International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

Defect Rate Prediction in Manufacturing Process Using K-Nearest Neighbor Algorithm

Muhammad David[1], * Muhammad Azka Firdaus [2]

[1,2]Faculty of Defense Science and Technology, Universitas Pertahanan Republik Indonesia, Bogor, Indonesia

Email address:
muhdavidd1808@gmail.com, muhammadazkafirdaus15@gmail.com

*Corresponding author: muhdavidd1808@gmail.com

Received: July 5, 2024; **Accepted**: August 16, 2024; **Published**: August 16, 2024

Abstract: The efficiency and effectiveness in the manufacturing industry are significantly impacted by artificial intelligence technology. An important application involves the improvement of product quality, which is measurable through the defects occurring during the production process. This research is aimed at predicting defects in the manufacturing process using the K-Nearest Neighbor (KNN) algorithm with various distance measurement methods, namely Euclidean, Minkowski, and Manhattan distances. The research methodology is composed of four stages: dataset collection, data preprocessing, modeling, and evaluation. The focus of this research is on the optimal K value and the conditions that yield the highest accuracy, considering various scenarios of training and test data splitting ratios and different random state values. The test results indicate that the Minkowski distance method, with a data division ratio of 80% for training data, 20% for test data, and a random state value of 32, provides the best performance, with an optimal K value of 10 and an accuracy of 86.41%.

Keywords: Defect Rate, K-Nearest Neighbor, Euclidean, Minkowski, and Manhattan

1. Introduction

The ongoing technological transformation has led to significant changes in the manufacturing sector, an essential cornerstone for economic progress in many countries. The digital transformation at the center of the industrial revolution has affected almost all manufacturing activities. The emergence of artificial intelligence technology has a favorable impact on manufacturing companies. With this technology, manufacturing companies are faced with great opportunities that can improve the company's operational efficiency. Artificial intelligence can provide intelligent and adaptive solutions to problems faced in the manufacturing environment. Artificial intelligence technology in manufacturing companies can improve operational efficiency in various ways. For example, artificial intelligence allows

DOI: 10.52362/ijiems.v3i2.1599

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

companies to optimize supply chains by predicting demand, managing inventory, and planning production more accurately. Not only that, the application of artificial intelligence can also be used to improve product quality, increase customer satisfaction, optimize production processes, and others. By utilizing the capabilities of artificial intelligence in data analysis and prediction, companies can make good decisions on all aspects of company activities [1]. One example that can be processed through artificial intelligence is the prediction of defects or failures in the manufacturing process. In manufacturing operations, product quality is one of the key factors determining a company's success. This quality can be measured by the defects or failures that occur during the production process. Such defects not only affect production efficiency but also impact the company's finances and reputation. Therefore, reducing the defect rate is one of the goals of the manufacturing process [2].

To reduce the level of disability, researchers conducted a research by applying artificial intelligence technology, namely the prediction of the level of disability in the manufacturing process. One of the artificial intelligence concepts that can predict defects or failures in the manufacturing process is the use of the Machine Learning approach. Machine learning is an approach to artificial intelligence that can help computers perform modeling based on experience and accurately predict future events. Machine Learning approaches can be classified into supervised learning and unsupervised learning [3]. This research process belongs to one type of supervised learning, namely classification, to predict the level of defects in the manufacturing process. This research uses one of the classification algorithms, namely K-Nearest Neighbor (KNN). The K-Nearest Neighbor algorithm is a data classification algorithm based on the closest distance from the training data with several k values of the nearest neighbors. The k value and distance model can affect the accuracy of the KNN algorithm [4].

In this research, researchers compared the Euclidean, Minkowski, and Manhattan distance calculation models in the KNN algorithm with several conditions in determining the nearest neighbor distance with a value of k = 2 to k = 14 using a dataset of production process defects in the manufacturing industry obtained from Kaggle. Thus, this research will get the most optimal distance model and k value in detecting production defects.

2. Research Method

The research methodology consists of several process stages: data collection, data preprocessing, model building, and evaluation of measurement metrics. Each stage has an important role in achieving the research goal, which is to develop a model that can accurately predict production defects.

Figure 1 shows the flow of the research conducted, illustrating each stage of the process, from data collection to the final evaluation of the model.

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

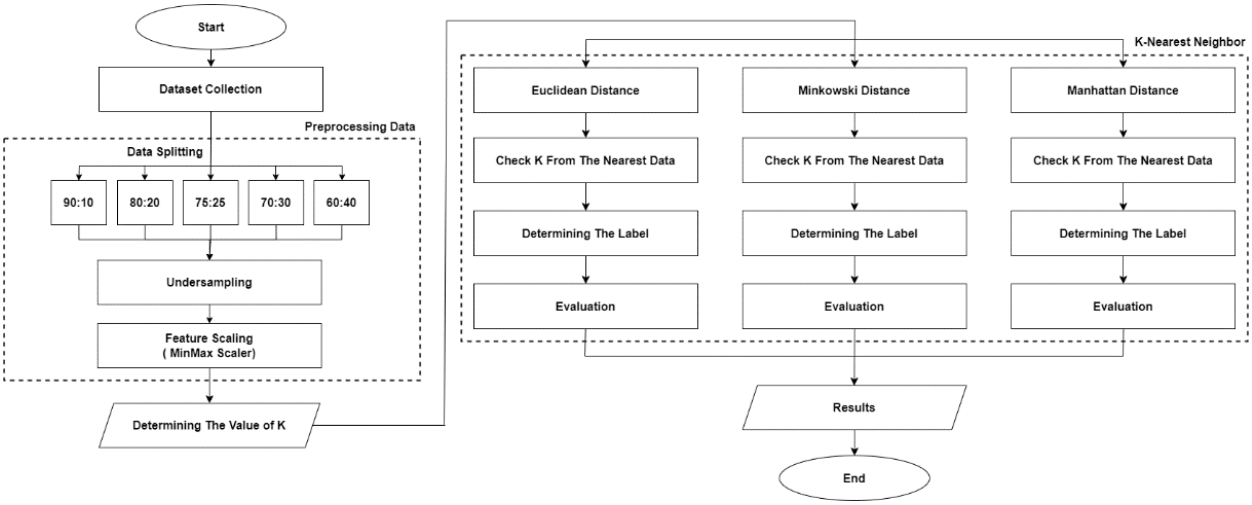

Figure 1. Research Process

The research process begins with data collection. This research uses a dataset of production defects in the manufacturing process obtained through the Kaggle website. This dataset contains 3,240 data consisting of 16 independent variables and 1 dependent variable and can be accessed at https://www.kaggle.com/datasets/rabieelkharoua. The variables that are the focus of this research are as follows.

Table 1. Variables of Predicting Manufacturing Defects Dataset

Variables		Data Type	Range
Production Metrics	Production Volume (X1)	Integer	100 to 1000 units/day
	Production Cost (X2)	Float	$5000 to $20000
Supply Chain and Logistics	Supplier Quality (X3)	Float	80% to 100%
	Delivery Delay (X4)	Integer	0 to 5 days
Quality Control and Defect Rates	Defect Rate (X5)	Float	0,5 to 5 defects
	Quality Score (X6)	Float	60% to 100%
Maintenance and Downtime	Maintenance Hours (X7)	Integer	0 to 24 hours
	Downtime Percentage (X8)	Float	0% to 5%
Inventory Management	Inventory Turnover (X9)	Float	2 to 10
	Stockout Rate (X10)	Float	0% to 10%
Workforce Productivity and Safety	Worker Productivity (X11)	Float	80% to 100%
	Safety Incidents (X12)	Integer	0 to 10 incidents
Energy Consumption	Energy Consumption (X13)	Float	1000 to 5000 kWh

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

and Efficiency	Energy Efficiency (X14)	Float	0,1 to 0,5
Additive Manufacturing	Additive Process Time (X15)	Float	1 to 10 hours
	Additive Material Cost (X16)	Float	$100 to $500
Defect Status (Y)		Integer	0 for Low Defects 1 for High Defects

The data preprocessing stage in this research includes data splitting and handling data imbalance. The initial process of data preprocessing in this research is data splitting. Data Splitting is dividing data into two or more parts to test models or algorithms. Generally, the dataset is divided into 2 (two) data, namely training data and test data. Training data is used to train the algorithm, while test data is used to evaluate the algorithm's performance [5]. Therefore, the data is divided into two stages/parts in this research. Data splitting in this research uses a ratio of 90:10, 80:20, 75:25, 70:30, and 60:40. After the data is divided, the next step is to perform under-sampling to overcome the imbalance of the dataset. The under-sampling method used in this research is Random Under-Sampling. Where the imbalance of the data used is seen in the amount of data for each level of disability, namely Low Defects as much as 517 data and High Defects as much as 2,723. Random Under-Sampling (RUS) calculates the difference between the majority and minority classes and then repeats the calculation results; during the repetition, the majority class data is randomly deleted so that the number of majority classes is the same as the minority class [6]. This process is essential to ensure the resulting model is not biased towards the majority class. In addition, feature scaling is performed using the Min-Max Scaler method to normalize the range of feature values so that all features have the same scale.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where,

x_{new} = New normalized data

x = Data to be normalized

x_{min} = The smallest value in an attribute

x_{max} = The largest value in an attribute

MinMaxScaler aims to improve the accuracy and performance of the model. The reason for using MinMaxScaler is that one of the variables or features in this research is price, where the price cannot have a negative value.

This research is continued with the prediction stage of production defects in the manufacturing process using the K-Nearest Neighbor (KNN) algorithm. The KNN algorithm is a

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

nonparametric classification method that is simple and effective in various cases. However, the performance of this algorithm is highly dependent on the selection of the K parameter (number of nearest neighbors) [7]. In this research, the KNN algorithm is implemented with several distance measurement methods, namely Euclidean distance, Minkowski distance, and Manhattan distance.

Euclidean distance is one of the most common metrics used to measure the distance between two points in Euclidean space, where this distance is the straight-line distance between two points. In the K-Nearest Neighbor algorithm context, the Euclidean distance is often used to measure the distance between data points. When using the K-Nearest Neighbor algorithm for classification cases, it needs to find the k closest data points to a given query point and then make predictions based on the labels or values of the neighboring points. The Euclidean distance is one way to measure how close each point is to the query point. The smaller the Euclidean distance, the more similar the two attributes are. If we have two points in two-dimensional space, P1 (x1, y1) and P2 (x2, y2), then the Euclidean distance between these points is calculated using the Pythagorean theorem [8, 9]. Euclidean distance is defined by:

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

Where,

d = Distance between x and y

x = x-point

y = y-point

n = Number of criteria used

x_i = The i criterion value of the first data compared

y_i = The i criterion value of the second data compared

Minkowski distance is a distance metric used to determine between two points in N-dimensional space. Minkowski distance is considered as a generalization of Euclidean distance and Manhattan distance. In the K-Nearest Neighbor algorithm context, the Minkowski distance is used to estimate the separation between data points [8, 10]. Minkowski distance is defined by:

$$d(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} \quad (3)$$

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

Where,

d = Distance between x and y

x = x-point

y = y-point

n = Number of criteria used

x_i = Critical value in the i dimension from points x

y_i = Critical value in the i dimension from points y

p = Parameters that can be adjusted

Manhattan distance is a distance calculation metric used to identify the most suitable case from the case base by measuring the sum of the absolute weights of the differences between the case under test and other cases in the case base [11]. Manhattan distance relates to the total horizontal and vertical distance between two points, unlike Euclidean distance, which determines the shortest distance between two points. Manhattan distance can calculate distances more accurately in cases where the points to be measured have the same or nearly the exact coordinates on their axes. Manhattan distance can be beneficial when working with high-dimensional data or features with varying scales [8, 9]. Manhattan distance is defined by:

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i| \quad (4)$$

Where,

d = Distance between x and y

x = x-point

y = y-point

n = Number of criteria used

x_i = The i criterion value of the first data compared

y_i = The i criterion value of the second data compared

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

The KNN algorithm will be trained by calculating the distance of test data and training data using various distance models to determine the closest distance and assign a new class to the test data based on the K value of the nearest neighbor. Parameter optimization is carried out through a series of tests with a range of K values between 2 and 14. The final stage in this research is evaluation. The evaluation metric used is accuracy. Accuracy is one of the evaluation metrics used to measure the performance of the classification model. This accuracy describes how accurate the model is in identifying a case correctly from all cases in the dataset. Accuracy is calculated using the following formula [11].

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} x\ 100\% \quad (5)$$

Where,

TP = True Positive (Number of correctly predicted positive data)

TN = True Negative (Number of correctly predicted negative data)

FN = False Negative (Number of positive data incorrectly predicted as negative)

FP = False Positive (Number of negative data incorrectly predicted as positive)

This formula shows the ratio between the number of correct predictions and the total number of cases in the dataset, giving a picture of the model's overall classification performance.

3. Results and Discussion

In this research, the modeling process is carried out to predict the level of defects in the manufacturing process. The level of defects analyzed is divided into two categories: Low Defect (0) and High Defect (1). Before modeling is done using the K-Nearest Neighbor (KNN) algorithm, data is preprocessed, which aims to optimize the quality of the dataset used so that the KNN algorithm can provide more accurate predictions. Modeling and analysis are performed using the Google Collaboratory platform, where the KNN algorithm is applied through several critical stages, including data splitting and feature normalization (feature scaling). This research focuses on scenarios based on different distance measurement methods in the KNN algorithm to find the conditions that can produce the highest accuracy. The two main factors varied in this research are the data splitting ratio (90:10, 80:20, 75:25, 70:30, and 60:40) and the random state (42, 32, and 24). The combination of these factors or conditions resulted in 25 research scenarios, each of which aimed to identify the optimal K value and the highest level of accuracy. From this total of 25 scenarios, each distance measurement method in the KNN algorithm, namely Euclidean, Minkowski, and Manhattan distances, was analyzed in 15 scenarios. By conducting an in-depth analysis of the various scenarios, this research aims to find the most suitable distance measurement method and K parameters for predicting defects in manufacturing

DOI: 10.52362/ijiems.v3i2.1599

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

processes and identify the scenarios that provide the best performance in terms of model accuracy. Therefore, in the following sections, the results of each scenario will be described in detail to evaluate how each distance method contributes to the overall KNN algorithm.

3.1. Results on the Euclidean Distance

The following results were obtained using the Euclidean Distance in the K-Nearest Neighbor algorithm with various conditions to predict production defects.

Table 2. Results on the Euclidean Distance

Research Scenario	Random State	Splitting Data		Best K Value	Accuracy
		Train Data	Test Data		
1	42	90%	10%	13	83,95%
2		80%	20%	13	84,25%
3		75%	25%	13	85,30%
4		70%	30%	11	84,77%
5		60%	40%	13	84,56%
6	32	90%	10%	11	85,18%
7		80%	20%	11	86,26%
8		75%	25%	11	86,04%
9		2470%	30%	13	85,80%
10		60%	40%	13	84,41%
11	24	90%	10%	7	82,40%
12		80%	20%	11	83,64%
13		75%	25%	13	83,95%
14		70%	30%	7	84,56%
15		60%	40%	13	84,41%

The research conducted a comprehensive analysis of the K-Nearest Neighbor (KNN) algorithm using the Euclidean distance metric across 15 research scenarios. These scenarios explored various combinations of training and test data proportions, along with different random state values, to identify the optimal conditions for achieving the highest predictive accuracy in assessing the status or level of disability in the manufacturing process. In the first set of five scenarios, a random state value of 42 was used, with the training data proportions ranging from 90% to 60%. The best K value consistently appeared as 13, except in the scenario with 70% training data, where the optimal K value was 11. The accuracy varied slightly across these scenarios, with the highest being 85.30% at a 75% training data split. The second set of scenarios replicated the data splits but used a random state value of 32. Here, the optimal K value remained mostly at 13, with one instance of 11 at the 70% training split. The accuracy increased in this set, with the highest observed at 86.26% when 80% of the data was used for training. In the final set of scenarios, the random state value was adjusted to 24. The results mirrored the earlier patterns,

DOI: 10.52362/ijiems.v3i2.1599

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

with the optimal K value again at 13, except for the 70% training data scenario, which favored K = 11. The accuracy was slightly lower in this set, with the highest being 84.56% for the 70% training data split. After analyzing all 15 scenarios, the research identified the optimal conditions for maximizing accuracy in predicting manufacturing process defects. The highest accuracy, 86.26%, was achieved with a training-to-test data ratio of 80:20 and a random state value of 32, where the best K value was 11. This finding suggests that these conditions are most effective for implementing the KNN algorithm with Euclidean distance in this specific predictive modeling context.

3.2. Results on the Minkowski Distance

The following results were obtained using the Minkowski Distance in the K-Nearest Neighbor algorithm with various conditions to predict production defects.

Table 3. Results on the Minkowski Distance

Research Scenario	Random State	Splitting Data		Best K Value	Accuracy
		Train Data	Test Data		
1	42	90%	10%	11	83,95%
2		80%	20%	11	84,25%
3		75%	25%	10	85,43%
4		70%	30%	10	84,77%
5		60%	40%	11	84,56%
6	32	90%	10%	11	85,18%
7		80%	20%	10	86,41%
8		75%	25%	13	86,29%
9		70%	30%	13	85,80%
10		60%	40%	13	84,56%
11	24	90%	10%	7	82,40%
12		80%	20%	11	83,64%
13		75%	25%	9	83,95%
14		70%	30%	10	84,67%
15		60%	40%	7	84,56%

The research conducted a comprehensive analysis of the K-Nearest Neighbor (KNN) algorithm using the Minkowski distance metric across 15 research scenarios. These scenarios explored various combinations of training and test data proportions, along with different random state values, to identify the optimal conditions for achieving the highest predictive accuracy in assessing the status or level of disability in the manufacturing process. In the first set of scenarios, with a random state value of 42, the training data proportions ranged from 90% to 60%. The optimal K values ranged between 10 and 11, with the highest accuracy of 85.43%

DOI: 10.52362/ijiems.v3i2.1599

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

achieved at a 75% training data split and a K value of 10. The second set of scenarios applied a random state value of 32. Here, the optimal K values varied between 10 and 13. The highest accuracy observed was 86.41%, achieved when 80% of the data was used for training and the best K value was 10. In the final set of scenarios, with a random state value of 24, the optimal K values varied more widely, ranging from 7 to 11. The highest accuracy in this set was 84.67%, achieved with 70% of the data allocated to training and a K value of 10. After analyzing all 15 scenarios, the research identified that the optimal conditions for maximizing accuracy were an 80:20 training-to-test data ratio, a random state value of 32, and a K value of 10. Under these conditions, the highest accuracy of 86.41% was achieved, indicating these settings are most effective for implementing the KNN algorithm with Minkowski distance in this specific predictive modeling context.

3.3. Results on the Manhattan Distance

The following results were obtained using the Manhattan Distance in the K-Nearest Neighbor algorithm with various conditions to predict production defects.

Table 4. Results on the Manhattan Distance

Research Scenario	Random State	Splitting Data		Best K Value	Accuracy
		Train Data	Test Data		
1	42	90%	10%	13	83,95%
2		80%	20%	13	84,25%
3		75%	25%	12	85,30%
4		70%	30%	10	84,77%
5		60%	40%	11	84,41%
6	32	90%	10%	11	85,18%
7		80%	20%	13	86,26%
8		75%	25%	13	86,29%
9		70%	30%	14	85,69%
10		60%	40%	13	84,49%
11	24	90%	10%	12	82,40%
12		80%	20%	12	83,64%
13		75%	25%	12	83,82%
14		70%	30%	11	84,46%
15		60%	40%	12	84,49%

The research conducted a comprehensive analysis of the K-Nearest Neighbor (KNN) algorithm using the Manhattan distance metric across 15 research scenarios. These scenarios explored various combinations of training and test data proportions, along with different random state values, to identify the optimal conditions for achieving the highest predictive accuracy in

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

assessing the status or level of disability in the manufacturing process. In the first set of scenarios, using a random state value of 42, the training data proportions ranged from 90% to 60%. The optimal K values varied from 10 to 13, with the highest accuracy of 85.30% achieved at a 75% training data split and a K value of 12. The second set of scenarios applied a random state value of 32. In this set, the optimal K values ranged from 11 to 14. The highest accuracy observed was 86.29%, achieved with a 75% training data split and a K value of 13. In the final set of scenarios, with a random state value of 24, the optimal K values varied between 11 and 12. The highest accuracy in this set was 84.49%, observed in both the 60% and 40% training data scenarios with a K value of 12. After analyzing all 15 scenarios, the research identified that the optimal conditions for maximizing accuracy were a 75:25 training-to-test data ratio, a random state value of 32, and a K value of 13. Under these conditions, the highest accuracy of 86.29% was achieved, indicating that these settings are most effective for implementing the KNN algorithm with Manhattan distance in this specific predictive modeling context.

3.4. Comparison of Best Results of Each Distance

After determining the optimal combination for the distribution of training data, test data, and random state that produces the highest accuracy value for each distance K-Nearest Neighbor algorithm used, the next step is to compare and select the K-Nearest Neighbor algorithm with several conditions that can provide the best accuracy value for each selected distance measurement method.

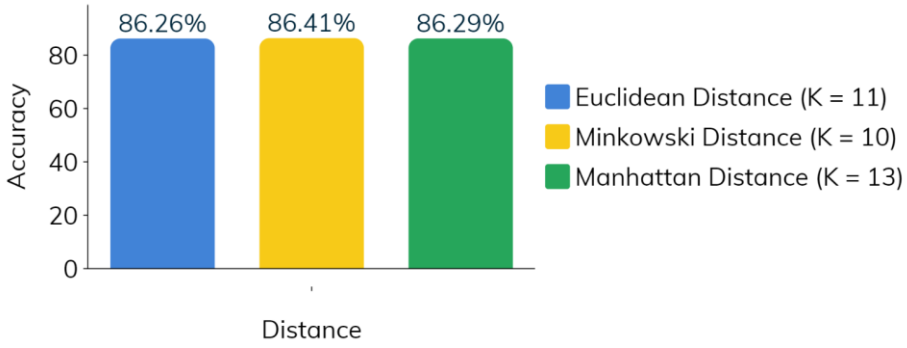

Figure 2. Comparison of the Highest Accuracy

As shown in Figure 2, the K-Nearest Neighbor algorithm's highest accuracy value at each distance is different. The Euclidean distance has a lower accuracy value than both the Minkowski distance and the Manhattan distance (Euclidean < Minkowski and Manhattan), the Minkowski distance has a higher accuracy value than both the Euclidean distance and the Manhattan distance (Minkowski > Euclidean and Manhattan), and the Manhattan distance has a higher accuracy value than both the Euclidean distance and the Manhattan distance (Euclidean < Manhattan < Minkowski). So, the K-Nearest Neighbor algorithm with Minkowski distance,

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

training data split up to 80%, test data split up to 20%, and a random state value of 32 with the best K value of 10 gives the best accuracy. The accuracy value obtained in the best K-Nearest Neighbor algorithm is 86.41%, indicating that the K-Nearest Neighbor algorithm has good accuracy in predicting defects in the manufacturing process.

4. Conclusion

Based on the test results and discussion in this research, it can be concluded that the use of distance models with several conditions, such as data division ratio and random state value, can affect the accuracy of the K-Nearest Neighbor algorithm in predicting the level of production defects in the manufacturing process. Of the three distances that have been tested, the type of distance in the K-Nearest Neighbor algorithm that can produce a high accuracy value is the Minkowski distance with the condition of splitting 80% training data, 20% test data, and a random state value of 32. Under these conditions, the Minkowski distance gets the best K value of 10, which can produce an accuracy value of 86.41%.

Acknowledgements

We wish to convey our profound appreciation to all individuals who have contributed, provided support, and assisted in achieving the successful culmination of our research.

References

[1] Y. Novita and R. Zahra, "Penerapan Artificial Intelligence (AI) Untuk Meningkatkan Efisiensi Operasional di Perusahaan Manufaktur: Studi Kasus PT. XYZ," in Jurnal Manajemen dan Teknologi, Vol. 1, No. 1, 2024, pp. 11-21.

[2] M. Rifaldi and W. Sudarwati, "Penerapan Metode Six Sigma dan FMEA Sebagai Usaha Untuk Mengurangi Cacat Pada Produk Bracket," in Seminar Nasional Sains dan Teknologi 2024, Fakultas Teknik Universitas Muhammadiyah Jakarta, (2024) April 30.

[3] T. Arifin, W. Sanjaya, I. M. M. Shahih, and E. Sopiah, "Pemanfaatan Algoritma Restricted Boltzmann Machines (RBM) Untuk Prediksi Dini Kanker Paru-Paru," in Jurnal Responsif, Vol. 5, No. 2, 2023, pp. 138-146.

[4] Iswanto, Tulus, and P. Sihombing, "Comparison of Distance Models on K-Nearest Neighbor Algorithm in Stroke Disease Detection," in Applied Technology and Computing Science Journal, Vol. 4, No. 1, 2021, pp. 63-68.

[5] Putri, C. S. Hardiana, E. Novfuja, F. T. P. Siregar, Rahmaddeni, Y. Fatma, and R. Wahyuni, "Komparasi Algoritma K-NN, Naïve Bayes dan SVM Untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," in Malcom: Indonesian Journal of Machine Learning and Computer Science, Vol. 3, Issue. 1, 2023, pp. 20-26.

[6] E. Irawan and R. S. Wahono, "Penggunaan Random Under Sampling Untuk Penanganan Ketidakseimbangan Kelas Pada Prediksi Cacat Software Berbasis Neural Network," in Journal of Software Engineering, Vol. 1, No. 2, 2015, pp. 92-100.

DOI: 10.52362/ijiems.v3i2.1599

International Journal of Informatics, Economics, Management and Science
http://journal.stmikjayakarta.ac.id/index.php/ijiems
E-ISSN: 2809-8471 (online), **P-ISSN:** 2809-9281 (Print)
DOI: 10.52362/ijiems.v3i2.1599
Volume 3 , Issue 2, August 2024, pp. 161-173

[7] G. H. Martono and N. Sulistianingsih, "Perbandingan Matriks Jarak Pada Algoritma K-NN Untuk Prediksi Penyakit Diabetes," in JoMI: Journal of Millennial Informatics, Vol. 2, No. 1, 2024, pp. 1-6.

[8] M. M. Abualhaj, A. A. Abu-Shareha, Q. Y. Shambour, A. Alsaaidah, S. N. Al-Khatib, and M. Anbar, "Customized K-Nearest Neighbors' Algorithm for Malware Detection," in International Journal of Data and Network Science., Vol. 8, Issue. 1, 2024, pp. 431-438.

[9] Salsabila, S. Martha, and W. Andani, "Komparasi Algoritma K-Nearest Neighbor Dengan Euclidean Distance dan Manhattan Distance Untuk Klasifikasi Stunting Balita (Studi Kasus : Puskesmas Kelurahan Parit Mayor)," in Buletin Ilmiah Matematika, Statistika dan Terapannya (Bimaster), Vol. 13, No. 2, 2024, pp. 285-292.

[10] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering Berbasis Chi-Square," in Jurnal Informatika: Jurnal Pengembangan IT (JPIT), Vol. 4, No. 1, 2019, pp. 20-24.

[11] R. Lina and D. C. Wati, "Klasifikasi Pengeluaran per Kapita di Tiga Provinsi Sulawesi Menggunakan K-Nearest Neighbor," in J Statistika, No. 16, No. 1, 2023, pp. 395-406.

[12] Sudriyanto, F. Syahro, and N. Fitriani, "Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, dan K-Nearest Neighbors Dalam Prediksi Harga Saham," in JARS: Journal of Advanced Research in Informatics, Vol. 2, No. 1, 2023, pp. 1-9.

[13] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity Dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," in Sainteks, Vol. 19, No. 2, 2022, pp. 141-155.

DOI: 10.52362/ijiems.v3i2.1599