JOIV : Int. J. Inform. Visualization, 8(3) - September 2024 1162-1167

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Determination of Training Participants in Community Work Training Centers Using the Naïve Bayes Classifier Algorithm

April Lia Hananto [a,*], Agustia Hananto [b], Baenil Huda [a], Aviv Yuniar Rahman [a], Elfina Novalia [a], Bayu Priyatna [a]

[a] *Faculty of Computer Science, Universitas Buana Perjuangan Karawang, East Teluk Jambe, Karawang, Indonesia*
[b] *Faculty of Informatics Engineering, Universitas Widyagama Malang, Lowokwaru, Malang, Indonesia*
Corresponding author: [*]aprilia@ubpkarawang.ac.id

***Abstract*—Community work training centers are skills training institutions that aim to improve the skills of the surrounding community by providing training programs that align with industry needs. Registration of training participants at the Al-Ikhwan Islamic Boarding School community work training centers often faces obstacles, namely, the selection process is still manual, so it takes a long time, and there is a possibility of errors. This study aims to apply the Naive Bayes Classifier Algorithm to determine whether applicants pass training at the Al-Ikhwan Islamic Boarding School community work training centers. This classification method is used to help optimize the applicant selection process by considering administrative factors, income, and training quotas. RapidMiner software is used as a tool to implement the algorithm. This study found that the Naive Bayes Classifier Algorithm can provide good accuracy results in determining applicants who pass the training selection. The test results show that the resulting model has an accuracy of 90.00% in determining passing training participants with data that has the highest chance of passing, namely data that has the attributes of the female gender, age 20 years, last education Senior High School/Vocational High School, student work/student, income 364,912, father's work as laborer, father's income 3912,280, mother's work as an IRT, and mother's income 885,964. This research increases efficiency and accuracy in determining training applicants at the Al-Ikhwan Islamic Boarding School community work training centers.**

***Keywords*— Data mining; classification; acceptance of participants; naïve bayes.**

Manuscript received 18 Jul. 2023; revised 2 Dec. 2023; accepted 18 Jan. 2024. Date of publication 30 Sep. 2024.

I. Introduction

Community work training centers are vocational training centers established in religious institutions or non-governmental educational institutions. The Al-Ikhwan Islamic Boarding School community work training centers were established at the Al-Ikhwan Islamic Boarding School. Community work training centers were established to provide technical skills training or vocational expertise to the demands of the labor market and the surrounding community. It is intended that these people have the provision to find a job or run their own business. Currently, the Al-Ikhwan Islamic Boarding School community work training centers have only one major in expertise, namely graphic design; through this department training, participants will receive material on design skills free of charge.

If someone wants to register for training, that person must contact the work training center manager's WhatsApp number. The contact person will ask for the registrant's data via a short message and then input it in a spreadsheet, which takes time. Whether or not applicants pass is determined based on administrative requirements, income, and training quotas. The work training centers manager must sort according to the requirements to determine the qualified participants. The registration and selection process, which only relies on visual observation, tends to be inefficient, less accurate and increases the possibility of human error [1]. Data mining techniques are needed to assist in determining the process of determining applicants who pass the training [2].

Data mining is finding valuable patterns or trends in a set of data [3]. The method used in data mining is classification. The classification method that will be applied is the Bayes theorem approach, known as the Naïve Bayes Classification method. The principle behind this method is to take advantage of statistical opportunities to combine previous knowledge with new information [4]. The prominent uniqueness of the Naive Bayes Classifier algorithm is the strong assumption about the independence between each event or criteria

involved. This algorithm has a simple structure and high speed in training and data classification [1], [5]. This study aims to explain the process and results of applying the Naive Bayes Classifier Algorithm in determining applicants who pass the training selection at the Al-Ikhwan Islamic Boarding School community work training centers.

II. Materials and Method

In determining applicants who pass the selection, we will later use an algorithm called Naïve Bayes Classifier.

A. Data Mining

Mining data is extracting large amounts of data and analyzing it to obtain something valid, new, and valuable. As a result, you can find patterns or formulas in the data [6]. Data mining can generally be grouped into two main categories: descriptive and predictive. Descriptive Mining is the process of finding important information from a database. Meanwhile, Predictive Mining is the process of determining data polarity by specifying several variables to estimate other variables in the future [7], [8].

B. Classification

Classification is a stage in data analysis that aims to find a model or function that can describe and differentiate the classes or concepts contained in the data [9]. Put, classification gives categories or labels to data based on certain qualities. Classification aims to produce a model to predict the class or category of the database/group of data [10]. There are three stages of classification: model construction, model application, and evaluation [11], [12].

C. Naïve Bayes Classifier

Naive Bayes is the most straightforward and most commonly used data mining classification [13], [14]. Naïve Bayes can predict a class's membership probability based on predictions of independent assumptions. The Naïve Bayes Classifier predicts future events' probability based on experience data [4], [15]. The prominent uniqueness of the Naive Bayes Classifier algorithm is the strong assumption about the independence between each event or criteria involved [16], [17]. This algorithm has a simple structure and high speed in training and data classification. This algorithm has a simple structure and has high speed in the training process and data classification [1], [18]–[20]. The Naïve Bayes formula used can be seen in Formula 1.

$$P\left(\frac{X}{Y}\right)=\frac{P(X)P\left(\frac{Y}{X}\right)}{P(Y)} \quad (1)$$

Where:
X = custom class
P(X/Y) = probability on the R hypothesis based on S conditions
P(X) = hypothesis probability of the R-value
P(Y/X) = probability S based on the value of the hypothesis R
P(Y) = probability value of S

D. Rapid Minner

Rapid Minner is software or software used to process data [21]. RapidMiner can be a solution for problems that require analysis of data mining, text mining, and prediction analysis [22]. In research conducted by Gagan Suganda, a scholarship recipient information system was successfully created using the Naïve Bayes Classifier algorithm. The results showed that the system achieved an accuracy of 96.56% in the first test, while in the second test, it reached 90.33%. Accuracy is included in the very good category. This study found that using this system's Naive Bayes Classifier algorithm resulted in high accuracy. After seeing the results of the research above, which showed a high level of accuracy in classification using the Naive Bayes Classifier method, the researchers decided to apply this method to the system to determine the acceptance of trainees. The expected results of this study are the percentage accuracy of applying the Naive Bayes Classifier Algorithm in determining applicants who pass the training selection at the Al-Ikhwan Islamic Boarding School community work training centers. The following are the Research Stages in this research which are shown in Fig. 1:

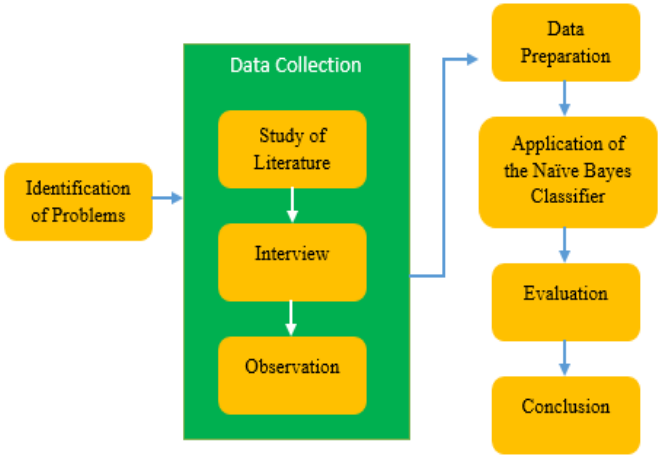

Fig. 1 Research Stage [15].

A. Identification of Problems

The first step is to gain an understanding of the goals and requirements of business processes, then translate this understanding into defining problems in data mining. After that, plans and strategies will be made to achieve these goals. At this stage, identification of problems that must be solved clearly will be carried out, as well as setting research objectives [15].

B. Study of Literature

Collecting and analyzing literature, books, national and international journals related to the Naive Bayes method, classification, and other related topics. This literature review helps to understand related research that has been done previously, recent advances, and relevant theoretical frameworks [23].

C. Data Collection

At this stage, literature study data collection was carried out. Literature study is the process of collecting and analyzing literature, books, national and international journals related to the Naive Bayes method, classification, and other related topics. This literature review helps to understand related research that has been done previously, recent advances, and relevant theoretical frameworks [23]–[25].

In addition to literature studies, data collection was also carried out through direct interviews with work training centers managers and studying documents available at work training centers.

D. Data Preparation

The data preparation stage is the stage in which there is a processing process from raw data to data that is ready to be processed [26], [27]. At this stage, the selection of tables, records, and data attributes to be used is carried out, including carrying out the process of cleaning and transforming the data so that it can become the appropriate input in the modeling stage [28], [29]. From this stage, one hundred data were obtained that are ready to be processed and already have pass and do not pass labels. Of the hundred data, data distribution was carried out: eighty for training and twenty for testing.

E. Application of the Naïve Bayes Classifier

In this stage, modeling is done by applying the naïve Bayes classifier algorithm using the RapidMiner Studio software.

F. Evaluation

Validation involves the process of evaluating the accuracy of a model by comparing the predictions generated using the model with actual results [30]. Evaluation of the accuracy of the classification model is critical because accuracy is an indicator of the model's ability to predict the target class [31]. Evaluation can be done through data testing, which has a certain value that is not owned by the training data [32]. Evaluation is done through accuracy assessment. Accuracy is the ratio of correct predictions (true) where the calculation can be seen in formula 2 [37]. Calculation of accuracy cannot be separated from precision and recall; precision is a metric that measures the accuracy of the results of a model. The formula is the ratio between the number of true positives and the total amount of data with a positive label, which can be seen in formula 3 [38]. Recall is a metric that measures a model's completeness level. The recall formula is the comparison between the number of true positives and the total number of examples that are actually positive, which can be seen in Formula 4 [33]. The following is the calculation formula:

Accuracy is calculated by dividing the number of correct predictions (True Positives and True Negatives) by the total amount of data [34].

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

Precision measures the number of correct positive predictions. It is calculated by dividing the number of True Positives by the total positive predictions (True Positives and False Positives) [35].

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

Recall measures the extent to which the model can capture all instances that should be positive. It is calculated by dividing the number of True Positives by the total instances that should be positive (True Positives and False Negatives). Recall is useful to ensure that the model does not miss positive instances that should be classified [36].

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

Information:
TP or True Positive = The number of positive and predicted data is correct as positive
FP or False Positive = The number of data is negative but, predicted as positive
FN or False Negatives = The number of positive data, but, is predicted as negative
TN or True Negative = Number of negative data and correctly predicted as negative

III. Results and Discussion

A. Data Collection

Data was collected through in-depth interviews and direct observation at the Al-Ikhwan Islamic Boarding School Community Work Training Center [40]. This process involves direct interaction with prospective trainees, allowing researchers to understand everyone's characteristics and background. From this stage, data relevant to the attribute criteria that are the research focus is obtained. The interview process allows researchers to gain direct insight from respondents, while observations provide valuable contextual information for understanding actual conditions in the field. This combination of methods ensures that the resulting dataset reflects a situation appropriate to the context of Al-Ikhwan Islamic Boarding School Community Job Training. From this stage, data is obtained with the following attribute criteria as follows:

- Gender
- Age
- Last education
- Work
- Income
- Status of Father & Mother
- Father's and Mother's Occupation
- Father and Mother's income.

B. Data Preparation

The obtained data is then cleaned and transformed to match the attributes needed in the modeling process [39]. From the preparation results, one hundred data can be entered into the modeling stage with the Naïve Bayes algorithm. Of the one hundred data obtained from the research object, eighty data were divided for training, and twenty data were used for testing. The following are the Dataset Specifications and Attributes Used for Training, which is shown in Table I:

TABLE I
Dataset specifications and attributes used for training

No	Attribute Name	Information	Amount
1	Gender	Gender namely Male and Female	W = 39, F = 41
2	Age	This is the age of the registrant	Range = 18 - 32 Average = 21.75 Standard Deviation = 4.1
3	Last education	The applicant's latest education, namely elementary, junior high, high school, D3, S1/D4	SD = 0 Junior High School = 24 SMA = 55 D3 = 0 S1/D4 = 1

No	Attribute Name	Information	Amount
4	Work	Occupations of applicants, namely Students, Entrepreneurs, & Private Employees	Student / Student = 62 Entrepreneur = 9 Private Employees = 9
5	Income	Income from applicants (in rupiah)	Range = 0 - 8000000 Average = 1218750 Standard Deviation = 2307670
6	Father status	The status of the registrant's father, namely Life & Death	Life = 74 Died = 6
7	Father's occupation	Occupation of the registrar's father	Civil servants = 12 Private Employees = 12 Labor = 22 Farmers = 2 Entrepreneur = 26 Not Working = 6
8	Father's income	Income from the registrar's father	Range = 0 – 15000000 Average = 4560000 Standard Deviation = 2728963
9	Mother Status	The status of the registrant's mother, namely Life & Death	Life = 61 Died = 19
10	Mother's job	Occupation of the registrar's father	civil servant = 2 Private Employees = 1 Labor = 8 Farmer = 1 Entrepreneur = 16 IRT = 32 Not Working = 19
11	Mother's income	Income from the registrar's mother	Range = 0 – 8000000 Average = 1325000 Standard Deviation = 2224860

The training dataset and testing data are presented as a Data Sharing table, which refers to the pro-positive and negative sentiment variables. The following is the Data Sharing shown in Table II:

TABLE II
DATA SHARING

Sentiment	Training Data	Data Testing	Amount
Positive	57	11	62
Negative	23	9	32
Amount	80	20	100

There are 52 Positive Sentiments in trending data and 11 Trending Data. Meanwhile, there were 23 negatives for data sharing for trending data and 9 for testing data. The number of Positive and negative Amounts is 100.

C. Application of the Naïve Bayes Classifier

Once the data is ready, modeling is carried out using the naïve Bayes classifier. Implementation of Naïve Bayes using RapidMinner Studio software. The model applied to carry out the trending and testing process with results that can be presented in Fig. 2 Modeling in the RapidMiner Application:

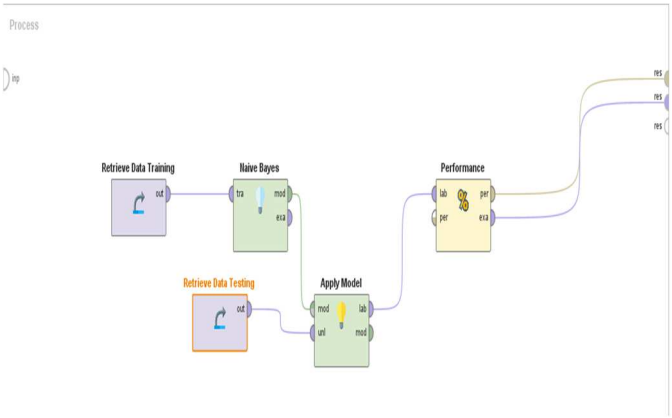

Fig. 2 Modeling on the RapidMiner Application [12].

D. Evaluation

From the results of implementing the naïve Bayes algorithm (modeling) in RapidMiner Studio, confusion matrix results were obtained. The following explains the results of Accuracy, Precision, and Recall, which are shown in Table III:

TABLE III
CONFUSION MATRIX

		Actual	
		Getaway	**Did not pass**
predictions	Getaway	9	0
	Did not pass	2	9

Where:
Accuracy = (9+9)/(9+9+0+2)
= 18/20
= 0.9 (90%)
Precision = 9/(9+0)
= 9/9
= 1 (100%)
Recall = 9/(9+2)
= 9/10
= 0.9 (90%)

Distribution of data obtained based on the attributes used as well as Gender, Age, Last education, Work, Father status, Father's occupation, Father's income, and Mother's job. The following are the results of the resulting data distribution shown in Table IV:

TABLE IV
DISTRIBUTION

Attributes	Parameter	Getaway	Did not pass
Gender	value=Male	0.439	0.565
Gender	value=Female	0.561	0.435
Gender	value=unknown	0	0.001
Age	mean	20.158	25,696
Age	standard deviation	2,541	4,626
Last education	value=SMP/MTs	0.351	0.174
Last education	value=SMA/SMK/MA	0.649	0.781
Last education	value=S1	0	0.044
Last education	value=unknown	0	0.001
Work	value=Student/Student	0.912	0.434
Work	value=Entrepreneur	0.07	0.217

Attributes	Parameter	Getaway	Did not pass
Work	value=Private Employee	0.018	0.348
Work	value=unknown	0	0.001
Income	mean	364912,617	3334782,609
Income	standard deviation	1221458,62	2958901,623
Father status	value=On	0.929	0.912
Father status	value=Dead	0.07	0.087
Father status	value=unknown	0	0.001
Father's occupation	value=Private Employee	0.193	0.044
Father's occupation	value=Entrepreneur	0.333	0.304
Father's occupation	value=PNS	0.018	0.477
Father's occupation	value=Labor	0.351	0.087
Father's occupation	value=-	0.07	0.087
Father's occupation	value=Farmers	0.035	0.001
Father's occupation	value=unknown	0	0.001
Father's income	mean	3912280,7	6165217,391
Father's income	standard deviation	2265586,5	3145792,095
Mother Status	value=On	0.719	0.869
Mother Status	value=Dead	0.281	0.131
Mother Status	value=unknown	0	0.001
Mother's job	value=IRT	0.403	0.433
Mother's job	value=-	0.28	0.13
Mother's job	value=Entrepreneur	0.158	0.304
Mother's job	value=Labor	0.123	0.044
Mother's job	value=PNS	0	0.087
Mother's job	value=Farmers	0.018	0.001
Mother's job	value=Private Employee	0.018	0.001
Mother's job	value=unknown	0	0.001
Mother's income	mean	885964,912	2413043,478
Mother's income	standard deviation	1520742,14	3186092,555

From the Distribution Table, it can be presented in the form of data visualization. The following is a visualization of the evaluation results shown in Fig. 3:

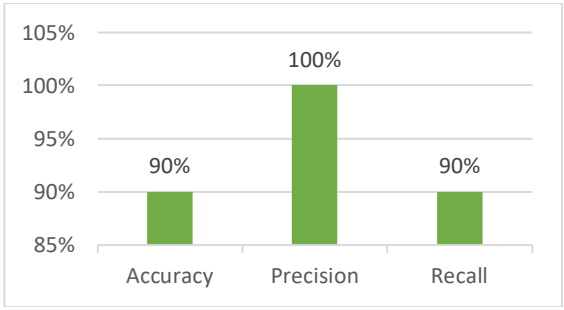

Fig. 3 Visualization of Evaluation Results

The results of using Molde Naïve Bayes on the dataset used in this research show 90% accuracy, 100% precision, and 90% recall. These results are explained in detail in point IV. Conclusion. The results of the prediction visualization accuracy can be seen in Fig. 4 below:

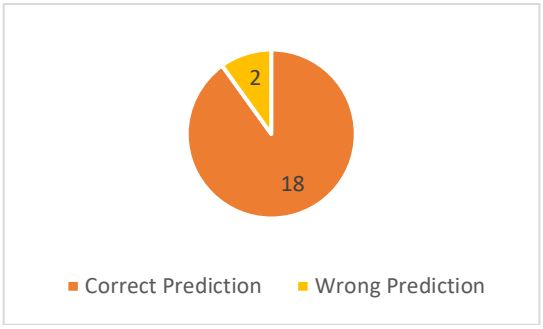

Fig. 4 Accuracy Visualization

IV. Conclusion

From the results of research using a total of one hundred data with eighty training data and twenty test data, it can be concluded that the method for determining whether or not to pass registration at the Al-Ikhwan Center Islamic Boarding School community work training using the Naïve Bayes algorithm has an accuracy of 90.00% so that it can be categorized as good with data that has the highest chance of passing, namely data that has the attributes female gender, age 20 years, highest school education/vocational school, student occupation, income 364,912, father's occupation as a laborer, father's income 3912,280, mother's occupation as a mother. household, and mother's income 885,964. Not much data was mis predicted; out of a total of 20 data mis predicted, there were 2 data.

The misclassification rate was minimal, with only 2 out of 20 test data inaccurately predicted. These results underscore the power and reliability of the Naïve Bayes algorithm in this context. Therefore, this research provides a valuable contribution to increasing efficiency and accuracy in the selection process of training applicants at the Al-Ikhwan Islamic Boarding School community work training center.

The findings show that this algorithm can effectively simplify the identification of candidates with a high probability of success, thereby optimizing the resources and efforts invested in training programs. Suggestions for future research could include increasing the number of datasets to determine better accuracy. You can also compare the algorithm's performance with other algorithms such as K-NN, Random Forest, and Tree.

References

[1] A. Rahman and A. Suryanto, "Implementasi Sistem Informasi Seleksi Penerima Beasiswa Dengan Metode Naive Bayes Classifier," *J. Penelit. Pendidik. Indones.*, vol. 2, no. 3, pp. 1–8, 2017.

[2] I. Issah, O. Appiah, P. Appiahene, and F. Inusah, "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," Decision Analytics Journal, vol. 7, p. 100204, Jun. 2023, doi:10.1016/j.dajour.2023.100204.

[3] I. Bahuguna, U. Kumar, K. Arunachalam, V. Shridhar, and A. Sharma, "Determinants of Natural Resources Based Microenterprises Performance in India's Western Himalayan Region: A Naïve Bayes Classifier Analysis," International Journal of Small and Medium Enterprises, vol. 6, no. 1, pp. 27–40, Jun. 2023,

doi:10.46281/ijsmes.v6i1.2032.
[4] Gagan Suganda, Marsani Asfi, Ridho Taufiq Subagio, and Ricky Perdana Kusuma, “Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (KIP) Kuliah Menggunakan Naïve Bayes Classifier,” JSiI (Jurnal Sistem Informasi), vol. 9, no. 2, pp. 193–199, Sep. 2022, doi: 10.30656/jsii.v9i2.4376.
[5] R. G. Almonte, C. R. Malizon, and J. N. Olimpiada, “Sentiment Analysis on Students’ Satisfaction of the Tanauan City College using Naïve Bayes Algorithm,” 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Nov. 2022, doi: 10.1109/iceccme55909.2022.9988637.
[6] R. Amalia, “Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode naïve bayes,” *J. Inform. dan Sist. Inf*, vol. 6, no. 1, pp. 33–42, 2020.
[7] A. Pebdika *et al.*, “Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima PIP,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, 2023.
[8] A. Lia Hananto et al., “Analysis Of Drug Data Mining With Clustering Technique Using K-Means Algorithm,” Journal of Physics: Conference Series, vol. 1908, no. 1, p. 012024, Jun. 2021, doi:10.1088/1742-6596/1908/1/012024.
[9] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, “Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review,” Faktor Exacta, vol. 13, no. 1, p. 35, Jun. 2020, doi: 10.30998/faktorexacta.v13i1.5548.
[10] H. K. Saputra, “Analisis Data Mining Untuk Pemetaan Mahasiswa yang Membutuhkan Bimbingan dan Konseling Menggunakan Algoritma Naïve Bayes Classifier,” Jurnal Teknologi Informasi dan Pendidikan, vol. 11, no. 1, pp. 14–26, Apr. 2018, doi:10.24036/tip.v11i1.104.
[11] N. W. Wardani and N. K. Ariasih, “Analisa komparasi algoritma decision tree c4. 5 dan naïve bayes untuk prediksi churn berdasarkan kelas pelanggan retail,” *Int. J. Nat. Sci. Eng.*, vol. 3, no. 3, pp. 103–112, 2019.
[12] J. K. Sethi and M. Mittal, “Efficient weighted naive bayes classifiers to predict air quality index,” Earth Science Informatics, vol. 15, no. 1, pp. 541–552, Jan. 2022, doi: 10.1007/s12145-021-00755-7.
[13] F. Prasetya and F. Ferdiansyah, “Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes,” Jurnal Sistem Komputer dan Informatika (JSON), vol. 4, no. 1, p. 132, Sep. 2022, doi: 10.30865/json.v4i1.4852.
[14] M. Wongkar and A. Angdresey, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter,” 2019 Fourth International Conference on Informatics and Computing (ICIC), Oct. 2019, doi: 10.1109/icic47613.2019.8985884.
[15] I. Kurniawan *et al.*, “Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 1, pp. 731–740, 2023.
[16] T. H. Roza et al., “Suicide risk classification with machine learning techniques in a large Brazilian community sample,” Psychiatry Research, vol. 325, p. 115258, Jul. 2023, doi:10.1016/j.psychres.2023.115258.
[17] R. Panigrahi et al., “Intrusion detection in cyber–physical environment using hybrid Naïve Bayes—Decision table and multi-objective evolutionary feature selection,” Computer Communications, vol. 188, pp. 133–144, Apr. 2022, doi: 10.1016/j.comcom.2022.03.009.
[18] A. L. . Hananto, A. P. . Nardilasari, A. . Fauzi, A. . Hananto, B. . Priyatna, and A. Y. . Rahman, “Best Algorithm in Sentiment Analysis of Presidential Election in Indonesia on Twitter”, Int J Intell Syst Appl Eng, vol. 11, no. 6s, pp. 473–481, May 2023.
[19] P. Padeli, A. Martono, and S. Sudaryono, “Application of Naive Bayes Model, SVM and Deep Learning Predicting,” CICES, vol. 9, no. 1, pp. 93–101, Feb. 2023, doi: 10.33050/cices.v9i1.2584.
[20] K. Airin Fariza Abu Samah, N. Farhanah Amirah Misdan, M. Nor Hajar Hasrol Jono, and L. Septem Riza, “The Best Malaysian Airline Companies Visualization through Bilingual Twitter Sentiment Analysis: A Machine Learning Classification,” *Int. J. Informatics Vis.*, vol. 6, no. 1, pp. 130–137, 2022.
[21] D. Novianti, “Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner,” *Parad. Komput. dan Inf.*, vol. 21, no. 1, pp. 49–54, 2019.
[22] E. Fitriani, “Perbandingan Algoritma C4. 5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan,” *Sistemasi*, vol. 9, no. 1, pp. 103–115, 2020.
[23] S. A. Yusuf and U. Khasanah, “Kajian Literatur Dan Teori Sosial Dalam Penelitian,” *Metod. Penelit. Ekon. Syariah*, vol. 80, pp. 1–23, 2019.
[24] Y. Bustomi, A. Nugraha, C. Juliane, and S. Rahayu, “Data Mining Selection of Prospective Government Employees with Employment Agreements using Naive Bayes Classifier,” Sinkron, vol. 8, no. 1, pp. 1–8, Jan. 2023, doi: 10.33395/sinkron.v8i1.11968.
[25] D. Fahrudy and S. ’Uyun, “Classification of Student Graduation using NaÃ¯ve Bayes by Comparing between Random Oversampling and Feature Selections of Information Gain and Forward Selection,” JOIV : International Journal on Informatics Visualization, vol. 6, no. 4, p. 798, Dec. 2022, doi: 10.30630/joiv.6.4.982.
[26] T. P. Prast, H. Zakaria, and P. Wiliantoro, “Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier,” *OKTAL J. Ilmu Komput. dan Sains*, vol. 1, no. 06, pp. 573–582, 2022.
[27] B. Marapelli, S. Kadiyala, and C. S. Potluri, “Performance Analysis and Classification of Class Imbalanced Dataset Using Complement Naive Bayes Approach,” 2023 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), Jan. 2023, doi: 10.1109/accthpa57160.2023.10083369.
[28] Z. M. Ali, N. H. Hassoon, W. S. Ahmed, and H. N. Abed, “The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification,” International Journal of Psychosocial Rehabilitation, vol. 24, no. 03, pp. 2143–2151, Feb. 2020, doi: 10.37200/ijpr/v24i3/pr200962.
[29] M. F. A. Saputra, T. Widiyaningtyas, and A. P. Wibawa, “Illiteracy Classification Using K Means-NaÃ¯ve Bayes Algorithm,” JOIV : International Journal on Informatics Visualization, vol. 2, no. 3, pp. 153–158, May 2018, doi: 10.30630/joiv.2.3.129.
[30] B. Sunarko et al., “Prediction of Student Satisfaction with Academic Services Using Naive Bayes Classifier,” 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Dec. 2022, doi: 10.1109/icitisee57756.2022.10057736.
[31] T. Arifin and S. Syalwah, “Prediksi keberhasilan immunotherapy pada penyakit kutil dengan menggunakan algoritma naïve bayes,” *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 1, pp. 38–43, 2020.
[32] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, and B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM,” vol. 7, no. 1, pp. 11–18, 2024.
[33] G. S, S. R, D. B. M, and J. J, “Improved Sentimental Analysis to the Movie Reviews using Naive Bayes Classifier,” 2022 International Conference on Electronics and Renewable Systems (ICEARS), Mar. 2022, doi: 10.1109/icears53579.2022.9752408.