Algorithms

3.1.1. Multiple Linear Regression Algorithm. The multiple linear regression (Multivariate.Linear.Regression) algorithm can be considered as an entry-level algorithm in machine learning. Its principle is relatively simple, and its application in practice is also very extensive. For a sample data set D with n-dimensional features (X1, X2, ⋯, Xn) (the number of samples is m), the expression of its corresponding multiple linear regression model is as follows:

\[ h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n = \sum_{i=0}^{n} \theta_i x_i. \]  

(1)

The matrix expression form of formula (1) is

\[ h_\theta(X) = \theta^T X. \]  

(2)

In the linear regression algorithm, we usually use the mean square error as the loss function and improve the fitting effect of the model by minimizing the loss function. The expression is as follows:

\[ J(\theta_0, \theta_1, \cdots, \theta_n) = \sum_{i=0}^{m} (h_\theta(x_0, x_1, x_2, \cdots, x_n) - y_i)^2. \]  

(3)

The matrix expression form of formula (3) is

\[ J(\theta) = \frac{1}{2} (X_\theta - Y)^T (X_\theta - Y). \]  

(4)

After determining the loss function, we usually adopt the gradient descent method or the least square method to determine the iterative formula of the \( \theta \) parameter, and their expressions are as follows:

\[ \theta = \theta - \alpha X^T (\theta X - Y), \]  

(5)

\[ \theta = (X^T X)^{-1} (X^T Y) \]  

(6)

The multiple linear regression algorithm has more than one data feature, and it is used to evaluate the linear correlation. Compared with the simpler univariate linear regression algorithm, the multiple linear regression algorithm has a wider range of applications, especially in the huge data information system of human resources, the generation of a decision is often accompanied by the influence of multiple factors. The linear regression algorithm is more scientific than using only one independent variable to predict the effect, and it is more in line with the needs of practical work.
3.1.2. K-Means Algorithm. K-means (K-means) algorithm is a commonly used machine learning clustering algorithm, and it is also widely used in the field of data mining. The main advantages of K-means are the principle is relatively simple, the implementation is also very easy, and the convergence speed is fast; clustering effect is better; the interpretability of the algorithm is relatively strong; the main parameter that needs to be adjusted is only the number of clusters K. And the main disadvantages of K-means are the selection of K value is not easy to grasp; it is difficult to converge for data sets that are not convex; if the data of each hidden category is unbalanced, such as the amount of data of each hidden category is seriously unbalanced, or the variance of each hidden category is different, the clustering effect is poor; using the iterative method, the result is only locally optimal, sensitive to noise and outliers.

We can apply the K-means algorithm to the employee performance appraisal management module of the human resource management system. First of all, we need to select the indicators related to employee performance appraisal according to the actual situation of the enterprise, such as employee communication and coordination ability, team leadership ability, innovation and development awareness, and the number of projects participated in, according to the level of importance to these assessment indicators. The total score of the assessment indicators is the comprehensive score of the employee. Then, we use the K-means algorithm to classify these scores as different performance levels. For example, if we set moxibustion to "3," the performance level can be set to "excellent," "moderate," and "low." The construction of this employee performance appraisal model can help enterprises to abandon the traditional subjective decision-making method and realize the objectivity of employee appraisal and evaluation. It is an important auxiliary tool for modern enterprises to achieve scientific management.

3.1.3. Naive Bayes Algorithm. It is one of the top ten algorithms in the field of data mining.

The core idea of naive Bayesian algorithm is to choose the decision with the highest probability, for example: for a new data point \( N(x, y) \), if you want to judge whether it belongs to category A or category B, you only need to consider which category has a higher probability of belonging. The specific rules are shown in formula (8):

1. If \( p_A(x, y) > p_B(x, y) \), the category of \( N \) is A
2. If \( p_A(x, y) < p_B(x, y) \), the category of \( N \) is B

To learn the Bayesian algorithm, we first need to understand the conditional independence formula. If \( x \) and \( y \) are independent of each other, then

\[
P(x, y) = P(x) P(y).
\]

Then, the conditional probability formula is

\[
P(y|x) = \frac{P(x, y)}{P(x)},
\]

\[
P(x|y) = \frac{P(x, y)}{P(y)}.
\]

In addition, the full probability formula can be expressed as

\[
P(x) = \sum_k P(x|y=y_k)p(y_k), \text{ where } \sum_k p(y_k) = 1.
\]

According to the conditional probability formula and the total probability formula, we can derive the Bayesian formula:

\[
P(y_k|x) = \frac{P(x|y_k)p(y_k)}{\sum_k P(x|y=y_k)p(y_k)}.
\]
In practical applications, for a given training dataset, we first need to learn the Bayesian algorithm to find the $y$ with the largest posterior probability.

### 3.2. Generalization Error and OOB Estimation of Random Forest

**Definition 1. Margin function.**

Assuming that the random forest is composed of $s$ decision trees $h_i(x), i = 1, 2, \cdots, s$, $y$ represents the correct category vector, and $z$ represents the wrong category vector, then, the calculation formula of the edge function $mg(x, y)$ as follows:

$$mg(x, y) = \text{avg}_y(I(h_i(x) = y)) - \max_{y \neq z} \text{avg}_y(I(h_i(x) = z)).$$  \hspace{1cm} (10)

Among them, $\text{avg}_y(\cdot)$ represents the average, $I(\cdot)$ represents the indicator function, and the edge function $mg(x, y)$ is the number of the maximum number of votes in the incorrect classification, which measures the classification model can correctly classify, the $mg(x, y)$, the confidence of the classification model.

**Definition 2. Generalization error.**

The generalization error measures the classification error rate of the random forest algorithm on the data set. The calculation formula of the generalization error $PE^*$ is as follows:

$$PE^* = P_{x,y}(mg(x, y) < 0).$$  \hspace{1cm} (11)

Among them, $P_{x,y}$ represents the definition space of probability, and $mg(x, y) < 0$ means that the sample is misclassified. It can be seen from the law of large numbers that the generalization error will tend to an upper bound, which shows that the RF algorithm will not follow the increase in the number of decision trees leads to the problem of overfitting, and the algorithm can effectively improve the shortcomings of single-classifier decision trees.

#### 3.2.1. Define 3 Out-of-Bag Data.

The random forest algorithm uses Bootstrap sampling to make about 37% of the samples in the original data set not sampled. These sample data that are not sampled are called out of bag (OOB) data. In the process of generating the training subset $D_i$ of each decision tree, the RF algorithm uses Bootstrap sampling with replacement to extract $n$ samples from the original data set $D$ with a capacity of $n$, one sample at a time, so some of the samples may be drawn multiple times, and some samples may not be drawn once. Assuming that each sample $x$ is independent of each other and the probability of being drawn each time is $1/n$, then after $n$ consecutive draws, the probability that a sample has never been drawn is $(1-1/n)^n$, when $n \rightarrow \infty$, the limit value of $(1-1/n)^n$ is about 0.368, which means that the training subset $D_i$ of each decision tree contains only about 63% of the samples in the original data set, and the remaining 37% of data never appears, this part of the data is OOB.
Definition 3. OOB estimation.

The out-of-bag data can be used to detect the pros and cons of the classification effect of it, and the obtained error value is called out-of-bag error (OOB error). Suppose the forest is composed of $s$ decision trees, record the OOB error of each tree as

$$\text{OOB error} (i = 1, 2, \ldots, i).$$

(12)

Then, the calculation formula of the out-of-bag error OOB error of the random forest algorithm is as follows:

$$\text{OOB error} = \frac{1}{s} \sum_{i=1}^{s} \text{OOB error}. \quad (13)$$

The datasets used in the experiment are new-thyroid2, ecoli1, ecoli2, ecoli3, glass0, wisconsin, and vehicle0. The above seven datasets are all from the KEEL database. The datasets can be downloaded from https://sci2s.ugr.es/keel/datasets.php URL to download. The characteristics of the dataset used in the experiment are described in Table 1, including the total number of samples in the data set, the number of attributes, the number of positive samples, the number of negative samples, and the imbalance ratio.

OOB estimation is an unbiased estimation of the random forest algorithm, which can replace the cross-validation method of the data set to avoid excessive time and space complexity, and is generally used to measure the generalization ability of classification models.

4. Experimental Design and Result Analysis

The datasets used in the experiment are new-thyroid2, ecoli1, ecoli2, ecoli3, glass0, wisconsin, and vehicle0. The above seven datasets are all from the KEEL database. The datasets can be downloaded from https://sci2s.ugr.es/keel/datasets.php URL to download. The characteristics of the dataset used in the experiment are described in Table 1, including the total number of samples in the data set, the number of attributes, the number of positive samples, the number of negative samples, and the imbalance ratio.
The BSMOTE algorithm processed by the SMOTE algorithm, the ADASYN algorithm, the Borderline-SMOTE (BLSMOTE) algorithm, the ISMOTE algorithm, and the KMSMOTE (KSMOTE) algorithm is compared to the random data set. Among them, the number of neighbors $k$ of positive samples in SMOTE algorithm, ADASYN algorithm, BLSMOTE algorithm, and ISMOTE algorithm is 5; the number of clusters in KSMOTE algorithm and BSMOTE algorithm is 5; the number of decision trees in random forest is 100, each decision tree. The number of attributes selected is $\log_2 M + 1$.

The performance of different algorithms on different data sets is shown in Table 2.

Figure 4 compares the changes in the OOB error value, $G$-mean value, $F$-measure value, and AUC value of the original dataset and the dataset processed by the six resampling algorithms and the random forest combined classification.

In order to analyze the difference between the ID3 algorithm, the C4.5 algorithm and the CART algorithm when the nodes are split, the following will compare the calculation results of the three different algorithms in selecting the split attribute of the root node of the decision tree. There are 15 samples in total, and each sample contains Figure 5 shows the comparison of the calculation results of three different node splitting algorithms with 9 attributes and 1 label.

It shows, according to the principle of the ID3 algorithm, the attribute with the largest gain should be selected as the splitting attribute of the node, so the algorithm will select the attribute $m4$ as the root node for splitting; according to the principle of the C4.5 algorithm, the GainRatio should be selected. The largest attribute is used as the split attribute of the node, so the algorithm will choose attribute $m5$ as the root node to split.

The dataset adopts tic-tac-toe, monkey-2, and breast, respectively, and uses ID3 algorithm, C4.5 algorithm, and CART algorithm as the node splitting algorithm of random forest to conduct 100 experiments. The number is 100, and the number of attributes selected for each decision tree is $\log 2M + 1$, where $M$ is the total number of attributes in the data set, the evaluation index is the OOB error estimate of the out-of-bag error, and the experimental result is the average of the OOB error of 100 experiments. The three algorithms on the three datasets are shown in Figure 6.

As can be seen from Figure 6, on three different data sets, the classification performance of random forest using ID3 algorithm and C4.5 algorithm is basically similar, and the random forest using CART algorithm is quite different from the former two. The OOB estimates on different datasets are sometimes high and sometimes low. The random forest classification performance using the CART algorithm is better on the breast dataset, but worse on the tic-tac-toe dataset.

Figure 7 shows the curve comparison of the classification accuracy of J48, RandomTree, SimpleCart, NBTree, and combined classifiers with different sample number. Figure 6 that the trends of the five overall curves tend to be consistent. The curve of combination of the combined classifier is higher than the curves of all the individual classifiers, indicating that the combined classifier is effective in improving the classification prediction accuracy of the decision tree, and the prediction result has high reliability.

The $G$-mean value, $F$-measure value, and AUC value of DT prediction model, KNN model, SVM model, RF prediction model, and BSMOTE-LPRF model are shown in Figure 8.

In Figure 8, the BSMOTE-LPRF prediction model has obvious performance advantages. Compared with the decision tree prediction model, the KNN prediction model, the SVM prediction model, and the traditional random forest prediction model, the BSMOTE-LPRF prediction model is in each performance evaluation. The indicators are obviously better than other models.
5. Conclusion

In view of the current situation that the human resources statistical management system is widely used in government management departments and various enterprises and has rapidly accumulated a large number of relevant data of enterprises and talents in recent years, use these data to rationally analyze the distribution of enterprises and talents in regions. The situation, implicit laws, and using these laws to provide policy or decision support for the government or enterprises have become the urgent concerns of the government and enterprises. Based on this research background, this paper takes data mining and random forest algorithm as the starting point, establishes the model of it, further focuses on the famous data mining platform, expands its open-source interface, and builds the management system model. Random forest algorithm is used to build a brain drain prediction model. And use the literature research method, data analysis method, and qualitative and quantitative combined analysis method to analyze the reasons for the brain drain, on this basis, put forward the main strategies to prevent the brain drain.

Data Availability

The figures and tables used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Opatha, “Teaching ethics in human resource management education: a study in Sri Lanka,” Prabandhan Indian Journal of Management, vol. 14, no. 1, pp. 8–24, 2021.
[2] S. Liu, “Human resource management of internet enterprises based on big data mobile information system,” Mobile Information Systems, vol. 2021, Article ID 1549342, 9 pages, 2021.
[3] Y. Y. Jing, M. Y. Yusliza, and O. O. Fawehinmi, “Green human resource management,” Benchmarking: An International Journal, vol. 27, no. 7, pp. 2005–2027, 2020.
[4] M. Shoaib, R. Žámečník, Z. Abbas, M. Javed, and A. U. Rehman, “Green human resource management and green human capital: a systematic literature review,” in Contemporary issues in business, management and economics engineering. Proceedings of the International Scientific Conference in Vilnius, vol. 9 no. 4, pp. 30–41, Lithuania, 2021.
[5] G. S. Saini, S. K. Dubey, and S. K. Bharti, “TraCard: a tool for smart city resource management based on novel framework algorithm,” Recent Advances in Computer Science and Communications, vol. 6, no. 3, pp. 28–32, 2020.
[6] H. S. Al-Dhaafri and M. S. Alosani, “The model of Brimob-Polres relationship: human resources management perspective; empirical evidence from public sector,” International Journal of Quality & Reliability Management, vol. 8, no. 9, pp. 70–83, 2021.
[7] I. V. Isaal, “Tracer study of BSBA (human resource management) graduates of UM Tagum College,” Human Resource Management, vol. 7, no. 2, pp. 81–90, 2021.
[8] H. C. D. Sancaka and S. Sarpan, “The model of Brimob-Polres relationship: human resources management perspective,” LPPM Universitas EkaKarya, vol. 9, no. 4, pp. 80–91, 2021.
[9] G. Bandi, T. S. Rao, and S. S. Ali, “Data analytics applications for human resource management,” in 2021 International
Conference on Computer Communication and Informatics (ICCCI), vol. 7, no. 4, pp. 23–31, Coimbatore, India, January 2021.

[10] N. A. Valcik, M. Sabharwal, and T. J. Benavides, “The theory of how HRIS human resource information system (HRIS) should work,” Management for Professionals, vol. 9, no. 11, pp. 77–82, 2021.

[11] J. Han, H. Cheng, Y. Shi, L. Wang, Y. Song, and W. Zhng, “Connectivity analysis and application of fracture cave carbonate reservoir in Tazhong,” Science Technology and Engineering, vol. 16, no. 5, pp. 147–152, 2016.

[12] S. L. V. Papineni, A. Mallikarjuna Reddy, S. Yarlagadda, S. Yarlagadda, and H. Akkineni, “An extensive analytical approach on human resources using random forest algorithm,” International Journal of Engineering Trends and Technology, vol. 69, no. 5, pp. 119–127, 2021.

[13] C. Kartiko, H. A. Arrasyid, and A. C. Wardhana, “Designing a mobile user experience student knowledge management system using lean UX,” Journal of Engineering and Applied Technology, vol. 2, no. 1, 2021.

[14] I. Y. Eremina, P. Hoa, and A. V. Chuprova, “Analysis and evaluation of human resource management systems in international petroleum companies,” vol. 6, no. 1, pp. 79–88, 2020.

[15] E. Bilyk, O. Khymych, Y. Blynda, and U. Lukashevska, “Tools to increase the efficiency of personnel recruitment in the system of human resources management in the it sphere,” National Association of Scientists, vol. 2, no. 32(59), pp. 24–31, 2020.

[16] Q. Liu and Z. Zhao, “Relationship model between human resource management activities and performance based on LMBP algorithm,” Security and Communication Networks, vol. 2022, Article ID 1125084, 11 pages, 2022.

[17] S. Selvam, “Career recommender system using random forest algorithm,” Journal of Critical Reviews, vol. 7, no. 17, p. 2020, 2020.

[18] B. Jitmaneeroj, “Does investor sentiment affect price-earnings ratios?”, Studies in Economics and Finance, vol. 34, no. 2, pp. 183–193, 2017.

[19] D. V. Balaji and J. Arunnehr, “Predictive analysis on HRM data: determining employee promotion factors using random forest and XGBoost,” Proceedings of International Conference on Deep Learning, Computing and Intelligence, vol. 737, no. 3, p. 139779, 2022.

[20] Y. Wu, Z. Wang, and S. Wang, “Human resource allocation based on fuzzy data mining algorithm,” Complexity, vol. 2021, Article ID 9489114, 11 pages, 2021.

[21] D. Zaharov and Y. Kashtanova, “Introduction of professional standards in the organization human resources management system,” Management of the personnel and intellectual resources in Russia, vol. 8, no. 6, pp. 34–37, 2020.