Journal of Knowledge Engineering and Artificial Intelligence (JKEAI)
Volume 1, Issue 1, P-ISSN: -, E-ISSN : -, Page 8-15

Predicting Student Preparation for Exams Using the Naive Bayes Method

Irkhas Saputra[1*], Mehtab Ali[2]
[1] Faculty of Engineering and Computer Science, Informatika, Universitas Teknokrat Indonesia, Lampung, Indonesia
[2]National Defence University Islamabad, Pakistan
Email: [1*]irkhas_saputra@teknokrat.ac.id, [2]mehtabali3797@gmail.com
[*)] Corresponding email

Abstract−The National Examination (UN) is a government policy in the field of education to determine the quality standards of educatioN. The function of the national examination is important to measure student competence and one of the considerations for selection to a higher level. The Naive Bayes algorithm is mostly used in spam message filtering, sentiment analysis, and recommendation systems. One of the main reasons for the use of this algorithm is due to its quick and easy implementation. The test was carried out using the WEKA application, using the confusion matrix method to determine the effectiveness of classification with model equations and can be calculated to find accuracy, sensitivity, specificity, positive predictive value (ppv), and negative predictive value (npv). The results of the evaluation and validation of the confusion matrix using training data and testing data showed the accuracy rate and error rate in the Naïve Bayes algorithm of 96.8254% and 3.1746%.

Keywords: Classification, Data Testing, Data Training, Naïve Bayes, Prediction

1. INTRODUCTION

The National Examination (UN) is a government policy in the field of education to determine the quality standards of education[1];[2];[3]. The function of the national examination is important to measure student competence and one of the considerations for selection to a higher level. The purpose of conducting national examinations is very good for improving the quality of education. The reality that has occurred in the field so far, the implementation of national examinations still causes many problems[4]. From year to year, the moment of the national exam has always been a concern for many parties. Not only students, but also parents and the school. Parents play a role in preparing their children by providing support and motivation. Meanwhile, the school is required to prepare strengthening activities for students through material deepening activities, material enrichment and exam training. Although now the UN score is not the only graduation benchmark, the best result is naturally a hope for all parties concerned.

The preparation of students before carrying out the national exam is very important for the school to pay attention to. Every student in facing the exam is different, some face it seriously and some face it casually. In order for students to be successful in the national exams, the school took a policy to hold national examination training or tryouts [5];[6]. A tryout is an evalution stage to face the real national exam. The questions contained in the tryout refer to the material in the national exam [7];[8]. The preparation of the questions was obtained from the material of classes I, II and III, and it is possible that the questions were obtained from various mass media sources that are still related to the lattice of the national examination. Of course, it is inseparable from the applicable curriculum, but there may be new things that have never been taught by teachers [9];[10];[11]. With the tryout activity, the school will know the results. It can be known to the students who are ready and not ready by looking at the results of the tryout. The school conducts try outs or practice exams from the four subjects, namely Indonesian, English, Mathematics and Science for students' readiness to face the national exam.

Using the Naïve Bayes algorithm method, it is hoped that it can solve the problem [12];[13];[14]. Naïve Bayes is a simple probabilistic classification that calculates a set by summing frequencies and combinations of values from a given dataset. The algorithm uses Bayes' theorem and assumes all the independent or non-interdependent attributes given by values on class variables [15], [16].

From the above problems, the author will create a data processing system to predict students' readiness to face national exams. The attribute used is tryout value data for the last 3 years. Then there are 3 kinds of labels used, namely ready, quite ready and not ready, where in the process of calculating the four subjects must have an assessment standard. Students are declared ready if the scores of all four subjects and the average score meet the minimum completion criteria (KKM) standards [17];[18];[19]. Then, students who are declared sufficiently prepared if one of the scores of the four subjects and the average is below the minimum completion criteria (KKM) standard. And students who are declared not ready if the scores of the four subjects and the average are more than one below the minimum completion criteria (KKM) standard.

Naive Bayes is a collection of algorithms compiled on the basis of Bayes' Theorem. Well, Bayes' Theorem itself is a mathematical model with a statistical and probability basis[20]–[22]. Although not a new thing, this algorithm remains relevant to machine learning (ML) that has developed recently, especially those that are still related to NLP or natural language processing problems [23];[24]. The Naive Bayes algorithm is mostly used in spam message filtering, sentiment analysis, and recommendation systems. One of the main reasons for the use of this algorithm is due to its quick and easy implementation. Naive Bayes is closely related to classification and machine learning[25], [26]. This algorithm is also often used in recommendation systems because it is considered efficient. Naive Bayes is a suitable method for binary and multiclass classification. This method, also known as the Naive Bayes Classifier, applies a supervised object classification technique in the future by assigning class labels to instances/records using conditional probabilities [27]. Conditional probability is a measure of the probability of an event occurring based on another event that has (assuming, preconceived, asserted, or proven) occurred. Classification of training data that has been labeled with classes.

2. RESEARCH METHODS

2.1 Research Framework

The research framework is basically a framework for the relationship between the concepts to be observed or measured through the research to be carried out. This study discusses the application of a prediction system using the Naïve Bayes method to determine student readiness in facing the national exam.

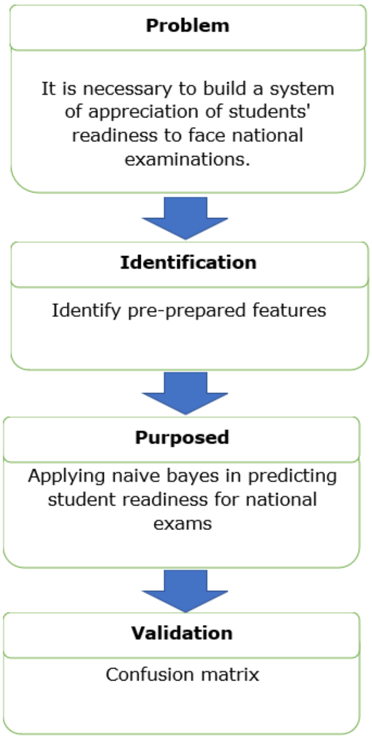

Figure 1. Research Framework

2.2 Naïve Bayes

Naïve Bayes is one of the algorithms found in the classification technique [28];[29];[30]. Naïve Bayes is a classification by probability and statistical methods proposed by the British scientist Thomas Bayes, namely predicting future odds based on previous experience so that it is known as Bayes' Theorem.

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)} \quad (1)$$

The theorem is combined with Naïve where it is assumed that the conditions between the attributes are mutually free. The Naïve Bayes classification assumes that the presence or absence of certain traits of a class has nothing to do with the traits of another class (Bustami, 2014).

Where:

X : Data with unknown classes

H : Data hypothesis is a specific class

P(H| X) : The probability of hypothesis H based on condition X (posteriori probability)

P(H): The probability of the hypothesis H (prior probability)

P(X): Probability X.

To explain the Naïve Bayes method, it is necessary to know that the classification process requires a number of instructions to determine what class is suitable for the analyzed sample.

3. RESULTS AND DISCUSSIONS

3.1 Data Preparation

The initial stage carried out in this study was the preparation of data obtained from SMP Negeri 3. The data obtained is tryout value data from the 2015/2016, 2016/2017, and 2017/2018 school years which totaled 419 datasets.
In the data, there are 8 attributes, namely Participant Number, Participant Name, Gender, UN Try Out Value (Indonesian, English, Mathematics, and Science), Average, and Description. But from the results obtained, there are some attributes that are not needed in the classification process. The picture below is a table diagram in grouping students with each school year.

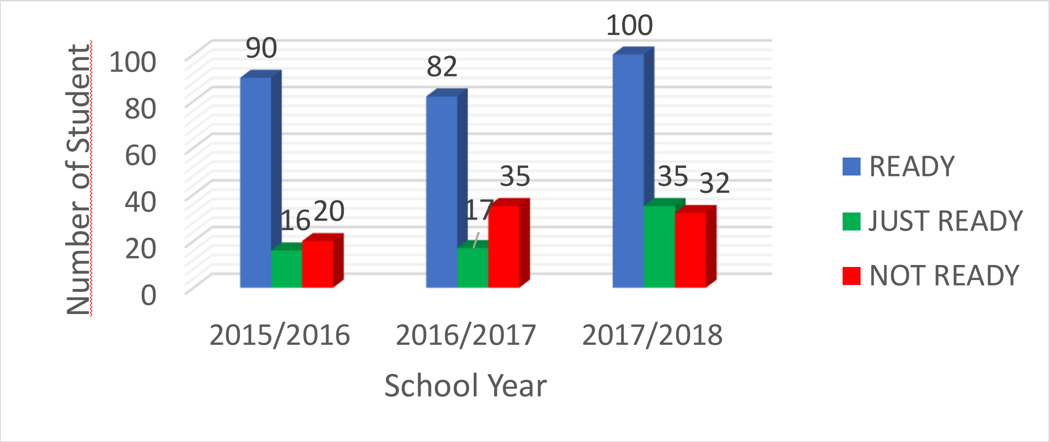

Figure 2. Grouping Diagram

3.2 Data Cleansing

From the results of the data obtained a number of data problems, especially incomplete data. For this reason, a data cleaning process is carried out to eliminate problematic data. The following is tryout value data used to determine student readiness. In this cleaning process, the deletion of records and variables is carried out because in the data processing there are variables that have no effect and are not used such as the Participant Name and Gender attributes. As contained in the research method, these attributes are not used in the prediction of determining student readiness, so they do not interfere during the process later. From the cleaning results, data were obtained of 419 datasets that will be used in the classification process.

3.3 Implementation of the Naïve Bayes Calculation

Training data is to determine whether a student is ready, sufficiently prepared, or not ready for the national exam. The classification of tryout un value data can be calculated if input is given in the form of Indonesian, English, mathematics, science and student readiness using the Naive Bayes algorithm. If new input is given, the data classification of the UN tryout value can be determined through the following steps:

a. Prepare training data, data to be used.
b. Count the number of classes
   P (Readiness=READY) = 182/293 = 0.62. The amount of READY data on the tryout un data divided by the amount of training data.
   P (Readiness=SUFFICIENTLY READY) = 44/293 = 0.15. The amount of sufficiently ready data on the tryout data un divided by the amount of training data.

P (Readiness=NOT READY) = 67/293 = 0.23. The amount of data NOT READY on the tryout un data divided by the amount of training data.

c. Counts the same number of cases with the same class
P(Indonesian = Good| Readiness = READY) = 71/182 = 0.39
P(Indonesian = Good| Readiness = SUFFICIENTLY READY) = 11/44 = 0.25
P(Indonesian = Good| Readiness = NOT READY) = 12/67 = 0.18
P(Indonesian = Sufficient| Readiness = READY) = 111/182 = 0.61
P(Indonesian = Sufficient| Readiness = SUFFICIENTLY READY) = 33/44 = 0.75
P(Indonesian = Sufficient| Readiness = NOT READY) = 52/67 = 0.78
P(Indonesian = Bad| Readiness = READY) = 0/182 = 0
P(Indonesian = Bad| Readiness = QUITE READY) = 0/44 = 0
P(Indonesian = Bad| Readiness = NOT READY) = 3/67 = 0.04

P(English = Good| Readiness = READY) = 89/182 = 0.49
P(English = Good| Readiness = QUITE READY) = 12/44 = 0.27
P(English = Good| Readiness = NOT READY) = 2/67 = 0.03
P(English = Sufficient| Readiness = READY) = 93/182 = 0.51
P(English = Sufficient| Readiness = QUITE READY) = 26/44 = 0.59
P(English = Sufficient| Readiness = NOT READY) = 43/67 = 0.64
P(English = Bad| Readiness = READY) = 0/182 = 0
P(English = Bad| Readiness = QUITE READY) = 6/44 = 0.14
P(English = Bad| Readiness = NOT READY) = 22/67 = 0.33

P(Math = Good| Readiness = READY) = 66/182 = 0.36
P(Math = Good| Readiness = QUITE READY) = 4/44 = 0.09
P(Math = Good| Readiness = NOT READY) = 0/67 = 0
P(Math = Enough| Readiness = READY) = 116/182 = 0.64
P(Math = Enough| Readiness = QUITE READY) = 20/44 = 0.45
P(Math = Enough| Readiness = NOT READY) = 17/67 = 0.25
P(Math = Bad| Readiness = READY) = 0/182 = 0
P(Math = Bad| Readiness = QUITE READY) = 20/44 = 0.45
P(Math = Bad| Readiness = NOT READY) = 50/67 = 0.75

P(IPA = Good| Readiness = READY) = 86/182 = 0.47
P(IPA = Good| Readiness = QUITE READY) = 12/44 = 0.27
P(IPA = Good| Readiness = NOT READY) = 8/67 = 0.12
P(IPA = Sufficient| Readiness = READY) = 96/182 = 0.53
P(IPA = Sufficient| Readiness = SUFFICIENTLY READY) = 30/44 = 0.68
P(IPA = Sufficient| Readiness = NOT READY) = 44/67 = 0.66
P(IPA = Bad| Readiness = READY) = 0/182 = 0
P(IPA = Bad| Readiness = QUITE READY) = 2/44 = 0.04
P(IPA = Bad| Readiness = NOT READY) = 15/67 = 0.22

P(Average = Good| Readiness = READY) = 45/182 = 0.25
P(Average = Good| Readiness = QUITE READY) = 0/44 = 0
P(Average = Good| Readiness = NOT READY) = 0/67 = 0
P(Average = Sufficient| Readiness = READY) = 137/182 = 0.75
P(Average = Sufficient| Readiness = SUFFICIENTLY READY) = 28/44 = 0.64
P(Average = Sufficient| Readiness = NOT READY) = 0/67 = 0
P(Average = Bad| Readiness = READY) = 0/182 = 0
P(Average = Bad| Readiness = QUITE READY) = 16/44 = 0.36
P(Average = Bad| Readiness = NOT READY) = 67/67 = 1

d. Calculating Testing Data
(Indonesian = 7.80, English = 6.80, Mathematics = 3.25, Science = 5.50, and average = 5.84)
Multiply all the results of the variables READY, SIMPLY READY, and NOT READY
P(READY) * P(Good| READY) * P(Enough| READY) * P(Bad| READY) * P(Enough| READY) * P(Enough| READY)
= (0.62)(0.39)(0.51)(0)(0.53)(0.75) = 0

P(QUITE READY) * P(Good| SIMPLY READY) * P(Enough| SIMPLY PREPARED) * P(Bad| SIMPLY READY) * P(Enough| SIMPLY READY) * P(Enough| JUST READY)
= (0.15)(0.25)(0.59)(0.45)(0.68)(0.64) = 0.004
P(NOT READY) * (Ok| NOT READY) * (Enough| NOT READY) * (Bad| NOT READY) * (Enough| NOT READY) * (Enough| NOT READY YET)
= (0.23)(0.18)(0.64)(0.75)(0.66)(0) = 0

e. Conclusion of the results of the specified class
From the results above, it can be seen that the highest probability value is in the class (P| SUFFICIENTLY PREPARED) so that it can be concluded that the status of the student falls under the classification of "SUFFICIENTLY PREPARED" to face the National Examination.

3.4 Evaluation and Validation

Classification methods can be evaluated based on criteria such as accuracy, speed, reliability, stability, and interpretability. After the data is processed, the data is tested for accuracy to see the performance of the method. The results of the model testing carried out, namely with the Naïve Bayes algorithm, were tested for the level of accuracy using the confusion matrix.

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Correctly Classified Instances         122               96.8254 %
Incorrectly Classified Instances         4                3.1746 %
Kappa statistic                          0.9292
Mean absolute error                      0.0707
Root mean squared error                  0.1288
Relative absolute error                 21.086  %
Root relative squared error             32.7606 %
Total Number of Instances              126

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     SIAP
                 0.875    0.018    0.875      0.875    0.875      0.857    0.997     0.980     CUKUP SIAP
                 0.900    0.019    0.900      0.900    0.900      0.881    0.997     0.986     BELUM SIAP
Weighted Avg.    0.968    0.005    0.968      0.968    0.968      0.963    0.999     0.995

=== Confusion Matrix ===

  a  b  c   <-- classified as
 90  0  0 |  a = SIAP
  0 14  2 |  b = CUKUP SIAP
  0  2 18 |  c = BELUM SIAP
```

Figure 3. Testing on the WEKA

From the picture above, it is divided into 2 data on the testing data, namely the wrong data and the correct data. 122 data are correct or 96.8254% and 4 data are incorrect or 3.1746%. From the calculation of the accuracy of the training data using testing data from 126 student data, 90 data were classified as ready predictions and turned out to be ready, 14 data were predicted to be quite ready and turned out to be quite ready, 2 data were predicted to be quite ready and turned out to be not ready, 2 data were predicted to be not ready and turned out to be quite ready, 18 data were predicted to be not ready and turned out to be not ready.

3.5 Validation Data Testing

Data testing is a set of data that is carried out on a Machine Learning algorithm by evaluating the performance of the algorithm. In the testing process, the performance of the algorithm will be tested using a testing set, where the testing set and training set are different data. Below is an image of the visualization of the classification of error data testing based on training data.

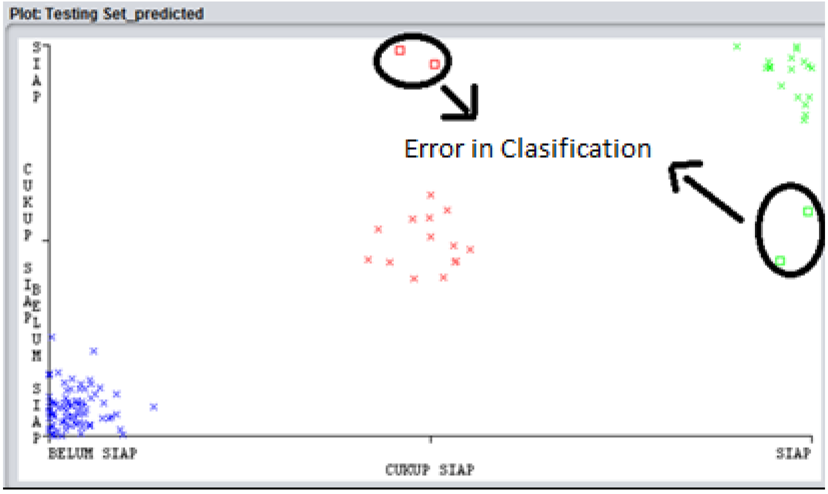

Figure 4. Error in Clasification

The image is a statistical error data in the testing set that is classified READY and turns out to be READY, classified READY and turned out to be QUITE READY, classified QUITE READY and turned out to be READY, classified QUITE READY and turned out to be QUITE READY, classified NOT READY and turned out to be NOT READY. The number of 120 testing data is denoted by the symbol X and the sum of 4 testing data is denoted by the symbol. The figure below is part of the testing data that is classified incorrect data (error).

The explanation of the correct data and false data (error), that is, the correct data means data that is classified or predicted ready and turns out to be READY, while the wrong data (error) means data that is classified or predicted ready and turns out to be QUITE READY. So, true data and incorrect data (error) have quite significant differences.

```
Plot : weka.classifiers.misc.InputMappedClassifier (Testing Set)
Instance: 54
             BIND : BAIK
             BING : BURUK
              MAT : CUKUP
              IPA : CUKUP
        RATA-RATA : CUKUP
prediction margin : -0.27954611912411925
    predicted KET : SIAP
              KET : CUKUP SIAP
```

Figure 5. Clasification Data Testing

The image is a statistical error data in the testing set that is classified READY and turns out to be READY, classified READY and turned out to be QUITE READY, classified QUITE READY and turned out to be READY, classified QUITE READY and turned out to be QUITE READY, classified NOT READY and turned out to be NOT READY. The number of 120 testing data is denoted by the symbol X and the sum of 4 testing data is denoted by the symbol. The figure below is part of the testing data that is classified incorrect data (error).

The explanation of the correct data and false data (error), that is, the correct data means data that is classified or predicted ready and turns out to be READY, while the wrong data (error) means data that is classified or predicted ready and turns out to be QUITE READY. So, true data and incorrect data (error) have quite significant differences.

4. CONCLUSION

The results of research on the prediction of student readiness to face national exams that are potentially ready, quite ready, or not ready can be drawn conclusions, namely data preparation is carried out by the process of data cleaning or data cleaning by eliminating unnecessary attributes. Furthermore, data in the form of numerical is changed to nominal. The test was carried out using the WEKA application, using the confusion matrix method to determine the effectiveness of classification with model equations and can be calculated to find accuracy, sensitivity, specificity, positive predictive value (ppv), and negative predictive value (npv). The results of the evaluation and validation of the confusion matrix using training data and testing data showed the accuracy rate and error rate in the Naïve Bayes algorithm of 96.8254% and 3.1746%.

REFERENCES

[1] V. Mariani, “Efektivitas Bimbingan Belajar Menghadapi Ujian Nasional Pelajaran Sosiologi SMA Muhammadiyah 2 Pontianak,” *J. Pendidik. Sosiol. dan Hum.*, vol. 4, no. 1, 2013.

[2] N. Iriawan and K. Fithriasari, “On the modeling of the average value of high school national examination in West Java using Bayesian hierarchical mixture normal approach,” in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 689–694.

[3] A. Amir, “The Contribution of National Examination Score Indonesia Language in High School and Reading Strategies toward Students Reading Comprehension Skill in Universitas Negeri Padang,” in *International Conference on Language, Literature, and Education (ICLLE 2018)*, 2018, pp. 229–233.

[4] I. K. Wijayanti and H. Retnawati, “Analisis profil kesiapan siswa SMA dalam menghadapi ujian nasional matematika di Kabupaten Temanggung,” *J. Pendidik. Mat. dan Sains*, vol. 6, no. 2, pp. 179–189, 2018.

[5] J. Harta, N. T. Rasuh, and A. Seriang, “Using HOTS-Based Chemistry National Exam Questions to Map the Analytical Abilities of Senior High School Students,” *J. Sci. Learn.*, vol. 3, no. 3, pp. 143–148, 2020, doi: 10.17509/jsl.v3i3.22387.

[6] N. V Obernikhina, T. S. Sanzhur, and L. V Gayova, “The Unified State Qualification Exam in Ukraine and Bogomolets National Medical University Under Covid-19 Conditions,” vol. 6, no. 1, 2021, doi: 10.19080/JOJPH.2021.06.555679.

[7] B. M. Modi, “CONSTRUCTION AND TRYOUT OF INQUIRY TRAINING MODEL WITH REFERENCE TO UPPER PRIMARY SCIENCE TEACHING”.

[8] F. Adiba, T. Islam, … M. K.-I., and undefined 2020, “Effect of corpora on classification of fake news using naive Bayes classifier,” *Researchlakejournals.Com*, vol. 1, no. 1, p. 80, 2020, [Online]. Available: https://researchlakejournals.com/index.php/AAIML/article/view/45

[9] M. Ainurrofiq, “Implementasi Algoritma Naive Bayes Untuk Menentukan Kesiapan Siswa Dalam Menghadapi Ujian Nasional,” *Skripsi Tek. Inform. Univ. Dian*, 2015.

[10] A. Wongrugsa, S. Kanjanawasee, and S. Ratchusanti, “The efficiency of an evaluation model for undergraduate vocationaeducation programs: An application of multiple evaluation approaches,” *Kasetsart J. Soc. Sci.*, vol. 43, no. 3, pp. 769–776, 2022, doi: 10.34044/j.kjss.2022.43.3.32.

[11] R. K. Zaman and R. Andriyanty, “Evaluation instrument development for scientific writing instruction with a constructivism approach,” *Tech. Soc. Sci. J.*, vol. 18, no. 21, pp. 235–243, 2021, [Online]. Available: https://techniumscience.com/index.php/socialsciences/index

[12] N. Salmi and Z. Rustam, “Naïve Bayes Classifier Models for Predicting the Colon Cancer,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052068.

[13] J. Galopo Perez and E. S. Perez, “Predicting Student Program Completion Using Naïve Bayes Classification Algorithm,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 13, no. 3, pp. 57–67, 2021, doi: 10.5815/ijmecs.2021.03.05.

[14] A. Tripathi, S. Yadav, and R. Rajan, “Naive Bayes Classification Model for the Student Performance Prediction,” *2019 2nd Int. Conf. Intell. Comput. Instrum. Control Technol. ICICICT 2019*, no. July 2019, pp. 1548–1553, 2019, doi: 10.1109/ICICICT46008.2019.8993237.

[15] A. Saleh, “Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga,” *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.

[16] S. Joshi, B. Pandey, and N. Joshi, “Comparative analysis of Naive Bayes and J48 Classification Algorithms,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 12, pp. 813–817, 2015.

[17] Y. F. Aziiza and T. Herman, “An analysis of the cause factors of unreachable standard minimum criteria on mathematics learning,” in *Journal of Physics: Conference Series*, 2020, vol. 1521, no. 3, p. 32066.

[18] A. R. Samsudin and D. E. Mascita, “The Development of Exposition Text Digital Teaching Materials for 8 th Grade Junior High School Students”.

[19] L. D. Haeruman, V. Serevina, and Y. E. Susanti, "Development of Interactive Ludo Games on Earth and Space Science Learning Material as High School Exercise Media," in *Journal of Physics: Conference Series*, 2022, vol. 2309, no. 1, p. 12091.

[20] D. Xiong, C. He, X. Liu, and M. Liao, "An End-To-End Bayesian Segmentation Network Based on a Generative Adversarial Network for Remote Sensing Images," *Remote Sensing*, vol. 12, no. 2. MDPI AG, p. 216, 2020. doi: 10.3390/rs12020216.

[21] D. Alita, I. Sari, A. R. Isnain, and S. Styawati, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *J. Data Min. Dan Sist. Inf.*, vol. 2, no. 1, pp. 17–23, 2021.

[22] L. Mescheder, "Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks," *34th International Conference on Machine Learning, ICML 2017*, vol. 5. pp. 3694–3707, 2017.

[23] M. Wongkar and A. Angdresey, "Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 2019, pp. 1–5.

[24] A. Kesumawati and D. T. Utari, "Predicting patterns of student graduation rates using Naïve bayes classifier and support vector machine," in *AIP conference proceedings*, 2018, vol. 2021, no. 1, p. 60005.

[25] E. Said Mohamed, A. A. Belal, S. Kotb Abd-Elmabod, M. A. El-Shirbeny, A. Gad, and M. B. Zahran, "Smart farming for improving agricultural management," *Egypt. J. Remote Sens. Sp. Sci.*, 2021, doi: https://doi.org/10.1016/j.ejrs.2021.08.007.

[26] Y. G. Shawai and M. A. Almaiah, "Malay language mobile learning system (MLMLS) using NFC technology," *Int. J. Educ. Manag. Eng.*, vol. 8, no. 2, p. 1, 2018.

[27] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

[28] F.-J. Yang, "An implementation of naive bayes classifier," in *2018 International conference on computational science and computational intelligence (CSCI)*, 2018, pp. 301–306.

[29] C. Slamet, R. Andrian, D. S. Maylawati, W. Darmalaksana, and M. A. Ramdhani, "Web scraping and Naïve Bayes classification for job search engine," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 288, no. 1, p. 12038.

[30] D. Berrar, "Bayes' theorem and naive bayes classifier," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. January 2018, pp. 403–412, 2018, doi: 10.1016/B978-0-12-809633-8.20473-1.