Journal of Science and Social Research ISSN 2615 – 4307 (Print)
April 2026, IX (2): 2828 – 2837 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

"COMPARATIVE STUDY OF KNN AND NAIVE BAYES ALGORITHMS WITH QUESTIONNAIRE DATA FOR STUDY PROGRAM RECOMMENDATION IN THE FACULTY OF ART AND DESIGN"

Surya Darma[1], Nita Syahputri[2], Nurhayati[3]
[1,3]Potensi Utama University, Medan
[2]Pembangunan Masyarakat Indonesia University, Medan
Email: [1]suryadarma090693@@gmail.com, [2]nieta20d@gmail.com,
[3]nurhayatimaulanaa@gmail.com.

Abstract: *The purpose of this study is to analyze and compare the performance of the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms in providing study program recommendations in the Faculty of Arts and Design. The data were obtained from 250 respondents through a questionnaire consisting of 20 indicators related to students' interests, abilities, creativity, technology, and career preferences. The research process included data preprocessing, data transformation, dataset splitting into training and testing data, modeling using the KNN and Naïve Bayes algorithms, and model performance evaluation using accuracy metrics. The data processing was carried out using the Python programming language on the Google Colab platform. The results showed that the KNN algorithm achieved an accuracy of 94%, while the Naïve Bayes algorithm obtained an accuracy of 92%. These findings indicate that the KNN algorithm performed better in classifying study program recommendations compared to the Naïve Bayes algorithm. It is expected that this research can serve as a foundation for developing a more effective decision support system to assist prospective students in selecting study programs that match their interests and abilities.*

Keywords: *K-Nearest Neighbor (KNN), Naïve Bayes, Machine Learning, Classification, Study Program Recommendation.*

Abstrak: Tujuan dari penelitian ini adalah untuk menganalisis dan membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes dalam memberikan rekomendasi program studi di Fakultas Seni dan Desain. Data diperoleh dari 250 responden melalui kuesioner yang terdiri dari 20 indikator yang berkaitan dengan minat, kemampuan, kreativitas, teknologi, dan preferensi karier mahasiswa. Proses preprocessing data, transformasi data, pembagian dataset menjadi data pelatihan dan pengujian, pemodelan menggunakan algoritma KNN dan Naive Bayes, dan evaluasi kinerja model dengan akurasi adalah bagian dari penelitian. Proses pengolahan data dilakukan pada platform Google Colab menggunakan bahasa pemrograman Python. Hasil penelitian menunjukkan bahwa algoritma KNN memiliki akurasi sebesar 94%, sedangkan algoritma Naïve Bayes memiliki akurasi sebesar 92%. Hasil ini menunjukkan bahwa algoritma KNN lebih baik dalam mengklasifikasikan rekomendasi program studi daripada algoritma Naïve Bayes. Diharapkan penelitian ini akan menjadi dasar untuk membuat sistem pendukung keputusan yang lebih baik yang membantu calon mahasiswa memilih program studi yang sesuai dengan minat dan kemampuan mereka.

Kata Kunci: KNN, Naïve Bayes, Machine Learning, Klasifikasi, Rekomendasi Program Studi.

INTRODUCTION

The development of information

technology and artificial intelligence has brought significant changes in various fields, including the world of education. One of the applications of this technology is in data-driven decision-making, particularly in helping individuals choose study programs that align with their interests, talents, and abilities. (Gyll, 2021), (Ayeni et al., 2024), (Pereira et al., 2025). Choosing the right study program is an important factor in determining someone's academic and career success in the future. (Schelfhout et al., 2021), (Missaghian, 2021), (Meyer et al., 2020).

In the Faculty of Arts and Design, there are various study programs with different characteristics and competencies, such as the Film and Television Study Program and Visual Communication Design. These two study programs have different focuses, with Film and Television emphasizing audiovisual production, while Visual Communication Design focuses on visual communication and graphic design.

Recommending the right study program for prospective students is a top priority for higher education institutions, especially in the Faculty of Arts and Design, which offers majors with diverse curriculum characteristics and graduate competencies. However, in practice, many prospective students face difficulties in making choices due to a lack of understanding of their interests and abilities, as well as the absence of an objective and structured decision support system. (Schelfhout et al., 2021), (Meens et al., 2020), (Reynolds et al., 2023).

Questionnaire data has long been used as the primary source of information to understand preferences, interests, and the determining factors for choosing study programs. (Zhao et al., 2025), (Bruen, 2021). In the era of data-driven decision making, exploring simple yet powerful classification algorithms such as K-Nearest Neighbors (KNN) and Naive Bayes can provide valuable insights into how questionnaire response patterns tend to guide study program recommendations, without requiring complex assumptions about variable relationships. (Jen, 2021), (Jin & Cutumisu, 2023), (Ha et al., 2023).

The use of these two algorithms allows for the comparison of classification performance relevant to the context of study program recommendations based on respondent characteristics, thereby helping to formulate recommendations that are more responsive to the needs of prospective students and the dynamics of the faculty.

This research aims to conduct a comparative study between the KNN and Naive Bayes algorithms in classifying study program preferences based on questionnaire data. Specifically: To examine the extent to which KNN and Naive Bayes can classify study program preferences in the Faculty of Arts and Design based on demographic variables, design interests, and medium preferences (e.g., conceptual media, field practice, or technical experiments). (Jen, 2021), (Jin & Cutumisu, 2023).

Evaluating the performance of both algorithms using accuracy, precision, recall, and F1-score metrics on training and test data. Identifying the most influential important factors (features) in study program recommendations and discussing their implications for curriculum development and mapping student interests to study paths.

Various studies have adopted simple algorithms such as KNN and Naive Bayes for classification based on questionnaire or survey data. KNN is an instance-based algorithm that uses the distance between samples to determine the class, thus not requiring strong data distribution assumptions but relying on the selection of the number of neighbors (k) and consistent feature scaling. (Jen, 2021), (Jin & Cutumisu, 2023), (Agarwal et al., 2021)

Meanwhile, Naive Bayes is a probabilistic class that assumes independence between features (naive independence) and often provides stable performance on data with relatively small sample sizes or when the relationships between features are not strongly

dependent on each other. The comparison of the two is beneficial for the context of study program recommendations. (Jen, 2021), (Jin & Cutumisu, 2023), (Agarwal et al., 2021), (Patulin, 2019), (Kurban et al., 2021).

KNN tends to be responsive to variations in questionnaire response patterns without requiring data distribution assumptions, thus capturing non-linear patterns that may emerge in student preferences. Naive Bayes can work well on questionnaire data with categorical or discretized features and provide clear probabilistic estimates of the generated recommendations. (Sari et al., 2023), (Tapidingan & Paseru, 2020).

The context of higher education and study program recommendations has long emphasized the importance of empirical data to understand student choices and the alignment between curriculum, facilities, and student interests. Although the specific literature on the combination of KNN-Naive Bayes on questionnaire data for study program recommendations in the Faculty of Arts and Design is still relatively limited, the comparative approach between these two algorithms is generally relevant to support data-driven decision-making practices in both administrative and curricular contexts.

This research uses questionnaire data that includes demographic variables (Name/Initials, Age, Gender, and Educational Background). Additionally, the questionnaire questions cover: the use of DKV interest variables, Film and TV interest, design skills, audiovisual skills, creativity, technology, career preferences, and prospective students' experiences.

The target variables are the Visual Communication Design study program, as well as Film and Television, which are available at the Faculty of Arts and Design. Data were collected thru closed and open questionnaires that have been processed for classification purposes. This research limits its focus to two main algorithms (KNN and Naive Bayes) for clear and replicable comparative purposes, and discusses their implications for academic policy and curriculum planning.

Based on the description, this research is conducted with the aim of analyzing and comparing the performance of the KNN and Naive Bayes algorithms in providing study program recommendations based on questionnaire data. The results of this research are expected to contribute to the development of a more effective and accurate decision support system, as well as assist prospective students in determining study program choices that align with their potential.

RESEARCH METHOD

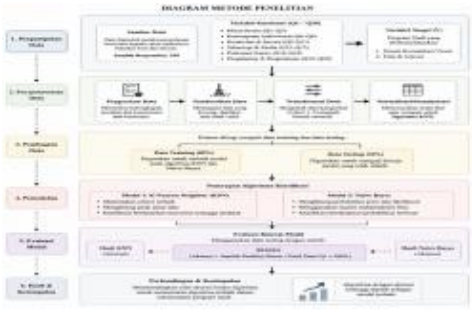


Figure 1 Research Methodology Diagram

Quantitative methods and a comparative approach are used in this study to compare the performance of the K-Nearest Neighbor (KNN) algorithm and Naïve Bayes in recommending study programs at the Faculty of Arts and Design. Research data were collected thru the distribution of questionnaires to 250 prospective students. The research instrument consists of twenty statements (Q1–Q20) that cover elements of design interest, audiovisual skills, creativity, technology, and career preferences. These statements are rated using a 1–5 Likert scale.

The independent variable (X) is the value from Q1 to Q20, and the dependent variable (Y) is the outcome of the Visual Communication Design and Film & Television study programs. The research

stages begin with data collection, preprocessing, transformation of categorical data into numerical form, and splitting the dataset into 80% training data and 20% testing data.

Modeling was conducted using the KNN and Naive Bayes algorithms on the Google Colab platform using the Python programming language. Then, the performance of both algorithms was evaluated using accuracy parameters to determine which algorithm has the best classification level for providing study program recommendations.

RESULTS AND DISCUSSION

This research uses primary data obtained thru the distribution of a structured questionnaire to 250 student respondents. The questionnaire instrument includes 20 statements (Q1-Q20) designed to represent the cognitive preferences, creative interests, and technical proficiency of prospective students in the domain of art and design. Respondents provide assessments using a 1-5 Likert scale to measure the degree of inclination toward certain variables.

In addition, the following questions were used in the survey:

Table 1 Questionnaire Statements

Statement	Description	Rating Scale
Q1	I am interested in graphic design activities such as creating posters or illustrations.	"1-5"
Q2	I enjoy processing visual elements such as color, typography, and layout.	
Q3	I keep up with the latest trends in visual design on social media or the internet.	
Q4	I am interested in the world of filmmaking and video production	
Q5	I enjoy watching and analyzing films or audiovisual content.	
Q6	I am interested in the filmmaking process (shooting, directing, etc.).	
Q7	I am capable of using design software such as Photoshop or Illustrator.	
Q8	I have drawing skills, both manual and digital.	
Q9	I understand the basic principles of design (composition, color, typography).	
Q10	I am capable of using video editing software such as Premiere or CapCut	
Q11	I understand the basic techniques of shooting (angle, lighting, etc.).	
Q12	I have made videos or short films.	
Q13	I am capable of generating creative ideas in creating visual works.	
Q14	I enjoy experimenting with new concepts in my creations.	
Q15	I am confident in the creative works I produce.	
Q16	I am accustomed to using digital technology in the creative process.	
Q17	I quickly learn new software or tools.	

Q18	I want to pursue a career in graphic/visual design.
Q19	I want to pursue a career in film, video, or television.
Q20	I have participated in training/courses or projects in the field of design or film.
Recommendation	Recommendation result

The questionnaire instrument consists of statements represented by variables Q1 to Q20 in Table 1. The study program recommendations for the Faculty of Arts and Design are made based on the scale values of several statement items grouped according to the characteristics of the study programs. Each statement is designed to measure specific elements related to the interests, abilities, creativity, and preferences of the respondents.

Meanwhile, the Film and Television Study Program is based on the scores of statement items Q1, Q2, Q3, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18, and Q20. These items indicate an interest in visual design, graphic design skills, visual creativity, understanding of digital technology, and career tendencies in visual communication design.

To collect data, questionnaires were distributed to prospective new students of the Faculty of Arts and Design. This process resulted in 250 respondent data used as the research dataset. Next, the collected data is processed using the previously planned research method stages. These stages include data selection, preprocessing, classification using the K-Nearest Neighbor (KNN) and Naive Bayes algorithms, and model performance evaluation to produce optimal study program results. Next, to facilitate the analysis and interpretation of the research results, the respondents' data will be presented in tabular form.

Table 2 Questionnaire Content

ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
1	1	5	4	4	1	4	4	3	4	1	1	4	5	5	2	5	5	4	4	5
2	3	4	5	2	5	5	2	1	4	2	4	4	1	2	4	5	2	1	1	2
3	2	5	2	3	4	4	4	1	3	2	3	1	1	4	4	3	3	3	3	5
4	3	1	5	3	5	5	3	2	1	3	3	3	3	2	5	1	4	3	4	2
5	3	1	3	1	1	3	1	3	3	3	1	5	3	4	4	4	2	4	2	3
6	2	2	4	5	1	4	1	2	4	2	3	1	4	2	4	1	1	4	3	4
7	4	2	4	2	4	3	3	1	3	4	1	3	2	2	5	4	3	4	3	4
8	2	5	2	2	5	2	3	5	2	1	5	3	3	3	3	4	1	4	5	2
9	5	4	4	4	5	5	5	2	4	2	3	3	4	3	5	1	3	1	3	5
10	2	4	3	5	4	1	1	3	4	4	2	4	4	3	2	4	5	2	5	4
11	2	1	4	1	4	1	1	4	3	3	4	3	1	2	5	5	1	1	1	1
12	3	4	2	4	5	5	5	5	1	4	5	2	4	2	1	5	3	4	4	2
13	4	5	5	4	2	3	2	4	1	3	4	5	5	1	2	4	1	1	3	4
14	4	1	5	5	2	5	2	1	3	4	1	3	5	5	1	1	2	1	2	2
15	4	2	5	1	4	4	3	4	1	1	3	4	3	2	2	5	4	4	2	4
16	1	2	2	2	2	1	1	3	4	1	4	4	4	2	2	2	2	4	5	2
17	2	3	2	4	4	3	2	4	1	1	5	1	1	4	1	3	3	5	3	5
18	5	2	5	1	2	1	2	5	3	2	5	1	2	3	4	2	1	2	1	3
19	5	4	5	4	1	5	4	1	1	3	2	3	3	5	1	4	1	2	2	3
20	5	3	5	5	3	4	2	5	4	2	3	1	4	5	5	5	4	4	3	3
21	3	3	2	1	4	1	4	3	5	5	1	2	2	4	4	5	1	2	2	1
22	1	2	2	3	5	4	1	3	2	2	1	3	2	5	2	3	2	3	5	2
23	1	1	4	1	5	3	3	4	3	2	3	2	3	5	3	2	1	3	1	2
24	5	2	3	5	2	2	4	1	2	3	3	5	3	1	4	4	1	1	4	1
25	3	5	3	5	4	3	5	2	5	5	5	5	4	3	4	1	2	1	3	4
26	3	4	5	3	3	3	3	5	3	3	5	3	1	4	5	2	3	3	1	5
27	2	3	1	5	2	2	1	3	5	3	1	3	2	5	5	4	4	1	3	4
28	2	4	3	4	1	2	5	1	5	2	1	1	4	2	3	2	5	3	5	3
29	4	5	2	4	3	4	3	3	1	1	3	2	3	2	4	4	1	4	2	1
30	4	4	3	4	4	3	5	5	5	1	3	2	3	3	2	1	4	3	2	3

Table 3 Recommendation Results Table

Value	Recommendation Result	Statement Source
1	Visual Communication Design	Q1 ,Q2, Q3, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18, Q20
2	Film and	Q4, Q5, Q6, Q7, Q8, Q9, Q13, Q14, Q15, Q16,

Television	Q17, Q19, Q20

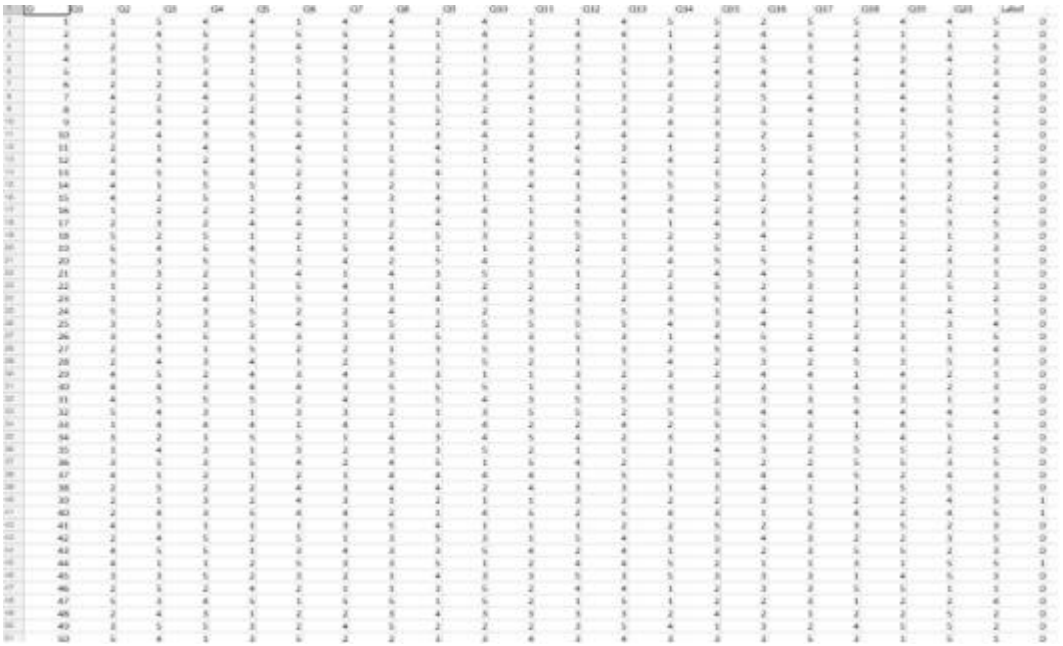
Pre-processing

At this stage, the data collected from the questionnaire results will be processed thru data processing, also known as data preprocessing. This will produce a structured, consistent dataset ready to be used in a Python-based system for the analysis process and testing the accuracy levels of the K-Nearest Neighbor (KNN) and Naive Bayes algorithms. The preprocessing stage is very important in machine learning research due to its significant impact on performance.

The "data cleaning" process is one of the preprocessing stages carried out. The purpose of this process is to clean the data of unnecessary elements, reduce data inconsistencies, and ensure that all data is formatted correctly so that the system can process it. This process is also carried out to reduce errors during the classification and model testing phases.

In this study, several features were converted from categorical or text data into numerical data to perform the "data cleaning" process. Because the KNN and Naive Bayes algorithms work better on numerical data, this data transformation is necessary. One of the attributes that underwent transformation is the study program recommendation attribute. Previously, features such as "Visual Communication Design" and "Film and Television" were converted into numerical representations so that the system could process them. The data transformation process is carried out with the assumption that the value "0" represents the Visual Communication Design Study Program (DKV), and the value "1" represents the Film and Television Study Program. The results of the data transformation are presented in Table 4.

Table 4 Recommendation Data Transformation



After the preprocessing stage, the dataset will be divided into two parts: "training data" and "test data." The purpose of this data splitting process is to train the classification model and test the algorithm's ability to make predictions on new data. The research data is divided into independent variables (features) and dependent variables (labels). The features from Q1 to Q20 are represented by the

value X, while the study program recommendation results are represented by the value Y.

After the variable separation process is complete, the data X is divided into "X_train" and "X_test", and the data Y is divided into "Y_train" and "Y_test". To ensure that the model can be trained and tested optimally, this division is carried out using the "split-train-test" method. From 250 respondent data, 80% is used as "training data" and the remaining 20% is used as "testing data." Thus, 200 data points are used for the model training process, and the other 50 data points are used for the model testing process.

Training data is used to create classification patterns in the K-Nearest Neighbor (KNN) and Naive Bayes algorithms. Meanwhile, testing data is used to evaluate the model's ability to provide accurate study program recommendation predictions. Next, the test results are evaluated using performance parameters such as accuracy, precision, recall, and confusion matrix. The goal is to determine which algorithm will have the best performance for this research.

Data Modeling

Data that has undergone the "data cleaning" process, "data transformation," and the division of the dataset into "training data" and "test data" will enter the modeling stage. This process is a key phase in the research because classification algorithms will be used to process the data that has been prepared earlier to generate predictions for the study program.

Two machine learning algorithms, K-Nearest Neighbor (KNN) and Naive Bayes, are used to perform the modeling process in this research. The KNN algorithm performs classification based on the distance between nearby data points. The system will compare the training data with the test data to calculate the nearest neighbors (*nearest neighbors*). Then, based on the majority of neighbors with the highest similarity, the system will assign a class. This method was chosen because it can recognize patterns based on the degree of similarity in the respondents' data characteristics.

However, the Naive Bayes algorithm is a probability-based classification technique that uses Bayes' Theorem for the prediction process. This algorithm calculates the probability of each class based on the characteristics of the respondents' data. It is well-known that Naive Bayes can handle classification data quickly and effectively.

At this modeling stage, both algorithms will be trained using the "training data" that has been prepared beforehand. Then, the created models will be tested using the "test data" to determine the algorithms' ability to provide appropriate study program recommendations. Subsequently, based on evaluation metrics such as accuracy, precision, recall, and confusion matrix, the prediction results from both methods will be compared to determine which algorithm has the best performance.

```
==== K-Nearest Neighbors ====
Accuracy  : 0.9400
Classification Report:
              precision    recall  f1-score   support

           0       0.94      1.00      0.97        47
           1       0.00      0.00      0.00         3

    accuracy                           0.94        50
   macro avg       0.47      0.50      0.48        50
weighted avg       0.88      0.94      0.91        50
```

Figure 2 KNN Algorithm Results

```
==== Naive Bayes ====
Accuracy  : 0.9200
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.98      0.96        47
           1       0.00      0.00      0.00         3

    accuracy                           0.92        50
   macro avg       0.47      0.49      0.48        50
weighted avg       0.88      0.92      0.90        50
```

Figure 3 Naïve Bayes Algorithm Results

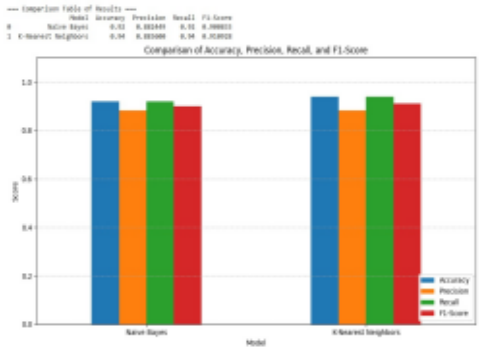


Figure 4 Comparison Chart

KNN and Naïve Bayes Algorithms

Based on the data processing results using the Python programming language on the Google Colab platform, it was found that the K-Nearest Neighbor (KNN) algorithm achieved an accuracy rate of 94%, while the Naïve Bayes algorithm obtained an accuracy of 92%. These results indicate that the KNN algorithm has better classification performance compared to Naïve Bayes in this study. The high accuracy of KNN is influenced by the method of calculating the distance between data points to determine classification based on the nearest neighbors, while Naïve Bayes uses a probabilistic approach with the assumption of independence between features. Thus, KNN is considered more effective in classifying data in this study.

CONCLUSION

The research results show that the K-Nearest Neighbor (KNN) and Naive Bayes algorithms can be used to classify study program recommendations in the Faculty of Arts and Design based on student questionnaire data. The KNN algorithm shows an accuracy of 94% and the Naive Bayes algorithm shows an accuracy of 92%.

Based on the comparison, the KNN algorithm outperformed Naïve Bayes in this study because it can classify data based on the proximity of characteristics among respondents more optimally. Meanwhile, Naïve Bayes also succeeded with the probability-based classification process. Therefore, the KNN algorithm is considered more efficient for use in the study program recommendation system of the Faculty of Arts and Design. This research is intended to serve as a reference in the process of developing decision support systems in the field of education that rely on machine learning.

DAFTAR PUSTAKA

Agarwal, A., Sharma, P., Alshehri, M., Mohamed, A. A., & Alfarraj, O. (2021). Classification model for accuracy and intrusion detection using machine learning approach. Peerj Computer Science, 7, e437. https://doi.org/10.7717/peerj-cs.437

AL-Bakri, N. F., Yonan, J. F., & Sadiq, A. T. (2022). Tourism Companies Assessment via Social Media Using Sentiment Analysis. Baghdad Science Journal, 19(2), 0422. https://doi.org/10.21123/bsj.2022.19.2.0422

Al Fayed, A. J., Darma, S., Sinabariba, Z., & Pardede, S. M. P. (2025). Comparison of Naïve Bayes, K-Nearest Neighbors, and Decision Tree methods for classifying heart disease risk factors. *Journal of Computer Science and Research (JoCoSiR)*, *3*(3), 81–88.

Al Fayed, A. J., Darma, S., Aqsha, M. H., Pardede, S. M. P., & Amin, M. (2026). Perbandingan kinerja algoritma machine learning dalam memprediksi tingkat stres mahasiswa berdasarkan faktor akademik dan non-akademik. *Journal of Science and Social Research*, *9*(1), 483–490. https://doi.org/10.54314/jssr.v9i1.5805.

Ayeni, O. O., Unachukwu, C. C., Hamad, N. M. A., Osawaru, B., & Adewusi, O. E. (2024). A multidisciplinary approach to STEM education: Combining HR, counseling, and mentorship. Magna Scientia Advanced Research and Reviews, 10(1), 351–360.

https://doi.org/10.30574/msarr.2024.10.1.0026

Bruen, M. (2021). Uptake and Dissemination of Multi-Criteria Decision Support Methods in Civil Engineering—Lessons from the Literature. Applied Sciences, 11(7), 2940. https://doi.org/10.3390/app11072940

Elrahman, A. A. (2024). Predicting Adults' Income using Naive Bayes Classifier. https://doi.org/10.21203/rs.3.rs-3890646/v1

Fokoué, E. (2018). To Bayes or Not To Bayes? That's no longer the question! https://doi.org/10.48550/arxiv.1805.11012

https://doi.org/10.30864/eksplora.v13i1.994

Guo, Y., Graber, A., McBurney, R., & Balasubramanian, R. (2010). Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. BMC Bioinformatics, 11(1). https://doi.org/10.1186/1471-2105-11-447

Gyll, S. P. (2021). Career development by design, not default: Creating a more efficient and data-driven process by connecting aptitude-based learner guidance to post-secondary pathways, competency-based credentials, and high-demand jobs. The Journal of Competency-Based Education, 6(1). https://doi.org/10.1002/cbe2.1236

Ha, K. M., Naseem, U., Keya, A. J., Maitra, S., Mithu, K., & Alam, Md. G. R. (2023). A Systematic Review on Airlines Industries based on Sentiment Analysis and Topic Modeling. https://doi.org/10.21203/rs.3.rs-3475984/v1

Jen, L. (2021). A Brief Overview of the Accuracy of Classification Algorithms for Data Prediction in Machine Learning Applications. Journal of Applied Data Sciences, 2(3), 84–92. https://doi.org/10.47738/jads.v2i3.38

Jin, H.-Y., & Cutumisu, M. (2023). Predicting pre-service teachers' computational thinking skills using machine learning classifiers. Education and Information Technologies, 28(9), 11447–11467. https://doi.org/10.1007/s10639-023-11642-7

Khotimah, B. K. (2022). Performance of the K-Nearest Neighbors method on identification of maize plant nutrients. Jurnal Infotel, 14(1), 8–14. https://doi.org/10.20895/infotel.v14i1.735

Kurban, H., Kurban, M., Sharma, P., & Dalkılıç, M. (2021). Predicting Atom Types of Anatase TiO₂ Nanoparticles with Machine Learning. Key Engineering Materials, 880, 89–94. https://doi.org/10.4028/www.scientific.net/kem.880.89

Kuyo, M., Mwalili, S., & Okango, E. (2021). Machine Learning Approaches for Classifying the Distribution of Covid-19 Sentiments. Open Journal of Statistics, 11(05), 620–632. https://doi.org/10.4236/ojs.2021.115037

Mavhemwa, P. M., Zennaro, M., Nsengiyumva, P., & Nzanywayingoma, F. (2024). Weighted naïve bayes multi-user classification for adaptive authentication. Journal of Physics Communications, 8(10), 105005. https://doi.org/10.1088/2399-6528/ad8a16

Meens, E., Bakx, A., Mulder, J., & Denissen, J. J. A. (2020). The Development and Validation of an Interest and Skill Inventory on Educational Choices (ISEC). European Journal of Psychological Assessment, 36(5), 817–828. https://doi.org/10.1027/1015-5759/a000546

Meyer, M. S., Cranmore, J., Rinn, A. N., & Hodges, J. (2020). College Choice: Considerations for Academically Advanced High School Seniors. Gifted Child Quarterly, 65(1), 52–74. https://doi.org/10.1177/0016986220957258

Missaghian, R. (2021). Social Capital and Post-Secondary Decision-Making Alignment for Low-Income Students. Social Sciences, 10(3), 83. https://doi.org/10.3390/socsci10030083

Pereira, N., Bright, S., Ozen, Z., Safitri, S., Castillo-Hermosilla, H., Matos, B. T. P. de, Karatas, T., & Fonseca, P. (2025). A Multitiered Approach to Computer Science Talent Development. Gifted Child Quarterly, 69(2), 130–146. https://doi.org/10.1177/00169862241307662

Patulin, E. P. (2019). Predictability of School Administrator's Job Satisfaction through Hybrid Segmentation-based Prediction Model. International Journal of Advanced Trends in Computer Science and Engineering, 8(3), 507–512. https://doi.org/10.30534/ijatcse/2019/26832019

Reynolds, J., Elliott, J. L., Castillo, K., Sliwak, R. M., & Halligan, C. S. (2023). I lost my mentor, now what? The experiences of counseling psychology women doctoral students who lost their mentor: Training and program implications. Qualitative Psychology, 10(2), 227–244. https://doi.org/10.1037/qup0000237

Sari, A. A., Pramesty, A. S., Malik, A. Z., Al-Hidayah, D. N. H., Alviani, L. S., & Assyifa, R. A. (2023). The Influence of Transformational Education Prediction on Softskills of Madrasa Student using Data Mining. Khazanah Journal of Religion and Technology, 1(1), 20–25. https://doi.org/10.15575/kjrt.v1i1.157

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., Derous, E., Fruyt, F. D., & Duyck, W. (2021). How interest fit relates to STEM study choice: Female students fit their choices better. Journal of Vocational Behavior, 129, 103614. https://doi.org/10.1016/j.jvb.2021.103614

Tapidingan, Y. C., & Paseru, D. (2020). Comparative Analysis of Classification Methods of KNN and Naïve Bayes to Determine Stress Level of Junior High School Students. Indonesian Journal of Information Systems, 2(2), 80–89. https://doi.org/10.24002/ijis.v2i2.303

Taslim, T., Handayani, S., & Fajrizal, F. (2023). Kinerja Komparatif Optimasi Algoritma Naive Bayes dalam Klasifikasi Teks untuk Uji Klinis Kanker. Eksplora Informatika, 13(1), 113–123.

Zhao, Z., Tang, X., & Hu, M. (2025). Establishment of a value assessment framework for orphan drugs in China: an application of the discrete choice experiment in multicriteria decision analysis. Frontiers in Pharmacology, 16. https://doi.org/10.3389/fphar.2025.1677627