Terbit *online* pada laman : https://ejournal.iainbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

Classification of Referral Decision Recommendations in Community Health Centers Using the K-Nearest Neighbor Approach

Leny Tritanto Ningrum[1], Nisrina Salsabila [2]*

[1,2]*Universitas Binaniaga Indonesia, Bogor, Indonesia*

Article Information

Article History:

Submitted: October 9, 2025
Revision: December 16, 2025
Accepted: December 20, 2025
Published: December 30, 2025

Keywords

Diabetes
Classification
Decision Support System

Correspondence

E-mail: lenytrinie@unbin.ac.id*

A B S T R A C T

Diabetes is a chronic disease with a high incidence rate that requires appropriate management, including determining patient referral decisions at community health centers. However, these decisions often still depend on the subjective assessment of medical personnel, resulting in an inaccurate and ineffective process of identifying diabetes patient management. The purpose and objective of this research and development is to identify diabetes patient management for referral decision recommendations at Puskesmas using the K-Nearest Neighbor (KNN) approach to obtain a more accurate and effective process and results so that Puskesmas can more quickly provide appropriate follow-up based on patient laboratory test results. The data used in this study was diabetes patient data at Puskesmas, using variables such as age, systolic and diastolic blood pressure, glucose tests, and referral to hospitals as the target class. The results of the research and classification evaluation using the Confusion Matrix in KNN modeling based on this data showed that the number of patients included in TP=41, TN=38, FP=1, and FN=4, with an accuracy of 94.02%, precision of 97.62%, recall of 91.11%, and F1-Score of 94.25%. These values are categorized as very good because they are able to predict classes correctly at the modeling stage. Thus, this study is considered feasible as a support for referral decision recommendations in identifying the treatment of diabetic patients at Puskesmas.

1. Introduction

Health has always been a major issue and concern for various parties. The increasingly complex and evolving nature of health issues necessitates health efforts in the form of integrated activities to maintain and improve health. In Indonesia, health efforts have been widely implemented, one of which is through health services provided directly to individuals or communities, such as community health Center that call Puskesmas.

Puskesmas, one of the health institutions in Indonesia and a primary health care facility, is currently facing various challenges, one of which is the increasing incidence of chronic diseases such as diabetes. Diabetes is characterized by an increase in blood sugar levels, which, if not treated properly, can lead to various serious complications.

According to the International Diabetes Federation (IDF), in 2021 there were more than 536.6 million people worldwide with diabetes. Indonesia itself ranked fifth as the country with the highest number of people with diabetes in 2021. The Institute for Health Metrics and Evaluation (IHME) reported that in 2019, diabetes was the third leading cause of death in Indonesia. In the 2018 Basic Health Research (Riskesdas), the prevalence of diabetes or diabetes mellitus (DM) was 10.9% and increased to 11.7% in 2023. The results of the 2023 Indonesian Health Survey (SKI) also stated that type 2 diabetes (DMT2) was the most common type of diabetes, accounting for 50.2% of the total weighted sample of 14,900 people.

This situation is further complicated by the fact that many patients are unaware of their health condition until the disease has reached a more serious stage. Patients with diabetes often only visit health facilities such as community health centers (Puskesmas) when their symptoms have worsened, requiring referral to a hospital for further treatment. However, the referral decision-making process often still depends on the subjective assessment of medical personnel, so the process of identifying diabetes patient treatment is not yet accurate and effective. Most patient data is also only used for administration and is not analyzed, even though this data can be used to assist in referral decision-making.

Data mining is presented as a solution with one of the relevant techniques in solving classification problems as a decision support to provide appropriate recommendations. Several previous studies have addressed similar issues with various applications of supervised learning algorithms, including Naive Bayes, Random Forest, Logistic Regression, and KNN. The research by [1] aims to develop a DMT2 prediction model using KNN with a dataset from the Mlati II Community Health Center, Sleman, with variables such as age, blood pressure, BMI, fasting blood sugar levels, and diabetes history. The results of the research evaluation show an accuracy rate of 88%. Meanwhile, research by [2] aims to develop a classification program using KNN to facilitate the initial diagnosis of DM with a dataset covering eight variables, including age, excessive thirst, weight loss despite regular eating patterns, high blood pressure, family history of diabetes, wounds that are difficult to heal, frequent urination at night, and blood sugar test results. The model built in this study produced an accuracy of 93%. However, these studies only focused on disease classification without considering further aspects such as referral decision recommendations. The difference between this study and previous studies is that the prediction results are not limited to classifying whether a patient will develop diabetes or not, but rather focus on the follow-up actions that need to be taken by medical personnel in order to make more accurate decisions, thereby providing greater benefits, especially for community health centers. Based on the various types of algorithms used, several supervised learning algorithms were explored in accordance with the existing dataset, and KNN showed better results than other supervised learning algorithms. Therefore, this study used the KNN algorithm for the modeling stage. The dataset used as the initial data is shown in Table 1.

Tabel 1. Sample Data of Diabetes Patient at Community Health Centers in 2024 with Uncertain Diagnoses

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
1	Anonymous 1	49	120	80	218	Not Referred
2	Anonymous 2	39	117	85	213	Not Referred
3	Anonymous 3	49	120	80	230	Not Referred
…	…	…	…	…	…	…
422	Anonymous 422	61	110	69	128	Not Referred
423	Anonymous 423	60	142	93	200	Not Referred
424	Anonymous 424	87	141	74	378	Not Referred

Table 1 is a sample of data on diabetes patients, specifically Type 2 Diabetes Mellitus, at the Community Health Center in 2024. Based on the 2024 Guidelines for the Management and Prevention of Type 2 Diabetes Mellitus in Indonesia, the diagnosis section states that patients with type 2 diabetes mellitus with sugar levels above 200 and elderly age are included in the hyperglycemic crisis category, requiring follow-up treatment for patients with these conditions. However, based on data from the Community Health Center, there are still inconsistencies where elderly patients with blood sugar levels greater than or equal to 200 are not referred. Based on this dataset, the variables used include age, systolic and diastolic blood pressure ( ), blood sugar tests, and hospital referral as the data class. According to[3], referral indications from PPK 1 (such as Puskesmas) to PPK 2 include patients with hypertension, hyperglycemia, elderly patients, or other conditions. However, there are discrepancies between referral decisions based on the data and the established indications.

Based on the above discussion, this study aims to obtain more accurate results in identifying diabetes patient management at Puskesmas for referral decision recommendations, to obtain a more effective process in identifying diabetes patient management for referral decision recommendations, to apply the KNN algorithm in identifying diabetes patient management for referral decision recommendations, and to measure the accuracy and effectiveness of the KNN algorithm in identifying diabetes patient management for referral decision recommendations.

2. Method

2.1 Data Mining

This study uses data mining techniques, which is the process of extracting patterns or knowledge from large data sets to obtain useful information that was previously unknown but has potential value [4]. One of the standard data handling processes in data mining is the Cross-Industry Standard Process for Data Mining (CRISP-DM) by implementing several stages of data management as follows [5].

a. Business Understanding, the scope of understanding the underlying business framework of a d a ta analysis;
b. Data Understanding, users must understand the substance of the data as a basis for analysis. Data that has attribute and value relationships must be interpreted so that entities can have a common thread;
c. Data Preparation, users must select, clean, construct, integrate, and present data formats;
d. Modeling, performed by fitting data or approaching models from data;
e. Evaluation, testing the accuracy of model selection;
f. Deployment, the characteristics built must be able to accommodate the needs.

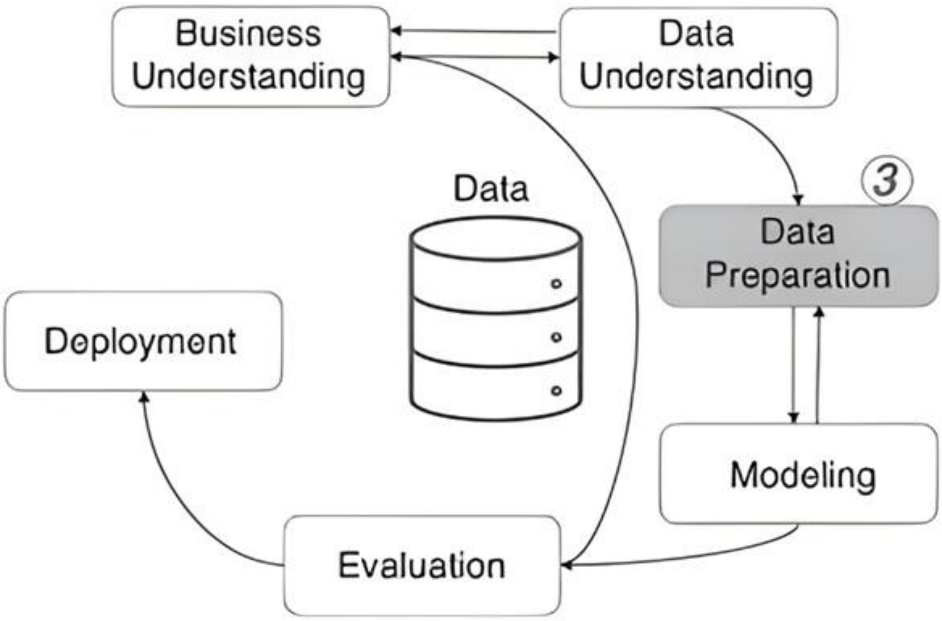

Figure 1. CRISP-DM Stages [5]

2.2 K-Nearest Neighbor (KNN)

One data mining technique that can be used to identify diabetes patient management for referral decision recommendations at community health centers is the classification technique. Classification is a technique in data mining that functions to group data into certain classes or categories according to their attributes [6]. The main objective of this technique is to build a model capable of predicting labels or categories of new data based on patterns learned from training data [7]. Through analysis of the attributes that influence the grouping, process, an understanding of the factors that determine the classification results can be obtained [8]. One classification algorithm is K-Nearest Neighbor (KNN).

KNN is a classification algorithm that works by utilizing labeled data as the basis for learning [9]. This algorithm is also used to calculate the probability of each class of selected testing data [10]. The process is carried out by measuring the similarity distance between the testing data and the training data [11][12]. Each distance calculation result is then associated with the original class of the training data involved in the calculation. Next, the K value is determined [13], which is a parameter in the KNN algorithm used in the study to determine the number of neighbors or data that become references for label prediction. From the K data, the probability of each class appearing is calculated, and the class with the highest probability is selected as the prediction for the testing data [14].

Since classification requires categorical results, the probability values from the model are converted into binary form (0 or 1) using a threshold, which is usually set to 0.5 by default [15]. Meanwhile, to calculate the distance between training data and testing data, the Euclidean Distance formula [16][17] can be used as follows [18] .

$$d(x,y) = \sqrt{(\sum_{i=1}^{n}(x_i - y_i)^2)} \tag{1}$$

Explanation:
d (x, y) = distance
n = data dimensioon
i = data variable
xi = data sample
yi = test data

2.3 Confusion Matrix

Classification model evaluation aims to show the extent to which the model is able to perform well in predicting or grouping (classifying) available data. One of the metrics widely used to evaluate classification models is the Confusion Matrix [19]. The Confusion Matrix is an evaluation method used to assess the performance of classification models by displaying the number of correct and incorrect predictions based on a comparison with actual values[20]. The components of the Confusion Matrix include True Positive (TP), which is the number of correct positive predictions; True Negative (TN), which is the number of correct negative predictions; False Positive (FP), which is the number of incorrect positive predictions; and False Negative (FN), which is the number of incorrect negative predictions[21]. In addition, four KNN evaluation techniques are also used, as follows [22]:

a. Accuracy, a measure indicating the percentage of correct predictions out of all predictions;

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

b. Precision, the ratio of correct positive predictions to the total number of positive predictions;

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

c. Recall, the ratio of correct positive predictions to the total number of positive samples;

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

d. F1 Score, the harmonic mean of precision and recall.

$$F1\ Score = 2 \times \frac{Precision \ \times Recall}{Precision + Recall} \tag{5}$$

3. Results and Discussion

After describing this research method, the research stages were adjusted based on the data mining concept using the CRISP-DM stages as follows:

3.1 Business Understanding

In the first stage, a needs analysis was conducted through interviews at the research site, namely at a community health center (Puskesmas) related to decision-making for referring diabetic patients. It was found that decision-making often still depended on the subjective assessment of medical personnel, so that the process of identifying the treatment of diabetic patients was not yet accurate and effective. Given these conditions, this study aims to apply the KNN algorithm to the identification of diabetes patient treatment to assist in making more accurate and effective referral decision recommendations.

3.2 Data Understanding

Through interviews with sources, namely medical personnel in the Non-Communicable Diseases (NCD) clinic who treat diabetes, documents were collected in the form of diabetes patient data, specifically DMT2 at the Puskesmas in 2024, totaling 424 records and approximately 6 variables. Of the 424 records, 161 patients were classified as "Referred" to the hospital and 263 patients were classified as "Not Referred." The following is a sample of the diabetes patient data collected, as shown in Table 2.

Tabel 2. Sample Data of Diabetes Patients at Community Health Centers in 2024

No	Patient Name	Date of Birth	Systolic	Diastolic	Height (CM)	Weight (kg)	Waist Circumferen	Blood Sugar Test	Refer to Hospital
1	Anonymous 1	08-08-1974	120	80	168,00	65,00	80	218	Not referred
2	Anonymous 2	07-11-1975	117	85	160,00	69,00	94	213	Not referred
3	Anonymous 3	11-03-1985	120	80	170,00	75,00	80	230	Not referred
…	…	…	…	…	…	…	…	…	…
423	Anonymous 423	04-12-1966	110	69	154,00	48,00	70	128	Not referred
424	Anonymous 424	08-12-1966	142	93	146,00	49,00	72	200	Referred
425	Anonymous 425	18-03-1971	141	74	168,00	74,00	75	378	Referred

3.3 Data Preparation

The patient data was selected to ensure that the data analyzed was valid. The selected variables consisted of patient name, date of birth, systolic and diastolic blood pressure, glucose test, and hospital referral as the target class variable. The patient name variable was hidden and not used in the modeling stage, while the date of birth variable was transformed into a numerical age variable to make it easier and more useful for analysis. Data cleaning was then performed to avoid duplicate and empty data. Of the total 425 records, 1 empty record was found, resulting in 424 valid records for analysis, as shown in Table 3.

Tabel 3. Sample Data Cleaning and Transformation of Diabetes Patients at the Community Health Center in 2024

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
1	Anonymous 1	49	120	80	218	Not Referred
2	Anonymous 2	39	117	85	213	Not Referred
3	Anonymous 3	49	120	80	230	Not Referred
…	…	…	…	…	…	…
422	Anonymous 422	61	110	69	128	Not Referred

423	Anonymous 423	60	142	93	200	Referred
424	Anonymous 424	87	141	74	378	Referred

Next, the data will be divided into training data and testing data based on the referral indications of DMT2 patients from PPK (Health Service Providers) 1 such as Puskesmas to PPK 2 such as hospitals. These referral indications include patients with hypertension (systolic and diastolic blood pressure > 140/90 mmHg), hyperglycemia (blood sugar level > 300 mg/dl), and elderly patients (aged > 55 years). In this study, training and testing data were selected manually based on whether they were appropriate or inappropriate for the referral class (Referred/Not Referred) , so they were considered to have a valid class.

The DMT2 patient dataset was divided into two parts: training data and testing data, with a data distribution ratio of 80:20, where 80% was used as training data, starting from data point 1 to data point- 339 with a total of 255 data points, and 20% as testing data starting from data point 340 to data point 424 with a total of 84 data points. The training data serves to train the model to be able to recognize patterns and relationships between variables better.

The next stage in data preparation is normalization, which involves rescaling the numerical values in the training and testing data to a range of 0 to 1 using Min-Max Normalization so that they are on a balanced scale, making the distance calculation and classification processes more accurate and fair [23]. The following are the normalized training and testing data as shown in Tables 4 and 5.

Table 4. Normalized Training Data

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
1	Anonymous 1	0,4328	0,2560	0,3511	0,2739	Not Referred
2	Anonymous 2	0,2836	0,2320	0,4043	0,2633	Not Referred
3	Anonymous 3	0,4328	0,2560	0,3511	0,2994	Not Referenced
…	…	…	…	…	…	…
337	Anonymous 337	0,8358	0,3360	0,3298	0,2675	Referred
338	Anonymous 338	0,6716	0,2400	0,2660	0,2590	Referred
339	Anonymous 339	0,7015	0,7040	0,4255	0,2548	Referenced

Table 5. Normalized Testing Data

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
340	Anonymous 340	0,6418	0,4080	0,4894	0,2569	Referred
341	Anonymous 341	0,7164	0,2320	0,3511	0,0000	Not Referred
342	Anonymous 342	0,9552	0,4000	0,2872	0,2675	Referred
…	…	…	…	…	…	…
422	Anonymous 422	0,6119	0,1760	0,2340	0,0828	Not Referred
423	Anonymous 423	0,5970	0,4320	0,4894	0,2357	Referred
424	Anonymous 424	1,0000	0,4240	0,2872	0,6136	Referred

3.4 Modeling

At this stage, the KNN algorithm is applied, where KNN is first trained using training data to determine the extent of the model's performance in making predictions. The training data used consists of 339 patient data, with 115 records with the hospital referral class "Referred" and 224 records with the hospital referral class "Not Referred". The initial stage *of modeling* is to determine the number of k as the closest neighbors. The process of determining the nearest neighbors can be done by determining the distance function in the data proximity matrix based on the calculation formula of the square root of the data count training [24] and using testing to measure the accuracy of the best k count using the *K-Fold Cross Validation* method [25]. K in KNN is determined by an odd number to avoid series results in determining the prediction class, and to determine the best K count in this study, the K-*Fold Cross Validation* technique by testing the accuracy at K=1, K=3, and K=5. The accuracy measurement results for K=1 obtained an accuracy of 94%, K=3 obtained an accuracy of 90%, and K=5 obtained an accuracy of 89%. Based on the results of the prediction accuracy measurement with *k-Fold Cross Validation*, the number of K for the nearest neighbor used in this study was determined to be 1 nearest neighbor. For KNN modeling, the training data sample is shown in Table 6 and the testing data sample is shown in Table 7.

Table 6. KNN Modeling Training Data Sample

No	Age	Systolic	Diastolic	Glucose Test	Refer to Hospital
1	0,4328	0,2560	0,3511	0,2739	Not Referred
2	0,2836	0,2320	0,4043	0,2633	Not Referred
3	0,4328	0,2560	0,3511	0,2994	Not Referred
…	…	…	…	…	…
337	0,8358	0,3360	0,3298	0,2675	Referred
338	0,6716	0,2400	0,2660	0,2590	Referred
339	0,7015	0,7040	0,4255	0,2548	Referred

Table 7. Sample Data Testing KNN Modeling

No	Age	Systolic	Diastolic	Glucose Test	Refer to Hospital
340	0,6418	0,4080	0,4894	0,2569	Referred
341	0,7164	0,2320	0,3511	0,0000	Not Referred
342	0,9552	0,4000	0,2872	0,2675	Referred
…	…	…	…	…	…
422	0,6119	0,1760	0,2340	0,0828	Not Referred
423	0,5970	0,4320	0,4894	0,2357	Referred
424	1,0000	0,4240	0,2872	0,6136	Referenced

Application of manual KNN modeling calculations using *Euclidean Distance* with training data and testing data to determine class predictions. The Euclidean formula can be seen in the equation below:

$$d(x,y) = \sqrt{(\sum_{i=1}^{n}(x_i - y_i)^2)}$$

The calculation of data distance using *Euclidean Distance* is used to measure the distance between *the test data* and all data in *the training data*, and then it will be ranked based on the closest distance and selected according to the predetermined K number to determine the class prediction. *The Euclidean Distance* calculation is applied to the first *test data* (data no. 340) against all *training data*. The following is an example of the distance calculation between data no. 340 and data no. 1.

$$d(x,y) = \sqrt{\begin{matrix}((0{,}6418 - 0{,}4328)^2 + (0{,}4080 - 0{,}2560)^2 + (0{,}4894 - 0{,}3511)^2 \\ +(0{,}2569 - 0{,}2739)^2)\end{matrix}} = 0{,}0293$$

The result of the *Euclidean Distance* calculation for the first test data (data point 340) with *the* first *training data* (data no. 1) produces a value of 0.0293. The distance calculation for this data is continued with the second *training data* (data no. 2) and so on until the distance of the data (data no. 340) to all training data is obtained. Then, all data distances are sorted and 1 data (K=1) is selected, which is the data with the closest distance to data no. 340. thus producing a prediction that the referral recommendation for data point 340 is the same as the closest data recommendation, as shown in Table 8.

Table 8. Results of Euclidean Distance Calculations and The ranking of the Closest Distance for The Test data (data poin 340)

No	Euclidean	Rank
1	0,2936	145

2	0,4081	263
3	0,2961	150
…	…	…
148	0,2201	57
149	0,2760	114
150	0,3751	245
…	…	…
311	0,5992	324
312	0,2835	126
313	0,0556	1
…	…	…
337	0,2615	97
338	0,2811	122
339	0,3086	170

Table 8 shows that the closest distance of *the testing* data (data no. 340) is with data no. 313, where the ranking column shows the closest ranking or order, which states that the referral recommendation for data no. 340 is the same as data no. 313, namely "Referred". The next modeling step is to calculate the data distance using *Euclidean Distance* for the testing data (data no. 341) with all *training data* to obtain the closest data distance ranking results as shown in Table 8.

After calculating the data distance between all *testing data* and all *training data*, the results of the comparison between the actual class and the KNN prediction class are obtained. These results show how the model works in classifying patients and reveal the extent of the model's consistency in predicting classes. The following modeling results show the distribution of actual and predicted classes in *the training data* as shown in Table 9.

Table 9. KNN Modeling Results on Training Data

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Actual	Predicted
340	Anonymous 340	0,7164	0,2320	0,3511	0,0000	Not Referenced	Not Referred
341	Anonymous 341	0,9552	0,4000	0,2872	0,2675	Referred	Referred
342	Anonymous 342	0,5970	0,4640	0,4894	0,2718	Referred	Referred
…	…					…	…
422	Anonymous 422	0,6119	0,1760	0,2340	0,0828	Not Referenced	Not Referred
423	Anonymous 423	0,5970	0,4320	0,4894	0,2357	Referred	Referred
424	Anonymous 424	1,0000	0,4240	0,2872	0,6136	Referred	Referred

3.5 Evaluation

Next, an evaluation of the KNN modeling was conducted to see the extent to which the algorithm was able to distinguish between patients who needed to be referred to a hospital and those who did not. This evaluation was the basis for determining whether the KNN that was built was good enough and suitable for use in the prediction stage for patient data, particularly as a referral decision support system.

The evaluation was carried out by calculating the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) [27] from the KNN modeling results [28]. These values show how much data was predicted correctly or incorrectly based on its actual class [29], so it can be concluded that TP=41, TN=38, FP=1, and FN=4. Table 9 below is an evaluation of the KNN modeling results with training data in the form of a Confusion Matrix.

Table 9. Confusion Matrix of KNN Modeling

	Prediction "Referred"	Prediction "Not Referred"
Actual "Referred"	41	4
Actual "Not Referred"	1	38

Based on these results, the evaluation metrics obtained were *accuracy* of 94.05%, *precision* of 97.62%, *recall* of 91.1%, and *f1-score* of 94.25%. The high *accuracy*, *precision*, *recall*, and *f1-score* values indicate that all patients with referral classes were successfully identified by the model. With these results, the KNN model that was built can be considered suitable for use in the prediction stage of the testing data [30] in order to identify diabetes patient treatment and provide referral decision recommendations.

3.6 Deployment

Since the modeling results on the training data are considered feasible, the next step is to deploy the KNN model and make predictions on new data [31]. This prediction process uses the entire training data as a reference, with the same parameter, namely K=1. The following is the training data used for predicting referral follow-up recommendations for type 2 diabetes mellitus patients, as shown in Table 10.

Tabel 10. Training Data for KNN Model Prediction

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
1	Anonymous 1	0,4328	0,2560	0,3511	0,2739	Tidak Dirujuk
2	Anonymous 2	0,2836	0,2320	0,4043	0,2633	Tidak Dirujuk
3	Anonymous 3	0,4328	0,2560	0,3511	0,2994	Tidak Dirujuk
…	…	…	…	…	…	…
422	Anonymous 422	0,6119	0,1760	0,2340	0,0828	Tidak Dirujuk
423	Anonymous 423	0,5970	0,4320	0,4894	0,2357	Dirujuk
424	Anonymous 424	1,0000	0,4240	0,2872	0,6136	Dirujuk

The new data to be tested on the KNN model is the result of laboratory tests on a 60-year-old patient with a systolic pressure of 129, diastolic pressure of 90, and blood sugar level of 220. The following are the steps for its application:

1. Normalize the data using the Min Max Normalization model. The test data before normalization can be seen in Table 11.

Table 11. Test Data for KNN Model Prediction

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
1	Anonymous XXX	60	129	90	220	?

Next, normalize the data in Table 11, and the test data values after normalization are shown in Table 12.

Table 12. Test Data for KNN Model Prediction Application After Normalization

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
1	Anonymous XXX	0,5970	0,3280	0,4574	0,2781	?

2. Calculate the distance of the new data using Euclidean Distance with all training data. The following is an example of calculating the distance between the new data to be predicted and the first data.

$$d(x,y) = \sqrt{\begin{matrix}((0{,}5970 - 0{,}4328)^2 + (0{,}3280 - 0{,}2560)^2 + (0{,}4574 - 0{,}3511)^2 \\ +(0{,}2781 - 0{,}2739)^2)\end{matrix}} = 0{,}2085$$

The result of calculating *the Euclidean Distance of* the new data with all *training data* yields a value of 0.2085. this data distance calculation is continued on the second *training data* (second data) and so on until the data distance to all training data is obtained. Then, sort all data distances and select 1 data (K=1) which is the data with the closest distance to the new data, thus producing a prediction that the referral recommendation for the new data is the same as the closest data recommendation, as shown in Table 13.

Table 13. Results of Euclidean Distance calculations and distance ranking order New data with all *training data*

No	Euclidean	Rank
1	0,2085	112
2	0,3324	297
3	0,2095	115
…	…	…
15	0,2054	106
16	0,0272	1
17	0,2815	228
…	…	…
422	0,3338	298
423	0,1168	21
424	0,5596	408

Table 13 shows that the closest distance of the new data is with the 16th data, where the ranking column shows the closest ranking or order, which states that the referral recommendation for the new data is the same as the 16th data, namely "Referred." Thus, the prediction results for new type 2 diabetes mellitus patients based on laboratory blood test results can be seen in Table 14.

Table 14. Prediction Results of Patient Referral Recommendations from KNN

No	Patient Name	Age	Systolic	Diastolic	Blood Sugar Test	Refer to Hospital
1	Anonymous XXX	60	129	90	220	Referred

Based on the results in Table 14, it can be seen that the KNN model is capable of providing supportive recommendations for patient referral follow-up based on blood test data from the laboratory and is consistent with historical data as reference data for referral recommendations.

4. Conclusion

The application of the KNN algorithm with a K value of 1 in identifying diabetes patient management for referral decision recommendations at the Community Health Center (Puskesmas) was able to provide fairly accurate prediction results, as shown by the accuracy value in KNN modeling of 94.05%, precision of 97.62%, recall of 91.11, and an F1-Score of 94.25%. Based on the high precision results, this study is considered feasible and can be used to support referral decision recommendations for diabetic patients at Puskesmas because it is able to identify patients who need to be referred for further treatment. Suggestions for further research include the use of a larger diabetic patient data set and more diverse variables, the use of other algorithms to improve performance and prediction accuracy, and the development of a referral follow-up recommendation application.

Reference

[1] A. Oktaviana, D. P. Wijaya, A. Pramuntadi, And D. Heksaputra, “Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” Inst. Ris. Dan Publ. Indones., Vol. 4, No. 3, Pp. 812–818, 2024, Doi: 10.57152.

[2] H. A. D. Fasnuari, H. Yuana, And M. T. Chulkamdi, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus: Studi Kasus : Warga Desa Jatitengah,” Antivirus J. Ilm. Tek.

Inform., No. Vol 16 No 2 (2022): November 2022, Pp. 133–142, 2022.

[3] Menteri Kesehatan Republik Indonesia, "Pedoman Nasional Pelayanan Kedokteran Tata Laksana Diabetes Melitus Tipe 2 Dewasa," P. 183, 2020.

[4] M. Z. Deny Jollyta, William Ramadhan, Konsep Data Mining Dan Penerapan, Pertama. Deepublish, 2020.

[5] S. J. Gellysa Urva, Desyanti, Isa Albanna, Muchamad Sobri Sungkar, I Made Agus Oka Gunawan, Iwan Adhicandra, Sahrul Ramadhan, Rifky Lana Rahardian, Herlawati, Rahmadya Trias Handayanto, Anak Agung Gede Bagus Ariana, Hartatik, Prima Dina Atika, "Penerapan Data Mining Di Berbagai Bidang : Konsep, Metode, Dan Studi Kasus," Pt. Sonpedia Publishing Indonesia, 2023. .

[6] B. Budiman, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Penelusuran Minat Calon Mahasiswa Baru," Nuansa Inform., Vol. 15, No. 2, Pp. 37–52, 2021, Doi: 10.25134/Nuansa.V15i2.4162.

[7] I. H. Witten, E. Frank, And M. A. Hall, Data Mining. 2023.

[8] Z. Setiawan Et Al., Buku Ajar Data Mining. Pt. Sonpedia Publishing Indonesia, 2023.

[9] F. Sodik And I. Kharisudin, "Analisis Sentimen Dengan Svm , Naive Bayes Dan Knn Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial Twitter," Prisma, Vol. 4, Pp. 628–634, 2021.

[10] L. Nur Aziza, R. Yuli Astuti, B. Akbar Maulana, And N. Hidayati, "Application Of The K-Nearest Neighbor Algorithm For Food Security Classification In Central Java Province," Malcom Indones. J. Mach. Learn. Comput. Sci. J., Vol. 4, No. 2, Pp. 404–412, 2024, [Online]. Available: Https://Journal.Irpi.Or.Id/Index.Php/Malcom/Article/View/1201%0ahttps://Journal.Irpi.Or.Id/Index.Php/Malcom/Article/Download/1201/553.

[11] M. A. Afrianto And M. Wasesa, "Booking Prediction Models For Peer-To-Peer Accommodation Listings Using Logistics Regression, Decision Tree, K-Nearest Neighbor, And Random Forest Classifiers," J. Inf. Syst. Eng. Bus. Intell., Vol. 6, No. 2, P. 123, 2020, Doi: 10.20473/Jisebi.6.2.123-132.

[12] N. Benayad, Z. Soumaya, B. D. Taoufiq, And A. Abdelkrim, "Features Selection By Genetic Algorithm Optimization With K-Nearest Neighbour And Learning Ensemble To Predict Parkinson Disease," Vol. 12, No. 2, Pp. 1982–1989, 2022, Doi: 10.11591/Ijece.V12i2.Pp1982-1989.

[13] N. Buslim, L. K. Oh, M. H. Athallah Hardy, And Y. Wijaya, "Comparative Analysis Of Knn, Naïve Bayes And Svm Algorithms For Movie Genres Classification Based On Synopsis.," J. Tek. Inform., Vol. 15, No. 2, Pp. 169–177, 2022, Doi: 10.15408/Jti.V15i2.29302.

[14] N. Wiliani And N. Fathurrahman, Identifikasi Cacat Pada Permukaan Mobil Menggunakan Metode K-Nearest Neighbors. Penerbit Nem, 2024.

[15] N. S. N. Az-Zahrani, H. K. A. Eloi, F. Salim, A.-Z. A. Ramadhani, C. Meysyanti, And L. N. A. Purwantiningsih, Python Untuk Analisis Data. Siega Publisher, 2025.

[16] D. Trianda, D. Hartama, And S. Solikhun, "Comparison Of Hyperparameter Tuning Methods For Optimizing K-Nearest Neighbor Performance In Predicting Hypertension Risk," J. Tek. Inform., Vol. 18, No. 1, Pp. 111–121, 2025, Doi: 10.15408/Jti.V18i1.42260.

[17] G. Liu, H. Zhao, F. Fan, G. Liu, Q. Xu, And S. Nazir, "An Enhanced Intrusion Detection Model Based On Improved Knn In Wsns," Sensors, Vol. 22, No. 4, Pp. 1–18, 2022, Doi: 10.3390/S22041407.

[18] A. S. Nasution, A. Alvin, A. T. Siregar, And M. S. Sinaga, "Knn Algorithm For Identification Of Tomato Disease Based On Image Segmentation Using Enhanced K-Means Clustering," Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control, Vol. 4, No. 3, 2022, Doi: 10.22219/Kinetik.V7i3.1486.

[19] M. K. Mayangsari, I. Syarif, And A. Barakbah, "Evaluation Of Stratified K-Fold Cross Validation For Predicting Bug Severity In Game Review Classification," Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control, Vol. 4, No. 3, Pp. 277–288, 2023, Doi: 10.22219/Kinetik.V8i3.1740.

[20] M. O. Khairandish, M. Sharma, V. Jain, J. M. Chatterjee, And N. Z. Jhanjhi, "A Hybrid Cnn-Svm

Threshold Segmentation Approach For Tumor Detection And Classification Of Mri Brain Images," Irbm, Vol. 43, No. 4, Pp. 290–299, 2022, Doi: 10.1016/J.Irbm.2021.06.003.

[21] S. U. Hassan, J. Ahamed, And K. Ahmad, "Analytics Of Machine Learning-Based Algorithms For Text Classification," Sustain. Oper. Comput., Vol. 3, No. February, Pp. 238–248, 2022, Doi: 10.1016/J.Susoc.2022.03.001.

[22] I. Fajri Et Al., Data Mining. Serasi Media Teknologi, 2024.

[23] P. P. Allorerung, A. Erna, M. Bagussahrir, And S. Alam, "Analisis Performa Normalisasi Data Untuk Klasifikasi K-Nearest Neighbor Pada Dataset Penyakit," Jiska (Jurnal Inform. Sunan Kalijaga), Vol. 9, No. 3, Pp. 178–191, 2024, Doi: 10.14421/Jiska.2024.9.3.178-191.

[24] S. Zhang, "Challenges In Knn Classification," Ieee Trans. Knowl. Data Eng., Vol. 34, No. 10, Pp. 4663–4675, 2022, Doi: 10.1109/Tkde.2021.3049250.

[25] J. L. Leevy, J. M. Johnson, J. Hancock, And T. M. Khoshgoftaar, "Threshold Optimization And Random Undersampling For Imbalanced Credit Card Data," J. Big Data, Vol. 10, No. 1, 2023, Doi: 10.1186/S40537-023-00738-Z.

[26] J. J. Pangaribuan, A. Maulana, And R. Romindo, "Unleashing The Power Of Svm And Knn: Enhanced Early Detection Of Heart Disease," Jitk (Jurnal Ilmu Pengetah. Dan Teknol. Komputer), Vol. 10, No. 2, Pp. 342–351, 2024, Doi: 10.33480/Jitk.V10i2.5719.

[27] W. E. Susanto And D. Riana, "Komparasi Algoritma," J. Speed, Vol. 8, No. 3, Pp. 18–27, 2016.

[28] Y. Yustikasari, H. Mubarok, And R. Rianto, "Comparative Analysis Performance Of K-Nearest Neighbor Algorithm And Adaptive Boosting On The Prediction Of Non-Cash Food Aid Recipients," Sci. J. Informatics, Vol. 9, No. 2, Pp. 205–217, 2022, Doi: 10.15294/Sji.V9i2.32369.

[29] A. Fahmi Limas, R. Rosnelly, And A. Nursie, "A Comparative Analysis On The Evaluation Of Knn And Svm Algorithms In The Classification Of Diabetes," Sci. J. Informatics, Vol. 10, No. 3, P. 251, 2023, Doi: 10.15294/Sji.V10i3.44269.

[30] M. R. Siregar, D. Hartama, And S. Solikhun, "Optimizing The Knn Algorithm For Classifying Chronic Kidney Disease Using Gridsearchcv," Jitk (Jurnal Ilmu Pengetah. Dan Teknol. Komputer), Vol. 10, No. 3, Pp. 680–689, 2025, Doi: 10.33480/Jitk.V10i3.6214.