Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13750
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Implementation of Random Forest Algorithm for Graduation Prediction

Fajar Riskiyono [1)*], Deni Mahdiana[2)]
[1)]Magister Ilmu Komputer, Universitas Budi Luhur
[2)]Sistem Informasi, Universitas Budi Luhur
[1)]fajarriskiyono@gmail.com, [2)]deni.mahdiana@budiluhur.ac.id

Submitted :Jun 12, 2024 | **Accepted** : Jul 4, 2024 | **Published** : July 10, 2024

Abstract: The prediction of student study duration is designed to support the study program in guiding students to graduate on time. Data mining techniques can be applied to make predictions, namely by using the Random Forest classification method. The selection of the method is because random forests have the ability to overcome noise and missing values, with this ability the random forest algorithm is able to overcome the problem of data that is used as a dataset. The stages used in this study are data collecting, namely collecting student data, the data selection stage of 300 students with five input data attributes including gender, age, marital status, job status, grade and one attribute as an output containing choices about on time and late. The next stage is preprocessing with the aim of eliminating duplication, noise, and missing values, the stage of data transformation by normalizing age attributes (young and old), grade (large and small). Then the data split stage 3 times, namely 50/50, 40/60, and 30/60, the modeling stage with random forest, and finally, the evaluation stage by analyzing the confusion matrix consisting of accuracy, precision, and recall. The results of the study show that the proposed model can do well with predictions, that is, with the same results for all three data splits. The test value is 100% accuracy, 100% recall, and 100% precision. This study contributed to other researchers to use random forests on student graduation, with 3 times the data split because of the very accurate success rate.

Keywords: graduation, random forest, prediction, data mining, student

INTRODUCTION

The development of the use of information technology and computers in the field of data management has caused the accumulation of very large amounts of data in various universities (Muhammad Sony Maulana et al., 2019). Research conducted at one of the universities in Jakarta shows that in 2016 and 2017 have different graduation rates and amount to 300 students.

If students graduate not on time, there must be factors that cause and will affect the ratio of students in a college (reza maulana & devy kumalasari, 2019). Various studies on graduation classifications that have been carried out indicate that there are many factors that affect the accuracy of student graduation classifications (Gede Suwardika & I Ketut Putu Suniantara, 2019) Graduation time and the imbalance between the number of applicants and graduates are major concerns of universities. This issue causes a buildup of data on students who have not graduated (Romdan Muhamad Ubaidilah et al., 2023).

The accuracy of student graduation in higher education is one of the criteria for assessing campus accreditation (Embun Fajar Wati et al., 2023). The more students who graduate on time, the better the performance of the college. The better the quality of the students (Embun Fajar Wati & Biktra Rudianto, 2022). The timeliness of graduation is also very important for students (Embun Fajar Wati et al., 2024), as a manifestation or form of their struggle for several years studying in college (Sri Diantika et al., 2024). Student graduation rates are difficult to predict early, resulting in late graduation (Zaskila Nurfadilla & Faisal, 2022). The prediction of student study duration is designed to support the study program in guiding students to graduate on time. In addition, if it is known that students who do not graduate on time from an early age, then supervisors and teaching lecturers can provide motivation, and so on that can provide encouragement for students who are predicted not to graduate on time so that these students can graduate on time (Jaya S. Saleh et al., 2022). Graduation analysis can also help universities in decision making (Wati et al., 2021).

This study tries to apply classification techniques to data mining to solve the old prediction problems of the study (I Made Budi Adnyana, 2021). In this problem, data mining techniques can be applied to make predictions, namely by using the Random Forest classification method. The choice of random forest algorithm as a

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13750
e-ISSN : 2541-2019
p-ISSN : 2541-044X

classification algorithm in the data mining section was chosen because random forests have the ability to overcome noise and missing values, with this ability the random forest algorithm is able to overcome the problem of data being used as a dataset. In addition, random forests can cope with large amounts of data and features (Pramudita Oktaviani et al., 2018).

Tree-based algorithms are machine learning methods used to solve problems that require large amounts of data (Muhammad Labib Mu'tashim et al., 2023). Random Forest is one of the most popular and powerful machine-learning (Juwariyem et al., 2024). The effectiveness of machine learning in pattern recognition, classification, prediction, and information retrieval can be obtained by measuring accuracy, recall, and precision. Accuracy is the percentage of correct predictions. Recalls are the percentage of actual positive cases. Precision is the percentage of positive prediction cases (Aria Hendrawan et al., 2021).

Taking into account technological developments and new methods available, this study aims to take a step forward in understanding and improving timely graduation predictions. It is hoped that the results of this study can provide new insights into better approaches to graduation prediction as well as contribute to the development of more effective educational strategies in the future (Indra Irawan et al., 2023). This research contributes in 2 (two) things, first scientifically by testing a random forest algorithm to predict student graduation, and second providing recommendations for universities to develop student admission models selectively (Aang Kisnu Darmawan et al., 2023).

LITERATURE REVIEW

The next research was carried out on the graduation data of one of the study programs at Telkom University, namely the S1 Informatics Engineering study program. The data used is academic data of alumni of regular S1 Informatics Engineering students with alumni status of the class of 2010 – 2017. Research only includes data related to academic data such as courses taken by students, average attendance, and academic grades. The data source used for this classification comes from queries taken in the data warehouse, the amount of data obtained is as many as 1118 data records, alumni students from 2010 to 2017. The student alumni dataset consists of 12 attributes which include student academic data such as NIM, mk_nilai_a, mk_nilai_ab, mk_nilai_b, mk_nilai_bc, mk_nilai_c, mk_nilai_d, mk_nilai_e, GPA, IPK_TPB, PRESENSI_MHS, STATUS. Based on the results of the tests that have been carried out, it can be concluded that 10-fold cross-validation, 5-tree, with 12 attributes is the optimal result of tests that have been carried out using random forests, performance values are also obtained based on the results of testing using micro average f1-score of 77% (Pramudita Oktaviani et al., 2018).

Random forest analysis on student graduation was also carried out at the Open University. The research data used in this study is secondary data in the form of graduation data in 2018 with the number of verified data as many as 133 graduates. The variables used are 1). Response variable (Y) stating student graduation status on time and not on time, 2). The independent variables consist of: gender (X1), major/study program (X2), GPA (X3), parental education (X4), parents' occupation (father and mother) (X5), parents' income (X6) and student status (X7). The application of the CART classification to the punctuality classification of Open University graduates shows that the variables used as the sorting of the CART classification tree and most determine the accuracy time of graduates are GPA, study program and student status. The results of the analysis showed that random forest was able to increase the accuracy of classification of inaccuracy in student graduation time which reached convergence with classification predictions reaching 93.23% (Gede Suwardika & I Ketut Putu Suniantara, 2019).

The use of random forest algorithms with the selection of data attribute features using wrapper and embedded methods carried out on the dataset of students who have completed undergraduate studies in higher education as many as 146 data that have 13 data attributes and 1 label. This feature selection method is carried out to obtain appropriate data attributes which will later be used in the process of classifying predictions of students who have graduated on time or not. The results showed that using the Random Forest Classification algorithm, the accuracy value was 73% and there was an increase in accuracy after using the wrapper and embedded feature selection method by 100% (Aria Hendrawan et al., 2021).

The results of classification using the Random Forest Algorithm using 1,351 data, then obtained the evaluation results with an accuracy value of 90.74% by dividing and testing as much as 80:20 The system successfully displayed data visualization to predict graduation on time by implementing the method (Zaskila Nurfadilla & Faisal, 2022). Research in the same year was also conducted at UKDLSM with a student graduation period in 2018 which amounted to 443 data records using 9 attributes. The accuracy value of the results of classifying student graduation datasets using the random forest algorithm increased along with increasing the use of the number of trees, namely with an accuracy of 90.00% using 50 trees (Jaya S. Saleh et al., 2022).

The data used as the next research object has an unbalanced number of data classes, to overcome this, a random oversampling (ROS) resampling technique is applied and also sets Split validation where there will be a division between learning data by 50% for 50% test data. To examine the model formed, the author uses evaluation metrics such as accuracy, precision, and recall. The results of the research show that the proposed model can well carry out predictions compared to other models, namely with a precision value of 87.05%, accuracy test of 90.04%,

*name of corresponding author

recall of 90.04%. From these results, it can be interpreted that the random forest algorithm is considered good in predicting the timeliness of a student's graduation (Sri Diantika et al., 2024).

From the percentage of accuracy, the largest value produced is 100% with increased accuracy after using the wrapper and embedded feature selection method (Aria Hendrawan et al., 2021). The second largest result, 93.23%, was carried out at the open university (Gede Suwardika & I Ketut Putu Suniantara, 2019). The accuracy results of the two are not far apart. In this study, the author also tried to explore the accuracy of predictions about student graduation with random forest with the split method.

METHOD

In this method section, it will be described in detail about the process tried in this research, from the starting point will be collected data used in research to how the flow or stages of research predict the timeliness of graduation of these students. The research flow chart can be seen in figure 1 below (Aang Kisnu Darmawan et al., 2023) (Sri Diantika et al., 2024):

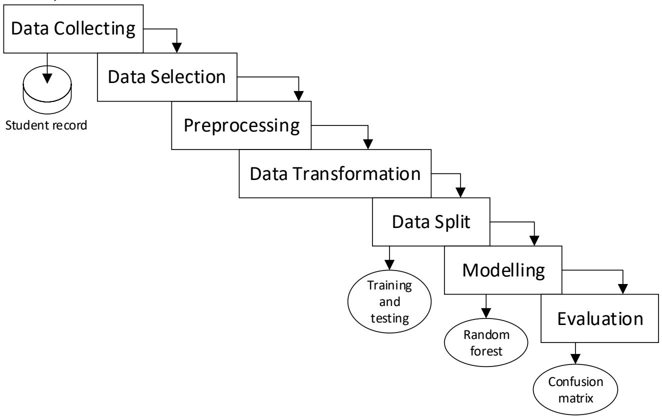

Figure 1. Research Flow

Data collecting

The data set used as the object or material of this research is data on the timeliness of graduation at one of the universities, but the name of the university is kept secret. The data used is secondary data, which is obtained from the results of literature studies (Embun Fajar Wati et al., 2024). This data was taken from 2016 to 2017 (Wati et al., 2021). The dataset contains a collection of student data. The amount of data used is 300 data. Each sample has 5 parameters and a class label that identifies it there are as many as 2 data classes.

Data selection

In this data selection process by eliminating some attributes that are not relevant to the purpose of the study. The selected attributes include 5 attributes because the information provided in it already represents the data needed to be used as research indicators.

Preprocessing

One process that should not be missed before going into the prediction stage is preprocessing. This stage is considered important because not all data used as objects in research are ready (Hermawan & Bellaniar Ismiati, 2020), Sometimes there is empty data to double data. This process is also tried so that the information to be used is suitable for research needs. Several preprocessing steps are also tested, such as filling in missing information, deleting duplicate data, checking information consistency, cleaning data, and correcting errors that may be contained in the dataset (Hikmah Dwiyanti Nasir et al., 2020). The initial preprocessing process carried out in this study was to carry out a data cleaning process, this process was carried out to eliminate data that had missing value and noise. Then also carried out double or double data cleaning.

After checking the data and it turns out that all 300 data are clean (no noise, missing, or duplicate data). The data consists of 5 (five) input data attributes including personal data (gender, age, marital status, job status) and academic data (grade) and 1 (one) attribute as output containing choices about being on time and late.

Data transformation

This stage of data transformation is carried out by converting data into a format suitable for data mining processing. In this research, rapidminer software wants data to be processed in the form of excel. The data is converted into categories.

Table 1. Data Transformation

Age	Young	19 – 24 years old
	Old	25 – 50 years old
Grade	Large	3 – 4 years old
	Small	1 – 2.9 years old

Source : (Embun Fajar Wati et al., 2024)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13750 p-ISSN : 2541-044X

Data Split

At this stage, data is divided into training and testing data using Rapidminer software with 50/50 data split. After completing the data preparation through the preprocessing process, the next step is to separate the data into two parts. The first part serves as learning data, which will be used to train the model, while the second part will be used as testing data. This learning data is used to instruct the model, while the testing data plays a role in evaluating the extent to which the proposed model is able to make predictions accurately (Sri Diantika et al., 2023). After completing the preprocessing stage, 300 data were found ready to use and 5 parameters with 2 data classes, namely the right class and the late class. This ready-to-use data is then divided into 2 parts, the data is used to teach the research model to recognize data and data to test the research model. Data split will be done 3 times with different percentages, namely 50/50, 40/60, 30/70.

Modelling

Before modeling with Rapid Miner, calculations with formulas are needed to form trees. The process of splitting the nodes is obtained by doing a calculation known as Information Gain to find the variables to be selected to grow the tree. To grow Information Gain, information theory calculations are carried out, namely Entropy. With the following equation (Zaskila Nurfadilla & Faisal, 2022):

$$Entropy\ (Y) = -\sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Where Y is the set of cases and p(c|Y) is the proportion of Y values to class c. Meanwhile, to get the Gain using the following equation:

$$Information\ Gain\ (Y, a) = Entropy\ (Y) - \sum_{v \in Values(a)} \frac{|Y_v|}{|Y_a|} Entropy\ (Y_v) \quad (2)$$

Where the value (a) is all that is possible in the set of cases a. Yv is a subclass of Y with class v related to class a. Yes are all values that correspond to the value a.

Once the data classes are balanced, the next step is to implement a random forest model. Random forest is a data mining method and is the result of the development of the decision tree method (Mao et al., 2020). The way it works, the random forest selects attributes randomly to create the number of K trees with different attributes without pruning. In a random forest, the test data will be tested on all constructed trees and then the most frequent outputs will be assigned repeatably until the resulting model meets the criteria.

Evaluation

The final step in this series of research stages is evaluation. At this stage of evaluation, test results are compared to determine better model performance. Performance evaluation of the constructed model involves examining the confusion matrix, which provides information about the correctness of data predictions against actual data. By using the confusion matrix, we can obtain Precision, Accuracy, and Recall values. The evaluation was carried out to answer the research question, namely the prediction of student graduation.

RESULT

After going through the process of collecting, selecting, and cleaning data, researchers obtained 300 data that will be broken down into training data and testing data prepared to be entered into research software, namely Rapidminer. The results of the data that has been cleaned are as in Table 2:

Table 2. Student Data

Gender	Age	Marital Status	Grade	Job Status
Male	Young	Married	Low	Employee
Female	Young	Single	Low	Employee
Male	Young	Married	Low	Unemployee
Male	Young	Single	Low	Employee
Male	Young	Married	Low	Unemployee
…	…	…	…	…
…	…	…	…	…
Male	Young	Married	Low	Unemployee
Female	Young	Single	Low	Employee
Male	Young	Married	Low	Unemployee

Source : (Embun Fajar Wati et al., 2024)

Calculation of the entropy value using equation 1 and to calculate the gain value using equation 2, the result is obtained as in the following table 3:

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13750
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 3. Results of Entropy and Overall Gain Calculation

Notes		Entropy	Gain
Amount		0,908	
Gender	Male	0,949	0,02915
	Female	0,604	
Age	Young	0,93	0,0092
	Old	0,722	
Marital Status	Married	0,918	0,290407
	Single	0,529	
Grade	Low	0,929	0,03474
	High	0	
Job State	Employee	0	0,759953
	Unemployee	0,419	

Node 1 is the gain with the highest value, namely Job Status with a value of 0.759953, so that a temporary decision tree is obtained, which is as shown in the following image:

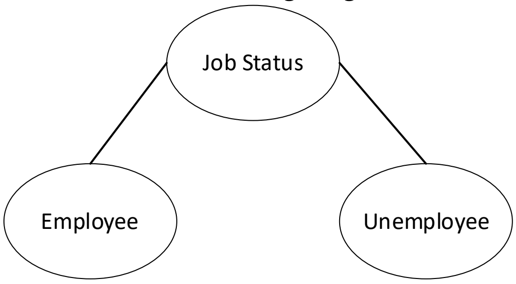

Figure 2. Tree Diagram 1

The next step is to calculate the entropy and gain of employee and unemployee nodes, and take random variables. After data collection, node employees have the same class target, namely on time. This means that the employee node will become a leaf node. Meanwhile, unemployee nodes will be calculated the entropy value and gain to get branching. The entropy and gain values of the unemployee nodes, on the gender, age, marital status, and grade attributes are all 0. Therefore, unemployee nodes become leaf nodes as well.

After the first tree is formed, the next stage is to make another tree by taking different variables using the equation formula 1 and 2 so that the trees formed in this method produce different shapes. The next stages of the tree are marital status, grade, gender, and age.

Modeling was done with the Random Forest algorithm using rapidminer as an analysis and evaluation platform (figure 3). This model will make it possible to predict new data based on existing and processed data, with the aim of achieving the performance of algorithms that have the highest accuracy.

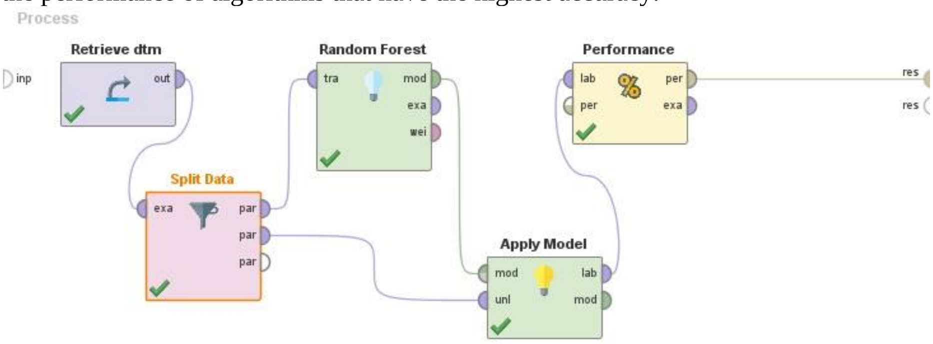

Figure 3. Modeling with Random Forest

Random Forest (RF) Accuration

To find out that the model makes predictions correctly about the timeliness of graduation according to its data class, it can be seen from the amount of accuracy produced by the model used. Table 4 describes the results of model testing of pass timeliness data using several data split techniques. The accuracy value obtained from the random forest algorithm model with the third data split technique is the same, which is 100%.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13750
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 4. RF Accuration

Training	Testing	Accuration
50%	50%	100%
40%	60%	100%
30%	70%	100%

Random Forest (RF) Precision

To understand the comparison between a positive true prediction and the amount of data that is predicted to be positive, we can observe the amount of precision produced by the model used. In table 5, the results of model testing of passable timeliness data using several data split techniques. The random forest algorithm model with the third data split technique is the same, which is 100%.

Table 5. RF Precision

Training	Testing	Presisi
50%	50%	100%
40%	60%	100%
30%	70%	100%

Random Forest (RF) Recall

To find out that the model makes predictions actually, namely predicting the timeliness of graduation according to the data class compared to the class of passing, it can be seen from the amount of recall generated by the model used. Table 6 describes the results of model testing of pass timeliness data using several data split techniques. The accuracy value for recall obtained from the random forest algorithm model with the third data split technique is also the same, which is 100%.

Table 6. RF Recall

Training	Testing	Recall
50%	50%	100%
40%	60%	100%
30%	70%	100%

DISCUSSIONS

Comparison of results contained in the confusion matrix shows a significant percentage of 100%. This can be because student data has been preprocessed, namely cleaning of missing values, duplicate data, and noise data, namely zero value. Various attributes used and also the large amount of data greatly affect the results of accuracy because the amount of data will train the model created so that better results will be obtained (Halim et al., 2023).

Comparison of confusion matrix of the three split data, can be seen in the following pictures (figure 4, figure 5, and figure 6). Figure 4 shows the first stage of the split.

accuracy: 100.00%

	true Tepat	true Terlambat	class precision
pred. Tepat	101	0	100.00%
pred. Terlambat	0	49	100.00%
class recall	100.00%	100.00%	

precision: 100.00% (positive class: Terlambat)

	true Tepat	true Terlambat	class precision
pred. Tepat	101	0	100.00%
pred. Terlambat	0	49	100.00%
class recall	100.00%	100.00%	

recall: 100.00% (positive class: Terlambat)

	true Tepat	true Terlambat	class precision
pred. Tepat	101	0	100.00%
pred. Terlambat	0	49	100.00%
class recall	100.00%	100.00%	

Figure 4. Confusion Matrix from Split Data 50/50

The results shown from the 50/50 data split in figure 4 are 100% accuracy, 100% precision, and 100% recal. Meanwhile, the division in table accuracy, table precision, and also table recall is 101 correct and 49 late. Figure 5 shows the second split stage.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13750 p-ISSN : 2541-044X

accuracy: 100.00%

	true Tepat	true Terlambat	class precision
pred. Tepat	122	0	100.00%
pred. Terlambat	0	58	100.00%
class recall	100.00%	100.00%	

precision: 100.00% (positive class: Terlambat)

	true Tepat	true Terlambat	class precision
pred. Tepat	122	0	100.00%
pred. Terlambat	0	58	100.00%
class recall	100.00%	100.00%	

precision: 100.00% (positive class: Terlambat)

	true Tepat	true Terlambat	class precision
pred. Tepat	122	0	100.00%
pred. Terlambat	0	58	100.00%
class recall	100.00%	100.00%	

Figure 5. Confusion Matrix from Split Data 40/60

The results shown from the 40/60 data split in figure 5 are 100% accuracy, 100% precision, and 100% recal. Meanwhile, the division of table accuracy, table precision, and also table recall is correct as many as 122 and late as many as 58. Figure 6 shows the third split stage.

accuracy: 100.00%

	true Tepat	true Terlambat	class precision
pred. Tepat	142	0	100.00%
pred. Terlambat	0	68	100.00%
class recall	100.00%	100.00%	

precision: 100.00% (positive class: Terlambat)

	true Tepat	true Terlambat	class precision
pred. Tepat	142	0	100.00%
pred. Terlambat	0	68	100.00%

recall: 100.00% (positive class: Terlambat)

	true Tepat	true Terlambat	class precision
pred. Tepat	142	0	100.00%
pred. Terlambat	0	68	100.00%
class recall	100.00%	100.00%	

Figure 6. Confusion Matrix from Split Data 30/70

The results shown from the 30/70 data split in figure 6 are 100% accuracy, 100% precision, and 100% recal. Meanwhile, the division of table accuracy, table precision, and also table recall is accurate as many as 142 and late as many as 68.

The yield remains 100% even though it is split three times. The analysis carried out by the author is that the data used has a simple pattern. Seen at the transformation stage, each attribute is only grouped into two categories. It is necessary to carry out data processing with random forests with diverse data, such as several attributes grouped into more than 2 attributes.

CONCLUSION

Research conducted to predict the timeliness of graduation using a random forest model provided results by applying split data or division between learning data and test data 3 times, namely 50/50, 40/60, and 30/70. To evaluate the model prepared, confusion matrix evaluation such as accuracy, recall, and precision are used. The results of the study show that the proposed model can do well with predictions, that is, with the same results for all three data splits. The test value is 100% accuracy, 100% recall, and 100% precision. With this value, the success rate for predicting the timeliness of student graduation will be more accurate.

The yield remains 100% even though it is split three times. The analysis carried out by the author is that the data used has a simple pattern. Seen at the transformation stage, each attribute is only grouped into two categories. It is necessary to carry out data processing with random forests with diverse data, such as several attributes grouped into more than 2 attributes. For a more accurate revision of the results, it is hoped that in the next research there are things that need to be tried, especially those that do not yet exist in this research, such as the accumulation of a larger number of datasets to be used as research objects so that the distribution of data can be balanced and representative. The split data method for random forests also needs to be tried for other datasets, in order to prove the accuracy results whether the same or different from this study.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13750 p-ISSN : 2541-044X

REFERENCES

Aang Kisnu Darmawan, Ivana Yudhisari, Anwari Anwari, & Masdukil Makruf. (2023). Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest. Jurnal Minfo Polgan, 12(2), 387–400.

Aria Hendrawan, Lenny Margaretta Huizen, Agusta Praba Ristadi Pinem, & Dinar Anggit Wicaksana. (2021). Implementasi Pemilihan Fitur Metode Wrapper dan Embedded dalam Prediksi Ketepatan Kelulusan Mahasiswa. Seminar Nasional Penelitian Dan Pengabdian Kepada Masyarakat (SNPPKM), 330–335.

Embun Fajar Wati, & Biktra Rudianto. (2022). Penerapan Algoritma KNN, Naive Bayes Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa. Jurnal Format, 11(2), 168–175.

Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Anggi Puspita Sari. (2023). Prediction of Student Graduation using the K-Nearest Neighbors Method. International Journal of Information System & Technology, 7(3), 211–216.

Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Anggi Puspita Sari. (2024). Improved Naive Bayes Algorithm with Particle Swarm Optimization to Predict Student Graduation. International Journal of Information System & Technology, 7(6), 386–391.

Gede Suwardika, & I Ketut Putu Suniantara. (2019). ANALISIS RANDOM FOREST PADA KLASIFIKASI CART KETIDAKTEPATAN WAKTU KELULUSAN MAHASISWA UNIVERSITAS TERBUKA. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 13(3), 179–186.

Halim, K., Erny Herwindiati, D., & Sutrisno, T. (2023). PENERAPAN METODE DECISION TREE UNTUK PRAKIRAAN CUACA KOTA BEKASI. Jurnal Ilmu Komputer Dan Sistem Informasi, 11(2). https://doi.org/10.24912/jiksi.v11i2.26026

Hermawan, L., & Bellaniar Ismiati, M. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. Jurnal Transformatika, 17(2), 188. https://doi.org/10.26623/transformatika.v17i2.1705

Hikmah Dwiyanti Nasir, Dahlia Nur, & Zawiyah Saharuna. (2020). Prediksi Tingkat Polusi Udara Dengan Data Mining. Prosiding Seminar Nasional Teknik Elektro Dan Informatika (SNTEI), 90–95.

I Made Budi Adnyana. (2015). PREDIKSI LAMA STUDI MAHASISWA DENGAN METODE RANDOM FOREST (STUDI KASUS : STIKOM BALI). CSRID Journal, 8(3), 201–208.

I Made Budi Adnyana. (2021). PENERAPAN TEKNIK KLASIFIKASI UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN NILAI AKADEMIK. Jurnal Teknologi Informasi Dan Komputer, 7(3), 480–485.

Ibnu Alfarobi, Taransa Agasya Tutupoly, & Ade Suryanto. (2018). KOMPARASI ALGORITMA C4.5, NAIVE BAYES, DAN RANDOM FOREST UNTUK KLASIFIKASI DATA KELULUSAN MAHASISWA JAKARTA. Jurnal Mitra Dan Teknologi Pendidikan, 4(1).

Indra Irawan, M Riski Qisthiano, Muhammad Syahril, & Pamuji M. Jakak. (2023). Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO. Jurnal Pengembangan Sistem Informasi Dan Informatika, 4(4), 26–36.

Jaya S. Saleh, Angelia M. Adrian, & Junaidy B. Sanger. (2022). Sistem Klasifikasi Kelulusan Mahasiswa dengan Algoritma Random Forest. Jurnal Ilmiah Realtech, 18(1), 10–14.

Juwariyem, Sriyanto, Sri Lestari, & Chairani. (2024). Prediction of Stunting in Toddlers Using Bagging and Random Forest Algorithms. Sinkron : Jurnal Dan Penelitian Teknik Informatika, 8(2), 947–955.

Mao, Y., He, Y., Liu, L., & Chen, X. (2020). Disease Classification Based on Eye Movement Features With Decision Tree and Random Forest. Frontiers in Neuroscience, 14, 1–11. https://doi.org/10.3389/fnins.2020.00798

Muhammad Labib Mu'tashim, Ati Zaidiah, & Bambang Saras Yulistiawan. (2023). Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta). Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA), 155–166.

Muhammad Sony Maulana, Raja Sabarudin, & Wahyu Nugraha. (2019). Prediksi Ketepatan Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi. JUSTIN (Jurnal Sistem Dan Teknologi Informasi), 7(3), 202–206.

Pramudita Oktaviani, Ibnu Asror, S. T. , M. T., & Dr. Moch. Arif Bijaksana. (2018). Analisis Implementasi Sistem OLAP dan Klasifikasi Ketepatan Waktu Lulus dan Undur Diri Mahasiswa Teknik Informatika Universitas Telkom Menggunakan Random forest. E-Proceeding of Engineering, 3564–3574.

reza maulana, & devy kumalasari. (2019). ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI STATUS KELULUSAN MAHASISWA AKADEMI BINA SARANA INFORMATIKA. Jurnal Informatika Kaputama (JIK), 3(2), 29–36.

Romdan Muhamad Ubaidilah, Zulfi Anugerahwati, Imaniar Ikko Mulya Rizky, & Sri Lestari. (2023). Prediksi Kelulusan Mahasiswa Berdasarkan Data Kunjung dan Peminjaman Buku menggunakan Rapid Miner dengan Metode C.45 dan Random Forest. I-Robot Jurnal, 7(2), 14–20.

Sri Diantika, Hiya Nalatissifa, Nurlaelatul Maulidah, Riki Supriyadi, & Ahmad Fauzi. (2024). Penerapan Teknik Random Oversampling Untuk Memprediksi Ketepatan Waktu Lulus Menggunakan Algoritma Random Forest. Computer Science (CO-SCIENCE), 4(1), 11–18.

Sri Diantika, Hiya Nalatissifa, Riki Supriyadi, Nurlaelatul Maulidah, & Ahmad Fauzi. (2023). Implementasi Multi-Class Gradient Boosting Untuk Mengklasifikasikan Jenis Hewan Pada Kebun Binatang. ANTIVIRUS: Jurnal Ilmiah Teknik Informatika, 7(1), 32–40.

Wati, E. F., Sari, A. P., Alawiah, E. T., Siregar, M. H., & Rudianto, B. (2021). Particle Swarm Optimization Comparison on Decision Tree and Naive Bayes for Pandemic Graduation Classification. 2nd International Conference on Advanced Information Scientific Development (ICAISD), 1–11.

Zaskila Nurfadilla, & Faisal. (2022). IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN RANDOM FOREST. AGENTS Journal of Artificial Intelligence and Data Science, 1(1), 1–8.