IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 28, No. 5, October, 2023, pp. 1205-1212
Journal homepage: http://iieta.org/journals/isi

An ID3 Decision Tree Algorithm-Based Model for Predicting Student Performance Using Comprehensive Student Selection Data at Telkom University

Sri Widaningsih, Wardani Muhamad, Robbi Hendriyanto, Heru Nugroho*

School of Applied Science, Telkom University, Bandung 40247, Indonesia

Corresponding Author Email: heru@tass.telkomuniversity.ac.id

https://doi.org/10.18280/isi.280508

Received: 10 April 2023
Revised: 10 July 2023
Accepted: 17 August 2023
Available online: 31 October 2023

Keywords:
student performance, decision tree, classification, Iterative Dichotomiser 3 (ID3), higher education

ABSTRACT

Telkom University, in its routine admission process, generates a rich dataset consisting of various attributes of prospective students. These attributes extend beyond academic parameters like the grade point average (GPA) from the final high school year, encompassing non-academic data such as parental occupation, income, student's gender, origin province, high school major, and school category. Previous research has predominantly focused on academic and sociodemographic data, such as GPA and family income, respectively, for predicting study performance. However, factors like school major, study program, and school category have often been overlooked. In this study, the objective is to utilize the comprehensive Student Selection Data (SMB) to devise a model for predicting the performance of students in their first semester at Telkom University. The aim is to address the issue of a low rate of on-time graduation by leveraging the untapped potential of SMB data. An Iterative Dichotomiser 3 (ID3) decision tree algorithm forms the backbone of the proposed model, enabling the classification of student performance based on a range of diverse attributes. Information gain-based feature selection revealed the five attributes with the greatest influence on student performance in the first semester: gender, grade point average from the final year of high school, study program, high school major, and school category. These findings underscore the potential of a more inclusive approach to student data analysis in predicting academic success in higher education.

1. INTRODUCTION

Telkom University, recognized as one of Indonesia's premier private institutions, places significant emphasis on student adaptation, successful graduation, and effective learning. The selection of new students (SMB) stands as a cornerstone in the university's mission to enroll promising candidates. In this endeavor, the utility of data mining and machine learning algorithms becomes evident, facilitating informed decision-making based on SMB activities and first-semester academic data. Furthermore, patterns discerned through data mining or machine learning can provide valuable insights into a student's potential performance during the first year.

The prediction of student performance has emerged as a central objective within higher education institutions and constitutes a significant area of research. Anticipating student achievement allows faculty to intervene proactively, offering support to students at risk of underperforming or withdrawing before examinations, thereby enhancing the institution's reputation [1, 2]. Despite the potential of SMB data in tracing student performance trends, its utilization within study programs at Telkom University has been limited. This underutilization correlates with a persisting concern—the low rate of on-time student graduation.

Previous studies have deployed student data to predict timely graduation, focusing primarily on academic attributes such as the grade point average from the final year of high school and the academic average or GPA, in addition to sociodemographic aspects like family income and gender [3-22]. However, factors like high school majors, study programs, and school categories have often been overlooked in the prediction of study performance [23]. Machine learning algorithms such as Decision Trees and Naïve Bayes, applied to university data, have shown potential in predicting student success, thereby bolstering academic and non-academic outcomes. A co-occurrence map using Vos Viewer, based on keywords and terms extracted from titles and abstracts obtained via Publish or Perish (PoP), reveals the prevailing focus of scientific publications, as depicted in Figure 1.

In the ongoing quest to enhance student achievement and retention rates, the current research aims to develop a decision tree model to predict student performance at Telkom University. Identifying the salient factors impacting academic performance in the first semester allows for targeted interventions to support struggling students and to improve overall success rates. The primary research question driving this study is: What are the key factors influencing student success at Telkom University, and how can a decision tree model be leveraged to predict student performance?

The overarching goal of this research is to devise a decision tree algorithm capable of predicting academic achievement based on diverse input data. The proposed decision tree analysis is intended to isolate the most critical determinants of student success. A comprehensive understanding of these factors will enable the institution to tailor its programs and processes more effectively to meet specific needs.

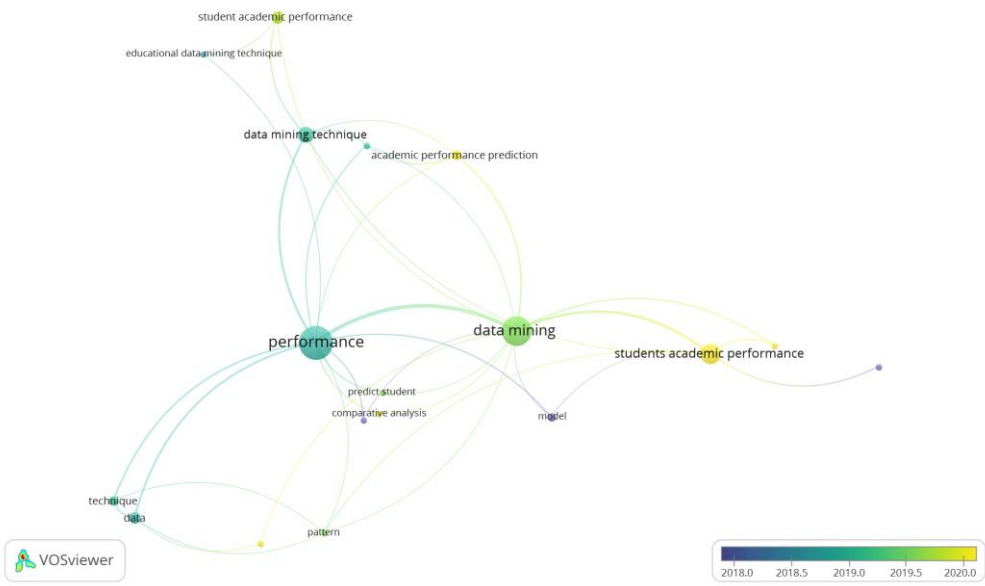

Figure 1. Co-occurrence maps base on vos viewer

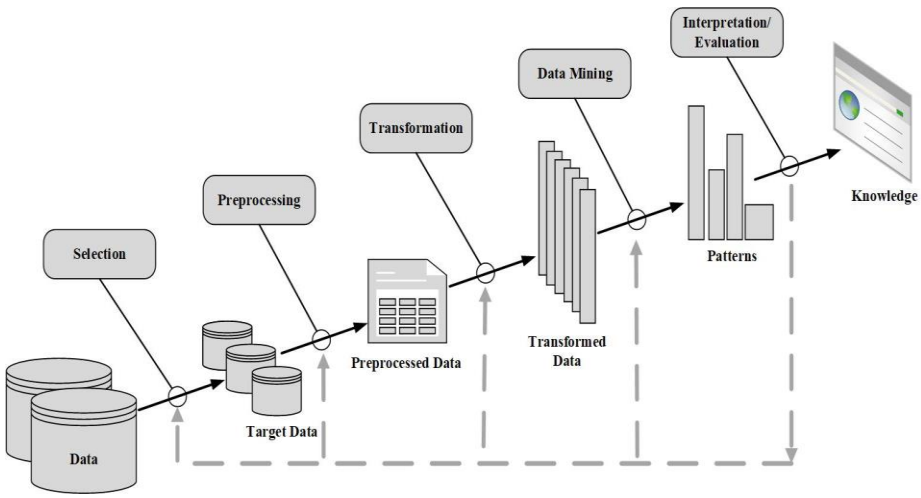

Figure 2. KDD process preparation steps

The remainder of this research is structured as follows: The 'Literature Review' section provides a concise overview of pertinent works, focusing on student performance prediction models based on machine learning algorithms. Subsequently, the 'Methodology' section elucidates the algorithm underpinning this research. The subsequent sections present the research outcomes, discussion, and contributions. The study concludes with a contemplation of research limitations and potential avenues for future exploration.

2. LITERATURE REVIEW

The pertinent literature within the scope of this research encompasses four principal areas—data mining, data analytics, decision tree algorithms, and student performance prediction using machine learning.

2.1 Data mining

Data mining is characterized by the application of rules, processes, and algorithms aimed at deriving actionable insights, identifying patterns, and unveiling relationships within voluminous data sets [24]. Often referred to as the Knowledge Discovery in Database (KDD) process, data mining is tasked with the analysis of extensive information repositories to uncover potentially valuable implicit knowledge [25]. Furthermore, the aggregation of substantial data allows data mining to discern patterns and trends, and reveal concealed relationships [26]. The functions of data mining, based on their utility, can be categorized into four groups: association, classification, clustering, and regression [27-29].

Data mining analysis typically employs three methodologies-classical statistics, artificial intelligence, and machine learning [30]. The comprehensive procedures for establishing the KDD process, as delineated by Plotnikova et al. [31], are depicted in Figure 2.

In the study "Predicting Student Success at Telkom University Using Decision Tree Algorithm," data mining is exploited for the analysis of a considerable dataset encompassing student information. Integral data concerning students' academic performance and sociodemographic

characteristics are collected. Post the data cleaning and preparation stage, a decision tree algorithm is deployed to predict student performance predicated on these parameters. The objective is to pinpoint primary determinants of student success and render insights instrumental in enhancing the student support at Telkom University.

2.2 Data analytic

Data analytics, a multidisciplinary field, involves both quantitative and qualitative data analysis to derive inferences, uncover novel information, or collect and validate data for decision-making and action [32]. As propounded by Breiman [33], the aim of data analytics is to forecast future responses to input variables and to discern the inherent relationships between response variables and input variables.

Data analytics is transitioning from small, simple data with hypothesis testing to large, complex data analysis intended for hypothesis-free knowledge and insight discovery. This shift marks a transition from the explicit era to the implicit age. The importance and complexity of data analytics have surged in several fields, including computer science, biology, medicine, finance, and homeland security. This paradigm shifts from explicit to implicit is crucial in the categorization of analytical data, particularly descriptive analytics, predictive analytics, and prescriptive analytics [32], as delineated in Table 1.

Table 1. Data analytic category

Category	Description
Descriptive Analytics	Describes a type of data analysis that typically employs statistics to transform raw data into usable information [32].
Predictive Analytics	A data-driven, model-driven strategy for producing what-if scenarios [34].
Prescriptive Analytics	Needs predictive analytics using deterministic and stochastic optimization techniques like the decision tree and Monte Carlo methods [34].

The process of decision-making within a business context can be facilitated by a diverse array of tools and strategies, each displaying varying degrees of sophistication and requisite expertise (Figure 3) [35].

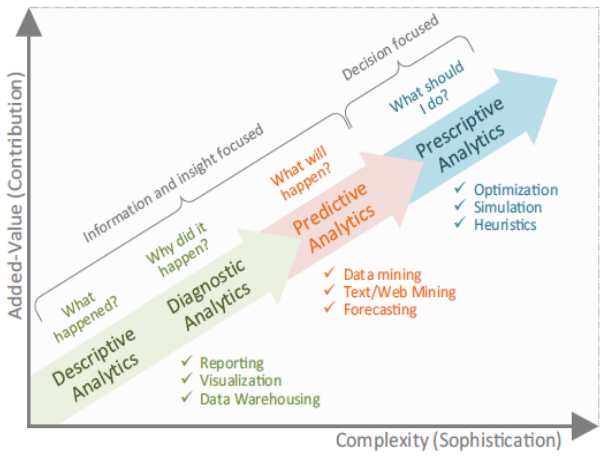

Figure 3. Identifying business analytics along the value proposition and computational sophistication dimensions [36]

The process of data analytics contributes to the identification of salient factors that influence student success at Telkom University. By comprehending these key predictors, the university is in a position to devise targeted interventions and support systems to augment student outcomes. Moreover, data analytics enables the researchers to address gaps in the extant literature by supplying tailored insights specific to the student population at Telkom University.

2.3 Decision tree algorithm

Within the realm of machine learning, decision trees are classified under supervised machine learning. The frequent employment of decision trees is attributed to their simplicity in implementation, analysis, and application to qualitative, quantitative, continuous, and discrete variables, in addition to their propensity to deliver accurate results [37]. The structure of the decision tree is depicted in Figure 4.

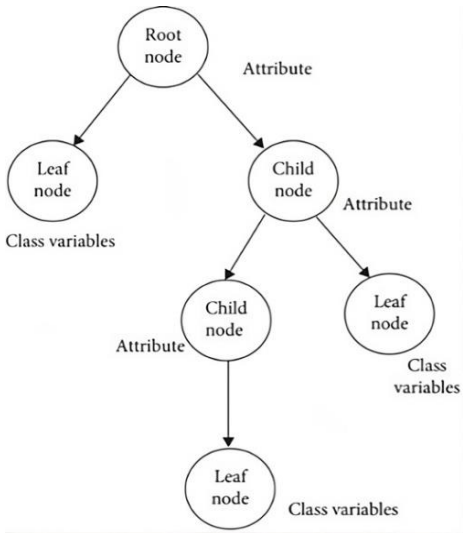

Figure 4. Decision tree structure [38]

The C4.5 algorithm, introduced by Singh, is among the decision-making algorithms [39] and constitutes an advanced version of Quinlan's ID3 algorithm. Furthermore, the J48 Algorithm is extensively utilized in data mining for decision tree classification. The C4.5 decision tree is also referred to as the J48 classifier. This algorithm organizes data following a top-down distribution. The final decision tree is constructed by dividing the data based on the attribute that offers the maximum information gain [40]. The ID3 algorithm is considered a rudimentary decision tree algorithm [41]. The growth of the ID3 algorithm ceases when all instances pertain to a single value of a target feature or when the optimal information gain does not exceed zero using information gain as a splitting condition. The ID3 algorithm does not incorporate a pruning mechanism and does not manage numeric attributes or missing data. The fundamental advantage of the ID3 algorithm lies in its simplicity, hence its prevalent usage in education [42].

2.4 Student performance prediction using machine learning

As described in Table 2, Recent literature has facilitated the consolidation of variables into fewer factors, including

previous academic performance (grade point average from the final year of high school, academic average or GPA) [16-18, 43, 44], sociodemographic characteristics (gender, family income, occupation) [19-22, 43, 44], and academic environment (type of program, faculty) [43].

Table 2. Students performance variables

Factor	Variable	Paper	Method
Academic	grade point average from the last year of high school, GPA	[16]	SVM, KNN, DT
		[17]	RF, NB, DT, SVM
		[18]	DT
		[43]	DT, ANN
		[44]	KNN, NB
Socio demographic	gender, family income, occupation	[19]	ANN, NB SVM, DT
		[20]	DT, NN, SVM
		[21]	SVM, ANN
		[22]	ANN, LR, KNN
		[43]	DT, ANN
		[44]	KNN, NB
Academic environment	type of program, faculty	[43]	DT, ANN

The focal research question of this study is: What are the pivotal factors influencing student success at Telkom University, and how can a decision tree model be harnessed for the prediction of student performance? Several domains, such as major in high school, study program, and high school category, are either unexplored or have received minimal attention in the evaluation of variables employed in prior research.

The variables "majors in high school," "study program," and "high school category" have not been exhaustively investigated in the existing literature. The research surrounding how these specific criteria impact student performance at Telkom University remains limited. These criteria can elucidate the effects of students' educational backgrounds and study interests on their performance and success in higher education. Comprehending how high school majors, university programs, and high school categorization influence student success can aid in designing educational interventions and support systems for different student cohorts.

To address this research gap, our study will probe into these factors that have been relatively under-studied and deploy the decision tree algorithm to ascertain their significance in predicting student success. By incorporating these variables into the analysis, it is anticipated that a broader understanding of the myriad factors affecting academic performance at Telkom University will be attained and more effective strategies to bolster student success in university, and their retention, will be devised.

3. RESEARCH METHODOLOGY

The research method is based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) [45-47]. The Cross-Industry Standard Process for Data Mining (CRISP-DM) is a popular and exhaustive framework for directing data mining operations. It gives an organized method to data mining jobs, from analyzing the problem to deploying the solution. CRISP-DM is indispensable in the field of data mining because it enables data scientists and analysts to prepare and execute projects effectively, hence preventing the omission of crucial steps. This research method consists of 4 stages: business understanding, data understanding, data preparation, and modeling. Figure 5 presents the sequence of steps that must be followed in implementing CRISP-DM.

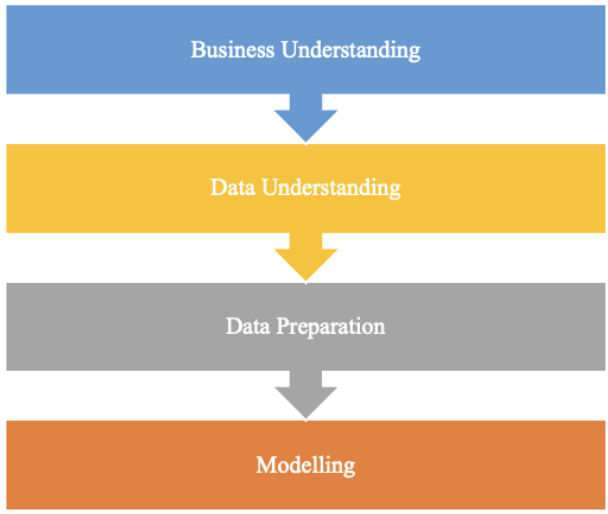

Figure 5. Research stages

3.1 Business understanding phase

This work employs a decision tree technique to build a decision tree model from SMB data. This approach is used to problems and data with various qualities to find the factors that influence the first-semester student success rate. From a commercial perspective, the admissions directorate can employ attribute identification as a factor in the marketing process for prospective new students by focusing on the characteristics that have the greatest impact on first-year academic achievement. In addition, for study programs, this decision tree model can predict the likelihood of failure for first-year students, thereby minimizing the number of withdrawals that can degrade the quality of study programs according to accreditation criteria.

3.2 Data understanding phase

In this phase, several activities are carried out, including initial data collection, description, exploration, and quality assurance. The objective of the data gathering procedure is to examine the structure of the data handled by the admissions unit and combine it with student academic data collected from the information system management unit and the study program. This data collection attempts to gain an initial understanding of the factors that determine the academic success of first-year students. The defined attributes (included in Table 3) will subsequently be utilized in the data mining procedure.

During this phase, data verification is performed to verify the data mining process is error-free.

(1) Ensure data sources include right values and no data anomalies.

(2) Verify that there are no missing data or values. If the record contains no entries, it will be removed.

(3) Ensure there is no duplication of data; if there is duplication, one of the duplicated data will be erased.

Table 3. SMB data attributes

Attribute	Information	Data Type
Father's Occupation	Type of father's occupation	Category
Mother's Occupation	Type of mother's occupation	Category
Gender	Student Gender (M/F)	Category
Province of Origin	Student's province of origin	Category
Highschool Major	Student's major at high school	Category
Study Program	Category of student study program	Category
School Category	Highschool Category	Category
Average Score	Score when registered as a student	Numerical
GPA of First Semester	GPA achievement index of 1st Semester	Category

3.3 Data preparation phase

In this phase, data processing is carried out as many as 16,776 records obtained from the 2018-2021 batch student data with 9 attributes. At this stage the process includes data selection, data cleaning and data transformation. The data selection stage aims to select the attributes that will be used in the formation of the decision tree model. At this stage there are two attributes that are not included in data processing, namely the salary attribute and the year of study attribute. The salary attribute is not used because the input data needs to be consistent, and much data must be included. The year of study attribute is only used as a marker because it only contains incoming batches of students. There are several simplified attributes in the data transformation process so that there are not too many types of filling in the attributes, such as the father/mother's occupation being grouped into working/not working and student study programs being grouped into social and engineering school, which initially quite a lot match the study programs the students come from. At this stage, numerical GPA data from the first semester were converted into categorical data because a decision model would be created based on SMB datas and academic data from the first semester using the following rules. IF(GPA=4; "Excellent"; IF(GPA) >3,5; "Very Good"; IF(GPA>3; "Good"; IF(GPA>2,5; "Fair"; IF(GPA>2; "Satisfactory"; "Passing")))). This regulation is based on the academic norms of Telkom University for referencing student accomplishment indices.

An outlier is a data point that deviates significantly from most other data points in a dataset. Unlike global outlier identification approaches that consider the entire dataset, LOF evaluates the degree of abnormality of each data point based on its immediate neighborhood. This makes LOF very useful for spotting outliers in datasets with variable densities and clusters. To obtain the final data, which will be processed in the subsequent stage as many as 13,581 records, outlier detection is also carried out at this stage using the Local Outlier Factor approach. This step is where the final data is obtained. The data distribution from each attribute mapped to the target class is categorized graphically in Figure 6.

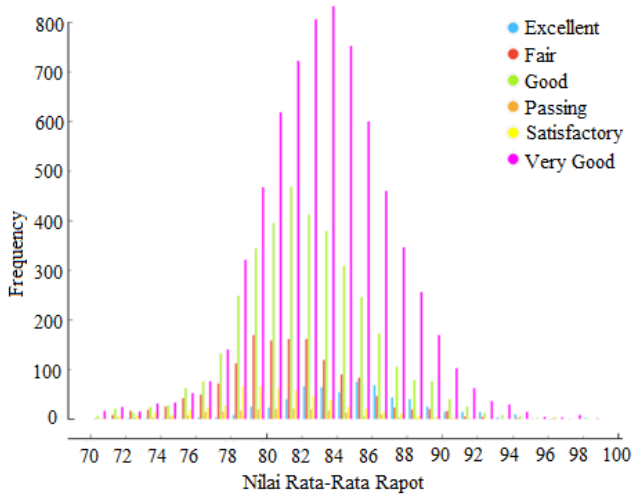

Figure 6. Data distribution for each attribute

3.4 Data modeling phase

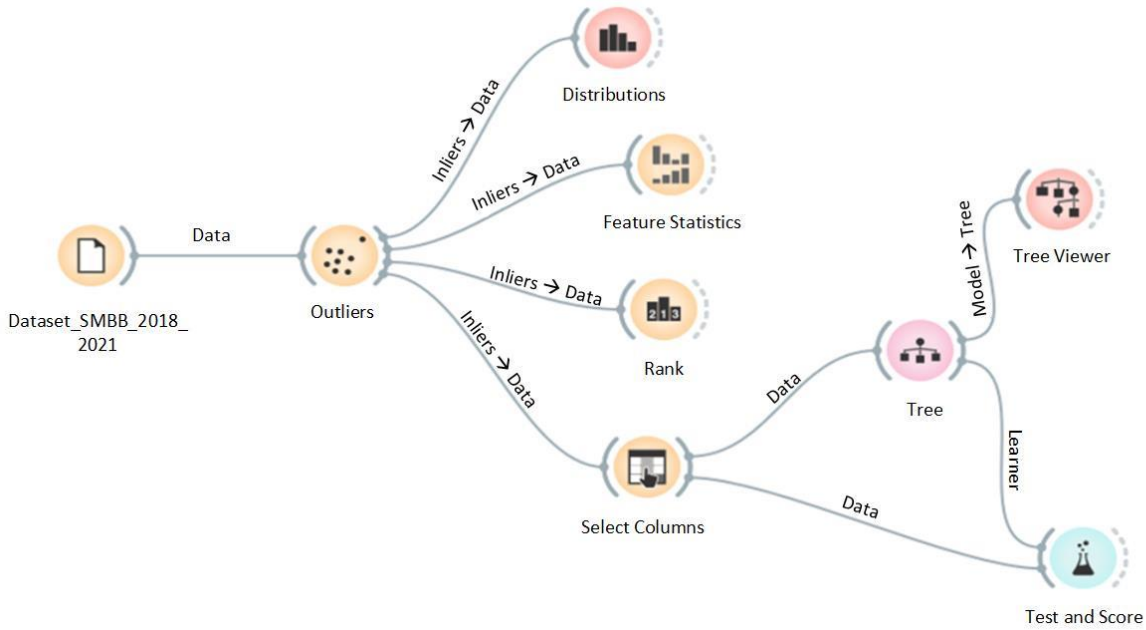

Figure 7. Feature processing and modeling using orange

During this phase, preprocessed data will be represented using a decision tree method. Nevertheless, a feature selection procedure will be conducted depending on the Information Gain earlier. This data modeling will make use of orange, a frequently utilized data mining tool in comparable investigations. Figure 7 depicts the outcomes of choosing features using orange.

The decision tree method was used for this study to predict student performance at Telkom University because to its numerous advantages that align with the aims of the research topic: (i) Interpretability: Decision trees provide a model that is highly interpretable. The resulting tree-like structure is simple to comprehend and interpret, making it ideal for finding the most influential indicators of student performance. (ii) Feature Importance: Decision trees provide a rating of feature importance that indicates which variables have the most influence on the target variable.

4. RESULT AND DISCUSSION

Based on the results of feature ranking using Information Gain (IG), the sequence of features that influence the modeling process is obtained, as shown in Table 4.

Based on this ranking, the five attributes with the highest IG scores were chosen: gender, average score, study program, high school major, and school category. The generated tree is displayed using the orange module's tree viewer. Because the resultant tree is highly detailed, we restrict the depth to four to make the tree model easy to read. Figure 8 depicts the outcomes of the tree model produced using the SMB dataset.

Gender influences academic performance in the first semester, according to the decision tree model. The Average Score is the second crucial feature for the male gender. Nonetheless, the decision tree shows that the expected value is still less than 50%. The female gender's success in the first semester is determined by the study program chosen. Whereas the figure for Bachelor of Social School is high at 71.3%, most female students receive the designation "Very Good".

Table 4. Feature rank

Feature	#	Info. Gain
Gender	2.0	0.51010096902557315
Average Score		0.04570195758572626
Study Program	2.0	0.045433258551514757
Highschool Major	7.0	0.011772790944482248
School Category	2.0	0.04448586323420045
Province of Origin	7.0	0.003314559738137657
Mother's Occupation	2.0	0.0010999044703932093
Father's Occupation	2.0	0.0004861422400905812 7

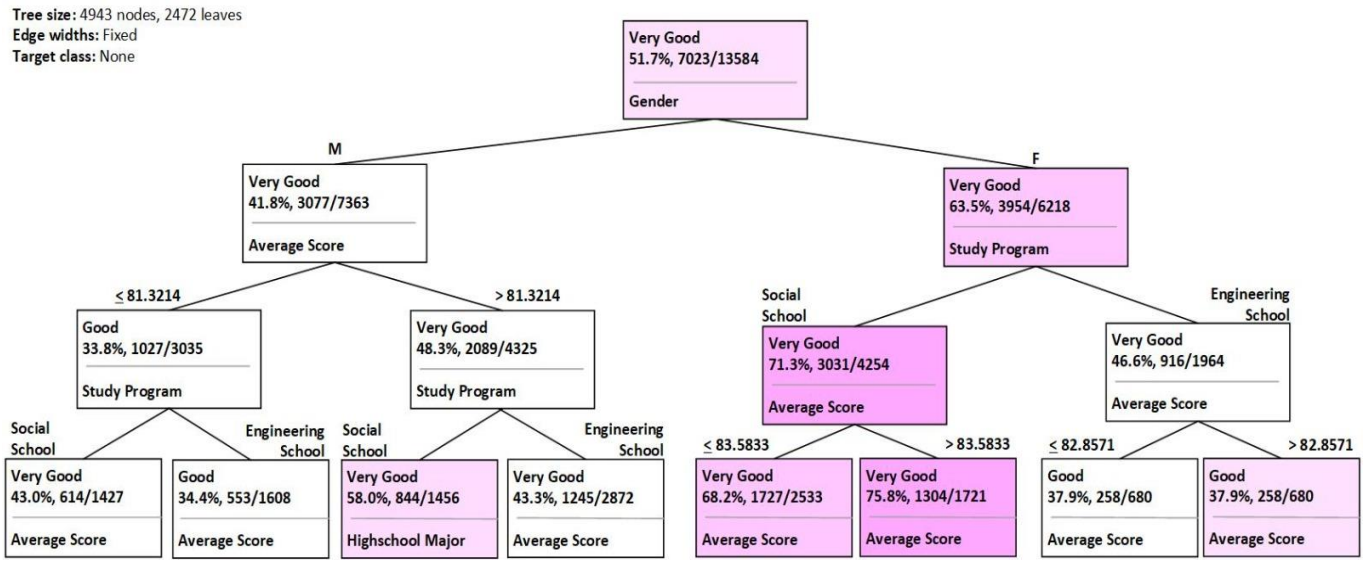

Figure 8. Telkom university SMB data decision tree model

5. CONCLUSIONS

Data mining has been widely employed in past study in the field of education to identify hidden knowledge and patterns about student academic performance, which can be seen from many factors such as father's/mother's occupation, gender, province of origin, high school major, study program, school category, average score and GPA of first semester. Student academic performance information can be used to forecast graduation rates or to measure the success rate of student adaptation to a higher education environment. The success rate of student studies in the first year, as expressed by grade point average scores in the first semester, was predicted in this study.

Experiments based on data from the Telkom University SMB unit generated these results. In addition, the five attributes that had the greatest impact on the student accomplishment index at the beginning of the academic year were identified: gender, average score, study program, high school majors, and school category. In addition, using the Decision Tree Algorithm, a model was created to determine how gender, average score, study program, high school major, and school category influenced the first semester success of the study.

As a result, the predictive model that is produced can determine student performance based on the dataset. The advantage is that the Telkom University SMB unit can use the prediction pattern as part of a marketing strategy to draw students with a propensity for student performance levels either in the first semester or in the first year of university. In addition, for future work, further exploration and

implementation of this technique are likely to yield significant improvements in the overall performance of the institution, with a particular focus on enhancing study programs to ensure higher rates of timely graduation for students and a reduction in student drop-out rates, addressing challenges often attributed to first-year failures.

REFERENCES

[1] Hashim, A.S., Awadh, W.A., Hamoud, A.K. (2020). Student performance prediction model based on supervised machine learning algorithms. In IOP Conference Series: Materials Science and Engineering, 928: 032019. https://doi.org/10.1088/1757-899X/928/3/032019

[2] Priya, S., Ankit, T., Divyansh, D. (2021). Student performance prediction using machine learning. In Advances in Parallel Computing Technologies and Applications, pp. 167-174. https://doi.org/10.3233/APC210137

[3] Pambudi, R.D., Supianto, A.A., Setiawan, N.Y. (2019). Prediction of student graduation based on academic performance using a data mining approach in the information systems study program, Faculty of Computer Science, Brawijaya University. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 3(3): 2194-2200.

[4] Astuti, I.P. (2017). Prediction of graduation time accuracy using the C4.5 data mining algorithm. Fountain of Informatics Journal, 2(2): 41-45. https://doi.org/10.21111/fij.v2i2.1067

[5] Orpa, E.P.K., Ripanti, E.F., Tursina, T. (2019). The prediction model for the start of a student's study period uses the C4.5 decision tree algorithm. JUSTIN (Jurnal Sistem dan Teknologi Informasi), 7(4): 272-278. https://doi.org/10.26418/justin.v7i4.33163.

[6] Sabilla, W.I., Putri, T.E. (2017). Prediction of student graduation time accuracy using k-nearest neighbor and naã-ve bayes classifier. Jurnal Komputer Terapan, 3(2): 233-240.

[7] Setiyani, L., Wahidin, M., Awaludin, D., Purwani, S. (2020). Predictive analysis of student graduation on time using the Naïve Bayes data mining method: Systematic review. Faktor Exacta, 13(1): 35-43. http://doi.org/10.30998/faktorexacta.v13i1.5548

[8] Romadhona, A., Suprapedi, S., Himawan, H. (2017). Prediction of student graduation on time based on age, gender and achievement index using a decision tree algorithm. Jurnal Cyberku, 13(1): 8.

[9] Subawa, I.B. (2019). Prediction of student graduation using Bayes' theorem. Jurnal Nasional Pendidikan Teknik Informatika, 8: 10.

[10] Kristanto, A., Sediyono, E. (2019). Predictive analysis of the level of success in higher education studies based on achievement using the iterative dichotomizer 3 (ID3) method (case study: FTI UKSW). Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 10(2): 433-444. https://doi.org/10.24176/simet.v10i2.2771

[11] Farida, I., Hendric, S.W.H.L. (2019). Prediction of student graduation patterns using data mining classification emerging pattern techniques. Petir, 12(1): 414. https://doi.org/10.33322/petir.v12i1.414

[12] Banjarsari, M.A., Budiman, I., Farmadi, A. (2016). Application of k-optimal in the KNN algorithm to predict on-time graduation for students of the FMIPA Unlam computer science study program based on GP up to semester 4. Klik-Kumpulan Jurnal Ilmu Komputer, 2(2): 159-173. http://doi.org/10.20527/klik.v2i2.26

[13] Munawir, M., Iqbal, T. (2019). Prediction of student graduation using the Naive Bayes algorithm (case study of 5 private universities in Banda Aceh). Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 3(2): 59-63. https://doi.org/10.35870/jtik.v3i2.77

[14] Rohman, A., Rochcham, M. (2019). Comparison of data mining classification methods for predicting student graduation. Neo Teknika, 5(1): 23-29. https://doi.org/10.37760/neoteknika.v5i1.1379

[15] Royan, S., Yulian, A., Syaechurodji, S. (2021). Implementation of data mining uses the naive Bayes method with feature selection to predict student graduation on time. Jurnal Ilmiah Sains Dan Teknologi, 5(2): 9-22. https://doi.org/10.47080/saintek.v5i2.1511

[16] Ajibade, S.S.M., Bahiah Binti Ahmad, N., Mariyam Shamsuddin, S. (2019). Educational data mining: Enhancement of student performance model using ensemble methods. In IOP Conference Series: Materials Science and Engineering, 551: 012061. https://doi.org/10.1088/1757-899X/551/1/012061

[17] Jalota, C., Agrawal, R. (2019). Analysis of educational data mining using classification. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, pp. 243-247. https://doi.org/10.1109/COMITCon.2019.8862214

[18] Castrillón, O.D., Sarache, W., Ruiz-Herrera, S. (2020). Prediction of academic performance through artificial intelligence techniques. Formación Universitaria, 13(1): 93-102. https://doi.org/10.4067/S0718-50062020000100093

[19] Hussain, M., Zhu, W., Zhang, W., Abidi, S.M.R., Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. Artificial Intelligence Review, 52: 381-407. https://doi.org/10.1007/s10462-018-9620-8

[20] Xu, X., Wang, J., Peng, H., Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. Computers in Human Behavior, 98: 166-173. https://doi.org/10.1016/j.chb.2019.04.015

[21] Nieto, Y., García-Díaz, V., Montenegro, C., Crespo, R.G. (2019). Supporting academic decision making at higher educational institutions using machine learning-based algorithms. Soft Computing, 23: 4145-4153. https://doi.org/10.1007/s00500-018-3064-6

[22] Wang, L., Yuan, Y. (2019). A prediction strategy for academic records based on classification algorithm in online learning environment. In 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), Maceio, Brazil, 2161, pp. 1-5. https://doi.org/10.1109/ICALT.2019.00007.

[23] Contreas-Bravo, L.E., Nieves-Pimiento, N., González Guerrero, K. (2023). Prediction of university-level academic performance through machine learning mechanisms and supervised methods. Ingenieria, 28(1): e19514. https://doi.org/10.14483/23448393.19514.

[24] Morabito, V. (2016). The future of digital business innovation. Springer International Publishing, 10: 978-3. https://doi.org/10.1007/978-3-319-26874-3.
[25] Han, J., Kamber, M. (2006). Data mining: Concepts and techniques, 2nd ed. in The Morgan Kaufmann series in data management systems. Amsterdam; Boston: San Francisco, CA: Elsevier, Morgan Kaufmann, 2006.
[26] Sumathi, S., Sivanandam, S.N. (2006). Introduction to data mining and its applications (Vol. 29). Springer.
[27] Hui, S.C., Jha, G. (2000). Data mining for customer service support. Information & Management, 38(1): 1-13. https://doi.org/10.1016/S0378-7206(00)00051-3
[28] Kao, S.C., Chang, H.C., Lin, C.H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. Information Processing & Management, 39(1): 133-147. https://doi.org/10.1016/S0306-4573(02)00019-5
[29] Nicholson, S. (2006). The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. Information processing & management, 42(3): 785-804. https://doi.org/10.1016/j.ipm.2005.05.008
[30] Girija, N., Srivatsa, S.K. (2006). A research study: Using data mining in knowledge base business strategies. Information Technology Journal, 5(3): 590-600. https://doi.org/10.3923/itj.2006.590.600
[31] Plotnikova, V., Dumas, M., Milani, F. (2020). Adaptations of data mining methodologies: A systematic literature review. PeerJ Computer Science, 6: e267. https://doi.org/10.7717/peerj-cs.267
[32] Cao, L. (2017). Data science: A comprehensive overview. ACM Computing Surveys (CSUR), 50(3): 1-42. https://doi.org/10.1145/3076253
[33] Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16(3): 199-231.
[34] Lam, D. (2014). A survey of predictive analytics in data mining with big data. Athabasca University.
[35] Banerjee, A., Bandyopadhyay, T., Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential? Vikalpa, 38(4): 1-12. https://doi.org/10.1177/0256090920130401
[36] Delen, D., Ram, S. (2018). Research challenges and opportunities in business analytics. Journal of Business Analytics, 1(1): 2-12. https://doi.org/10.1080/2573234X.2018.1507324
[37] Kaur, H., Wasan, S.K. (2006). Empirical study on applications of data mining techniques in healthcare. Journal of Computer Science, 2(2): 194-200. https://doi.org/10.3844/jcssp.2006.194.200
[38] Hajjej, F., Alohali, M.A., Badr, M., Rahman, M.A. (2022). A comparison of decision tree algorithms in the assessment of biomedical data. BioMed Research International, 2022: Article ID 9449497. https://doi.org/10.1155/2022/9449497
[39] Singh, K., Sulekh, R. (2017). The comparison of various decision tree algorithms for data analysis. International Journal of Engineering and Computer Science, 6(6): 21557-21562. https://doi.org/10.18535/ijecs/v6i6.03
[40] Khanom, N.N., Nihar, F., Hassan, S.S., Islam, L. (2020). Performance analysis of algorithms on different types of health related datasets. In Journal of Physics: Conference Series, 1577: 012051. https://doi.org/10.1088/1742-6596/1577/1/012051
[41] Quinlan, J.R. (1986). Induction of decision trees. Machine Learning, 1: 81-106. https://doi.org/10.1007/BF00116251
[42] Maimon, O.Z., Rokach, L. (2014). Data Mining with Decision Trees: Theory and Applications. 81. World Scientific.
[43] Alsalman, Y.S., Halemah, N.K.A., AlNagi, E.S., Salameh, W. (2019). Using decision tree and artificial neural network to predict students' academic performance. In 2019 10th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, pp. 104-109. https://doi.org/10.1109/IACS.2019.8809106
[44] Das, A.K., Rodriguez-Marek, E. (2019). A predictive analytics system for forecasting student academic performance: Insights from a pilot project at eastern Washington university. In 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, pp. 255-262. https://doi.org/10.1109/ICIEV.2019.8858523
[45] Marban, O., Mariscal, G., Segovia, J. (2009). A data mining & knowledge discovery process model. In Data Mining and Knowledge Discovery in Real Life Applications. https://doi.org/10.5772/6438
[46] Gusnina, M., Salamah, U. (2022). Student performance prediction in Sebelas Maret University based on the random forest algorithm. Ingenierie des Systemes d'Information, 27(3): 495-501. https://doi.org/10.18280/isi.270317
[47] Venkatachalam, B., Sivanraju, K. (2023). Predicting student performance using mental health and linguistic attributes with deep learning. Revue d'Intelligence Artificielle, 37(4): 889-899. https://doi.org/10.18280/ria.370408