Contemporary Mathematics
http://ojs.wiserpub.com/index.php/CM/

Research Article

Constructing a Predictive Model for High School Students' Enrolment Results

Jong-Yih Kuo[1], Yi-Fang Chung[1], Hui-Chi Lin[1*], Ping-Feng Wang[2]

[1]Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei 10608, Taiwan
[2]Department of Artificial Intelligence and Computer Engineering, National Chin-Yi University of Technology, Taichung 411030, Taiwan

E-mail: huchlin@ntut.edu.tw

Received: 22 December 2023; **Revised:** 1 February 2024; **Accepted:** 26 February 2024

Abstract: Nowadays, data-informed decision-making can assist high-school educational decision-making, especially student test scores, and further education data, then these data can be used to inform decision-making in schools and boost overall school performance. We aimed to assess the relationships among high school students' admission scores, academic performance at school, and psychological tests in the school system. The relationships were used to predict the departmental fields and universities that students would select for advanced studies in the future. This research will utilize data mining algorithms, including Naïve Bayes, SVM, and Neural Network classifiers, to extract and analyze hidden patterns in students' academic records and credentials. When the Bayesian Network model was used to predict whether a student's university is public or private, our results showed 78.46% of the predicted results. Predicting the accuracy of the top three departmental fields, the SVM model achieved the highest performance with an accuracy of 69.77% .Given that students' academic performance varies across high schools, each school should develop its customized predictive model using only its students' data. This will help students believe in and pursue their interests.

***Keywords*:** educational data mining, bayesian network, support vector machine, neural network, prediction

MSC: 68T07, 68T09, 65Z05, 62C10

1. Introduction

Data mining is defined as the method of obtaining information and analyzing patterns in large sets of data [1]. Educational Data Mining (EDM) is the process of obtaining knowledge specifically from types of data extracted from educational settings. computational exploration of the underlying historical data could lead to the identification of some consistent patterns that depict the systematic relationships be-tween variables. Due to school managing students every year, schools generally use management systems to record large amounts of data, such as basic student information, curriculum planning and academic performance it's also important to the significance of school management has more increased due to an increase in the number of students and knowledge and expectation. High school students will be affected by many factors when choosing a university department. Some people make choices based on their interests, some people

DOI: https://doi.org/10.37256/cm.5320244149

follow the advice of relatives and friends or follow the expectations of their parents, and some people evaluate the future development of the department and employment trends.

The study provided by Raghavendran et al. [2] shows there are many different problems to face at high school and mentions that acquiring admission to their dream university is one of their main worries. Keser et al. [3] think there are many reasons for students to choose a school and mention that factors such as family, friends, instructors, school environment, class-room environment, social status, and lessons have positive or negative effects on the student during this process. Lin [4] showed that some students also consider development trends and employment opportunities, and some combine their own subjective ideas with external factors. Students can be guided by psychological tests or participating in counselling to help them find their interests. Alghamdi et al. [5] conducted a study that three models were constructed using different algorithms: Naïve Bayes (NB), Random Forest (RF), and Decision Tree J48 algorithms was to predict the achievement of early secondary students based on these factors, the school administration may arrange counselling and guidance to the related students, and it will help improve their success rate.

Under the context of the curriculum reforms in Taiwan, there have been few studies on the impact of the reformed multienrolment system on students' higher education in Taiwan. However, so far, there has been little discussion about it. If the field or faculty that students choose can be predicted based on the results of various psychological tests in the education system, the information could be used as a reference. It could act as a guide for helping students choose their further studies. The broad purpose of this research is to support educational decision-making by determining study choices based on student data and giving useful feedback. Because the uncertainty of students' admission into universities or higher educational institutions after their examinations is a significant research issues, Karthika et al. [6] emphasize the importance of simulating and providing a convenient environment for students regarding their chances of admission to universities. This proposed model by Sharma et al. [7] alleviates the uncertainty in the process of university or higher educational institution admissions. Rawal et al. [8] study presents a joint venture of Naïve Bayes Classification and Kernel Density Estimations (KDE) to predict students' admission into universities or other higher education institutions. This research study aims to reduce the uncertainty associated with gaining admission to universities/institutions by considering previous credentials and other essential parameters. Matar et al. [9] proposed a study focusing on one of the machine learning algorithms, Naïve Bayes, to predict whether a student is getting or failing to get admission to universities. Jie et al. [10] showed that the Naïve Bayes is the predictor classification algorithm and is used to study historical data to generate comprehensive and precise analysis.

Considering the above, this study employs data mining methods to utilize the school's information system for recording students' scores in the junior high school entrance examination before enrollment, and the regular assessment scores after registration. Data analysis, including entrance examination scores and student learning history, helps to understand students' learning outcomes and predict their likelihood of admission to public schools.

The main contributions of this work are as follows:

1. Considering the practical implementation, we utilize the existing school administrative system to store student data in recent years, such as students' pre-enrolment exam scores, three-year high school grades, career guidance materials, and learning history, and select suitable data. Proposes the original data, use classification algorithms, and construct a model specifically for predicting high school students' enrolment based on their abilities and interests.

2. We propose a method that utilizes Bayesian network, Support Vector Machine, and Neural Network to examine the impact of data selection on model accuracy, the variation in probability values within each probability ranking, and other factors. This approach aims to gain insights into the factors influencing decision-making during model forecasting, in order to optimize model performance. To predict whether students will be admitted to public schools or be able to choose a suitable department according to their own interests.

3. Since the promotion of the 108 Curriculum in China, the high school curriculum has focused on self-directed learning to enable students to fully explore their own interests and goals. We analyzed and design features that may affect whether they are chosen as an ideal college. We experimentally prove that the Bayesian network, Support Vector Machine, and Neural Network are used as two algorithms that can predict high school graduation student success in admission to an ideal college. Utilizing SVM and selecting RBF kernel to handle multi-class classification problems, observation students choose their favorite department.

4. Finally, we propose the predicted results of further education and the information produced in the process of prediction. Our study contributes to extending the early identification of students who are involved in this process and understanding the students' outcomes of further education.

This study constructed a prediction model for high school students for use by each school in developing its own prediction model. The model uses student data stored in the school administration system of a school and selects suitable information as factors affecting the choice of higher education. The data can be used for the classification and construction of the predictive model; this model can be used to predict high school graduation student success in admission to an ideal college.

The rest of this paper is divided into the following sections. Section 2 introduces the related literature, and introduces terms related to the education field and algorithms related to this research. Section 3 introduces the research process, acquired datasets, data prepossessing methods, and application methods of data exploration algorithms used in this research. Section 4 explains the method of implementation and describes the results of the experiment and the analysis report. Section 5 concludes and discusses directions for future research.

2. Related work

2.1 *Departmental exploration scale*

The Departmental Exploration Scale is one of the psychological tests designed and conducted by the College Entrance Examination Canter (CEEC) for high school students. The test changes the students' preference for 30 kinds of knowledge into scores from 0 to 15 as the knowledge score and then performs Pearson's correlation with the knowledge scores of each department to measure the suitability. This can be used by students as a reference for future selection of a university (CEEC) [11].

2.2 *Standard classification of education*

There are two ways to classify the university's disciplines. One of them is established by Taiwan's Ministry of Education for college students and professors. Based on the International Standard Classification of Education published by the United Nations Educational, Scientific and Cultural Organization (UNESCO), departments are divided into 11 fields, 27 narrow fields, 93 detailed fields, and 174 sub-detailed fields. The other way to classify university disciplines has been established by the CEEC and applied to psychological tests, such as the Interest Scale and the Departmental Exploration Scale. The latter is mostly used and is familiar to high school students. According to such factors as learning content, interest, and future development, the departments are classified into 18 discipline clusters and 126 detailed fields. However, each discipline cluster is not exclusively independent, and thus, some departments may be classified into two discipline clusters simultaneously; for this reason, the classification result might not be suitable. Therefore, the first method to classify departments is used in this research.

2.3 *Data mining*

Data mining involves effectively analyzing a database and finding potentially useful information. The step of Knowledge-Discovery in Database (KDD) is used to analyze and search for rules, as shown in Figure 1. Predictive data mining can be divided into descriptive and predictive categories. Educational data mining encompasses various subfields, such as predicting student achievement, analyzing teaching deficiencies, assessing students' adaptive learning abilities, and automatically identifying disparities in academic outcomes. These subfields can assist educators in designing courses to enhance teaching effectiveness [12], providing administrators with resources and tools for decision-making and organization. It also focuses on the impact of students' evaluation scores and their activities to demonstrate the relationship with academic performance [13, 14].

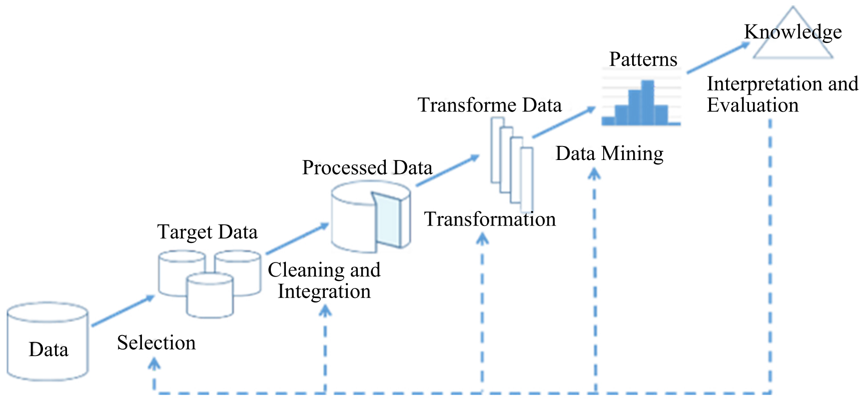


Figure 1. The Knowledge-Discovery in Database (KDD) process

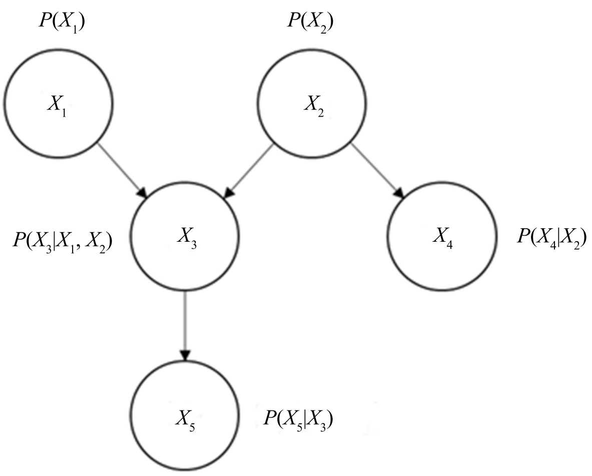
Almasri et al. [15] proposed a predictive model that would be used to predict the performance of the students, especially those having low performance, and understand the reasons behind such results. Miguéis et al. [16] proposed that early classification of students based on their potential academic performance could be a useful strategy for reducing failure, promoting better performance, and better managing the resources of higher education institutions. Jia et al. [17] emphasized that predicting students' academic performance is very important for their future development. Educational data mining encompasses various subfields, such as predicting student achievement, analyzing teaching defects, assessing students' adaptive learning abilities, and automatically identifying disparities in academic outcomes. Huang et al. [18] among others, proposed a hybrid method for student performance prediction, mainly as a methodological reference for student performance prediction contributing to student performance management. These studies show the versatility of the development of EDM, and growing research on the topic can help re-searchers better understand educational structures and assess learning effectiveness.

2.4 *Classification*

Classification is one of the supervised learning techniques in data mining. The process is based on the attributes of the data, which are then used to create a reliable classification model or classification rules. We can calculate the accuracy of the classification result and use this model or rule to predict the class of new, unclassified data. Several classification algorithms have been used, including Naïve Bayes Classifier, Decision Tree, Bayesian Network, k-nearest neigh-bours (KNN), Support Vector Machine (SVM), Random Forests, and Neural Network by Irfiani et al. [19]. The related research on classification is very rich. Taking the research in recent years as examples, Irfiani et al. [19] collected such attributes as students' cognitive value, affective value, and the number of times they participated in mock exams and used Decision Tree and Naïve Bayes Classifier to predict whether students' performance improved. This helped reduce the number of students who failed and improve the overall quality of education. Kondo [20] extracted such behavior as logins, browser notifications, and initiating or submitting homework can be tracked from the login records of the online learning management system. They used the Bayesian Network to under-stand students' learning status and predict the probability of better performance. Ragab et al. [21] used three traditional techniques-neural network, decision tree, and a simple Bayesian classifier-to construct a model for predicting student learning effects. Combining the proposed method with different classifiers, such as Boosting, Random Forest, Bagging, and voting algorithms, the results can more accurately predict student learning outcomes and effectively over-come the decline in academic performance. Kusumawardani et al. [22] proposes, based on deep learning transformer encoding, to sequentially predict the student's final performance based on log activities provided by an LMS. The results show that the model could predict at an early stage with an accuracy of withdrawn versus pass-distinction classes. Ismail et al. [23] proposed a deep learning method for the internal assessment of students. The use of machine learning techniques to forecast student performance has proven useful for identifying underachievers and enabling tutors to implement corrective actions earlier.

2.5 *Bayesian network*

The Bayesian Network is developed based on Bayes' theorem, as shown in equation (1). It is a probabilistic graph model. A directed acyclic graph is formed from the nodes representing the random variable and the arrows representing the causal relationship be- tween two random variables, as shown in Figure 2. This yields the conditional probability distribution of each node and the conditional probability table can be used to show the relationship between each "child node" and its "parent node." Mathematically, Yang et al. [24] by the joint probability distribution of the Bayesian Network is expressed, as shown in equation (2), to represent the set of local conditional probability distributions $P(Xi \mid pa(Xi))$ of each node and its parent node.



$$P(X_1, X_2, \ldots, X_5) = P(X_1)P(X_2)P(X_3 \mid X_1, X_2)P(X_4 \mid X_2)P(X_5 \mid X_3)$$

Figure 2. Example of Bayesian Network architecture

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \tag{1}$$

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{N} P(X_i|pa(X_i)) \tag{2}$$

$P(X_1, X_2, \ldots, X_5) = P(X_1)P(X_2)P(X_3 \mid X_1, X_2)P(X_4 \mid X_2)P(X_5 \mid X_3)$

Bayesian network has the advantages of clear and intuitive problem solving, and has become a research hotspot in academia. It has been widely used in pattern recognition, decision support, medical diagnosis, data mining, sensor fusion, and automatic control, etc. Athani et al. [25] used the Naïve Bayesian (NB) algorithm to predict student academic success and behavior. The goal of this study is to use data ex-traction techniques to help educational institutions gain insight into their educational level, which can also be useful in enhancing the academic performance of students. Shi et al. [26] propose a dropout prediction model based on Bayesian networks (Dropout Prediction Bayesian Net-work, DPBN), use collect the individual learning data of learners and form a feature vector to per-form dropout prediction using Bayesian network inference. Kerdvibulvech et al. [27] used the skin-colored region of a user is segmented by adaptively and automatically applying a Bayesian classifier. Afterward, a matching algorithm is used to determine the probabilities of the fingertips based on their basic elements. Following this, a particle filter is used with a deterministic clustering algorithm to track fingertips. Based on these characteristics, we believe that Bayesian Networks can be used in this experiment to predict the field or department that students will choose. By utilizing the results of various psychological tests within the existing education system, we can provide valuable reference information.

However, performing Bayesian inference in complex models will always be a challenge. The art of probabilistic modeling lies in balancing two objectives: creating models that accurately capture the generative processes behind our data, while also ensuring that these models allow for efficient inference algorithms. Linderman et al. [28] used an approach that combines the GLM with flexible graph-theoretic priors governing the relationship between latent features and neural connectivity patterns. Fully Bayesian inference via Pólya-gamma augmentation of the resulting model allows us to classify neurons and infer latent dimensions of circuit organization from correlated spike trains. Chutisant et al. [29] proposed a vision-based method for tracking the fingerings made by guitar players. Utilize a Bayesian classifier to learn color probabilities from a small training image set and then adaptively learn the color probabilities from online input images. By applying this method, the first appealing aspect is that it can alleviate the burden associated with manually generating a large amount of training data. Xu et al. [30] develop a novel learning method to deal with the finite resolution of system log file time stamps, without losing the benefits of our continuous time model. Use continuous time Bayesian networks (CTBNs) and avoid specifying a fixed time interval. Build generative models from historic non-attack data, and flag future event sequences whose likelihood under this norm is below a threshold. Kerdvibulvech et al.[31] introduced a marker less framework for tracking a colored remote-control car by integrating a Bayesian classifier into particle filters. This feature adds the useful abilities of automatically initializing tracks and recovering from tracking failures in a dynamic background.

2.6 *Support vector machine*

SVM can be used for classification and regression analysis. By finding a decision boundary, the margin between the two classes can be maximized. If the decision boundary is a hyperplane, it indicates a linear classification problem. For nonlinear classification problems, the data can be mapped to a higher-dimensional space H (Hilbert space, also known as feature space) through a kernel function composed of nonlinear mapping function φ to make it linearly separable. The definition of the kernel function is shown in equation (3). $K(X_i, X_j)$is the kernel function, φX_i is the training vector, $\varphi(\mathrm{X}i)$ is the mapping function, and $\varphi(X_i)^T$ $\varphi(\mathrm{X}j)$ represents the inner product of X_i and X_j mapped to Hsu et al. [32] mentioned that the four basic kernel functions include Linear Kernel, Polynomial Kernel, Radial Basis Function Kernel (RBF Kernel), and Sigmoid Kernel.

$$K(X_i, X_j) \equiv \varphi(X_i)^T \varphi(X_j) \tag{3}$$

SVM is a binary classifier, but it can also be used to deal with multiclass classification problems. We can use the results of Huang et al. [18] who proposed one-against-one and one-against-all approaches to construct multiple binary classifiers in order to obtain the most suitable classification results for multiple classes. Cardona et al. [33] proposed the application of support vector machines (SVM) to predict degree completion within three years by STEM community college students, then used SVM enabling the classification of the input variables into expected classes, completion and not completion, by maximizing the margin between the points from the different classes in order to constrain the misclassification.

Most current methods for predicting students' performance are focused on achieving the highest prediction accuracy, but they overlook the pedagogical principles underlying the predictors and the technical limitations of teachers in actual teaching. This paper proposes a comprehensive comparison of methods and a comparative study using data to explore a prediction model for high school students' enrollment. In this paper, we selected Bayesian networks and SVM as classification algorithms, when predicting students' future choice of subject areas, consider which attributes will affect the prediction results. Therefore, in addition to considering both students' academic ability and interest as attributes, experiments are also conducted to solely focus on academic ability and interest separately. This helps to identify the current trend of various factors that affect students' choice of subject areas.

2.7 *Neural network*

Neural Network (NN) is a model that imitates the structure and functionality of biological neural networks [34]. It consists of many interconnected neurons, where each neuron functions like a nerve cell in a biological neural network, as shown in Figure 3. Each neuron has a threshold (b), which is added to the received signal value. After undergoing a transformation through an activation function, the output is passed to the next neuron. The signal received by a neuron is the sum of the products of the input values X_i from the previous layer's neurons and their corresponding weights W_i. The formula for the neuron's output can be seen in equation (4).

$$y = f\left(\sum_{i=1}^{m} W_i X_i\right) \tag{4}$$

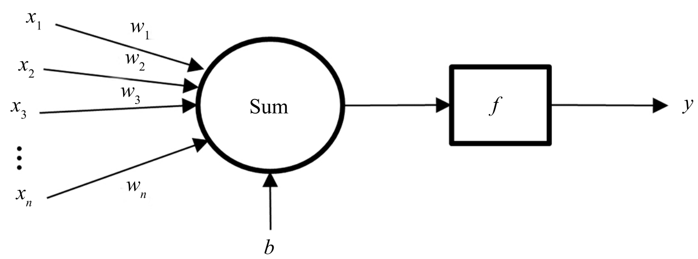


Figure 3. Diagram of Neuron architecture

The Multilayer Perceptron (MLP) consists of an Input Layer, one or more Hidden Layers, and an Output Layer. It is a type of Feedforward Neural Network, where signals are transmitted from the Input Layer to the Hidden Layer and then to the Output Layer, with no loops between neurons. The Input Layer receives feature vectors of data, and the number of neurons in this layer corresponds to the number of features in the data. The Hidden Layer are located between the Input and Output Layers, and the number of layers and neurons is adjusted based on the complexity of the model. The Output Layer provides the final output results, and the number of neurons in this layer is equal to the number of classes in the prediction results. Each neuron in every layer is only connected to the neurons in the adjacent layers, both above and below. Figure 4 shows the architecture of a Multilayer Perceptron, where X represents the input features, and Y represents the output results. The value of each neuron in the Hidden Layer transmitted to the neurons in the next layer can be calculated using equation (5). In this equation, Y_j represents the output of the j neuron, f is the activation function of the neuron, net_j is the weighted sum of the j Hidden Layer neuron, W_ij is the weight corresponding to the i input for the j neuron, X_i is the input signal to the i neuron, and b_j is the threshold of the j neuron. N_{INPUT} denotes the number of input features, and N_{HIDDEN} represents the number of neurons in the Hidden Layer. To compute the output of the Output Layer, one only needs to modify the range of i to the number of neurons in the last Hidden Layer and the range of j to the number of neurons in the Output Layer.

$$y_j = f(net_j) = \sum_{i}^{N} W_{ij} X_i - b_j,\ j = 1 \cdots N_{INPUT},\ j = 1 \cdots N_{HIDDEN} \tag{5}$$

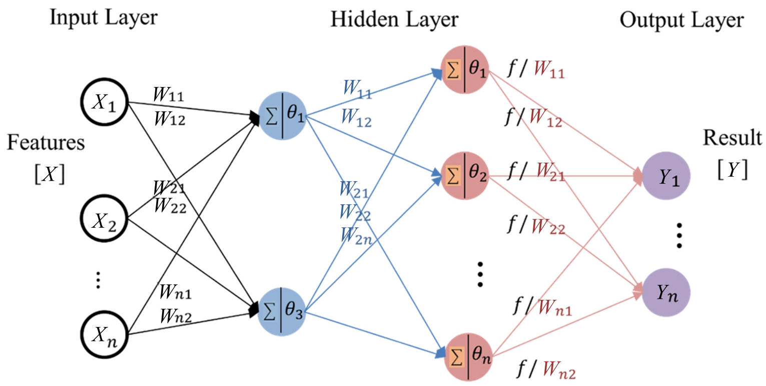


Figure 4. Diagram of Multiplayer Perception architecture

3. Research methodology

3.1 *Research process*

In a first stage, is the collection of raw data. The objective is to understand the data and their variables. The analysis also includes data pre-processing and data cleaning. The research process is described in Figure 5.

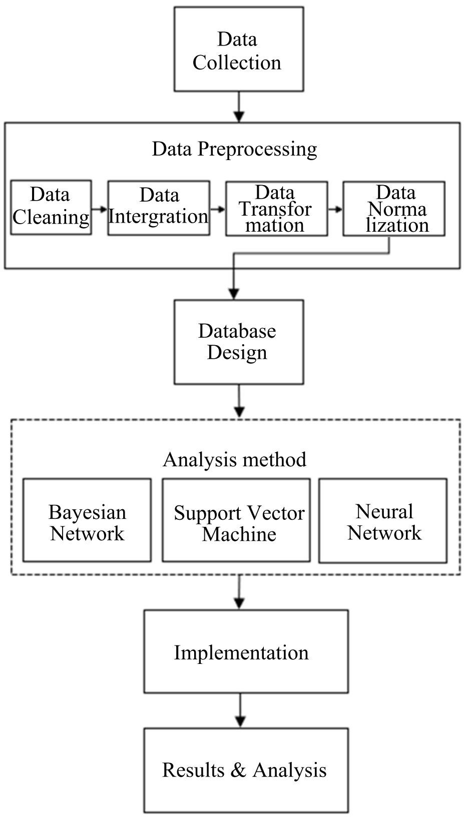


Figure 5. Research flow diagram

3.2 *Data collection*

The data used in this study were sourced from the Ministry of Education's Department of K-12 Education Administration high school and vocational school administrative systems database, as well as from the academic information system of a specific school. The data provided by the K-12 Administration includes students' basic information, class details, course enrollment records, academic history, subject average scores, and semester grades, among other types of data. The specific high school provides information on students' entrance examination scores, interest test data from psychological assessments, knowledge scores from the major exploration scale, academic program fit results from the major exploration scale, subject competency scores, and college admission information for various departments and schools. The data tables include the Department Exploration Scale Knowledge Score and the Department Exploration Scale Student Category Adaptation Result. A total of 23 data tables and 218 columns were redesigned and planned for predictive analysis.

3.3 *Data preprocessing*

The quality of data and the extraction of features greatly impact the results of machine learning analysis. This study focuses on four types of data preprocessing techniques:

3.3.1 *Data cleaning*

In this phase, we use the approach proposed by Samuel et al. [35] to conduct the pre-processing and handle imperfect data. First, this research excluded null data and data that did not match the attributes from the database. Clear data with missing values, inconsistent data types, or illegal characters in the original data, so that the data becomes a data value and type suitable for model access. In this research, missing data, such as NULL data in the data samples in Table 1 were deleted. The data is inconsistent with the data type and contains invalid values, such as the "Absent" of the data sample in Table 2.

Table 1. Data sample (1) before cleaning

Class	No	School Name	Department	Entry Score
301	1	Hsuan Chuang University	Department of Applied Psychology	5B
301	3	NULL	NULL	2B3C
302	4	Chinese Culture University	Graduate Institute of History	3B2C
307	18	NULL	NULL	4B1C

Table 2. Data sample (2) before cleaning

Registration No	Chinese Score	English Score	Math A Score	Math B Score	Society Score	Science Score
O 30101	8		-	2	11	-
O 30103	11	2	3	3	12	6
O 30119	Absent	Absent	-	-	Absent	-
O 30702	10	10	3	-	-	8

3.3.2 *Data integration*

Combine data from various sources, such as databases, multi-dimensional data, or files, into a single database. The data for this research comes from two sources: the K-12 Education Administration, MOE and a high school.

3.3.3 *Data transformation*

In this research, the rankings and total number of students in each semester were converted into percentage rankings. Furthermore, the total scores from the different sections of the General Scholastic Ability Test (GSAT) were transformed into five standard grades, as shown in Tables 3 and 4.

Table 3. Semester ranking score transformation

Semester	School Ranking (%)
First Semester	10
Second Semester	11
Third Semester	16
Forth Semester	5

Table 4. Subject score transformation

Subject	School Ranking (%)
Chinese	65
English	11
Math	16
Geography	68
History	76
Civics	68
Physics	58
Chemical	40
Biology	81
Geoscience	58

Additionally, convert students' original scores in Chinese, English, mathematics, social studies, and science from the academic test or academic test simulation into letter grades. The grade range is from 0 to 15. The five-level standard for the social group is determined by the total score of four subjects: Chinese, English, Mathematics, and Social Studies, corresponding to the five-level standard for all candidates in the social group. The five-level standard for the natural group is determined by the total score in four subjects: Chinese, English, mathematics, and natural science. The benchmark for all candidates in the natural sciences group is to attain a total score in Chinese, English, and Mathematics that aligns with the national standard for English and Mathematics. The GSAT score transformation is presented in Table 5.

Table 5. GSAT score transformation

Student	Attribute	Fifteen Standards	Five Standards
A	Chinese	10	2
	English	12	4
	Math & Social	27	5
B	Chinese	10	3
	English	13	5
	Math & Social	17	3

3.3.4 *Data normalization*

Due to the significant differences in the range of features in the original data, certain features with larger values may dominate the entire dataset and influence the importance of other features. Therefore, in this study, the Min-Max Scaler method is used for feature scaling, which scales the original feature values proportionally into the range [0, 1], as shown in Equation (6).

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \in [0,\ 1] \tag{6}$$

3.4 *Analysis method*

After pre-processing, the data were divided into a training dataset and a test dataset. The data mining algorithms used include Bayesian Networks, SVM and Neural Network. In Bayesian Networks, structural learning is utilized to identify the influential relationships between nodes and to construct the most appropriate network architecture. These methods search for optimal network architectures from all possible configurations formed by the network nodes. They employ a scoring method to evaluate the quality of each network architecture and assign scores to all potential architectures based on their suitability for the data. Finally, the architecture with the highest score is selected for implementing the Bayesian network. The process is shown in Figure 6. When using SVM, RBF Kernel was used to classify the data; grid search is used to find the optimal hyperparameters of a model, and probability estimation in obtaining the probability of classification into each class. The process is shown in Figure 7. When using NN, the grid search method will try all the preset hyperparameter range combinations, and bring the various permutations and combinations of hyperparameters into the model for training. And finally, select the best hyperparameter combination. The process is shown in Figure 8.

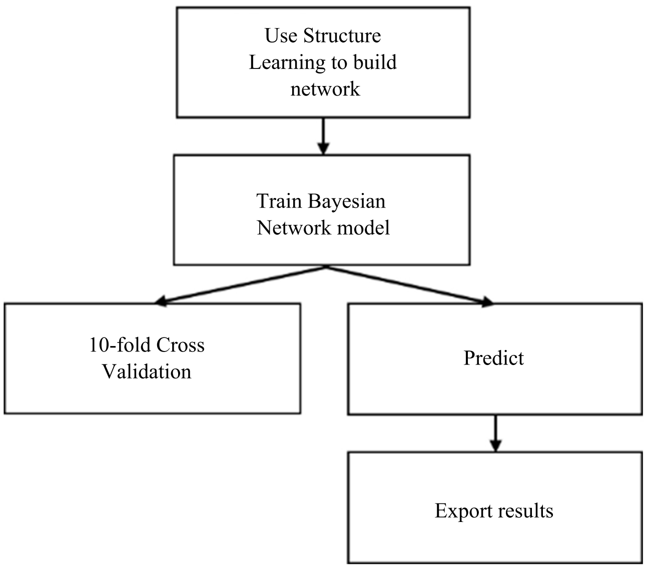


Figure 6. Analysis flow of Bayesian Network

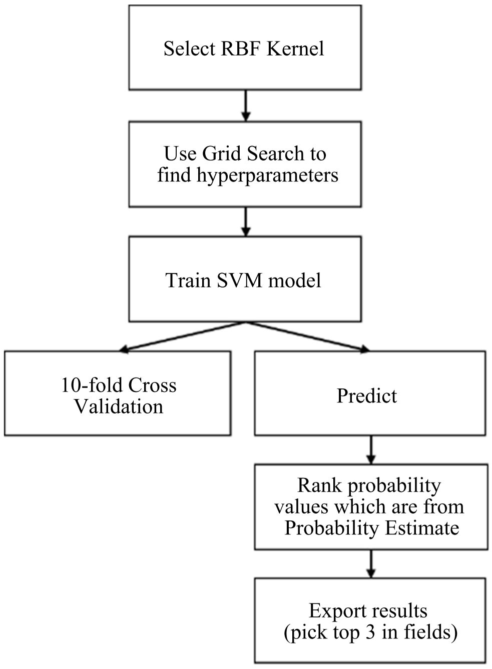


Figure 7. Analysis flow of support vector machine (SVM)

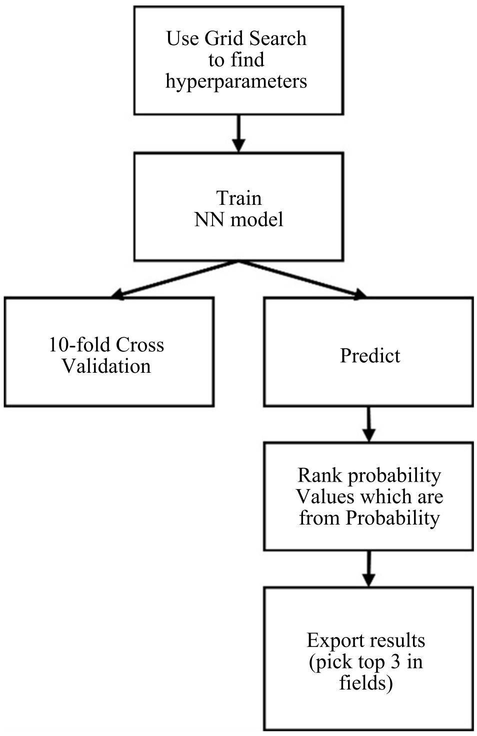


Figure 8. Analysis flow of Neural Network (NN)

3.4.1 *Structure learning*

In structure learning, numerous learning algorithms can construct an appropriate Bayesian Network structure from the data. This research utilized the K2 algorithm, which is a score-based algorithm, to construct the network structure. The K2 algorithm can effectively narrow down the search range, and it offers the advantages of enhancing learning efficiency and faster construction speed, ultimately leading to a better structure.

3.4.2 *The radial basis function kernel*

Hsu et al. [32] showed that the use of The Polynomial Kernel has more hyperparameters than RBF Kernel. When there are more hyperparameters, the model becomes more complicated. For these reasons, this research selected RBF Kernel as the kernel function of SVM.

3.4.3 *Grid search*

The classification performance of SVM depends on the choice of kernel function and the selection of appropriate hyperparameters. In this research, RBF Kernel was selected as the kernel function. The two main hyperparameters that need to be adjusted are C and γ. C (Cost) is a penalty parameter used to balance the size of the classification margin and the accuracy; γ is a parameter of RBF Kernel, which affects the number of support vectors. Grid Search is a method of adjusting hyperparameters. Hsu et al. [32] use an exhaustive search method to find a set of hyperparameters with the best performance within a specified range so that SVM can be trained with the adjusted hyperparameters. It can help improve the accuracy of the model, predict results, and aid in cross-validation to prevent over-fitting. The Grid Search used in this research was implemented before training the SVM model to find the C and γ that would give the model the best performance.

3.4.4 *Probability estimation*

The prediction of departmental fields involves a multi-class classification problem. The factors for choosing departments are quite diverse, and multiple classes are listed as the predicted result. The classification results are ordered based on the probability of being sorted into each class. The probability estimate is not implemented under standard SVM. However, Platt [36] proposed a post processing method that uses the sigmoid function to map the output of a single SVM to the posterior probability, while Lin et al. [37] improved the algorithm to deal with the problem of overflow so that SVM can be used to implement the probability estimate.

3.4.5 *Cross-validation*

Cross-validation is one of the most widely used data resampling methods to assess the generalization ability of a predictive model and to prevent over-fitting. K-Fold cross-validation is a form of cross-validation. After k times of validation, the average of all accuracy is the result of cross-validation. This research used K-fold cross-validation and assigned that use a value of 10.

4. Implementation method

4.1 *Introduction of Implementation*

n this study, the Bayesian Network, SVM, and Neural Network were selected as the classification algorithm, and the model was implemented in Python. We used Bnlearn to learn the structure of the Bayesian Network, Pomegranate to implement the Bayesian Network model, and applications LibSVM Chang et al. [38] present solving SVM optimization problems were used to implement the SVM model and Pandas was used to read and operate the data and export the data to Microsoft Excel.

4.2 *Experiment result*

The accuracy of this research in predicting whether a student's university is public or private is demonstrated in Table 6. The study found that the three training models used in this study accurately predict the establishment of the university. Among them, the Bayesian Network achieved a cross-validation accuracy of 86.25% and a prediction accuracy of 78.46%, making it the best model.

Table 6. The Accuracy of Predicting Students' Universities

Algorithm	Accuracy of cross-validation	Accuracy of predicted results
Bayesian Networks	86.25%	78.46%
SVM	84.40%	71.67%
Neural Network	81.71%	76.74%

Furthermore, the data used in the model for this research covers students' interests and academic abilities. The students' interest indicators are based on the 30 knowledge scores from the psychological test results, organized by the school counselor as the features. Use the GSAT score and the grade ranking of each subject as the indicators of academic ability. Predicting which departmental field students may choose as their professional focus from 11 available options. The students' interest indicators are based on the 30 knowledge scores from the psychological test results, which are organized by the school counselor as the features. Use the GSAT score and the grade ranking of each subject as indicators of academic ability. Predicting which departmental field students may choose from 11 professional fields. Table 7 presents the prediction accuracy of the department's field for high school students who have been admitted.

Table 7. The Accuracy of Predicting Students' Departmental Fields

Algorithm	Accuracy of cross-validation	Accuracy of predicted results
Bayesian Networks	26.05%	20.93%
SVM	32.29%	34.84%
Neural Network	34.90%	27.91%

The results indicate that, in addition to students' interests and academic abilities, other factors influence their selection of university departments. The probability of each category prediction is rarely dominant. It is possible that the second and third fields also have a high probability of prediction, which can lead to lower prediction accuracy.

Therefore, this research uses probability estimation to determine the prediction probability of each field and selects the top three fields with the highest probability as the correct prediction result. It means that if the model's predictions match the second and third probabilities, they are also considered correct. The table in Table 8 displays the predicted accuracy of the top three departmental fields. The Bayesian Networks model achieved an accuracy of 62.79%. The Neural Network model achieved an accuracy of 67.44%. While the SVM model achieved an accuracy of 69.77%, it is the highest among the three models.

Table 8. Performance of field of department (top three)

Algorithm	Accuracy of cross-validation	Accuracy of predicted results
Bayesian Networks	61.86%	62.79%
SVM	63.90%	69.77%
Neural Network	65.24%	67.44%

In addition, this study compared the best experimental results obtained by applying the Naïve Bayes Algorithm prediction model to predict student learning. The comparison included the works of Sutoyo et al. [39], Sembiring et al. [40], Rawal et al. [8], and Matar et al. [9] using relevant literature for comparison. The comparison results are shown in Table 9.

Table 9. Comparison with Other Studies

	This study	Sutoyo et al. [39]	Sembiring et al. [40]	Rawal et al. [8]	Matar et al. [9]
Accuracy of predicted results	78.46%	73.73%	70.83%	70.46%	72%

Table 9 presents a comparison of the experimental results from this research with those of related studies. The table demonstrates that the Naïve Bayes Algorithm model proposed in this research achieves the highest prediction accuracy (78.46%), which is considered high compared to accuracies obtained in most other research studies addressing similar problems. As per the previous research study, Sutoyo et al. [39] used the Naïve Bayes algorithm to predict a model with an accuracy level of 73.73%, and Sembiring et al. [40] verified 70.83%. In our research, the predictive model produced a high level of accuracy of 78.46%. This indicates that it is superior to other machine learning algorithms. Furthermore, the researcher concluded that using algorithms to construct a Support Vector Machine model based on Naïve Bayes classification for distinguishing continuous students' performance could be beneficial. The study suggested that predicting academic performance based on previous performance could assist students in gaining admission to public or private universities. We also take students' interests (using 30 knowledge scores) and academic abilities (using the scores of each course during the school period and scores of GSAT) into account to make predictions. In comparison, the other two researchers only apply this to students' entrance into universities/institutions and use only three admission parameters, namely, academic ability, interest, and GSAT scores. Rawal et al. [8] used GRE, GPA, and RANK as parameters, and was limited to datasets gathered over one semester and for one course by Matar et al. [9]. We propose a method with dataset comprised of more courses to improved accuracy.

The research found a gap between the accuracy of predictive models and recent trends in academic admission uncertainty. This is because different institutions are using various parameters, such as entrance tests, GRE scores, and language scores. The results suggest that these are all factors affecting students' achievement. Institutions can fully leverage learning analytics by adopting an analytical approach over a few years, rather than a one-off. We had access to the 2017 course data for training purposes, which was validated using the 2020 data. If we had more historical datasets for training the classifier, we might have been able to fully utilize the tools to enhance the predictive accuracy of our model. Therefore, our ultimate conclusion is that educational institutions should adopt a long-term perspective on educational data mining (EDM) to effectively identify meaningful patterns in the data.

4.3 *Experiment analysis*

4.3.1 *Model evaluation*

In addition to accuracy, precision and recall are indicators for evaluating model performance, as shown in Tables 10 and 11.

Table 10. Evaluation of Model Performance for Predicting Students' Universities

Algorithm	Accuracy	Precision	Recall
Bayesian Networks	86.25%	78.23%	74.67%
SVM	84.40%	78.74%	55.12%
Neural Network	81.71%	76.74%	50.01%

Table 11. Evaluation of Model Performance for Predicting Students' Departmental Fields

Algorithm	Accuracy	Precision	Recall
Bayesian Networks	26.05%	20.93%	13.10%
SVM	32.29%	38.84%	23.76%
Neural Network	34.90%	27.91%	23.61%

4.3.2 *Comparison with the results of other algorithms*

In addition to the Bayesian, SVM, and Neural Network, we also utilized Python to implement the classification models of Random Forest and KNN. The accuracy comparison is presented in Tables 12 and 13. Results from Table 12 indicate that when predicting a university, the Bayesian Network demonstrates the best performance among algorithms, while the performance of other algorithms is also commendable. Results from Table 13 indicate that the Neural Network demonstrates the best performance among the algorithms for predicting the departmental field, while the Bayesian Network performs the worst. By employing multiple algorithms for prediction, we were able to identify suitable algorithms for different prediction tasks.

Table 12. Comparison of Other Algorithms Predicting Students' Universities

Algorithm	Accuracy of cross-validation	Accuracy of predicted results
Bayesian Networks	86.25%	78.46%
SVM	84.40%	71.67%
Random Forest	78.00%	75.00%
KNN	79.43%	75.00%
Neural Network	81.71%	76.74%

Table 13. Comparison of Other Algorithms Predicting Students' Departmental Fields

Algorithm	Accuracy of cross-validation	Accuracy of predicted results
Bayesian Networks	26.05%	20.93%
SVM	32.29%	34.84%
Random Forest	28.52%	16.92%
KNN	26.29%	23.08%
Neural Network	34.90%	27.91%

4.3.3 *Attribute selection*

e considered which attributes affect the prediction results when forecasting the departmental field. Therefore, in addition to using students' academic ability and interest as attributes simultaneously, we also experimented with using academic ability only and interests only. The purpose of these experiments was to identify the factors that could influence students' choice of department. Table 9 presents the experimental results. Thirty knowledge scores were used to measure interest, while the scores from each course during the school term and GSAT scores were used to assess academic ability. The top three models for predicting the probability of the field of the department are shown in Table 14.

Table 14. The Impact of Selected Attributes on Accuracy when Predicting Students' Departmental Fields

Algorithm	Interest,Academic ability	Interest	Academic ability
Bayesian Networks	62.79%	62.79%	41.86%
SVM	69.77%	65.12%	58.14%
Neural Network	69.77%	69.77%	48.84%

As shown in Table 14, Bayesian Networks models incorporate both the attributes of interest and academic ability, utilizing the same attribute for the characteristic of interest. The accuracy of predicting the student's field department is 62.79%, notably exceeding that of using academic ability feature attributes. Additionally, SVM models utilize two key attributes, namely interest and academic ability, to predict students' field of study department with an accuracy of 69.77%. This level of accuracy is notably higher than the prediction based solely on interest or a single characteristic attribute of academic ability. Furthermore, neural network models integrate both the relevant features and academic abilities, utilizing the same attributes for the characteristics of interest. The accuracy of predicting a student's field of study is 69.77%, notably exceeding that of using academic ability characteristics.

On the whole, Support Vector Machine (SVM) and Neural Network models demonstrate the highest accuracy in predicting students' departments and fields by utilizing two key attributes: interest and academic ability. However, neural network models outperform Bayesian Networks and SVM models in predicting students' departments and fields using only a single characteristic attribute of interest, achieving higher accuracy. However, when using a single characteristic attribute of academic ability to predict students' subject fields, SVM models perform the best, significantly outperforming Bayesian networks and neural network models.

When predicting students' university choices, the attributes to be considered include each student's admission score, academic performance at school, progress trend, and GSAT score. In this section, the GSAT scores were specifically divided into three attributes: majoring in social studies, majoring in science, and the main subject. Here, we utilized Bayesian Networks, which can produce high accuracy for the experiment. The experiment results are presented in Table 15. We found that using all three attributes simultaneously (which also served as the method for predicting students' universities in this research) yielded more accurate results compared to the other experiments.

Table 15. The Impact of Selected Attributes on Accuracy when Predicting Students' Universities

Algorithm	Society subject, science subject, and main subject	Society subject, science subject	Main subject
Bayesian Networks	69.77%	67.44%	69.77%
SVM	79.07%	76.74%	74.42%
Neural Network	76.74%	76.74%	76.74%

As shown in Table 15 of the Impact of Selected Attributes on Accuracy when Predicting Students' Universities, the analysis is based on three key attributes: society subject, science subject, and main subject. The SVM models demonstrate an accuracy of 79.07% in predicting students' universities, which is significantly higher than the accuracy of the other two models. When comparing the two feature attributes of the Society subject and the Science subject, the impact of these specific attributes on the accuracy of predicting students' university choices reveals that SVM and Neural Network models achieve an accuracy of 76.74%, notably higher than Bayesian Networks. When predicting students' university choices based on the single feature attribute of the main subject, neural network models outperform Bayesian networks and SVM models in terms of accuracy. They achieve a 76.74% accuracy rate, demonstrating the impact of selected attributes on accuracy.

4.3.4 *Hyperparameter selection*

When using SVM to predict, the optimal hyperparameters C, γ were obtained by Grid Search using the grid. By tool of LibSVM. Many models were generated during Grid Search. The performance of each model can be shown graphically, by which the hyperparameter interval of a good model can be clearly known, as shown in Figure 9 and Figure 10. Figure 9 shows it is the process of optimizing hyperparameters, constantly adjusting C, γ, and then finding the best hyperparameters to make the model, and we find that the best the highest accuracy rate 85.1064%, and then finding the best hyperparameters to make the model with the highest accuracy. In this paper, we aim to predict whether a student will attend a public or private university in the future. The attributes considered include the student's admission score, 3-year school performance, academic performance progress trend, and academic test scores. The accuracy was found to be higher than that of the other two experimental groups. Figure 10 shows the experiment for selecting both interest and academic ability as attributes. We conduct a finer grid search accuracy rate C, γ, and then finding the best hyperparameters to make the model, and we find that the best the highest accuracy rate 34.7518%, and between C, and γ finding the best hyperparameters rate 3.05175%. The prediction accuracy was higher than when using only interest or academic ability. This means that both interest and academic ability have a degree of influence on students' choices. The optimal hyperparameters are presented in Table 16.

Table 16. The Optimal Hyperparameters of Predictive Models

Predicted item	C	γ
University	128	7.8125e-3
Departmental field	32	3.0517578125e-5

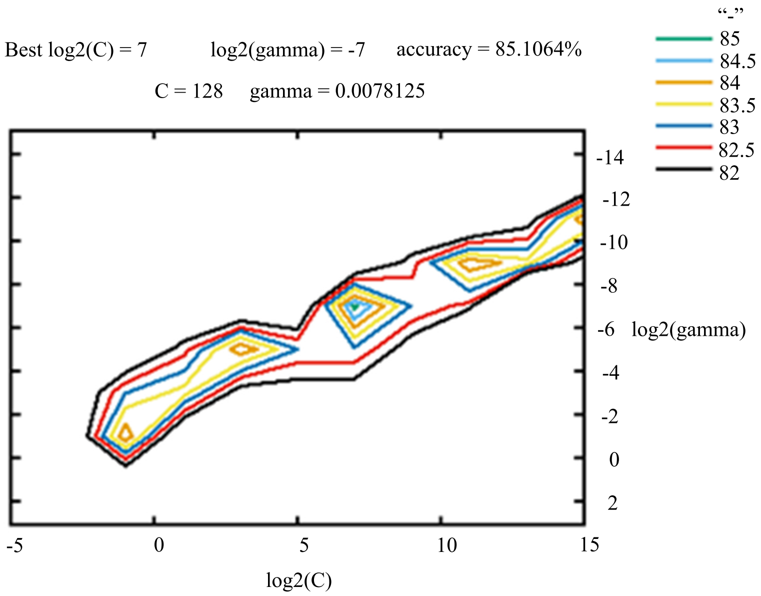


Figure 9. The performance of each model constructed by hyperparameters in predicting the establishment of a university

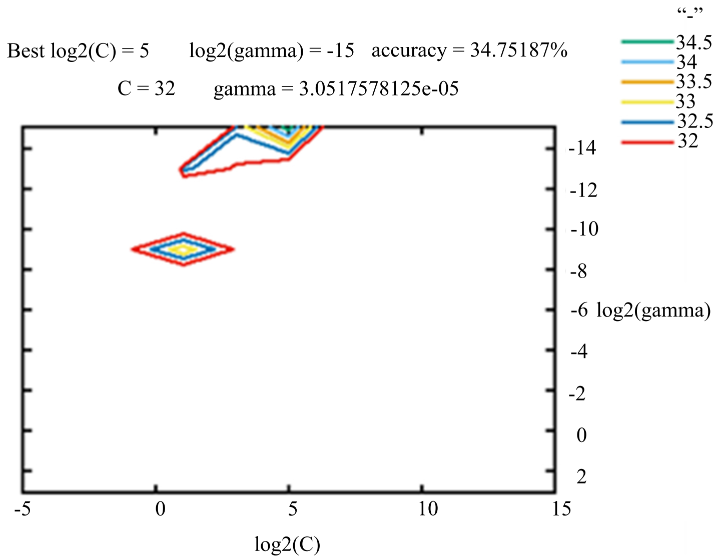


Figure 10. The performance of each model constructed by hyperparameters in predicting the field of the department

4.3.5 *Probability comparison of predicted results*

When implementing SVM, the probability value of classifying into each class could be obtained using probability estimation. Figure 11 displays the average probability values for each ranking in predicting the establishment of the university, after arranging the class results in descending order of probability value. The standard classification of education levels is used to predict the rankings of prospective students in their chosen fields of study.

To illustrate, Table 17 depicts the model's prediction that two students will attend public and private schools in the future. The results indicate that the first student has an 80% likelihood of attending a public school and a 20% likelihood of attending a private school, while the second student has a 70% likelihood of attending a private school and a 30% likelihood of attending a public school. After averaging the results, the probability for the first high-probability class is 75%, and for the second high-probability class is 25%.

Table 17. Averaging probability of each probability rank

Prediction result	Probability with rank	
	No.1 (Highest probability value class)	No.2 (second highest probability value class)
Student 1	Public: 0.8	Private: 0.2
Student 2	Private: 0.7	Public: 0.3
Average	0.75	0.25

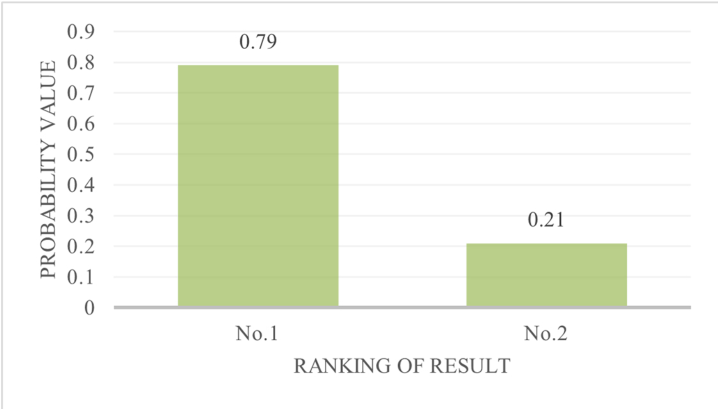


Figure 11. The averaged probability values of each ranking in predicting establishment of university

The average probability values for each ranking in predicting the departmental field are shown in Figure 12. The calculation method is the same as that in Figure 11. The probabilities of the 1st to 11Th highest probabilities in the 11 class areas are summed, and then the average is calculated. Taking rank No.1 in Figure 12 for an explanation, the probability value of 0.26 represents that the probability value of the department with the highest probability value predicted by the model is equal to 0.26 in the average case. For example, if the model predicts a student will most likely choose a department related to the education domain, the model means the student will have a 26% probability on average of choosing a department related to the education domain.

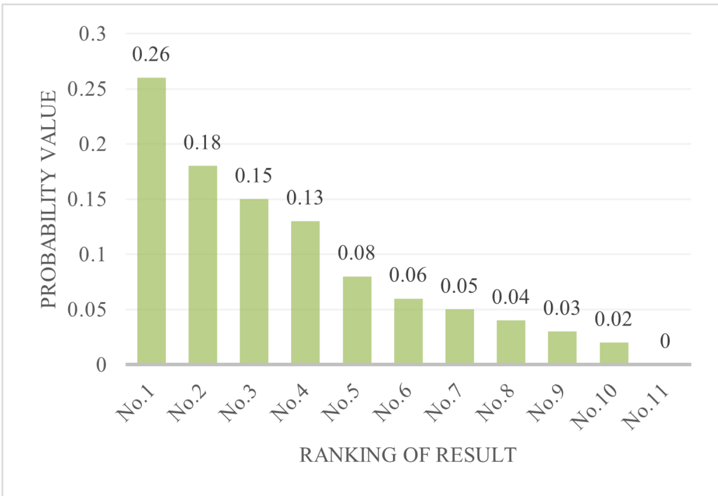


Figure 12. The averaged probability values of each ranking in predicting field of department

These two graphs indicate a significant disparity in the probability values for each ranking in predicting a university, while the disparities in predicting the departmental field are relatively small. The probability values for the second, third, and even fourth places in Figure 12 are still quite high compared to the first place. Based on this result, it can be inferred that the students' actual choices are likely to be the second, third, and fourth places. As a result, the accuracy is low when only predicting the first place.

4.3.6 *Accuracy of rankings in the field of the department*

When predicting the departmental field, the probability value of each student going to each standard classification of education levels was sorted. Figure 13 displays the accuracy of each ranking after sorting the probability values. Using the first probability ranking of 27.69% as an example, which represents all predicted students, the final selection model predicts that the student will choose the department with the highest probability of 27.69%. As shown in Figure 13, the accuracy of the first and second places exceeded 20%, while the accuracy of the third to eleventh places gradually decreased. The higher-ranking outcome makes it more likely for students to choose the departmental field.

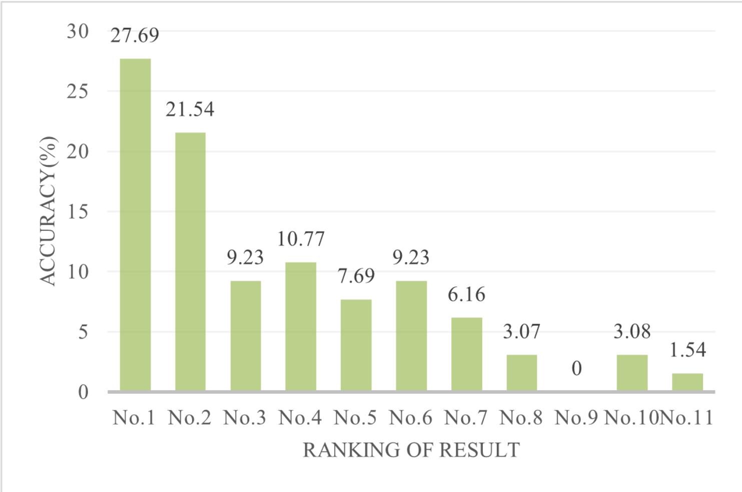


Figure 13. Accuracy of each ranking in predicting field of department

5. Conclusion

The new Curriculum Guidelines for 12-Year Basic Education encourage students to explore their interests across various courses. This research proposes a predictive model for high school students, which can be used by individual schools to develop their own prediction models in order to make early predictions about student outcomes. Considering that the academic levels of students vary in each school, and the teaching environment and curriculum also differ, each high school only utilizes the data of its own students to develop a model tailored to its specific school. The school utilizes a significant amount of annual data to assess the progress of implementation and to develop improvement strategies based on the data. This approach aims to investigate students' learning behaviors.

Implementing Bayesian Networks, Support Vector Machines (SVM), and Neural network, and then using the model with the highest accuracy could help make predictions about these choices, providing accurate predictive results. This could help schools and teachers provide students with appropriate guidance. In this study, when constructing a classification model using three data mining algorithms, including the K2 algorithm is utilized to identify the relationship between students' entrance scores, school grades, academic progress, decline trends, and academic test scores. This is done to identify the network structure suitable for each school. Based on this structure, a model is trained to predict whether a student will enroll in a public or private university in the future. Among them, the Bayesian Networks model achieved the highest accuracy of 86.25%. When predicting the department's field, The SVM model demonstrates improved accuracy by utilizing probability estimation to identify the top three fields with the highest probability, achieving an accuracy of 64.77%.

We used data from students at a high school. One of the strengths of our study is that we considered the students' interests, which were measured using 30 types of knowledge scores, and their academic abilities, which were assessed through scores from various courses and subjects during their studies. This enabled us to compare the performance of the models in different areas and to discuss which variables are more relevant in the prediction models based on these fields of knowledge. When using a Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel and Neural Network (NN) for multiclass classification problems, the prediction results can be categorized into 11 different department fields. This classification is based on students' interest in various subjects, their performance in each course, and their performance in academic tests. Despite finding the best hyperparameters through grid search and applying them to the model, the accuracy of the model remains quite low. Through probability estimation, it is possible to calculate the likelihood of each student being classified into each category. After sorting, the student's volunteering activities can still be observed and quantified.

Therefore, the study predicts that more information and resources are needed to support education policy and planning. Teachers can use data as part of practical advice and comprehensive, evidence-based guidance on key issues in classroom practice, including relationships, behaviors, and strengthening group counseling in grades 8 and beyond.

This research has some limitations. We only consider student achievement and predict academic test results. Under the higher education system in Taiwan, as in many others internationally, students are required to rank their degree options in order of preference. Based on their grades, a system can predict whether a student will enroll in a public or private university in the future. Students may end up studying their first choice if they have a very good grade, but they may also end up pursuing studies that they had ranked as a lower option. In addition, when predicting projects, it may be beneficial to categorize the classification results of public and private universities into popular public, unpopular public, popular private, and unpopular private universities. This approach would better align with the actual direction of student selection.

For a more comprehensive analysis and predictions, future research could incorporate additional data attributes such as students' study records, school club experience, leadership experience, certificates obtained, awards received, and scores of College Admission Practical Examinations, among others. Personal traits could also be considered by incorporating the results of the Interest Scale developed by the CEEC. Potentially influential factors, such as family background and rural-urban disparities, could also be included. For the existing attributes, future research could consider aggregating or averaging the scores of all subjects to determine which colleges students are eligible to apply to, based on criteria set by the colleges. If a student meets the qualifications, the college will invite them to participate in the second stage of the screening process. It is possible to remove one attribute at a time for more rigorous experiments in order to identify the best combination of attributes. With the current 12 years of basic education, the implementation of a new curriculum has also led to changes in the way students pursue higher education. Whether the new syllabus will affect students and the department are topics for further research and discussion.

Acknowledgement

This research was funded by Ministry of Science and Technology, R.O.C. program MOST 111-2221-E-027-133.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Han J, Kamber M, Pei J. *Data mining: Concepts and techniques*. 3rd ed. Burnaby, Canada: Simon Fraser University; 2012. Available from: https://9780123814791/data-mining-concepts-and-techniques.

[2] Raghavendran CV, Pavan Venkata Vamsi C, Veerraju T, Veluri RK. Predicting student admissions rate into university using machine learning models. *Machine Intelligence and Soft Computing*. 2021; 151-162. Available from: https://doi.org/10.1109/ICCMC51019.2021.9418279 [Accessed 21st January 2021].

[3] Keser SB, Aghalarova S. HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students. *Education and Information Technologies*. 2022; 27: 4521-4552. Available from: https://doi.org/10.1007/s10639-021-10780-0 [Accessed 1st May 2022].

[4] Lin L. Some influential factors and its casual relationships about undergraduates' college choices and major selections. *Journal of Educational Research and Development*. 2010; 6(3): 223-256.

[5] Alghamdi AS, Rahman A. Data mining approach to predict success of secondary school students: A saudi arabian case study. *Education Sciences*. 2023; 13(3): 293. Available from: https://doi.org/10.3390/educsci13030293 [Accessed 9th March 2023].

[6] Karthika S, Sairam N. A naïve bayes classifiers for educational qualification. *Indian Journal of Science and Technology*. 2015; 8(16): 1-5. Available from: https://doi.org/10.17485/ijst/2015/v8i16/75353 [Accessed 30th January 2015].

[7] Sharma S, Pandey SK, Garg K. A Machine learning for prediction in academics. *International Journal of Recent Technology and Engineering (IJRTE)*. 2020; 15(2): 443-461. Available from: https://doi.org/10.17485/ijst/2015/v8i16/75353 [Accessed 1st January 2020].

[8] Rawal A, Lal BG. Predictive model for admission uncertainty in high education using naïve bayes classifier. *Journal of Indian business research*. 2023; 15(2): 262-277. Available from: https://doi.org/10.1108/JIBR-08-2022-0209 [Accessed 1st January 2023].

[9] Matar NA, Matar W, AlMalahmeh T. A predictive model for students admission uncertainty using naïve bayes classifier and kernel density estimation (KDE). *International Journal of Emerging Technologies in Learning (iJET)*. 2022; 17(8): 75-96. Available from: https://doi.org/10.3991/ijet.v17i08.29827 [Accessed 26th April 2022].

[10] Jie H, Li Y, Pu X, Ye J. Pentraxin 3, a predicator for 28-day mortality in patients with septic shock. *The American Journal of the Medical Sciences*. 2017; 353(3): 242-246. Available from: https://0.1016/j.amjms.2017.01.003 [Accessed 11th January 2017].

[11] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2011; 2(3): 1-27. Available from: https://doi.org/10.1145/1961189.1961199 [Accessed 6th May 2011].

[12] Hsia TC, Shie AJ, Chen LC. Course planning of extension education to meet market demand by using data mining techniques-an example of Chinkuo technology university in Taiwan. *Expert Systems with Applications*. 2008; 34(1): 596-602. Available from: https://doi.org/10.1016/j.eswa.2006.09.025 [Accessed 1st January 2008].

[13] Injadat M, Moubayed A, Nassif AB, Shami A. Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Systems*. 2020; 200: 105992. Available from: https://doi.org/10.1016/j.knosys.2020.105992 [Accessed 20th July 2008].

[14] Alhassan AM. *Using data mining techniques to predict students' academic performance*. Jeddah, Saudi Arabia: King Abdulaziz University; 2020.

[15] Almasri A, Celebi E, Alkhawaldeh RS. EMT: Ensemble meta-based tree model for predicting student performance. *Scientific Programming*. 2019; 2019: 3610248. Available from: https://doi.org/10.1155/2019/3610248 [Accessed 24th February 2019].

[16] Miguéis VL, Freitas A, Garcia PJV, Silva A. Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*. 2018; 115: 36-51. Available from: https://doi.org/10.1016/j.dss.2018.09.001.

[17] Miguéis VL, Freitas A, Garcia PJV, Silva A. Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*. 2018; 115: 36-51. Available from: https://doi.org/10.1016/j.dss.2018.09.001.

[18] Huang S, Wei J. Student performance prediction in mathematics course based on the random forest and simulated annealing. *Scientific Programming*. 2022; 2022: 9340434. Available from: https://doi.org/10.1016/j.dss.2018.09.001 [Accessed 23th March 2022].

[19] Irfiani E, Elyana I, Indriyani F, Schaduw FE, Harmoko DD. Predicting grade promotion using decision tree and naïve bayes classification algorithms. *2018 Third International Conference on Informatics and Computing (ICIC)*. Palembang, Indonesia; 2018. Available from: https://10.1109/IAC.2018.8780431.

[20] Kondo N, Hatanaka T. An approach for estimation of students' learning states using bayesian network with LMS log data. *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)*. Japan; 2018. p.474-477. Available from: https://10.1109/IIAI-AAI.2018.00102.

[21] Ragab M, Abdel Aal AMK, Jifri AO, Omran NF. Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques. *Wireless Communications and Mobile Computing*. 2021; 2021(1): 6241676. Available from: https://doi.org/10.1155/2021/6241676 [Accessed 7th December 2018].

[22] Kusumawardani SS, Alfarozi SAI, Jifri AO, Omran NF. Transformer encoder model for sequential prediction of student performance based on their log activities. *IEEE Access*. 2023; 11: 18960-18971. Available from: https: //10.1109/ACCESS.2023.3246122.

[23] Ismail SSI, Mansour RF, Abd El-Aziz RM, Taloba AI. Efficient E-mail spam detection strategy using genetic decision tree processing with NLP features. *Computational Intelligence and Neuroscience*. 2022; 2022: 7710005. Available from: https://doi.org/10.1155/2022/7710005 [Accessed 24th March 2022].

[24] Yang Q, Li Z, Jiao H, Zhang Z, Chang W, Wei D. Bayesian network approach to customer requirements to customized product model. *Discrete Dynamics in Nature and Society*. 2019; 2019: 1026-0226. Available from: https://doi.org/ 10.1155/2019/9687236 [Accessed 15th July 2019].

[25] Athani SS, Kodli SA, Banavasi MN, Hiremath PGS. Student academic performance and social behavior predictor using data mining techniques. *2017 International Conference on Computing, Communication and Automation (ICCCA)*. India; 2017. Available from: https://doi.org/10.1155/2019/9687236.

[26] Shi S, Zhang S, Hao J, Chen K, Wang J. MOOC dropout prediction based on bayesian network. In: Xu Y, Yan H, Teng H, Cai J, Li J. (eds.) *Machine learning for cyber security*. Cham, Switzerland: Springer; 2023.

[27] Kerdvibulvech C. A methodology for hand and finger motion analysis using adaptive probabilistic models. *International Journal of Embedded Systems*. 2014; (2014): 1-9. Available from: https://doi.org/10.1186/ s13639-014-0018-7 [Accessed 3rd July 2014].

[28] Linderman S, Adams RP, Pillow JW. Bayesian latent structure discovery from multi-neuron recordings. *Advances in Neural Information Processing Systems*. 2016; 2016: 29. Available from: https://proceedings.neurips.cc/paper_ files/paper/2016/file/708f3cf8100d5e71834b1db77dfa15d6-Paper.pdf.

[29] Kerdvibulvech C, Saito H. Guitarist fingertip tracking by integrating a bayesian classifier into particle filters. *Advances in Human-Computer Interaction*. 2008; 2008: 384749. Available from: https://doi.org/10.1155/2008/ 384749.

[30] Xu J. *A continuous time bayesian network approach for intrusion detection*. Riverside, CA, USA: University of California; 2010.

[31] Kerdvibulvech C. Real-time augmented reality application using color analysis. *2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI)*, Austin, TX, USA; 2010. p.29-32. Available from: https://10.1109/ SSIAI.2010.5483927.

[32] Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. *Datascienceassn*. 2013; 2013: 1-16. Available from: https://www.csie.ntu.edu.tw/~cjlin/talks/freiburg.pdf.

[33] Cardona TA, Cudney E. Predicting student retention using support vector machines. *Procedia Manufacturing*. 2019; 39: 1827-1833. Available from: https://doi.org/10.1016/j.promfg.2020.01.256.

[34] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Procedia Manufacturing*. 1943; 5: 15-133. Available from: https://doi.org/10.1007/BF02478259.

[35] Samuel YT, Hutapea JJ, Jonathan B. Predicting the timeliness of student graduation using decision tree C4.5 algorithm in universitas advent Indonesia. *2019 12th International Conference on Information & Communication Technology and System (ICTS)*. Indonesia; 2019. p.276-280. Available from: https://10.1109/ICTS.2019.8850948.

[36] Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*. 1999; 10(3): 61-74.

[37] Lin HT, Lin CJ, Weng RC. A note on Platt's probabilistic outputs for support vector machines. *Advances in Large Margin Classifiers*. 2007; 68: 267-276. Available from: https://doi.org/10.1007/s10994-007-5018-6 [Accessed 25th June 2007].

[38] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011; 2(3): 2157-6904. Available from: https://doi.org/10.1145/1961189.1961199 [Accessed 6th May 2011].

[39] Sutoyo E, Almaarif A. Educational data mining untuk prediksi kelulusan mahasiswa menggunakan algoritme Naïve Bayes classifier. *Jurnal RESTI*. 2020; 4(1): 95-101.

[40] Sembiring MT, Tambunan RH. Analysis of graduation prediction on time based on student academic performance using the Naïve Bayes Algorithm with data mining implementation (Case study: Department of Industrial Engineering USU). *IOP Conference Series: Materials Science and Engineering*. 2021; 1122(1): 012069. Available from: https://dx.doi.org/10.1088/1757-899X/1122/1/012069.