NASKAH PUBLIKASI (*MANUSCRIPT*) IMPLEMENTASI ALGORITMA ROUGH SET DAN NAÏVE BAYES DALAM MEMPREDIKSI INDEKS PRESTASI MAHASISWA BERDASARKAN STRESS AKADEMIK

IMPLEMENTATION OF ROUGH SET AND NAÏVE BAYES FOR PREDICTING GRADE POINT AVARAGE BASED ON ACADEMIC STRESS

Andhika Pratama, Rofilde Hasudungan



DISUSUN OLEH : ANDHIKA PRATAMA 1811102441010

PROGRAM STUDI S1 TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

SAMARINDA

Naskah Publikasi (*Manuscript*) Implementasi Algoritma Rough Set dan Naïve Bayes dalam Memprediksi Indeks Prestasi Mahasiswa Berdasarkan Stress Akademik

Implementation of Rough Set and Naïve Bayes for Predicting Grade Point Avarage Based on Academic Stress



Andhika Pratama, Rofilde Hasudungan

Disusun Oleh : Andhika Pratama 1811102441010

PROGRAM STUDI S1 TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

SAMARINDA

HALAMAN PENGESAHAN

IMPLEMENTASI ALGORITMA ROUGH SET DAN NAÏVE BAYES DALAM MEMPREDIKSI INDEKS PRESTASI MAHASISWA BERDASARKAN STRESS AKADEMIK

NASKAH PUBLIKASI

DISUSUN OLEH:

ANDHIKA PRATAMA

1811102441010

Dosen Pembimbing

Rofilde Hasudungan, S.kom., M.Sc NIDN : 1107048601 Penguji

Naufal Azmi Verdikha, S.kom., M.Eng

NIDN : 1114048801

Dekan ito, MT., Ph.D IDN. 0610116204

Ketua Program Studi

S.Kom., M.Cs N. 1124098902

Implementation of Rough Set and Naïve Bayes Algorithms for Predicting Grade Point Avarage Based on Academic Stress

Andhika Pratama¹, Rofilde Hasudungan^{2*}

^{1,2} Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Jl. Ir. H. Juanda No.15, Sidodadi, Kec.Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124, Indonesia

*Email Corresponding Author: 1811102441010@umkt.ac.id

ABSTRACT

Abstract – Due to a variety of factors that are being dealt with, including academic stress, student capability to perform well is negatively impacted. By making predictions about a student's performance since the beginning of the day, one can assist in predicting the kind of help that will be needed for a student who is expected to have low performance so that they can have high performance the following day. This study will use the academic stress factor to predict performance using the Rough Set, Naive Bayes, and four other lagging parameters. The data used for this study consisted of 113 student data. The results of this research demonstrate that the Rough Set algorithm combined with Naive Bayes successfully predicted with an overall accuracy of 73%, whereas alone applying Naive Bayes successfully predicted with an accuracy level of 70%.

Keywords: Naïve Bayes, Rough Set, prediction of student achievement, Academic stress

1. Introduction

Grade Point Avarage (GPA) is a Avarage credit figure that represents the value of the learning process each semester. It may also be used as a benchmark or indicator of achievement in the master's degree program for one semester [1]. Those with high GPAs are more likely to be able to pay attention in class, and vice versa students with lower GPAs are much more likely to struggle to pay attention in class.

Students who must meet a variety of academic requirements go through academic stress. Scholastic stress is described as a state when a person finds it difficult to handle academic expectations and views them as an inconvenience [2]. Academic stress is triggered by a variety of academic stressors [3]. Academic stressors include things like the pressure to perform well academically, the duration of the study period, the number of assignments, the achievement level, and test anxiety [3].

Several studies have been conducted on academic stress that have already been conducted, such as the one mentioned above [4] using multinominal logistic regression analysis to examine the propensity to get good GPA [5] relating to the human engagement between academic stress and the use of a swift return.

The Rough Set and Nave Bayes algorithms will be employed in this research, and two experiments will be conducted first, excess attributes will be reduced using the Rough Set method, and then previously decreased attributes will be categorized using the Nave Bayes algorithm. Without lowering the data using the Rough Set first, the two direct parameters are categorized using Naive Bayes.

In this study, the Rough Set algorithm, which was developed in the 1980s and is a version of a Fuzzy Set theory, will be employed. A mathematical method is called the Rough Set Algorithm may be applied to handle uncertainty resulting from faulty data. The benefit of Rough Set is the capability to execute database searches despite the database's incompleteness of information [6]. Another benefit of the algorithm is its effectiveness in uncovering hidden data patterns. It

can reduce the size of the data collection, give a thorough evaluation of the data, and create decision-making rules from the data. Additionally, both quantitative and qualitative data may be used with it.

In addition to that, the previous research also employs the Naive Bayes Classifier classification method. Classification by Nave Bayes is a long-running probabilistic classification technique that is designed to be used with the assumption that each of the two class in question is independent. When using Naive Bayes classification, larger data sets will result in less error rates being discovered during training. In addition to that, Naive Bayes classification has been proven to be accurate and fast when performed to a large number of bases of data [7].

Based on recent research, the researchers decided to use Rough Set and Nave Bayes as the subject of their study to predict the index of a master's student's performance based on academic stress at the University of Muhammadiyah Kalimantan Timur, Indonesia, This study is anticipated to assist create a better educational environment by educating academic supervisors and allowing them to track students' temporary grade point average growth. In this study, experiment A, which uses the Rough Set to first reduce the parameters before being categorized using Nave Bayes, will also be conducted. Furthermore, the parameters in experiment B are explicitly categorized using Nave Bayes rather than the Rough Set.

2. Research Method

2.1 Rough Set

Teori Rough Set till this day other approach for unknown reasoning [8]. Similar to how fuzzy set theory is not a substitute for classical set theory but rather a part of it. The concept of fuzziness may be considered as being specifically implemented in rough set theory, where the imprecision is communicated by set boundaries rather than partial membership as it is in fuzzy set theory.

Approximations are topological, interior, and closure processes that may be used to establish the idea of rough set in general.

After data gathering is complete, the goal of rough set research is to create classification methods[9]. Following deductions, the rules have been established in this case.

For instance, if a patient has the flu, their symptoms are the same but they are not visible, therefore according to medical understanding, these symptoms constitute the sickness. These medical skills are referred to as a set or basic set (concept). Each notion may be categorized in terms of experience and understanding concepts since these fundamental concepts can be joined to create new concepts.

The basic set or also known as the firm set (initial set), a set other than the basic set, will be referred to as a rough set or Rough Set (unclear, imprecise). The borders of the area show the difference between both the basic set and the rough set in the basic set, there are components that are outside of that set but may be members of many other sets, whereas in the coarse set, there are elements that are outside of that set but may be members of other sets. Rough Set will define the theory with approximations and the theory without approximations. The two definitions are not similar since in this case the membership function is not the method's basic concept.

2.1.1 Information Table

Rough sets, often known as information systems or information tables, represent data in a tabular format. Information tables will be employed in this research. Data is organized in rows and columns in an information table, where each row is a reference to an item or entity and each column denotes a quality or characteristic of that object. A data table is described as , where is the universality set, is the attribute set, is the domain of attribute a, and is the information table .

Table 1 example information table

Respondents	А	В	С	D	E	Category
R1	Appropriate	Very Suitable	Appropriate	Appropriate	Non- Compliant	Very satisfied
R2	Appropriate	Appropriate	Appropriate	Appropriate	Appropriate	Satisfied
R3	Appropriate	Very Compliant	Non- Compliant	Non- Compliant	Non- Compliant	Unsatisfied
R4	Appropriate	Appropriate	Non- Compliant	Appropriate	Appropriate	Very satisfied
R5	Appropriate	Appropriate	Appropriate	Appropriate	Non- Compliant	Satisfied
R6	Appropriate	Appropriate	Appropriate	Non- Compliant	Appropriate	Unsatisfied

Description :

A = sufficient study spaces.

- B = Environment is very suitable for the learning and growing process.
- C = Secure vehicle parking lot.
- D = The cleanliness of the study room is maintained.
- E = Large parking lot.

2.1.2 Indiscernible Relation

For example is a data table with information, and let be a subdivision of. If and only if, two items are defined B-indiscernible (indistinguishable/equivalent) only if and when . Indiscernible Relation is the process of gathering data based on a relationship with a parameter that is currently being used.

Table 2 Atribut Grouping

Decision/category	Responden
Very satisfied	R1,R4
Satisfied	R2,R5
Less satisfied	R3,R6

2.1.3 Reduction

Utilize the properties found in the information table to create the indiscernibility relation between the items. To create a connection, though, not all of the qualities are required. Because occasionally it's necessary to remove extraneous properties or options. The qualities that have no effect on the relationship will be deleted. Equation (1) has been used to evaluate each attribute's consistency.

$$Y(a,b) \frac{amount of data}{all data}$$
(1)

γ = data consistency

(a,b) = the parameter used

If a parameter or attribute has a level of consistency is 1 and anything below will result in its elimination since it is considered to be a feature that has no bearing on relational data.

2.2 Naïve Bayes

The Naive Bayes method is a technique for doing thorough probabilistic classification by adjusting the frequency and adjusting the numeric values from the available pool of data. Nave Bayes is based on the assumption that when an attribute is given a numerical value in a straightforward manner, the value of the attribute in conventional form will be lower if it is given an output, and the probabilities connected with that value will be calculated jointly. Naive Bayes requires training data to reduce parameter estimates used in classification processes, however this approach consistently has higher performance than anticipated in complex real-world scenarios.

The advantage of the Naive Bayes algorithm is that it just requires a little amount of data, presuming that the presence or absence of certain attributes of one class has no bearing on other classes [10]. The (2) agreement is a result of the Naive Bayes method.

(2)

Description:

P(H|E): Probability that a particular hypothesis occurrence H will occur if a given circumstance E occurs.

P(E|H): Probabilities on item E will change in connection to hypothesis H

P(H): Probability that hipotesis H will occur without any intervening triggers.

P(E): Probability that event E will occur without consideration of other hypotheses or events.

The single most important factor that is used as a k-factor in a classification algorithm is the tingkat akurasi. While a system is being qualified, it is anticipated that it will be capable of classifying every set of data with certainty, but it is not indicated that the accuracy of the program's operation will be 100% [12]. Use of equality (3) is made to reduce the arousal threshold.

$$Accuracy = \frac{the amount of data predicted correctly}{All data}$$
(3)

2.3 Classification Model

As seen in Figure 1, this study will develop a categorization model. In this research model, data processing comes first, then data collection. Additionally, the processed data will be used in two trials, one to remove parameters that are unaffected by the rough set and the other to classify the resulting data that use the Naive Bayes method. The Nave Bayes method was used to instantly classify the data in the second experiment. The two experiments' outcomes will be reassessed and compared to determine which experiment has the highest degree of accuracy.

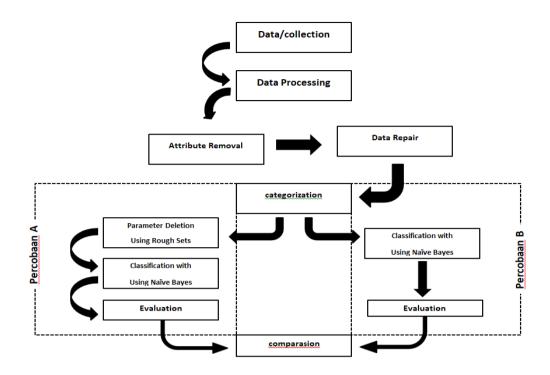


Figure 1 Design Models

3. Result and Discussion

3.1 Data collection

This study retrieved data from Informatics Engineering students from 2019 to 2021, data collection used a questionnaire, then it would be distributed to students in the form of a Google Form, with the attributes used based on research conducted by [1] and [11]. The attributes used and a description of the contents of each attribute are in table 3.

Table 3	8 Attributes	Used
---------	--------------	------

NO	Question	Value/Answer Choice
1	Student Name	Student name
2	Student Identification Number (NIM)	NIM
3	Gender	man
		woman
4	Facilities to conduct adequate learning, both from students and from campus	 very inappropriate Not compliant Appropriate Fits perfectly
5	The campus has a very suitable environment to carry out the learning and teaching process	 very inappropriate Not compliant Appropriate

NO	Question	Value/Answer Choice
		 Fits perfectly
6	The socialization ability of brothers is quite active	 very inappropriate Not compliant Appropriate Fits perfectly
7	Always make a schedule or divide time tailored to the needs of a student	 very inappropriate Not compliant Appropriate Fits perfectly
8	During college, parents always pay attention	 very inappropriate Not compliant Appropriate Fits perfectly
9	I get tired easily when doing college assignments	 very inappropriate Not compliant Appropriate Fits perfectly
No	Question	Value/Answer Choice
10	I feel that the tasks given are very much	 very inappropriate Not compliant Appropriate Fits perfectly
11	I felt that there were too many exams in the lecture to take	 very inappropriate Not compliant Appropriate Fits perfectly
13	Academic achievements and grades are everything to me	 very inappropriate Not compliant Appropriate Fits perfectly
14	I feel guilty to my parents if I get low test scores	 very inappropriate Not compliant Appropriate Fits perfectly
15	I found it very difficult to keep up with the lessons in college	 very inappropriate Not compliant Appropriate Fits perfectly

NO	Question	Value/Answer Choice
17	I feel anxious if I can't meet the expected test score target	 very inappropriate Not compliant Appropriate Fits perfectly
18	I lack confidence due to my grades and academic performance	 very inappropriate Not compliant Appropriate Fits perfectly
19	I feel anxious if I can't meet the expected test score target	 very inappropriate Not compliant Appropriate Fits perfectly
21	I have trouble sleeping when I remember the time for the exam/test is approaching	 very inappropriate Not compliant Appropriate Fits perfectly
22	Last semester's IP	IP obtained last semester
23	GPA	GPA

3.2 Rough set feature selection

Feature selection using the Rough Set algorithm while utilizing text (1). Out of Fifteen feature , around Five feature were succeeded in reducing into to the Ten feature that rough set deemed to be the best feature or attribute. Aspect or feature that was developed by Rough Set is a feature that generally does not impact the validity of the data it contains. Results of the redaction may be shown in Table 3.

Table 4 reduction results

Item	Information
A2	The campus environment is very suitable for carrying out the teaching and learning process
A15	I have trouble sleeping when I remember that the exam / test time is approaching
A6	I get tired easily when doing college assignments
A12	I feel uneasy when I can't meet my expected test score targets

Item	Information
A5	Parents always pay attention during their studies
A11	I find it very difficult to take lessons in lectures
A10	I feel guilty to my parents if I get a low test score.
A8	I feel there are too many exams in college that must be passed.
A4	Always make a schedule / divide time according to the needs of a student
A13	I lack confidence due to grades and academic performance

3.3 Repairing Data

Checking for incomplete data must be done before making any changes to the data, and if any are identified, the incomplete data must be erased.

3.4 Data Categorization

At this stage categorization will be carried out on the GPA value data (Grade Point Average). Respondents will be divided into three parts, the categorization is based on the academic guidebook of the Muhammadiyah University of East Kalimantan, namely the categories Satisfying, Very satisfying, and With praise (cum laude). The following is a description of the categorization contained in the table 5.

Table 5 Categorization

Value	Category
2,76 - 3,00	Satisfying
3,01 – 3,50	Very Satisfying
>3.51	With Praise (laude)

3.5 Rough Set is used to choose the optimal settings.

The next step is to process the selection of the best parameters after all the data has been gathered. The Rough Set method is used for parameter selection in order to deal with parameter ambiguities, accuracy issues, and uncertainty. By removing attribute-specific constraints, such as those that monitor correctness, and by reducing the number of attributes in the data collection.

3.6 Naïve Bayes classification

The categorization procedure for the qualities that the rough set has decreased must now be carried out as the following step. The research model in this study was conducted using Google Coolab, as seen in Figure 1. In this study, data were employed in a 70:30 ratio, with 70% of the data used for training and 30% for testing. The experiment's goal is to ascertain the impact of feature selection performed by the Rough Set algorithm on the classification outcomes performed by the Naive Bayes algorithm in order to compare the findings of the two experiments that have been conducted. The outcomes are displayed in Table 4.

	precision	recall	F1-score
With Praise	0,74	1	0,85
Satisfying	0	0	0
Very Satisfying	0	0	0
Accuracy			0,73

Table 6 Experiment A

Based on Table 1, the accuracy in experiment A by classifying using the naïve bayes algorithm collaborated with the Rough Set yielded an accuracy value of 73%. And in experiment B the data directly classified using Naïve Bayes without the elimination of attributes by the Rough Set resulted in an accuracy value of 70%. As shown in Table 3.

Table 7 Experiment B

	precision	recall	F1-score
With Praise	0,73	0,96	0,83
Satisfying	0	0	0
Very Satisfying	0	0	0
Accuracy			0,70

3.7 Evaluation

Formula (3), which would be part of the theoretical foundation, is used in this evaluation process employing a ratio of 70:30 to ascertain the accuracy outcomes. It was successful in reducing from 15 characteristics to 10 attributes based on experiment A, which was performed utilizing 15 attributes with three distinct kinds of categories in the Rough Set. Also, the characteristics left over after the Rough Set algorithm is gone are the traits that will be categorized using Nave Bayes. After that, conduct experiment B, in which the 70:30 ratio is still used but no prior attribute elimination method was used as in experiment A. The Naive Bayes algorithm will categorize the data right away using 15 attributes and three different sorts of groupings. Accuracy comparison is in.

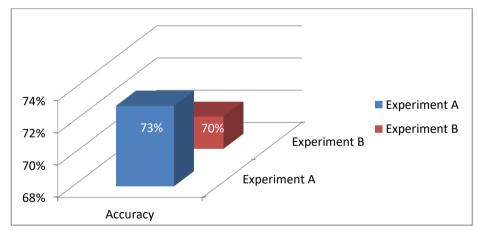


Figure 2 Accuracy

4. Conclusion

4.1 Conclusion

Conduct a study on how students' stress levels impact their educational outcomes. This study examined the impact of students' academic stress on accomplishment as shown by GPA scores. 113 student data with a total of Fifteen parameters were used in this study. The Rough Set and Nave Bayes model was employed in this investigation. It produced findings with an accuracy rate of 73% when features had first been selected by Rough Set and then categorized by Nave Bayes, whereas results that were directly classified without attribute selection by Rough Set were 70% less accurate.

4.2 suggestion

This study makes recommendations for improving accuracy by including some of the proper features or a larger data sample. It is also important to identify the right indicators, thus it is important to pay attention to this process. Knowing the metrics that will influence a student's academic performance will enable decisive action to be done.

5. Reference

- U. S. Karyanus Daely and A. Manurung, "Analisis Statistik Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Mahasiswa," *Saintia Mat.*, vol. 1, no. 5, pp. 483–494, 2013.
- [2] M. Barseli, I. Ifdil, and N. Nikmarijal, "Konsep Stres Akademik Siswa," J. Konseling dan Pendidik., vol. 5, no. 3, pp. 143–148, 2017, doi: 10.29210/119800.
- [3] E. Sayekti, "Efektifitas Teknik Self-Instruction Dalam Mereduksi Stress Akademik Pada Siswa Kelas Xi Ma Yarobi Kec. Grobogan, Kab. Grobogan Tahun 2016/2017," pp. 1–123, 2017, [Online]. Available: http://erepository.perpus.iainsalatiga.ac.id/1634/1/ENGGAR SAYEKTI PAI 111 11 132.pdf.
- [4] A. Fauzi, "Indeks Prestasi Mahasiswa Program Studi," Anal. PENGARUH STRES DAN Motiv. TERHADAP Indeks PRESTASI Mhs. Progr. Stud. Manaj. Univ. RIAU Kepul. Achmad, vol. 1, no. 1, pp. 76–97, 2021.
- [5] S. M. Jannah and J. Empati, "Akademik Pada Mahasiswa Yang Menjalani Pembelajaran Jarak Jauh Akibat Covid-19," vol. 19, no. 2020, pp. 255–259, 2020.
- [6] S. Samaray, "Implementasi Algoritma Rough Set dengan Software Rosetta untuk Prediksi Hasil Belajar," J. Eksplora Inform., vol. 11, no. 1, pp. 57–66, 2022, doi: 10.30864/eksplora.v11i1.498.
- [7] Erene Fajrila, "Perbandingan klasifikasi ketepatan waktu kelulusan mahasiswa menggunakan regresi logistik biner dan naïve bayes classifier," 2018.
- [8] Z. Pawlak, "Rough sets," Int. J. Comput. Inf. Sci., vol. 11, no. 5, pp. 341– 356, 1982, doi: 10.1007/BF01001956.
- [9] W. Maharani, "Analisis Performansi Algoritma Rough Adaptive Neuro-

Fuzzy Inference System," vol. 2008, no. Snati, pp. 39–42, 2008.

- [10] R. Hasudungan and W. J. Pranoto, "Implementasi Teorema Naïve Bayes Pada Prediksi Prestasi Mahasiswa," J. Rekayasa Teknol. ..., vol. 5, no. 1, pp. 10–16, 2021, [Online]. Available: http://ejournals.unmul.ac.id/index.php/INF/article/view/4996.
- [11] A. I. Sihotang, "Adaptasi Alat Ukur Stres Akademik Untuk Remaja," 2021,[Online]. Available: https://repositori.usu.ac.id/handle/123456789/43870.

Naspub: Implementation Of Rough Set And Naïve Bayes Algorithms In Predicting Grade Point Avarage Based On Academic Stress

by Andhika Pratama

Submission date: 04-Apr-2023 09:13AM (UTC+0800) Submission ID: 2055140528 File name: In_Predicting_Grade_Point_Avarage_Based_On_Academic_Stress.docx (116.41K) Word count: 3102 Character count: 15792

Naspub: Implementation Of Rough Set And Naïve Bayes Algorithms In Predicting Grade Point Avarage Based On Academic Stress

ORIGINALITY REPORT			
1% SIMILARITY INDE	1% INTERNET SOURCES	0% PUBLICATIONS	0% STUDENT PAPERS
PRIMARY SOURCES			
		Line data da instituta	
"Inve lightv SUV u	min, Sarjito, Aniq stigation of alumi veight outer hood using finite eleme shing, 2018	num alloy for I panel of local	compact

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		