Prediction of Grade Point Average (GPA) for Students at Informatics and Computer Engineering Education – Universitas Negeri Jakarta during Online Learning Using Naive Bayes Algorithm

Miftahul Jannah[1]*, Widodo[2], Hamidillah Ajie[3]
Informatics and Computer Engineering Education Study Program, Faculty of Engineering [1], [2], [3]
Universitas Negeri Jakarta
Jakarta, Indonesia
miftahuljnnh294@gmail.com[1], widodo@unj.ac.id [2], hamidillah@unj.ac.id [3]

***Abstract*— The transition of learning models from face-to-face to online learning has had several impacts on student learning, reflected in their academic achievements. This study aims to determine the performance of the algorithm model using data mining classification techniques in predicting the Semester Grade Point Average (GPA) of Informatics and Computer Engineering Education students, at Universitas Negeri Jakarta during online learning. The prediction employed the Naive Bayes algorithm and the dataset obtained by collecting questionnaires from 2020 and 2021 batches. The total data obtained is 155 records with 13 (thirteen) attributes in the form of 1 (one) ID attribute including NIM, 11 (eleven) regular attributes including gender, college entrance, smartphone facilities, network conditions, preferred online applications, interest in learning, learning attitudes, learning creativity, parental support, study groups, and other activities outside of lectures during online learning, and 1 (one) the label attribute namely the Semester Grade Point Average for students in 3rd and 5th semester. The evaluation of this research involved the confusion matrix and the ROC (Receiver Operating Characteristic) curve. Confusion matrix resulted in an accuracy of 75%, precision of 28.33%, and recall of 26.43%. The ROC curve resulted in an AUC value of 0.679, indicating the category of poor classification. This study also applied the SMOTE data balancing technique, leading to a confusion matrix evaluation with 88.46% accuracy, 57.43% precision, and 52.14% recall. Furthermore, the ROC curve resulted in an AUC value of 0.809 which is categorized as a Good classification.**

Keywords*— Prediction, Data Mining, Naive Bayes, Online Learning, Grade Point Average

I. INTRODUCTION

The COVID-19 pandemic that has affected Indonesia since early 2020 has impacted several sectors of community life, including education. The transition from face-to-face to online learning has had several effects on its implementation. Several studies have found that online learning has a positive impact on the improvement of students' academic achievements. These studies explain that students' academic performance during online learning tends to be higher than before the implementation of online learning [1]. Therefore, online learning can be a promising instructional model to enhance academic performance [2]. The academic achievement of students during online learning is reflected by their Grade Point Average (GPA). Academic performance in this research is assessed based on the Semester Grade Point Average (GPA).

The Semester Grade Point Average (GPA) provides an evaluation of students' learning outcomes for a specific semester, and it is accessible at the end of that semester. The GPA serves as a reference in calculating the achievement of students' learning outcomes at the end of their studies, expressed in the Cumulative Grade Point Average (CGPA). The students' CGPA is a crucial indicator used to determine graduation and serves as a measure of the quality of an educational institution [3]. The importance of the CGPA encourages students to obtain the best possible GPA. Many factors can influence the acquisition of students' GPAs, including factors originating from the students' internal aspects, such as physiological and psychological factors [4]. The shift in the learning paradigm from face-to-face to online has affected students' ability to adapt physically, psychologically, and socially. Physical, psychological, and social factors influence students' ability to adapt during online learning [5]. These three factors are significant components that influence student learning outcomes in their final evaluation in the form of GPA. However, this research only focuses on examining the psychological influence. Investigating the psychological aspects of predicting social studies is beneficial for students in determining their future graduation and valuable for supervisors in providing assistance and planning their studies [6].

Psychological factors can be a reference for predicting students' GPA during online learning because they influence student's academic performance, as reflected in the Semester Grade Point Average (GPA). Academic achievements can be impacted by interest, motivation, time management skills, family relationships, living conditions, social conditions, students' ability to adapt to the learning environment, and the teaching methods of instructors [5].

Classification is widely used in making predictions. One

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i1.1958, Copyright ©2024
Submitted : September 12, 2023, Revised : January 30, 2024, Accepted : January 31, 2024, Published : February 15, 2024

frequently employed algorithm for classification is the Naive Bayes algorithm. The Naive Bayes algorithm is a simple classification algorithm that has the advantage of handling data with irrelevant attributes and producing high accuracy [7].

Based on the background, this research applied prediction using the classification data mining technique with the Naive Bayes algorithm to predict the Grade Point Average (GPA) of students during online learning. This study predicted the semester Grade Point Average using psychological factors that impact students' learning outcomes during online learning. The prediction utilizes data from students of Universitas Negeri Jakarta, majoring in Informatics and Computer Engineering Education, from the 2020 cohort who attended online lectures in the 5th semester and the 2021 cohort who attended online lectures in the 3rd semester. The data collected includes 13 attributes, consisting of 1 (one) ID attribute, namely the student ID (NIM), 11 (eleven) regular attributes, namely gender, admission pathway, smartphone facilities, network conditions in their region, preferred online learning applications, interest in learning during online learning, learning attitude, learning creativity, parental support, study groups (online discussions), and other activities outside of lectures during online learning. Additionally, there is 1 (one) class label attribute, which is the Grade Point Average (GPA) of the students in the 3rd and 5th semesters [5]. The results of this research are expected to accurately predict the Grade Point Average (GPA) of students during online learning, which can contribute to assisting students, mentors, universities, and decision-makers in making better policies for future online learning.

II. LITERATURE REVIEW

A. Naive Bayes

The Naive Bayes Classifier algorithm, commonly known as Naive Bayes, is one of the classification algorithms in data mining. Naive Bayes utilizes statistical principles in making predictions through simple probabilistic theory with the assumption of strong attribute independence (naive). Implication of the presence or absence of other features in the data is not correlated with that feature [8]. The classification process with the Naive Bayes algorithm begins by training the dataset and testing it against the dataset to be examined. The advantages of the Naive Bayes algorithm are that it can use little training data to estimate the parameters needed for classification, is fast and space efficient, and is tough against irrelevant attributes [9].

B. Information Gain

Information gain is a measure of how effective an attribute is in classification. Information gain is a method for attribute selection in the classification process by assigning weight values to each attribute. Information gain is based on the concept of entropy to identify the best attribute. Entropy measures uncertainty, meaning that the higher the entropy, the higher the uncertainty [10]. Information gain is widely used in feature selection due to its fast nature [11]. Feature selection is performed to optimize classification performance by removing attributes that are not relevant to the classification results, thereby improving classification performance and increasing model accuracy.

C. Confusion Matrix

The Confusion Matrix is an evaluation method commonly used to calculate the accuracy of data mining models [12]. The Confusion Matrix evaluates the classification model by categorizing data into true or false. A matrix of prediction results compared with the actual class, which is the input of the actual values [13]. The Confusion Matrix performs calculations with 4 (four) outputs, including accuracy, precision, and recall. Accuracy is the ratio of correctly identified cases to the total number of cases. Precision is the ratio of true positive cases, and recall is the ratio of true positive cases [12].

D. ROC Curve

The ROC curve is one type of performance metric for classification techniques. The ROC curve can be used to measure the accuracy of a diagnostic system or the predictive performance of a model [14].

The calculation method for the area under the ROC curve is referred to as the Area Under the Curve (AUC). AUC is the area under the ROC curve and serves as a measure of diagnostic test accuracy and model prediction performance [14]. AUC values always range between 0.0 and 1.0. If the resulting AUC is <0.5, then the evaluated classification model has low accuracy and is identified as a very poor model. The larger the area, the better the classification value [15].

For data mining classification, AUC values are divided as follows:

a. 0.90 – 1.00 = Very good classification
b. 0.80 – 0.90 = Good classification
c. 0.70 – 0.80 = Adequate classification
d. 0.60 – 0.70 = Poor classification
e. 0.50 – 0.60 = Incorrect classification

E. SMOTE (*Synthetic Minority Oversampling Technique*)

SMOTE is one of the widely used oversampling methods to address imbalanced data in data mining. SMOTE works by adding synthetic data to the minority class and balancing it with the majority class. The advantage of using SMOTE is that it does not cause information loss and can improve the prediction accuracy of the minority class. However, the disadvantage of SMOTE is the occurrence of excessive generation (overgeneralization), which may lead to synthetic data from SMOTE spreading into both majority and minority-class regions. Result in a decrease in classifier performance, leading to low prediction accuracy [16].

This research is based on relevant studies that predict the Cumulative Grade Point Average (CGPA) using the Naive Bayes algorithm in its modeling. The difference lies in the attributes used in the relevant study, which are the demographic factors of students. The study concluded that the prediction model accuracy with the Naive Bayes algorithm was 74.47% [17].

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i1.1958, Copyright ©2024
Submitted : September 12, 2023, Revised : January 30, 2024, Accepted : January 31, 2024, Published : February 15, 2024

III. METHODOLOGY

This research goes through several stages starting from data collection, data preprocessing, attribute selection, data cleaning, data balancing, modeling using the Naive Bayes algorithm, and evaluation.

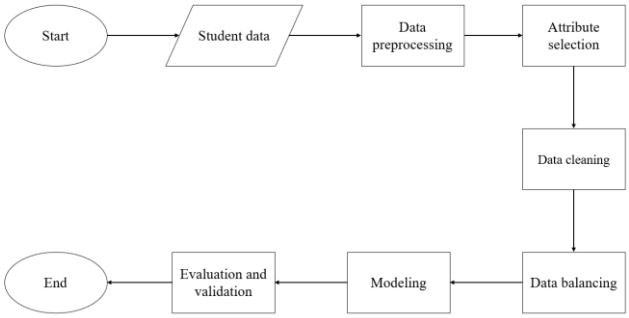

Fig. 1 Research Structure

A. Data Collection

The data used in this research are the results of data collection using a questionnaire. The questionnaire was distributed to relevant parties in the study, namely students of Informatics and Computer Engineering Education at Universitas Negeri Jakarta from the 2020 and 2021 cohorts. The questionnaire ditribution was conducted online using the Google Form platform. This research also collects data by searching, analyzing, and extracting information from relevant research as needed. Literature sources for this research include reference books and relevant scientific journals.

B. Data Preprocessing

The research begins with organizing the student data obtained to ensure it aligns with the required format. The questionnaire data was collected through Google Forms organized using Microsoft Excel. Save the file in the .xlsx format. Additionally, data type declarations and roles are define in the RapidMiner application to process the data for generating predictions and algorithm accuracy results.

C. Attribute Selection

This stage involves the selection of attributes to be used in predictive modeling. Attribute selection aimed to obtain relevant attributes. In this study, the attributes to be used are 13 attributes, including 1 (one) ID attribute, namely the student ID (NIM), and 11 (eleven) regular attributes, namely gender, admission pathway, smartphone facilities, network conditions in their area, preferred online learning applications, interest in learning during online learning, learning attitude, learning creativity, parental support, study groups (online discussions), and other activities outside of lectures during online learning. Additionally, there is 1 (one) class label attribute, which is the Grade Point Average (GPA) of students in the 3rd and 5th semester.

Attribute selection is carried out for 10 regular attributes by calculating the weight or influence of each attribute using the Information Gain operator in RapidMiner to determine the relevant attributes. The relevant attributes are then used in the subsequent stages of the research.

D. Data Cleaning

The next stage is data cleaning. Unclean data refers to data containing impurities in the form of missing values [18]. Applying data mining to dirty data can yield less accurate results in the analysis process. Data cleaning is carried out to ensure that the prediction process produces better accuracy and prevents errors. Data cleaning is done by filling in or removing missing values [10].

E. Data Balancing

Data balancing is a stage of balancing data when each class is indicated to be imbalanced. The data balancing process involves using one of the data balancing techniques, namely SMOTE. In the RapidMiner application, you can use an operator called SMOTE Upsampling. SMOTE Upsampling increases the size of the minority class to be equivalent to the majority class by generating new data in the form of synthetic data.

F. Modeling

In this stage, predictive modeling is carried out using the Naive Bayes algorithm in RapidMiner. In this study, the data is divided into 2 parts, with 70% as training data and 30% as testing data. The data division was performed using the Split Data operator in RapidMiner.

G. Evaluation and Validation

This stage evaluates the prediction results obtained from applying the Naive Bayes method in the classification process. Evaluation is conducted using the Confusion Matrix and ROC curve (Receiver Operating Characteristic). Performance values to be used include accuracy, precision, recall, and the AUC value from the ROC curve, allowing the determination of the accuracy of the model built to predict the students' semester Grade Point Average during online learning. The higher the performance values, the better the performance of the generated prediction model.

IV. RESULT AND ANALYSIS

This section will explain the results of the conducted research, starting from the data collection process, data preprocessing, attribute selection, data cleaning, data balancing, modeling using the Naive Bayes algorithm, and evaluation.

A. Data Collection

The initial stage in the research is organizing and arranging the collected data from the distributed questionnaires. The total number of data records obtained from the questionnaire survey is 155 respondents, with a breakdown of data based on the Semester Grade Point Average (GPA) categories including 125 records for GPA > 3.50; 25 records for GPA 2.75 – 3.50; 3 records for GPA 2.00 – 2.75; and 2 records for GPA < 2.00. The data is then downloaded and saved in Excel format (.xlsx). Data attributes that are not needed for the research are removed, including their columns. The attribute names, which were originally in the form of questions to facilitate respondents in completing the survey, are changed to shorter names for easier

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i1.1958, Copyright ©2024
Submitted : September 12, 2023, Revised : January 30, 2024, Accepted : January 31, 2024, Published : February 15, 2024

understanding during data processing in RapidMiner.

In the RapidMiner application, the data types and roles of each data are declared to be processed for generating predictions and accuracy. The declaration of data types and roles is done by first importing the data into RapidMiner. The data types used in this study are polynomial and binomial. Meanwhile, the roles used are ID and Label roles. The names of the attributes, data types, and roles are further detailed in Table I.

TABLE I. DATA TYPE AND ROLE

No.	Name	Data Type	Role
1.	NIM	Polynominal	ID
2.	Gender	Binominal	Regular
3.	Admission pathway	Polynominal	Regular
4.	Adequate smartphone facilities	Binominal	Regular
5.	Network conditions in their area	Polynominal	Regular
6.	Preferred online learning applications	Binominal	Regular
7.	Interest in learning during online learning	Polynominal	Regular
8.	Learning attitude	Binominal	Regular
9.	Learning Creativity	Polynominal	Regular
10.	Parental support	Polynominal	Regular
11.	Study groups (online discussions)	Binominal	Regular
12.	Other activities outside of lectures during online learning (organizations, work, etc.)	Polynominal	Regular
13.	Semester Grade Point Average (GPA) in the 3rd and 5th semester	Polynominal	Label

B. Attribute Selection

In this stage, the selection of 11 (eleven) regular attributes used the information gain technique. In the RapidMiner application, the operator for information gain is called the "Weight by Information Gain" operator. This operator produced weight values for each regular attribute. The weighting of each attribute using Information Gain is shown in Figure 2.

attribute	weight
Fasilitas Smartphone yang Memadai	0.019
Aplikasi Daring yang Disukai	0.019
Kegiatan Lain di Luar Perkuliahan Selama Pembelajaran Daring	0.020
Jenis Kelamin	0.023
Kondisi Jaringan di Daerahnya	0.026
Minat Belajar Saat Pembelajaran Daring	0.028
Jalur Masuk	0.036
Sikap Belajar	0.038
Kreativitas Belajar	0.052
Dukungan Orang Tua	0.053
Kelompok Belajar	0.054

Fig. 2 Result of Information Gain for Attributes

After creating the attribute selection design modeling using the "Weight by Information Gain" operator in the RapidMiner application, the weighting results for each attribute can be observed in Figure 2. The attributes with the lowest weight values are achieved by two attributes, namely adequate smartphone facilities and preferred online learning applications, both with the same weight value of 0.019. Based on these information gain results, the two attributes with the lowest weights, adequate smartphone facilities and preferred online learning applications, not included in the subsequent data processing stage. Therefore, in this study, the attributes used for the next stage can be seen in Table II.

TABLE II. LIST OF ATTRIBUTES AFTER ATTRIBUTE SELECTION

No.	Name	Data Type	Role
1.	NIM	Polynominal	ID
2.	Gender	Binominal	Regular
3.	Admission pathway	Polynominal	Regular
4.	Network conditions in their area	Polynominal	Regular
5.	Interest in learning during online learning	Polynominal	Regular
6.	Learning attitude	Binominal	Regular
7.	Learning Creativity	Polynominal	Regular
8.	Parental support	Polynominal	Regular
9.	Study groups (online discussions)	Binominal	Regular
10.	Other activities outside of lectures during online learning (organizations, work, etc.)	Polynominal	Regular
11.	Semester Grade Point Average (GPA) in the 3rd and 5th semester	Polynominal	Label

C. Data Cleaning

The next stage is data cleaning. In the RapidMiner application, the operator used to remove duplicate data is the "Remove Duplicates" operator. This operator works by selecting the attribute that filters its duplicate data. In this study, the filtered attribute is the NIM attribute because it is an ID attribute. Additionally, the "Remove Duplicates" operator provides the option to "treat missing values as duplicates" in the Parameters tab, allowing the cleaning of missing data to be performed by checking this option.

This data cleaning stage resulted in 9 (nine) dirty data records being cleaned or removed by the "Remove Duplicates" operator. The deleted data are those with duplicate NIMs, indicating them as duplicate data. Thus, the remaining data consists of 146 clean records ready for use in the next stage, which is algorithm modeling.

D. Naive Bayes Modeling

Before processing the data using the Naive Bayes algorithm, the data is divided into 2 (two) parts consisting of training data and testing data. Data division is done with the RapidMiner operator named Split Data. The data is divided into 70% as training data and 30% as testing data.

The results of processing the Naive Bayes algorithm in the form of a comparison of the actual semester grade index data with the predicted semester grade index data are shown in Table III.

TABLE III. PREDICTION RESULTS

No.	NIM	Semester Grade Point Average (GPA)	Prediction Results
1.	1512620070	>3,50	>3,50
2	1512620005	> 3,50	> 3,50

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i1.1958, Copyright ©2024
Submitted : September 12, 2023, Revised : January 30, 2024, Accepted : January 31, 2024, Published : February 15, 2024

3	1512620062	> 3,50	> 3,50
4	1512620044	> 3,50	2,00 - 2,75
5	1512621028	> 3,50	> 3,50
:	:	:	:
40	1512621005	> 3,50	> 3,50
41	1512621065	> 3,50	> 3,50
42	1512621033	> 3,50	> 3,50
43	1512621085	> 3,50	2,75 - 3,50
44	1512621054	> 3,50	> 3,50

E. Evaluation

1) Confusion Matrix

Evaluation of the Naive Bayes algorithm prediction model using a confusion matrix provides a classification results table containing the predicted values and actual values shown in the following table IV.

TABLE IV. CONFUSION MATRIX CLASSIFICATION RESULT

Prediction	Actual			
	>3,50	2,75 – 3,50	2,00 – 2,75	< 2,00
>3,50	**32**	6	1	1
2,75 – 3,50	2	**1**	0	0
2,00 – 2,75	1	0	**0**	0
<2,00	0	0	0	**0**

Based on the confusion matrix classification in Table IV, the results of the confusion matrix evaluation are obtained in the form of accuracy, precision, and recall show in Table V.

TABLE V. CONFUSION MATRIX EVALUATION RESULT

Category	Accuracy (%)	Precision (%)	Recall (%)
>3,50	-	80,00 %	91,43 %
2,75 – 3,50	-	33,33 %	14,29 %
2,00 – 2,75	-	0,00 %	0,00 %
<2,00	-	0,00 %	0,00 %
Average	**75,00 %**	**28,33 %**	**26,43 %**

Based on Table IV, the results of the confusion matrix evaluation showed an accuracy of 75.00%, precision of 28.33%, and recall of 26.43%.

2) ROC (Receiver Operator Characteristic)

Evaluation using the ROC curve applies testing four times, which involves converting the labels from multi-class to binary class to enable processing with the ROC curve. The algorithm evaluation yields algorithm performance in terms of AUC values for each binary class, which is then averaged to obtain the overall AUC value for all classes [19].

The obtained AUC values are as follows: positive class > 3.50 is 0.717, positive class 2.75 – 3.50 is 0.593, positive class 2.00 – 2.75 is 0.535, and positive class < 2.00 is 0.872. The overall AUC accumulation is shown in Fig 3.

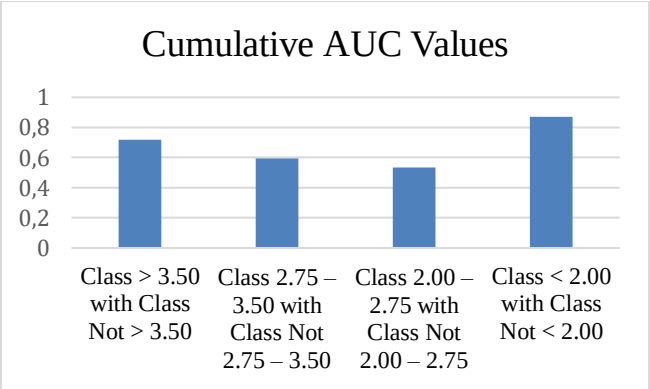

Fig. 3 Cumulative AUC Values

Based on Fig 3, the results of the ROC curve evaluation show an average AUC value of 0.679. This value falls within the range classified as poor based on the AUC values.

3) Confusion Matrix Using SMOTE

This research utilizes the data balancing technique, namely SMOTE, to achieve better performance. The result of the confusion matrix classification show in the following Table VI.

TABLE VI. CONFUSION MATRIX CLASSIFICATION RESULT USING SMOTE

Prediction	Actual			
	>3,50	2,75 – 3,50	2,00 – 2,75	< 2,00
>3,50	**27**	6	0	0
2,75 – 3,50	2	**1**	1	0
2,00 – 2,75	0	0	**0**	0
<2,00	1	0	0	**30**

The results of the confusion matrix evaluation using SMOTE show in Table VII.

TABLE VII. CONFUSION MATRIX EVALUATION RESULT USING SMOTE

Category	Accuracy (%)	Precision (%)	Recall (%)
>3,50	-	82,50 %	94,29 %
2,75 – 3,50	-	50,00 %	14,29 %
2,00 – 2,75	-	0,00 %	0,00 %
<2,00	-	97,22 %	100,00 %
Average	**88,46 %**	**57,43 %**	**52,14 %**

Based on Table VII, evaluation of the confusion matrix using SMOTE resulted in an accuracy of 88.46%, precision increased to 57.43%, and recall of 52.14%.

4) ROC Curve Using SMOTE

The ROC curve obtained from the application of SMOTE also yielded increased values compared to before the application of SMOTE. The cumulative AUC values with SMOTE show in Fig 4.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i1.1958, Copyright ©2024
Submitted : September 12, 2023, Revised : January 30, 2024, Accepted : January 31, 2024, Published : February 15, 2024

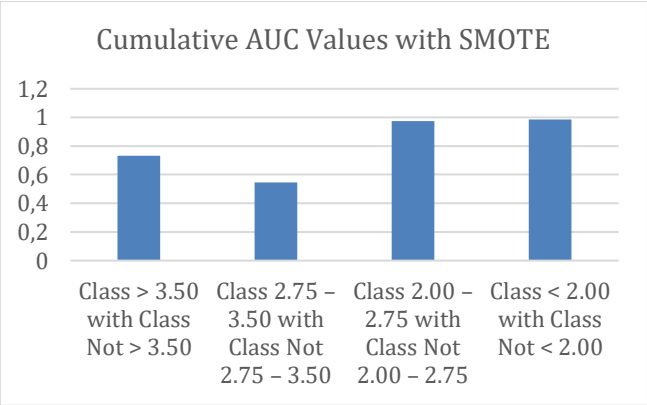

Fig. 4 Cumulative AUC Values with SMOTE

Based on Fig 4, the evaluation with the ROC curve results in an average AUC value of 0.809. This value falls within the range of good classification based on the interpretation of AUC classification values.

F. Result Analysis

After conducting testing and analysis of the prediction results and the performance of the Naive Bayes algorithm in predicting the Semester Grade Point Average (GPA) for students at Informatics and Computer Engineering Education, Universitas Negeri Jakarta, the results indicate that the evaluation and validation using the confusion matrix testing method resulted in an accuracy of 75% for unbalanced data and 88.46% for balanced data using SMOTE. The precision values obtained are 28.33% and 57.43% with SMOTE. The recall values obtained are 26.43% without SMOTE and 52.14% for balanced data using SMOTE.

The evaluation results of the Naive Bayes algorithm with the ROC curve yield an AUC value of 0.679 for unbalanced data. This value interprets that the classification produced is poor. In contrast, the AUC value obtained after balancing the data using SMOTE is 0.809, indicating that the classification results are categorized as good.

Therefore, it concluded that the results of the Semester Grade Point Average (GPA) prediction model for students using the Naive Bayes algorithm have good classification performance when applied to a balanced dataset. Thus, the model can be used to predict students' GPA during online learning. In an imbalanced dataset, the accuracy results obtained are not quite satisfactory, but they do not fall into the category of failed classification, making them still usable for predictions. However, to achieve optimal results and better performance, it is advisable to collect balanced or preprocess data using data balancing methods such as SMOTE to generate better model performance and accuracy.

V. CONCLUSION

Based on this research, the Naive Bayes algorithm in predicting students' Semester Grade Point Average (GPA) during online learning yielded confusion matrix performance values of 75% accuracy, 28.33% precision, and 26.43% recall. Meanwhile, performance using the ROC curve resulted in an AUC value is 0.679, categorized as poor classification. The prediction performance with SMOTE data balancing produced an accuracy of 88.46%, precision of 57.43%, and recall of 52.14%. Furthermore, the evaluation results using the ROC curve with balanced data through SMOTE yielded an AUC value of 0.809, categorizing the prediction model as a good classification.

Based on the results of this research, future research can employ other classification algorithms to obtain more accurate prediction results.

REFERENCES

[1] Y. N. Saputra, "Dampak Perkuliahan Daring terhadap Prestasi Belajar Mahasiswa Teologi Sekolah Tinggi Teologi Cipanas," *Andragogi: Jurnal Diklat Teknis Pendidikan dan Keagamaan*, vol. 9, no. 2, hlm. 154–164, Des 2021, doi: 10.36052/andragogi.v9i2.241.

[2] S. DH, Abd. Hafid, Mujahidah, dan Kasma, "Pengaruh Pembelajaran Daring Terhadap Prestasi Belajar Siswa Kelas IV Sekolah Dasar," *Jurnal Pendidikan & Pembelajaran Sekolah Dasar*, vol. 1, no. 3, hlm. 359–365, 2022, [Daring]. Tersedia pada: https://ojs.unm.ac.id/jppsd/index

[3] S. Karlina, R. S. Hayati, C. P. Danari, dan N. Nuryati, "Pengaruh Jumlah Jam Belajar Tambahan terhadap Indeks Prestasi Kumulatif Mahasiswa: Studi Kasus di Politeknik Negeri Bandung (Selama Masa Pandemi Covid-19)," dalam *Prosiding The 12th Industrial Research Workshop and National Seminar*, Bandung, 2021, hlm. 1574–1579.

[4] L. Sitinjak dan A. U. K. Kadu, "Faktor Internal Dan Eksternal Yang Mempengaruhi Kesulitan Belajar Mahasiswa Semester IV AKPER Husada Karya Jaya Tahun Akademik 2015/2016," *Jurnal AKademi Keperawatan Husada Karya Jaya*, vol. 2, no. 2, 2016.

[5] N. Laila, "Aspek Psikologi Pembelajaran Daring Masa Pandemi COVID-19 Dengan Capaian Indeks Prestasi Kumulatif Mahasiswa Vokasi," *Jurnal Ilmiah Pamenang - JIP*, vol. 2, no. 2, hlm. 7–16, 2020, doi: 10.53599.

[6] M. N. Faruqhy, D. Andreswari, dan J. P. Sari, "PREDIKSI PRESTASI NILAI AKADEMIK MAHASISWA BERDASARKAN JALUR MASUK PERGURUAN TINGGI MENGGUNAKAN METODE MULTIPLE LINEAR REGRESSION (STUDI KASUS: FAKULTAS TEKNIK UNIVERSITAS BENGKULU)," *Jurnal Rekursif*, vol. 9, no. 2, hlm. 172–183, 2021, [Daring]. Tersedia pada: http://ejournal.unib.ac.id/index.php/rekursif/

[7] T. Arifin dan D. Ariesta, "Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization," *Jurnal Tekno Insentif*, vol. 13, no. 1, hlm. 26–30, Apr 2019, doi: 10.36787/jti.v13i1.97.

[8] Syarli dan A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *Jurnal Ilmiah Ilmu Komputer*, vol. 2, no. 1, 2016, [Daring]. Tersedia pada: http://ejournal.fikom-unasman.ac.id

[9] I. Bagus, A. Peling, N. Arnawan, I. Putu, A. Arthawan, dan I. Janardana, "Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm," *International Journal of Engineering and Emerging Technology*, vol. 2, no. 1, hlm. 53–57, 2017, Diakses: 9 Mei 2022. [Daring]. Tersedia pada: https://ojs.unud.ac.id/index.php/ijeet/article/download/34457/20766

[10] Suyanto, *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Informatika, 2019.

[11] R. B. Pereira, A. Plastino, B. Zadrozny, dan L. H. C. Merschmann, "Information Gain Feature Selection for Multi-Label Classification," *Journal of Information and Data Manaagement*, vol. 6, no. 1, 2015.

[12] E. P. K. Orpa, E. F. Ripanti, dan Tursina, "Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Jurnal Sistem dan Teknologi Informasi*, vol. 7, no. 4, 2019.

[13] Mustakim dan G. Oktaviani, "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa," *Jurnal Sains, Teknologi dan Industri*, vol. 13, no. 2, hlm. 195–202, 2016, [Daring]. Tersedia pada: http://ejournal.uin-suska.ac.id/index.php/sitekin

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i1.1958, Copyright ©2024
Submitted : September 12, 2023, Revised : January 30, 2024, Accepted : January 31, 2024, Published : February 15, 2024

[14] V. Nykänen, I. Lahti, T. Niiranen, dan K. Korhonen, "Receiver operating characteristics (ROC) as validation tool for prospectivity models - A magmatic Ni-Cu case study from the Central Lapland Greenstone Belt, Northern Finland," *Ore Geol Rev*, vol. 71, hlm. 853–860, 2015, doi: 10.1016/j.oregeorev.2014.09.007.

[15] F. S. Nugraha, M. J. Shidiq, dan S. Rahayu, "Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, hlm. 149–156, Agu 2019, doi: 10.33480/pilar.v15i2.601.

[16] N. P. Y. T. Wijayanti, E. N. Kencana, dan I. W. Sumarjaya, "SMOTE: Potensi Dan Kekurangannya Pada Survei," *E-Jurnal Matematika*, vol. 10, no. 4, hlm. 235, Nov 2021, doi: 10.24843/mtk.2021.v10.i04.p348.

[17] A. Desiani, S. Yahdin, dan D. Rodiah, "PREDIKSI TINGKAT INDEKS PRESTASI KUMULATIF AKADEMIK MAHASISWA DENGAN MENGGUNAKAN TEKNIK DATA MINING," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 7, no. 6, hlm. 1237–1244, 2020, doi: 10.25126/jtiik.202072493.

[18] Y. B. Samponu dan Kusrini, "Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu," *Jurnal ELTIKOM*, vol. 1, no. 2, hlm. 56–63, 2017.

[19] N. Gotlieb *dkk.*, "A Multi-Class Neural Network Machine Learning Algorithm to Diagnose Graft Pathology in Liver Transplant Recipients." [Daring]. Tersedia pada: https://ssrn.com/abstract=4323740

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v13i1.1958, Copyright ©2024
Submitted : September 12, 2023, Revised : January 30, 2024, Accepted : January 31, 2024, Published : February 15, 2024