Jurnal Sistem Informasi Bisnis 02(2025)

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/71807

Implementation of the Ensemble Machine Learning Algorithm for Student Dropout Prediction Analysis

Winarsih[a,b,*], Heri Sutanto[c], Aris Puji Widodo[d]

[a] Doctoral Program of Information System, School of Post Graduate Studies, Diponegoro University, Jl. Imam Bardjo S.H., No. 5, Pleburan, Semarang, Indonesia 50241
[b] Department of Informatics, Faculty of Communication and Information Technology, Universitas Nasional, Jl. Sawo Manila No. 61, Jakarta, Indonesia 12520
[c] Department of Physics, Faculty of Science and Mathematics, Diponegoro University, Jl. Prof. Soedarto, S.H., Tembalang, Semarang, Indonesia 50275
[d] Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Jl. Prof. Soedarto, S.H., Tembalang, Semarang, Indonesia 50275

Submitted: February 11th, 2025; Revised: March 14th, 2025; Accepted: April 5th, 2025; Available Online: May 31st, 2025
DOI: 10.14710/vol15iss2pp159-166

Abstract

Educational Data Mining provides an effective approach to tackle numerous issues within the education sector, including the capacity to perform predictive analyses regarding student attrition based on academic information. In this research, data from the Open University Learning Analytics dataset (OULAD), which is publicly accessible, has been employed, which encompasses student information collected during online learning. We apply various Machine Learning models, including Decision Trees, Naïve Bayes, Logistic Regression, and ensemble approaches like Random Forest and AdaBoost. Among the models tested, Random Forest (RF) achieved the highest accuracy of 89.37%, along with a precision of 89.57% and a recall of 93.86%, using the data splitting approach. When employing an alternative evaluation model, specifically K-Fold Cross Validation, the maximum F1 score achieved was 9.45%. In summary, the ensemble machine learning algorithm, specifically Random Forest (RF), exhibited strong performance in predicting student academic achievement quality.

Keywords : OULAD; education data mining; random forest; decision trees; naïve bayes; logistic regression.

1. Intruduction

Educational institutions have a vital role in nurturing and enhancing the potential of students with the aim of shaping individuals who are creative, ethical, knowledgeable, and responsible. These institutions also contribute to maintaining students' performance in achieving these goals. Student academic performance is a critical factor in assessing whether the educational objectives have been met, typically evaluated through the students' Grade Point Average (GPA) and on-time graduation. Actively engaging students during the learning process significantly influences the improvement of academic quality within an educational institution. The research conducted aims to identify elements that can impact the quality of student learning performance (Hameed & Akhtar, 2021). The objective is to identify the most effective techniques for analyzing patterns and deriving conclusions that can improve the quality of student learning and facilitate the prediction of students who may be at risk of discontinuing their education. (Alhothali et al., 2022).

Various methods must be employed to maintain the quality of student learning performance in good condition. These methods include periodic evaluation of the learning process, objective assessments conducted by educators, and student learning monitoring carried out by the education providers' internal teams. One of the issues that can be addressed through educational data processing is predicting student graduation, which determines each student's status in each subject as either "dropout" or "non-dropout." The prediction results are derived from the analysis and evaluation of student learning over one semester. Student data is processed to create a dataset that can be used for modeling using Machine Learning algorithms. Student data is extracted and patterns are formed using a model that links students' habits, behaviors during the learning process, family circumstances, and other factors influencing their learning outcomes. This data is then compiled into a dataset, with students classified as either "dropouts" or "non-dropouts." These values function as attributes for analysis through Machine Learning algorithms. The prediction of student graduation is expected to serve as a benchmark for students to enhance their academic performance. Additionally, forecasting students' academic performance can act as an early indicator, enabling education providers to devise optimal strategies for

*) Corresponding author: winarsih@students.undip.ac.id

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/71807

enhancing the learning process. (Bagunaid et al., 2022) .

2. Literature Review

Within the scope of educational data analysis, and specifically in the field of Educational Data Mining (EDM), various analytical techniques are employed to tackle the challenges and issues present in education. EDM is a subset of Data Mining employed to analyze data in the context of education. One of the most common topics related to applying EDM is the identification of students' successes and failures in their educational journey. This entails forecasting final grades and identifying students who are at risk of leaving school. The ability to predict student performance and address the challenges they encounter is a vital and highly beneficial practice, advantageous to both students and educational institutions. These insights can be leveraged to enhance the quality of education by providing intensive academic guidance by educators, gaining a detailed understanding of students' potentials and shortcomings, and addressing individual student issues that significantly impact the educational process. When properly planned and executed, this can aid educational institutions in improving the learning process and formulating effective strategies. Several research initiatives have been conducted to evaluate the effectiveness of Educational Data Mining in this context (Rodríguez-Hernández et al., 2021), which are:

Research conducted by Ali Al-Zawqari et al. (2023), titled 'A Flexible Feature Selection Approach for Predicting Student Academic Performance in Online Courses,' focuses on predicting student performance using four criteria: Pass-Fail, Distinction-Pass, Distinction-Fail, and Withdraw-Pass. The study compares feature dimensionality reduction using all features alongside various machine learning and deep learning algorithms (Al-Zawqari et al., 2022). The second research their study titled 'Artificial Neural Networks in Academic Performance Prediction: Systematic Implementation and Predictor Evaluation,' Carlos Felipe R. and his team aimed to explore how neural networks can be systematically applied to predict the quality of student academic achievement in higher education institutions (Rodríguez-Hernández et al., 2021). The third research conducted by Vito Renò and his team, as detailed in 'Learning Analytics: Analysis of Methods for Online Assessment,' aims to assess online learning methodologies and their effectiveness through a binary evaluation approach, categorizing student outcomes as either passing or failing (Renò et al., 2022). This research leverages a dataset made publicly available from the Open University Learning Analytics (OULAD), The data set covers aspects of student engagement in online courses and records their activities and interactions within the Virtual Learning Environment (VLE), information about students in utilizing learning materials provided by the institution. education, detailed student biodata information, and exam results obtained from various tests taken by students, research using OULAD data will predict student involvement in the online learning process whether it has a big influence on determining the status of each student in each subject, namely "drop out " or "didn't drop out" (Renò et al., 2022).

3. Research Method

In this study, the data mining process is structured using the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, where there are several steps that must be taken to obtain evaluation results related to the student's academic achievement status, namely determining whether the student "dropped out" or "did not drop out" (Barros et al., 2019). Six stages of Data Mining development used in research methodology are shown in Figure 1.

3.1. Bussines Understanding

There are several methods that need to be employed to enhance student learning performance, primarily through the implementation of regular assessments of the learning process. The objective of this study is to scrutinize student forecasts for each subject, determining whether students are categorized as "dropout" or "non-dropout." The objective of this research is to determine the extent to which certain factors influence student dropout rates and to evaluate how student engagement with course content and participation in class might affect their overall grades. Research endeavors investigating the correlation between students' ultimate grades and

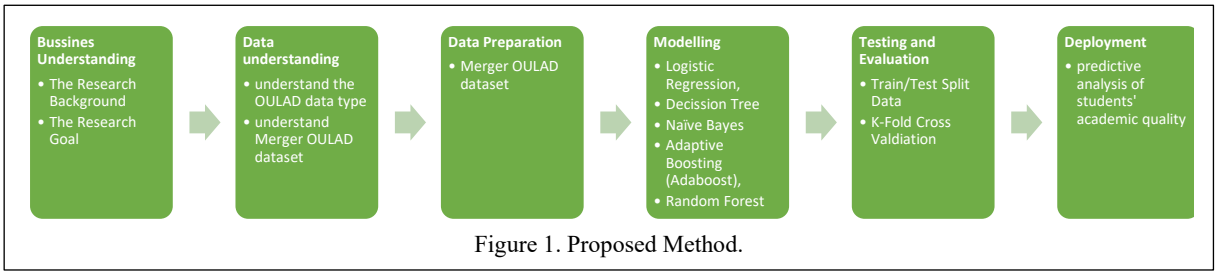

Figure 1. Proposed Method.

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/71807

their engagement in the course material employed machine learning algorithms, ultimately establishing that students who displayed a high level of engagement in quizzes and the material tended to achieve higher grades in the final examination (Jawad et al., 2022).

3.2. Data Understanding

The Open University Learning Analytics dataset (OULAD) is used in this analysis. It comprises detailed information on students collected through their online learning sessions and known for having the largest undergraduate student enrollment in the UK. Since its establishment in 1969, it has emerged as the largest academic institution in the UK and is also among the largest in Europe, having enrolled 2 million students throughout its history. As can be understood from the name, the Open University is primarily populated by off-campus students. This dataset features data on students' online learning experiences, capturing their activities and interactions within the Virtual Learning Environment (VLE), information on students using learning materials provided by educational institutions, detailed information on student biodata, and exam results from various types of tests. done by students, Each class has more than 500 students, Each course offered by this university includes two assessments per semester, with records kept on how many students succeed or fail in these exams. Student learning is assessed through a virtual learning environment and encompasses seven educational programs, Out of these, four are specifically concentrated on the areas of Science, Technology, Engineering, and Mathematics (STEM), The other three areas are dedicated to Social Sciences. The dataset records multiple offerings of each course across different semesters, where each offering is called a 'presentation,' identified by the years 2013 or 2014 and semesters 'B' or 'J,' commencing in February or October. It includes data from a total of 32,593 students, with each student linked to four distinct categories of information within OULAD (Renò et al., 2022). The dataset student infor form OULAD is listed in Table 1.

3.3. Data Preparation

The dataset that has been obtained will be analyzed to see if there are missing values or null data, the null data will be input automatically (imputation) using the fill method by replacing the null column with the value in the previous or next column, then the scattered dataset will be merged into one type of dataset. "The dataset comprises various data types: nominal data, which includes attributes such as gender, region, disability, starting month, code module, and code category, and ordinal data, which includes elements like highest educational attainment, IMD band, and age range, these features will be encoded using the method one hot encoding and label encoding, the labels available in this dataset are Pass, Distinction, Fail, and drop out, the dataset will be made into two labels, namely drop out and not drop out, drop out is taken from the label drop out, not drop out taken from the labels Pass, distinction and fail (Rodríguez-Hernández et al., 2021).

3.4. Modeling

Categorization at this stage will leverage both classic machine learning algorithms and sophisticated ensemble methods. Prediction analysis is conducted by identifying distinguishing patterns between hate speech and non-hate speech through the assessment of label similarities.(Daza Vergaray et al., 2023).

3.4.1. Logistic Regression

The classification algorithm is employed to ascertain if there is a connection between discrete or

Table 1. Dataset student info from OULAD

Code Module	Code Presentation	Id	Sex	Region	Highest Education	Imd Band	age_band	Attempt	Credit	Disability	Final Result
AAA	2013J	11391	M	East Anglian Region	HE Qualification	90-100%	55<=	0	240	N	Pass
AAA	2013J	28400	F	Scotland	HE Qualification	20-30%	35-55	0	60	N	Pass
AAA	2013J	30268	F	North Western Region	A Level or Equivalent	30-40%	35-55	0	60	Y	Withdrawn
AAA	2013J	31604	F	South East Region	A Level or Equivalent	50-60%	35-55	0	60	N	Pass
AAA	2013J	32885	F	West Midlands Region	Lower Than A Level	50-60%	0-35	0	60	N	Pass
...	...	...	...	...	...	...	...	...	...	...	...
GGG	2014J	2640965	F	Wales	Lower Than A Level	Oct-20	0-35	0	30	N	Fail
GGG	2014J	2645731	F	East Anglian Region	Lower Than A Level	40-50%	35-55	0	30	N	Distinction
GGG	2014J	2648187	F	South Region	A Level or Equivalent	20-30%	0-35	0	30	Y	Pass
GGG	2014J	2679821	F	South East Region	Lower Than A Level	90-100%	35-55	0	30	N	Withdrawn

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/71807

continuous features and the likelihood of discrete output outcomes. Logistic regression can be categorized into two primary types: single logistic regression, which utilizes one input variable, and multiple logistic regression, which incorporates multiple input variables. The f (Ika Alfina, Rio Mulia, Mohamad Ivan Fanany, 1999). Logistic Regression is calculated using Equation 1.

$$In\left(\frac{\acute{P}}{1-\acute{P}}\right) = B_0 + B_1X \quad (1)$$

where

In : natural logarithm
$B_0 + B_1X$: equation OLS
$\acute{P}$: logistic probability for binary classification.

3.4.2. Decision Tree (DT)

The calculations in this algorithm use a decision tree structure by modeling several possibilities so that an alternative is found to solve the problem. Decision trees are able to eliminate data if it has no connection to the calculation. (Khanday et al., 2022). Decision tree algorithm is expressed in Equation 2.

$$E(S) = \sum_{i=1}^{C} - P_i log_2 P_i \quad (2)$$

where S is the initial condition, i is a set of classes in S, namely yes and no, P_i is the probability of an event S.

3.4.3. Naïve Bayes (NB)

Classification algorithm to calculate probability values by adding up the frequencies and combinations of values in the data. Naïve Bayes estimates class probabilities by evaluating the data's features (Taamneh et al., 2023). Referred to as 'naïve,' this probabilistic classifier assumes that the presence of one feature does not influence the presence of others. (Khanday et al., 2022). The Naïve Bayes (NB) formula is presented in Equation 3.

$$P(x) = \frac{P(c)\,P(c)}{P(x)} \quad (3)$$

where P(c|x) is class probability estimated by the predictor. P(c) is probability of the class based on earlier observations. P(x|c) is likelihood of a class as predicted by the model. P(x) is initial probability assigned by the predictor before new data is considered.

3.4.4. Random Forest (RF)

The tree structure resulting from this algorithm is very complex for data that has many variables.(Sawangarreerak & Thanathamathee, 2020). Random forest is a collection of classifiers in the form of trees {h(**x**, θ k), k = 1, . . .} where *θk* The method involves using a distributed vector, where each decision tree selects the most prevalent category present in the input *X* (Jawad et al., 2022).

(1)

3.4.5. Ensemble Machine Learning

Ensemble techniques involve the use of multiple machine learning algorithms to perform classification tasks, with the objective of achieving superior accuracy compared to the use of a single algorithm. Key types of ensemble strategies are bagging, boosting, and stacking.(Khanday et al., 2022).

3.4.6. Adaptive Boosting (Adaboost)

Adaboost, or Adaptive Boosting, represents an algorithm used in ensemble learning techniques in machine learning, leveraging the principle of boosting. This algorithm boosts the classification performance of a machine learning model, resulting in a more durable and effective classifier, which is subsequently designated as a weak learner (Tsai & Hung, 2021).

(2)w

3.5. Testing and Evaluation

The assessment strategy includes the train/test split technique and K-Fold Cross Validation. The train/test split method breaks the dataset into training and testing segments. The dataset is divided using an 80:20 ratio, where 80% is allocated to the training set and 20% to the testing set. Conversely, the K Fold Cross Validation approach generates a singular group of data designated for testing purposes. This specific dataset will be utilized throughout the testing phase according to the defined values... K (Mastour et al., 2023).

3.6. Deployment

At this stage, a report or data mining implementation process will be carried out with the aim of providing an overview of the conclusions of the data mining algorithm calculations.

4. Results And Analysis

The results obtained from the encoding process using the one hot encoding method are as shown in Figure 2.

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/71807

```
# Function to Encode the Categorical Data Columns
# Parameters 1. df :- Dataframe 2. Column_name :- Feature to encode

def categorical_encoding(df, column_name_list=[]):

    for column_name in column_name_list:
        print(df[column_name].unique())
        categorical_columns = pd.get_dummies(df[column_name], prefix = column_name,
                                             prefix_sep = '_', drop_first = False)
        df = pd.concat([df, categorical_columns], axis = 1)
        df = df.drop(column_name, axis = 1)
    return df


# Function to Label Encode the data
# Parameters :- 1. data :- Dataframe 2. columns_list :- List of columns to label encode

def labelEncoder(data, columns_list):
    for col in columns_list:
        encoder = LabelEncoder()
        data[col]  = encoder.fit_transform(data[col])
    return data
```

Figure 2. Function used to perform encoding

In this dataset, there is a presence of categorical data, which refers to data characterized by words or non-measurable numbers. Examples of nominal data types include gender, region, disability, Starting_Month, code_module, Code_Category, while ordinal data types include features like highest_education, imd_band, and age_band. The one-hot encoding method will be applied to the highest_education, imd_band, and age_band features using the function mentioned above, resulting in a dataset similar to the one presentedin Table 2.

Table 2. The encoding results use the one hot encoding method

region_North Region	region_North Western Region	region_Scotland	region_South East Region	region_South Region	region_South West Region	region_Wales	region_West Midlands Region	region_Yorkshire Region	disability_N	disability_Y	Starting_Month_February
0	0	0	0	0	0	0	0	0	1	0	0
0	0	1	0	0	0	0	0	0	1	0	0
0	1	0	0	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	1	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...
0	0	0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	1	0	0

As a result of the one hot encoding method above, the data will be changed from several columns consisting of each parameter value. The appropriate parameters will be labeled 1 and those that do not correspond will be labeled 0. The encoding results use the Label Encoding method is shown in Table 3.

As a result of the Label Encoding method above, in the label encoding we use the number 0 for the amount of data in the feature. For example, the age-band feature has several types of data, namely 55<=, 0-35, 35-55, etc. which will be numeric values such as 1,2,3,4, etc.

Table 3. The encoding results use the Label Encoding method

6	imd_band	age_band
1	9	2
1	2	1
0	3	1
0	5	1
0	5	0
...	...	...
0	1	0
0	4	1
0	2	0
0	9	1
1	5	1

The provided content pertains to a traditional machine learning algorithm model and a machine learning ensemble implemented in a Python application. In the realm of traditional machine learning algorithms, this study populates algorithm parameters with null or empty values. However, in the case of the ensemble algorithm, parameter values are provided for each individual algorithm, with the intent of enhancing predictive performance, as shown in Table 4.

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/71807

Table 4. Traditional and Ensemble Machine Learning Algorithm Models

LGREGG	LogisticRegression()
RF	RandomForestClassifier()
DT	DecisionTreeClassifier()
NB	GaussianNB()
adaboost	AdaBoostClassifier(n_estimators=100)
Lgregg	LogisticRegression()

The results provided reflect the Accuracy, Precision, Recall, and F1 scores calculated from a machine learning algorithm, employing a data split strategy where 80% of the dataset was used for training and the remaining 20% for testing purposes. Following this, a retesting procedure was conducted employing the K-Fold Cross Validation model with K set to 10.

Upon examining the accuracy values presented earlier, it was observed that The Random Forest (RF) algorithm recorded an accuracy of 89.37% using the data splitting method. However, with K-Fold Cross Validation, it achieved its best performance with an accuracy of 88.75% is shown in Figure 3.

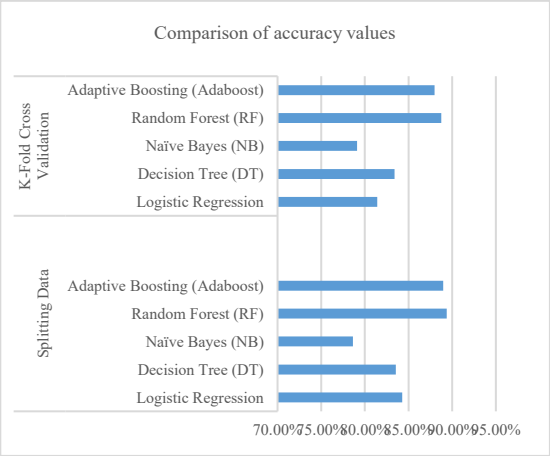

Figure 3. Comparison of Accuracy Values Using Splitting and K-Fold Cross Validation Methods

It can be verified that Random Forest (RF) algorithm achieved a precision rate of 89.57% when utilizing the data splitting method. However, when the K-Fold Cross Validation method was employed, the Random Forest (RF) algorithm attained the highest precision, specifically 88.75% as shown in Figure 4.

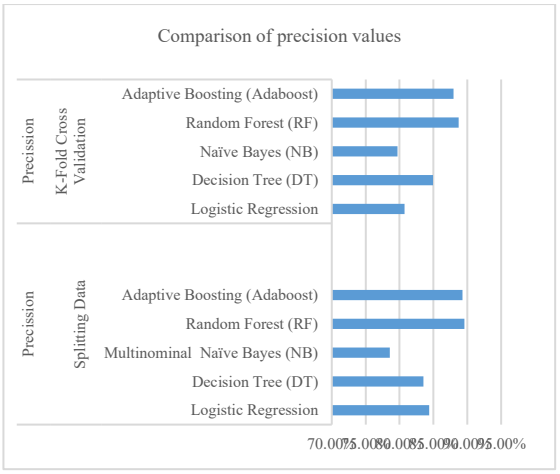

Figure 4. Comparison of Precision Values Using Splitting and K-Fold Cross Validation Methods

However, by using the data splitting method, the Random Forest (RF) algorithm delivered an accuracy of 89.62%, whereas the K-Fold Cross Validation method provided the highest accuracy at 93.86% as shown in Figure 5.

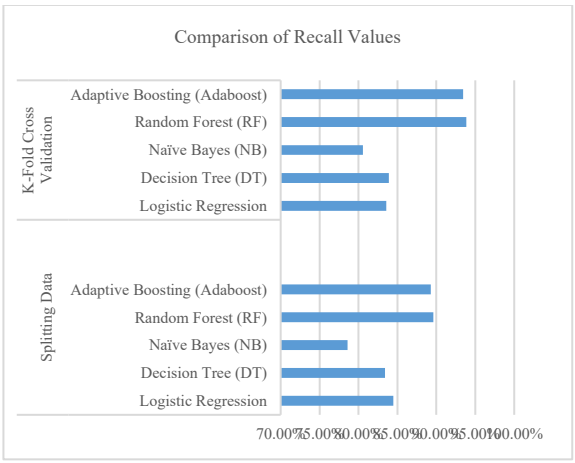

Figure 5. Comparison of Recall Values Using Splitting and K-Fold Cross Validation Methods

When employing the data splitting method, the Random Forest (RF) algorithm shows an accuracy rate of 89.37%. In contrast, this algorithm achieves its best performance with K-Fold Cross Validation, reaching an accuracy of 94.45% as shown in Figure 6.

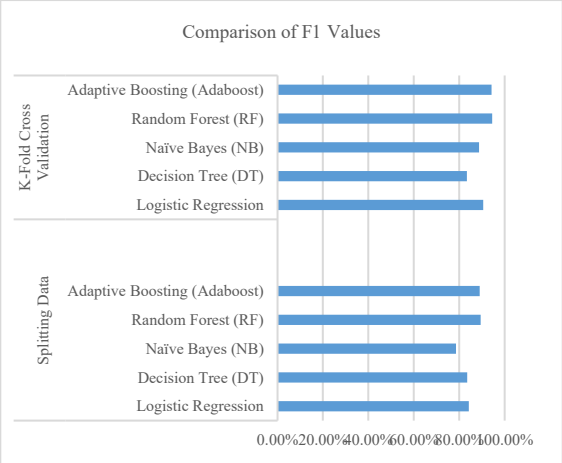

Figure 6. Comparison of F1 Values Using Splitting and K-Fold Cross Validation Methods

5. Conclusion

This research focuses on predictive analysis to measure the risk of students withdrawing from their studies, by utilizing the Open University Learning Analytics (OULAD) dataset. The data set was processed by categorical feature coding, with one-hot coding used for nominal categories and label coding for ordinal categories. Analysis results involving machine learning models with five algorithms—Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB), and ensemble methods such as Random Forest (RF) and Adaboost—show that Random Forest (RF) outperforms the algorithm others in terms of accuracy, precision, recall, and F1 score. The data split approach provides superior performance, achieving 89.37% accuracy, 89.57% precision, and 93.86% recall, in contrast to the K-Fold Cross Validation method. Meanwhile, the highest F1 score is obtained using the K-Fold Cross Validation method, reaching 94.45%. In conclusion, it can be inferred that the ensemble machine learning algorithm, specifically Random Forest (RF), exhibits strong predictive performance when applied to this dataset.

References

Al-Zawqari, A., Peumans, D., & Vandersteen, G. (2022). A flexible feature selection approach for predicting students' academic performance in online courses. *Computers and Education: Artificial Intelligence*, *3*(November), 100103. https://doi.org/10.1016/j.caeai.2022.100103

Alhothali, A., Albsisi, M., Assalahi, H., & Aldosemani, T. (2022). Predicting Student Outcomes in Online Courses Using Machine Learning Techniques: A Review. *Sustainability (Switzerland)*, *14*(10), 1–23. https://doi.org/10.3390/su14106199

Bagunaid, W., Chilamkurti, N., & Veeraraghavan, P. (2022). AISAR: Artificial Intelligence-Based Student Assessment and Recommendation System for E-Learning in Big Data. *Sustainability (Switzerland)*, *14*(17). https://doi.org/10.3390/su141710551

Barros, T. M., Neto, P. A. S., Silva, I., & Guedes, L. A. (2019). Predictive models for imbalanced data: A school dropout perspective. *Education Sciences*, *9*(4). https://doi.org/10.3390/educsci9040275

Daza Vergaray, A., Miranda, J. C. H., Cornelio, J. B., López Carranza, A. R., & Ponce Sánchez, C. F. (2023). Predicting the depression in university students using stacking ensemble techniques over oversampling method. *Informatics in Medicine Unlocked*, *41*(June). https://doi.org/10.1016/j.imu.2023.101295

Hameed, M., & Akhtar, N. (2021). Student Performance Prediction in Intelligent E-Learning for Tertiary Education How to Cite: Mustafa Hameed and Nadeem Akhtar (2021). Student Performance Prediction in Intelligent E-Learning for Tertiary Education. International Journal of Computational I. *International Journal of Computational Intelligence in Control*, *13*(2), 293–299.

Ika Alfina, Rio Mulia, Mohamad Ivan Fanany, Y. E. (1999). *Hate Speech Detection in the Indonesian Language: A Dataset and Preliminary Study*. 473–481.

Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. *Sustainability (Switzerland)*, *14*(22). https://doi.org/10.3390/su142214795

Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., & Malik, S. H. (2022). Detecting twitter hate speech in COVID-19 era using machine learning and ensemble learning techniques. *International Journal of Information Management Data Insights*, *2*(2), 100120. https://doi.org/10.1016/j.jjimei.2022.100120

Mastour, H., Dehghani, T., Moradi, E., & Eslami, S. (2023). Early prediction of medical students' performance in high-stakes examinations using machine learning approaches. *Heliyon*, *9*(7), e18248. https://doi.org/10.1016/j.heliyon.2023.e18248

Renò, V., Stella, E., Patruno, C., Capurso, A., Dimauro, G., & Maglietta, R. (2022). Learning

Analytics: Analysis of Methods for Online Assessment. *Applied Sciences (Switzerland)*, *12*(18), 1–10. https://doi.org/10.3390/app12189296

Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, *2*(December 2020). https://doi.org/10.1016/j.caeai.2021.100018

Sawangarreerak, S., & Thanathamathee, P. (2020). Random forest with sampling techniques for handling imbalanced prediction of university student depression. *Information (Switzerland)*, *11*(11), 1–13. https://doi.org/10.3390/info11110519

Taamneh, M. M., Taamneh, S., Alomari, A. H., & Abuaddous, M. (2023). Analyzing the Effectiveness of Imbalanced Data Handling Techniques in Predicting Driver Phone Use. *Sustainability (Switzerland)*, *15*(13). https://doi.org/10.3390/su151310668

Tsai, J. K., & Hung, C. H. (2021). Improving adaboost classifier to predict enterprise performance after covid-19. *Mathematics*, *9*(18), 1–10. https://doi.org/10.3390/math9182215