A Decision Model to Predict the Selection of Education Streams – A Case Study for Trang Province, Thailand

Bunjira Makond

Faculty of Commerce and Management, Prince of Songkla University, Trang, Thailand

Abstract **– The research aims to develop a classification model to predict students' decisions relating to which education stream to follow and to identify the relevant factors to such decision-making. The model created used the decision tree (DT) technique with data collected from 800 samples of year-10 students in Trang province, Thailand. Of those, 428 had decided to study in general education while the other 372 were studying vocational courses. This study identified six relevant variables, which were influential in students' decision making; comprising students' academic performance (i.e., mathematical cumulative grade point average in junior high school; science cumulative grade point average in junior high school; and grade point average for all subjects in junior high school); parents' education and profession (i.e., father's level of education; mother's occupation); and students' gender. In terms of performance and employment, the DT model has high performance regarding sensitivity and g-mean. Moreover, it has advantages compared to the other models.**

Keywords **– decision tree, classification, decision-making, general education stream, vocational education stream**

DOI: 10.18421/TEM113-44
https://doi.org/10.18421/TEM113-44

Corresponding author: Bunjira Makond,
Faculty of Commerce and Management, Prince of Songkla University, Trang, Thailand.
Email: bunjira.m@psu.ac.th

Received: 18 May 2022.
Revised: 06 August 2022.
Accepted: 11 August 2022.
Published: 29 August 2022.

1. Introduction

Recently, there has been a gap between the consequences of education and the needs of the labor market in Thailand. Because of the expansion in Thailand's automotive and auto parts industries, there has been a high demand in the labor market for workers with vocational education. Meanwhile, evidence from [1] indicates that the number of students choosing vocational education is lower than those choosing general education. In 2009, approximately 38% of all upper secondary students were enrolled in the vocational stream, and despite the government's attempts in 2010 to make vocational secondary education more appealing, the ratio of students enrolled in vocational to general secondary schools has remained steady at approximately 40:60 [2]. Furthermore, as higher education grows in popularity, accessibility, and affordability in Thai public schools, many students choose to continue their education at universities. Accordingly, the number of graduated students entering the workforce has expanded considerably to the point where the labor market is struggling to meet the growing demand for new graduates, whereas Thailand actually needs vocational-level workers more than those with university-degrees. As a result, Thailand's labor market has been distorted by a mismatch between educational attainment and labor market requirements.

After finishing compulsory education, students in Thailand must select between vocational and general education streams for their continued education. It is critical to make the proper choice about which stream to pursue in upper secondary school in order to achieve success, both in the rest of people's education, as well as in their future careers and their earnings from them. Therefore, students need to be effectively counseled and guided when making decisions about which educational path to choose. Identification of factors, which influence students' making a decision to study in either vocational or general education, will help those giving advice to students deciding about their future education to

provide effective guidance to those students. Likewise, various factors can be influential in classifying students who will make a decision to study in either vocational or general education. Classification modeling assists in the identification of critical components.

In education domain, most previous studies have used traditional statistical methods to investigate the factors with the purposes for inferring relationships between variables. For instance, [3] found five major elements that influence a student's decision to seek a university education using factor analysis. In order to investigate factors that influence the decision to study abroad using data obtained in an empirical study, descriptive statistics, factor analysis and T tests were the main statistical tools used in the analysis [4]. In addition, a classification model was developed using binary logistic regression analysis in order to classify prospective students as enrolees or non-enrolees. This statistical method identified that six factors influenced the prospective students' decisions to enroll [5]. The study's goals were to look into their post-secondary education choices and assess the factors that influenced their choices for students who chose vocational education over university following high school. The significance of important influential elements was investigated using Binary Logistic Regression [6].

However, such statistical techniques necessitate certain assumptions, such as the assumption that the data being analyzed is normally distributed. They also have inherent assumptions and predefined correlations that, if violated, can produce inaccurate results. Though such assumptions can be violated if the technique used is robust, such violations should be few; otherwise, the use of statistical methods without first ensuring that the data utilized is consistent with these assumptions can lead to significant hypothesis testing mistakes [7].

By contrast, machine learning (ML) is a subcategory of computer science and artificial intelligence. ML models are non-parametric, meaning they are not based on predetermined assumptions. On the other hand, traditional statistical approaches need considerable assumptions [8]. ML learns a model from observations rather than being programmed with rules. In order to achieve the task, algorithms learn a model from a training set based on sample input-output pairings and translate an input to an output so that the task can be achieved appropriately with new, undetermined inputs.

The DT is a supervised learning approach used in machine learning to create classification systems based on a large number of factors or to create prediction algorithms for a response variable. The technique forms an upturned tree structure graph, which splits a collection of instances into branch-like segments. The method is non-parametric, which means it can handle big, complex datasets without imposing a complex parametric framework. Furthermore, as compared to typical predictive models of non-statisticians' interpretability, DT has an advantage since it offers a model in the form of a graphical structure [9].

In recent years, various DT algorithms have been widely employed in the education domain for classification problems with different purposes. In the study conducted by [10], a DT method was utilized to predict students' final grade point average based on their grades in preceding classes. The researchers used the DT algorithm to predict student achievement in vocational high school [11]. The study of [12] investigated and evaluated the method of using DT algorithms in conjunction with a student questionnaire to identify elements that influence student success or failure. The goal of [13]'s work was to propose a prediction model for students' on-time graduation using the C4.5 algorithm, which takes into account four factors: department, GPA, English score, and age. In the study of [14], the DT method was employed to classify students into distinct groups such as brilliant, average, and weak, with a high performance when compared to another technique. Hong et al. [15] created DT models to account for students' proclivity to pursue Technical and Vocational Education and Training (TVET) following high school. Respondents for this study include 428 secondary school students from Kedah, Malaysia. The Decision Tree was one of the classification techniques used to analyze the actual enrollment data of TVET training provider organizations in Punjab, Pakistan [16]. In this study, the factors that can explain the attitudes and judgments of individuals towards vocational and general education in Romania were adopted by the decision tree method [17].

The goal of this research is to create a DT model for classification of which education stream to pursue and to discover the significant classification factors that influence students' decision-making on education streams. Furthermore, the study intends to evaluate the DT model's performance against that of other supervised machine learning models based on a variety of matrices.

The rest of the article is organized as follows. Materials and methods, including data and data collection, data preprocessing, supervise machine learning methods, and model evaluation, are described in Section 2. The experiments and results are presented in Section 3. Discussion and conclusion are provided in Sections 4 and 5, respectively.

2. Materials and Methods

2.1. Data and Data Collection

Since no existing dataset that would allow the researcher to build a suitable classification model was available, primary data collection was necessary. The population adopted in this study was all 3,476 students in Trang province, Thailand who chose to attend public secondary schools in tenth grade and the 1,197 students who, at the same stage of education, had decided to study in a first-year course leading to a vocational certificate at a vocational technical college. The schools were chosen using a stratified random sampling approach, followed by systematic random selection to pick the tenth-grade pupils from each school. Meanwhile, the first-year vocational certificate students were chosen using the systematic random sampling technique from all vocational schools since the number of vocational schools in Trang is lower than the number of higher secondary schools. The final sample of 800 students from which data was collected consisted of 428 students studying in the general stream at the upper level of secondary schools and 372 respondents studying in the vocational stream.

The factors concerning which data were obtained were based on the literature study included the demographic factors, namely, gender [18], SES [19], family size [20], and parental marital status [20], [21], all of which have been suggested to have an impact on students' educational choices. In addition, data relating to grade point average, and previous academic achievement, which have also been found to affect students' decisions [18] were collected. In this study, SES was indicated by parental education, parental occupation and family income; family size was indicated by the total number of siblings, as well as the number of siblings who were actively enrolled in a Bachelor's degree program or below. Previous academic achievement was indicated by the respondents' cumulative grade point average in mathematics and science in junior high school, as well as their total grade point average in all subjects. The data was collected based on the students' responses to the items in the questionnaire relating to those indicators, which were as adopted as predictor variables. Meanwhile, the response variable was the students' decision between the general and vocational education streams.

2.2. Data Preprocessing

Data preprocessing is a crucial step in the data mining process that converts raw data into a usable format. Data cleaning, data transformation, and data reduction are the three primary processes in preprocessing [22]. The implementation of the procedures, on the other hand, is limited to dealing with data issues. The data obtained for some of the variables in this study was full and useful, whereas data for others, such as parental marital status, father's highest education level, and mother's highest education level, was in various states. Although any number of values for variables can be used, it is better to define a variable with fewer values since this simplifies the decision-making process. As a result, data transformation was required to reduce the number of values for those variables. Table 1. presents a list of the attributes, their labels, values, and counts.

2.3. Supervise Machine Learning Methods

2.3.1. Decision Tree

A well-known supervised learning approach in machine learning is the decision tree (DT), which analyzes the set of class-labeled training instances and represents the data in a tree structure graph and is commonly utilized as a technique of classification [23]. DT is typically made up of three components: a root node, internal nodes, and leaf nodes. The tree begins with a root node that includes all of the instances and has no incoming edges. All other nodes, on the other hand, have only one incoming edge. An internal node is a node with outgoing edges that divides the instances into at least two groups based on the value of the predictor variable. Similarly, each of the leaf nodes is given a class that represents the response variable's most appropriate value. The conversion of DTs to classification rules is simple.

Growing and pruning are the two phases in the induction of decision trees, in general. The training data is used to create the tree using certain criteria in order to identify a sequence of splits that divide the training data into smaller subsets with pure class labels in the growth stage. Pruning the tree is an important step in improving the model's computational efficiency and classification accuracy. Pruning decreases the size of the tree (i.e., the number of nodes) and, as a result, its complexity, as well as the model's overfitting.

This study used the C4.5 method, which uses a top-down and recursive splitting technique based on the idea of information entropy to create DT models from a collection of training data. According to Han and [24], assume that a training set D has m unique values for the class label attribute, C_i (for i = 1, 2,…,m). The number of instances in D and $C_{i,D}$ are indicated by $|D|$ and $|C_{i,D}|$, correspondingly. The expected information is expressed below:

$$Info(D) = -\sum_{i=1}^{m} \frac{|C_{i,D}|}{|D|} \log_2 \frac{|C_{i,D}|}{|D|} \quad (1)$$

Table 1. Predictor variables

Attribute	Label	Value	Count
Gender	students' gender	"f" denotes female	463
		"m" denotes male	337
Math_gpa	mathematical cumulative grade point average in junior high school	below 2.00	98
		2.00-2.99	394
		above 2.99	308
Sci_gpa	science cumulative grade point average in junior high school	below 2.00	69
		2.00-2.99	400
		above 2.99	331
GPA	grade point average for all subjects in junior high school	below 2.00	30
		2.00-2.99	372
		above 2.99	398
Status	parental marital status	Completeness (i.e. both parents are in the family)	584
		incompleteness (i.e. a family in which one or both parents are missing)	216
Edu_f	father's level of education	below bachelor's degree	667
		bachelor's degree	84
		above bachelor's degree	39
Edu_m	mother's level of education	below bachelor's degree	658
		bachelor's degree	122
		above bachelor's degree	20
F_occ	father's occupation	"occ1" denotes the owner or merchant of a business	140
		"occ2" denotes a government officer or employee of a state-owned firm	93
		"occ3" denotes agriculture	296
		"occ4" denotes employee	52
		"occ5" denotes general laborer	219
M_occ	mother's occupation	"occ1" denotes the owner or merchant of a business	181
		"occ2" denotes a government officer or employee of a state-owned firm	81
		"occ3" denotes agriculture	286
		"occ4" denotes employee	42
		"occ5" denotes general laborer	210
Income	family income	<฿15,000	324
		฿15,001 - ฿30,000	294
		฿30,001-฿45,000	90
		>฿45,000	92
Quan_sib	the quantity of siblings	1	94
		2	393
		>2	313
Edu_sib	the total number of siblings enrolled in a bachelor's degree program or below	1	269
		2	392
		>2	139
Stream	students' selection of education streams	"g" denotes general education	428
		"v" denotes vocational education	372

If the instances in D are divided in accordance with attribute A having ν distinct values as observed from the training data, the attribute A would split D into D_j partitions where $j = 1, 2, \ldots, \nu$. The expected information required to classify an instance from D according to the partitioning by A is measured by

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j) \quad (2)$$

The information gain is defined as the following:

$$\text{Gain}(A) = Info(D) - Info_A(D) \quad (3)$$

C4.5 algorithm uses gain ratio, which is an extension to information gain, as an attribute selection measure. The gain ratio of A is defined as follows:

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (4)$$

$$\text{where, } SplitInfo(A) = -\sum_{i=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right) \quad (5)$$

Since the DT is a learned approach that selects attributes as part of the learning process, the attribute with the highest gain ratio is chosen. In this study, the WEKA machine learning software [25], which has a classifier named J48 that implements the C4.5 algorithm, was used to construct the DT model.

The J48 classifier employs two pruning techniques: pre-pruning and post-pruning. When pre-pruning is used, the tree's growth is stopped when a particular condition is met. Similarly, in the post-pruning stage, the tree will be entirely created first, followed by the replacement of the last sub-trees with leaves, based on a comparison of the tree's error before and after the sub-trees were replaced. The pruning process is influenced by the confidence factor and the minimum number of objects in a single leaf. While pruning the decision tree, the confidence factor represents a threshold of allowable inherent error in data. Lowering the threshold allows for more pruning and, as a result, more general models to be generated. It is possible to specify the minimum number of objects in a single leaf to obtain simpler models with a bigger number of samples. These factors can also be used to tune decision trees to make them simpler and smaller [26].

2.3.2. Naive Bayes

Based on Bayes' theorem, Naive Bayes (NB) is a data mining approach that used as a supervised-learning classification algorithm. It uses a probabilistic learning approach that combines previous knowledge with observed data [27]. NB assumes that conditionally independent variable values provide the class of response variable. Furthermore, the class has no parents, yet the class is the single parent of each variable. The Bayes' theorem is used to determine the posterior probability, which is then used to predict the class with the highest posterior probability [28]. Bayes' theorem can be expressed as follows:

$$P(y_j | X_i) = \frac{P(X_i | y_j) P(y_j)}{P(X_i)} \quad (6)$$

where $X_i, i = 1, 2, \ldots, n$ which consists of k attributes, i.e., $X_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$; every instance is considered to be a member of only one class y_j. In this study, $y_j \in \{y_1, y_2\}$, $P(y_j)$ represents prior probability, $P(X_i)$ represents evidence, $P(X_i | y_j)$ represents conditional probability, and $P(y_j | X_i)$ represents posterior probability.

2.3.3. Artificial Neural Network

A mathematical model called an Artificial Neural Network (ANN) tries to mimic the structure and function of biological neural networks. ANN is a supervised-learning technique that has been used to solve a wide range of issues. McCulloch and Pitts proposed the artificial neuron model in the 1940s, which was later generalized in many ways. The most widely used method is shown in Figure 1.

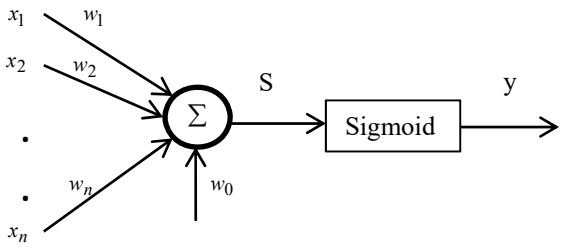

Figure 1. Artificial neuron

According to Figure 1., the neuron calculates the weighted sum of n inputs, adding a threshold value as defined in equation (7), and then applies an activation function as expressed in equation (8) to the result to compute the output y [29].

$$S = \sum_{i=1}^{n} w_i x_i + w_0 \quad (7)$$

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

$$y = f(S) \quad (9)$$

Generally, ANNs comprise three layers: input, hidden, and output. Each layer has a number of nodes. The nodes in the three layers are connected and the connections are assigned weights between nodes. In this study, a multilayer perceptron (MLP) which is a class of feed-forward ANN with a back-propagation algorithm used in training data was employed [30].

2.3.4. Logistic Regression

Logistic regression (LR) is a statistical classification approach used in data mining. The approach is employed to look into the relationship between a set of predictor factors represented by $X' = (x_1, x_2, \ldots, x_p)$ and a set of dichotomous outcome variable with values of 1 and 0 (herein, 1 denotes general education and 0 denotes vocational education). The conditional probability of the answer Y given the predictor variables is $P(Y|X) = \pi(x)$, which is calculated as follows:

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p)}}{1+e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p)}} \quad (10)$$

when $0 \le \pi(x) \le 1$.

When the value of the probability $\pi(x)$ is in the range [0, 1], odds can be used to convert the probability x to a real number. The following equation may be used to express the odds:

$$\text{odds} = \frac{\pi(x)}{1-\pi(x)} \quad (11)$$

The logit transformation, often known as the natural logarithm of the odds, is defined as follows:

$$ln\frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \quad (12)$$

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ is a set of parameters obtained using the maximum likelihood approach.

The interpretation of the model based on an odds ratio associated with the effect of a one unit change in x_j when the other predictor variables in the model are held constant is denoted as e^{β_j} [31].

2.4. Model Evaluation

The accuracy, sensitivity, specificity, and geometric mean (g-mean) were used to assess the models' performance. Accuracy is a popular metric for classification performance, and it's usually expressed as a percentage of all examples successfully categorized by the model. Sensitivity refers to a model's ability to appropriately classify students into the vocational education stream. Specificity represents a model's ability to appropriately classify students into the general education stream. Sensitivity and specificity are inversely proportional, which means that the specificity decreases with increasing sensitivity and vice versa. The g-mean, on the other hand, is a metric that assesses the balance between vocational and general education stream classification performance. Even if the vocational education stream cases are accurately classified as vocational education stream, a low g-mean indicates poor performance in the classification of general education stream cases.

Moreover, the performances of the proposed DT model were compared to the performances of LR, NB, and ANN methods, which are also supervised-learning methods, since those methods have been commonly applied in the classification of problems in different educational domains and have produced effective performances.

3. Experiments and Results

This is a description of the experiments that were conducted in this study. The goal of this study is to develop a DT model for determining which education stream to pursue and to identify the main categorization elements that influence students' education stream decisions. Furthermore, the study will compare the performance of the DT model to that of other supervised machine learning models, i.e., Naive Bayes, Artificial Neural Networks, and Logistic Regression, using a variety of matrices.

The machine learning algorithms were run on the dataset using WEKA (a software package used for knowledge analysis). WEKA is a library of algorithms that can be used to analyze data. After data preprocessing, the dataset was converted into a format usable by the WEKA. A training set and a testing set were created from the original dataset, which consisted of 12 variables and a response variable. The training set had 80% of the data and was used to train the models, while the testing set contained the remaining data and was used to assess the models' performance.

3.1. Decision Tree Model Development

Initially, the J48 classifier built a huge initial model that was over-fitting. Likewise, this experiment focused only on pruned trees as they looked for the most interpretable patterns without losing predictive performance. The pruning approach was used in this study to improve the model's computational efficiency and classification accuracy by adjusting two parameters: the confidence factor (C) and the minimum number of instances per node (minNumObj: M). WEKA's default J48 decision tree prunes based on sub-tree raising, with a confidence factor of 0.25 and a minimum number of objects of 2. In order to explore the optimal DT model, diverse values of C, which was defined in a range of 0.05 to 0.3 with each additional step being 0.05, and M,

which was set from 2 to 16 by increments of 2. The most interpretative model without loss of predictive performance is developed when the values of C and M are 0.2 and 8, respectively. Accordingly, the J48 algorithm based on information gain constructed the tree, which is composed of six variables and is named DT_J48, as presented in Figure 2.

Moreover, one of the benefits of the DT model is its ability to be shown as a graph-like tree structure; the interpretation of the DT_J48 model can also be presented in the form of classifying rules as follows:

1. If Edu_f = "above bachelor's degree" then student would decide to study in general stream
2. If Edu_f = "bachelor's degree" then student would decide to study in general stream
3. If Edu_f = "below bachelor's degree" and Math_gpa = "below 2.00" and gender = "m" then student would decide to study in vocational stream
4. If Edu_f = "below bachelor's degree" and Math_gpa = "below 2.00" and gender = "f" and Sci_gpa = "below 2.00" then student would decide to study in general stream
5. If Edu_f = "below bachelor's degree" and Math_gpa = "below 2.00" and gender = "f" and Sci_gpa = "2.00-2.99" then student would decide to study in vocational stream
6. If Edu_f = "below bachelor's degree" and Math_gpa = "below 2.00" and gender = "f" and Sci_gpa = "above 2.99" then student would decide to study in general stream
7. If Edu_f = "below bachelor's degree" and Math_gpa = "2.00-2.99" then student would decide to study in vocational stream
8. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "2.00-2.99" then student would decide to study in vocational stream
9. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "below 2.00" then student would decide to study in general stream
10. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "above 2.99" and Sci_gpa = "below 2.00" then student would decide to study in general stream
11. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "above 2.99" and Sci_gpa = "2.00-2.99" and Occ_m = "occ3" then student would decide to study in general stream
12. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "above 2.99" and Sci_gpa = "2.00-2.99" and Occ_m = "occ4" then student would decide to study in general stream
13. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "above 2.99" and Sci_gpa = "2.00-2.99" and Occ_m = "occ1" then student would decide to study in vocational stream
14. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "above 2.99" and Sci_gpa = "2.00-2.99" and Occ_m = "occ5" then student would decide to study in vocational stream
15. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "above 2.99" and Sci_gpa = "2-2.99" and Occ_m = "occ2" then student would decide to study in general stream
16. If Edu_f = "below bachelor's degree" and Math_gpa = "above 2.99" and GPA = "above 2.99" and Sci_gpa = "above 2.99" then student would decide to study in general stream

3.2. Model Comparisons

The dataset containing the six variables utilized in DT_J48 was separated into training and testing sets in order to evaluate the DT model. Later, the default LR and NB algorithms in WEKA were implemented on the training set to construct LR_J48 and NB_J48 models, respectively. Likewise, the ANN algorithm in WEKA is performed to construct the ANN_J48 model by setting the hidden layers, momentum rate, and learning rate of the ANN's parameters. In this study, the values of the hidden layers, momentum rate, and learning rate are o, 0.7, and 0.3, respectively. The experiment of each model was conducted 100 times. The performances of all models in terms of the mean values of accuracy, sensitivity, specificity, and g-mean are presented in Table 2.

Statistical analysis is utilized to find the differences in predictive performance among the models. The differences in performance were observed using the one-way ANOVA test, in which the model was treated as a factor. Multiple comparison tests are also performed to determine whether models are distinct. Because the equality of variance of data is not assumed, the Tukey's Honestly Significant Difference (HSD) test is employed to identify the different models. The significance level is set at 0.05 for the entire difference test. The ANOVA results in Table 3. show the significant differences among the models in terms of sensitivity, specificity, and g-mean since the returned p-value (0.000) is lower than the defined p-value (0.05), while the models are insignificantly different regarding accuracy.

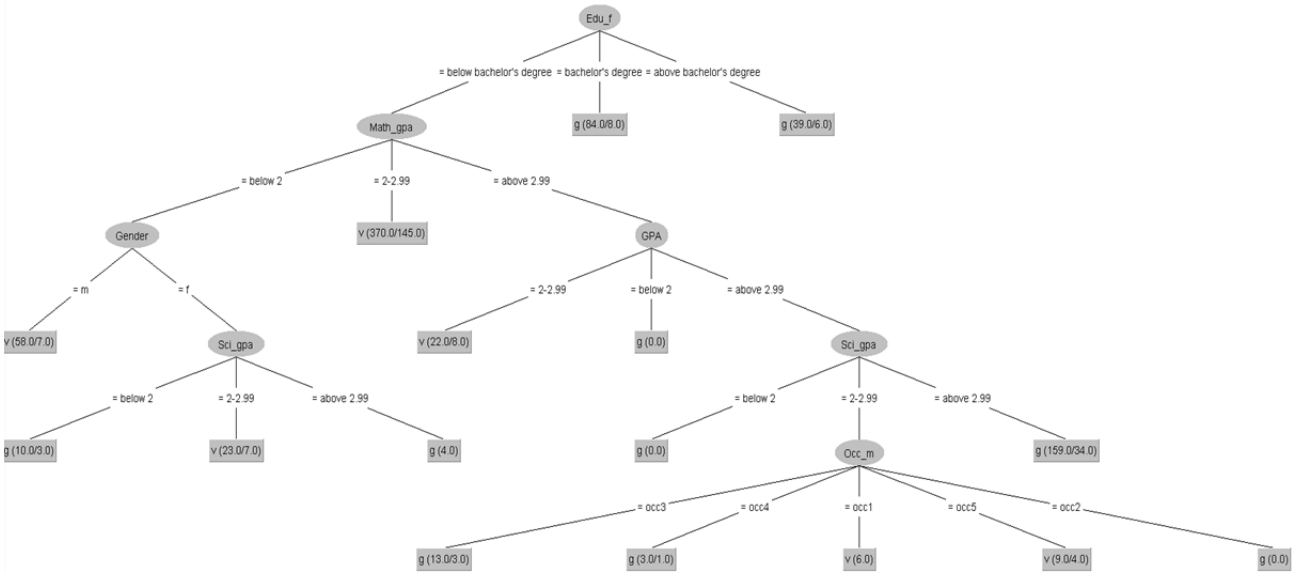

Figure 2. Decision tree model (DT_J48)

Table 2. The performance comparison of models' results

Model	accuracy	sensitivity	specificity	g-mean
ANN_J48	68.011%	0.794	0.581	0.644
DT_J48	67.639%	0.787	0.580	0.673
LR_J48	67.553%	0.734	0.624	0.675
NB_J48	66.840%	0.718	0.625	0.669

Table 3. One-way ANOVA for all metrics

Metric	Sources of variance	Sum of Squares	df	Mean Square	F	p-value
accuracy	Between Groups	71.944	3	23.981	1.966	.118
	Within Groups	4829.777	396	12.196		
	Total	4901.721	399			
sensitivity	Between Groups	.427	3	.142	9.784	.000
	Within Groups	5.763	396	.015		
	Total	6.190	399			
specificity	Between Groups	.193	3	.064	9.917	.000
	Within Groups	2.574	396	.006		
	Total	2.767	399			
g-mean	Between Groups	.065	3	.022	3.522	.015
	Within Groups	2.419	396	.006		
	Total	2.484	399			

Table 4. Tukey's HSD test for sensitivity

Model	Different group	
	1	2
NB_J48	0.718	
LR_J48	0.734	
DT_J48		0.787
ANN_J48		0.794

Table 5. Tukey's HSD test for specificity

Model	Different group	
	1	2
DT_J48	0.581	
ANN_J48	0.580	
LR_J48		0.624
NB_J48		0.625

Table 6. Tukey's HSD test for g-mean

Model	Different group	
	1	2
ANN_J48	0.644	
NB_J48	0.669	0.669
DT_J48		0.673
LR_J48		0.675

The results of Tukey's HSD test in Tables 4.-6. show the different models in different groups, while the indifferent models are listed in the same group. The results of the Tukey's HSD test for sensitivity in Table 4. show that the performance of DT_J48 and ANN_J48 models is indifferent; meanwhile, their performance is higher than that of NB_J48 and LR_J48 models in terms of sensitivity. From the

results of the Tukey's HSD test for specificity presented in Table 5., it is evident that the performance of DT_J48 and ANN_J48 models is indifferent; however, they have lower performance than NB_J48 and LR_J48 models regarding specificity. The Tukey's HSD test for g-mean in Table 6. presents the insignificant differences of the NB_J48, DT_J48, and LR_J48 models in terms of g-mean; however, DT_J48 and LR_J48 models are obviously better than ANN_J48.

4. Discussion

The DT_J48 model, based on the findings, depicts the relationships among predictors in the form of a tree structure graph to elucidate the elements that impact decision-making. In the DT_J48 model, the significant six predictor variables were used to classify the response variable, namely, students' selection of education streams. Those were grouped into students' academic achievement (i.e., mathematical cumulative grade point average in junior high school; science cumulative grade point average in junior high school; and grade point average for all subjects in junior high school); parents' education and profession (i.e., father's level of education; mother's occupation); and students' gender.

Previous academic achievement, for which the proxies are cumulative junior high school grade point averages in mathematics, science, and all-courses, has been previously proven to be significant factors in students' selection of education streams. This study backs up the findings of [32], who discovered that students with low academic performance were more likely to choose vocational education. Moreover, parents' education and profession were also shown to be important determinants in students' decisions about which educational stream to pursue in this study, which matches earlier findings in the literature. Hahs-Vaughn [33] discovered that parental education influenced their children's academic careers. Furthermore, the findings support those of [34], who found that parents' education and profession have a substantial impact on secondary school kids' career choices, implying a choice between general and vocational schooling. The model also revealed that gender has a significant effect on students' educational stream choices, which was consistent with the findings of [32], who discovered that males were more likely to enroll in vocational programs than females.

In terms of performance and employment, DT_J48 has high performance regarding sensitivity and g-mean, meaning that the model can correctly classify students into the vocational education stream. Although the proposed DT was not superior to the other machine learning methods from all metrics, it has the following advantages when compared to the other models: 1) the ability to depict the connections between variables in a graph with a tree structure, facilitating users' understanding and interpretation of the model [35]; 2) the ability to handle different types of predictor variables [35]; 3) the ability to analyze data without testing parametric assumptions [36]; 4) the ability to handle datasets when the training or testing data has missing values [37]; 5) the ability to reduce data preparation effort [38]. In contrast, LR models entail basic assumptions, including error independence, linearity of logit values for continuous variables, lack of multicollinearity, and lack of very influential outliers. Further, the NB technique entails the assumption of independence among the predictor variables, although in practice, NBs' attribute independence assumption is often violated. Finally, ANNs have many advantages; however, ANN is commonly applied to derive results from data without any evidence as to how the results are obtained, and an ANN is, therefore, often described as a black box [39].

5. Conclusion

The purpose of this research was to create a model capable of classifying students' decisions relating to education streams. The model created was derived using a J48 algorithm based on information gathered from 800 pupils in province of Trang. There were 12 predictor factors and a response variable in the data set. During the creation of the model, a post-pruning procedure was utilized to simplify the tree. To clarify the factors that influence the decision-making process, the DT_J48 model shows the interactions among predictors in the shape of a tree-like graph. Eventually, students' academic performance (i.e., mathematical cumulative grade point average in junior high school, science cumulative grade point average in junior high school, and grade point average for all subjects in junior high school), parents' education and occupation (i.e., father's level of education, mother's occupation), and students' gender were all taken into classifying the decisions made relating to education stream. The findings of this research will be beneficial in giving effective counsel to students, as well as educationalists and those in government agencies creating suitable policies. In terms of performances, DT is superior to LR, NB, and ANN. Further, DT is simply, transparency, and friendly user model.

The Bayesian network (BN) model which is a powerful tool to represent stochastic events in a simple and graphically readable representation. Furthermore, the BN model is capable of visualizing

and interpreting variable interactions, as well as calculating complex multivariable probability distributions of heterogeneous variables as interpretable local probabilities. Furthermore, the BN is used in both linear and non-linear relationships, as well as interactions. In the future study, it is interesting to purpose the BN model to predict the students' decision-making on education streams, and moreover, the performance of the proposed BN model is compared to the other machine learning methods.

Acknowledgements

This study received full financial support from the Thailand Research Fund (TRF), Grant number MRG6080165.

References

[1]. Office of the Education Council. (2017). The national scheme of education B.E. 2560-2579 (2017-2036). Retrieved from: http://www.onec.go.th/us.php/home/category/CAT0000196 [accessed: 25 April 2022].

[2]. Faviere, F. (2015). *Marketization of Education, Reproduction of Social Inequality And Violence: The Case of Vocational Students in Bangkok* (Doctoral dissertation, Facultad Latinoamericana de Ciencias Sociales (Argentina).

[3]. Sojkin, B., Bartkowiak, P., & Skuza, A. (2012). Determinants of higher education choices and student satisfaction: the case of Poland. *Higher education*, *63*(5), 565-581.

[4]. Mercy, M. (2009). Comparative analysis of factors influencing the decision to study abroad. *African Journal of Business Management*, *3*(8), 358-365.

[5]. Ozturk, O. (2019). A logistic regression analysis of factors affecting enrollment decisions of prospective students of distance education programs in Anadolu University. *Turkish Online Journal of Distance Education*, *20*(1), 145-160.

[6]. Khowchernklang, J., Madilokkovit, C., & Siribanpitak, P. (2021). Factors Influencing High School Students' Decision in Eastern Economic Corridor area (EEC) to Choose Vocational School over University: Binary Logistic Regression Analysis. *Asia Social Issues*, *14*(6), 250133-12.

[7]. Hoekstra, R., Kiers, H. A., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)?. *Frontiers in psychology*, *3*, 137.

[8]. Wahab, L., & Jiang, H. (2019). A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. *PLoS one*, *14*(4), e0214966.

[9]. Miller, B., Fridline, M., Liu, P. Y., & Marino, D. (2014). Use of CHAID decision trees to formulate pathways for the early detection of metabolic syndrome in young adults. *Computational and mathematical methods in medicine*, *2014*.

[10]. Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting students final GPA using decision trees: a case study. *International journal of information and education technology*, *6*(7), 528.

[11]. Putri, G. (2020). Implementation of the C4. 5 Algorithm to Predict Student Achievement at SMK Negeri 6 Surakarta. *Indonesian Journal of Informatics Education*, *4*(2).

[12]. Hamoud, A., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*, 26-31.

[13]. Yuliansyah, H., Imaniati, R. A. P., Wirasto, A., & Wibowo, M. (2021). Predicting Students Graduate on Time Using C4. 5 Algorithm. *Journal of Information Systems Engineering and Business Intelligence*, *7*(1), 67-73.

[14]. Vasani, V. P., & Gawali, R. D. (2014). Classification and performance evaluation using data mining algorithms. *International Journal of Innovative Research in Science, Engineering and Technology*, *3*(3), 10453-10458.

[15]. Hong, C. M., Ch'ng, C. K., & Roslan, T. N. (2021). Application of decision tree in classifying secondary school students' tendencies to choose TVET in Malaysia. *Turkish Journal of Computer and Mathematics Education*, *12*(3), 3002-3012.

[16]. Hassan, R. H., & Awan, S. M. (2019). Identification of trainees enrollment behavior and course selection variables in technical and vocational education training (TVET) program using education data mining. *International Journal of Modern Education and Computer Science*, *11*(10), 14.

[17]. Matei, M. M. M., Mocanu, C., & Zamfir, A. M. (2018). Educational paths in Romania: choosing general or vocational education. *HOLISTICA–Journal of Business and Public Administration*, *9*(2), 127-136.

[18]. Igbinedion, V. I. (2011). Perception of factors that influence students' vocational choice of secretarial studies in tertiary institutions in Edo State of Nigeria. *European Journal of Educational Studies*, *3*(2), 325-337.

[19]. Misran, N., Sahuri, S. N. S., Arsad, N., Hussain, H., Zaki, W. M. D. W., & Aziz, N. A. (2012). The influence of socio-economic status among matriculation students in selecting university and undergraduate program. *Procedia-Social and Behavioral Sciences*, *56*, 134-140.

[20]. Hansen, T. D., & McIntire, W. G. (1989). Family Structure Variables as Predictors of Educational and Vocational Aspirations of High School Seniors. *Research in Rural Education*, *6*(2), 39-49.

[21]. Gunderson, M. M. (2004). *A study of the influence vocational education has on students' ultimate academic success* (Doctoral dissertation, University of Central Florida).

[22]. Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised leaning. *International journal of computer science*, *1*(2), 111-117.

[23]. Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1), 1-21.

[24]. Han, J., & Kamber, M. (2006). Data mining: concepts and techniques, 2nd. *University of Illinois at Urbana Champaign: Morgan Kaufmann*.

[25]. Eibe, F., Hall, M. A., & Witten, I. H. (2016). The WEKA workbench. Online appendix for data mining: practical machine learning tools and techniques. In *Morgan Kaufmann*. Morgan Kaufmann Publishers.

[26]. Stiglic, G., Kocbek, S., Pernek, I., & Kokol, P. (2012). Comprehensive decision tree models in bioinformatics. *PloS one*, *7*(3), e33812.

[27]. Swetapadma, A., & Yadav, A. (2016). Protection of parallel transmission lines including inter-circuit faults using Naïve Bayes classifier. *Alexandria Engineering Journal*, *55*(2), 1411-1419.

[28]. Othman, M. F. B., & Yau, T. M. S. (2007). Comparison of different classification techniques using WEKA for breast cancer. In *3rd Kuala Lumpur international conference on biomedical engineering 2006* (pp. 520-523). Springer, Berlin, Heidelberg.

[29]. Morariu, D., Crețulescu, R., & Breazu, M. (2017). The WEKA multilayer perceptron classifier. *International Journal of Advanced Statistics and IT&C for Economics and Life Sciences*, *7*(1).

[30]. Lorena, A. C., Jacintho, L. F., Siqueira, M. F., De Giovanni, R., Lohmann, L. G., De Carvalho, A. C., & Yamamoto, M. (2011). Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*, *38*(5), 5268-5275.

[31]. Rodríguez, G. (2013). *Lecture notes on generalized linear models*. Retrieved from: https://www.math.ntnu.no/emner/TMA4315/2013h/lecture-notes.pdf [accessed: 28 March 2022].

[32]. Agodini, R., Uhl, S., & Novak, T. (2004). Factors That Influence Participation in Secondary Vocational Education. MPR Reference No. 8879-400. *Mathematica Policy Research, Inc.* Retrieved from: https://eric.ed.gov/?id=ED518742 [accessed: 20 April 2022].

[33]. Hahs-Vaughn, D. (2004). The impact of parents' education level on college students: An analysis using the beginning postsecondary students longitudinal study 1990-92/94. *Journal of College Student Development*, *45*(5), 483-500.

[34]. Udoh, N. A., & Sanni, K. B. (2012). Parental background variables and the career choice of secondary school students in Uyo local government area, Nigeria. *Mediterranean journal of social sciences*, *3*(1), 497-497.

[35]. Lytvynenko, T. I. (2016). Problem of data analysis and forecasting using decision trees method. *Problemy prohramuvannya, (vyp.)*), 220-226.

[36]. Zohair, A., & Mahmoud, L. (2019). Prediction of Student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-18.

[37]. Patidar, P., & Tiwari, A. (2013). Handling Missing Value in Decision Tree Algorithm. *International Journal of Computer Applications*, *70*(13), 31-36.

[38]. Brunello, A., Marzano, E., Montanari, A., & Sciavicco, G. (2019). J48SS: A novel decision tree approach for the handling of sequential and time series data. *Computers*, *8*(1), 21.

[39]. Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, *49*(11), 1225-1231.