

# **FREE ELECTIVE COURSE RECOMMENDATION WITH MBTI PERSONALITY AND DATA MINING BASED ON STUDENTS' PERFORMANCE**

Thongchai Kaewkiriya,  
Panyapiwat Institute of Management, Thailand  
tkaewkiriya@gmail.com

Phuwadol Viroonluecha,  
Top Sky Group, Thailand  
phuwadolviro@gmail.com

---

## **ABSTRACT**

Free elective courses are required credits for college student's graduation but sometimes these courses can complicate course choices for students. Well-advising is needed for students to select the course. This paper aimed to create the free elective course recommendation using classification approaches of data mining such as J4.5 Decision Tree, Logistic Model Tree, Random Forests and Gradient Boosted Trees. The recommendation is personalised so attributes of data reflect on an individual that are personality type based on Myers-Briggs type indicator revealing students' preferences and their learning style. Moreover, student's performance at college e.g., enrolled courses and its grade is applied to the study. The label consists 9 of free elective courses which are identical requirements and time schedule. The optimal performance algorithm for this research is the Logistic Model Tree with 70.19% of accuracy and 12.2 seconds of run time whilst the original methods of the Logistic Model Tree which are Logistics Regression and Decision Tree performed around 50% of accuracy.

**Keyword:** Free electives, MBTI, Data Mining, Course Recommendation, Learning Analytics

---

## **1. INTRODUCTION**

College students have an opportunity to choose desirable courses in addition to the core courses in their degree structure which called "free elective courses". Students are free to choose and can take literally any class they want and it counts toward their degree. It could be a supplemental course in the student's major area of study, a class on a subject student has been wanting to find out or a unique class that sounds like it would be a ton of fun and would not require excessively work. The number of free elective courses that students will have to take will depend on pursuing a degree. Each college requires a specific minimum of credits to accomplish a bachelor's degree, however, each field of study has its own core prerequisites. Some degree program might be so tight with its core subjects that there is only space for one or couple of free electives in the suggested degree structure. Other programs might be more flexible. Obviously, in case of a student is pursuing a double major or some other blend degree program, student's choices will be undeniably constrained.

On the other hand, choosing a free elective course can complicate course selection for students. Well-advising is necessary for supporting students to select the right free elective courses and maintain retention rates and their graduation rates. The free elective course recommendation is based on several features such as personal features (e.g., age, school's grade point average, hometown) and academic features (e.g., student major, academic level, previously enrolled courses and its grade).

Educational data mining and recommendation system are applied to improve the course advising for students and educators. These data mining and recommendation approaches are not only for prediction what course is suitable for a particular student. Collaborative filtering is one of the methods based on matrix factorization. This approach is using for a movie and product recommendation [19] and also applied for grade forecast [12] and course recommendation [17]. In addition, either supervised machine learning (e.g., linear and logistic regression, decision tree, artificial neural networks) or unsupervised machine learning (e.g., K-means clustering, hierarchical clustering, DBSCAN).

The Myers-Briggs type indicator (MBTI) is a widely used personality assessment [18] for comprehending the personal learning type which includes four pairs of opposite poles that are extroversion, sensation, feeling, and perception as personality dimensions [6]. In Thailand, the MBTI assessment frequently uses in the educational field and research, such as the relationships between the MBTI and student's academic performance [5,14], the personalised course design by MBTI [3].

In this paper, we present the course prediction as the course recommendation for free electives in the bachelor degree program in modern trade business management at the Panyapiwat Institute of Management. The research contribution is mapping the students' MBTI personality traits with students' academic performance aiming to predict the free choosing course of the degree. Furthermore, classification machine learning technique was applied to this study. The paper was divided into six sections. The first section is an introduction. The background and related works located in section two and three. The fourth part is the process and algorithm. The noticeably different results were compared and researchers have discussed the results and conclusion in the fifth section in this paper. The last section is the references to the study.

## 2. BACKGROUND

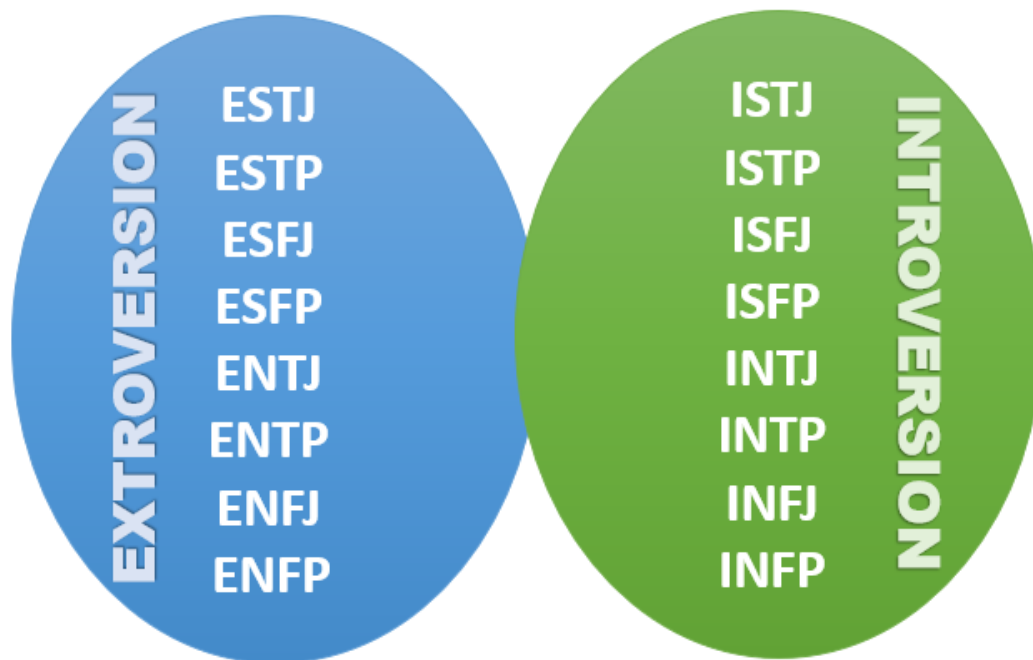
### 2.1 Free Elective Course in Bachelor Degree in Modern Trade Business Management

The bachelor programme in modern trade business management at the Panyapiwat Institute of Management has required 138 academic credits with four years of study. The degree structure includes general subjects – 33 credits, core subjects – 99 credits and free elective subject – 6 credits. The student will have an opportunity to choose two subjects as their desirable choice. These free electives are credits which students need to graduate, but students can select any course they want. The subject either does or does not relate to student degree's core subjects or major requirements. The student's chance for free electives are on their semesters 6 and 7 in ordinary course structure.

## 2.2 Myers-Briggs Type Indicator

The Myers-Briggs Type Indicator (MBTI) is a psychological personality assessment constructed by Katharine Cook Briggs and Isabel Briggs Myers [16] based on Carl Jung's theory of personality types. The purpose of this assessment is indicating different human psychological preferences underlie interests, needs, values, and motivation which lead to career's and learning's choice.

The four opposite pairs of preferences: introversion (I) and extroversion (E), intuition (N) and sensing (S), feeling (F) and thinking (T), and perception (P) and judging (J) is combined to one personality traits to describe the tester. These abbreviations are applied to all sixteen types which are representing in Figure 1.



**Figure 1.** The 16 personality types of MBTI

In addition, various individual learning styles are also based on these four pairs of personality [13] which are

1. Extroversion – extroverts prefer to learn by discussion and interaction with others. They acquire new information by participating with the surroundings.
2. Introversion – introverts prefer to quietly ponder, to have privacy. New information for these people arises from finding ideas and concepts from within their mind.
3. Intuition – people with intuition tend to learn in an atmosphere that gives importance to the implications and relationship for each other. Deep understanding is more valuable than careful observation. The pattern of recognition has occurred from nature for this type of person.
4. Sensing – individuals with sensing perception are happy to learn when having a good explanation and step by step. Most of them recognize today occurrences. They obtain a concrete experience and then transform into abstraction

5. Feeling – they pursuing the truth, natural form, the logical principles, and deductive logic of nature.
6. Thinking – they focus on issues and causes that can be made as they consider the motives of others.
7. Perception – people with this type will develop themselves in a flexible learning environment that encourages new and interesting ideas. The judgment will occur at the moment of the situation but awareness will occur after that
8. Judging – people with this type will develop themselves when information is organized in systematic and structured manner.

## 2.3 Decision Tree and Related Algorithms

The Decision Tree (DT) algorithm is a nonparametric supervised learning technique used for classification and regression problems. The algorithm structures by nodes, branches and leaf nodes [11]. Nodes are if-conditions of the decision; Branches are node interconnection that demonstrates possible features of each node; Leaf nodes are results of the decision. There are several algorithms based on DT's concept such as ID3, C4.5, and CART.

Whereupon, a variety of enhancing algorithms based on DT was invented. The Random Forests (RF) and Gradient Boosting Trees (GBT) algorithms are from related method to DT. The RF is an ensemble data mining technique for classification and regression that build from a multitude of decision trees [1]. The GBT is a technique for building predictive models from combining weak learners together [4]. These methods are developed to avoid DT's over fitting that frequently occurs in an experiment.

## 2.4 Logistic Model Tree

The logistic model tree (LMT) [15] is a classification model with an associated supervised training algorithm that combines logistic regression and C4.5 decision tree in which the information gain is used for splitting and the LogitBoost algorithm is used to produce a logistic regression model at every node in the tree.

The basic LMT induction algorithm uses cross-validation to find a number of LogitBoost iterations that does not overfit the training data and the CART algorithm is used for pruning. The LogitBoost algorithm performs additive logistic regression with least-squares fits for each class  $C_i$  as follows:

$$L_c(x) = \sum_{i=1}^D \beta_i x_i + \beta_0$$

where  $D$  is the number of conditioning factors and  $\beta_i$  is the coefficient of the  $i$ -th component in the input vector  $x$ . The posterior probabilities in the leaf nodes of the LMT are processed using the linear logistic regression method.

$$p(C|x) = \exp(L_c(x)) / \sum_{c'=1}^C \exp(L_{c'}(x))$$

where  $C$  is the number of classes and the least-square fits  $L_C(x)$  are transformed such that  $\sum_{C'=1}^C \exp(L_{C'}(x)) = 0$ .

### 3. RELATED WORKS

Student performance's prediction is one of the most well-known topics in educational data mining. **Modelling and predicting students' academic performance using data mining techniques** was proposed by A. Mueen et al. [2], which aimed to predict and analyse students' academic performance based on their academic record and forum participation. They used three algorithms for this research that is (naïve Bayes, neural network, and decision tree). The result is naïve Bayes classifier well performed with 86% of accuracy. M. A. Al-Barrak and M. Al-Razgan [10] conducted the research in final GPA prediction for IT students with the J48 decision tree as an algorithm for modelling.

Predicting to predict student course selections in higher education using institutional data was proposed by I. Ognjanovic et al. [8]. The student's preferences are modelled using the **Analytical Hierarchy Process (AHP)**, to predict student course selection. There are 1061 instances in the study. The findings suggest that a student's grade point average relative to the grades of the courses they are considering for enrolment was the most important factor in determining future course selections.

The recommender systems for university elective course recommendation using Pearson Correlation Coefficient and Alternating Least Square was proposed by K. Bhumichitr et al. [9]. The experiment with Alternating Least Square outperformed than the other with 86 per cent of accuracy.

Latent semantic analysis (LSI) and Doc2vec was applied in Course recommendation based on semantic similarity analysis [7]. H. Ma et al. built the system based on content similarities and deployed their system and 20 students were satisfied with recommended courses from the system. Their algorithm succeeded to find similar courses among the 537 courses we analysed.

**This research aims to build the course recommendation for free elective course selection based on 16 personality types of MBTI assessment and students' academic performance using data mining techniques such as decision tree's family algorithms.** The course recommendation could help students to choose their free elective courses by their personality and performance and support students towards their graduation.

### 4. METHODOLOGY

In this research, we conducted in RapidMiner, Weka, Spreadsheet as tools. The research's sections include:

#### 4.1 Data Gathering

There are two main steps in this process. **First, we collected students' data which contains students' profile, course enrolment and course grades from the database of registration office.** The enrolment and grade's data from graduates in major of modern trade business management at the Panyapiwat Institute of Technology in 2014-2017. Some students were in general programme with four years study but some were in a continuing programme with two years of study for who has a high vocational degree.

Second, MBTI assessment was collected from an online questionnaire since students are graduated and live in various areas in Thailand so it is a convenient way for an interviewee. The interviewees are graduates of the major of modern trade business management at the Panyapiwat Institute of Technology who started the degree in 2014.

## 4.2 Data Processing

The collected data from two sources (institute's database and online questionnaire) was mapped into one table by the student's ID, then the data was eliminated missing and duplicated. After that, the data was normalised and standardised into numeric. As a result, there are 36 attributes and 1251 instances after cleansing the data.

Table 1 demonstrates the attributes of this study after data processing. There are student's cumulative GPA from high school, hometown, grades from 32 courses in semesters 1-5, MBTI personality type and a free elective course of student's choice as a label. Researcher separated data into two parts, one for training the model and other for evaluating the model by the ratio 8:2 and we will use all features in the next step.

**Table 1.** Attributes of the study

Attribute	Detail
Cumulative GPA from High School	4.00-point grading system (from 0.00 to 4.00)
Hometown	Bangkok and 76 provinces
Grades from 32 courses	8-scale grading system (A, B+, B, C+, C, D+, D, F)
MBTI Personality Type	16 types (as depicts in table 1)
Chosen Free Elective Course	9 courses with identical conditions of choice (e.g. time schedule, requirement)

## 4.3 Modelling and Evaluation

The proposed model in this paper is **Logistic Model Tree (LMT)** performing classification which is a supervised learning method that obtains predictive results through gradually improved estimations. Accuracy was used to evaluate the performance of this model.

$$\%Accuracy = 100 - \%Error$$

The percentage of error can find by finding a relative error

$$Relative\ error = |(X_{mea} - X_t) / X_t|$$

then

$$\%Error = Relative\ error \times 100$$



Where  $X_{\text{mea}}$  is measure value and  $X_t$  is a true value.

The evaluation was done by the 10-folds cross-validation. To compare the results, researchers use original algorithms of LMT which is logistic regression and decision tree, powerful modified decision tree which is the Random Forest Algorithm (RF) and the Gradient Boosted Trees (GBT). The Deep Neural Networks are including in the experiment since the algorithm is popular with current data mining's research. These algorithms are in the same group for classification problem-solving. The performance's comparison will demonstrate in the next section.

## 5. RESULT AND DISCUSSION

### 5.1 Result

The result of this research depicts in Table 2. The performance of LMT was significant from other algorithms. The accuracy of LMT was notable from the remaining by 70.19 %. The original algorithms of LMT presented the only hairbreadth from a middle line – LR 50.16 % and DT 50.20%. The complex DT was well performed as well – GBT 69.77 % and RF 65.01 %. The runtime is important to the modelling to as we have tons of data and need to rapidly obtain a result. The DNNs has not performed well with 64.42% of accuracy and 10.46 seconds of process time. LMT can process the data within 12.2 seconds with the highest accuracy of the research. On the other hand, the fastest modelling technique in this study is DT but its accuracy too low to be used.

**Table 2.** The performance's comparison

Algorithm	Accuracy (%)	Run time (sec.)
Logistic Regression	50.16	19.94
Decision Tree	50.20	<u>2.63</u>
Logistic Model Tree	<u>70.19</u>	<u>12.2</u>
Gradient Boosted Trees	69.77	77.92
Random Forest	65.01	106.57
Deep Neural Networks	64.42	10.46

### 5.2 Discussion and Conclusion

This study pointed out the problem of choosing free elective courses in college degree which have not many academic credits but it would reflect on students' transcript hence students must choose their choice thoroughly. The free elective course recommendation based on four pairs' personality types of MBTI assessment and students' academic performance using Logistic Model Tree techniques could help students in this problem.

The framework of the research is based on two sources of data that are students' course enrolment and course grades from the institute's database and the online questionnaire includes the MBTI assessment. Then, the researcher mapped it into a table and model it with several machine learning algorithms.

The LMT technique performed better than the GBT technique only 0.42% of the accuracy but if we compare their run times, the GBT technique had longer process time than the LMT about one minute that means more data and features, the GBT will

require more time to calculate. So, the optimal result of this experiment is using the LMT with 70.19% of accuracy and 12.2 seconds of its run time. In future research, we would like to improve the research by adding other attributes, feature importance and other personality assessment.

## 6. REFERENCES

- [1] A. Liaw and M. Wiener, Classification and regression by Random Forest. *R news*, 2(3), 18-22, 2002.
- [2] A. Mueen, B. Zafar, and U. Manzoor, Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science*, 8(11), 36-42, 2016.
- [3] A. Tananchai, The Personality of Students Studying the Social Etiquette and Personality Development Course by Myers Briggs Type Indicators (MBTI) Theory. *Swarm and Evolutionary computation*, 3(2), 1-18, 2017.
- [4] B. Gupta et al. Taxi travel time prediction using ensemble-based random forest and gradient boosting model. *Advances in Big Data and Cloud Computing, Singapore, 2018*.
- [5] C. Iaosanurak et al., The Relationships Between Gender Majors' Subject Academic Achievement and Personality with Learning Style of Student Teachers. *Journal of graduate studies Valaya Alongkorn Rajabhat University*, 11(1), 33-43, 2017.
- [6] G. Picciano, Theories and Frameworks for Online Education: Seeking an Integrated Model. *Online Learning*, 21(3), 166-190, 2017.
- [7] H. Ma et al., Course recommendation based on semantic similarity analysis. *International Conference on Control Science and Systems Engineering, Beijing, 2017*.
- [8] I. Ognjanovic, D. Gasevic, and S. Dawson, Using institutional data to predict student course selections in higher education. *The Internet and Higher Education* 29, 49-62, 2016.
- [9] K. Bhumichitr et al., Recommender Systems for university elective course recommendation. *International Joint Conference on Computer Science and Software Engineering, Nakhon Si Thammarat, 2017*.
- [10] M. A. Al-Barrak and M. Al-Razgan, Predicting students final GPA using decision trees: a case study. *International Journal of Information and Education Technology*, 6(7), 528-533, 2016.
- [11] M. Muchová, J. Paralič and M. Jančuš. An approach to support education of data mining algorithms. *International Symposium on Applied Machine Intelligence and Informatics, Herl'any, 2017*.
- [12] M. Sweeney, J. Lester, and H. Rangwala, Next-term student grade prediction. *International Conference on Big Data, California: Santa Clara, 2015*.
- [13] N. Anusornthangkun, Effective factors of the learning of basic computer use in order to create designs for learners with different learning backgrounds. *Suan Sunandha Rajabhat University, Bangkok, 2012*
- [14] N. Anutharun and N. Romphoree, Correlation Between MBTI Personalities and Academic Achievement of Nursing Students of Siam University. *Journal of Nursing, Siam University*, 11(20), 18-27, 2010.
- [15] N. Landwehr, M. Hall, and E. Frank, Logistic model trees. *European Conference on Machine Learning, Berlin: Heidelberg, 2003*.



- [16] P. Basavaraj and I. Garibay, A Personalized Course Navigator Based on Students' Goal Orientation. *ACM Conference on Supporting Group Work, Florida: Sanibel Island, 2018.*
- [17] S. Ray and A. Sharma, A collaborative filtering based approach for recommending elective courses. *International Conference on Information Intelligence, Systems, Technology and Management. Springer, Berlin: Heidelberg, 2011.*
- [18] S. Zamir et al., Relationship between Personality and Occupational Stress among Academic Managers at Higher Education Level. *Research on Humanities and Social Sciences*, 4(7), 21-28, 2014.
- [19] Y. Koren, R. Bell, and C. Volinsky, Matrix factorization techniques for recommender systems. *Computer*, 8, 30-37, 2009.