Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

Naïve Bayes-based Student Graduation Prediction Model: Effectiveness and Implementation to Improve Timely Graduation

Ketut Jaya Atmaja[1)*], I Putu Yoga Indrawan[2)], I Made Dwi Putra Asana[3)] , I Kadek Wawan[4)], Ayu Gde Chrisna Udayanie[5)]
[1)*2)3)4)]Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia
[5)]Universitas Bali Dwipa
[1)*]ketutjayaatmaja@instiki.ac.id, [2)]yoga.indrawan@instiki.ac.id, [3)]dwiputraasana@instiki.ac.id,
[4)]kdwawan08@gmail.com, [5)]ayuchrisna@balidwipa.ac.id

ABSTRACT

Studies in an educational institution, when the lack of timely graduation of students in each batch and the number of students in each batch, causes an imbalance between incoming students and outgoing students and causes a decrease in accreditation from the campus, this should not continue to happen, the solution to dealing with this problem as an early detection of students who graduate on time is to predict the length of the student study period they have. Therefore, researchers will discuss the design of a prediction system for graduating on time using the Naïve Bayes method, to predict student graduation so that there is no imbalance of incoming and outgoing students. The construction of this system also uses the Naïve Bayes method and the CRISP-DM (Cross Industry Standard Process Data Mining) development method. In this case study, the Naïve Bayes method predicts data into 2, namely 1 (graduated on time) and 0 (did not graduate on time) by labeling the data used. In this model using 3247 data with the selection of features, namely semester achievement index 1 (ips1), ips2, ips3, ips4, ips5, semester credit units1 (credits1), credits2, credits3, credits4, credits5, semester credit units not passed 1 (skstidaklulus1), skstidaklulus2, skstidaklulus3, skstidaklulus4, skstidaklulus5 and labels. Using these feature variables results in model performance with 80% accuracy, with 80% accuracy it can be said that the model works well.

Keywords: Naïve Bayes; Prediction; Timely Graduation

1. INTRODUCTION

Higher education plays an important role in preparing competent human resources who are ready to face global challenges. One indicator of the success of higher education institutions is the on-time student graduation rate. A high on-time graduation rate not only reflects the quality of education provided, but also significantly affects the reputation and accreditation of the institution.

However, many higher education institutions face challenges in achieving optimal on-time graduation rates. The disparity between the number of students who enter and those who graduate on time can reduce accreditation and negatively impact public perception of the institution. Therefore, an effective solution is needed to detect and predict student graduation early, so that appropriate interventions can be made.

The National Higher Education Standards stipulate that one of the criteria for assessing campus accreditation is student academic achievement, which includes on-time graduation and student cumulative grade point average (GPA). When there is a comparison between the minimal number of students graduating on time each batch and the total number of students each batch, this should not continue to happen because it can have a negative impact on campus assessment. In addition, based on PerBAN-PT Regulation No. 8 Th 2022, on-time students also tend to have better quality graduates, which is one of the dimensions of assessment at the output quality point. Therefore, the number of on-time students can be one of the factors taken into account in the campus accreditation process.

Monitoring the study period of students carried out by the study program is only using analysis and calculation of the number of credits and the value of IPS, GPA. If students get an IPS below 2.75 in each semester, there will be action given to students by the study program, namely by conducting counseling guidance. The study program has

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

always monitored students who have passed the study period above four years (8 semesters) or in other words exceed the provisions to graduate on time.

To overcome this, it is necessary to have a system that can predict students who graduate on time. The technique used to determine the prediction system uses data mining. Data Mining is an activity that includes collecting and using historical data to find regularities, patterns, or relationships in large data sets (Asana, 2020). The naïve bayes data mining method is a probability and statistical method that predicts future opportunities based on previous historical data (Setiyani, 2020). The Naive Bayes method works by calculating class probabilities on training data, then using the probabilities to predict classes on test data (Witten, 2023).

This research develops a Naïve Bayes-based student graduation prediction model. The Naïve Bayes method was chosen because of its simplicity and its ability to provide accurate results for data classification. In this study, the prediction system was developed using a structured and systematic CRISP-DM (Cross Industry Standard Process for Data Mining) approach. The causes of students graduating not on time are due to several things, namely the first is the length of study, taking college leave, many courses that are not passed, semester grade point average (IPS) and cumulative grade point average (GPA) that have not met the target. The maximum IPS and GPA of students will allow students to take the maximum number of credits between 20-24 credits. So that students will be faster in achieving the fulfillment of the required number of credits as one of the graduation requirements, which is a total of 144 credits.

2. LITERATURE REVIEW

The first research by Fatah (2018) used the Naïve Bayes method. The main problem is the absence of a graduation prediction pattern at UIN Raden Fatah Palembang and the non-utilization of large data. The results show a graduation classification accuracy of 82.08%, which is influenced by the lack of data complexity. The second study by Putra and Kamayani (2020) identified the problem of many students not graduating on time, causing a buildup in the number of students, lack of classrooms, and parking lots. This model shows 80.19% accuracy, 80.26% recall, 92.75% precision, and 86.05% F-Measure. The third study by Mulyadi and Sugiarto (2021) found a decrease in the percentage of on-time graduation from 92.7% to close to 70%. The prediction model using Naïve Bayes produces 93.75% accuracy, 93.75% precision, 100% recall, and AUC 1,000 (excellent classification). The fourth study by Siswanto (2019) emphasized the importance of universities producing quality and competitive scholars. The results of student performance evaluation stored in the academic database are used to support management decisions. The Naïve Bayes method obtained an accuracy of 95.14%. The fifth study by Setiyani et al. (2020) found that the on-time graduation rate affects the quality and accreditation of universities. The Naïve Bayes method produces accuracy above 90% with various attributes, including GPA as the main determinant of prediction. Based on literature review or previous research, on average most researchers use the naïve bayes method. However, GAP current research is previous research did not use the application of CRISP-DM development and the testing process using confusion matrix simultaneously. Therefore, this research proposal also predicts students who graduate on time using the naïve bayes method which is carried out on the campus XYZ What distinguishes between previous researchers and this researcher's proposal is using the application of CRISP-DM development and the testing process will be carried out using confusion matrix simultaneously.

3. METHOD

System Overview

The general description of the system to be built is that there are two main processes, namely the training and testing processes. Where the first step divides the dataset into two, namely training data and test data. After that, data preprocessing is carried out to convert raw data into a form of data that is easy to understand and data selection is carried out to sort out data attributes that are in accordance with the needs of the timely graduation prediction process. The next step is the data mining process, the results of data modeling are then tested and evaluated using the confusion matrix method. If the modeling results are as expected, the model is used to predict student graduation on time using test data. An overview of the system to be created is shown in Figure 1.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

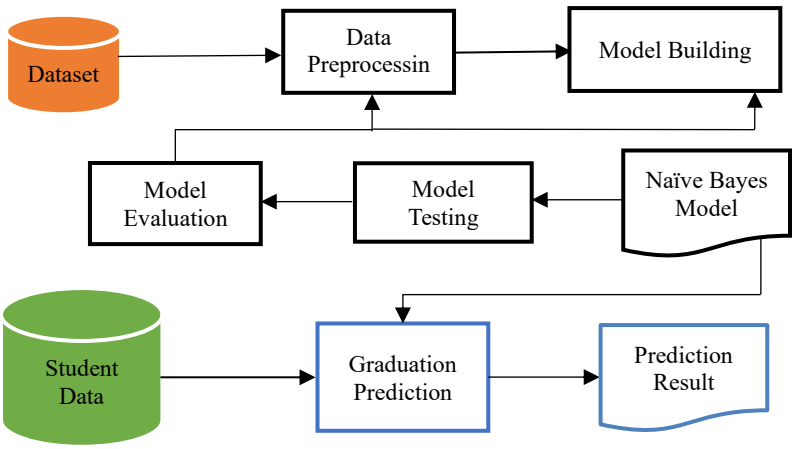

Fig.1 System Overview

System Development Method

In this research, the system development method uses the CRISP-DM development model. This development model consists of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. CRISP-DM is a standard data mining process for universal problem-solving strategies from businesses or research units [9]. Figure 3 shows the flow of the CRISP-DM method [10].

Fig.2 CRISP-DM Development Model

The six stages that will be carried out are as follows:

1) Business Understanding

The problem raised in this study is related to the on-time graduation of students. Where at this time when the minimum number of students graduates on time each generation and the number of students each generation, causing an imbalance between the number of incoming students and graduating students. this should not continue to happen because it can have a negative impact on campus assessment and accreditation. One solution to deal with this problem is to predict the length of student study period. This system is needed to keep the student study period from exceeding the specified limit and reduce the number of students who drop out (DO).

2) Data Understanding

The data used in this study is data obtained from the XYZ campus where the data is in the form of student graduation data from 2015 to 2019. The data has 7 attributes consisting of nim, name, major, IPS, SKS, SKS not

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

passed, gender, student status.

3) Data Preparation
In this phase, the selection and processing of data to be used is carried out. This aims to get optimal model performance by deleting or changing irrelevant data in the dataset.

4) Modeling
At this stage, the system modeling technique is determined, system modeling can be used in the data processing process by the system where it is hoped that the resulting model will be able to predict student graduation on time with a high level of accuracy. This research uses the naïve bayes method.

5) Evaluation
At this stage, it is carried out to evaluate the performance of the model that has been generated using the confusion matrix method.

6) Deployment
This stage is the stage of implementing the resulting model into a system that will be used for the prediction process of student graduation on time.

Model Evaluation

At the stage of evaluating the model used in this research, namely using the evaluation method, namely confusion matrix, to find out how the accuracy produced by the model. Confusion Matrix is a performance measurement for machine learning classification problems where the output can be two or more classes. Confusion Matrix is a table with 4 different combinations of predicted and actual values. There are four terms that represent the results of the classification process in the confusion matrix: True Positive, True Negative, False Positive, and False Negative. Table 1 is a table of the confusion matrix in question.

Table 1
Confusion Matrix Table

		Observed	
		TRUE	FALSE
Predicted Class	TRUE	True Positive (TP)	False Positive (FP)
	FALSE	False Negative (FN)	True Negative (TN)

Based on the confusion matrix above, the accuracy, precicion, and recall values are calculated. Accuracy: the accuracy value is the ratio between the correctly classified data and the entire data. The accuracy value can be obtained by the equation. Precision: Precision value describes the number of correctly classified positive category data divided by the total positive classified data, Precision can be obtained by the equation: Recall: Meanwhile, recall shows the percentage of positive category data that is correctly classified by the system.

4. RESULT

System Development With CRISP-DM Method

In the flow and process of system development, researchers use the flow development method in data mining applications, namely CRISP-DM (Cross Industry Standard Process Model For Data Mining). The CRISP-DM method consists of several methods, Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

Business understanding

Business understanding there is a model building process where in this process to get results about the student prediction system on time with the aim of focusing on modeling with data recognition techniques to be able to predict the prediction of student graduation on time. There are several features that influence the process of predicting student graduation on time, namely semester achievement index 1 (ips1), ips2, ips3, ips4, ips5, semester credit units1 (credits1), credits2, credits3, credits4, credits5, semester credit system does not pass 1 (skstidaklulus1), skstidaklulus2, skstidaklulus3, skstidaklulus4, skstidaklulus5 and labels. Understanding this feature is intended so that the information

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

generated can be useful for the XYZ college, this also makes it easier for the college to record students who graduate on time and not on time.

Data Understanding

Data understanding where the process is carried out when making or building this model, the data used is data obtained from the college or XYZ agency, the data is in the form of student data from 2015 to 2019. From 2015 to 2018 data consists of students who graduated on time and students who did not graduate on time. The data obtained is data that contains several variables or attributes: nim, name, study program, ips1, ips2, ips3, ips4, ips5, ips6, ipk, total semester, credits1, credits2, credits3, credits4, credits5, credits6, credits not passed1, credits not passed2, credits not passed3, credits not passed4, credits not passed5, gender, region of origin, employment status, label. Furthermore, the data is stored in a .csv format file so that it can be read by the system when performing the modeling process.

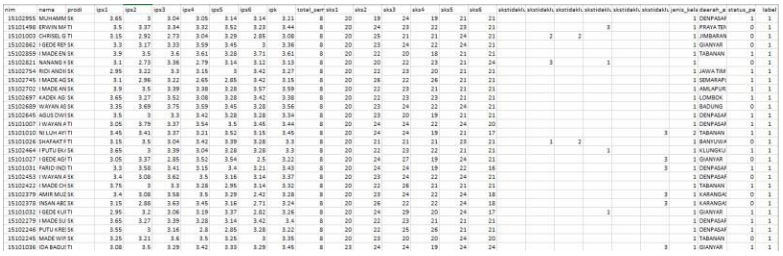

Fig.3 Prediction Data Image

Data Preparation

Before being used in the train and validation process, student data must be processed first, in order to be able to produce the best prediction model when modeling. The data that has been formatted into .csv will be imported into google collab to be processed before the modeling stage. The process carried out is data selection so that it becomes part of the data features that will be used when making prediction models. The data features used are semester achievement index 1 (ips1), ips2, ips3, ips4, ips5, semester credit units1 (credits1), credits2, credits3, credits4, credits5, semester credit system not passed 1 (skstidaklulus1), skstidaklulus2, skstidaklulus3, skstidaklulus4, skstidaklulus5 and labels. After the feature selection process is done, there are many columns or variables that are not very useful in model building. Columns that are not useful are deleted to facilitate activities during the Data Preprocessing process. So that the remaining columns include: semester achievement index 1 (ips1), ips2, ips3, ips4, ips5, semester credit units1 (credits1), credits2, credits3, credits4, credits5, semester credit system not passed 1 (skstidaklulus1), skstidaklulus2, skstidaklulus3, skstidaklulus4, skstidaklulus5 and labels. In Figure 4.2 are columns and examples of data features that will be used for the next process.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3246 entries, 0 to 3245
Data columns (total 16 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   ips1            3246 non-null   float64
 1   ips2            3246 non-null   float64
 2   ips3            3246 non-null   float64
 3   ips4            3246 non-null   float64
 4   ips5            3246 non-null   float64
 5   sks1            3246 non-null   int64
 6   sks2            3246 non-null   int64
 7   sks3            3246 non-null   int64
 8   sks4            3246 non-null   int64
 9   sks5            3246 non-null   int64
 10  skstidaklulus1  3246 non-null   int64
 11  skstidaklulus2  3246 non-null   int64
 12  skstidaklulus3  3246 non-null   int64
 13  skstidaklulus4  3246 non-null   int64
 14  skstidaklulus5  3246 non-null   int64
 15  label           3246 non-null   int64
```

Fig.4 Data Features

After selecting features, the dataset will be processed by splitting data, the splitting process is carried out to determine the amount of data used in the training process and the testing process. The use of Random state serves to

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

randomize the data used. The division of data is very influential on the model

Modeling

At the modeling stage where the system modeling technique is determined. In the process of building this model using the Naïve Bayes method. The method that can be used to predict student graduation is the Naïve Bayes method. Where by using the naïve bayes method when predicting student graduation on time because of its accuracy. The model building process is also carried out by collecting training data to be processed into a model.

After the training process, the model is re-evaluated using test data that has been prepared previously. This testing process aims to measure the performance of the model. Whether the model has been able to predict well or not. To see the results of the model performance, a confusion matrix is used which consists of accuracy, precision, and recall. The results of model performance using data splitting are 80% - 20%. The results of the model performance measured using the confusion matrix method by producing accuracy, precision, and recall values can be seen in the following figure:

```
accuration
              precision    recall  f1-score   support

           0       0.87      0.54      0.66       232
           1       0.79      0.95      0.86       418

    accuracy                           0.81       650
   macro avg       0.83      0.75      0.76       650
weighted avg       0.82      0.81      0.79       650
```

Fig.5 Naïve Bayes Model Performance Results

Meanwhile, accuracy shows that the Accuracy value is the ratio between correctly classified data and the total data .. where there are 126 true positives, 309 true negatives, 106 false positives and 19 false negatives. and the results obtained are 81% accuracy value. The Precision value describes the number of correctly classified positive category data divided by the total positive classified data and where the true positive is 126 data and the positive phase is 106 data and the results obtained are for graduating on time 79%, and not graduating on time 87%. Meanwhile, recall shows the percentage of positive category data that is correctly classified by the system. where the true positive is 126 data and the negative flase is 19 data. and the results obtained are for passing on time 95%, and not passing on time 54%.

Evaluation

The evaluation stage is intended to evaluate the performance of the model built in the training and testing stages. The evaluation stage is also carried out in order to determine the performance of the running model whether the performance of the model is good or not. This evaluation process is carried out by entering, then the data is predicted again, whether the performance of the model can predict well or not. In this evaluation, researchers used 327 data, so that the final results of this prediction produced output in the form of numbers 0 and 1, where the number (0) is said not to pass on time while the number (1) is declared to pass on time. The results of the prediction output are as follows:

```
predict
[1 1 1 1 0 1 1 1 0 1 1 1 0 0 1 0 0 0 1 1 1 0 1 1 1 1 1 1 0 1 1 1 0 1 1 0 1
 1 0 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 0 1 1 1 1 1]
```

Fig.6 Model Performance Evaluation

5. DISCUSSIONS

Confusion Matrix

Confusion Matrix is a performance measure for machine learning classification problems where the output can be two or more classes. Confusion Matrix is a table with 4 different combinations of predicted and actual values. There are four terms that represent the results of the classification process in the confusion matrix: True Positive, True Negative, False Positive, and False Negative.

From the results using confusion matrix by using data splitting from cross fold validation 10 stages of 3247 data used in modeling. Cross fold validation to estimate the prediction error in evaluating the performance of the model that has

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

been built.

Cross Fold Validation

At this testing stage, cross fold validation is used, where cross fold validation is used to estimate prediction errors in evaluating the modeling stage that has been built so that overfitting and underfitting do not occur. Cross fold validation is also used to obtain maximum accuracy results from the modeling built. At this stage, researchers use stage 10 of cross fold validation so that each data is processed and separated into two data subsets, namely training data and test data, using 10 cross fold validation also provides a good accuracy estimate because each data subset is processed. Applying cross fold validation will also produce accuracy scores.

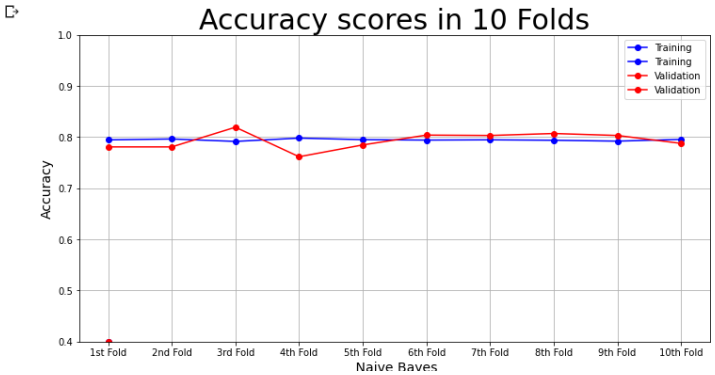

Fig.7 Acuracy scores Cross Fold Validation

Figure 7 shows the accuracy visualization of cross fold validation 10, the visualization above shows the stability of the algorithm from training data, validation data and there is no overfitting and underfitting.

```
{'Training Accuracy scores':
array([0.79452055, 0.79623288, 0.79152397,
0.79794521, 0.79494863,
       0.79409247, 0.79460847, 0.79375267,
0.79204108, 0.79546427])
'Validation Accuracy scores':
array([0.78076923, 0.78076923, 0.81923077,
0.76153846, 0.78461538,
       0.80384615, 0.8030888 , 0.80694981,
0.8030888 , 0.787644791)
```

Fig.8 Accuracy Training and Validation Score

Figure 8 displays the results of the training accuracy score and validation accuracy score using 10 cross fold validation where the accuracy results average 79% for training accuracy score and 80% for validation accuracy score.

Roc (Receiver Operating Characteristic) And AUC (Area UnderCurve)

Area under the curve (AUC) is an area under the receiver operating characteristic (ROC). Receiver operating characteristic (ROC) is a curve that results from the trade-off between sensitivity and specificity at various cut-off points. The AUC value is theoretically between 0 and 1. The AUC value provides an overview of the overall measurement of the suitability of the model used. The larger the area under curve, the better the variable under study in predicting the event. The code of the ROC and AUC generation can be seen:

ROC and AUC are also square-shaped regions whose values are always 0 and 1. Random perfomance also produces AUC values, the curve obtained is a diagonal line between point (0,0) and point (1,1). If the resulting AUC is less than 0.5, the AUC value provides an overview of the overall measurement of the suitability of the model used.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

The larger the area under curve, the better the variable under study in predicting the event. The following is a graph to determine the ROC and AUC of the model fit measurement.

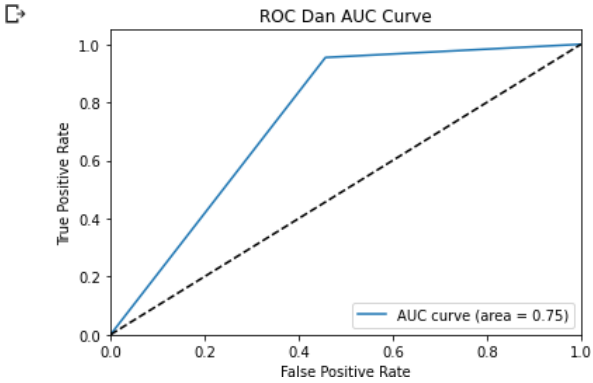

Fig.9 ROC and AUC

If the resulting AUC is less than 0.5, then the evaluation of the model that has provided an overview of the overall measurement of the suitability of the model used is very poor and if the resulting AUC is greater, namely 0.5 to 1, the better the variables studied in predicting students on time, the evaluation is good and the percentage results obtained are 75% (medium).

6. CONCLUSION

From the results of the research that has been done, several conclusions are obtained which are explained as follows: In building a prediction model for on-time graduation, the first stage is to prepare data that will be used during the modeling process. The data used is data from the Agency, namely data on students who graduated on time and not on time, which amounted to 3247 data. In preparing the data, the selection of features used in the training process, model validation is carried out. After selecting features, it is continued with the data splitting process where to get training and testing data, this data splitting uses cross fold validation to systemize prediction errors so that overfitting and underfitting do not occur. In the training process, training data is processed and processed into a model. In the process of building this model using the Naïve Bayes method. The method that can be used to predict student graduation is the Naïve Bayes method. Where by using the Naïve Bayes method when predicting student graduation on time because of its accuracy. The model building process is also carried out by collecting training data to be processed into a model. The AUC value provides an overview of the overall measurement of the suitability of the model used. The greater the area under curve, the better the variable studied in predicting student graduation on time and getting a percentage of 75% (medium). Based on the results of the training process using cross fold validation, it can be said that the evaluation of this model is stable at 79%. Measurement of model performance using the confusion matrix method by producing 81% accuracy of the data used. The model that has been built will be saved into a file called pickle and formatted into .pkl which will then be used in building a student graduation prediction system website on time using the Naïve bayes method.

7. REFERENCES

Amalia, R. 2020. "Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode naïve bayes". Jurnal Informatika dan Sistem Informasi, *6*(1), 33–42.

Asana IMDP, Ganda Wiguna IKA, Atmaja KJ, Sanjaya IPA. *FP-Growth Implementation in Frequent Itemset Mining for Consumer Shopping Pattern Analysis Application*. Mantik [Internet]. 2020Nov.24 [cited 2023Apr.5];4(3):2063-70. Available from: https://iocscience.org/ejournal/index.php/mantik/article/view/1075

Asri, A., Arifin, A., dkk. 2022. "2775-801X (Online)", *2*(1), 21–26.

Chohan, S., Nugroho, A., dkk. 2020. "Analisis Sentimen Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling Technique", *22*(2).

Fatah, R. 2018. "Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah", *4*, 49–62.

Firdaus, A. F., Saedudin, R., dkk. 2021. "Implementasi Metode Klasifikasi Naive Bayes Implementation of Naive

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4408

Submitted : 27 July 2024
Accepted : 30 July 2024
Published : 31 July 2024

Bayes Classification Method in Predicting", *8*(5), 9274–9279.

Ganda, L. H., Bunyamin, H., dkk. 2021. "Penggunaan Augmentasi Data pada Klasifikasi Jenis Kanker Payudara dengan Model Resnet-34", *3*, 187–193.

Gustanto, A. D., Rismawan, T., dkk. 2022. "Implementasi Metode Naïve Bayes dan Certainty Factor dalam Diagnosis Hama dan Penyakit Tanaman Anggrek Bulan Berbasis Android". Justin (Jurnal Sistem dan Teknologi Informasi), *10*(1), 180–188. https://doi.org/10.26418/justin.v10i1.51983.

Heriyanto, E., Kumalasarinurnawati, E., dkk. 2018. "Skripsi Implementasi Kecerdasan Buatan Pada Game Menggunakan Metode Pathfinding Dengan Game Engine Unity3D". Jurnal SCRIPT, *5*(2), 56–62. diambil dari https://ejournal.akprind.ac.id/index.php/script/article/view/641.

Latifah, R., Wulandari, E. S., dkk. 2019. "Model Decision Tree Untuk Prediksi Jadwal Kerja Menggunakan Scikit-Learn". Jurnal Universitas Muhammadiyah Jakarta, 1–6. diambil dari https://jurnal.umj.ac.id/index.php/semnastek/article/download/5239/3517.

Mardiana, L., Kusnandar, D., dkk. 2022. "ANALISIS DISKRIMINAN DENGAN K FOLD CROSS VALIDATION UNTUK KLASIFIKASI KUALITAS AIR DI KOTA PONTIANAK", *11*(1), 97–102.

Marudut, V., Siregar, M., dkk. 2018. "Menurut Turangan et . al (2017) insentif merupakan salah satu jenis penghargaan yang dikaitkan dengan prestasi kerja . Semakin tinggi prestasi kerja semakin besar pula insentif yang diterima . Sudah menjadi kebiasaan bahwa setiap perusahaan harus meneta", *7*, 87–94.

Mulyadi, C., dan Sugiarto, L. 2021. "Penggunaan algoritma naïve bayes untuk prediksi ketepatan waktu lulus mahasiswa diploma 3 STMIK Cipta Darma Surakarta". Teknomatika, *11*(01), 21–30. diambil dari http://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/512.

Prabowo, W. A., dan Wiguna, C. 2021. "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM". Jurnal Media Informatika Budidarma, *5*(1), 149. https://doi.org/10.30865/mib.v5i1.2604.

Prasanta, M. R., Pranata, M. Y., dkk. 2022. "Rancang Bangun Quadcopter Drone Untuk Deteksi Api Menggunakan YOLOv4". Cyclotron, *5*(1). https://doi.org/10.30651/cl.v5i1.10013.

Pratiwi, B. P., Handayani, A. S., dkk. 2021. "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix". Jurnal Informatika Upgris, *6*(2), 66–75. https://doi.org/10.26877/jiu.v6i2.6552.

Putra, D. A., dan Kamayani, M. 2020. "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes di Program Studi Teknik Informatika UHAMKA", *5*(2502), 34–40. https://doi.org/10.22236/teknoka.v5i.331.

Retnoningsih, E., dan Pramudita, R. 2020. "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python". Bina Insani Ict Journal, *7*(2), 156. https://doi.org/10.51211/biict.v7i2.1422.

Saluky, S. 2018. "Tinjauan Artificial Intelligence untuk Smart Government". ITEJ (Information Technology Engineering Journals), *3*(1), 8–16. https://doi.org/10.24235/itej.v3i1.22.

Setiyani, L., Wahidin, M., dkk. 2020. "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review". Faktor Exacta, *13*(1), 35. https://doi.org/10.30998/faktorexacta.v13i1.5548.

Sholeh, M., Rachmawati, R. Y., dkk. 2022. "Penerapan Regresi Linear Ganda Untuk Memprediksi Hasil Nilai Kuesioner Mahasiswa Dengan Menggunakan Python", *11*(1), 13–24.

Siswanto, E. 2019. "Optimasi Metode Naïve Bayes Dalam Memprediksi Tingkat Kelulusan Mahasiswa Stekom Semarang", *6*(1), 1–6.

Sukerta Wijaya, I. W., Harjumawan Wiratmaja KS., I. G., dkk. 2021. "Program Menghitung Banyak Bata pada Ruangan Menggunakan Bahasa Python". TIERS Information Technology Journal, *2*(1). https://doi.org/10.38043/tiers.v2i1.2840.

Utami dan Hidayat 2018. "Bab Ii Landasan Teori". *Journal of Chemical Information and Modeling*, *53*(9), 8–24.

Witten IH, Frank E, Hall MA. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, USA.

Yoshua, I. A., Pragantha, J., dkk. 2020. "Aplikasi Pengukur Kelayakan Tempat Tinggal Dengan Menggunakan Metode Naive Bayes". Jurnal Ilmu Komputer dan Sistem Informasi, *8*(1), 79–83.

* Corresponding author