Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 41-49
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Classification of Smoking Dependency Levels Using the K-NN Algorithm

Alfin Kurniawan[1,*], Diva Ukhwa Mawarni Lubis[1], Pray Argalova Sinaga[1], Rahma Mawaddah Purba[1], Irfan Sudahri Damanik[2]

[1] Information System Study Program, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
Jl. Jenderal Sudirman Street, Block A No. 1/2/3, Pematangsiantar 21127, North Sumatra, Indonesia
[2] Master of Informatics Study Program, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
Jl. Jenderal Sudirman Street, Block A No. 1/2/3, Pematangsiantar 21127, North Sumatra, Indonesia
sEmail: [1,*]alfinkurniawan2407@gmail.com, [2]divaukhwamawarnilubis@gmail.com, [3]prayargalovasinaga@gmail.com, [4]rahmamawaddahp@gmail.com, [5]irfansudahri@amiktunasbangsa.ac.id.
Correspondence Author Email: emailpenuliskorespondensi@email.com
Submitted: **01/06/2026**; Accepted: **12/06/2026**; Published: **30/06/2026**

Abstract−Smoking is a major public health issue due to the addictive effects of nicotine, which can lead to varying levels of dependence among individuals. Identifying smoking dependence levels is important to support prevention and intervention efforts. This study aims to classify smoking dependence levels using the K-Nearest Neighbor (K-NN) algorithm. The dataset used in this research was obtained from Kaggle and consists of 103 respondent records with attributes including age, gender, smoking peers, smoking family members, age of smoking initiation, number of cigarettes consumed per day, and daily allowance. The research process involved data preprocessing, including data cleaning, categorical data transformation, and Min-Max normalization, followed by data splitting into 80% training data and 20% testing data. Classification was performed using the Euclidean Distance metric, while model performance was evaluated using accuracy, precision, recall, and F1-score. Experimental results show that the K-NN algorithm with K = 5 achieved an accuracy of 95.24%, precision of 95.71%, recall of 95.24%, and F1-score of 95.10%. Although K values of 1, 2, and 3 produced higher accuracy, K = 5 was selected because it provided better stability and generalization while reducing sensitivity to noise and outliers. The findings demonstrate that the K-NN algorithm is effective for classifying smoking dependence levels and can be utilized as a decision-support tool in smoking behavior analysis and public health initiatives.

Keywords: Smoking Dependence; K-Nearest Neighbor; Classification; Data Mining; Nicotine Addiction

1. INTRODUCTION

Cigarettes are tobacco-based products derived from the *Nicotiana* plant or its synthetic derivatives, containing nicotine and tar, and are used by inhaling the smoke produced by burning them (Anggarawati, 2021; Rachmat et al., 2013; Wahyuni, 2022). Among adolescents, smoking behavior is generally influenced by social circles, curiosity, and peer pressure, making it difficult to avoid this habit. Smoking is known to cause various health disorders, such as respiratory diseases, heart disease, stroke, cancer, and tuberculosis (Marito et al., 2024; Rahmadi & Lestari, 2013). In addition to affecting physical health, the addictive nature of nicotine leads to dependence, making it difficult to quit smoking even when the risks are known (Kampung et al., 2017). Efforts to quit smoking are generally undertaken through gradual reduction of consumption, redirecting the urge to smoke, and avoiding smoking environments, with success rates influenced by individual motivation and social support (Salsabila, 2022). Globally, smoking remains a major health issue, with the number of adult smokers reaching approximately 1.3 billion people (Ilmiah et al., 2020).

In the ASEAN region, Indonesia has the highest prevalence of smokers and is also one of the world's largest tobacco producers, with the majority of adult smokers consuming kretek cigarettes (Sahelvi et al., 2025). According to a 2011 WHO report, Indonesia ranked fifth globally in the number of smokers in 2007. Chemicals in tobacco contribute to the development of hypertension by damaging the lining of blood vessels, thereby increasing the risk of atherosclerotic plaque formation (Arvelia et al., 2022; Kesehatan et al., 2025). These conditions are influenced by nicotine, which accelerates heart rate and constricts blood vessels, as well as carbon monoxide, which reduces oxygen levels in the blood, thereby increasing the heart's workload (Amelia et al., n.d.; Octavian et al., n.d.). The effects of smoking are felt not only by active smokers but also by passive smokers and the surrounding environment, and they even contribute to economic problems (Sawitri et al., 2020). Nicotine dependence also leads to withdrawal symptoms, such as irritability, restlessness, difficulty concentrating, sleep disturbances, increased appetite, and even depression when someone attempts to quit smoking (Rachmat et al., 2013).

Various studies have examined the level of nicotine dependence among users of conventional cigarettes and e-cigarettes, particularly among adolescents and young adults. Although developed as an alternative to conventional cigarettes, e-cigarettes still have the potential to cause nicotine dependence. In a study conducted by Abdullah et al., the levels of nicotine dependence in both groups of users were compared, and the results showed no significant difference. Therefore, e-cigarettes cannot yet be considered safer in terms of the risk of dependence. These findings indicate that the nicotine contained in e-cigarettes still contributes to the development of addictive behavior (Marpaung & Marbun, 2021).

Given this issue, an analytical method is needed that can objectively and systematically classify levels of nicotine dependence. One technique widely used in machine learning is classification, which is the process of grouping data into specific categories based on their characteristics by utilizing labeled data as a reference to accurately determine the class of new data (Neighbor- , et al., 2023; Status, Balita, et al., 2024). One supervised learning algorithm frequently used for this purpose is K-Nearest Neighbor (K-NN), which works by identifying a set of nearest neighbors based on distance calculations, such as Euclidean Distance, and then assigning the class of new data based on the most dominant category

among those neighbors (Maiyanti et al., 2023; Munawaroh et al., 2024; No et al., 2025). As an instance-based learning or lazy learning method, K-NN does not build a model during the training phase but instead stores all training data and performs the classification process when test data is input into the system (Aprilia et al., 2024; Status, Bayi, et al., 2024). Although simple in concept and implementation, this algorithm is capable of producing good classification performance across various problems. However, its accuracy is highly influenced by the quality and characteristics of the features used, making the feature selection and weighting process a critical factor in improving the efficiency and accuracy of classifying labeled data (Bayes et al., 2022; Dan & Genetika, 2023; Putri et al., 2023).

Based on the explanation outlined above, the objective of this study is to classify the level of smoking dependence. This classification process aims to determine the extent of an individual's dependence on cigarettes. The expected outcome of this study is to provide a clear picture of smokers' levels of dependence, thereby serving as a foundation for decision-making and more targeted efforts in preventing and addressing smoking dependence.

2. RESEARCH METHODOLOGY

2.1 Dataset

The dataset for this study was obtained from Kaggle under the name "*Smoker Data*," published by M. Zulkarnaen. This dataset contains information regarding the characteristics and smoking habits of the respondents, which was utilized to classify the level of smoking dependence using the K-Nearest Neighbor (K-NN) algorithm. The variables used include age, gender, status of smoking friends, status of smoking family members, age at which smoking began, daily cigarette consumption, and daily allowance category. All these attributes were selected because they are considered to be related to smoking behavior and an individual's level of dependence on cigarettes.

The processing results showed that the dataset consisted of 103 respondents grouped into three levels of smoking dependence: low, moderate, and high. The diversity of characteristics among each respondent allowed the K-NN algorithm to learn data patterns based on the proximity between objects to determine the appropriate dependence category. Before the classification process was carried out, the data was first processed through a *preprocessing* stage that included data cleaning, conversion of categorical data to numerical form, and normalization of attribute values so that distance calculations in the K-NN algorithm could run more optimally.

Table 1. Sample Dataset of Smoking Dependency Levels

Age	Gender	Has Smoker Friends	Smoking Family Members	Age When Smoking Began	Cigarettes Per Day	Daily Allowance
16	Male	Yes	Yes	14	15	Medium
17	Women	No	No	N/A	0	Low
15	Male	Yes	No	13	5	Low
16	Women	Yes	Yes	15	8	Medium
17	Male	Yes	Yes	12	20	Height
15	Male	No	No	N/A	0	Medium
16	Male	Yes	Yes	16	10	Height
15	Female	No	No	N/A	0	Low
17	Male	Yes	No	15	30	Height
16	Women	Yes	No	14	6	Moderate

A portion of the sample data used in the study is presented in Table 1. Each respondent is represented by a single data row containing various characteristics, including age, gender, the smoking environment around the respondent, age at which smoking began, the number of cigarettes consumed per day, and the daily allowance category. All of these attributes were used as input data to classify the level of smoking dependence using the K-Nearest Neighbor (K-NN) algorithm.

2.2 Research Framework

This study aims to classify the level of smoking dependence based on the respondents' characteristics and smoking habits. The research process was systematically structured, beginning with problem identification and culminating in a classification that serves as a foundation for understanding the level of smoking dependence. Before the K-Nearest Neighbor (K-NN) algorithm was applied, the data was first processed and tested through several stages to obtain an optimal classification model.

The research framework in Figure 1 illustrates the research flow, beginning with problem identification and dataset collection. The obtained data then undergoes preprocessing, including data cleaning and normalization to improve data quality. Next, the dataset is divided into training data and testing data. The classification process is performed using the K-NN algorithm based on the attributes available in the dataset. The model's performance is then evaluated using a confusion matrix, accuracy, precision, recall, and F1-score. The final stage of the research involves analyzing the evaluation results and drawing conclusions based on the findings obtained.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 41-49
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Figure 1. Research Flowchart for Smoking Dependency Level Classification

Based on Figure 1, the research process was carried out systematically, starting from the data collection stage to model evaluation. The interrelationships between the stages were designed to ensure that the K-NN algorithm could effectively classify smoking dependence levels and produce optimal accuracy.

2.3 K-NN Algorithm

The K-Nearest Neighbor (K-NN) method is a technique that utilizes the k nearest neighbors of training data to determine classification results, thereby grouping data based on the proximity of a data point to its neighbors (Kunci, 2023; No et al., 2024; Pringsewu & Tree, n.d.) . The determination of the nearest distance is performed by first separating the data into two categories: training data and test data. Once both sets are obtained, the distance between each test data point and the training data is calculated using Euclidean Distance. Euclidean Distance is the distance between two points or coordinates calculated using the Pythagorean theorem. The Euclidean formula is used to calculate the distance between two points, namely the point in the training data (x) and the point in the test data (y), as shown in the following equation.

$$D_q = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_n - b_n)^2} \quad (1)$$

Explanation:
a_i = the value of the i-th attribute in the new data
b_i = the value of the i-th attribute in the old data
n = number of attributes
D_q = distance between the new data and the old data

This formula is the Euclidean Distance, which is a method for calculating the distance between two data sets: the new data (query) and the old data (dataset). In K-NN, the data with the smallest distance is considered the most similar, so it is used as the nearest neighbor to determine the class.

3. RESULTS AND DISCUSSION

The level of smoking dependence in this study was classified using the K-Nearest Neighbor (K-NN) algorithm based on respondent characteristics found in the smoking behavior dataset. Before the classification process was carried out, the data first underwent a preprocessing stage to improve data quality. Next, the dataset was divided into training and testing data, after which the K-NN algorithm was applied and its performance evaluated using a confusion matrix along with several evaluation metrics. The test results showed that smoking dependence patterns could be well identified by the K-NN algorithm based on the attributes used. Additionally, several values of K were tested to determine the most optimal parameter so that the algorithm's effectiveness in classifying data could be determined more accurately.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 41-49
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

3.1 Preprocessing Data

During the *preprocessing* stage, *data cleaning* was performed to ensure the quality of the dataset used in the study remained high. This process focused on identifying and handling *missing values* present in several attributes. As shown in Figure 1, a number of *missing values* were still present before data cleaning; however, after the *data cleaning* process was completed, all missing data were successfully addressed, making the dataset complete and ready for use in the classification stage. This stage is crucial in the application of the K-Nearest Neighbor (K-NN) algorithm because the calculation of proximity between data points is performed using the Euclidean distance. The presence of missing data can affect the distance calculation results and degrade model performance. Therefore, the data cleaning process is conducted to improve the quality of the dataset, ensuring that the classification results obtained are more accurate and optimal.

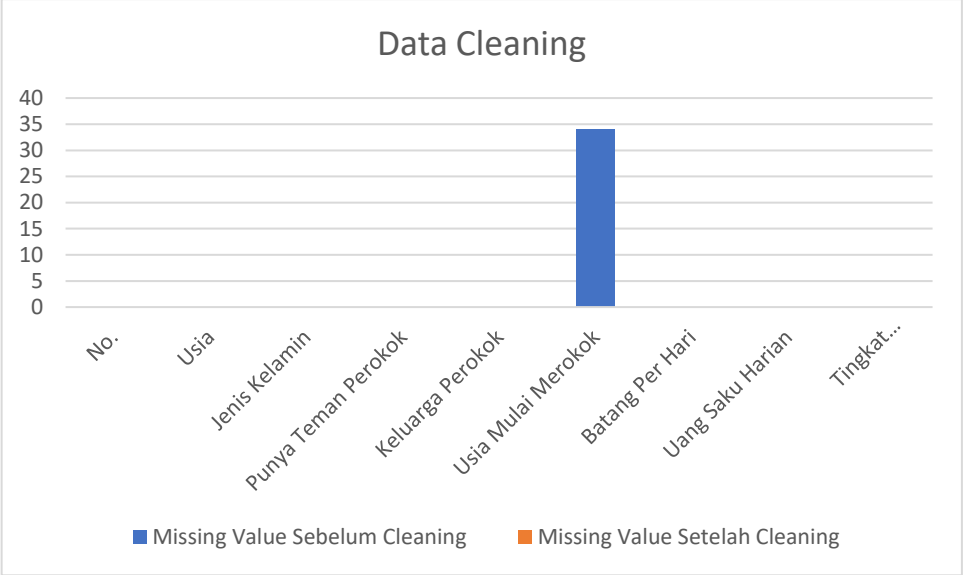

Figure 1. Data Cleaning Process and Handling of Missing ValuesAfter the data cleaning process is completed, the dataset is still stored in its original form with a diverse range of values for each attribute. Some of the data used in the study prior to the normalization stage is shown in Table 1.

Table 1. Before Normalization

Age	Gender	Has Friends Who Smoke	Family Members Who Smoke	Age When Smoking Began	Daily Allowance	Daily Allowance
16	Male	Yes	Yes	14	15	Medium
17	Women	No	No	N/A	0	Low
15	Male	Yes	No	13	5	Low
16	Women	Yes	Yes	15	8	Medium
17	Male	Yes	Yes	12	20	Height

Based on Table 1, numerical attributes such as age, age at which smoking began, and the number of cigarettes consumed per day have different ranges of values. Additionally, some attributes are still categorical data, such as gender, smoking status of friends, and smoking status of family members. These differences in data scale and form can affect the distance calculation process in the K-NN algorithm, so that attributes with larger values have the potential to exert a more dominant influence than other attributes.

Differences in the value ranges between attributes are addressed through a normalization process using the Min-Max Scaling method. With this method, all attribute values are transformed into the range 0–1 without altering the existing data distribution pattern.

Table 2. After Normalization

Age	Gender	Has Friends Who Smoke	Family Members Who Smoke	Age When Smoking Began	Cigarettes Per Day	Daily Allowance
0.5	0	1	1	0.875	0.5	0.5
1	1	0	0	0	0	0
0	0	1	0	0.8125	0.166666666666667	0
0.5	1	1	1	0.9375	0.266666666666667	0.5
1	0	1	1	0.75	0.666666666666667	1

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 41-49
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

In Table 2, the normalization results show that all numerical attributes now have a uniform scale, while categorical attributes have been converted into numerical form. This uniformity of scale allows each attribute to contribute equally to the Euclidean distance calculation, thereby making the K-Nearest Neighbor (K-NN) algorithm's classification performance more optimal and accurate.

3.2 Data Split (Training and Testing)

Before the K-Nearest Neighbor (K-NN) algorithm is used, the preprocessed data is first divided into training data and testing data. This division aims to allow the model to learn patterns from the available data and test its ability to classify new data that has never been used before. In this study, the *hold-out* method was applied with a composition of 80% training data and 20% testing data.

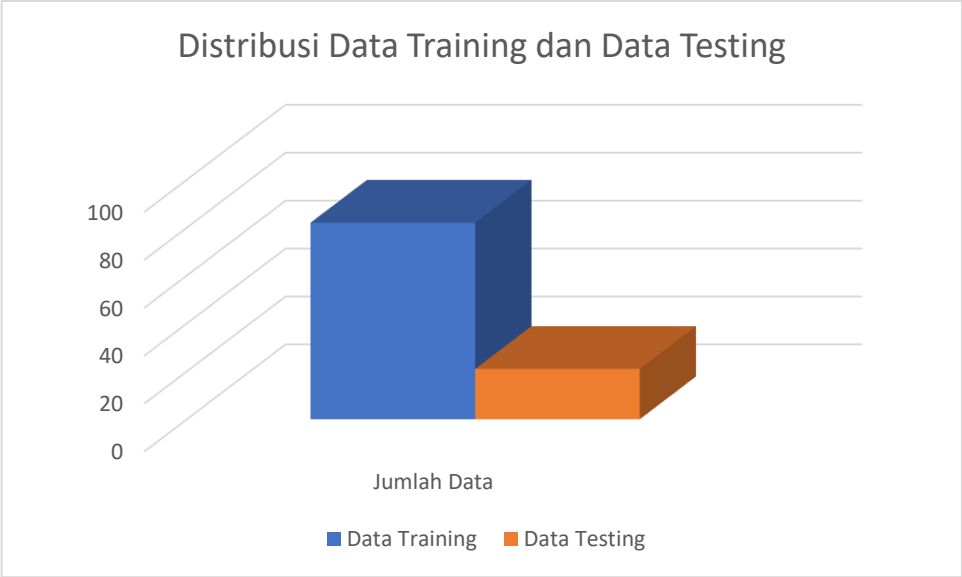

Figure 2. Distribution of Training and Testing Data in the Research Dataset

As shown in Figure 2, out of a total of 103 respondent data points, 82 were allocated as training data and 21 as testing data. This allocation provides the K-NN algorithm with a greater opportunity to learn the patterns and characteristics of smoking dependency levels during the training phase. Meanwhile, the testing data is used to assess the model's ability to make predictions on data that has not been processed previously. This balanced data division supports a more objective model evaluation, so that the classification process can produce more optimal performance that is representative of the actual data conditions.

3.3 Application of the K-NN Algorithm

After the data has been split into training and testing data, the next step is to apply the K-Nearest Neighbor (K-NN) algorithm. Data classification is determined based on the number of nearest neighbors obtained through Euclidean distance calculations. To improve model performance and produce more optimal classifications, the algorithm parameters are adjusted according to the research requirements.

Table 3. K-NN Algorithm Implementation Parameters

Parameter	Value
Algorithm	K-NN
Value of K	5
Training Data	82
Testing Data	21
Distance Method	Euclidean Distance

Based on Table 3, the K-Nearest Neighbor (K-NN) algorithm was applied to the classification process with a value of K set to 5, where the determination of the test data class was performed through voting by the five nearest neighbors with the highest similarity levels. Proximity measurements between data points were performed using the Euclidean Distance method because it is effective in calculating distances for numerical attributes. The selection of K = 5 was based on a balance between model accuracy and generalization ability. Although K = 1, K = 2, and K = 3 yielded the highest accuracy of 100%, a value of K that is too small tends to make the model more sensitive to noise and data variations. Therefore, K = 5 was chosen because it still produces high accuracy (95.24%) while providing a more stable and representative classification process by considering the five nearest neighbors, thereby improving the model's performance in classifying smoking dependency levels on new data.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 41-49
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

3.4 Classification of Smoking Dependency Levels.

After the model parameters were set, the test data was classified to determine the level of smoking dependence for each respondent. The classification results were determined by comparing the characteristics of the test data with the training data that was geographically closest.

Table 4. Results of Smoking Dependency Level Classification Using the K-NN Algorithm

No	Aktual	Prediksi
1	2	2
2	0	0
3	2	1
4	1	1
5	1	1

Table 4 presents some of the classification results obtained using the K-NN algorithm, where the *actual* column shows the true class of each respondent and the *prediction* column shows the classification results generated by the model. Based on the test results, most of the data was successfully classified according to its original class, indicating that the model is able to recognize patterns in the data quite well. Although there are still some prediction errors, such as data that actually belongs to class 2 but is predicted as class 1, the number of such errors is relatively small compared to the number of correct predictions. These results indicate that the K-NN algorithm has good capability in identifying the level of smoking dependence based on the characteristics possessed by each respondent.

3.5 Model Evaluation

The model's success is measured through an evaluation process using *a confusion matrix* and performance metrics such as *accuracy*, *precision*, *recall*, and *F1-score*. This evaluation process is applied to the entire *testing* dataset that has undergone the previous classification stage.

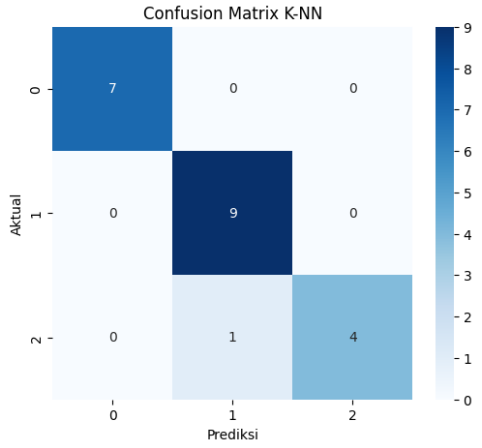

Figure 3. Confusion Matrix of Classification Results Using the K-Nearest Neighbor (K-NN) Algorithm

Based on the confusion matrix in Figure 3, most of the data has been successfully classified into the correct class. This is indicated by the dominance of values on the main diagonal, which represent correct predictions, while the values off the diagonal indicate classification errors, the number of which is relatively small. This indicates that the model has a good ability to distinguish between low, moderate, and high levels of smoking dependence, so the errors that occur do not significantly affect the model's overall performance.

Table 5. K-NN Model Evaluation Metrics

Metrik	Nilai (%)
Accuracy	95.24
Precision	95.71
Recall	95.24
F1-Score	95.10

Based on Table 5, the model's performance was evaluated using four metrics. The evaluation results show that accuracy reached 95.24%, indicating that the majority of the test data was successfully classified correctly. The model's prediction accuracy was also very good, as evidenced by a precision value of 95.71%. The model's ability to identify data across each smoking dependence category is reflected in a recall value of 95.24%, while the balance between precision and recall is demonstrated by an F1-score of 95.10%. Overall, these results confirm that the K-NN algorithm demonstrates excellent capability in classifying levels of smoking dependence.

3.6 Analysis of the Optimal K Value

The K value is one of the main parameters in the K-NN algorithm because it determines the number of nearest neighbors used as a reference in the classification process. To achieve optimal performance, several K values were tested and compared to determine the value that yields the best results.

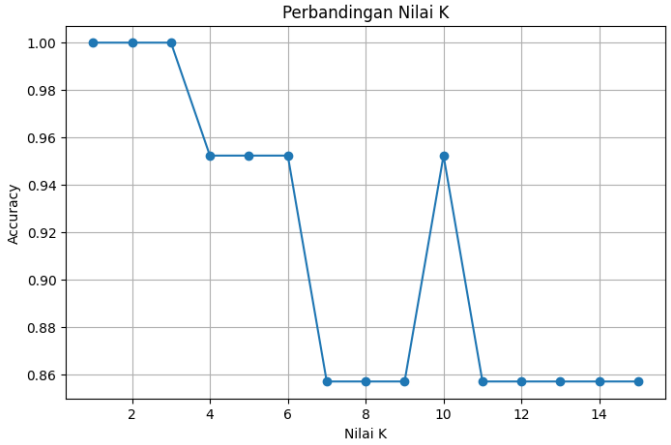

Figure 4. Comparison of K-NN Model Accuracy at Various K Values

Based on Figure 4, the K value was found to have a significant impact on the model's accuracy. The highest accuracy of 100% was achieved at K=1, K=2, and K=3, while increasing the K value caused accuracy to drop to as low as 85.71% in some tests. These results indicate that the dataset exhibits clear patterns, allowing optimal classification to be achieved with a small number of nearest neighbors. Conversely, using a larger K value involves more data in the voting process, thereby reducing the local characteristics that are important in the classification process.

Table 6. Accuracy Test Results Based on Variations in K Values

K	Accuracy (%)
1	100.00
2	100.00
3	100.00
4	95.24
5	95.24
6	95.24
7	85.71
8	85.71
9	85.71
10	95.24
11	85.71
12	85.71
13	85.71
14	85.71
15	85.71

Based on Table 6, testing various values of K shows that K=1, K=2, and K=3 yield the highest accuracy of 100%, while increasing the value of K tends to result in a decrease in accuracy . These results indicate that using a relatively small value of K is more suitable for this dataset because it provides more optimal classification performance compared to larger values of K.

Although the highest accuracy of 100% was achieved at K = 1, K = 2, and K = 3, this study set K = 5 as the primary parameter in the model. This decision was based on the ability of larger K values to minimize the influence of noise and outliers, resulting in classification results that are more stable and consistent. Furthermore, the accuracy achieved by K = 5 remains very good at 95.24%, making this value sufficiently representative to demonstrate the model's performance in classifying levels of smoking dependence.

3.7 Analysis and Discussion

The research results show that the K-Nearest Neighbor (K-NN) algorithm is capable of classifying smoking dependence levels effectively, as indicated by an accuracy of 95.24%. The high *precision*, *recall*, and *F1-score* values indicate that the model can effectively distinguish between low, moderate, and high smoking dependence categories. This performance is supported by the *preprocessing* steps, which include handling *missing values*, transforming categorical data, and applying Min-Max Scaling normalization, ensuring the data is more consistent and suitable for distance calculations in the K-NN algorithm.

Based on testing various values of K, K = 5 was selected because it provides high accuracy while also yielding better generalization capabilities compared to smaller values of K. Additionally, factors such as age at which smoking began, the number of cigarettes consumed per day, social environment, and family circumstances were found to influence the level of smoking dependence. Overall, the research results demonstrate that K-NN is an effective method for identifying the level of smoking dependence and has the potential to be utilized as a decision-making tool in the field of health.

4. CONCLUSION

This study applied the K-Nearest Neighbor (K-NN) algorithm to classify the level of smoking dependence based on respondent characteristics, including age, gender, smoking friends, smoking family members, age at which smoking began, the number of cigarettes consumed per day, and daily allowance. Before the classification process was conducted, the data first underwent preprocessing stages, including data cleaning, categorical data transformation, and normalization using the Min-Max Scaling method to improve data quality and optimize Euclidean distance calculations. Test results showed that the K-NN model was able to produce excellent classification performance with an accuracy of 95.24%, precision of 95.71%, recall of 95.24%, and an F1-score of 95.10% at K = 5. Although K = 1, K = 2, and K = 3 achieved the highest accuracy of 100%, K = 5 was selected because it provides a better balance between accuracy and model stability and is more resistant to the effects of noise and outliers. Based on these results, the K-NN algorithm proved effective in classifying levels of smoking dependence and has the potential to be utilized as a tool to support smoking behavior analysis and decision-making in efforts to prevent and manage smoking dependence.

ACKNOWLEDGMENT

The author would like to thank all parties who have provided support and contributions to the implementation of this research. Special thanks are extended to the Information Systems Study Program and the Master of Informatics Study Program at STIKOM Tunas Bangsa for the facilities and academic support provided. The author also appreciates the supervising lecturer and colleagues who have provided input, suggestions, and assistance throughout the research process up to the preparation of this article. It is hoped that the results of this research will be beneficial and serve as a reference for future research.

REFERENCES

Amelia, R., Nasrul, E., & Basyar, M. (n.d.). *Artikel Penelitian Hubungan Derajat Merokok Berdasarkan Indeks Brinkman dengan Kadar Hemoglobin*. *5*(3), 619–624.

Anggarawati, T. (2021). *Hubungan Ketergantungan Rokok Dengan Kadar Karbonmonoksida Udara Ekspirasi Pada Mahasiswa Akper Kesdam Iv / Diponegoro Semarang*. *3*(2).

Aprilia, T., Informatika, T., & Sri, U. S. (2024). *Klasifikasi Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor dan Metode Naive Bayes*. *4*(2), 156–163. https://doi.org/10.54259/satesi.v4i2.3167

Arvelia, Y., Septa, E., Safitri, A., Sari, A., Anggraini, D., & Suryani, K. (2022). *Persepsi Remaja Yang Berhenti Merokok Dengan Studi Deskriptif*. *5*(1), 26–30. https://doi.org/10.52774/jkfn.v5i1.90

Bayes, N., Jabbar, M. A., Hasmin, E., Susanto, C., & Musu, W. (2022). *Komparasi Algoritma Decision Tree , Naive Bayes , dan K-Nearest Neighbors dalam Klasifikasi Kanker Payudara*. *14*(3), 258–270.

Dan, K. N. N., & Genetika, A. (2023). *Sistem rekomendasi musik spotify menggunakan knn dan algoritma genetika*. *7*(4), 2585–2591.

Ilmiah, J., Sandi, K., & Kriswiastiny, R. (2020). *ARTIKEL PENELITIAN Hubungan Merokok Dan Riwayat Keturunan Dengan Kejadian Hipertensi*. *9*(1), 30–36. https://doi.org/10.35816/jiskh.v10i2.214

Kampung, R., Rawalele, B., Ilmu, D., Masyarakat, K., Kedokteran, F., Yarsi, U., Pusat, J., & Rawalele, B. (2017). *Faktor-faktor yang Berhubungan dengan Perilaku Merokok pada Factors Associated with Teenager ' s Smoking Behavior at*. *5*(March), 194–198.

Kesehatan, P., Bahaya, T., Bagi, M., Hipertensi, P., Education, H., The, A., Of, D., For, S., & Hypertension, P. W. (2025). *INTEGRITAS : Jurnal Pengabdian INTEGRITAS : Jurnal Pengabdian*. *9*(2), 462–470.

Kunci, K. (2023). *Indonesian Journal of Computer Science*. *12*(1), 3746–3756.

Maiyanti, S. I., Zayanti, D. A., Andriani, Y., Suprihatin, B., Desiani, A., Salsabila, A., & Marselina, N. C. (2023). *Perbandingan Klasifikasi Penyakit Kanker Paru-paru menggunakan Support Vector Machine dan K-Nearest Neighbor*. *18*(1), 54–62.

Marito, Y., Yanti, A., Siburian, K., & Sianturi, T. E. (2024). *Upaya Mengatasi Masalah Perilaku Mahasiswa Merokok Menurut Teori Psikoanalisis*. *5*.

Marpaung, D. A., & Marbun, M. (2021). *Sistem Pendukung Keputusan Penentuan Tingkat Kecanduan Masyarakat Terhadap Rokok dengan Metode Fuzzy Mamdani*. *4*(1), 102–107.

Munawaroh, S., Rosyidah, U. A., & Yanuarti, R. (2024). *Klasifikasi Tingkat Kecemasan Atlet Sebelum Bertanding Menggunakan Algoritma K – Nearest Neighbor (KNN) Berbasis Website*. *5*(2).

Neighbor, K., Muzakir, A., Desiani, A., & Amran, A. (2023). *Klasifikasi Penyakit Kanker Prostat Menggunakan Algoritma Naïve Bayes Classification of Prostate Cancer Using Naïve Bayes and K-Nearest Neighbor Algorithms*. *12*(148), 73–79. https://doi.org/10.34010/komputika.v12i1.9629

No, V., Alwanda, A. Y., Utami, E., Yaqin, A., & No, V. (2024). *Infotek : Jurnal Informatika dan Teknologi Analisis Klasifikasi Konsentrasi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor Universitas Hamzanwadi , yang berlokasi di Nusa Tenggara Barat , telah membuktikan dirinya sebagai salah satu institusi pendidik*. *7*(2).

No, V., Amri, Z., Rodi, M., Wathani, M. N., Bagja, A., & No, V. (2025). *Infotek : Jurnal Informatika dan Teknologi Prediksi Diabetes Menggunakan Algoritma K-Nearest (KNN) Teknik SMOTE-ENN Infotek : Jurnal Informatika dan Teknologi Diabetes*

merupakan penyakit tidak menular yang secara serius memengaruhi sistem kesehatan besa. *8*(1), 193–204.

Octavian, Y., Setyanda, G., Sulastri, D., & Lestari, Y. (n.d.). *Artikel Penelitian Hubungan Merokok dengan Kejadian Hipertensi pada Laki- Laki Usia 35-65 Tahun di Kota Padang*. *4*(2), 434–440.

Pringsewu, K., & Tree, D. (n.d.). *Komparasi Penerapan Adaboost Pada K-NN Dan Decision Tree Untuk Prediksi Penyakit Hati*. 907–917.

Putri, A., Hardiana, C. S., Novfuja, E., Try, F., & Siregar, P. (2023). *Comparison of K-NN , Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction Komparasi Algoritma K-NN , Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir*. *3*(April), 20–26.

Rachmat, M., Syafar, M., Rachmat, M., Thaha, R. M., Syafar, M., Promosi, B., Perilaku, I., Kesehatan, F., & Universitas, M. (2013). *Perilaku Merokok Remaja Sekolah Menengah Pertama Smoking Behavior at Junior High School*. *7*(11), 502–508. https://doi.org/10.21109/kesmas.v7i11.363

Rahmadi, A., & Lestari, Y. (2013). *Artikel Penelitian Hubungan Pengetahuan dan Sikap Terhadap Rokok Dengan Kebiasaan Merokok Siswa SMP di Kota Padang*. *2*(1), 25–28.

Sahelvi, E., Cikita, P., & Sapitri, R. M. (2025). *Comparison of K-Nearest Neighbors and Random Forest Algorithms for Recommendations for a Healthy Lifestyle in Prevent Heart Disease Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung*. *5*(July), 830–840.

Salsabila, N. N. (2022). *Jurnal Ekonomi Kesehatan Indonesia* Gambaran Kebiasaan Merokok Di Indonesia Berdasarkan Indonesia Family Life Survey 5 (Ifls 5) Gambaran Kebiasaan Merokok Di Indonesia Berdasarkan Indonesia Family Life Survey 5. *7*(1). https://doi.org/10.7454/eki.v7i1.5394

Sawitri, H., Maulina, F., Kharima, R., & Aqsa, D. (2020). Karakteristik Perilaku Merokok Mahasiswa Universitas Malikussaleh 2019. *6*(1), 78–86.

Status, K., Balita, G., Deteksi, U., Stunting, D., & Neighbor, M. K. (2024). *G-Tech : Jurnal Teknologi Terapan*. *8*(4), 2777–2786.

Status, K., Bayi, G., Algoritma, M., Puskesmas, K. P., Hernita, A., Yulandari, A., & Nur, S. K. (2024). *The Indonesian Journal of Computer Science*. *13*(6), 10281–10290.

Validation, C., Galih, K. S. P., Galih, K. S. P., Validation, C., & Kunci, K. (2023). *1 1 , 2*. 5*(2), 294–300.

Wahyuni, F. (2022). *Pengaruh hipnoterapi terhadap tingkat ketergantungan rokok pada perokok aktif 1*. *7*(1).