PREDICTING AND ANALYZING THE STUDENTS' LENGTH OF STUDY-TIME USING SUPPORT VECTOR MACHINE

Teny Handhayani[1]; Lely Hiryanto[2]

[1,2]Computer Science Department, Faculty of Information Technology, Tarumanagara University
Jln. Letjen S. Parman No. 1, DKI Jakarta 11440, Indonesia
[1]tenyh@fti.untar.ac.id; [2]lelyh@fti.untar.ac.id

Received: 24th February 2017/ **Revised:** 17th March 2017/ **Accepted:** 24th March 2017

Abstract **-** The length of study-time is one of the important issues in higher education. The goal of this research was to predict and analyze the length of study-time in the early stage of Computer Science students in X University. The research proposed Mutual Information (MI) as feature selection method and Support Vector Machine (SVM) as a classification method. There were two different sections of the experiments. The first experiment used two class targets that were grouped in 'on time group' and 'late group'. The experiment result shows that the proposed method produces accuracy around 85%. The second experiment used three class targets, 'on time group', 'late group', and 'very late group'. The experiment result of the proposed method produces accuracy around 80%. Mutual Information (MI) does not only successfully raise the accuracy but also uncovers the relationship between subjects and the class targets.

Keywords: Support Vector Machine, Mutual Information, length of study-time

I. INTRODUCTION

Students' grades are one of the important information in academics. Every university stores those in the database. The students' grade dataset has some useful information. It does not only list the students' transcripts but also contains a pattern of the data for further analysis. The collection of students' grade dataset can be used to build a system to predict the length of students time and the students' performance. Predicting the students' performance is useful for academic workers and institution to improve the learning and teaching process (Shahiri *et al*., 2015). Moreover, predicting the students' length of study-time is important for the academic worker and institution to help the students to arrange their study plan.

The length of study-time is a part of the important issue in Indonesia higher education system. It is a duration of study that spent by the students from the first semester up to the maximum of the academic year. According to the government of the Republic of Indonesia (Dirjen Belmawa, 2016), there is a different length of study-time. For example, the full-time bachelor degree students need around 4 to 7 years to finish their degree. Full-time students in bachelor degree have 3,5 years or 7 semesters as the minimum academic year and 7 years or 14 semesters as the maximum academic year is. Moreover, the duration of a semester is around 5 months. The bachelor degree students who fail to finish their study in 7 years will be expelled from the university. Then, they are labeled as drop out.

Indonesian higher education system usually starts the academic semester in September and February every year. The length of study time is not the only criteria for the students to receive bachelor degree status. There are some academic and non-academic requirements that must be fulfilled to be bachelor degree graduate. However, the length of study-time has an important role for the students and their institution. It is also one of the criteria to evaluate the performance of higher education systems by the government. Research related to the length of study-time behavior and academic achievement has been conducted by Ukpong and George (2013).

Some research in educational data mining has been done using various methods. Ogunde and Ajibade (2014) predicted the grade of university students using ID3 decision tree algorithm. They used students' data such as sex, students' entry grade, entrance examination score, and grade obtained in the graduation. The result showed that the performance of ID3 algorithm using IF-THEN rules had produced accuracy of 79,56%. Moreover, Shahiri *et al*. (2015) analyzed the performance of decision tree using IF-THEN rules, Neural Network, Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine to predict the performance of students based on some features. The features were students' Cumulative Grade Point Average (CGPA), internal and external assessments, extra-curricular activities, students' demographic, high school background, social interaction, psychometric factors, and scholarship. The researchers concluded that Neural Network and decision tree performed higher accuracy than other methods.

Then, Taruna and Pandey (2014) compared the performance of decision tree, Naïve Bayes, Naïve Bayes Tree, K-Nearest Neighbor and Bayesian Network for predicting students' grade in four classes for engineering students. There is also a research conducted by Mouri *et al*. (2016). They used Bayesian Network to predict students' final grade using e-book logs data. Next, Bo *et al*. (2015) implemented deep learning for predicting students' performance for junior high school students. Meanwhile, Liu and Cheng (2016) proposed Machine Learning Feature Selection (MLFS) and Support Vector Machine (SVM) to analyze students' academic achievement for the elementary school. Moreover, the research of educational data mining for predicting employability of IT graduates has been done by Piad *et al*. (2016). They identified that IT core, IT professional and gender were variables that had significant features in predicting IT employability. They applied logistic regression which produced 78,4% of accuracy. Moreover, there is also a research regarding the unsupervised method using K-Means that has been applied in mapping students'

performance by Harwati *et al*. (2014). They used dataset consisting of some features. Those features were gender, national origin, parental job, Grade Point Average (GPA), optimization grade, and grade of production planning and control. This research mapped the students into three clusters, namely performance of low students, average students, and smart students.

The aim of this research is to develop a computer system for predicting the length of study-time and analyze the data for decision support system. The system is expected to predict the length of study-time after the students finished their study in the fourth semester. The researchers use a dataset from the X University. The name of the university is hidden to cover the private information. This research is focused on predicting and analyzing the bachelor degree students majoring in Computer Science in X University. In this research, the researchers are interested in conducting the research in Computer Science department. It is because according to the information from the faculty, some Computer Science students have difficulties in the first and second year. Thus, some of them leave or change their major in the early stage of the academic year. Based on this condition, the researchers use the students from the first to the fourth semester.

In X University, the length of study-time for a bachelor degree is 3,5 years to seven years. In this university, there are two groups regarding the length of study-time. The first groups are the students who have a length of study-time about 3,5 to 4 years. It is called as 'on time group'. On the other hand, the students who finish their degree in 5 to 7 years is named 'late group'. Moreover, the others who need more than 7 years or leave their study without completing the rules is grouped into 'drop out'.

The research, analyzes the list of subjects that have an important effect on the students' grade. The research proposes Support Vector Machine (SVM) and Mutual Information (MI). SVM is a powerful classifier (Cristianini & Taylor, 2000). It implements kernel method that can be used to handle non-linear separable data. It has been successfully implemented to predict the performance of faculty member as stated by Deepak *et al*. (2016). Meanwhile, Mutual Information (MI) is useful to measure the relationship between two variables. It can work without affected by the data distribution (Smith, 2015). Some researches related to MI for feature selection can be found in Alzubaidi *et al.* (2016), Gad and Rady (2015), and Li *et al*. (2015).

This research is different from the other related works. It predicts the length of study-time for Computer Science students based on their grades from the first to the fourth semester. In general, student's performance is estimated based on their grade in these semester periods. This research will predict the graduation for each student. The result of this research is necessary for students and academic members, especially for the academic planning. The advantage of this research is this research reveals a list of subjects that contribute more in assigning the length of study-time. Moreover, it reveals the relationship between subjects and its contribution on students' length of study-time.

II. METHODS

This research consists of several main phases. The first phase is feature selection. The researchers apply Mutual Information (MI) to select the appropriate features. After feature selection, the data is divided into training data and testing data randomly. The second phase is predicting the class of length of study-time using Support Vector Machine (SVM). The model is formed in training phase using training data, and the classification uses testing data. In this research, SVM module used is in Scikit-learn (2016). Figure 1 shows the flow chart of this research.

Then, research uses the dataset from database in X University. It is a dataset of Computer Science students in the year of 2008 to 2012. The data consist of 240 alumni and 25 subjects. Table 1 shows the list of 25 subjects in the first to the fourth semester. The subjects are chosen from the mandatory subjects for Computer Science students. The list of subjects is collected by considering a recommendation from the Head of the Computer Science department. The list of subjects are the features, and the length of study-time is the class target. Meanwhile, Table 2 shows the sample of the data. The dataset is the weight of students' grade which is ranged from 0 to 4. Table 3 describes the weight, grade, and the annotation of grades.

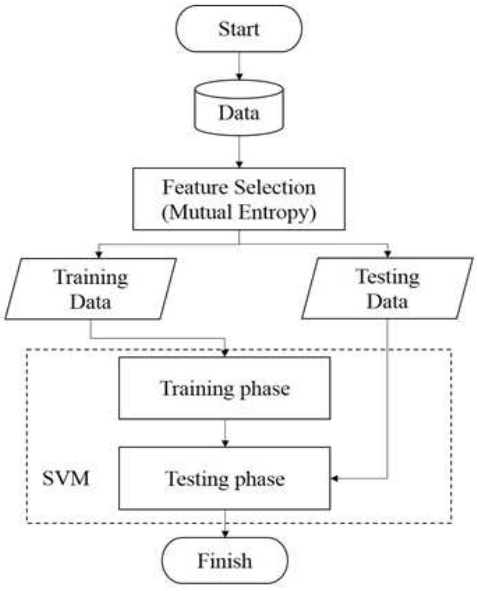

Figure 1 Flow Chart of the Research

Table 1 List of Subjects

No	Code	Subjects	Semester
1	N1	Basic Algorithm	1
2	N2	Calculus I	1
3	N3	Discrete Mathematics	1
4	N4	Management and Computer Organization	1
5	N5	Introduction to Computer	1
6	N6	Logic Information	1
7	N7	Advanced Algorithm	2
8	N8	Information Systems	2
9	N9	Linear Algebra	2
10	N10	Statistics 1	2
11	N11	Digital System	2
12	N12	Operating System	2
13	N13	Human Computer Interaction	2
14	N14	Algorithm Analysis	3
15	N15	Statistics 2	3
16	N16	Physics Mechanics	3
17	N17	Database	3
18	N18	Graph Theory	3
19	N19	Introduction to Artificial Intelligence	3
20	N20	Object Oriented Programming and Java 1	3
21	N21	Differential Equations	4
22	N22	Visual Programming using Visual Basic .Net	4
23	N23	Data Structure	4
24	N24	Computer Network 1	4
25	N25	Physics Electric Wave	4

Table 2 Sample Data

Student ID	S1	S2	S3	S4	S5	Class
ID001	2,45	2,04	2,71	2,21	2,99	1
ID002	2,15	2,32	1,73	2,04	2,53	2
ID003	2,24	2,10	2,50	3,40	4,00	3
ID004	1,71	1,68	2,25	2,25	3,20	3
ID005	2,73	2,07	3,38	2,43	2,53	2
ID006	2,50	3,04	3,13	1,50	3,00	2
ID007	2,85	2,20	3,08	2,50	2,62	2
ID008	3,32	3,04	4,00	3,05	4,00	1
ID009	2,94	2,40	4,00	2,51	3,53	1
ID010	3,28	2,23	4,00	2,84	3,78	1

Table 3 Students' Grade Annotation

No	Score	Grade	Annotation
1	W = 4	A	Excellent
2	3 ≤ W< 4	B	Good
3	2 ≤ W< 3	C	Satisfactory
4	1 ≤ W< 2	D	Fair
5	0 ≤ W< 1	E	Failed

*W is weight of grade

This research implements two different class targets. There are two class targets based on the duration of study, namely two classes and three classes. The detail of class target is explained in Table 4. The length of study-time is measured in year. The classification of two classes is following the rule in X University in determining the 'on time group' and 'late group'. The three classes is a suggestion from the researchers because the late group has longer range of duration, so it might be proper to define the new group. The three classes group represent 'on time group', 'late group', and 'very late group'.

Table 4 Class Target Criteria

Two Classes	Three Classes
Class 1: 3,5 ≤ Duration ≤ 4	*Class 1: 3,5 ≤ Duration≤ 4*
Class 2: 4,5 ≤ Duration ≤ 7	*Class 2: 4,5 ≤ Duration≤ 5*
	Class 3: 5,5 ≤ Duration≤ 7

Mutual Information (MI) measures the relationship between two variables. The high score of MI about variable indicates that those two variables have a close relationship. Meanwhile, the low score describes that there is a weak relationship between them. The Mutual information is computed using equation (1) (Zhang *et al.,* 2012).

$$MI(A,B) = \sum_{a \in A, b \in B} p(a,b) \log \frac{p(a,b)}{p(a)p(b)} \quad (1)$$

MI is implemented to select features which have a close relationship to the length of study-time. The features that have high MI score means that those have a close relationship to the class target. For each pair of feature and the class target, it is measured by MI. The average MI score is used as a threshold. The score between feature and class target which are bigger is chosen as a feature, while the others are removed. Figure 2 shows the algorithm of feature selection based on Mutual Information.

For each features f_i and class c:
    $MI_i = Mutual\ Entropy(f_i, c)$
m = average(MI)
If $MI_i \geq m$
    $SF_i = f_i$
Return SF

Figure 2 Feature Selection Using Mutual Information

Support Vector Machine (SVM) is an algorithm introduced by Vapnik (Cristianini & Taylor, 2000). It can be implemented for classification and regression. SVM uses kernel method to handle the nonlinearly separable data by mapping the data in high dimensional. SVM computes optimum hyperplane separating the dataset in minimum error (Cristianini & Taylor, 2000). The original SVM classifies the data into two classes, +1 and -1. For instance, the N data is $\{[x_1, y_1], \ldots, [x_n, y_n]\}$. The x_i is feature and y_i is class label of x_i. A hyperplane can be described using equation (2) (Liu & Zheng, 2005). If the training data are linearly separable, SVM creates optimal hyperplane to separate the two classes. If the data of two classes are separable, it can be computed using equation (3) (Suykens *et al.,* 2002). On the other hand, if the data is non-linearly separable, it can be computed using equation (4).

$$w^T x + b = 0 \quad (2)$$

$$\begin{cases} w^T x + b \geq 1, if\ y_k = +1 \\ w^T x + b \leq -1, if\ y_k = -1 \end{cases} \quad (3)$$

$$\begin{cases} w^T \varphi(x) + b \geq 1, if\ y_k = +1 \\ w^T \varphi(x) + b \leq -1, if\ y_k = -1 \end{cases} \quad (4)$$

$$\varphi(.): \mathbb{R}^n \rightarrow \mathbb{R}^{n_k}.$$

Figure 3 illustrates the optimum hyperplane on SVM. In Figure 3(a), the data is perfectly separated by a linear hyperplane. Meanwhile, Figure 3(b) describes the kernel method to separate the nonlinearly separable data on SVM.

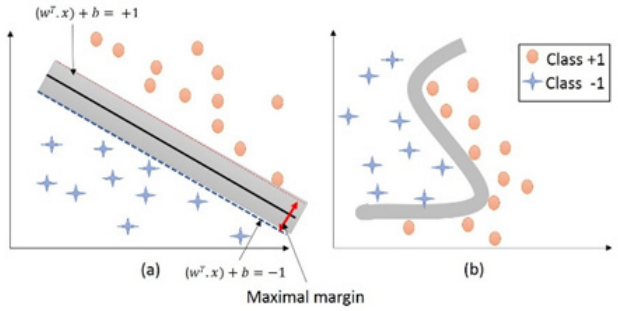

Figure 3 Linear and Non-Linear Hyperplane on SVM

Although originally SVM is only available for classification of two different class targets, it has been developed for classifying more than two class targets or multi-class classification. The common algorithms for multi-class SVM are one-against-all, one-against-one, and Directed Acyclic Graph SVM (Hsu & Lin, 2002).

In one-against-one SVM, it creates $\frac{k(k-1)}{2}$ classifiers, where k is the number of classes. For instance, there are 3 different classes, so the one-against-one SVM creates 6 classifiers. The hyperplane is constructed from two classes that are chosen from k-classes. Table 5 shows the one-against-one SVM model classifier (Liu, Wang, & Zheng, 2007).

Table 5 One-against-One SVM

$y_i = +1$	$y_i = -1$
Class A	Class B
Class A	Class C
Class B	Class A
Class B	Class C
Class C	Class A
Class C	Class B

In SVM one-against-all, there are N data $\{[x_1, y_1], \ldots, [x_n, y_n]\}$ which the x_i is feature and y_i is class label of x_i. The y_i is multi-class, $y_i \in \{1, 2, \ldots, M\}$. The one-against-all SVM creates M binary of SVM classifiers. Each classifier segregates one class from the other classes. The ith of SVM is trained using all training data of the ith class that belongs to positive label, and the others are signed as a negative label as stated by Liu & Zheng (2005). Table 6 shows one-against-all SVM model classifier.

Table 6 SVM One-Against-All

$y_i = +1$	$y_i = -1$
Class A	NonClass A
Class B	NonClass B
Class C	Non Class C
...	...
Class m	NonClass m

III. RESULTS AND DISCUSSIONS

In the feature selection step, the researchers use MI to select the features which have a strong relationship to the length of study-time. MI score is computed between each feature and the length of study-time. The high MI score indicates that the feature has a strong relationship to the length of study-time. After computing all MI scores for 25 features, the average MI score is 0,24. In this research, the feature selection method is choosing the features which have MI score ≥ 0,24. The outcome of the feature selection phase is 12 subjects that can be seen in Table 7. In Table 7, Discrete Mathematics has the highest MI score. It describes that there is the strongest relationship between Discrete Mathematics and to the length study-time.

Table 7 Feature Selection Result

No	Code	Subjects	MI Score
1	N3	Discrete Mathematics	0,30
2	N11	Digital System	0,29
3	N7	Advanced Algorithm	0,28
4	N16	Physics Mechanics	0,28
5	N1	Basic Algorithm	0,27
6	N10	Statistics 1	0,27
7	N24	Computer Network 1	0,27
8	N12	Operating System	0,27
9	N19	Introduction Artificial Intelligence	0,26
10	N20	Object Oriented Programming and Java 1	0,26
11	N23	Data Structure	0,25
12	N21	Differential Equations	0,24

The researchers conduct two experiments. The first experiment is using two class targets, and the other is implementing three class targets. The class targets are determined in Table 4 based on the length of study-time. The system is developed using the scikit-learn module (Scikit-learn, 2016).

The dataset consists of 240 instances and 25 subjects (features). Table 8 shows the data distribution of each class. There is 69,17% in 'on time group' and 30,83% in 'late group'. The experiments are repeated 50 times to select 70% training data and 30% testing data. The data selection considers the distribution of each class for fairness reason. In each experiment, the researchers choose the training data and testing data randomly. The training and testing data are chosen once for each experiment for the fairness. It means that the experiment before and after feature selection use the same training and testing data. This technique is applied to all algorithms.

Table 8 Data Distribution

No	Length Study-Time (year)	The Number of Instances
1	3,5	45
2	4,0	121
3	4,5	44
4	5,0	11
5	5,5	10
6	6,0	7
7	6,5	0
8	7,0	2

The detail of the first experiment result is shown in Table 9. The decision tree and Gaussian Naïve Bayes are used to compare the performance of SVM. The experiment result shows that feature selection using MI only improve the accuracy slightly. It happens to SVM, decision tree, and Gaussian Naïve Bayes. The highest increasing accuracy is reached by SVM. It is around 2%. SVM shows the best accuracy among decision tree and Gaussian Naïve Bayes.

Table 9 Experiment Result of Two Classes

No	Methods	Before Feature Selection		After Feature Selection	
		Avg.Acc.	Std.	Avg.Acc.	Std.
1	SVM Linear Kernel	83,64%	0,04	85,72%	0,04
2	Decision Tree	79,39%	0,04	80,97%	0,04
3	Gaussian Naïve Bayes	84,33%	0,04	85,03%	0,04

The second experiment uses three class targets that are defined by the researchers. In this experiment, SVM multi-class classification used is from scikit-learn. There are two methods, namely one vs one SVM and one vs rest SVM. Both methods use linear kernel. The experiment result shows that there is a slight rising accuracy after applying feature selection method. After feature selection phase, the accuracy of SVM increases about 3%. On the other hand, the accuracy of decision tree and Gaussian Naïve Bayes only rise 0,33%. Both the first and second experiments produce small deviation standard of accuracy. The small deviation standard shows that the accuracy of each experiment remains stable. Table 10 shows the result of the second experiments.

Table 10 Experiment Result of Three Classes

No	Methods	Before Feature Selection		After Feature Selection	
		Avg. Acc.	Std.	Avg.Acc.	Std.
1	SVM One Vs One	77,2%	0,05	80,58%	0,03
2	SVM One Vs Rest	77,2%	0,04	80,82%	0,04
3	Decision Tree	76,41%	0,05	76,74%	0,04
4	Gaussian Naïve Bayes	78,88%	0,05	79,21%	0,05

In the first and second experiments, the researchers prefer to use the linear kernel in SVM. It is because the linear kernel is the easiest kernel method. It does not require tuning parameter kernel that needs further research. In the first and second experiments, the performance of SVM reaches the best accuracy than the other methods. It might be caused by the dataset which is nonlinearly separable data. SVM works by mapping the dataset into feature vectors in high dimensional, so the data which are impossible to be separated in input space can be classified properly in there.

To analyze the relationship among subjects, the researchers compute MI score between them. The average of MI score is around 0,7. It shows that some of the subjects have a strong relationship with others. The network containing the subjects has MI score $\geq 0,8$. The interesting subjects are the subjects which have degree ≥ 3 in the network. Figure 4 shows that Advanced Algorithm has the highest degree in the network. It means that Advanced Algorithm influences some subjects such as Data Structure, Database, Physics Mechanics, Physics Electric Wave, Introduction to Artificial Intelligence, Differential Equations and Object Oriented Programming and Java 1. Advanced Algorithm is also affected by Basic Algorithm, Management and Computer Organization, and Introduction to Computer. Moreover, Introduction to Artificial Intelligence has the second highest degree in the network. It has a close relationship with Management and Computer Organization, Introduction to Computer, Advanced Algorithm and Object Oriented Programming and Java 1. Figure 4 also shows the Mutual Information network of several subjects

The highest MI score is 0,85 which computed by Advanced Algorithm and Object Oriented Programming, and Java 1. On the other hand, the MI score between Basic Algorithm and Advanced Algorithm is 0,83. In Figure 5, there is a direct relationship between Basic Algorithm and Advanced Algorithm, and Advanced Algorithm, Object Oriented Programming and Java 1. It also shows an indirect relationship of Basic Algorithm, Object Oriented Programming and Java 1. It might be affected by the rule of the Computer Science department in X University. Based on this rule, the students are allowed to enroll in Advanced Algorithm class after they have successfully passed Basic Algorithm. Furthermore, the students must get minimum grade C in Advance Algorithm if they want to enroll in Object Oriented Programming and Java 1.

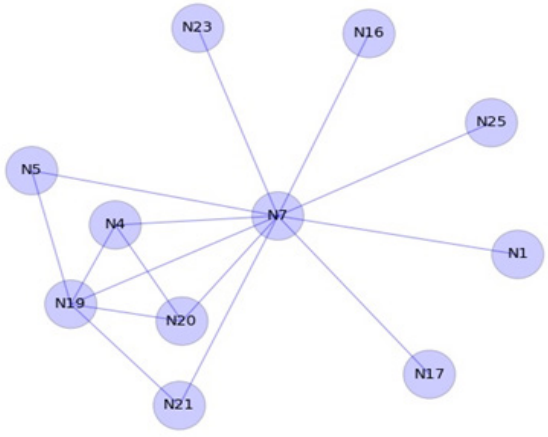

Figure 4 Mutual Information Network among Selected Subjects

Figure 5 shows the scatter plot of the selected subject after feature selection. It shows the relationship between the weight of the grade and length of study-time. The C1, C2, and C3 are the group of Class 1, Class 2 and Class 3 which are explained in Table 4. The data are taken from the first outcome of the academic report, so some students have the weight of grade less than 2,0. It is the standard to pass the subjects. The students who have eight grades which are less than 2,0 have to enroll the subject in the other next semester. Re-enrolling for the same subject in the next semester usually affects the length of study-time. In Figure 5, the group of students who graduate on time (less than or equal to 4 years) mostly have a high weight of grade. On the other hand, the group of students who graduate more than 4 years, some of them have the weight of grade less than 2,0. It means that they need longer time to finish the study.

In addition, the researchers analyze the grade of subjects to find more information. Statistics 1, Physics Mechanic, and Advanced Algorithm produce the percentage of small grades more than 20%. While, Differential Equation, Introduction to Artificial Intelligence and Object Oriented Programming and Java 1 have percentage of grade ≤ 2,0 around 10% to 16%. By reaching the minimum requirement grade in those subjects in early semesters, the students can graduate on time. However, the lower grade in particular subjects might be caused by the content of the courses, the instructors' performance, and the background of the students. Table 11 shows the list of subjects which have the percentage of the weight of grade ≤ 2,0.

Moreover, feature selection based on the MI only increases the accuracy slightly. In Table 7, MI score of each feature and the length of study-time is less than 0,5. It means that the features do not have a strong relationship with the length of study-time. In fact, the length of study-time is not only affected by the subjects from the first to the fourth semester but also the other subjects in other semesters. There are non-academics factors that contribute to the length of study-time. There are personal identification,

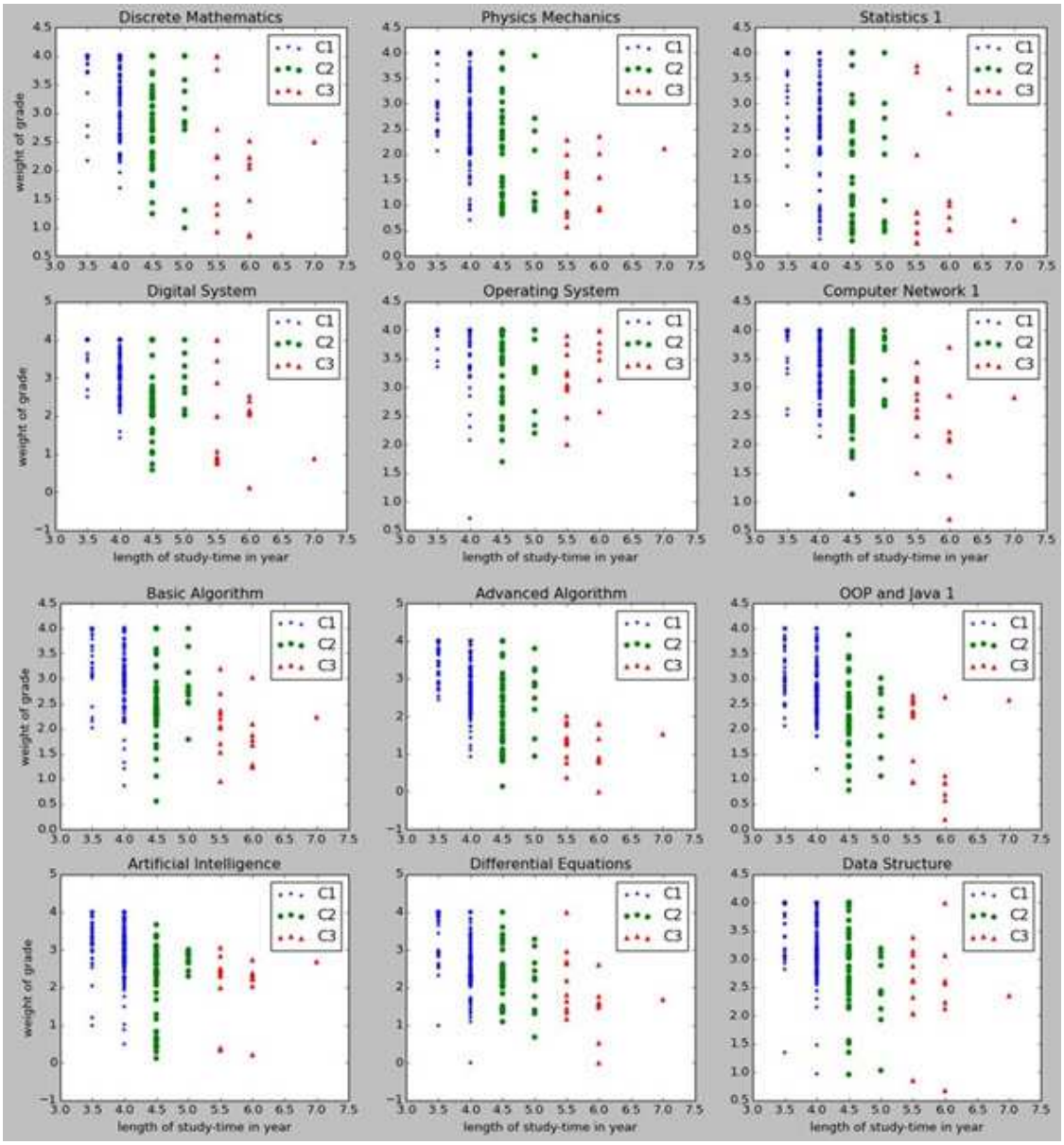

Figure 5 Scatter Plot of Selected Features

students' background, demographic, and psychology. Those are excluded in this research, but they are important information for the students.

Table 11 List of Subject
with Weight of grade less than 2,0

Group 1	Group 2	Subject
8,33%	17,08%	Statistics 1
6,25%	17,50%	Physics Mechanics
3,75%	16,67%	Advanced Algorithm
5,83%	10,83%	Differential Equations
3,33%	8,75%	Introduction Artificial Intelligence
1,67%	10,00%	Object Oriented Programming and Java 1
2,08%	6,25%	Basic Algorithm
0,83%	6,67%	Digital System
0,83%	5,42%	Discrete Mathematics
1,25%	3,33%	Data Structure
0,00%	2,92%	Computer Network 1
0,42%	0,42%	Operating System

IV. CONCLUSIONS

This research is for predicting and analyzing the length of study-time for the Computer Science students in X University. The researchers use the dataset of the weight of grade of the particular subjects and the length of study-time (in year). In this research, the researchers implement Mutual Information (MI) to select the proper subjects that have high contribution in the length of study-time and Support Vector Machine (SVM) to predict the length of study-time. The outcome of feature selection process is 12 subjects.

The experiments are done in two sections. In the first experiment, the researchers use two class targets to predict the length of study-time. The performance of SVM produces 83,64% of accuracy. After feature selection, the accuracy of proposed method reaches 85,72%. In the second experiment, the researchers propose three class targets. The accuracy of SVM is around 77% and 80% for before and after feature selection respectively. The performance of the proposed method is higher than the performance of decision tree and Gaussian Naïve Bayes.

Feature selection using MI is successfully implemented to select the subjects which have a close relationship to the class target. It is also can be used to detect the list of subjects that contribute more to the length of study-time. In the future research, it is necessary to include the non-academic factors that might determine the length study-time. Furthermore, it is necessary to conduct further study to analyze the main problem that causes the lower grade in particular subjects.

REFERENCES

Alzubaidi, A., Cosma, G., Brown, D., & Pockley, A. G. (2016). Breast cancer diagnosis using a hybrid genetic algorithm for feature selection based on mutual information. In *2016 International Conference on Interactive Technologies and Games (ITAG)* (pp. 70-76). IEEE.

Bo, G., Rui, Z., Guang, X., Chuangming, S., & Li, Y. (2015). Predicting students performance in educational data mining. In *2015 International Symposium on Educational Technology (ISET)* (pp. 125-128). IEEE.

Cristianini, N., & Taylor, J. (2000). *An introduction to Support Vector Machines and other kernel-based learning methods.* New York: Cambridge University Press.

Deepak, E., Pooja, G. S., Jyothi, R. N., Kumar, S. V., & Kishore, K. V. (2016). SVM kernel based predictive analytics on faculty performance evaluation. In *2016 International Conference on Inventive Computation Technologies (ICICT)* (pp. 1-4). IEEE.

Dirjen Belmawa. (2016). *Direktorat jenderal pembelajaran dan kemahasiswaaan Kemristekdikti*. Retrieved February 22nd, 2017, from http://belmawa.ristekdikti.go.id/2016/03/04/kemristekdikti-sosialisasikan-permen-nomor-44-tahun-2015-tentang-sn-dikti/

Gad, W., & Rady, S. (2015). Email filtering based on supervised learning and mutual information feature selection. In *2015 Tenth International Conference on Computer Engineering & Systems (ICCES)* (pp. 147-152). IEEE.

Harwati, Alfiani, A. P., & Wulandari, F. A. (2014). Mapping student's performance based on data mining approach. In *The 2014 International Conference on Agro-industry (ICoA): Competitive and Sustainable Agroindustry for Human Welfare* (pp. 173-177). ELSEVIER.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 415-425.

Li, Y., Ma, X., & Yang, M. (2015). Improved feature selection based on normalized mutual information. In *2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)* (pp. 518-522). IEEE.

Liu, W. X., & Cheng, C. H. (2016). A hybrid method based on MLFS approach to analyze students' academic achievement. In *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 1625-1630). IEEE.

Liu, Y., & Zheng, Y. F. (2005). One-against-all multi-class SVM classification using reliability measures. In *International Joint Conference on Neural Networks* (pp. 849-854). Montreal: IEEE.

Liu, Y., Wang, Rui, & Zheng, Y. S. (2007). An improvement of one-against-one method for multi-class Support Vector Machine. In *Sixth International Conference on Machine Learning and Cybernetics* (pp. 2915-2920). Hongkong: IEEE.

Mouri, K., Okubo, F., Shimada, A., & Ogata, H. (2016). Bayesian network for predicting students' final grade using e-book logs in university education. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)* (pp. 85-89). IEEE.

Ogunde, A. O., & Ajibade, D. A. (2014). A data mining system for predicting university students' graduation grades using ID3 decision tree algorithm. *Journal*

of Computer Science and Information Technology, *2*(1), 21-46.

Piad, K. C., Dumlao, M., Ballera, M. A., & Ambat, S. C. (2016). Predicting IT employability using data mining techniques. In *2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC)* (pp. 26-30). IEEE.

Scikit-Learn. (2016). *Scikit-Learn*. Retrieved December 10th, 2016 from http://scikit-learn.org/stable/

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting student's performance using data mining techniques. In *The Third Information Systems International Conference* (pp. 414–422). Procedia Computer Science.

Smith, R. (2015). A mutual information approach to calculating nonlinearity. *Stat*, *4*(1), 291-303.

Suykens, J. A., Gestel, T. V., Brabanter, J. D., Moor, B. D., & Vandewalle, J. (2002). *Least Square Support Vector Machines*. London: World Scientific

Taruna, S., & Pandey, M. (2014). An empirical analysis of classification techniques for predicting academic performance. In *2014 IEEE International Advance Computing Conference (IACC)* (pp. 523-528). IEEE.

Ukpong, D. E., & George, I. N. (2013). Length of study-time behaviour and academic achievement in social studies education students in the University of Uyo. *International Education Studies*, *6*(3), 172.

Zhang, X., Zhao, X. M., He, K., Lu, L., Cao, Y., Liu, J., ... & Chen, L. (2012). Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics*, *28*(1), 98-104.