JITE (Journal of Informatics and Telecommunication Engineering)

A Comparative Study of Student Satisfaction Levels on Online Learning Using K-NN and Naïve Bayes

Hilda Mutiara Nasution1)*, Mustakim2), Inggih Permana3), M. Afdal4)

1,2,3,4) Department of Information System, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

*Coresponding Email: 11950324870@students.uin-suska.ac.id

Abstrak

Merebaknya pandemi Covid-19 di Indonesia menyebabkan pembatasan aktivitas sosial manusia untuk meminimalisir penularan. Kegiatan belajar-mengajar juga terdampak ketika siswa harus tinggal di rumah dan mengikuti pembelajaran jarak jauh berdasarkan Peraturan Pemerintah Nomor 21 Tahun 2020, kebijakan Pembatasan Sosial Berskala Besar (PSBB), yang dikeluarkan pada tanggal 31 Maret 2020. Hal ini mendorong munculnya aplikasi-aplikasi pendukung pembelajaran seperti Zoom, Google Classroom, Google Meet, E-Learning, dan masih banyak lagi. Namun, budaya belajar yang baru ini membutuhkan adaptasi agar dapat diimplementasikan secara efektif. Dalam proses adaptasi tersebut, peneliti ingin mengukur tingkat kepuasan mahasiswa dan mencari tahu algoritma terbaik untuk mengklasifikasikan tingkat kepuasan mahasiswa. Pengukuran ini menggunakan dua algoritma data mining yaitu K-Nearest Neighbor (K-NN) dan Naïve Bayes, dengan mahasiswa Universitas Islam Negeri Sultan Syarif Kasim Riau sebagai objek penelitian. Algoritme yang berbeda memiliki kekuatan dan kelemahan yang berbeda-beda dalam menangani jenis data dan tugas klasifikasi tertentu. Dengan membandingkan kedua algoritma tersebut, kita dapat menilai kemampuan generalisasinya. Sebuah model yang berkinerja baik pada data pelatihan tetapi gagal menggeneralisasi ke data yang tidak terlihat mungkin tidak seefektif algoritme yang lebih kuat yang menunjukkan kinerja generalisasi yang lebih baik. Klasifikasi K-NN dengan nilai k = 3 mendapatkan hasil yang baik. Berdasarkan hasil penelitian diperoleh kesimpulan bahwa K-NN lebih optimal dalam mengklasifikasikan tingkat kepuasan mahasiswa dibandingkan dengan Naïve Bayes dengan perbandingan akurasi sebesar 85% : 80%, precision sebesar 85% : 84%, dan recall sebesar 99% : 93%.

Kata Kunci: Klasifikasi, Perbandingan, K-NN, Naïve Bayes, Tingkat Kepuasan

Abstract

The outbreak of the Covid-19 pandemic in Indonesia led to restrictions on human social activities to minimize transmission. Teaching-learning is also affected when students must stay home and follow distance learning based on Government Regulation Number 21 of 2020, the Large-Scale Social Restrictions (PSBB) policy, issued on March 31, 2020. This has led to the emergence of learning support applications such as Zoom, Google Classroom, Google Meet, E-Learning, and many more. However, this new learning culture requires adaptation for effective implementation. During the adaptation process, researchers want to measure the level of student satisfaction and find out the best algorithm for classifying the level of student satisfaction. This measurement uses two data mining algorithms, K-Nearest Neighbour (K-NN) and Naïve Bayes, and the Islamic State University of Sultan Syarif Kasim Riau students as the research object. Different algorithms have varying strengths and weaknesses in handling specific data types and classification tasks. By comparing both algorithms, we can assess their generalization capabilities. A model that performs well on training data but fails to generalize to unseen data may not be as effective as a more robust algorithm that exhibits better generalization performance. K-NN classification with a value of k = 3 gets good results. Based on the study results, the conclusion is that K-NN is more optimal in classifying student satisfaction levels than Naïve Bayes with an accuracy ratio of 85% : 80%, precision of 85% : 84%, and recall of 99% : 93%.

Keywords: Classification, Comparison, K-NN, Naïve Bayes, Satisfaction Level

I.   INTRODUCTION

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) outbreak, which is the cause behind the Covid-19 phenomenon, began in Wuhan, Hubei Province, China, on January 26, 2020 (Wu et al., 2020). As of August 17, 2022, in Indonesia, 6.297.484 people have been infected, and 157,296 people have died (World Health Organization Indonesia, 2022). Government Regulation Number 21 of 2020, the Large-Scale Social Restrictions (PSBB) policy, was issued on March 31, 2020. The Covid-19 pandemic has brought unprecedented disruption to the world of education, with school closures impacting 1.2 billion learners and adolescents worldwide (UNESCO, 2020).

Students are confined to their homes due to the application of an online learning system that aims to minimize the transmission of the virus. This new educational culture led to the emergence of alternative applications supporting distance learning, such as Zoom, Google Classroom, e-Learning, Microsoft Teams, Google Meet, and many more. E-learning has become a significant force in education and has been implemented on a large scale in higher education (Al-fraihat et al., 2020). Unfortunately, many educational institutions, educators, and students need more time to prepare for this new experience (Maqableh & Alia, 2021).

Delivering knowledge during the pandemic has become a new challenge, and many lecturers are still designing the best approaches and solutions to overcome this crisis. Similar to traditional learning in general, e-learning also provides a cooperative spirit of collaboration via the online platform and a sense of "togetherness" (Prasetyo et al., 2021). Distance learning at the Islamic State University of Sultan Syarif Kasim Riau lasted for ± 5 academic semesters. During that time, almost 26,459 students accessed supporting applications. In e-learning, lecturers can give instructions to students, upload learning materials, and create a place to collect assignments and deadlines.

By measuring student satisfaction, educational institutions can gain valuable insights, which enables them to identify areas needing improvement and make necessary adjustments to enhance the learning experience. By understanding student difficulty, instructors can adapt their teaching strategies to better align with students' preferences and needs. Ensuring student satisfaction can enhance an institution's brand image, help institutions meet accreditation requirements, and showcase their commitment to providing high-quality online education.

However, on the other hand, according to (Hasan & Bao, 2020), (González-Betancor et al., 2021), and (van de Werfhorst et al., 2022), during the pandemic, there are specific gaps in digital learning. The digital imbalance and lack of access to new technology are perceived to make it difficult for students from low-income communities. Not all students live in areas that have internet coverage. Technical issues such as poor internet connectivity or platform glitches may disrupt the learning process. Online learning can cause feelings of isolation as students miss out on social interactions found in traditional classrooms. The lack of face-to-face contact with peers and instructors may impact student satisfaction. Unlike face-to-face learning, online learning usually relies on digital communication tools. While video conferencing and chat platforms can provide some level of interaction, they cannot completely replace the benefits of face-to-face communication. The reduced ability to ask questions in real-time or engage in spontaneous discussions can impact student satisfaction.

Online learning requires students to manage their time effectively. The absence of a fixed schedule can be challenging for students who struggle with discipline. These challenges can negatively impact student satisfaction. Although the Covid-19 pandemic in some countries has improved, including Indonesia, evaluating online learning is very important. Online learning is an alternative to learning during the pandemic. Still, it must evaluate to improve its effectiveness as a more flexible learning option. In the face of the possible permanence of online learning implementation, there needs to be clear regulations and standards to ensure good quality of online learning. Data mining is required to classify student satisfaction levels with online learning. Data mining is the extraction of procedural modalities and other helpful information from sizable data sets (Mostafa & Mahmoud, 2022).

Classification techniques in data mining are suitable for preparing much information and are used to organize recently accessed information. Naïve Bayes has been used by (Natuzzuhriyyah et al., 2021) to classify student satisfaction levels using RapidMiner with 76.92% accuracy, 100% precision, 57.14% of recall and 0.881 or close to 1 of AUC, so the resulting model is good. The same research conducted by (Yanti & Kriestanto, 2022) shows satisfied respondents of the testing data classification from respondent number

R89 to R93 towards the online learning system during the Covid-19 pandemic and (Samuel & Dewi, 2019) the attributes used are content quality, relevanance, privacy, ease to operate, speed, visual appeal, online completeness, and customer service; shows the classification of determining user satisfaction using the Naïve Bayes method gets the greatest accuracy value with training data testing. Other student satisfaction level research has been conducted (Faisal & Nurhayati, 2020) using K-Nearest Neighbor to get an accuracy of 98%, a recall of 86.67%, a precision of 100%, and an AUC of 0.75. The same research was conducted by (Diansyah, 2022) based on the test results has an accuracy rate of 94.12% with k = 5 as the optimal k value.

The difficulty in this research is collecting the data because measurement methods relying on self-reporting, such as questionnaires, may be subject to response biases and inaccurate reporting. Student's responses may be influenced by factors such as social desirability bias or mood at the time of response. Students may have varying criteria for assessing satisfaction, making establishing a standardized measurement approach difficult. Therefore, this research uses K-Nearest Neighbour which is hereafter abbreviated as K-NN and Naive Bayes to classify student satisfaction levels because of the simplicity yet efficiency and the ability to handle mixed data. K-NN and Naïve Bayes are well-established and widely used classification algorithms, and they have been successfully applied to various domains and have a strong presence in the literature. K-NN is a non-parametric instance-based algorithm that classifies data based on the similarity of its neighboring instances. Naïve Bayes, on the other hand, is a probabilistic algorithm that applies Bayes' theorem and assumes independence between features.

KNN effectively captures local patterns and can adapt well to varying densities in the feature space, which is valuable for identifying similarities between student's satisfaction levels. It can handle numerical and categorical features, making it applicable to many student satisfaction factors. Naive Bayes is well-suited when the dependencies between features have a minimal impact on the classification performance. It can handle large datasets with high-dimensional feature spaces, making it scalable for analyzing various factors contributing to student satisfaction.

Additionally, by comparing the performance of both algorithms, the research can provide insights into their relative strengths and weaknesses for classifying student satisfaction levels in an online learning environment. This research is distinct from previous studies is that this research compares two classification algorithms, Naïve Bayes and K-NN. The dataset that is the basis for grouping is communication, student assessment, learning atmosphere, and material delivery. Another reason this research differs from previous research because is that it focuses on examining the effectiveness of these algorithms in classifying data related to communication, student assessment, learning atmosphere, and material delivery. By comparing the performance of both algorithms, we can gain insights into their respective strengths and limitations in handling this specific dataset.

Therefore, this research aims to find the best algorithm for classifying satisfaction levels and can be used as advice or consideration by related parties based on the analysis results that have been obtained. The analysis results obtained from the research can serve as a reference and consideration for decision-making processes, allowing stakeholders to make informed choices regarding the classification of student satisfaction levels. This information can guide the development and improvement of online learning programs, leading to enhanced student experiences and improved educational outcomes.

II.   RESEARCH METHOD

A.   *Slovin Technique*

The Slovin technique is used to determine the minimum sample size of a population, provided that the population is relatively large. The first thing that needs to be done is to set the confidence level or margin of error (%) of the facts or the significance level of error tolerance (0,..) that will occur. The margin of error is an indicator of the accuracy of an estimate. Typically, the "margin of error" that is often used is 5% or a confidence level of 0.95 (Mohr et al., 2022). The formula for determining the sample size according to Slovin, as in (1) below (W.-C. Yang et al., 2020) :

$$n = \frac{N}{1+N\,(e)^2}.$$ (1)

B. Simple Random Sampling

The sample data is chosen randomly and purely by chance. Hence the quality of the sample is not affected since every member has an equal chance of being selected as the sample. This type of sampling is most suitable for highly homogeneous populations (Bhardwaj, 2019). Simple random sampling requires carefully defining the population from which the sample is drawn (Golzar, 2022).

C. Hold-Out

The hold-out method randomly divides data into two separate sets: training and test sets. The data is divided multiple times; for each division, the host selects one predictor; then, the predictors gained by different divisions are combined (Maillard et al., 2021). Usually, about two-thirds of the data is distributed to the training set, and the rest one-third is distributed to the testing set. The training set is used to obtain a model. The model's accuracy is evaluated with the testing set. Data sharing can do with percentages such as 90:10, 50:50, 80:20, 70:30, and 75:25. (Awwalu et al., 2019).

D. Data Collection Technique

This research uses five stages, as shown in Fig.1 below; the first stage is planning. At this stage, identification of problems occurring, especially the impact of Covid-19 on the learning process, in the form of an evaluation. Furthermore, determining the objectives that serve to clarify the framework of what is the target of this research. At this stage, researchers also look for sources or references relevant to the research methods, from scientific articles, books, proceedings, and so on, which will be used as research references. Then determine the limitations of the problem, which aims to make the scope of a problem or discussion to be carried out focused and stay consistent with the research.

The second is data collection. The data source in this study was collected through an online questionnaire. A questionnaire is one of the tools commonly used to collect information such as a form containing a series of questions filled in by respondents to provide the information researchers need for research (Taherdoost, 2021). The questionnaire was distributed to respondents via online media, namely Whatsapp because the author could distribute questionnaires on target to students who had participated in online learning. The data collection for this study was conducted over 2 months, specifically from October 8 to December 8, 2022. The target population was students from the Islamic State University of Sultan Syarif Kasim Riau who had experienced online learning. A closed online questionnaire comprising 13 questions was used to ensure a representative sample (Yanti & Kriestanto, 2022).
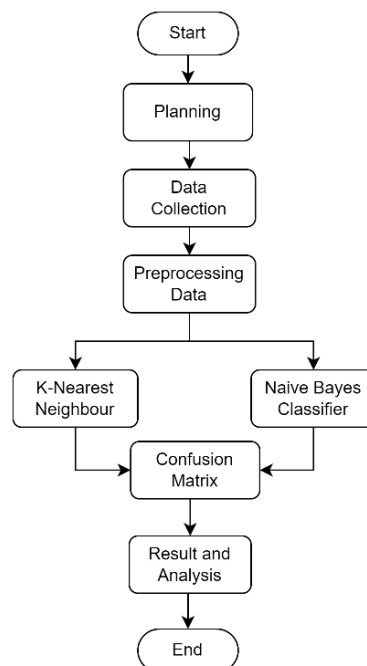

Figure 1 Research Method

The Slovin technique with 10% margin of error (Ali, 2019; Anderjovi et al., 2022);(Bimaruci et al., 2020) and simple random sampling was used to determine the appropriate sample size. According to the calculations, the minimum sample size required for this research was 100 data points. However, to improve the accuracy of the results, enhance sample representation, and mitigate potential unrepresentativeness, the researchers opted to utilize a larger sample size of 140 data points. As explained by (Spinde et al., 2021) and (S. T. Noor et al., 2021)in their paper, a larger dataset is needed to improve research results. By increasing the sample size beyond the minimum requirement, the study aimed to enhance the precision and reliability of the findings. This larger sample allows for more robust analysis and strengthens the generalizability of the results to the broader population of Islamic State University students who have experienced online learning.It is worth noting that by utilizing the Slovin technique and employing simple random sampling, the researchers ensured that each participant had an equal chance of being selected, minimizing potential bias and increasing the validity of the study's conclusions.

Once the data is collected, the next stage is pre-processing data following the stages in data mining. First, data cleaning is done to avoid incomplete data and prevent data duplication and then combines clean data into a data set. After this stage, the data is processed using K-NN and Naïve Bayes then the results of both algorithms are tested using Confusion Matrix. The last stage is results and analysis, here drawing conclusions and describing the results of observations and knowing how the classification results.

The questionnaire's research attributes were measured using a Likert scale. The surveyed participants were instructed through an instruction to indicate their level of agreement (from strongly disagree to strongly agree) with each question (item) on a metric scale. (Hassan, 2019). The research attributes are shown in Table 1, and the Likert scale used to measure the research attributes is shown in Table 2 below:

Table 1 Attribute and Value

No	Attribute	Value
1	Material Delivery (**A**)	1,2,3,4,5
2	Communication (**B**)	1,2,3,4,5
3	Student Assessment (**C**)	1,2,3,4,5
4	Learning Atmosphere (**D**)	1,2,3,4,5

Table 2 Rating Scale

Rating Scale	Value
Very Good	5
Good	4
Quite Good	3
Less Good	2
Not Good	1

After preprocessing, the data obtained is ready for classification, as shown in Table 3 below, based on the research dataset below, the second data is that students from Science and Technology faculty gave a score of 3 (quite good) to the attribute of material delivery, a score of 4 (good) for communication, a score of 2 (less good) or less for student assessment, and a score of 2 (less good) or less for learning atmosphere.

Table 3 Research Dataset

No	Faculty	A	B	C	D
R1	Agriculture and Animal Husbandry	2	4	2	3
R2	Science and Technology	3	4	2	2
R3	Science and Technology	3	4	3	3
R4	Science and Technology	3	3	3	4
R5	Education and Teacher Training	3	4	4	3
R6	Usul al-Din	3	3	3	3
R7	Science and Technology	3	4	3	3
R8	Sharia and Law	4	4	4	4
R9	Science and Technology	3	4	3	4

No	Faculty	A	B	C	D
R10	Economics and Social Sciences	2	2	2	2
...	...	...	...	...	...
R139	Psychology	1	1	1	1
R140	Da'wa and Communication	3	3	2	3

III. RESULTS AND DISCUSSION

A. *K-Nearest Neighbor (K-NN)*

K-Nearest Neighbor is a query-triggered but improvised learning procedure. The algorithm is only performed when the test data is predicted, setting an appropriate k value and searching for the K nearest neighbors (Zhang & Li, 2021). Besides being easy to understand, the algorithm is also versatile, covering a wide range of applications. Apart from its simplicity, as a simple classifier that does not generate trained models but keeps or remembers training examples in exchange (Karam et al., 2022). To calculate the squared Euclidean distance using (2):

$$d\,(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{(q_{1-}p_1)^2 + (q_{2-}p_2)^2 + \cdots + (q_{n-}p_n)^2}. \tag{2}$$

One hundred forty data were processed using K-NN, which had already passed the preprocessing stage. In this method, the K value aims to determine the number of nearest neighbors of the training set. The dataset is divided into 70% for the training set and 30% for the testing set (Vrigazova, 2021). Data processing is done using Python with various K-value experiments. The classification results of various K values are then tested with a confusion matrix, as shown in Fig. 2, Fig. 3, Fig. 4, and Fig. 5 below:
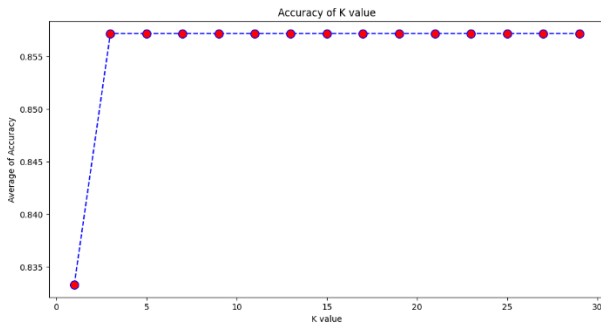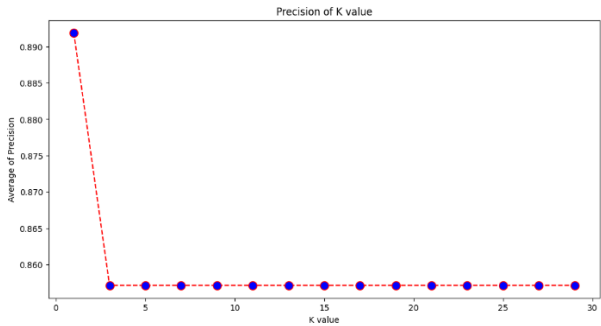

Figure 2 K-NN Confusion Matrix Accuracy

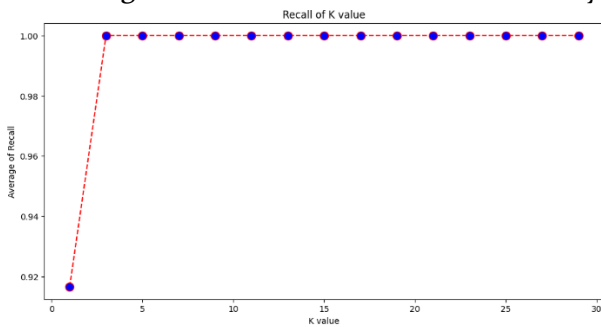**Figure 3** K-NN Confusion Matrix Precision
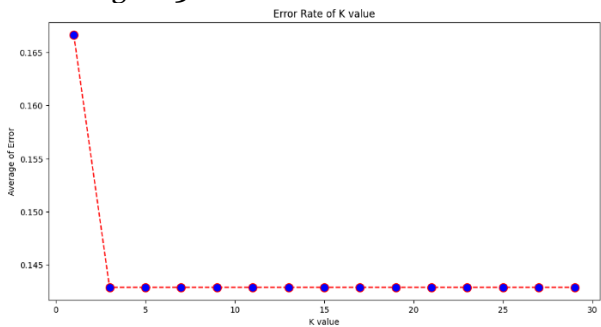

Figure 4 K-NN Confusion Matrix Recall

Figure 5 K-NN Confusion Matrix Error

As can be seen from the above four figures, starting from the value of K=3 and above, the graph begins to form a straight line on the diagram, even when the value of k is increased, indicating that the accuracy of the model no longer increases with a more significant number of nearest neighbors. This indicates that up to any value of K, the results will always be the same and form a straight line. In this case, K=3, which produces a straight line, is the optimal k value for the K-NN model.

B. Naïve Bayes

Naïve Bayes is a probabilistic classifier that relies on the Bayes theorem and assumes feature independence for a particular class (Uddin & Ahmed, 2020). Naïve Bayes called the conditional independence assumption, assumes that all attributes are independent given the output class (Sari et al., 2019). To solve the problem using Naïve Bayes can be done with the following (3) (Patel & Parikh, 2020):

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}. \qquad (3)$$

Another classification algorithm used in this research is Naïve Bayes. It uses the same data as in the classification process with K-NN. Naïve Bayes has widely been adopted to predict results under uncertainty (F. J. Yang, 2018).

C. Confusion Matrix

The confusion matrix is an easy and powerful tool to demonstrate the performance of a classifier and is easy to understand the results. The performance of any model or algorithm can be evaluated with the confusion matrix (Yun, 2021). The performance of a classification model can be measured by its accuracy (Gong, 2021). KNN effectively captures local patterns and can adapt well to varying densities in the feature space, which is valuable for identifying similarities between students' satisfaction levels. Naive Bayes is well-suited when the dependencies between features have a minimal impact on the classification performance. The K-NN confusion matrix shows that out of the total data points, 36 are classified as true positives, 0 as false negatives, 6 as false positives, and 0 as true negatives. The accuracy is calculated as 85%. The precision, which measures the proportion of correctly classified positive examples among the total predicted positive example, is also 85%. The recall, also known as sensitivity or true positive rate, measures the proportion of correctly classified positive examples among the true positive example, which is 99%.

Then, the Naïve Bayes confusion matrix shows that 34 data points are classified as true positives, 2 as false negatives, 6 as false positives, and 0 as true negatives. The accuracy is calculated as 80%. The precision, which indicates the accuracy of positive predictions, is 84%. The recall, which measures the model's ability to find all the positive instances, is 93%. The results of both classifications are then tested using the confusion matrix as shown in Fig. 6, and Fig. 7 below:
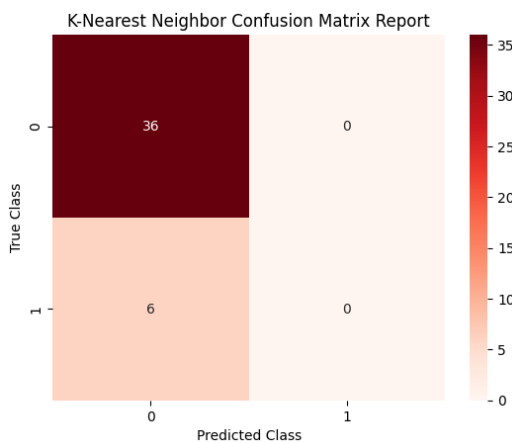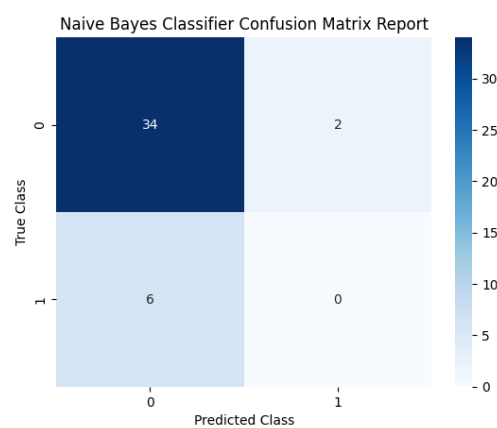

Figure 6 Confusion Matrix K-NN

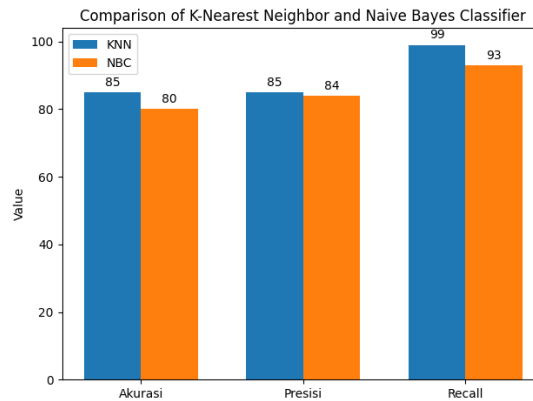**Figure 7** Confusion Matrix Naïve Bayes

Figure 8 Comparison of both algorithms

Based on its higher accuracy and recall values, the results are that the K-NN algorithm may be more effective in classifying student satisfaction levels with online learning than the Naïve Bayes algorithm. However, the significance of these differences and the selection of the best model ultimately depend on the specific research question, the context of the study, and other relevant considerations.

D.   Discussion

While higher accuracy and recall values indicate that K-NN may perform better in this particular scenario, other factors should be considered. The assumption of independence made by the Naïve Bayes algorithm might not hold in all cases. If the features used for classification are not truly independent, it can lead to suboptimal performance. As a non-parametric algorithm, K-NN makes no strong assumptions about the underlying data distribution. Additionally, the interpretability of the model can be an essential consideration. Naïve Bayes provides precise probabilities and allows for interpretability, making understanding the factors contributing to the classification decisions easier. K-NN, on the other hand, does not provide such straightforward interpretability.

In conclusion, while the statement suggests that K-NN may be more effective in classifying student satisfaction levels with online learning based on higher accuracy and recall values, selecting the best model requires considering the specific research question, the context of the study, computational efficiency, assumptions made by the algorithms, dimensionality of the dataset, and interpretability requirements.

IV.   CONCLUSION

Based on the objectives of this research through the data analysis and processing results, the best classification algorithm to classify student satisfaction with online learning is K-NN. It shows that K-NN can produce more accurate results. Confusion Matrix is used to validate the classification results. It shows that the comparison of the accuracy of K-NN and Naïve Bayes is 86%: 80%, precision comparison 85%: 91%, and 100% recall comparison: 86%. Although Naïve Bayes has a higher precision value than K-NN, K-NN is superior in accuracy and recall. The perfect recall of K-NN indicates that the model can correctly classify all students who should belong to a particular category (e.g delighted students). It shows the reliability of the K-NN in recognizing positive cases.

For further exploration research there are numerous other classification algorithms that could be considered. Algorithms such as Decision Trees, Random Forests, Support Vector Machines, or Neural Networks may offer alternative approaches to classifying student satisfaction. Implement different cross-validation techniques, such as k-fold cross-validation or stratified cross-validation, to assess the generalization performance of the classification model. And supplement the quantitative analysis with qualitative research methods, such as interviews or surveys, to gain a deeper understanding of the reasons behind student satisfaction or dissatisfaction in online learning.

REFERENCES

Al-fraihat, D., Joy, M., & Sinclair, J. (2020). Evaluating E-learning systems success : An empirical study. Computers in Human Behavior, 102(March 2019), 67–86. https://doi.org/10.1016/j.chb.2019.08.004

Ali, H. (2019). Purchase Decision and Repurchase Models: Product Quality and Process Analysis (Case Study of House Ownership Credit Financing in Permata Sharia Bank Jakarta). Scholars Bulletin, 05(09), 526–535. https://doi.org/10.36348/sb.2019.v05i09.006

Anderjovi, S., Hamzah, M. L., Maita, I., & Khairil Ahsyar, T. (2022). User Satisfaction Analysis of E-Learning Using End User Computing Satisfaction in Covid 19.

Awwalu, J., Ogwueleka, F., & Nonyelum, O. F. (2019). On Holdout and Cross Validation: A Comparison between Neural Network and Support Vector Machine. International Journal of Trend in Research and Development, 6(2), 2394–9333. www.ijtrd.com

Bhardwaj, P. (2019). Types of sampling in research. Journal of the Practice of Cardiovascular Sciences, 5(3), 157. https://doi.org/10.4103/jpcs.jpcs_62_19

Bimaruci, H., Havidz, H., Hudaya, A., & Ali, H. (2020). MODEL OF CONSUMER TRUST ON TRAVEL AGENT ONLINE: ANALYSIS OF PERCEIVED USEFULNESS AND SECURITY ON RE-PURCHASE INTERESTS (CASE STUDY TIKET.COM). 1(1). https://doi.org/10.38035/DIJEFA

Diansyah, S. (2022). Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN). Jurnal Sistim Informasi Dan Teknologi, 4(1), 7–12. https://doi.org/10.37034/jsisfotek.v4i1.114

Faisal, S., & Nurhayati. (2020). Implementation of K-Nearest Neighbor Algorithm for Customer Satisfaction. Buana Information Technology and Computer Sciences (BIT and CS), 1(2), 27–32. https://doi.org/10.36805/bit-cs.v1i2.886

Gong, M. (2021). A Novel Performance Measure for Machine Learning Classification. International Journal of Managing Information Technology, 13(1), 11–19. https://doi.org/10.5121/ijmit.2021.13101

González-Betancor, S. M., López-Puig, A. J., & Cardenal, M. E. (2021). Digital inequality at home. The school as compensatory agent. Computers and Education, 168(July 2020). https://doi.org/10.1016/j.compedu.2021.104195

Hasan, N., & Bao, Y. (2020). Impact of "e-Learning crack-up" perception on psychological distress among college students during COVID-19 pandemic: A mediating role of "fear of academic year loss." Children and Youth Services Review, 118(July). https://doi.org/10.1016/j.childyouth.2020.105355

Hassan, S. (2019). Online Survey-based Research Using Likert-Scale Questionnaire as an Educational Tool during COVID-19: From Design and Development to Validation Process and Interpretation. International Journal of Science and Research, 10, 2319–7064. https://doi.org/10.21275/SR21512174451

Karam, A. F., Helmy, A. M., & Mohammed, A. (2022). An approach to enhance KNN based on data clustering using K-medoid. Conference: International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC-2022).At: Egypt (MIUCC-2022)., June, 49–55. https://doi.org/10.1109/MIUCC55081.2022.9781724

Maillard, G., Arlot, S., & Lerasle, M. (2021). Aggregated Hold-Out. In Journal of Machine Learning Research (Vol. 22). http://jmlr.org/papers/v22/19-624.html.

Maqableh, M., & Alia, M. (2021). Evaluation online learning of undergraduate students under lockdown amidst COVID-19 Pandemic: The online learning experience and students' satisfaction. Children and Youth Services Review, 128(July), 106160. https://doi.org/10.1016/j.childyouth.2021.106160

Mohr, D. L., Wilson, W. J., & Freund, R. J. (2022). Principles of Inference. In Statistical Methods (pp. 123–167). Elsevier. https://doi.org/10.1016/b978-0-12-823043-5.00003-5

Mostafa, A. A. N., & Mahmoud, H. E. A. (2022). Review of Data Mining Concept and its Techniques. International Journal of Academic Research in Business and Social Sciences, 12(6). https://doi.org/10.6007/ijarbss/v12-i6/13135

Natuzzuhriyyah, A., Nafisah, N., & Mayasari, R. (2021). Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naïve Bayes. JISKA (Jurnal Informatika Sunan Kalijaga), 6(3), 161–170. https://doi.org/10.36805/technoxplore.v6i2.1377

Noor, S. T., Asad, S. T., Khan, M. M., Gaba, G. S., Al-Amri, J. F., & Masud, M. (2021). Predicting the Risk of Depression Based on ECG Using RNN. Computational Intelligence and Neuroscience, 2021. https://doi.org/10.1155/2021/1299870

Noor, S., Tajik, O., & Golzar, J. (2022). Simple Random Sampling. International Journal of Education & Language Studies, 1(2), 78–82. https://doi.org/10.22034/ijels.2022.162982

Patel, B. A., & Parikh, A. (2020). Impact Analysis of the Complete Blood Count Parameter using Naive Bayes. Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020, 7–12. https://doi.org/10.1109/ICICT48043.2020.9112533

Prasetyo, Y. T., Ong, A. K. S., Concepcion, G. K. F., Navata, F. M. B., Robles, R. A. V., Tomagos, I. J. T., Young, M. N., Diaz, J. F. T., Nadlifatin, R., & Redi, A. A. N. P. (2021). Determining factors affecting acceptance of e-learning platforms during the covid-19 pandemic: Integrating extended technology acceptance model and delone & mclean is success model. Sustainability (Switzerland), 13(15), 1–16. https://doi.org/10.3390/su13158365

Samuel, Y. T., & Dewi, K. (2019). Penggunaan Metode NAÏVE BAYES Dalam Mengukur Tingkat Kepuasan Pengguna Terhadap Online System Universitas Advent Indonesia The Use of Naïve Bayes Method in Measuring User's Satisfaction With Adventist University of Indonesia's Online System. https://www.online.unai.edu.

Sari, E. Y., Wierfi, A. D., & Setyanto, A. (2019). Sentiment Analysis of Customer Satisfaction on Transportation Network Company Using Naive Bayes Classifier. 2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia, CENIM 2019 - Proceeding. https://doi.org/10.1109/CENIM48368.2019.8973262

Spinde, T., Rudnitckaia, L., Hamborg, F., & Gipp, B. (2021). Identification of Biased Terms in News Articles by Comparison of Outlet-Specific Word Embeddings. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12646 LNCS, 215–224. https://doi.org/10.1007/978-3-030-71305-8_17

Taherdoost, H. (2021). Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects. International Journal of Academic Research in Management (IJARM), 2021(1), 10–38. https://hal.science/hal-03741847

Uddin, M. A., & Ahmed, M. S. (2020). Modified naive Bayes classifier for classification of protein-protein interaction sites. Journal of Bioscience and Agriculture Research, 26(02), 2177–2184. https://doi.org/10.18801/jbar.260220.266

UNESCO. (2020). COVID-19 – Education is the bedrock of a just society in the post-COVID world. https://webarchive.unesco.org/web/20220102202635/https://en.unesco.org/news/covid-19-education-bedrock-just-society-post-covid-world

van de Werfhorst, H. G., Kessenich, E., & Geven, S. (2022). The digital divide in online education: Inequality in digital readiness of students and schools. Computers and Education Open, 3(July), 100100. https://doi.org/10.1016/j.caeo.2022.100100

Vrigazova, B. (2021). The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. Business Systems Research, 12(1), 228–242. https://doi.org/10.2478/bsrj-2021-0015

World Health Organization Indonesia. (2022). Coronavirus Disease 2019 (COVID-19). Situation Report - 93., WHO Indonesia.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

Yang, F. J. (2018). An implementation of naive bayes classifier. Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018, 301–306. https://doi.org/10.1109/CSCI46756.2018.00065

Yang, W.-C., Lin, C.-N., & Hu, P.-H. (2020). Willingness and Obstacles of Food and Farming Education in Leisure Farm Management: Viewpoint of Experience Activities. OALib, 07(06), 1–14. https://doi.org/10.4236/oalib.1106485

Yanti, S., & Kriestanto, D. (2022). ANALISIS KEPUASAN MAHASISWA TERHADAP SISTEM PEMBELAJARAN ONLINE PADA MASA PANDEMIC COVID 19 DI STMIK AKAKOM DENGAN METODE NAIVE BAYES. Jurnal Teknologi Informasi Dan Ilmu Koputer, 20(1), 23–31.

Yun, H. (2021). Prediction model of algal blooms using logistic regression and confusion matrix. International Journal of Electrical and Computer Engineering, 11(3), 2407–2413. https://doi.org/10.11591/ijece.v11i3.pp2407-2413

Zhang, S., & Li, J. (2021). KNN Classification with One-step Computation. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2021.3119140