# FORECASTING STUDENTS' EMPLOYMENT RATE UNDER THE OBE MODEL

XiaokeDeng, Huakun Hou, Meiyu Jin and Linbo Zhai*

School of Information Science and Engineering, Shandong Normal University, Jinan, China

***ABSTRACT***

With the rapid development of information technology and its penetration into the field of education, it has become a worthwhile research topic to explore how to combine Outcome Based Education(OBE) model with data mining techniques to obtain predictions of secondary school students' employment rates. Therefore, this paper first collected attribution data affecting the employment rate of secondary school students, and then used three models to predict the highly correlated attribution data. Finally, based on the analysis and prediction results, the OBE model was implemented for the corresponding classes. The experimental results found that the combination of machine learning algorithms and educational practices facilitated the development of secondary education and significantly increased the employment rate of secondary school students.

***KEYWORDS***

Education data mining, OBEmodel, Student employment prediction

## 1. INTRODUCTION

Data mining techniques (EDM) is an emerging area of research applied to educational background (Bakhshinategh et al., 2018). EDM uses various amounts of data obtained from educational institutions, such as student data and behavioural records, to uncover valuable information to help teachers better understand student behaviour and use it to improve teaching and learning environments. In today's era of knowledge economy, students are a key factor in the socio-economic growth of any country, so maintaining their performance is vital. The development of EDM has driven the development of traditional education. Analyzing data related to student performance not only meets the diverse and growing needs of teachers and learners, but also plays an important role in driving innovation in education.

OBE is a new model of education based on learning outputs that emphasizes the process of instructional design and delivery. He aims at the highest level of competence that students can achieve through a certain stage of learning, in particular knowledge, skills and values, which means that learning is best achieved by first identifying what needs to be achieved.

Existing machine learning methods are widely used to predict student performance and show shown good performance, but are rarely integrated with educational reality. In this paper, first of all, feature selection is used to reduce the number of variables in the input model, and then Decision Tree (DT), Artificial Neural Network (ANN) and Support Vector Machines(SVM) in

David C. Wyld et al. (Eds): CMLA, GRAPH-HOC, CIoT, DMSE, ArIT, WiMoNe, CSEIT – 2023
pp. 37-47, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131004

the machine learning algorithm are used to predict the employment rate of students. Finally, according to the prediction results and the actual education, teaching arrangements are reasonably planned. The experimental results show that the combination of feature selection and ANN model can improve the accuracy of prediction, and on this basis, the combination of OBE model can greatly improve students' achievement and employment rate.

The paper is organized as follows: Section 2 presents the related work. The data analysis and methods are presented in Section 3. Section 4 gives the OBE model. Section 5 discusses the results of the experiment. Section 6 provides conclusions and future work.

## 2. RELATED WORKS

Predictive analysis is an activity that uses past data to build models to predict future events and behaviors. It is widely used in the field of education. Ortiz-Lozano et al. (2018) used three machine learning methods, namely SVM, ANN and DT, as well as multiple linear regression for developing predictive models to discover the potential relationship between the overall e-learning system availability and its predictors, however valuable data indicating the level of student engagement was not available. Dahman et al. (2019) used three machine learning models were used to predict the ability of adult learners to decide to continue with an English for speakers of other languages (ESOL) course, but only for English language. Pojon, Murat (2017) used Linear Regression, DT, and Naive Bayes classification (BNs) to predict whether a student would be successful, but focused more on the impact of feature engineering selection on predictive performance. Tan, Mingjie et al. (2015) selected students' personal characteristics and academic performance as input attributions, and use ANN, DT and BNs to establish a model to predict students' dropout potential, but ignore the low accuracy of a single model. Li X et al.(2021)proposed a technique for predicting the employment situation of graduates based on a long short-term memory (LSTM) recurrent neural network, but resulted in low accuracy with random interference noise. Chen, Yawen (2023) used seven machine learning algorithms to predict binary and multi-classification education data, investigating student performance in different application scenarios and demonstrating the applicability of different machine learning methods to different classification prediction tasks, but there is no educational reality.

This paper considers that judging learning outcomes is a more comprehensive measure compared to evaluating only grades, which is consistent with the claim in Wang, Peipei, et al.(2022) that learning outcomes represent a key factor in learning success. Different from the current work, this paper selects three different machine learning algorithms to make predictions on the same data set with the aim of determining the effectiveness of machine learning in education. On this basis, the OBE model is introduced to combine machine learning with the education scenario, providing a brand new perspective for the application of machine learning in the education scenario.

## 3. DATA ANALYSIS

### 3.1. Data Collection

Each of the basic subjects offered by the school has a corresponding competency that students are required to master, which is important for the promotion of original innovation and also affects their future employment rate. Table 1 shows the competencies that students are required to master for basic courses of schools formulated by the state. The dataset selected for this study was the grades and employment of 300 students from six classes of computer application in a secondary school in Jinan for the class of 2022 for all subjects over three years, together with the behavioral

characteristics, parental participation characteristics, demographic characteristics and other factors obtained through questionnaire. Table 2 describes the data set used in this article.

Table 1 Competencies required of students in foundation courses

| Courses | Capabilities |
|---|---|
| Chinese | Understand and use language correctly |
| Math | Logical thinking skills and support for other subjects |
| English | Develop students' ability to apply in scenarios |
| Moral | Formation of good ideological perceptions |
| PE | Promoting student fitness and health |
| History | Developing students' historical thinking and concepts |

Table 2 Properties of the data set

| Column | Description |
|---|---|
| Sex | Sex of student |
| Age | Age of student |
| Moed | Mother's education |
| Questions | Number of questions answered |
| Satisfaction | Satisfaction with the school |
| Discussion | Number of discussion groups attended |
| Correct | Correctness rate of assignments |
| Study time | Weekend study time |
| Absences | Number of absences |
| Satisfaction | Parents' satisfaction with the school |
| Willingness | Willingness to go on to higher education |
| Foundation courses | Chinese Math English Moral PE History |
| C | C language programming |
| Web | Web Design and production |
| Ps | Photo shop |
| Graphics | Graphics and Image Processing |
| Computer | Computer Basics |
| Pass | Student employment |

## 3.2. Data Processing

There are deficiencies in the data set used in this study. In order to improve the accuracy of prediction results, these incomplete data should be processed specifically. Common methods for missing data are manual filling, constant value replacement, mean replacement, median replacement or ignoring tuples and most likely value replacement. Manual fill methods require significant time costs and constant value replacement is a replacement using the same variables, but is prone to unknown errors. In this paper, the regression clustering difference method is

adopted to speculate the missing value of the existing data with the regression method, and then cluster the data set after filling the missing value. For the data in the same cluster, the regression method is applied again to calculate the missing value for filling. Since the clustering attribute is taken into account, the method allows for more accurate estimation of missing values in the dataset compared to other methods.

After filling in the missing values, feature transformation is required, which is an important part of data pre-processing. Data transformation is used to transform the data into a form that meets the requirements of the algorithm. The basic methods include: feature construction, standardization, discretion and so on. Discretion is adopted in this paper, which discrete some numerical data into standard attributes.

Therefore, in order to map different features to the same scale, this paper reduces some attribute data in proportion and puts the data in a smaller area. After standardization, individual features of the original data are converted to the same quantitative level, allowing for comprehensive measurement and analysis.

### 3.3. Feature Selection and Correlation Analysis

Feature selection plays a very important role in machine learning. It means to select some of the most important feature values from the original features, reduce the dimension of the data set, improve the accuracy of the prediction model and provide a better understanding and explanation for the model. Figure 1 shows the factors that affect the employment rate of computer application students. Using all the factors as input variables of the model will affect the predictive performance of the model. Therefore, this paper selects the most important factors affecting employment through correlation analysis by Meek (Meek et al., 2003), and the correlation coefficient formula is:

$$Correl(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 * \sum (y_i - \bar{y})^2}}$$

where (X,Y) is a pair of variables, $x_i$ is the value of X, $\bar{x}$ is the average of X, $y_i$ is the value of Y, $\bar{y}$ is the average of Y.

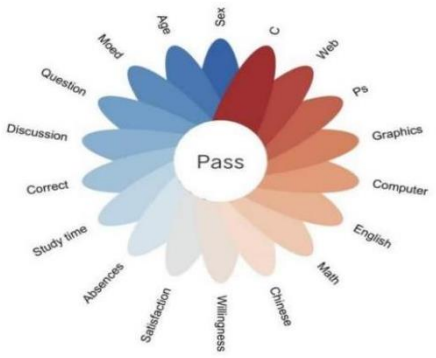


Figure.1 Factors affecting computer application students' employment rate

The colour of the lines pointing to the different indicators in the chord diagram represents the correlation, and the thickness of the lines represents the size of the correlation. The visualization is shown in Figure.2.

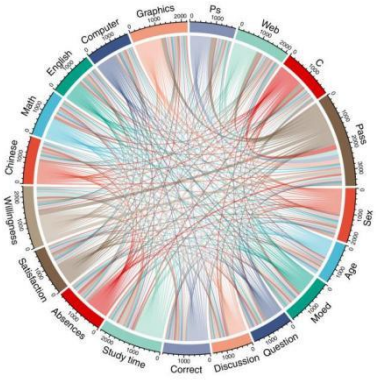


Figure.2 Correlation analysis between different indicators.

It can be seen from Figure.2 that in addition to the strong correlation between each factor and itself, the Pass factor has a strong correlation with C, Ps and Web courses, but has little correlation with Correct factors and Chinese. In this paper, C, Ps, Web and Willingness are selected as the relevant features of the model prediction.

### 3.4. Prediction Model

In this subsection, we apply three classical models for employment prediction, including DT, KNN and SVM. The goal is to select an optimal prediction model, and teachers input students' relevant indicators to obtain prediction results, so as to facilitate timely adjustment of teaching strategies.

The DT model is constructed using a top-down greedy algorithm resulting in a recursive construction. At each internal node, the best test attribute is selected to classify the training sample tree. A decision tree is a tree diagram where the nodes indicate where we select attributes and ask questions, the edges indicate the answers to the questions and the leaves indicate the actual class labels. Generally, it involves three basic steps: feature selection, decision tree generation and decision tree pruning. (Yağcı & Mustafa,2022).This model structure is shown in Figure. 3.

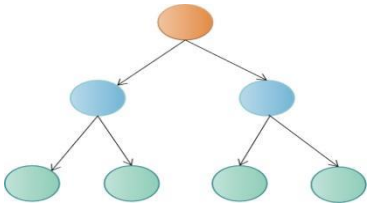

Figure. 3 Structure of Decision Tree

The ANN is an algorithm model used for processing complex information. It can form a "knowledge" input-output mapping relationship through sample training learning, without knowing the equation used to describe the mapping or functional relationship between the input and output. (Xiao, Wen et al., 2022). This model structure is shown in Figure. 4.

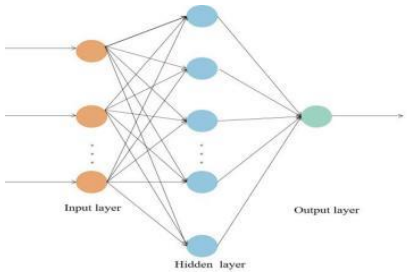


Figure.4 Structure of BPNN

The SVM is a supervised machine learning algorithm used to predict multivariate classification and regression models (Mitra et al., 2022).Each data item is drawn in N-dimensional space, and the kernel is used to non-linearly represent the training information into higher-dimensional feature space, in an attempt to find a hyperplane to separate classes, while minimizing classification errors and maximizing edges. This model structure is shown in Figure.5.

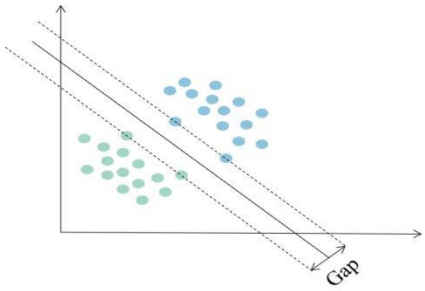


Figure.5. Structure of SVM

## 4. The OBE Model

This study verifies that three machine learning algorithms can be used to predict the future employment of students based on existing educational data from institutions.In order to improve the employment rate of students, we introduce the OBE model and re-plan the teaching in the following way.

### 4.1. Teaching and Learning Mode

The traditional teaching and learning mode only uses classroom time, which is proved to be far from enough in practice, while the OBE concept advocates a blended teaching teaching mode. Before class, students can preview the course content through online platforms such as Rain Class. After class, they can practice and complete their homework through the platform to achieve the purpose of consolidating knowledge and testing knowledge mastery. Teachers can monitor and analyze students' learning process, track students’ learning records, and answer questions online with the help of the platform's data such as student access statistics and average scores in homework exams.

### 4.2. Teaching Methods

The OBE model changes the traditional teaching method based on teachers and supplemented by students into one based on students' independent inquiry and supplemented by teacher's guidance, giving full play to students' main role and improving students' ability of self-inquiry.

### 4.3. Form of Evaluation

The evaluation form based on the OBE concept emphasizes multi-dimensional and continuous process evaluation, which reduces the pressure of teachers' teaching management and motivates students to learn. The procedural evaluation items mainly include attendance, preview, online learning hours, homework, classroom performance, experiments, unit tests and grades, which are mainly represented inFigure.6.

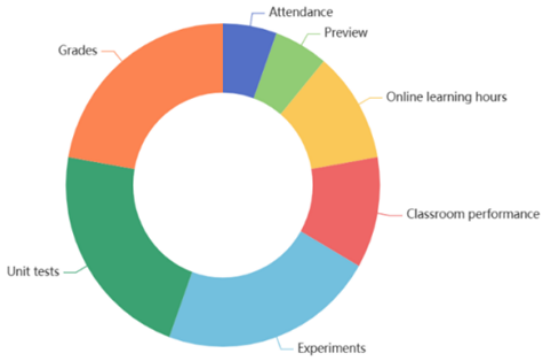


Figure.6 Evaluation form under OBE model

## 5. EXPERIMENTAL RESULTS

### 5.1. Experimental Results of the Three Prediction Models

In order to compare the performance of the three prediction models, Accuracy, Precision, Recall and F-measure (Powers & David MW, 2020) are used to detect the performance of each prediction model, and the experimental results are analyzed and discussed. The performance comparison results are shown in Figure.7.

Accuracy is the ratio of the number of correctly predicted instances to the number of instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (2)$$

Precision is the ratio of correctly predicted positive instances to total predicted positive instances.

$$Precision = \frac{TN}{TN + FP} \times 100 \quad (3)$$

Recall is the ratio of correctly predicted positive instances to all instances in the actual class.

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (4)$$

F1-Measure conveys a balance between recall and precision.

$$F1 - Measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (5)$$

where true positive rate (TP) is the number of instances correctly predicted to be positive, false positive rate (FP) is the number of instances incorrectly predicted to be positive, true negative rate (TN) is the number of instances correctly predicted to be negative, and false negative rate (FN) is the number of instances incorrectly predicted to be negative.

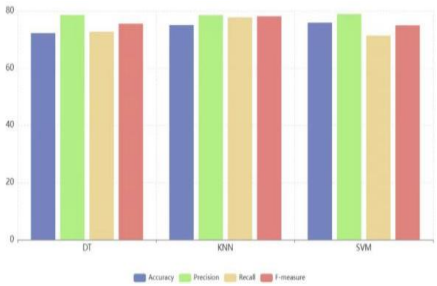


Figure.7 Comparison of the performance of the three prediction models

From Figure 7, it can be seen more intuitively that KNN is superior to the other two models in four performance indicators, among which the accuracy rate can reach 75.37%, the accuracy rate can reach 79.5%, the recall rate is 77.73%, and the F-measure is 79%. KNN shows a better prediction effect in predicting student achievement because the neural network has the hidden layer structure and the weight parameters can be dynamically adjusted, which is more suitable for exploring the potential relationship between certain courses and employment.

Therefore, the KNN model can be used to better predict the employment outcomes of students, which can provide teachers with a reference for improving the education model and facilitating students' learning, thus effectively improving the achievement of talent cultivation goals.

### 5.2. Experimental Results After the Implementation of the OBE Model

In order to assess the teaching effectiveness after the implementation of the OBE teaching model and to analyze the feasibility of the model from the students' perspective, teachers conducted a questionnaire survey on students after the end of each year's course. From December 2020 to December 2022, 500 questionnaires were distributed to computer application students and 480 valid questionnaires were returned. Some of the topics of the questionnaire survey are shown in Table 3.

Table 3 Some of the topics of the questionnaire survey

| Appraisal items | Very satisfied | Satisfied | Not really satisfied |
|---|---|---|---|
| Satisfaction with the OBE model | 63.3% | 36.7% | 0 |
| Satisfaction with stimulating learning | 61% | 36.9% | 2.1% |
| Satisfaction with online learning resources | 65.6% | 30.2% | 4.2% |
| Satisfaction in mastery of knowledge | 59.6% | 37.5% | 2.9% |
| Helping to increase satisfaction with learning | 71.1% | 27.3% | 1.6% |
| Satisfaction with teachers' teaching organization and design | 91.4% | 8.6% | 0 |

Table 3 shows that 63.3% of the students were very satisfied and 36.7% were satisfied with the OBE teaching model. Students had a high degree of recognition that the course in terms of stimulating interest in learning, online learning resources, mastering knowledge and helping to improve learning skills. At the same time, students' satisfaction with teachers' teaching organization and design is as high as 91.4%. From the survey, it can be clearly felt that the OBE model is favored and recognized by students.

In the two years that the OBE model has been integrated into the classroom, the diversified evaluation reasonably reflects the comprehensive effect of students' learning. The statistical results of the distribution of students' grades and employment status at the end of this course are shown in Figure.8.

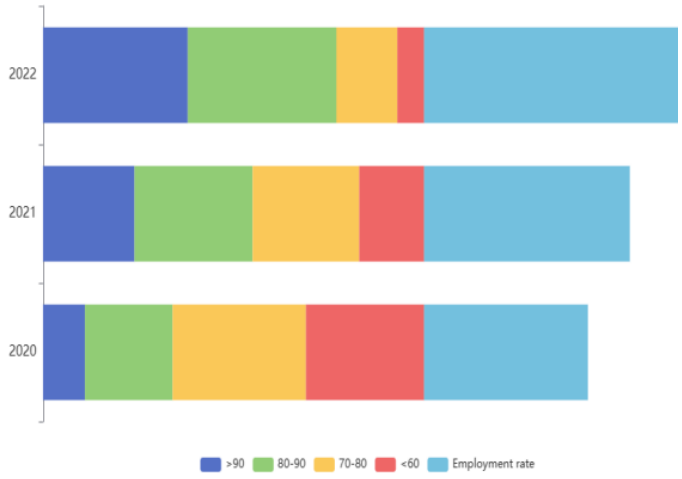


Figure.8 Distribution of students' academic performance and employment rates

From Figure.8, it can be seen that the overall distribution of students' grades is reasonable, with the number of students scoring greater than 90 increasing and the number of students scoring less than 60 decreasing year by year, while the employment rate has increased significantly, thus indicating that this education model has good teaching and learning effects.

## 6. CONCLUSIONS

This paper collects the data set of computer application major of a college. Firstly, we select the characteristics of factors affecting the employment rate of students through correlation analysis. Then, we predict the employment rate of students through DT, ANN and SVM. The results show that the combination of feature selection and ANN can significantly improve the prediction performance. Finally, we infiltrate the reverse instructional design based on the OBE education model into the computer application major of schools. Through comparison, it is found that the employment rate and performance of the class after the implementation of the OBE education model are significantly improved.

In future work, it is hoped that the combination of machine learning algorithms and the OBE education model will improve the quality of teaching and learning in the respective professions, taking into account the country's economic situation and labour market trends, and further increase the employment rate of students.

## REFERENCES


[1] Bakhshinategh, Behdad, et al. "Educational data mining applications and tasks: A survey of the last 10 years." *Education and Information Technologies 23.1* (2018): 537-553.

[2] Ortiz-Lozano, José María, et al. "University student retention: Best time and data to identify undergraduate students at risk of dropout." *Innovations in education and teaching international* (2018).

[3] Dahman, Mohammed R., and Hasan Dağ. "Machine learning model to predict an adult learner's decision to continue ESOL course." *Education and Information Technologies* 24.4 (2019): 2429-2452.

[4] Pojon, Murat. *Using machine learning to predict student performance*. MS thesis. 2017.

[5] Tan, Mingjie, and Peiji Shao. "Prediction of student dropout in e-Learning program through the use of machine learning method." *International journal of emerging technologies in learning* 10.1 (2015).

[6] Li X, Yang T. Forecast of the employment situation of college graduates based on the LSTM neural network[J]. Computational Intelligence and Neuroscience, 2021, 2021: 1-11.

[7] Chen, Yawen, and Linbo Zhai. "A comparative study on student performance prediction using machine learning." *Education and Information Technologies* (2023): 1-19.

[8] Wang, Peipei, et al. "Complexity-based attentive interactive student performance prediction for personalized course study planning." *Education and Information Technologies* (2022): 1-23.

[9] Tseng, Shin-Mu, Kuo-Ho Wang, and Chien-I. Lee. "A pre-processing method to deal with missing values by integrating clustering and regression techniques." Applied Artificial Intelligence 17.5-6 (2003): 535-544.

[10] Meek, M. E., et al. "A framework for human relevance analysis of information on carcinogenic modes of action."*Critical reviews in toxicology* 33.6 (2003): 591-653.

[11] Yağcı, Mustafa. "Educational data mining: prediction of students' academic performance using machine learning algorithms." *Smart Learning Environments* 9.1 (2022): 1-19.

[12] Xiao, Wen, Ping Ji, and Juan Hu. "A survey on educational data mining methods used for predicting students' performance." *Engineering Reports* 4.5 (2022): e12482.

[13] Mitra, Anisha, et al. "Students Performance Prediction Using Educational Data Mining." *Internet of Things and Its Applications*. Springer, Singapore, 2022. 171-183.

[14] Siddique, Ansar, et al. "Predicting academic performance using an efficient model based on fusion of classifiers." *Applied Sciences* 11.24 (2021): 11845.

[15] Yassine, S.; Kadry , S.; Sicilia, M.A. A framework for learning analytics in moodle for assessing course outcomes. *In Proceedings of the 2016 IEEE Global Engineering Education Conference (EDUCON), Abu Dhabi*, UAE, 10–13 April 2016; IEEE: Piscataway , NJ,USA, 2016; pp. 261–266.

## AUTHORS

**Xiaoke Deng** is currently pursuing the masters degree with the School of Information Science and Engineering, Shandong Normal University. Her current research interests include Education, artificial intelligence.

**Huakun Hou** is currently pursuing the masters degree with the School of Information Science and Engineering, Shandong Normal University. He current research interests include Education, artificial intelligence.

**Meiyu Jin** is currently pursuing the masters degree with the School of Information Science and Engineering, Shandong Normal University. Her current research interests include Education, artificial intelligence.

**Linbo Zhai** (Member, IEEE) received his B. S. and M.S. degrees from School of Information Science and Engineering at Shandong University in 2004 and 2007, respectively. He received his Ph.D. degree from School of Electronic Engineering at Beijing University of Posts and Telecommunications in 2010. From then on, he worked as a teacher in Shandong Normal University. His current research interests include cognitive radio, crowd sourcing and distributed network optimization.
*corresponding author

© 2023 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.