Implementation of ML Rough Set in Determining Cases of Timely Graduation of Students

Sepyan Purnama Kristanto¹, Reza Syehma Bahtiar², Meriksa Sembiring³, Hidayatulah Himawan⁴, Lukman Samboteng⁵, Hariyadi⁶, I Ketut Suparya⁷

¹Politeknik Negeri Banyuwangi, Indonesia. Email: sepyan@poliwangi.ac.id
²Universitas Wijaya Kusuma Surabaya, Surabaya, Indonesia. Email: syehma_fbs@uwks.ac.id
³Universitas Pembangunan Panca Budi, Medan, Indonesia. Email: meriksa@dosen.pancabudi.ac.id
⁴UPN Veteran Yogyakarta, Indonesia. Email: if.iwan@upnyk.ac.id
⁵STIA LAN Makassar, Indonesia. Email: lukmansamboteng@yahoo.co.id
⁶Muhammadiyah University of West Sumatera, Indonesia. Email: hariefamily@yahoo.co.id
⁷Sekolah Tinggi Agama Hindu Negeri Mpu Kuturan Singaraja, Indonesia. Email: Iketutsuparya@gmail.com

Abstract. The Rough Set method is part of machine learning that analyzes the dataset's uncertainty used to determine the attributes of important objects (classification). This study aimed to extract information from the rough set method using the Rough Set Exploration System (RSES) application in the case of determining student graduation on time. The attributes used are Student Status (A1), Gender (A2), Grade Point Average (A3) and Graduation Status (A4). Sources of data obtained from the International Journal of Information System & Technology (IJISTECH) paper. The results of the application of the Rough Set method in determining the predictions of students who will graduate on time produce new knowledge, namely the graduation status based on the Grade Point Average. There are 3 Reductions with the 41 Rule. But overall, the attributes that affect the level of determining students who graduate on time are Student status (A) and Grade Point Average (C).

1. Introducing

Students are intellectuals who carry out the learning process in higher education. Every student certainly eagerly desires to graduate and graduate on time. In addition, tertiary institutions have an obligation to carry out Graduation activities for students who have completed their studies during lectures [1]. Student retention and completion rates for degrees have received critical attention over the years in the higher education literature [2]. In this paper, a classification will be carried out on determining student graduation on time using machine learning algorithms. Machine learning is a collection of programming algorithms that are used to optimize computer or system performance based on pre-existing sample data [3]. There are 7 steps in machine learning, including collecting data, preparing input data, analyzing input data, human involvement, training algorithms, testing algorithms...
and using them. Machine Learning (ML) is one part of the artificial intelligence (AI) algorithm [4]. Many machine learning methods are often used to solve computing problems [5]. Machine learning algorithms have brought about significant changes in the AI field. Machine learning supports human discernment in a special way [6]. Among some of the well-known machine learning algorithms include: Support vector machine [7], logistic regression [8], naive bayes algorithm [9], decision tree [10], boosted tree [11], random forest [12], algoritma KNN [13], rough set [14], etc. This paper proposes a Machine Learning algorithm with the Rough Set method for determining student graduation on time.

Rough Set is a mathematical technique developed by Pawlack and used for data classification analysis in table form [15], and extracting ambiguity in exchange for the boundary of membership values [16]. This method is efficient for handling uncertain information [17]. The Rough set methodology is especially good when used in the field of artificial intelligence, as it can be applied as a component of a hybrid solution in data mining and machine learning [18]. The Rough Set method has been widely used to solve many complex problems. A Iqbal, et al (2020) presented an improved dynamic reduct discovery technique based on the rough set theory. In this technique, the reduction is selected, optimized, and then generalized through the Parallel Feature Sampling (PFS) algorithm. The results of this study indicate that the proposed approach outperforms the newest approaches in terms of efficiency and effectiveness. Overall, the average accuracy achieved was 96%, and a 46.13% reduction in execution time, which was observed by the proposed algorithm compared to contemporary approaches being compared [19]. T M Hossain, et al (2020) select Extra Tree Classifier (ETC) based feature for sorting important attributes and three different electrofacies extracted from dendrogram plots using selected attributes. They proposed the Rough Set theory. Based on the white box classification approach to extracting electrofaceous patterns in the form of decision rules that allow geoscience researchers to correlate electrofacies with lithofacies from extracted Rough Set (RS) rules [20]. B S Shylajla and R Bhaskar (2020) using Rough Set machine learning to optimize the utilization of virtual machines in cloud computing. The machine learning-based virtual machine selection approach integrates migration control mechanisms that increase the efficiency of the selection strategy. Experiments were carried out with various real machine workload situations to provide evidence and effectiveness of the proposed method. Exploration results show that the proposed approach streamlines virtual machine utilization and reduces energy consumption and increases service level agreement violations to achieve better performance [21].

Based on these related researches, it is proposed to implement the use of the Rough Set method in the Determination Case for Student Graduation on Time. Because student graduation also affects accreditation assessments, so it is necessary early to find out what parameters influence a student to complete his studies on time.

2. Methodology

2.1. Sample Data

Sources of data used in the study were obtained from the output of a paper in the International Journal of Information Systems & Technology (IJISTECH). The sample data contains 150 records [22]. The sample data is processed using the Rough Set Exploration System (RSES) to obtain a decision rule, which can be seen in table 1.

| No | Student Name       | Student Status | Gender | Grade Point (GP) | GPA | Graduation Status |
|----|--------------------|----------------|--------|------------------|-----|--------------------|
| 1  | Vinkhi F Saragih   | Student        | Male   | 3.7 3.5 3.3 3.4 | 3.4 | 3.7 3.9 3.4 | 3.4 | On time |
| 2  | Wahyu Prasetyo     | Student        | Male   | 3.7 3.4 3.6 3.6 | 3.9 | 3.3 3.9 3.5 3.8 3.6 | 3.6 | Late |
| 3  | Eko Supiandi       | Work           | Male   | 3.2 3.4 3.5 3.6 | 3.3 3.6 3.3 3.3 3.3 3.3 3.4 | 3.4 | Late |
| 4  | Natal Ingot Siruit | Work           | Male   | 3.3 3.3 3.6 3.1 | 3.1 3.3 3.2 | 3.8 3.4 | 3.4 | Late |
| 5  | Petrus R. Siombing | Student        | Male   | 3.4 3.3 3.6 3.1 | 3.6 2.8 3.6 4.0 3.4 | 3.4 | Late |
| 6  | Abisaleh Zebua     | Work           | Male   | 3.2 3.0 3.3 3.0 | 3.1 3.0 | 3.3 3.6 3.2 | On time |
| 7  | Ricky H Simatupang | Work           | Male   | 2.9 2.6 2.7 3.0 | 3.3 2.7 3.3 3.8 3.0 | 3.0 | Late |
| 8  | Hari Susanto       | Work           | Male   | 3.3 3.0 3.1 2.9 | 3.0 3.3 3.4 3.6 3.2 | 3.2 | Late |
| 9  | Suci AnggunTari    | Work           | Female | 3.1 3.0 3.3 3.0 | 3.1 3.0 3.3 3.6 3.2 | 3.2 | On time |
Based on the sample data in table 1, only important attributes that represent the contents of the table will be taken. Among them are Status_Student, Gender, Grade Point Average (GPA) and Graduation Status.

2.2. Eligibility Criteria Analysis

The condition attributes used in the study were Status_Student, Gender, Grade Point Average (GPA) and Graduation Status. Meanwhile, the attribute of the decision is Graduation Status. The following is a list of attributes used in determining the determination of on-time graduation which can be seen in table 2.

| No | Student Name       | Student Status | Gender | Grade Point (GP) | GPA | Graduation Status |
|----|--------------------|----------------|--------|------------------|-----|-------------------|
|    |                    |                |        | Sem 1  | Sem 2  | Sem 3  | Sem 4  | Sem 5  | Sem 6  | Sem 7  | Sem 8  |       |
| 10 | Cicci Suryani      | Student        | Female | 3.1    | 3.0    | 3.2    | 3.1    | 3.6    | 3.3    | 3.6    | 3.3    | Ontime |
| 141| Hotni M Saragih    | Work           | Female | 3.3    | 3.0    | 3.0    | 3.4    | 3.4    | 3.0    | 3.6    | 3.3    | Ontime |
| 142| Fadlan Suhanda     | Student        | Male   | 3.5    | 3.2    | 3.1    | 3.1    | 3.0    | 3.0    | 3.1    | 2.0    | 3.0    | Late   |
| 144| Farius Waruwu      | Student        | Male   | 3.3    | 3.1    | 3.5    | 3.4    | 3.3    | 3.6    | 3.0    | 4.0    | 3.4    | Ontime |
| 145| Nurul Fadillah     | Work           | Female | 3.0    | 3.3    | 3.3    | 3.4    | 3.1    | 3.3    | 3.9    | 3.3    | Ontime |
| 146| Bagus Wijaya       | Work           | Male   | 3.2    | 3.0    | 3.0    | 3.3    | 2.9    | 3.3    | 3.3    | 3.1    |        |
| 147| Karmen T Sinaga    | Work           | Male   | 3.6    | 3.8    | 3.6    | 3.3    | 3.1    | 3.3    | 3.6    | 3.5    |        |
| 148| Dian Permatasari   | Student        | Female | 3.1    | 3.4    | 3.4    | 3.3    | 3.6    | 3.7    | 3.3    | 2.3    | 3.3    | Ontime |
| 149| Dewi Novika Sari   | Student        | Female | 3.4    | 3.1    | 3.6    | 3.1    | 3.1    | 3.3    | 3.1    | 3.7    | 3.3    |        |
| 150| Fitri N Urika      | Work           | Female | 3.0    | 3.0    | 3.6    | 3.3    | 3.1    | 3.3    | 3.4    | 3.7    | 3.3    | Ontime |

Source: Processed Data [22]

2.3. Research Procedure

The procedure of the Rough Set method can be represented as shown in Figure 1.

![Figure 1. Rough Set Procedure](image-url)
b. Pre-Processing
   This stage is carried out to eliminate problems that can interfere with the results of the data process, because sometimes in the data there are various problems that can interfere with the results of the process of extracting information from the data itself, such as missing values, redundant data, outliers, or inappropriate data formats. according to the system.

c. Reduct
   The Reduct process is obtained by maintaining a minimal set of attributes that have interesting information. The decision rules were then concluded based on this Reduct process.

d. Testing Data
   At this stage, the student graduation data sample will be tested using RSES tools.

e. Useful Knowledge
   At this stage new knowledge has been found based on the data that has been tested.

3. Results and Discussion
   The results of the analysis are in the form of an explanation of the problem solving algorithm of the Rough Set method based on Figure 1 which has been presented previously.

3.1. Initial Data
   Initial data is obtained from table 1 (Student Graduation Data Sample). This data still needs to be sorted to get the appropriate attributes so that it can be processed to the next stage.

3.2. Pre-Processing (Decision System)
   Pre-Processing (Decision System) Student Graduation Data Samples are carried out to eliminate unnecessary attributes, such as Student Name and Grade Point (GP) Sem 1 - Sem 8. The results of Student Graduation Data Pre-processing can be seen in table 3. Then the data from the pre-processing results are entered into the RSES tools.

| No | Student Status (A) | Gender (B) | GPA (C) | Graduation Status |
|----|--------------------|------------|---------|-------------------|
| 1  | Student            | Male       | 3,4     | On time           |
| 2  | Student            | Male       | 3,6     | Late              |
| 3  | Work               | Male       | 3,4     | Late              |
| 4  | Work               | Male       | 3,4     | Late              |
| 5  | Student            | Male       | 3,4     | Late              |
| 6  | Work               | Male       | 3,2     | On time           |
| 7  | Work               | Male       | 3,0     | Late              |
| 8  | Work               | Male       | 3,2     | Late              |
| ...| ...                | ...        | ...     | ...               |
| 143| Student            | Male       | 3,3     | On time           |
| 144| Student            | Male       | 3,4     | On time           |
| 145| Work               | Female     | 3,3     | On time           |
| 146| Work               | Male       | 3,1     | Late              |
| 147| Work               | Male       | 3,5     | On time           |
| 148| Student            | Female     | 3,3     | On time           |
| 149| Student            | Female     | 3,3     | On time           |
| 150| Work               | Female     | 3,3     | On time           |
3.3. Reduct
The reduct process is carried out after the data pre-processing is complete. After the steps in Figure 2 have been carried out, the next step is to reduct based on the table that has been entered in RSES.

3.4. Testing Data
Data testing is carried out after the reduct process is complete and based on the results of the reduct to produce Knowledge. Data testing is done to obtain generated rules, and it is shown in Figure 4.

3.5. Useful Knowledge
The Reduct results obtained are used to produce Knowledge by referring to the Decision System table, which is an Information Systems that already has decisions or results based on assumptions based on its attributes that meet the terms and conditions. The value of each existing result attribute comes from a sample of data that is converted into an eligibility criterion.
Figure 5. Display RSES for Obtaining Useful Knowledge

Figure 5 is a view of the relationship process between tables, reducts and generated rules using the RSES tools. The resulting knowledge is in the form of generate rules which can be seen in Figure 6.

Figure 6. Rules Useful Knowledge
Figure 6 is a useful knowledge that yields 41 rules. After conducting the test, the results of the analysis carried out can produce optimal decisions in predicting the level of student understanding of the subject in the form of rules or rules patterns that are formed so that they become useful information in decision making. Of the 4 attributes used, attributes that affect the case of Student Graduation Determination On Time is Student (A) and Grade Point Average (C).

4. Conclusion

It can be concluded that several things, namely the application of the Rough Set (RS) method in predicting the level of student understanding of the subject can be applied. The results of the Rough Set method using the RSES application can produce information to make more optimal decisions so that they can provide policies on cases of determining student graduation on time in order to maintain the quality and quality of education. The results of the application of the Rough Set method in determining student graduation on time, produce new knowledge, namely the graduation status based on the Grade Point Average, there are 3 Reductions with 41 Rules.

References

[1] C. A. Dahlvig, J. E. Dahlvig, and C. M. Chatriand, “Institutional Expenditures and Student Graduation and Retention,” *Christian Higher Education*, vol. 19, no. 5, pp. 352–364, 2020.

[2] M. Millea, R. Wills, A. Elder, and D. Molina, “What matters in college student success? Determinants of college retention and graduation rates,” *Education*, vol. 138, no. 4, pp. 309–322, 2018.

[3] S. Badillo et al., “An Introduction to Machine Learning,” *Clinical Pharmacology and Therapeutics*, vol. 107, no. 4, pp. 871–885, 2020.

[4] X.-D. Zhang, *Machine Learning. In: A Matrix Algebra Approach to Artificial Intelligence*. 2020.

[5] J. Tang, X. Zhang, W. Yin, Y. Zou, and Y. Wang, “Missing data imputation for traffic flow based on combination of fuzzy neural network and rough set theory,” *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 0, no. 0, pp. 1–16, 2020.

[6] N. Dutta, U. Subramaniam, and S. Padmanaban, “Mathematical models of classification algorithm of Machine learning,” *International Meeting on Advanced Technologies in Energy and Electrical Engineering*, pp. 1–2, 2019.

[7] D. A. Pisner and D. M. Schnyer, *Support vector machine*. Elsevier Inc., 2019.

[8] L. Connelly, “Logistic Regression,” *MEDSURG Nursing*, vol. 29, no. 5, p. 353, 2020.

[9] H. Zhang, L. Jiang, and L. Yu, “Class-specific attribute value weighting for Naive Bayes,” *Information Sciences*, vol. 508, pp. 260–274, 2020.

[10] M. M. Ghiasi, S. Zendehboudi, and A. A. Mohsenipour, “Decision tree-based diagnosis of coronary artery disease: CART model,” *Computer Methods and Programs in Biomedicine*, vol. 192, no. 105400, pp. 1–14, 2020.

[11] R. T. Selvi and I. Muthulakshmi, “Modelling the map Reduct based optimal gradient boosted tree classification algorithm for diabetes mellitus diagnosis system,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[12] M. Schonlau and R. Y. Zou, “The random forest algorithm for statistical learning,” *Stata Journal*, vol. 20, no. 1, pp. 3–29, 2020.

[13] S. K. Nayak, M. Panda, and G. Palai, “Realization of optical ADDER circuit using photonic structure and KNN algorithm,” *Optik - International Journal for Light and Electron Optics*, vol. 212, no. 164675, pp. 1–8, 2020.

[14] T. M. Hossain, J. Watada, I. A. Aziz, and M. Hermana, “Machine Learning in Electrofacies Classification and Subsurface Lithology Interpretation: A Rough Set Theory Approach,” *Applied Sciences (Switzerland)*, vol. 10, no. 5940, pp. 1–16, 2020.

[15] J. H. Sihotang, “Analysis of Service Satisfaction Level Using Rough Set Algorithm,” *Infokum*, vol. 8, no. 2, pp. 50–56, 2020.

[16] M. Riaz, B. Davvaz, A. Firdous, and A. Fakhar, “Novel concepts of soft rough set topology
with applications,” *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 4, pp. 3579–3590, 2019.

[17] W. Wei and J. Liang, “Information fusion in rough set theory: An overview,” *Information Fusion*, vol. 48, pp. 107–118, 2019.

[18] W. Li, X. Jia, L. Wang, and B. Zhou, “Multi-objective attribute reduction in three-way decision-theoretic rough set model,” *International Journal of Approximate Reasoning*, vol. 105, pp. 327–341, 2019.

[19] A. Iqbal, M. S. Raza, M. Ibrahim, A. Baz, H. Alhakami, and M. A. Saeed, “An Improved Approach for Finding Rough Set Based Dynamic Reducts,” *IEEE Access*, vol. 8, pp. 173008–173023, 2020.

[20] T. M. Hossain, J. Wataada, M. Hermana, and I. A. Aziz, “Supervised Machine Learning in Electrofacies Classification: A Rough Set Theory Approach,” *Journal of Physics: Conference Series*, vol. 1529, no. 052048, pp. 1–6, 2020.

[21] B. S. Shylaja and R. Bhaskar, “Rough-set and machine learning-based approach for optimised virtual machine utilisation in cloud computing,” *IET Networks*, vol. 9, no. 6, pp. 279–283, 2020.

[22] A. Pradipta, D. Hartama, A. Wanto, S. Saifullah, and J. Jalaluddin, “The Application of Data Mining in Determining Timely Graduation Using the C45 Algorithm,” *IJISTECH (International Journal of Information System & Technology)*, vol. 3, no. 1, p. 31, 2019.

[23] Z. Abbas and A. Burney, “A Survey of Software Packages Used for Rough Set Analysis,” *Journal of Computer and Communications*, vol. 4, no. 9, pp. 10–18, 2016.