ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse48952020.pdf
https://doi.org/10.30534/ijatcse/2020/48952020

Classification of the Final Project Utilized a Modified Naïve Bayes Algorithm

Terttiaavini[1], Tedy Setiawan Saputra[2], Anisa Fitriani[3]
[1]Universitas Indo Global Mandiri, South Sumatera - Indonesia, avini.saputra@uigm.ac.id
[2]Universitas Indo Global Mandiri, South Sumatera - Indonesia, tdyfaith@gmail.com
[3]Universitas Indo Global Mandiri, South Sumatera - Indonesia, annisafitriani@gmail.com

ABSTRACT

The student is a part of the national aspiration which will be lead to future advancement. The student should has high integrity as a reflection of the academic ability which thereafter pursued their study at university. The science absorption during study measurable base on the academic grade. To proved how science has been accepted by the student, it could be measured from the student thesis. However, the relationship between the academic grade and the student research topic is mostly irrelevant. Mastery of science and research themes has no correlation, therefore the student's ability to accomplished the research was not properly. In order to map the research field, the coursework and student research field interest can be done by grade classification. The data classifier technique can use the simple Naive Bayes method with fairly accurate output. It could be an intake for the student to choose the research topic which suits their academic grade.

Key words : Students, Thesis, Naïve Bayes, Classification.

1. INTRODUCTION

The pandemic COVID 19 has been damage the economic sector and education as well. University is a center of intellectualism development for nation generation sustainability [1] and it should become attention to improve and enhance the quality in all aspects.

The enhancement of the university quality can be done by improving the student creativity, skill and team workability [2], performance lecturer enhancement [3] and information technology enhancement [4], time discipline and the parent support [5].

The student intellectual could be measured base on the academic achievement that affected by the discipline and support from campus. The student career development can be set up before by the student successful forecasting during the study period. Some research methods could be applied for forecasting the student study successful, such as Tree Decision Method [6] [7], Deep Learning method [8][9] , Neural network [10][11] [12][13], Expert System [14], Symbiotic structure learning algorithm and feed-forward neural-network Method [15], Best-Worst Method [9], etc. It shows how interested the researcher to create a model for student career development.Thus methods are a part of the data mining method. Data mining is a statistic technic that processing the big data by regression, association rule, clustering, classification, forecasting and sequence analysis in order to extract the data become valuable information.

The study object is the last semester student of the Computer Science school at Indo Global Mandiri University, South Sumatera Indonesia. The current problem which occurs is the percentage of the student who can accomplish the final assignment on schedule with grade B is 53%, and the student can graduate out of schedule is 47%. Grade A or B for the thesis will be awarded for the student who can settle the problem by the right problem-solving.

One of the ways to identify science acceptance during the study period is by doing an assessment of the student thesis. The ability to solved the problem scientifically is affected by experience and the mastery of the knowledge. The student's ability to stimulate and also support from campus can help the student to accomplished the thesis properly [16].

One of the universities effort to assist the student to accomplish their thesis is to give them the input to choose a research topic which suits to their talent and academic achievement [14]. Normally, the student determined the research topic base on the trend, lecturer instruction, and the school mate suggest as well. But sometimes, it does not works, since the student does not well understand the theoretics.

Therefore, there is needed a model to determined the research topic so that the student can choose the right research topic base on the academic achievement. To create a prediction model, one of the methods utilized is the classification technique. some classification methods at data mining are Naïve Bayes Classifier, Decision tree, Neural Networks, and Support Vector Machine [17].

In this study, the classification method utilized is Naïve Bayes Method. The Naïve Bayes Method often utilized to predict the student academic activity [17][18][19][20][21][22][23] dan environment social [24].

Generally, classification on the naïve Bayes method resulting in 2 targets, anyway in the adaptation on the case study in the real condition, it could be developed and resulting in 3 targets [25]. Based on the results of research comparing the performance of the Naïve Bayes Method with other methods, it shows that the performance of the Naïve Bayes Method is much better [26] [27] [28]
Besides that, the naïve Bayes method is already replicated in any disciplined study [29] to settled the classification data problem. And also, the calculation formula of the probability value is applicable to the more accurate result.
This study is expected can be given an input to the student of the computer science school at Indo Global Mandiri University as an intake to take the best research topic in order to minimalize student lateness in the accomplished thesis.

2. METHODOLOGY

2.1 Research Approach

This study is feedback from some research on the same topic and method. Step to complete this research shown on the Figure 1.

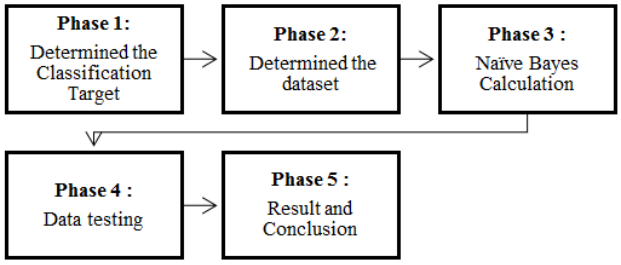

Figure 1 : Research Approach

2.2 Naive Bayes method

Naïve Bayes is a classification method by calculating the probability of each level. Naïve Bayes was famous by a British scientist is Thomas Bayes.
The Naive Bayes algorithm predicts future events based on previous events.Bayes naive algorithm produces classification accuracy values. This algorithm is quite simple but produces strong (naive) assumptions with an incidence of incidence [30]
Naïve Bayes method, generally results in 2 targets [25], anyway, for this study, it has 4 targets
Using the formula that is

$$P(H|X) = \frac{P(X|H)\ P(H)}{P(X)} \quad \text{.....}$$

Where can it be explained that : X = unknown class data; H = data hypothesis X is a specific class; P(H) = Probability Hypothesis H (previous probability); P(X) = Probability of X; P (H |X) = Probability of hypothesis H based on X conditions (posterior probability); P (X |H) = Probability of X based on H conditions.

2.3. Relationship between coursework and research theme

Base on the student mark sheet, and so the student choose the coursework which supported the research theme. The research theme determined base on the curriculum at computer science school. Identification of the coursework mastery, if the coursework gets grade B. The correlation between coursework and research theme, can be explained on the table 1.

Table 1 : The relationship between courses and research themes

coursework	research theme
Data mining, Decision supporting system	Smart System
Computer Graphic, Games programming, Smart games	Computer graphics and animation
Database, Interface programming, Knowledge Engineering, Enterprise system	Database
Computer network, Network Programming, Network wireless, Machine Education	Networking

2.4. The relationship between the research theme and the research topic

In the process of proposing a thesis title, students must determine the research theme of their research interest. Choosers of themes related to the research topic.The relationship between research themes and research topics can be explained in the table 2.

Table 2 : The relationship between the research theme and the research topic

research theme	research topic
Smart System	Application of algorithms, Case-based Decision Support Systems, Artificial Intelligence
Computer Graphic and animation	Games
Database	System base on knowledge, warehouse, Enterprise System
Networking	System Security, Monitoring and remote

3. RESULT DAN ANALYSIS

In this phase, explain how the algorithm naïve bayes work on the dataset. In line to the research approach, each step explained as below:

Phase 1: Pre-processing
This study utilized 150 student records of the computer science school between 2017 and 2019. Each record was taken 13 point data. The point data comes from 13

coursework which explained on table 1, they are a Games programming (X1), computer network (X2), Data mining (X3), Interface programming (X4), Database (X5), Computer Graphic (X6), Network Programming (X7), Machine Education (X8), Knowledge Engineering (X9), Smart Games (X10), Network wireless (X11), Decision supporting system (X12), Enterprise system (X13).

All data which consisted of grade A, B, C, and E. all data are transformed into numeric data with criteria A = 4, B = 3, C = 2, E = 1. If the student did not take the coursework because it's the optional coursework, given value 1. This step is also known as preparation data. In this study data which show only 20 data. The result transformed the data described in table 3.

Table 3 : The transformation of 13 student grade data

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
1	3	3	2	3	4	4	4	4	3	4	3	4	1
2	3	2	3	3	3	3	1	3	3	3	3	4	4
3	4	3	3	3	3	3	1	3	3	3	3	4	4
4	3	2	3	3	3	3	1	3	3	1	3	4	3
5	4	4	2	4	4	3	4	3	3	4	3	4	1
6	4	4	2	4	4	3	4	3	2	3	3	2	1
7	4	4	3	4	3	2	1	3	2	3	3	3	4
8	3	3	4	3	3	4	4	3	2	3	3	3	3
9	3	4	2	4	3	3	4	3	2	3	3	2	4
10	2	3	3	4	3	2	4	3	2	3	2	2	3
11	4	4	3	4	4	4	4	3	3	3	3	3	1
12	4	4	4	3	4	3	4	3	3	3	2	3	1
13	3	4	3	4	4	3	4	3	3	3	3	3	3
14	3	4	2	4	4	4	4	2	4	4	2	2	1
15	3	4	2	4	3	3	4	3	2	2	3	3	1
16	3	4	3	4	4	3	4	3	2	2	3	3	1
17	3	4	2	3	2	2	3	3	3	3	3	3	2
18	3	4	3	4	4	2	1	4	3	4	4	4	4
19	3	3	2	4	3	3	3	2	3	4	3	3	1
20	3	3	3	4	3	4	1	4	3	4	3	4	3

The next step, early classification process for each record. It was aimed to grouped the database on a certain research topic. The early classification is shown in table 4.

Table 4.: Initial classification for datasets

No	Initial Classification	No	Initial Classification
1	c4	11	c1
2	c3	12	c2
3	c3	13	c3
4	c3	14	c2
5	c4	15	c2
6	c4	16	c4
7	c2	17	c2
8	c2	18	c3
9	c2	19	c4
10	c2	20	c1

Classification Code c1 = Smart System, c2 = Computer, Graphic and Animation, c3 = database, c4 = networking.

Phase 2: Naive Bayes Calculation

The Naïve Bayes calculation is calculated P(H), P(H|X), the probability value for each class with utilized the naïve Bayes formula, as below:

a) Probability calculation each class P (H)

P(H = c1) = 12/120 = 0,10
P(H = c2) = 48/120 = 0,40
P(H = c3) = 32/140 = 0,27
P(H = c4) = 28/120 = 0,23

b) Determination of the probability of each subject variable P(X|H)

The calculation for each variable in each course uses the Naive Bayes formula. In this case, the total of courses = 13, then the calculation is done 13 times for each class. The results are shown on table 5.

Table 5 : Probability determination for each variable of the coursework P(X|H)

	P(c1)	P(c2)	P(c3)	P(c4)
X1 = 4	0,33	0,17	0,25	0,29
X1 = 3	0,33	0,75	0,75	0,71
X1 = 2	0,33	0,08	0,00	0,00
X1 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X2 = 4	0,33	0,50	0,25	0,43
X2 = 3	0,67	0,50	0,50	0,57
X2 = 2	0,00	0,00	0,25	0,00
X2 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X3 = 4	0,00	0,17	0,00	0,00
X3 = 3	1,00	0,42	0,75	0,14
X3 = 2	0,00	0,42	0,25	0,86
X3 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X4 = 4	0,33	0,58	0,50	0,29
X4 = 3	0,33	0,33	0,38	0,71
X4 = 2	0,33	0,08	0,13	0,00
X4 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X5 = 4	0,33	0,17	0,38	0,57
X5 = 3	0,33	0,75	0,63	0,43
X5 = 2	0,33	0,08	0,00	0,00
X5 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X6 = 4	0,67	0,33	0,00	0,14
X6 = 3	0,00	0,42	0,75	0,86
X6 = 2	0,33	0,25	0,25	0,00
X6 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X9 = 4	0,00	0,08	0,25	0,00
X9 = 3	0,67	0,50	0,63	0,71
X9 = 2	0,33	0,42	0,13	0,29
X9 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X10 = 4	0,33	0,08	0,25	0,57
X10 = 3	0,67	0,83	0,50	0,29
X10 = 2	0,00	0,08	0,13	0,14
X10 = 1	0,00	0,00	0,13	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X11 = 4	0,00	0,00	0,13	0,00
X11 = 3	1,00	0,75	0,88	1,00
X11 = 2	0,00	0,25	0,00	0,00
X11 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X12 = 4	0,67	0,08	0,63	0,29
X12 = 3	0,33	0,50	0,38	0,57
X12 = 2	0,00	0,42	0,00	0,14
X12 = 1	0,00	0,00	0,00	0,00

	P(c1)	P(c2)	P(c3)	P(c4)
X13 = 4	0,00	0,17	0,50	0,00
X13 = 3	0,33	0,17	0,50	0,00
X13 = 2	0,33	0,08	0,00	0,00
X13 = 1	0,33	0,58	0,00	1,00

Based on the results of these calculations, the probability value for each class can be used for data testing.

Phase 4: Data Testing

Data testing was carried out using 30 sample data that had been prepared. The purpose of Data testing is to ascertain whether the initial classification is accurate. Data testing is done by multiplying the value of the course variable with the value of the probability. The following is an explanation of four sample Data testing as an example below:

Data uji 1:
X1 = 4; X2 = 4; X3 = 2; X4 = 4; X5 = 4; X6 =3 ; X7 = 4;
X8 = 3; X9 =3; X10 = 4; X11= 3; X12= 4; X13 = 1; k = c4

Data uji 2 :
X1 = 2; X2 = 3; X3 = 3; X4 = 4; X5 = 3; X6 =2 ; X7 = 4;
X8 = 3; X9 =2; X10 = 3; X11= 2; X12= 2; X13 = 3; k = c2
Data uji 3 :
X1 = 3; X2 = 3; X3 = 3; X4 = 4; X5 = 3; X6 =4 ; X7 = 1;
X8 = 4; X9 =3; X10 = 4; X11= 3; X12= 4; X13 = 3; k = c1
Data uji 4 :
X1 = 4; X2 = 3; X3 = 3; X4 = 4; X5 = 4; X6 =1 ; X7 = 1;
X8 = 4; X9 =4; X10 = 4; X11= 3; X12= 3; X13 = 4; k = c3

The calculation is aim to get a higher classification value. Each data examined resulted in value for each class. The higher class value is identified as the research topic. The table 6 describes the data examination result.

Tabel 6: Data examination

	Test data			
	1	**2**	**3**	**4**
c1	22,33	17,00	22,42	15,67
c2	17,67	17,67	16,25	13,17
c3	20,13	14,13	21,25	20,13
c4	24,14	10,00	17,00	17,43

The results of these calculations are obtained: test data 1 has the highest value at c4 with a value of 24.14; test data 2 the highest score is c2 with a value of 17.67; test data 3 the highest value is c1 with a value of 22.42 and test data 4 the highest value is c3 with a value of 20.13

Base on the higher value identification can determine the research topic. The research topic is compared with the initial classification. If there has the same result, that means the initial classification is accurate. The same step is implemented on all data examination. The pattern of the examination result can use to make a prediction of the student research topic base on the academic grade.

4. CONCLUSION

The results prove that the naïve Bayes method can be applied to produce a pattern of selecting research topics based on 4 choices, namely intelligent systems, computer graphics and animation, databases, and networks. A further study of the results of this study is to create a pattern in the form of a matrix so that it can be implemented into a programming language.

REFERENCES

[1] E. de Matos Pedro, J. Leitão, and H. Alves, "Bridging intellectual capital, sustainable development and quality of life in higher education institutions," *Sustain.*, vol. 12, no. 2, 2020.

[2] E. Bryndin, "Creative Innovative Higher Education of Researchers with Flexible Skills and Synergy of Cooperation," *Contemp. Res. Educ. English Lang. Teach.*, vol. 1, no. 1, pp. 1–6, 2019.

[3] T. Terttiaavini and R. Wiryasaputra, "Pengembang Sistem Informasi Kinerja Dosen berbasis WEB dalam upaya meningkatkan Kompetensi Dosen di Universitas Indo Global Mandiri," *Inform. Glob.*, vol. 4, no. 2, pp. 42–53, 2013.

[4] S. Newell, L. F. Edelman, D. S. Staples, J. Webster, and O. Henfridsson, "Changes in the Information Technology Field: A Survey of Current Technologies and Future Importance," *J. Inf. Syst. Appl. Res.*, vol. 12, no. 3, pp. 2008–2008, 2008.

[5] Sana, I. F. Siddiqui, and Q. A. Arain, "Analyzing Students' Academic Performance through Educational Data Mining," *3C Tecnol. Glosas innovación Apl. a la pyme*, no. May, pp. 402–421, 2019.

[6] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, pp. 177–194, 2017.

[7] M. Wasil, A. Sudianto, and Fathurrahman, "Application of the Decision Tree Method to Predict Student Achievement Viewed from Final Semester Values," in *Journal of Physics: Conference Series*, 2020, vol. 1539, no. 1, pp. 1–6.

[8] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Transfer learning from deep neural networks for predicting student performance," *Appl. Sci.*, vol. 10, no. 6, pp. 1–12, 2020.

[9] W. P. Shoong, S. A. Asmai, and T. C. Chuan, "Stock prices time series forecasting by deep learning using three-point moving gradient," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 6622–6630, 2020.

[10] S. Qu, K. Li, S. Zhang, and Y. Wang, "Predicting Achievement of Students in Smart Campus," in *IEEE Access*, 2018, vol. 6, pp. 60264–60273.

[11] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," *SN Appl. Sci.*, vol. 1, no. 9, 2019.

[12] A. Yağci and M. Çevik, "Prediction of academic achievements of vocational and technical high school (VTS) students in science courses through artificial neural networks (comparison of Turkey and Malaysia)," *Educ. Inf. Technol.*, vol. 24, no. 5, pp. 2741–2761, 2019.

[13] A. Heryati, Erduandi, and Terttiaavini, "Penerapan Jaringan Saraf Tiruan Untuk Memprediksi Pencapaian Prestasi Mahasiswa," in *Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018*, 2018, pp. 8–9.

[14] Terttiaavini, *Potensial Ability*, 1st ed. Palembang: NoerFikri Offset, 2017.

[15] A. Dinesh Kumar, R. Pandi Selvam, and K. Sathesh Kumar, "Review on prediction algorithms in educational data mining," *Int. J. Pure Appl. Math.*, vol. 118, no. 8, pp. 531–537, 2018.

[16] C. L. Mbato and A. Cendra, "Efl undergraduate students ‟ self -regulation in thesis writing : help-seeking and motivation-regulation," *J. English Lang. Educ.*, vol. 5, no. 1, pp. 66–82, 2019.

[17] S. Umadevi and K. S. J. Marseline, "A survey on data mining classification algorithms," in *Proceedings of IEEE International Conference on Signal Processing and Communication, ICSPC 2017*, 2018, no. July, pp. 264–268.

[18] M. Makhtar, H. Nawang, and S. N. W. Shamsuddin, "Analysis on Students Performance Using Naïve," *J. Theor. Appl. Inf. Technol.*, vol. 31, no. 16, pp. 3993–4000, 2017.

[19] E. Sugiharti, S. Firmansyah, and F. R. Devi, "Predictive evaluation of performance of computer science students of unnes using data mining based on naÏve bayes classifier (NBC) algorithm," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 4, pp. 902–911, 2017.

[20] Haviluddin, N. Dengen, E. Budiman, M. Wati, and U. Hairah, "Student Academic Evaluation using Naïve Bayes Classifier Algorithm," in *Proceedings - 2nd East Indonesia Conference on Computer and Information Technology: Internet of Things for Industry, EIConCIT 2018*, 2018, pp. 104–107.

[21] F. Razaque *et al.*, "Using naïve bayes algorithm to students' bachelor academic performances analysis," *4th IEEE Int. Conf. Eng. Technol. Appl. Sci. ICETAS 2017*, vol. 2018-Janua, pp. 1–5, 2018.

[22] J. T. Rosado, A. P. Payne, and C. B. Rebong, "EMineProve: Educational Data Mining for Predicting Performance Improvement Using Classification Method," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 649, no. 1, pp. 1–8.

[23] A. Tripathi, S. Yadav, and R. Rajan, "Naive Bayes Classification Model for the Student Performance Prediction," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2019*, 2019, pp. 1548–1553.

[24] S. S. Athani, S. A. Kodli, M. N. Banavasi, and P. G. S. Hiremath, "Student Academic Performance and Social Behavior Predictor using Data Mining Techniques," in *International Conference on Computing, Communication and Automation (ICCCA2017)*, 2017, pp. 170–174.

[25] F. J. Yang, "An implementation of naive bayes classifier," in *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018*, 2018, pp. 301–306.

[26] I. A. A. Amra and A. Y. A. Maghari, "Students Performance Prediction Using KNN and Naïve Bayesian," in *International Conference on Information Technology (ICIT)*, 2017, pp. 909–913.

[27] M. Wati, W. Indrawan, J. A. Widians, and N. Puspitasari, "Data mining for predicting students' learning result," in *Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017*, 2018, vol. 2018-Janua, pp. 1–4.

[28] N. Yusof, S. A. Rosidi, N. K. Ibrahim, and A. E. B. Ali, "Thematic textual hadith classification: An experiment in rapidminer using support vector machine (SVM) and naïve bayes algorithm," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5967–5972, 2020.

[29] J. Zhou, X. Li, and H. S. Mitri, "Classification of Rockburst in Underground Projects : Comparison of Ten Supervised Learning Methods," *J. Comput. Civ. Eng*, vol. 04016003–1, pp. 1–19, 2016.

[30] Y. An, S. Sun, and S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics," in *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017*, 2017, no. 1, pp. 635–638.