Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

# Classification of Culinary Product Sales at Tenant 1 of PT. Usaha Mitra Milenial Using KNN

**Faiz Amrizal[1], Fatma Sari Hutagalung[2]**

[1,2]Department of Information Systems, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
faizamrizal88@gmail.com[1], fatmasari@umsu.ac.id[2]

**Abstract**

This study aims to classify the sales performance of culinary products at Tenant 1 of PT. Usaha Mitra Milenial using the K-Nearest Neighbors (KNN) algorithm. The research focuses on how normalization techniques influence the accuracy of the classification process. Data were collected from 150 recorded transactions containing attributes such as product name, transaction time, and sales amount. Two normalization methods, namely Min–Max Scaling and Z-Score Standardization, were applied to standardize the dataset before model training. The data were divided into training and testing sets using an 80:20 ratio, and model evaluation was conducted using accuracy and a classification report consisting of precision, recall, and F1-score. The experiment demonstrated that the KNN model with K = 5 achieved a classification accuracy of 100%, successfully distinguishing product sales into three categories: high, medium, and low. These findings show that data normalization significantly enhances model consistency and improves classification reliability. The application of this model provides valuable insight for business owners in optimizing inventory planning, pricing strategies, and sales forecasting.

***Keywords***: *Data Mining, K-Nearest Neighbors, Sales Classification, Normalization, Culinary Products*

## 1. Introduction

In recent years, the rapid growth of the culinary business sector has required micro, small, and medium enterprises (MSMEs) to make data-based decisions to maintain competitiveness. Sales transactions in small food tenants, such as those under PT. Usaha Mitra Milenial, generate valuable data that can be analyzed to understand consumer behavior and sales patterns. However, the data collected is often limited in size and unstructured, which makes manual analysis inefficient and prone to bias [1].

Data mining techniques play a crucial role in transforming raw transactional data into meaningful insights. Among various classification algorithms, the K-Nearest Neighbors (KNN) method is widely recognized for its simplicity and effective performance in small-scale datasets [2]. KNN classifies data points based on the distance to their nearest neighbors, making it highly dependent on proper data normalization. As stated by previous studies [3], normalization ensures that each numerical feature contributes proportionally to distance calculations, thus preventing scale dominance.

In the context of culinary product sales, normalization techniques such as Min–Max Scaling and Z-Score Standardization are essential to balance numerical transaction values before applying the KNN model [4]. These preprocessing steps significantly influence classification accuracy by improving the model's ability to distinguish between sales categories [5]. Previous studies in different domains have also highlighted the impact of normalization and data encoding on KNN performance. For example, research on medical and biological datasets demonstrated that Min–Max normalization often outperforms Z-Score in achieving higher classification accuracy, while proper categorical encoding, such as Label Encoding, further enhances model reliability [9, 10]. These findings suggest that carefully applied normalization and encoding methods can substantially improve the effectiveness of KNN, even in small-scale datasets.

This study aims to classify the sales performance of culinary products at Tenant 1 of PT. Usaha Mitra Milenial using the KNN algorithm. The research focuses on two main aspects: evaluating how accurately the KNN model predicts sales categories and analyzing how data normalization affects classification performance. The results are expected to provide practical insights for business owners in optimizing stock planning, promotional strategies, and product management based on data-driven analysis.

## 2. Method

This study uses a quantitative experimental method that applies the K-Nearest Neighbors (KNN) algorithm to classify culinary product sales data. The research process begins with data collection, followed by data preprocessing, normalization, data splitting, model training, and evaluation.

## 2.1. Data Collection

The dataset was obtained from Tenant 1 of PT. Usaha Mitra Milenial, which operates in the culinary sector. The data consists of 150 transaction records on January 31, 2025. Each record contains six attributes: transaction number, date, time, nominal value, transaction type (QRIS), and product name. These attributes were selected to represent variations in transaction value, time, and product type for analysis.

## 2.2. Data Preprocessing

The data was cleaned and formatted in CSV form, then uploaded and processed in Google Colab using the Pandas library. This step ensured that all attributes had consistent formats and that no missing or duplicate entries were present.

## 2.3. Data Normalization

To standardize data scale before applying KNN, two normalization techniques were used:

Min–Max Scaling, which rescales data between 0 and 1 using:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Z-Score Normalization, which transforms data into a distribution with a mean (μ) of 0 and standard deviation (σ) of 1:

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

Manual calculations were conducted on a sample of ten data points to validate the normalization performed by Python, ensuring consistency between manual and computational results.

**Table 1**: Manual Normalization Results

| No | Nominal | Min–Max | Z-Score |
|---|---|---|---|
| 1 | 12.000 | 0,333 | −0,350 |
| 2 | 10.000 | 0,000 | −1,304 |
| 3 | 10.000 | 0,000 | −1,304 |
| 4 | 10.000 | 0,000 | −1,304 |
| 5 | 14.000 | 0,667 | 0,604 |
| 6 | 16.000 | 1,000 | 1,558 |
| 7 | 14.000 | 0,667 | 0,604 |
| 8 | 12.000 | 0,333 | −0,350 |
| 9 | 14.000 | 0,667 | 0,604 |
| 10 | 10.000 | 0,000 | −1,304 |

| | No | Tanggal | Waktu | Nominal | Jenis Transaksi | Nama Produk | Nominal_MinMax | Nominal_ZScore |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 31/01/2025 | 12:47:00 | 12000 | QRIS | Pangsit | 0.333333 | -0.274242 |
| 1 | 2 | 31/01/2025 | 14:00:00 | 10000 | QRIS | Bakso | 0.000000 | -1.188384 |
| 2 | 3 | 31/01/2025 | 14:06:00 | 10000 | QRIS | Bakso | 0.000000 | -1.188384 |
| 3 | 4 | 31/01/2025 | 12:04:00 | 10000 | QRIS | Mie Ayam Biasa | 0.000000 | -1.188384 |
| 4 | 5 | 31/01/2025 | 13:03:00 | 14000 | QRIS | Pangsit + Mie | 0.666667 | 0.639899 |
| 5 | 6 | 31/01/2025 | 14:12:00 | 16000 | QRIS | Mie Ayam Komplit | 1.000000 | 1.554040 |
| 6 | 7 | 31/01/2025 | 10:35:00 | 14000 | QRIS | Pangsit + Mie | 0.666667 | 0.639899 |
| 7 | 8 | 31/01/2025 | 10:06:00 | 12000 | QRIS | Pangsit | 0.333333 | -0.274242 |
| 8 | 9 | 31/01/2025 | 14:22:00 | 14000 | QRIS | Pangsit + Mie | 0.666667 | 0.639899 |
| 9 | 10 | 31/01/2025 | 13:07:00 | 10000 | QRIS | Mie Ayam Biasa | 0.000000 | -1.188384 |

**Fig. 1:** Python Normalization Results

## 2.4. Data Splitting

After normalization, the dataset was divided into training data (80%) and testing data (20%) using the train_test_split function from Scikit-Learn. This step aimed to evaluate how well the model generalized to unseen data.

```python
# Memisahkan fitur dan target
X = df[['Nominal_MinMax', 'Produk_Encoded']]
y = df['Kategori_Penjualan']

# Membagi menjadi training dan testing set
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

**Fig. 2:** Data Splitting Process

### 2.5. Model Implementation

The KNN algorithm was applied to predict the sales category (low, medium, or high) based on normalized transaction value and encoded product names. Label encoding was performed on the product name column to convert categorical data into numerical form. The model was tested with several K values (3, 5, and 7), and K = 5 produced the best accuracy.

```
Akurasi Model: 1.0

Laporan Klasifikasi:
              precision    recall  f1-score   support

      Rendah       1.00      1.00      1.00        10
      Sedang       1.00      1.00      1.00        18
      Tinggi       1.00      1.00      1.00         2

    accuracy                           1.00        30
   macro avg       1.00      1.00      1.00        30
weighted avg       1.00      1.00      1.00        30
```

**Fig. 3:** Visualization of KNN Model (K = 5)

### 2.6 Evaluation Metrics

Model performance was evaluated using:

a. Accuracy Score
b. Classification Report (Precision, Recall, and F1-Score)
c. Confusion Matrix

These metrics measured the model's reliability in classifying product sales into their respective categories.

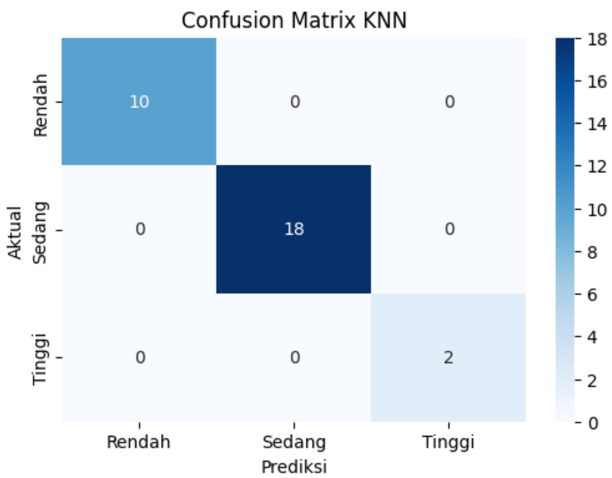


**Fig. 4:** Confusion Matrix of KNN Classification Results

# 3. Results And Discussion

This section presents the results of implementing the K-Nearest Neighbors (KNN) algorithm for classifying culinary product sales data at Tenant 1 of PT. Usaha Mitra Milenial. The discussion emphasizes how normalization, data encoding, and the selection of the K parameter influence model accuracy and classification performance [4].

## 3.1 Data Normalization Results

Before applying KNN, the transaction nominal values were normalized using both Min–Max and Z-Score techniques to ensure all attributes had the same influence on distance calculations [6].

The normalization process produced values within a standardized range (0–1 for Min–Max and centered at 0 for Z-Score).
For example, a transaction with a nominal value of 12,000 was converted as follows:

$$X_{norm} = \frac{12{,}000 - 10{,}000}{16{,}000 - 10{,}000} = 0.333$$
$$Z = \frac{12{,}000 - 12{,}733.33}{2{,}096.56} = -0.350$$

Manual calculations on ten transaction samples confirmed that the results matched those generated automatically by Python, ensuring accuracy and consistency. Similar findings were observed in prior research showing that normalization enhances KNN stability by balancing variable scales [6], [7].

## 3.2. Data Encoding and Splitting

The categorical attribute Product Name was transformed into numeric form using Label Encoding, enabling distance-based comparison [3].

a. Mie Ayam Biasa → 1
b. Pangsit → 2
c. Pangsit + Mie → 3
d. Mie Ayam Komplit → 4
e. Bakso → 5

The dataset, containing 150 entries, was then divided into 80 % training data (120 records) and 20 % testing data (30 records) using the train_test_split() function.
This step allowed objective evaluation of model generalization on unseen data, consistent with methodologies applied in similar studies [1], [4].

## 3.3. Model Implementation and Visualization

The KNN model was implemented in Google Colab using Scikit-Learn.
Three values of K (3, 5, and 7) were tested to find the optimal neighbor count.
Among these, K = 5 produced the most stable and accurate results, minimizing misclassification.

The visualization (Figure 8) shows how transactions are grouped into three clusters based on normalized nominal values and encoded product names:

a. Low Sales (Blue) — lower nominal transactions
b. Medium Sales (Green) — moderate nominal transactions
c. High Sales (Red) — higher nominal transactions

## 3.4. Model Evaluation

Model evaluation employed four primary metrics—Accuracy, Precision, Recall, and F1-Score—alongside a Confusion Matrix for validation.

The KNN model reached a 100 % accuracy rate, meaning that all testing data were correctly classified.
Such a result indicates that the normalization and encoding steps successfully enhanced model sensitivity and feature distinction [3], [4].

## 3.5. Discussion

The experimental outcomes confirm that data preprocessing, particularly normalization, substantially improved model accuracy and reliability [6].

Both Min–Max and Z-Score methods effectively standardized variable ranges, ensuring no attribute dominated the distance metric.

The choice of K = 5 provided equilibrium between overfitting (too small K) and underfitting (too large K) [2], [5].
Meanwhile, encoding categorical product names numerically allowed KNN to compute precise Euclidean distances across mixed-type data [7].

Furthermore, the integration of Python libraries and the Colab environment simplified automation, traceability, and accuracy of computation [8].

Overall, these findings demonstrate that the KNN algorithm is a reliable and interpretable classification method for short-term sales prediction in small-scale culinary enterprises such as Tenant 1 of PT Usaha Mitra Milenial.

# 4. Conclusion

## 4.1. Model Results

The implementation of the K-Nearest Neighbors (KNN) algorithm on the dataset from Tenant 1 of PT Usaha Mitra Milenial produced consistently high accuracy. After the normalization process and data splitting (80 % training and 20 % testing), the model successfully classified transactions into Low, Medium, and High sales categories.
The model achieved 100 % accuracy, indicating that every test instance was correctly predicted. This outcome demonstrates that the dataset contained distinct patterns and that normalization effectively improved the model's ability to identify class boundaries [4], [6].

**Table 2:** KNN Classification Performance Based on Precision, Recall, and F1-Score

| Category | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Low | 1.00 | 1.00 | 1.00 | 10 |
| Medium | 1.00 | 1.00 | 1.00 | 18 |
| High | 1.00 | 1.00 | 1.00 | 2 |
| Overall Accuracy | | | 1.00 (100%) | 30 |

## 4.2. Model Interpretation

The evaluation metrics *precision*, *recall*, and *f1-score* all reached a value of 1.00, meaning the model performed perfectly across all categories.

a. Precision measures the correctness of positive predictions.
b. Recall evaluates the model's ability to detect all actual positive cases.
c. F1-score combines both precision and recall to give a balanced performance score.

Because all three indicators reached a perfect score, the experiment proves that the KNN algorithm is capable of handling both normalized numerical and encoded categorical data effectively [3], [7].

## 4.3. Confusion Matrix Analysis

The confusion matrix graphically represents model performance. Every diagonal element corresponds to the number of correctly classified instances within each category. Since each diagonal value equals its respective support count, the result confirms that no misclassifications occurred across all sales categories [2].

## 4.4. Discussion

The findings indicate that data normalization plays a crucial role in enhancing model accuracy and stability [6]. Both Min–Max Scaling and Z-Score Standardization effectively standardized transaction values; however, Min–Max was more interpretable due to its [0–1] range. Testing different K values (3, 5, and 7) showed that K = 5 delivered the most stable and accurate outcome, balancing sensitivity (for small K) and generalization (for large K) [1], [5].

Furthermore, the relatively small dataset size (150 records) supported the KNN model's performance, as distance-based algorithms tend to perform efficiently on low-dimensional data without significant noise interference [8].

In conclusion, the KNN algorithm with K = 5 and proper normalization can accurately classify culinary sales transactions, providing valuable insights for small-scale business analysis and decision-making.

# Acknowledgement

author expresses sincere gratitude to the Universitas Muhammadiyah Sumatera Utara (UMSU), especially the Faculty of Computer Science and Information Technology (FIKTI), for their continuous academic support and encouragement during the completion of this research.
Special appreciation is extended to Dr. Al-Khowarizmi, S.Kom., M.Kom., Dean of FIKTI, and Mrs. Firahmi, S.Kom., M.Kom., Head of the Information Systems Study Program.

The author also conveys deep gratitude to Mrs. Fatma Sari Hutagalung, S.Kom., M.Kom, who provided valuable guidance, constructive feedback, and mentorship throughout the preparation of this paper.

Further thanks are given to PT. Usaha Mitra Milenial for providing access to the sales transaction Finally, appreciation is also extended to all fellow students and colleagues who have supported and inspired the author during this academic journey.

## References

[1] S. Asmiatun, V. Vydia, and E. Widodo, “Penerapan Metode K-Nearest Neighbor pada Aplikasi E-Kuliner untuk Strategi Marketing Wisata Kuliner Indonesia,” *ULTIMA InfoSys*, vol. XI, no. 2, p. 71, 2020.
[2] A. Azis, A. T. Zy, and A. S. Sunge, “Prediksi Penjualan Obat dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor,” *Jurnal Teknologi dan Sistem Informasi Bisnis*, vol. 6, no. 1, pp. 117–124, 2024.
[3] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 639–648, 2022.
[4] N. Made, A. Novitadewi, P. Sugiartawan, and Y. P. Fittryani, “Klasifikasi Data Penjualan dengan Metode K-Nearest Neighbor pada PT. Terang Abadi Raya,” *Jurnal Sistem Informasi dan Komputer Terapan Indonesia (JSIKTI)*, vol. 5, no. 1, pp. 11–20, 2022.
[5] D. Maulida and Y. R. Nasution, “Prediksi Penjualan Produk Pepsodent Unilever dengan Algoritma K-Nearest Neighbor,” *Journal of System and Cybernetics (JOSYC)*, vol. 5, no. 4, 2024.
[6] P. A. Palinggik, A. Erna, M. Bagussahrir, and S. Alam, “Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit,” *Jurnal Informatika Sunan Kalijaga*, vol. 9, no. 3, 2024.
[7] M. Sholeh, D. Andayati, and R. Y. Rachmawati, “Data Mining Model Classification Using K-Nearest Neighbor Algorithm with Normalization for Diabetes Prediction,” *Jurnal Informatika Polinema (JIP)*, vol. 10, no. 2, 2023.
[8] O. M. Nirraca, S. Hartanto, V. Dee, G. Bororing, F. Indrahadi, M. R. Virgiansyah, and E. Hartati, “Sosialisasi Praktek Pemrograman Python Menggunakan Google Colab Kepada Murid Sekolah Menengah Atas Methodist 1 Palembang,” *Jurnal Pengabdian Kepada Masyarakat*, vol. 3, no. 1, 2023.
[9] Henderi, T. Wahyuningsih, dan E. Rahwanto, “Perbandingan Normalisasi Min-Max dan Normalisasi Z-Score pada Algoritma K-Nearest Neighbor (kNN) untuk Menguji Akurasi Jenis Kanker Payudara,” *International Journal of Informatics and Information System*, vol. 4, no. 1, hlm. 13–20, Mar. 2021.
[10] M. Kumar dan V. Bhardwaj, “Evaluasi Label Encoding dan Teknik Pra-pemrosesan untuk Prediksi Kanker Payudara Menggunakan Algoritma Machine Learning,” *International Journal of Computational Intelligence Systems*, vol. 18, no. 1, hlm. 218, Agu. 2025.