Nat. Volatiles & Essent. Oils, 2021; 8(6): 2821-2833

NVEO
Natural Volatiles &
Essential Oils

Machine Learning Techniques For The Ordinary National Educational Test (O-Net) Prediction: Case Of Small Sized Schools In Nakhon Sawan Province, Thailand

Wudhijaya Philuek[1] , Surasret Sornchai[2] , Thitiphong Raksarikorn[3] , Sirirat Janyarat[4]

[1]Faculty of Education, Nakhon Sawan Rajabhat University, Nakhon Sawan, 60000 Thailand

[2]Faculty of Education, Nakhon Sawan Rajabhat University, Nakhon Sawan, 60000 Thailand

[3]Rongtateewittaya School, Uthai Thani, Thailand.

[4]Innovative Learning Center, Srinakharinwirot University, Bangkok 10110 Thailand

Abstract:

The objectives of this research were to analyze and predict the outcomes of Ordinary National Educational Testing (O-NET) and to compare the results of Machine Learning techniques with 4 data analysis techniques, namely K-Nearest Neighbor technique, Logistic Regression technique, Decision Tree technique, and Support Vector Machine technique. They also use Microsoft Excel 2013 and Jupyter. The data used to forecast trends of O-NET scores were scores for each subject of schools in KrokPhra District, Nakhon Sawan Province, namely Thai, math, science and English by showing counting data for the quality level of the data used for the forecast analysis. The trend of O-NET scores showed that most of the students' data were at a fair level. The average scores in each subject showed that the average scores for most Thai subjects were at good level. The average scores in most math subjects were good. The average scores in most science subjects were good. And the average scores of most English subjects were at a good level, respectively. And comparing the results of the machine learning method, it was found that the Support Vector Machine technique was the best predictor. It has 65.75% accuracy and the model has more efficiency than other techniques.

Keywords: Machine Learning, Ordinary National Educational Testing (O-NET), K- Nearest Neighbor, Logistic Regression, Decision Tree, Support Vector Machine.

I. INTRODUCTION

Education is important for everyone's life; learners are learning all the time and never stop. Knowledge and education are what all learners seek to use in their implementation in working life and creating success life, therefore it is undeniable that education has an influence on today's global society and expecting to find one's self through education [1]. Learning development is an important factor in organizing the process for learners to enter the standards and to develop themselves. Teachers need to change their role from being a knowledge teller to being a facilitator. The Ordinary National Educational Test (O-NET) is a process that plays an important role in being a tool that reflects the quality of teaching and learning management in schools and the quality of learners. Assessing the quality of basic education of students in each school to bring information in preparing a plan to develop learners to be able to Read, Write, and know how to think. The analysis was to measure the learning achievement in each Grade at the end of three years of study. The test will be taken at the highest level of that class, which is around the end of February or the beginning of March each year. The Ordinary National Educational Test (O-NET) is administered by the National Institute of Educational Testing to use the information to improve and develop students individually before finishing their Grades in the following year. Students can use their scores to be used as a component of the entrance examinations for Grade 7 and Grade 9, or even as a component of class promotion, so students will be able to graduate in each Grade. Students must take the Ordinary National Education Test (O-NET) is a particular duty. Grade 12 must test the O-NET because in addition to being a national quality assessment according to the basic education curriculum. Schools will bring the scores to fill in the “Por Por 1” Form in the national assessment results report. Students can apply the scores to select the Faculty or University for admission to higher education [2]. In addition, the O-NET for students in Grade 6, the school must review the results of the examination because the scores of the five subject areas tested in the first two phases, the students scored less than half of the full score, especially Mathematics and English, and some of them had 0 points which reflected that the quality of students graduating from Grade 12 was not at a satisfactory level. Noting the problem with the child's inability to improve performance could be the exams used in the O-NET were not assessed in line with the classroom teaching which emphasizing on analytical thinking and rationalization [3]. The researcher has seen that the results of the study of O-NET scores of children in the country are lower than the current set standards of the country and there is a statistical increase every year. Therefore, it is necessary to study in order to improve and encourage students to be interested and aware of the O-NET, to develop themselves to have higher learning achievements, and to develop teachers to understand the students and the curriculum. This study was the application of four analytical and forecasting techniques for analysis and forecasting to study the results of O-NET scores in schools of KrokPhra District, and the achievement of each school to see how the O-NET scores in each year's examinations and to improve the school more effective by using O-NET scores to analyze and predict the results by Machine Learning method [4]. Kanchanawasee[5] defined achievement as learning according to a pre-determined plan resulting from the teaching and learning process at any one time. At the initial stage, studies of variables influencing outcomes were mostly based on the fundamental theory that educational outcomes influenced important cognitive variables, including intelligence, language aptitude computational ability, etc., but recent research studies have concluded that intelligence influences only

about 25-50% of academic achievement. The use of intelligence scores alone could not accurately predict educational achievement in later periods, so many other factors influencing the study results were interested in studying the results [6]. A student's academic achievement is related not only to cognitive factors, but also to other factors such as socio-economic factors, motivation, and others that are not related to intelligence, with the components influencing academic achievement. There are many people who have studied it, such as Bloom's concept [7]. Data Mining is an evolution of data storage and interpretation. From the simple storage of data to the storage in the form of a database that can extract information to data mining that can discover the knowledge hidden in the data such as Support Vector Machine [8] K-Nearest Neighbors (K-KNN) [9], Logistic Regression, and Decision Tree [10]. The purpose of this research was to analyze and predict the outcomes of Ordinary National Educational Testing (O-NET) and to compare the results of Machine Learning techniques.

II. MATERIALS ANDMETHODOLOGY

The research on Machine Learning techniques for the Ordinary National Educational Test (O-NET) prediction: case of small sized schools in Thailand has the following steps:

1. Procedures for conducting research

1.1 Study of past O-NET scores

1.2 Study of programs used

1.3 Study the school area of the target group

1.4 The results of O-NET scores of each school were analyzed and compared with the total average of all schools and the national average for each school in each subject.

1.5 Apply forecast analysis techniques to analyze and forecast.

2. Scope of research

Data used for analytics in this research were the O-NET scores of small-sized schools in KrokPhra District, Nakhon Sawan Province under the Office of Nakhon Sawan Primary Educational Service Area 1 and the National Central Value.

3. Research tools

3.1 Working information

3.1.1 Individual O-NET scores for the academic year 2016, 2017, 2018, 2019, and 2020

3.1.2 The average academic results of Thai, Mathematics, Science, and Foreign Languages (English) over 3 years (Grades 4-6) of individual students.

3.1.3 Data Analysis Program

3.2 Working equipment

3.2.1 Computer

3.2.2 Techniques used in data analysis 4 techniques as follows:

- Support Vector Machine
- K - Nearest Neighbors (K-NN)
- Logistic Regression

- Decision Tree

3.2.3 Spreadsheets Program

3.2.4 Jupyter Program

3.2.5 O-NET scores and 5-year average grades used in the analysis

4. Method of collecting data

Step 1: Request for O-NET scores for schools in Krok Phra District

Step 2: Bring the letter that is requested for O-NET scores to be sent to schools in KrokPhra District.

Step 3: Get the O-NET score results from all schools that have contacted you back.

Step 4: Identify factors affecting academic achievement.

Step 5: Select the program to be analyzed.

5. Data analysis

Vanich [11] proposed the used of data mining development process by Cross-Industry Standard Process for Data Mining (CRISP-DM), which has a research step as follows.

Step 1: Business Understanding

It defines the purpose and requirements of the project and then converts that purpose into the form of a data mining problem. At this stage, the researcher studied from the past O-NET scores of each school which encountered some problems, for example, the students' O-NET scores each year did not meet the goals set by each school. The O- NET is not consistent with school grades. To ensure that the scores are consistent with the goals of the school and in accordance with the results obtained from the school to meet the goals. and know the cause of the problem as efficiently as possible.

Step 2: Data Understanding

The researcher studied the existing data and considered the data and the feasibility of data analysis. There are methods as follows.

Collect Initial Data, this procedure begins with collecting O-NET scores and individual grades as data stored in the National Educational Testing Institute database. All O-NET score data is taken from the National Institute of Education Testing and School Registration Web site, then a mock database is created to support the data to be imported and to create a new file. Then bring the data from the real database into the database that has been prepared.

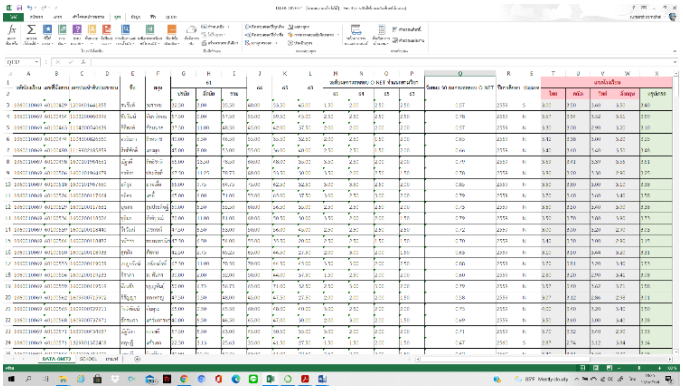

Fig I Sample of raw data

Step 3: Data Preparation

It is the preparation of the data by preparing the raw data available as the data that will be used in the remaining steps, as well as selecting the appropriate variables to analyze and converting the variables to be in the same format to provide the data. The modeling in this section, the researcher used the data of O-NET scores for the academic year 2016 - 2020 of 313 students from 12 schools and the average grades over 3 years (Grade 4-6) of 4 subjects. The data is divided in 3-step process. Data Selection is the selection of fields or related data. Those interested in data mining may be columns from multiple tables forming the dataset of interest.

TABLE I Quantity of information for each school

List	School Code	School Name	Quantity
1	1060010069	Wat Nakhang	50
2	1060010070	Ban Khaopoon	18
3	1060010073	Ban Nernsala	18
4	1060010074	Ban Khongmoung	26
5	1060010075	Ban Khao ThumPra	28
6	1060010076	Wat NernKapree	14
7	1060010078	Wat ThasootJareansil	27
8	1060010079	Wat Manorat	28
9	1060010085	Wat Bang Maphor	21
10	1060010090	Wat Saladang	35
11	1060010091	Ban KrajunkNgam	28
12	1060010092	Ban Hadsoong	20
Total			**313**

From Table I, the amount of data for each school, it was found that most of the students' data, when sorted in descending order, can be described as follows: Wat Nakhang 50 students, Wat Saladang 35 students, Ban Khao ThumPra, Ban KrajunkNgam and Wat Manorat 28 students, Wat ThasootJareansil

27 students, Ban Khongmoung 26 students, Wat Bang Maphor 21 students, Ban Hadsoong 20 students, Ban Khaopoon and Ban Nernsala 18 students, and Wat Nern Kapree 14 students respectively.

TABLE II An overview of the O-NET results benchmark

Quality level	Score	Score Range								Score Range	
		Thai		Mathematics		Science		English		Criterion	
Excellent	4.00	88.00	100.00	90.00	100.00	88.00	100.00	87.00	100.00	88.25	100.00
Very Good	3.50	75.80	87.99	75.00	89.99	75.00	87.99	75.00	86.99	75.20	88.24
Good	3.00	63.70	75.79	60.00	74.99	62.50	74.99	62.00	74.99	62.05	75.19
Quite Good	2.50	51.60	63.29	45.00	59.99	50.00	62.49	50.00	61.99	49.15	62.04
Moderate	2.00	39.50	51.59	30.00	44.99	38.00	49.99	37.00	49.99	36.13	49.14
Fair	1.50	20.01	29.99	15.00	29.99	25.00	37.99	22.00	36.99	20.50	33.74
Improvement	1.00	10.01	20.00	10.00	14.99	10.00	24.99	10.01	21.99	10.01	20.49
Unsatisfied	0.00	0.00	10.00	0.00	9.99	0.00	9.99	0.00	10.00	0.00	10.00

From Table II, the overall comparison of O-NET results has the following criteria:

88.25 - 100 is at an excellent level.
75.2 - 88.24 is at very good level
62.05 - 75.19 is at good level
49.15 - 62.04 is at quite good level
36.13 - 49.14 is at moderate level
20.50 - 33.74 is at fair level
10.05 - 20.49 is at a level that should be improved
0 - 10.00 is at the level, should be greatly improved

Data Transformation was the transformation of data into a format that is ready for use in the analysis of the data transformation process in the selection process to be suitable for the data mining process. An analytical selection feature for correlation in O-NET analysis to be used in comparison with school grades by using the O-NET analysis method, the results were matched with the school-level results with the following variables.

TABLE III Sample data for use in modeling

Id	Thai	Math	Science	English	class
1	3	3.5	3.2	3.5	Moderate
2	3.27	3.54	3.52	3.21	Quite Good
…	…	…	…	…	…

From Table IIIshows sample data before use in modeling class columns, derived from the use of O-NET results comparison criteria to be used as answers used in modeling based on school grades in Thai, Mathematics, Science and English subjects.

Step 4 Modeling Phase

A suitable model was selected with 80% of the data for modeling and 20% of the data for the model performance testing. At this stage, the researchers selected the following data mining techniques and algorithms.

- Support Vector Machine techniques
- K - Nearest Neighbors (K-KNN) techniques
- Logistic Regression techniques
- Decision Tree techniques

Step 5 Evaluation Phase (Model Evaluation)

It is an assessment of the models used in all analyzes to determine the appropriateness of the model's application of how accurate the model is to predict. Based on support, accuracy, precision, recall, and f1-score.

Step 6 Deployment Phase

This research is the application of geographic information system and O-NET results analysis to develop learners. The results of data mining from this research can be used to predict O-NET scores in advance for each school. The results obtained from the model.

III. DATAANALYSIS AND INTERPRETATION

The researcher would like to present results of the data analysis as follows:

1. The results of the forecast analysis of the results of the basic national educational testing (O-NET), the results of the forecast analysis of the results of the basic national educational testing (O-NET) were analyzed as follows:

Table IV The data for the quality level to predict the O-NET scores

Class	Quality Level	Quantity	Percentage
Fair	Fair	147	46.96
Good	Good	3	0.96
Improvement	Improvement	27	8.63
Moderate	Moderate	87	27.80
Quite_Good	Quite_Good	49	15.65
Total		**313**	**100.00**

From Table IV showing the counting data of the quality level of the data used to predict the trend of O-NET scores for the subjects of schools in KrokPhra district, it was found that most of the students' data were at the fair level, representing 46.96% of the total number of 147 person.

Table V The average scores in each subject.

	Thai		Mathematics		Science		English	
class	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
Fair	3.13	0.29	2.83	0.32	2.97	0.34	2.74	0.37
Good	3.20	0.00	3.72	0.15	3.87	0.16	3.22	0.07
Improvement	2.89	0.33	2.57	0.31	2.73	0.31	2.56	0.35
Moderate	3.52	0.22	3.39	0.25	3.42	0.26	3.32	0.29
Quite Good	3.59	0.15	3.50	0.20	3.53	0.21	3.47	0.24

From Table Vshowing the average scores for each subject, it was found that the mean scores for most Thai subjects were at a good level, with a mean of 3.20, and a standard deviation of 0.00. The mean was 3.72, the standard deviation was 0.15, the mean score for most of science subjects was good, the mean was 3.87, the standard deviation was 0.16, and the average English subjects were mostly good. is at 3.22, the standard deviation is 0.07.

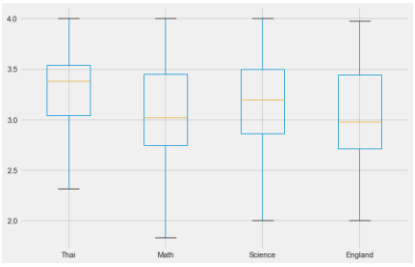

Fig. II Average score of each course

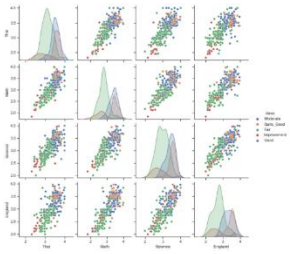

Fig. **III**The distribution of the data.

2. Comparison of model test results by 4 techniques

The comparison results of model testing by 4 techniques as shown in Table VI.

Table VI Model test results

Model	Mean Accuracy	S.D.
K-Nearest Neighbors (KNN)	0.6475	0.4410
Logistic Regression (LR)	0.6425	0.4455
Decision Tree (DT)	0.5925	0.4585
Support Vector Machine (SVM)	0.6575*	0.4375

From Table VI, the model test results showed that the most efficient model was Support Vector Machine (SVM) with accuracy of 65.75 and Standard Deviation of 43.75, followed by K Nearest Neighbors (KNN) with accuracy of 64.75 and Standard Deviation of 44.10, followed by Logistic Regression (LR) with accuracy of 64.25 and Standard Deviation of 44.55, and Decision Tree (DT) with accuracy of 59.25 and Standard Deviation of 45.85, respectively.

Table VII Results of the K-Nearest Neighbors (KNN) Model test.

Class	Precision	Recall	f1-score	Support
Fair	0.75	0.94	0.83	32.00
Improvement	0.00	0.00	0.00	7.00
Moderate	0.43	0.71	0.54	14.00
Quite Good	0.00	0.00	0.00	10.00
accuracy			**0.64**	**63**
macro avg	**0.3**	**0.41**	**0.34**	**63**

From Table VII, the results of the K-Nearest Neighbors (KNN) model test shows that the overall model efficiency forecast is 64.75 Percent accurate. Considering each class, it was found that the Fair class has the precision of 0.75, the recall value of 0.94, and Moderate classes has the precision of 0.43 and the recall of 0.71, respectively.

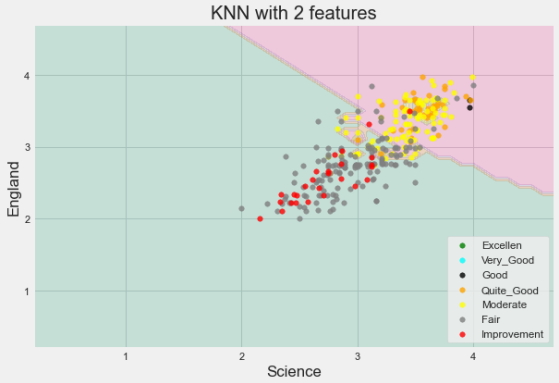

Fig. IV Graph of K-Nearest Neighbors (KNN)Spots

Table VIII Results of the Logistic Regression (LR)Model test.

Class	precision	recall	f1-score	support
Fair	0.75	0.94	0.83	32
Improvement	0	0	0	7
Moderate	0.43	0.71	0.54	14
Quite Good	0.00	0.00	0	10
Accuracy			**0.64**	**63**
macro avg	**0.30**	**0.41**	**0.34**	**63**

From Table VIII, the results of the Logistic Regression (LR) model test, it was found that the overall forecasting model's efficiency was 64.25 Percent accurate. Considering each class, it was found that the Fair class has the precision of 0.75 and the recall value was 0.94,and the Moderate class has the precision of 0.43 and the recall value of 0.71, respectively.

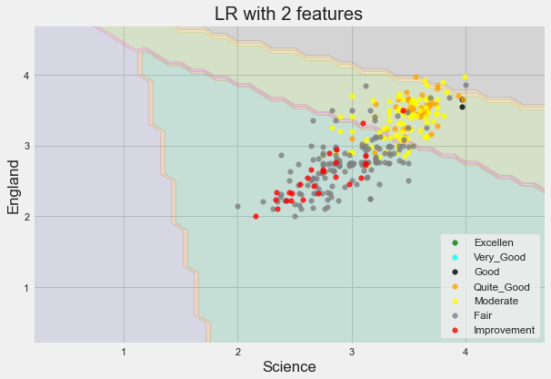

Fig. V Graph of Logistic Regression (LR) Spots

Table IX Results of the Decision Tree (DT)Model test

Class	precision	recall	f1-score	support
Fair	0.74	0.72	0.73	32
Improvement	0	0	0	7
Moderate	0.56	0.64	0.6	14
Quite Good	0.6	0.6	0.6	10
accuracy			**0.59**	**63**
macro avg	**0.48**	**0.49**	**0.48**	**63**

From Table IX, the Decision Tree (DT) model test results showed that the overall forecasting model performance was 59.25 Percent accurate. Considering each class, it was found that the Fair class has the precision of 0.74, a recall value of 0.72, and a Moderate class has the precision of 0.72. 0.56 has the recall of 0.64, and the class Quite Good has the precision of 0.60 and the recall of 0.60, respectively.

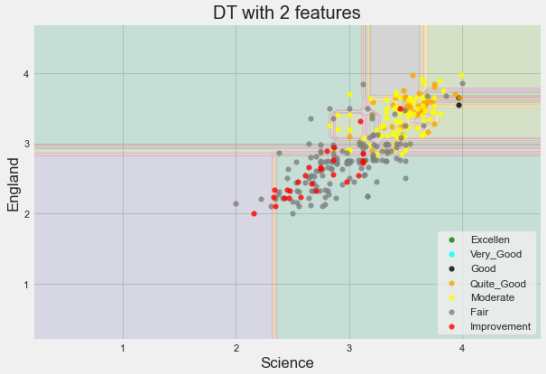

Fig.VI Graph of Decision Tree (DT) Spots

Table X Results of the Support Vector Machine (SVM) Model test

Class	precision	recall	f1-score	support
Fair	0.77	0.94	0.85	32
Improvement	0	0	0	7
Moderate	0.42	0.71	0.53	14.00
Quite Good	0.00	0.00	0.00	10.00
accuracy			**0.65**	**63.00**
macro avg	**0.30**	**0.41**	**0.34**	**63.00**
weighted avg	**0.48**	**0.63**	**0.55**	**63**

From Table X, the test results of the Support Vector Machine (SVM) model showed that the overall model performance forecast was 65.75 Percentaccurate. Considering each class, it was found that the Fair class has the precision of 0.77, the recall value of 0.94, and a Moderate class has the precision of 0.42 with the recall value of 0.71, respectively

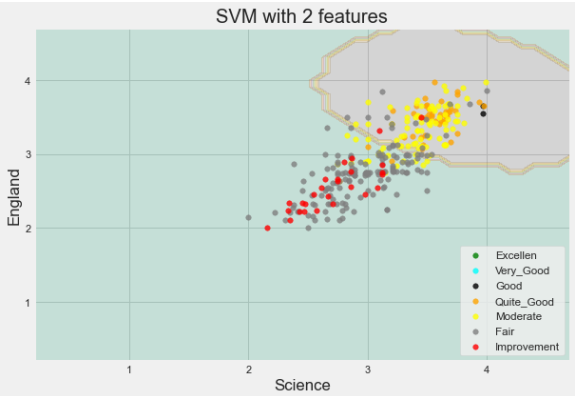

Fig. V Graph of Support Vector Machine (SVM) Spots

III. CONCLUSION

The research of Machine Learning Techniques for the Ordinary National Educational Test (O-NET) Prediction: Case of Small Sized Schools in Nakhon Sawan Province, Thailand found that the data analysis by Support Vector Machine technique has a 65.75% accuracy of the overall model forecast can forecast and classify the level of the best score results. The summary of the data used to predict the trend of O-NET score for Thai language, Mathematics, Science, and English subjects of schools in KrokPhra District found that the quality level of the data used to predict the trend of the O-NET scores showed the most of students' information was at the level of average scores in each subject, which the average scores for most Thai language subjects was good, the average scores for Mathematics subjects was good, the average scores for Science subjects was good, and the average scores for English language subjects was good. The comparison of Machine Learning technique, the results of the model tests used for the analysis revealed that the Support Vector Machine Techniques was more accurate than other models from the tests which had the model efficiency was 65.75%, which was consistent with Thamsiri and Meesat[4] when comparing to K-Nearest Neighbors (KNN), Logistic Regression, and Decision Tree [12].

REFERENCES

[1] Phithaktanatchkul, E. (2015). Education in Today's World. Retrieved May 15, 2021, from http://www.stou.ac.th/study/sumrit/1-58(500)/page9-158(500).html.

[2] Phanpruek, S. (2011) NSTDA Must Not Cause Children to Lose Their Rights. Retrieved May 15, 2021, from https://mgronline.com/qol/detail/9540000032352.

[3] Panthayak, C. (2019). Data Analysis of O-NET Exam Scores Case Study of BangpakokWittayakom School using Data Mining Techniques. Bachelor of Science Thesis in Computer Science. The Faculty of Science and Technology, Kanchanaburi Rajabhat University.

[4] Thamsiri, D. and Meesaj, P. (2010). The Classification of Data Using a Collaborative Decision-Making Method based on the Neural Network, Decision Tree, and Support Vector Machine Techniques in Combination with Selection of Suitable Agents by Genetic Algorithm. Academic Journal of King Mongkut's University of Technology, North Bangkok, Vol. 21, No. 2, May-Aug. 2011: 293-303.

[5] Kanchanawasee, S. (2001) Selection of Appropriate Statistics for Research (4th Edition). Bangkok: Boonsiri Printing.

[6] Anastasi, A. (1961). Psychological Tests: Uses and Abuses. Teachers College Record, 62, 389–393.

[7] Luangsawas, W., Thiravanichtrakul, S., and Rakngam, C. (2018) Factors Affecting Learning Achievement of Students Under Rayong Primary Educational Service Area Office 1. Journal of Education Science, The Faculty of Education, Mahamakut University, Vol. 6, No. 1 (January – June 2018).

[8] Zhang Y. (2012) Support Vector Machine Classification Algorithm and Its Application.

[9] Tang, J., Xia, H., Zhang, J. et al. Deep Forest Regression Based on Cross-layer Full Connection. Neural Computer & Application Journal 33, 9307–9328 (2021). https://doi.org/10.1007/s00521-021-05691-7

[10] Wang T., Li Z., Yan Y., Chen H. (2007) A Survey of Fuzzy Decision Tree Classifier Methodology. In: Cao BY. (eds) Fuzzy Information and Engineering. Advances in Soft Computing, vol 40. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-71441-5_104 In: Liu C., Wang L., Yang A. (eds) Information Computing and Applications. ICICA 2012. Communications in Computer and Information Science, vol 308. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34041-3_27

[11]Vanich, S. (2015) Data Mining. Retrieved May 15, 2021, from https://sajeegm301.blogspot.com/2015/11/data-mining.html

[12] MohdSojak M.R., MohdRazman M.A., P. P. Abdul Majeed A., Musa R.M., Abdul Ghani A.S., Iskandar I. (2019) The Identification of Oreochromis Nilotic us Feeding Behaviour Through the Integration of Photoelectric Sensor and Logistic Regression Classifier. In: Kim JH., Myung H., Lee SM. (eds) Robot Intelligence Technology and Applications. RiTA 2018. Communications in Computer and Information Science, vol 1015. Springer, Singapore. https://doi.org/10.1007/978-981-13-7780-8_18