Analysis of Key Features Affecting The Effectiveness of English Language Learning among Undergraduate Students Utilizing The Data Mining Techniques

Chidchanok Kaewpanitch* and Sutep Tongngam**

Received: September 7, 2023 Revised: November 8, 2023 Accepted: November 15, 2023

Abstract

In this paper there are two folds of study. One was to compare two data mining techniques, which were Random Forest and Decision Tree, while the other was to pick the better one to analyse the key features affecting the effectiveness of English language learning at the undergraduate level.

With the same dataset, the Random Forest had shown that it was more accurate than the other. When applying the Random Forest to the collected data, the results revealed that the factors impacted on the effectiveness of English language learning at undergraduate level of the selected university students were Enjoying learning English, Asking teachers immediately when they didn't understand the content of the lesson, Interest and attention to English, Students that studied in groups to review English lessons with friends when the exam was approaching, Students that invested extra time into research and studying English beyond the classroom, respectively. They could be the keys to success in English learning for the other undergraduate students.

Keywords: Data Mining, Decision Tree, Random Forest, Academic Effectiveness, Learning English

* Graduate School of Applied Statistics, National Institute of Development Administration
148 Serithai Road, Khlong-Chan, Bangkapi, Bangkok 10240, THAILAND.
E-mail: chidchanok.kae@stu.nida.ac.th

** Assistant Professor, Graduate School of Applied Statistics,
National Institute of Development Administration
148 Serithai Road, Khlong-Chan, Bangkapi, Bangkok 10240, THAILAND.
E-mail: Sutep.t@nida.ac.th

การวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิผลการเรียนวิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล

ชิดชนก แก้วพานิชย์* และ สุเทพ ทองงาม**

รับวันที่ 7 กันยายน 2566 ส่งแก้ไขวันที่ 8 พฤศจิกายน 2566 ตอบรับตีพิมพ์วันที่ 15 พฤศจิกายน 2566

บทคัดย่อ

ในบทความวิจัยนี้แสดงผลการศึกษาเป็นสองขั้นตอน โดยเริ่มจากการเปรียบเทียบประสิทธิภาพของการทำเหมืองข้อมูลด้วยเทคนิค ตัวแบบป่าไม้ตัดสินใจ (Random Forest) และตัวแบบต้นไม้ตัดสินใจ (Decision Tree) จากนั้น นำเทคนิคที่ดีกว่ามาใช้วิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิผลการเรียนวิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี

ผลการเปรียบเทียบโดยใช้ข้อมูลชุดเดียวกัน พบว่า เทคนิค ตัวแบบป่าไม้ตัดสินใจ ให้ค่าความแม่นยำสูงกว่า และเมื่อใช้เทคนิคดังกล่าวทำการประมวลผลข้อมูล ปรากฏว่า ความสนุกกับการเรียนภาษาอังกฤษมีความสัมพันธ์ การถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน ความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน การจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ และการศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน มีความสัมพันธ์กับประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักศึกษาจากมากไปหาน้อย ตามลำดับ ซึ่งคุณลักษณะเหล่านี้จะเป็นกุญแจสำคัญสู่ประสิทธิผลในการเรียนการสอนภาษาอังกฤษในระดับปริญญาตรี

คำสำคัญ: การทำเหมืองข้อมูล ตัวแบบต้นไม้ตัดสินใจ ตัวแบบป่าไม้ตัดสินใจ ประสิทธิผลทางการศึกษา การเรียนภาษาอังกฤษ

* คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์
เลขที่ 148 ถนนเสรีไทย แขวงคลองจั่น เขตบางกะปิ กรุงเทพฯ 10240
อีเมล: chidchanok.kae@stu.nida.ac.th

** ผู้ช่วยศาสตราจารย์ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์
เลขที่ 148 ถนนเสรีไทย แขวงคลองจั่น เขตบางกะปิ กรุงเทพฯ 10240
อีเมล: Sutep.t@nida.ac.th

Introduction and Objectives

"English" is a popular foreign language and a common language that conveys meaning to almost four corners of the world. Over 2,000 million people or one third of the global population mainly use English in their communication (Thanakorn and Sangiam, 2016). Therefore, it is crucial to promote the Thais especially those in higher education who will contribute to the country's future strength. Their English skills should be developed, and their English language learning should be upgraded for higher efficiency. Based on the ranking of English in 100 countries worldwide where the English language skills are not used as a main language by the EF (Education First) Institute, the criteria to measure the level of English is divided into five levels for the use of English ranked from Very high to High, Moderate, Low, and Very low, respectively. It was revealed that the English of the Thais belonged to the last group of Very Low with the score of only 47.61, which decreased for three consecutive years since 2017. The ranking of English of the Thais in 2020 was found that the use of English of the Thais continuously decreased or down to the rank of 89 (Education First, 2020). Therefore, the promotion of the skills in English in higher education must be planned from the compulsory basic courses starting with the understanding of the behavior and motivation which affects English language learning and is probably a factor that promotes the English skills of the Thais for the better future. The research (Supika, 2018) was conducted to find the correlation between the factors of attitude, motivation, and behavior of English language learning of the students at Phuket Rajabhat University. The study results revealed that the factors of attitude, internal motivation, and learning behavior linked with success in English language learning. Good attitude had the most influence on the effectiveness of English language learning. Motivation also had a major role in successful learning of foreign language or the second language although learners' motivation could change all the time, depending on the context of the language learning (Mitra, 2016). At the same time, Data mining techniques started to attract more interest in finding the correlation within the dataset and apply them in many dimensions including education. For example, the use of the Decision Tree technique to forecast the study results of engineering students in India to acquire the information of the students who tended to fail in the study and plan necessary measures beforehand (Kabra & Bichkar, 2011). Data mining is therefore an instrument that will increase the opportunity for in-depth analysis to know the major factors to plan, accommodate,

and create the paradigm of learning between various correlations based on targets. Thus, the researchers came up with the concept to study the feasibility of utilizing the information from behavior, attitude, and motivation together with the process of data mining to classify and forecast the types of features or behaviors that would affect the effectiveness of English language learning.

This study used the case study of the students of the Faculty of Geoinformatics, Burapha University. The aim was to achieve the effectiveness of the basic English language learning of the students who did not take English as a major, as well as prepare and upgrade the knowledge level of English language both in daily life and for efficient self-development in various dimensions.

Objectives

The objective of this research is to study and analyze the key features affecting the effectiveness of English language learning among undergraduate students utilizing the data mining techniques between the Decision Tree technique and the Random Forest technique.

Research Questions

Which data mining techniques between the Decision Tree technique and the Random Forest technique is appropriate to be used to analyze the key features affecting the effectiveness of English language learning among undergraduate students?

What type of features that enable undergraduate students to have good and effective basic English language learning?

Research Scope

Scope of Content

This research studied the key features affecting the effectiveness of English language learning among undergraduate students at the Burapha University covering four dimensions namely 1) Grades of basic English class and the number of class missing in the class 2) General information of the students 3) Information of behavior affecting English language learning and 4) Attitude and motivation in English language learning.

Scope of Population

This research collected data from the sample of 299 undergraduate students, Faculty of Geoinformatics, Burapha University during the academic years 2018-2020 and with the grades of the University's basic English.

Scope of Methods

This research used the data processing and comparison of the two techniques of Data Mining namely Decision Tree and Random Forest to analyze the key features affecting the effectiveness of English language learning of undergraduate students and measure the effectiveness of the model with the Cross Validation and Confusion Matrix which collected the data with close-ended questionnaire.

Scope of Time

This research spent six months for the study, data collection, and data processing between January to June 2021.

Definition of Terms

Effectiveness in English language learning means students with very good results or grade A of the basic English language learning at Burapha University.

Education First (EF) institute means leading foreign language institute. It is a large education organization that ranks the countries with the ability to use English recognized worldwide.

Application of the data mining technique means utilization of the process of the data mining technique by adapting for appropriateness of the undergraduate students with the results of the basic English language learning at Burapha University.

Key features mean the data of the specific features or behavior relevant to the dataset to be used to construct the algorithm with the data mining technique.

Literature Review

Attitude and Motivation Affecting English Language Learaning

Attitude is crucial in English language learning. Visser (2008) stated that the success in language learning did not depend on intelligence or expertise alone but good attitude in learning as well. Similarly, to the research of Gardner (1975) who stated that attitude constituted a major factor in successful foreign language learning similarly to learning of other courses. It must start with good attitude towards the goal to lead to effective learning process and to learners' good behavior. Gardner & Lambert (1972) suggested that the attitude in foreign language learning was divided into two types namely 1) Integrated attitude is the feeling of admiration for a language and the intention to create familiarity as part of the society that used the language and 2) Functional attitude was the intention to use language for ones' utility in developing various dimensions. Apart from attitude, motivation was also crucial to the success in English language learning. Spraat et al. (2005) explained that motivation significantly influenced success or failure in language learning. This was in accordance with Broussard & Garrison (2004) who stated that the learners with high level of motivation in learning English would have higher learning effectiveness. The research of Napat Woothiwongsa (2014) recommended three levels of strategy to motivate English language learning namely 1) Language level such as attitude towards the culture of language owners and own success 2) Learner level such as the desire to succeed and self-confidence and 3) Level of learning situation such as motivation of interesting content of the course, lecturer, and learners. Therefore, the motivation of all forms derives from good attitude and has positive impact on English language learning (Supika, 2018; Deeden, 2013).

Behavior Affecting English Language Learning

Learning behavior means technique or learning method used by learners to succeed in foreign language learning (Ellis, 1997). This was in line with Pornpot Petchtaveeporndej (2004) who gave the meaning of learning behavior as activities, response, or learning methods of learners to develop knowledge, skills, and attitude to achieve the learning goal by expressing satisfaction and efforts for better development. Once a habit, it will lead to learning success. If learners had behavior or learning technique which was appropriate to themselves, they would increase their learning proficiency. The behavior

or technique would develop and increase the skills in language learning (Oxford, 1990). There were numerous research works in foreign countries on attitude, motivation, and behavior of English language learning. For example, Chen & Huang (2003) found that highly skilled students in language had better and more frequent behavior and learning strategies than the students with lower language skill. This was consistent with the research of Wu (2008) who found that highly skilled students in language in Taiwan had learning strategies such as planning, communication, and more cultural learning of language. Moreover, Broussard & Garrison (2004) studied the relations between motivation and effectiveness in English language learning of primary school pupils in the first and third years. It was found that motivation in learning had positive impact on the effectiveness of English language learning. It could be said that attitude, motivation, and good behavior had positive impact on the success of English language learning.

Concept of Data Mining

Data mining is a process that uses calculation with big data by bringing out complicated and important data. The objective is to find the model of data, survey big dataset, construct the model that explains key features of data, and forecast the trend of the data (Springer Link, 2013). One of the popular uses of data mining is Classification which is to construct the model that classifies data from the learning dataset already classified in order to classify new data the type of which is not yet known (Surapong, 2018). For example, the classification of the opportunity that a client will cancel services, etc. The Decision tree technique is a type of data mining based on the Classification of data with the model resembling a tree consisting of nodes and links. Each node of the Decision Tree consists of condition using any variable of the data. The decision will start from the Root node of the tree up to Child node and finally to Leaf node , which is the node that will identify the type of the data (Surapong, 2018).

The Random Forest technique is one the techniques that use the Machine learning with the features to classify the data of the tree structure to construct a model and forecast the objective. The Random Forest technique is widely recognized by biomedical and pharmaceutical research due to its highly effective forecast and accuracy. The algorithm of Random Forests is the model that forecasts the trees, each of which will depend on the vector that is randomly picked, independently from each other, with

the same dispersion of the whole tree. Napaporn Sirikulviriya (2011) cited in Wanvisa (2017) who stated that Random Forest was a technique developed and enhanced from the technique of Decision Tree with the use of only one tree to classify the types of desired data to ensure that the model had the highest efficiency in forecast accuracy.

Cross Validation is the method to divide the dataset to construct and test the model by dividing it into the K Disjoint Subsets and each Subsets did not have the same data. Then, it would start with the data of k-1 subsets in developing the model and use the remaining part of 1 subset as the data for the test. Repeat the activity with the number of k times. Therefore, in developing the model each time, the k-1 subset and the remaining 1 subset would be used. The same data would not be used to assess the effectiveness of the model. The mean of the assessment of the number of k times would be used. This method was called k-fold Cross Validation. It could be seen that each data would be used as the data for the test package only one time and would jointly construct the model of k-1 time. In case there in not much data, the k = n could be determined by while n was the total data. It means that each time, the data n-1 would be used to construct the model and the remaining data of 1 would develop the model (Surasit, 2015).

The measurement of the efficiency of accuracy in classifying the types consistent with the actual data was conducted through the calculation of the ratio of the number of data that the algorithm made completely correct forecast compared with the entire number of the data. Precision in correct forecast of the classification of the types of data was calculated only the ratio of the data that the algorithm made forecast of what happened with the data that the algorithm forecast as actual. The Recall means the ability of the variable in classifying the types of data and the F-Measure measures the efficiency by combining the calculation of Precision and Recall and the combined average (Surapong, 2018).

Research Conceptual Framework

Based on the relevant concepts and theories, the research framework could be determined to study and analyze the key features affecting the effectiveness of English language learning among undergraduate students utilizing the data mining technique as in Figure 1.

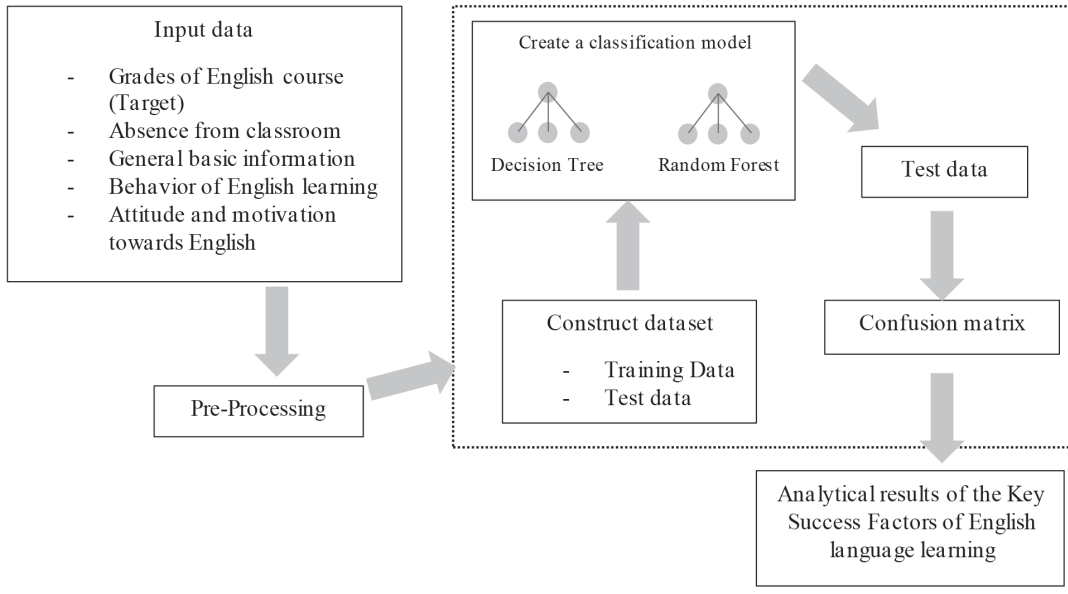


Figure 1: Research Conceptual Framework

Population and Sample

This research determined the population of 462 students of Faculty of Geoinformatics, Burapha University during the academic years 2018-2020 with the results of basic English of the University (Office of the Registrar, Burapha University as of 4 May 2021). The smallest size of the sample was 210 students according to the concept of Krejcie and Morgan (1970) to reduce the discrepancy from collection of data and for more accurate results. The researchers therefore increased the size of the sample to 299 students.

Methodology

Based on the study and the analysis of the key features affecting the effectiveness of English language learning among undergraduate students by using the data mining technique, the researchers applied the process of data mining development (CRISP-DM) to serve as the research guideline. The operation was divided into five processes as follows:

Study and Data Collection

This research was collected through the questionnaire consisting of data on English language learning of the students, basic data of the students, behavior that affected English language learning, attitude and motivation. The data collection aimed to seek the key features which could identity the types of features that were the reasons that the undergraduate students had effective English language learning by using the analytical process of data mining, application of the Binary Classification (Cortez and Silva, 2008). The dataset consisted of 31 Input Attributes and one Target Attribute shown in Table 1. The researchers determined the grades of basic English of students with Binary Classification, classified into the group of students with effective English language learning (Students earning grade A) and the group of students with results of learning of general, basic English (students earning other grades).

Data Preparation

The inspection of the incompleteness of the data collected from the questionnaire of 299 samples was done using Microsoft Excel program to delete the irrelevant parts and modify the language of all the questions from Thai to English as in Figure 2.

Table 1: Dataset Relevant to the Success of English Language Learning for Data Processing

Features	Explanation	Modified Value
GENDER	Gender of the students	{F, M}
AGE	Age of the students	{numeric 18 to 25}
H-COM	Possessed computer and internet for personal use	{yes, no}
LOC	Location of habitation	{urban, rural}
ABSENCE	Number of times of absence from English class	{numeric 1 to 10}
S-ENG	Extra lessons in English	{yes, no}
F-TIME	Free time from class	{very little, little, moderate, high, very high}

Table 1: Dataset Relevant to the Success of English Language Learning for Data Processing (cont.)

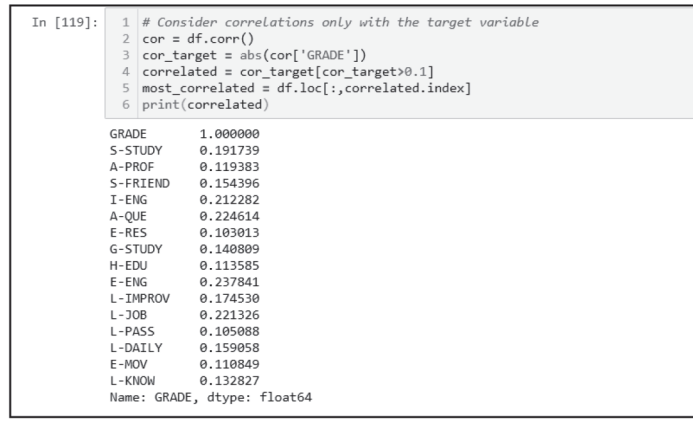
Features	Explanation	Modified Value
H-FREQ	Time to go out with friends	{very little, little, moderate, high, very high}
S-STUDY	Free time to practice English by oneself	{very little, little, moderate, high, very high}
A-PROF	Asked lecturers immediately when the students did not understand the content of the lesson	{very little, little, moderate, high, very high}
E-PLAN	Planning of English language learning	{yes, no}
P-SELF	Preparation before attending English language class	{yes, no}
S-FRIEND	Preferred to sit next to the friend who was attentive to learning	{yes, no}
F-FRIEND	Had foreign friend with whom to practice English skills	{yes, no}
I-ENG	Showed interest and attentiveness in English learning during lecture	{very little, little, moderate, high, very high}
T-NOTES	Took notes during lecture	{very little, little, moderate, high, very high}
A-QUE	Answered questions while learning English	{very little, little, moderate, high, very high}
A-SCHOOL	Revised English lessons after class	{very little, little, moderate, high, very high}
E-RES	Conducted additional study and research on English apart from lecture	{very little, little, moderate, high, very high}
C-STUDY	Studied in groups to revise English lessons with friends upon the approach of the examination	{very little, little, moderate, high, very high}
H-EDU	Desired to study at higher level	{very little, little, moderate, high, very high}
F-SUPP	Family support for English language learning	{yes, no}

Table 1: Dataset Relevant to the Success of English Language Learning for Data Processing (cont.)

Features	Explanation	Modified Value
E-ENG	Enjoyed learning English	{very little, little, moderate, high, very high}
L-IMPROVE	Interest in learning English for self-development	{very little, little, moderate, high, very high}
L-JOB	Interest in learning English for finding jobs in the future	{very little, little, moderate, high, very high}
L-PAREN	Interest in learning English due to parents' wish	{very little, little, moderate, high, very high}
L-PASS	Interest in learning English to pass the examination	{very little, little, moderate, high, very high}
L-DAILY	Interest in learning English for daily communication	{very little, little, moderate, high, very high}
E-MOV	Joy in reading books and watching films in English	{very little, little, moderate, high, very high}
L-GRADE	Learning English was important due to the wish to have good grades	{very little, little, moderate, high, very high}
L-KNOW	Learning English was important because English enhanced more knowledge	{very little, little, moderate, high, very high}
Grade	University's grades of basic English	(A, B, C, D, F)

Then, the Jupyter program and the Python language were used to convert data for appropriateness prior to entering the process of data mining through the methods of Factorization to replace each dataset with the features in the form of message (Label) or figures (Numerical) as the data was continuously related and not a lot of different data was dispersed. The survey was also conducted of all data with 32 attributes to forge understanding and to be used to consider the statistical trend within the dataset. The ratio of the group of students with effective English language learning totaled 42 students and the group of students with the grades of basic general English totaled 257 persons. The survey revealed that most of the sample were female more than male and almost all students had computer and internet for personal use. At the same time, they did not take extra lessons

in English language. Then, the researchers selected only key features to represent the dataset based on Pearson's Correlation. The researchers selected the attributes with the value of over 0.1 for analysis. The remaining 15 attributes were correlated to the grades of basic English as in Figure 2.



```
In [119]:
# Consider correlations only with the target variable
cor = df.corr()
cor_target = abs(cor['GRADE'])
correlated = cor_target[cor_target>0.1]
most_correlated = df.loc[:,correlated.index]
print(correlated)

GRADE       1.000000
S-STUDY     0.191739
A-PROF      0.119383
S-FRIEND    0.154396
I-ENG       0.212282
A-QUE       0.224614
E-RES       0.103013
G-STUDY     0.140809
H-EDU       0.113585
E-ENG       0.237841
L-IMPROV    0.174530
L-JOB       0.221326
L-PASS      0.105088
L-DAILY     0.159058
E-MOV       0.110849
L-KNOW      0.132827
Name: GRADE, dtype: float64
```

Figure 2: Data of the Features Selected to Construct the Model in the Jupyter Program

Construct the Model

The Sklearn library of Python was used to determine the ratio of the dataset of 229 samples divided into two parts namely training dataset of 70% for instruction (209 samples) and the test dataset of 30% for testing of model (90 samples) as it determined the basic ratio of the data testing (Alls, Karakurt, and Melli, 2000). Then, the algorithm was chosen with the Decision Tree and the Random Forest techniques for testing and parameter adaptation through the dataset of 15 attributes (Variable X) to determine the grade of the students' basic English as the target attributes (Variable Y) and divide the effectiveness of English language learning.

Assess the Effectiveness of the Model

The test of effectiveness of the model would be based on the consideration of the most appropriate model between the Decision Tree technique and the Random Forest technique. The researchers used the method of k-fold cross validation by dividing the training dataset and test dataset equally into ten parts. The nine parts would be used for the training dataset and the remaining one part would be used for the test

dataset. Each part derived from random sampling and training and testing would be conducted repeatedly for ten times. The researchers used the Confusion matrix as an instrument to assess the results of the effectiveness of the model divided by types. The effectiveness of the model was conducted based on Accuracy, Precision, Recall, and F1 score in order to find the accuracy of the model with the highest value.

Utilization

The research results from the analysis revealed the effective features or behavior of undergraduate students in English language learning. The features from the analysis based on the data mining could be used to plan the students' English language learning, which would benefit the enhanced analysis for higher effectiveness of the students' other courses.

Study Results

Results of the Comparison of the Model

The results of the construction and the testing of the model based on the Decision Tree technique and the Random Forest technique revealed that the Random Forest technique yielded the forecast accuracy of the effectiveness of the students' English language learning equaled to 85.14% whereas the forecast accuracy of the Decision Tree technique was 78%. These were assessed from the Accuracy Mean which calculated the accuracy from the mean and the standard deviation of all the test results based on the 10-fold cross validation as shown in Table 2.

Table 2: Testing Results of the Model Based on the K-fold Cross Validation

Algorithm	Accuracy Mean	Accuracy STD
Random Forest	85.14	6.26
Decision Tree Classifier	78.00	5.26

Testing was conducted of the accuracy of the model once again with the test method of Confusion Metrix by finding the accuracy from the ratio of the classification of types with Training 70% and Testing 30% to identify the level of accuracy in classifying the types of data. It was calculated from the ratio of the number that the models correctly

classified the types with the total amount of data. It was found that the Random Forest technique still had high accuracy with the value of 0.84 and the accuracy of the Decision Tress technique equaled 0.69. The comparison of the effectiveness of the models revealed that the model developed with the Random Forest technique provided accuracy to analyze the key features affecting the effectiveness of English language learning of undergraduate students more than the model developed with the Decision Tree technique. The researchers chose the model from the Random Forest technique to analyze 15 key features.

Analysis Results of the Key Features Using the Random Forest Technique

Based on the data processing with the Random Forest technique to analyze the features that identified the effectiveness of the students' English language learning, the sample of the students of Faculty of Geoinformatics, Burapha University were used. They consisted of 299 students with 15 features (Variable X). It was found that the results of the variables of the first five features mostly correlated to the effectiveness of the students included enjoyment of English language learning which had a 0.1194 correlation to the effectiveness of the students' English language learning, and asking questions to lecturer immediately when not understanding the content of learning had a 0.1149 correlation to the effectiveness of the students' English language learning. The interest and attentiveness in English language learning while the lecturer was teaching had a 0.0884 correlation to the effectiveness of the students' English language learning. The study in groups to review English lessons with friends when the examination was approaching had a 0.0864 correlation to the effectiveness of the students' English language learning. The additional study or research on English apart from the lecturers' teaching had a 0.0747 correlation to the effectiveness of the students' English language learning. The graph derived from the analysis using the Random Forest technique was in Figure 3.

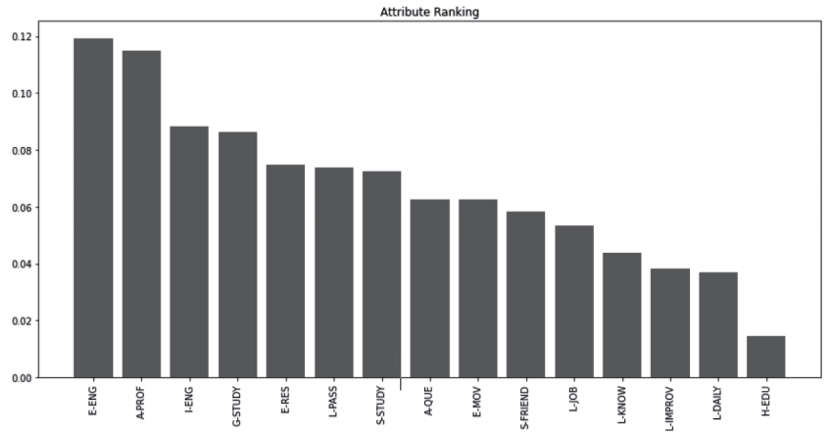


Figure 3: Graph of the Attribute Ranking Which Identified the Effectiveness of the Students' English Language Learning

Analytical results of the features identifying the effectiveness of the students' English language learning compared the ratio of the group of students with effectiveness of English language learning (grade A of basic English) and the group of students with the grades of the basic general level of English as in Table 3.

Table 3: Ratio of Students with the Grade A of Basic English of the First Five Ranks

Features	Level of Frequency (%)				
	Very Little	Little	Moderate	High	Very High
E-ENG	12.5	11.11	6.25	27.08	**40.0**
A-PROF	15.09	7.14	17.14	**27.77**	**27.27**
I-ENG	0	11.11	7.23	27.78	**30.43**
G-STUDY	**21.74**	16.28	13.48	10.00	0
E-RES	14.81	11.26	12.76	10.34	**43.75**

The results of the testing of effectiveness of the model developed with the Random Forest technique using the Confusion Metrix in measuring the effectiveness revealed that the Accuracy Score of the model in classifying the types equaled to 86% and measuring the Precision, Recall, F-Score equaled to 87%, 94%, and 92%, respectively. They demonstrated the perspective of the ability of the model to classify the group of students with the performance of basic general English (determined grades B, C, D, F

of basic English). At the same time, the perspective of the ability of the model to classify the group of students with effectiveness in English language learning (determined grade A of basic English) equaled to 50%, 15%, and 24% respectively. This part demonstrated that for the features identifying the effectiveness of English language learning, the data might be too little.

Conclusion and Discussions

Data mining contributes to the extraction of interesting knowledge from data and widespread use. Many research studies use data mining to improve the quality of education and increase the effectiveness of instructions. For this study, the test was performed to compare the effectiveness of the models between the Decision Tree technique and the Random Forest technique to select the model appropriate to the analysis of the features affecting the effectiveness of English language learning among undergraduate students. The model was developed by using the data collected from the questionnaire after the experts' testing of precision. The study results revealed that the Random Forest technique was more accurate and effective than the Decision Tree technique. This was in accordance with the research of R (2015) who explained that the Random Forest technique was a group of Decision Tree consisting of numerous Decision Trees. The algorithm would compare and analyze each Decision Tree separately. As a result, the rate of the mistakes of the Random Forest technique was smaller than the Decision Tree technique which rendered the forecast more accurate.

As for the results of the algorithm of ranking the key features in correlation to the effectiveness of English language learning of undergraduate students with the Random Forest technique, the key study results from the analysis of the first five ranking could be summarized as follows:

1) In terms of enjoyment in learning English, the study results revealed that the level of great enjoyment had direct correlation to the students with effectiveness of English language learning. It demonstrated that the group of students with effectiveness of English language learning usually enjoyed learning English. This was consistent with the research of Lucardie (2014) who stated that the motivation and enjoyment in learning any subject would connect outcome, effectiveness, and success in learning the subject. Many factors that could create the positive learning environment to motivate enthusiasm

and enjoyment in learning English. The lecturers could use the results of the study to modify the curriculum in the classroom, pleasurable activities such as English exercise, video games, and social activities to forge friendly atmosphere, and support better English language learning of the students.

2) In terms of asking lecturers immediately when they did not understand the content of the lesson, the study results revealed that the students tending to seek assistance or ask the lecturers when learning English were usually the group of students with the effectiveness of English language learning. This was in accordance with Prathan (1994) who stated that the technique to forge understanding in learning was to prepare in advance and ask questions to acquire real knowledge and understanding. Therefore, the lecturers could specifically allocate time to construct exercise or activity to raise questions. This activity could provide the opportunity for both students and lecturers to ask questions. Moreover, asking questions would ensure the students' success and the lecturers' learning from the students as well.

3) In terms of interest in learning English, the study results revealed that the students with effectiveness of English language learning were interested in learning at a high level compared with the students with normal learning results. This was in accordance with the academic research of Renninger et al. (1992) who explained that students with more interest in learning tended to have better grades than students with low interest in learning. Therefore, the arrangement of appropriate form of instruction and the creation of interesting learning atmosphere could be an instrument that would enhance interest in the English language learning. Lecturers could organize the activities that students might learn English apart from textbooks in general. They would render the English language learning more interesting and enjoyable. There were numerous methods of arranging the classrooms that were more interesting to students such as learning through films and TV programs, etc.

4) Forming groups to revise the English lessons with friends before the examination revealed that the students with effectiveness of the English language learning usually did not form group to review the lessons. This behavior could be analyzed that the students with effectiveness of the English language learning might clearly allocate their time, as well as revise the lessons, and study or research for additional information on English. Therefore,

it was not necessary to revise the lessons before the examination. The academic research showed that the strategy of forming in group had both negative and positive impacts on students' learning and development. It depended on the methods used in the classroom (Chen, Hwang, & Lin, 2019). The review of lessons near the examination date might have both positive and negative impacts. Lecturers could use this information to observe the students' behavior such as from handing in homework, assigning work, or motivating the students to allocate time for examination preparation, etc.

5) Additional study or research on English apart from lecturers' teaching revealed that the students with effectiveness of English language learning were interested in additional study or research on English which constituted an important behavior of learning. The English language learning outside the classroom for additional review of lessons was in accordance with the concept of Maxwell-Jolly (2011). It showed that extra time used for education outside the classroom could improve motivation and participation of students, as well as promote social relationship between students and families. These results could be used to arrange for instruction that might focus on providing additional information sources and necessary instruments through classrooms.

Recommendations for future research

1. This research was conducted and tested by using the sample from the Faculty of Geoinformatics, Burapha University. Therefore, the results of the development of the sample derived from one group of samples which might not be sufficiently inclusive. Therefore, to acquire the effectiveness in forecasting the model in the future study, the samples may have to be dispersed and added for more inclusive data collection.
2. Measurement of the effectiveness of the English language learning among undergraduate students may have additional factors that can be used to determine the criteria to measure the effectiveness of the students' English language learning better than the grades of individual course such as construction of the model to test the level of English and compare the scores with the grades received from the university, etc.

3. Study should be conducted on other causal variables expected to impact the effectiveness of English language learning in order to compare and plan effective and inclusive learning. There may be other factors that affect the effectiveness of students' English language learning which can be additionally analyzed such as grades of other courses, skills, number of learning hours a day, social factors, and quality of instruction, etc.

References

Alizadeh, M. (2016). The Impact of Motivation on English Language Learning. *International Journal of Research in English Education*, *11*.

Broussard, S. C., & Garrison, M. E. B. (2004). The relationship between classroom motivation and academic achievement in elementary school-aged children. *Family and Consumer Sciences Research Journal, 33*(2), 106-120.

Chanasap, W. (2017). *Applying Data Mining Techniques for Career Guidance of Undergraduate Students, Faculty of Archaeology, Silpakorn University*. Proceedings of the 7th National and International Graduate Studies Conference: Thailand 4.0 Innovations for Sustainable Development (NGSC&IGSC).

Chen, I., & Huang, S. (2003). Language learning strategy use differences between high and low proficiency learners – an example from a technology college in Taiwan. *Journal of Humanities of Changhua Teachers College, 2*, 301-321.

Cortez, P., & Silva, A. (2008). *Using data mining to predict secondary school student performance*. Proceedings of 5th Annual Future Business Technology Conference, Porto, 2008.

Ellis, R. (1997). *The study of second language acquisition*. Oxford University Press.

Gardner, R. C. (1975). Motivational variables in second language learning. *Proceedings of the Canadian Association of Applied Linguistics*, 45-73.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Rowley, Massachusetts: Newbury House.

Kabbra, R. R., & Bichkar, R. S. (2011). Performance Prediction of Engineering Students using Decision Trees. *International Journal of Computer Applications*, 8.

Krejcie, R.V., & Morgan, D.W. (1970). Determining sample size for research activities. *Educational and Psychological measurement*. (1970):607-610.

Lucardie, D. (2014). *The impact of fun and enjoyment on adults' learning*. *Procedia - Social and Behavioral Sciences, 142,* 439-446.

Maxwell-Jolly, J. (2011). *English Learners and Out-of-School Time Programs: The Potential of OST Programs to Foster EL Success*. ENGLISH LEARNERS AND OUT-OF-SCHOOL TIME PROGRAMS, 1-12.

Niratsai, S. (2022). Attitudes, Motivation, and English Learning Behavior. *Journal of Humanities and Social Sciences, 9*(2), 160.

Omer, F. A., Karakurt, E., & Melli, P. (2000). Data Mining For Database Marketing At Garanti Bank. *WIT Transactions on Information and Communication Technologies, 25*, 93-107.

Oxford, R. L. (1990). *Language Learning Strategies: What every teacher should know*. New York: Newbury House/ Harper Collins.

Petchrat, P. (2004). *Factors Related to Learning Performance of Students at Kasetsart University*. Unpublished Master's Thesis, Srinakharinwirot University.

Renninger, A. K., Hidi, S., & Krapp, A. (1992). *The Role of Interest in Learning and Development*. New York: January 17, 2014 (1st Edition).

Spraat, M., Pulvemess, A., & William, M. (2005). *The TMT teaching knowledge test course*. Cambridge: Cambridge University Press.

Suebpong, E. (2022). *Data Mining*. Bangkok.

Suwannapruit, T., & Totarat, S. (2016). Developing Foreign Language Learning Strategies to Enhance Listening and Speaking Skills in Undergraduate Students. *Veridian E-Journal*, Silpakorn University, 2.

Visser, P. S., Holbrook, A. L., & Krosnick, J. A. (2008). *Knowledge and Attitudes. Handbook of Public Opinion Research*. Thousand Oaks, CA: Sage Publications.

Wattanawat, P. (1994). *Learning Is Not Difficult If You Want to Learn*. Bangkok: Charoenphol Printing.

Wu, M. M. F. (2008). Language learning strategy use of Chinese ESL learners of Hong Kong - findings from a qualitative study. *Electronic Journal of Foreign Language Teaching, 5*(1), 68-83.

Wutthiwongsa, N. (2014). Strategies to Foster Motivation: Developing English Language Skills. *Journal of Management, 34*(1), 89-97