

Evaluation and Design of Employment Decision Analysis System Based on Artificial Intelligence

Peng Li^{1,a}, Lina Yuan^{2,b,*}

¹*School of Fine Arts and Design, Changchun Humanities and Sciences College, Changchun, Jilin, 130117, China*

²*Ideological and Political Theory Teaching and Research Department, Changchun Humanities and Sciences College, Changchun, Jilin, 130117, China*

^a253482280@qq.com, ^b25065600@jlu.edu.cn

**Corresponding author*

Keywords: Data Mining Technology, Information Management System, Decision Tree, Graduate Employment, Decision Support, Employment Forecast

Abstract: Data mining technology based on artificial intelligence is already a hot topic of research nowadays. This article adopts a B/S architecture to develop and design a college graduate employment decision analysis system, combined with artificial intelligence data mining technology, in order to solve the problem of graduate employment decision-making and management in schools. The main work includes: (1) an online management information system is designed and implemented for graduates, which enables automatic computer processing and orderly management of cumbersome employment tasks; (2) the experiment shows that this paper can be applied to the analysis and prediction of the employment situation of graduates, and fully realizes the entire process of data classification mining, determining the objects and target data collection of data mining. Data preprocessing technology, using C4.5 classification algorithm to generate decision trees, tested with 249 employment analysis information data from 2022, with an accuracy rate of 80%.

1. Introduction

People from all walks of life have begun to attach importance to the auxiliary management of computer information systems. More and more companies are using data mining technology to complete different activities of decision making in their operations, providing determination for the business management [1]. In the field of education, data mining methods have been successfully implemented in higher education and have become an interesting field of educational data mining research. Student performance is an important component of school management's attention. Data mining applications can help academic management systems investigate and identify outstanding student and dropout groups in universities, and it can help identify the most important attributes that contribute to student performance. Higher education institutions can acquire in-depth and comprehensive knowledge to strengthen their curriculum planning, evaluation, evaluation planning, and decision-making [2].

In the research of this field, researchers have used various artificial intelligence algorithms and data mining technologies to process and analyze employment market data. Chen Jin demonstrated for the key implementation of data mining in educational management decision support systems [3]. At the same time, many scholars used natural language processing technology to analyze and process recruitment information, job descriptions and other text data, improving the efficiency and accuracy of data mining [4]. Some scholars have combined multiple algorithms to design intelligent employment decision analysis systems that can quickly provide job recommendations and employment suggestions based on user needs and backgrounds [5].

This article is based on data mining technology and adopts a B/S architecture to develop and design a college graduate employment decision analysis system based on artificial intelligence data mining technology, to solve the problem of graduate employment information management and employment decision-making in schools, and analyze data mining technology and decision tree classification algorithms, and apply data mining technology to the employment decision analysis system, analyzing the role of data mining in the decision-making system [6-7].

2. Data Mining Methods

Decision tree-based methods and neural network-based methods are two commonly used methods in data mining, both of which can be used for tasks such as classification and prediction [8].

2.1. Decision Tree-based Approach

A decision tree is a structured diagram composed of nodes, where each node corresponds to a specific feature, and each branch denotes a possible value or condition for that feature. The final endpoints of the tree are the leaf nodes, which indicate the outcome or decision based on the values of the features. By constructing a decision tree, the category or label to which the input data belongs can be predicted based on its characteristic values [9-10]. The process of building a decision tree is basically 3 processes. The first operation is feature selection, which requires selecting the features that best distinguish the data categories as nodes; then, in decision tree generation, it is necessary to generate a decision tree by recursively dividing the data set; finally, in decision tree pruning, it is necessary to avoid overfitting by pruning operations. The method is highly explanatory and easy to understand and operate, and also has high efficiency and accuracy [11]. It is widely used in tasks such as classification, prediction and feature selection.

2.2. Overview of Decision Tree Algorithms

Three forms of entropy are applied in the decision tree, namely information entropy, conditional entropy and mutual information. Entropy is a measure of the uncertainty of information and knowledge in massive amounts of data. The uncertainty of information knowledge can determine the size of entropy, and the two are positively correlated, while the uncertainty and complexity of data information are determined by probability. If there is only one type of information, the highest entropy value of purity is 0. If there are two types of information with the same quantity, the degree of confusion is highest; the purity is lowest; the entropy value is 1. The following is the calculation formula for information entropy in the unary model. c is the number of features; P is the proportion of features in the total number, and T is a random variable:

$$E(T) = \sum_{i=1}^c - p_i \log_2 p_i \quad (1)$$

The following is the calculation formula of conditional entropy $E(T, X)$. $p(c)$ is the proportion of each eigenvalue, and $E(c)$ is the information entropy of the eigenvalue. The lower the value of conditional entropy, the smaller the uncertainty of the binary model.

$$E(T, X) = \sum_{c \in \chi} p(c)E(c) \quad (2)$$

The decision tree algorithm essentially calculates data, classifies attributes based on their similarity, and constructs a tree like graph. Starting from the top root node, each sub node is separated from top to bottom until the data attributes cannot be further split. When performing classification recursion, selecting appropriate features is the most important step in constructing node decision trees.

2.3. Neural Network-based Approach

The approach is a machine learning method used for tasks such as classification, prediction, and pattern recognition. A neural network is a connected system consisting of many neurons, which process input data and output results by connecting weights and activation function [12-13]. Neural network-based methods typically include three stages: forward propagation, error backpropagation, and weight update. In forward propagation, the input data passes through a neural network to obtain the output result. In error backpropagation, the gradient of weights is calculated based on the error between the output result and the actual result. In weight update, the weights of the neural network are updated based on gradients. Neural network-based methods have advantages such as strong adaptability and the ability to handle nonlinear problems, while also having disadvantages such as long training time and requiring a large amount of data.

3. Data Mining Theory

3.1. Concepts of Data Mining Technology

Data mining discovers and extracts hidden patterns, relationships and trends in data by applying machine learning, statistics, data and information visualization and other technologies [14]. Data mining is a branch or subfield of artificial intelligence that provides necessary data support and predictive models for artificial intelligence. In the fields of machine learning and deep learning, data mining is an important process for training algorithms, as algorithms require a large amount of data to learn and optimize [15].

3.2. Functions of Data Mining

Data mining is a process of automatically discovering hidden patterns and relationships from a large amount of data [16-17]. In this process, data mining technology would apply a series of algorithms and technologies. These techniques are described below:

(1) Cluster analysis is the process of dividing objects in a dataset into different groups based on their similarity, to master its structure and characteristics.

(2) Classification analysis is achieved by learning the classification labels of existing data samples and predicting the classification labels of new unknown data, thereby achieving data classification.

(3) Time series analysis is the modeling and analysis of time series data, discovering the periodicity, trend, seasonality, and other regularities of the data, and making predictions and decisions.

(4) Text mining includes text classification, text clustering, emotion analysis, entity recognition, keyword extraction and other technologies [18].

The employment management department of college graduates can introduce data mining to analyze the employment trends and actual employment situation of students from a large amount of college graduates' employment data, providing better assistance for future employment problems of college students.

3.3. Data Mining Environment

A data mining environment is a software tool that provides a range of mining and analysis capabilities that can help users discover useful information and knowledge from large amounts of data [19-20]. The schematic diagram of the data mining environment is shown in Figure 1:

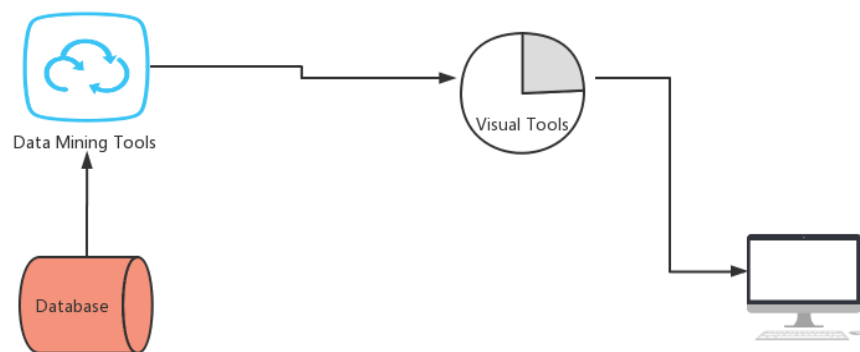


Figure 1: Data mining environment block diagram

3.4. Introduction to the Data Mining Process

The general content of each step in the process is as follows:

- (1) Identifying business objects refers to clarifying the problems and goals that data mining needs to solve.
- (2) Data preparation refers to the pretreatment of raw data such as selection, cleaning, conversion, etc., to prepare the data set for data mining.
- (3) Data mining is the technique of discovering relationships, trends and patterns from data sets.
- (4) Result analysis is the evaluation, interpretation, and validation of data mining results to ensure their accuracy, reliability, and practicality.
- (5) Knowledge assimilation refers to the combination of knowledge and information obtained from data mining with business practice, forming a certain knowledge system and application model to support business decision-making and optimization.

3.5. Application of Data Mining Technology in Graduate Management Information System

Data mining technology can be widely applied in graduate management information systems. For example, clustering analysis algorithms can be used to classify graduates in order to distinguish different types of graduates and provide different services for them; Classification analysis algorithms can be used to predict and analyze the employment positions of graduates, in order to help them better choose their career direction. Text mining technology can be used to extract key information from graduates' resumes and cover letter to help human resources departments better understand the ability and quality of graduates. In summary, applications combine this technology

with systems to provide better service and support for graduates

4. Design and Implementation of an Employment Decision Analysis System Based on Data Mining Technology

4.1. Design Purpose

After the successful design of this system, the system stores various information of graduates, employers and students' employment, etc, which is very rich in content. Based on this, relevant data mining is needed. There are many available mining algorithms among them. This project mainly focuses on the characteristics of the data in each module of the system, and uses association rules and decision tree classification methods to mine existing data. The main objectives of achieving this system are as follows:

(1) It realizes scientific and efficient management of graduates' information and their employment information.

(2) The correlation results obtained from data mining analysis can provide decision-making support to college leaders, enabling them to tailor more appropriate training models for the holistic development of students.

4.2. System Requirements and Functional Evaluation

This paper chooses to use Microsoft SQLServer 2000 relational database as the basis to realize the graduate information, employment information, employers and other aspects of the centralized storage. Then, the system is able to realize the operations of adding and deleting, modifying different types of their information processes, Finally, through association rules and data mining, it can provide support for students' employment and other aspects. the system design needs to be combined with data business processes to realize system development and management from several aspects, so that its prediction results can meet the relevant requirements, as shown in Figure 2:

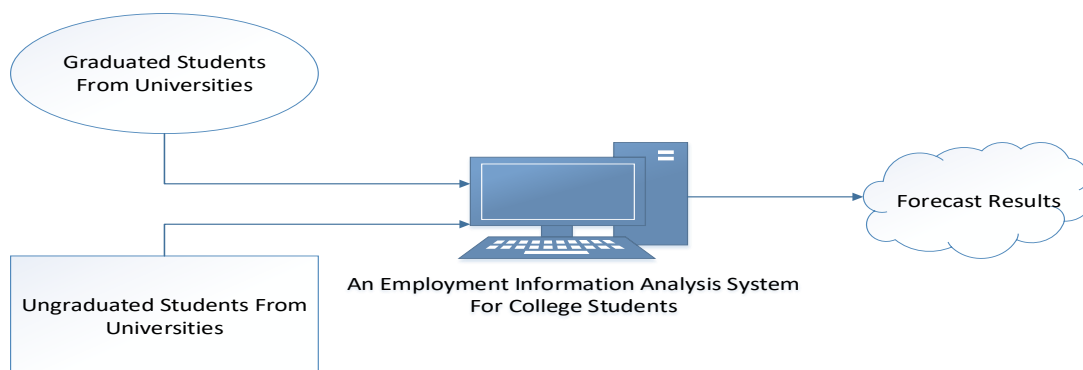


Figure 2: System data flow chart

4.3. System Functional Design

Due to the relatively difficult job analysis and decision-making work of college students, Therefore, in the process of system development and design, this article not only needs to adhere to the principles of orderliness and clear goals, but also to ensure that each module and its relationship are in a high degree of aggregation and low degree of coupling. The functional modules of the system are designed for 5 main parts, as shown in Figure 3.

The specific objectives can be achieved as follows:

- (1) People collect detailed basic information of graduates and employers.
- (2) Graduates can promptly inquire about all employers' information, and employers can also have a comprehensive understanding of all graduate information.
- (3) People feedback employment information of graduates to college leaders and managers for making relevant decisions.
- (4) It realizes the management and use of the system by college administrators and employment authorities.

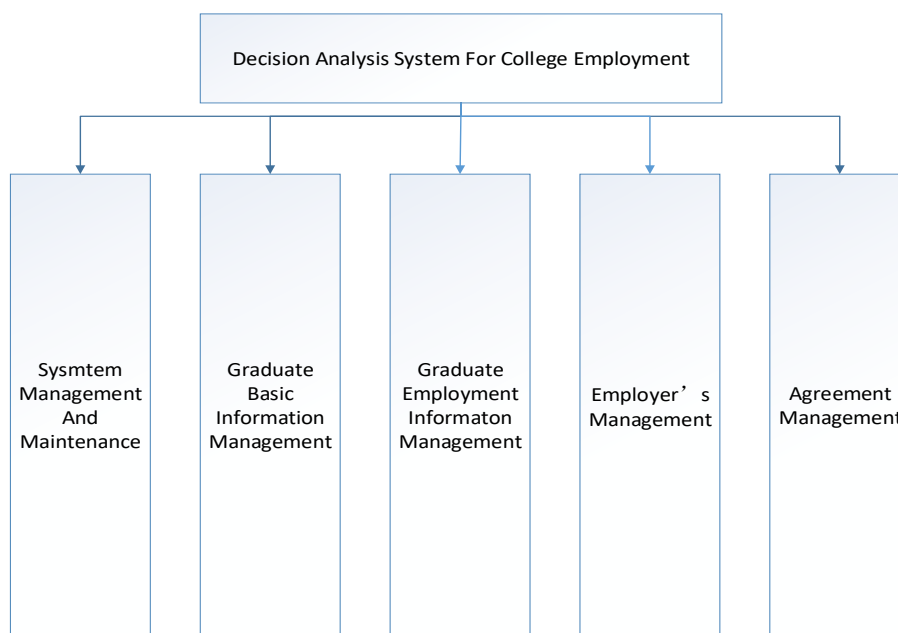


Figure 3: Schematic diagram of system functional modules

4.4. Implementation of Database Tables

This article chooses to use the Microsoft SQLServer 2000 database for database design, and the data table design is shown in Table 1:

Table 1: Employee information management system user information table

Field name	field type	field length
Username	Char	18
password	Char	18
identity	Char	18
permission	Char	5

4.5. Implementation of Main System Function Templates

Through the design analysis of the system, the main functional modules of the system were designed using JAVA programming, as shown in Figure 4.

1) System Management and Maintenance Module

The design of this module can achieve operations such as adding, querying, and logging out users, as well as user login/exit, password modification, and permission modification through system settings, as shown in Figure 5.

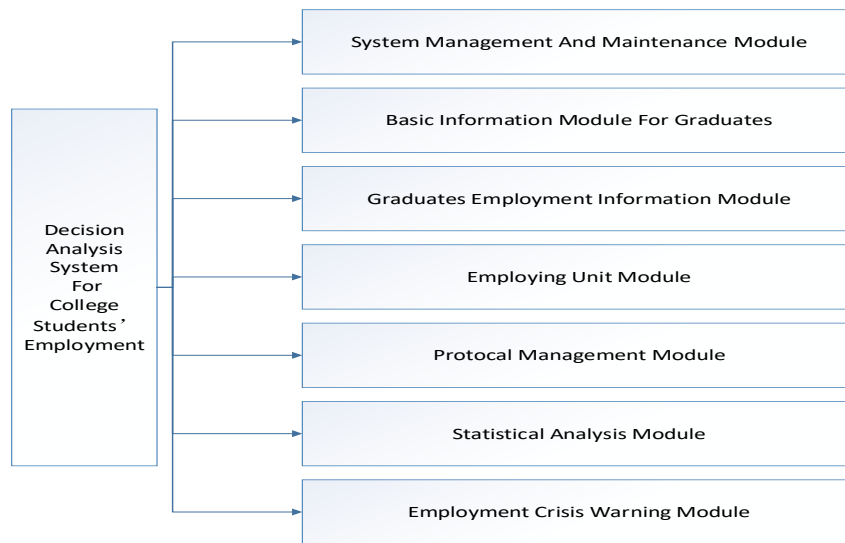


Figure 4: Design diagram of the main functional modules of the system

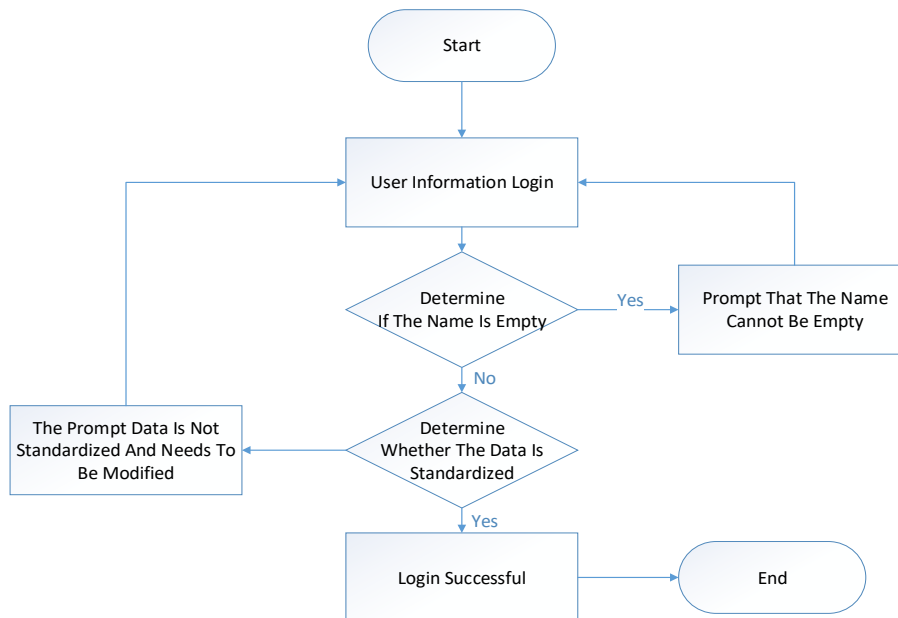


Figure 5: User login management

2) University Graduate Information Management Module

This module is mainly responsible for the management and maintenance of college graduates' information, including the release of graduate profile information, employment information, personal information, recruitment information from employers, and related information such as students' job search intentions, in order to comprehensively understand and analyze the situation of graduates.

3) Protocol Management Module

This module is mainly responsible for the management and maintenance of employment agreements, including functions such as entering, reviewing, and storing the agreements.

4) Statistical Evaluation Module

This module provides statistics on the employment rate, employment industry distribution and employment salary of graduates.

5) Employment Crisis Warning Module

This module implements real-time monitoring of student employment information and employment situation in the college student employment analysis and decision-making system. The relevant student employment information analysis reports uploaded by the statistical analysis module have been processed to build corresponding early warning modules from various perspectives such as talent demand, supply, professional demand, and current employment situation, so as to provide assistance in judging the employment information of school students. The application of this module can also provide corresponding guidance and suggestions for students' employment evaluation, employment prediction, and warning.

4.6. Student Position Forecast

The current basic information, curriculum information and other characteristics of the student, the characteristics of previous students and employment information and other data are analyzed and predicted through random forest decision tree, Bayesian model, GBDT and other algorithms to get the post prediction of the student. The prediction of student positions is shown in Figure 6:

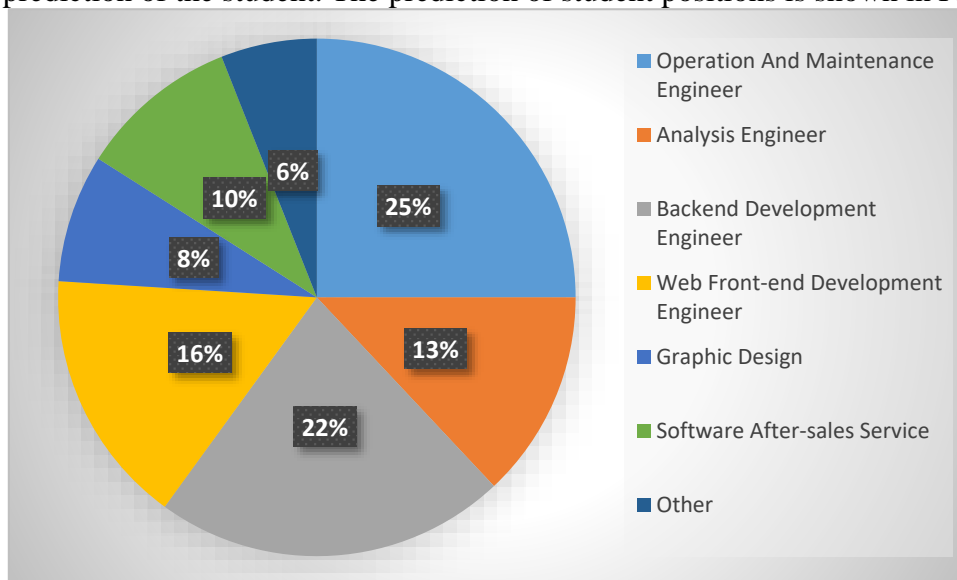


Figure 6: Student position forecast

5. Application of Decision Tree Algorithm in Employment Evaluation Decision System

5.1. Classification Algorithm

Classification technology is an important algorithm in data mining technology, which can classify and predict data, helping decision-makers better understand the characteristics and trends of data, and providing support and basis for decision-making. In the employment analysis and decision-making system for college graduates, classification technology can analyze and predict the employment situation of graduates by using specific classifiers to provide corresponding guidance and support. Classification technology has the advantages of predictability and discreteness. By classifying and predicting data, it is possible to predict its trends and patterns, providing early warning and preventive measures for decision-making. At the same time, the classification technology can also carry out discretization processing to convert continuous data into discrete data sets, so as to facilitate the classification and analysis of data.

5.2. Data Preprocessing

Before preprocessing a large amount of data on college graduates, it is necessary to collect their information, such as basic information and employment information, before corresponding processing can be made. Data preprocessing is divided into two aspects: data integration and data conversion. The basic information of college graduates and the employment information data of graduates are sourced from different databases. In this case, the database must be integrated first, and then its information can be collected. Then, the data is converted and the existing information is Discretization.

5.3. Data Classification Model

Through the application and analysis of the ID3 algorithm and C4.5 algorithm of the decision tree under the same processing mode, it is found that the C4.5 algorithm has more significant performance advantages, and the algorithm can fully guarantee the accuracy of attribute analysis. Therefore, this paper chooses to use the C4.5 algorithm to set and divide the data in order to build an employment information analysis module for college graduates, so that the employment situation of college graduates can be mined and analyzed. The classification rules are as follows:

Example 1: Classification rules for education majors in comprehensive universities

(1) Type of IF institution=comprehensive university+discipline=education+year-on-year change in the number of graduates=growth+average teaching and administrative room per student=meeting THEN employment rate=average;

(2) Type of IF institution=comprehensive university+discipline=education+year-on-year change in the number of graduates=growth+average teaching and administrative room per student=critical student teacher to metallurgical sector THEN employment rate=poor;

Example 2:

(3) IF institution type=teacher training institution+discipline=engineering+year-on-year change in the number of graduates=growth+teaching and research equipment per student=meeting standards THEN employment rate=average;

(4) IF institution type=teacher training institution+discipline=engineering+year-on-year change in the number of graduates=reduction+student teacher ratio=qualified THEN employment rate=poor.

6. Conclusions

In summary, this article designs a college graduate employment decision-making analysis system, to achieve employment management decisions for college students. In the process of system design and development, advanced computer technology and decision tree theory were fully integrated, achieving database creation and providing guarantee for the storage of system data. Finally, the classification and prediction rules obtained were tested using 249 employment statistical analysis data from 2022 as test data. The results show that the accuracy of data classification is 80%. The reliability of classification prediction rules is high. This indicates that the data mining technology can provide many conveniences for information management and employment decision-making of college graduates.

References

- [1] Jian, L. (2022). Design of enterprise human resources decision support system based on data mining. *Soft Comput*, 26(2), 10571–10580.
- [2] Fairos, W., Wan Yaacob, W. F., Azlin, S., Md Nasir, S. A., Faizah, W., Sobri, N., Mara, C., & Kelantan, Malaysia.

- (2019). Supervised data mining approach for predicting student performance. *Indonesian Journal of Electrical Engineering and Computer Science*, 16(3), 1584-1592
- [3] Chen, J. (2021). Horizontal model of higher education management policy support system based on data mining. In J. MacIntyre, J. Zhao, & X. Ma (Eds.), *The 2020 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy*, 36(2), 836-840.
- [4] Gary K Y Chan. (2022). AI employment decision-making: integrating the equal opportunity merit principle and explainable AI. *AI & Society*, 36(1), 321-329.
- [5] Zhu, S. (2022). Design of a College Student Employment Information Management System Based on Data Mining Technology. *Information Recording Materials*, 23(11), 220-222.
- [6] Lian, J & Huang, W. (2022). Design and exploration of a university employment decision-making system based on big data. *Science and Technology and Innovation*, 23(1): 28-31
- [7] Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, 24(3), 567-598.
- [8] Joseph, S. I. T., & Thanakumar, I. (2019). Survey of data mining algorithm's for intelligent computing system. *Journal of trends in Computer Science and Smart technology (TCSST)*, 1(01), 14-24.
- [9] Wang, S. (2019). Research on the Decision Model of Graduate Employment Information Management Based on Data Mining. *Journal of Northeast Electric Power University*, 39(05): 86-90.
- [10] Ye, Z. (2019). Research on the Application of Decision Tree Based Data Mining Technology in Employment Information Management System. *Science and Technology Communication*, 11 (16): 132-134.
- [11] Kurnia, R., Triayudi, A., & Rahman, B. (2020). Employee Assessment Decision Support System Using the MOORA Method at the National University: Employee Assessment Decision Support System Using the MOORA Method at the National University. *Jurnal Mantik*, 4(1), 562-571.
- [12] Gandomi, A., & Haider, M. (2018). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- [13] Huang, B. (2021). Research on Employment Prediction of University Graduates through Data Mining. *Application of Microcomputers*, 37 (11), 171-173
- [14] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., & Vlahavas, I. (2018). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 16(5), 145-155.
- [15] Kakad, S., Kadam, R., Deshpande, P., Karde, S., & Lalwani, R. (2020). Employee attrition prediction system. *Int. J. Innov. Sci., Eng. Technol.*, 7(9), 7-15.
- [16] Purwanto, T., Abdillah, L., & Agustini, E. (2020). Knowledge management system analysis and planning for employee recruitment and training. *Journal of Electrical, Electronics and Informatics (JEEI)*, 4(2), 38-43.
- [17] Sugiyarti, E., Jasmi, K. A., Basiron, B., Huda, M., Shankar, K., & Maselena, A. (2018). Decision support system of scholarship grantee selection using data mining. *International Journal of Pure and Applied Mathematics*, 119(15), 2239-2249.
- [18] Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining techniques applied in educational environments: Literature review. *Digital Education Review*, 33(1), 235-266.
- [19] Hussain, S., Dahan, N. A., Ba-Ahwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447-459.
- [20] Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of medical systems*, 43(1), 1-15.