Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Flood Risk Prediction Using the K-Nearest Neighbors (KNN) Method with Machine Learning Optimization in Jakarta Province

Arnoldus Raja Ino[1], Andi Taufik[2]
Universitas Bina Sarana Informatika, Jakarta. Jl. Kramat Raya No.98, Jakarta Pusat, Indonesia

Article Info
Keywords:
Jakarta,
K-Nearest Neighbors, Machine Learning,
Flood Prediction.

ABSTRACT
This study aims to predict flood risk in the Jakarta Province using the K-Nearest Neighbors (KNN) algorithm. The model is designed to classify regions into three risk levels—low, moderate, and high—based on historical features such as water level and annual flood frequency. Data from 2013 to 2024 provided by the Jakarta Regional Disaster Management Agency (BPBD) served as the basis for model training. The normalization process using Z-Score and Min-Max Scaling proved essential in enhancing KNN performance. Cross-validation (5-Fold Cross Validation) revealed that K=3 yielded the best results. The model achieved accuracy rates between 96% and 99%, with precision, recall, and F1-score exceeding 95%. These results demonstrate that KNN can reliably map flood risks based on numerical data trends. Predictions for the years 2026 to 2028 indicate an increased flood risk in several areas, particularly Central, North, and West Jakarta, which fall into the high-risk category. East and South Jakarta are classified as moderate to high risk, while the Thousand Islands region is expected to remain in the low-risk category. This model can be used as a foundation for developing data-driven mitigation strategies and early warning systems.

Corresponding Author:
Arnoldus Raja Ino
Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, Jakarta Pusat, Indonesia
tarsisiusarnoldus@gmail.com

INTRODUCTION

Indonesia, as a tropical country, experiences only two seasons: the rainy season and the dry season. During the rainy season, high rainfall intensity often leads to flooding disasters, especially in urban areas such as Jakarta, which is characterized by a dense population and a complex drainage system. Flooding has become an annual problem that significantly impacts infrastructure, economic activities, and the social lives of communities. One of the primary indicators of flood events is the sudden rise in river water levels during heavy rainfall. Therefore, an accurate prediction system is essential to detect potential flooding early (Sandiwarno, 2024).

Several studies have shown that floods are influenced by various factors such as rainfall intensity, soil conditions, and the accumulation of waste in waterways (Restu, Siswa, & Pranoto, 2024). Advances in information technology have enabled new approaches to flood prediction using machine learning algorithms. The K-Nearest Neighbors (KNN) method is recognized as a simple yet effective classification algorithm that can process historical data—

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

such as water level and flood events—to predict the likelihood of future disasters. KNN is capable of identifying patterns in past data and classifying current conditions based on similarity (Roihan, Sunarya, & Rafika, 2020).

However, implementing KNN in institutional settings such as the Jakarta Regional Disaster Management Agency (BPBD DKI Jakarta) presents challenges, including limitations in data quality and integration into existing early warning systems (Rafi Nahjan, Heryana, & Voutama, 2023). Moreover, few studies have specifically examined river water levels and historical flood data as primary variables, or how such models can be applied within the framework of disaster mitigation policies (Angreni & Pusadan, 2024). A literature-based approach is therefore needed to develop a prediction model that is better aligned with local conditions and institutional needs (Angreni & Pusadan, 2024).

This study aims to analyze the potential application of the K-Nearest Neighbors (KNN) method in constructing a flood risk classification prediction model for the Jakarta area. The main focus lies in utilizing water level data and historical flood incidents as key parameters. Through a literature review approach, this study is expected to offer both theoretical and practical contributions to the development of a responsive, accurate, and locally grounded flood prediction system, which can also serve as a basis for urban disaster mitigation policies.

METHODS

This study employed a quantitative predictive approach using the K-Nearest Neighbors (KNN) classification algorithm to model flood risk levels in the Jakarta Province. The model was constructed based on two primary variables: the water level (in centimeters) and the annual frequency of flood incidents per administrative unit (RW). The overall research procedure involved several critical stages, from data collection and preprocessing to model evaluation and prediction.

Data Collection and Preprocessing

The dataset used in this study was obtained from the Jakarta Regional Disaster Management Agency (BPBD DKI Jakarta), encompassing the period between 2013 and 2024. The data included detailed flood records across various RWs, such as water depth, frequency of flood events per year, and location identifiers. Missing values were removed, and inconsistent formats, such as water levels presented in ranges, were standardized by averaging them into single numeric values.

Normalization Process

To ensure consistency and comparability among features, the numerical data were normalized using two techniques: Z-Score and Min-Max Scaling. These methods were implemented to improve distance calculations within the KNN algorithm by rescaling the data and reducing feature dominance. For Z-Score normalization, the mean and standard deviation of each variable were used to produce standardized scores. In Min-Max Scaling, each feature value was transformed to a range between 0 and 1.

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Data Partitioning

The dataset was split into training and testing subsets, with 80% of the data used for model training and the remaining 20% for testing. This was done to prevent overfitting and to evaluate the model's generalization capabilities.

KNN Modeling and Cross-Validation

The KNN model was implemented using the Python programming language. To determine the optimal number of neighbors (K), the model was tested across several K values (1, 3, 5, and 7), and performance was assessed using 5-Fold Cross-Validation. The best K value was selected based on the highest average accuracy achieved during validation.

Model Evaluation

The classification performance of the KNN model was measured using standard evaluation metrics: accuracy, precision, recall, and F1-score. A confusion matrix was also generated for each administrative region to visualize classification outcomes. These metrics were essential to ensure the robustness and reliability of the model in classifying flood risks.

Prediction for Future Periods

Finally, the validated model was used to predict flood risk categories for the years 2026 through 2028. The results were interpreted to identify areas with increasing flood risks and to provide recommendations for flood mitigation planning.

RESULTS AND DISCUSSION

This study aims to develop a model capable of predicting flood risk using the K-Nearest Neighbors (KNN) algorithm, based on historical data on water levels and flood frequency in Jakarta Province, including the Thousand Islands, covering the period from 2013 to 2024. The model is designed to classify regions into three levels of flood risk: low, moderate, and high.

At the initial stage, the process began by training the model using historical data and identifying the most appropriate value of K for each region. The determination of the optimal K value was conducted through cross-validation, which helped identify the best parameter combinations to achieve the highest accuracy on the test data. The results indicated that most regions performed best with lower K values (between 1 and 3), suggesting that flood risk classification in Jakarta is highly influenced by local neighborhood characteristics.

Once the training phase was completed, the model was tested using unseen data (test data) to objectively evaluate its predictive capability. The testing results showed that the average model accuracy across all regions exceeded 95%, with precision, recall, and F1-score values also consistently high. These outcomes demonstrate that the developed KNN model is highly reliable in distinguishing areas with low, moderate, and high flood risk.

Determination of Optimal K Value

Cross-validation tests on the training dataset revealed that the most suitable K value varied by region. A summary of the optimal K values used in the prediction and classification model—as determined by feature selection and accuracy assumptions from cross-validation using K-fold values ranging from 1 to 3—is presented in Table 1,

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Table 1. Optimal K Value Determination for Each Region

Region	Optimal K Value	Highest Cross-Validation Accuracy (%)
Central Jakarta	3	97%
North Jakarta	1	99%
West Jakarta	1	99%
South Jakarta	1	96%
East Jakarta	1	97%
Thousand Islands	1	100%

Furthermore, the interpretation of the KNN model performance based on cross-validation results is aligned with the X and Y axes of the data categories illustrated in the graph previously presented in the research methodology section (Figure 1).

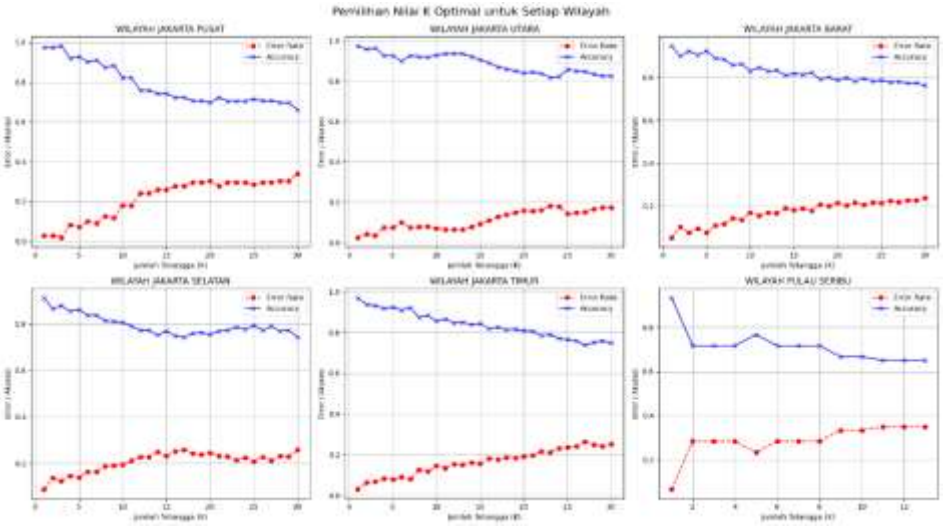


Figure 1. Selection of Value (K)

The graphs illustrate the performance of the KNN algorithm across five regions in Jakarta, with the X-axis representing the number of neighbors (K) and the Y-axis representing the error rate and accuracy. The optimal K value, which minimizes the error rate and maximizes accuracy, varies by region, ranging from approximately 15 to 30 for Central, North, West, and South Jakarta, and between 8 and 10 for the Thousand Islands. This indicates that the higher the K value applied in the KNN model, the lower the accuracy and the higher the error rate across all regions.

Flood Frequency Trend Statistics

Based on the diagram displaying the annual frequency of elevated water levels during flood events in Jakarta from 2013 to 2024, several findings reveal a fluctuating pattern in the rise and decline of flood occurrences. These trends are visualized in Figure 2

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

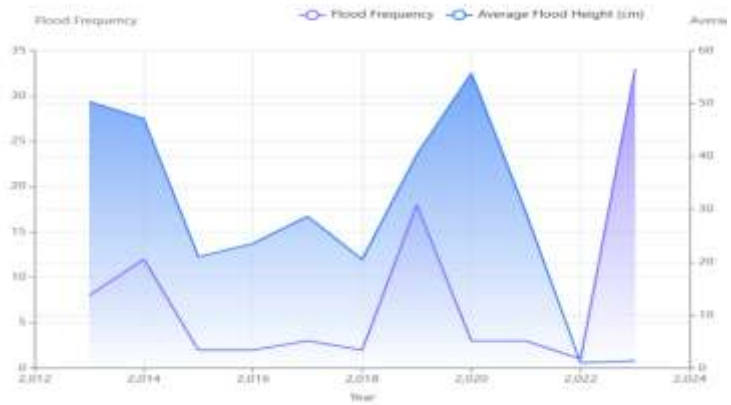


Figure 2. Graph Of Average Flood Events Per Year

The illustration in Figure 2 serves as a reference for standard deviation in the normalization process using Z-score and Min-Max Scaling, particularly in calculating the average range of accuracy scores within flood event categories. The observed deviation values include 30, 40, 50, 60, 70, and 80, with 30 aligning with the threshold for maximum cross-validation accuracy, and 40 through 80 representing the upper range of the data distribution.

Z-Score and Min-Max Scaling Normalization Results

The Z-score normalization was calculated using the following formula:

$$Z = \frac{X - \mu}{\sigma}$$

Where the data values (X) include 30, 40, 50, 60, 70, and 80; the mean (μ) is 50; and the standard deviation (σ) is 15.81. The normalization results are:

a. If X = 60: Z = (60 – 50) / 15.81 = 0.63
b. If X = 30: Z = (30 – 50) / 15.81 = –1.26
c. If X = 40: Z = (40 – 50) / 15.81 = –0.63
d. If X = 50: Z = (50 – 50) / 15.81 = 0.00
e. If X = 70: Z = (70 – 50) / 15.81 = 1.26
f. If X = 80: Z = (80 – 50) / 15.81 ≈ 1.90

In contrast, Min-Max Scaling was calculated using the following formula:

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Manual calculations were conducted to transform the data range and ensure no feature dominates the distance calculation in the KNN algorithm. Using the same values (30, 40, 50, 60, 70, 80), the results are:

a. If X = 50: X_scaled = (50 – 30) / (70 – 30) = 20 / 40 = 0.50
b. If X = 30: X_scaled = (30 – 30) / (70 – 30) = 0 / 40 = 0.00
c. If X = 40: X_scaled = (40 – 30) / (70 – 30) = 10 / 40 = 0.25
d. If X = 60: X_scaled = (60 – 30) / (70 – 30) = 30 / 40 = 0.75
e. If X = 70: X_scaled = (70 – 30) / (70 – 30) = 40 / 40 = 1.00

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

f. If X = 80: X_scaled = (80 – 30) / (70 – 30) = 50 / 40 = 1.25

Based on these calculations, the author determined the normalized mean values for each method, as summarized in Table 2, which presents the standardized outputs of both Z-score and Min-Max Scaling processes.

Table 2. Normalization Results Using Z-Score and Min-Max Scaling

Water Level (TMA)	Z-Score	Min-Max Scaled
30 cm	–1.26	0.00
40 cm	–0.63	0.25
50 cm	0.00	0.50
60 cm	0.63	0.75
70 cm	1.26	1.00
80 cm	1.90	1.25

Based on the data, the normalization results provide interpretive values derived from both Z-score and Min-Max Scaling formulas. Specifically, a water level value of 30 is mapped to a normalized score of 0 (indicating the lowest risk), while a value of 70 corresponds to a normalized score of 1 (representing a high risk). A value of 80 is mapped beyond the maximum normalized range, with a score of 1.25 (indicating the highest observed level). All other values are proportionally mapped between 0 and 1 to reflect their respective accuracy contributions within the model's predictive categorization.

Data Processing in Python

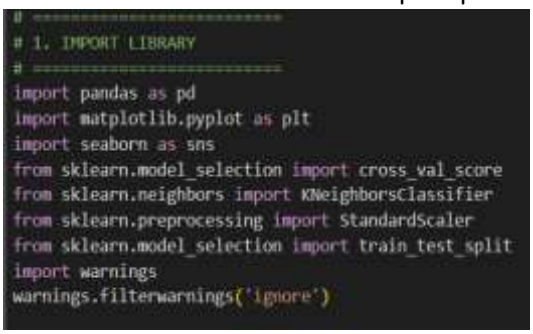
The model testing was conducted by classifying flood disaster predictions across the Jakarta area using machine learning techniques. The evaluation utilized Python as the development environment, with the implementation of the K-Nearest Neighbors (KNN) algorithm.

Importing Required Libraries

The initial step involved importing the necessary libraries for data processing and analysis. This included pandas for data handling, matplotlib.pyplot and seaborn for visualization, along with other essential libraries. The corresponding source code for importing libraries is shown in Figure 3.

Importing Libraries

The first step involved importing the necessary libraries for analysis. This included pandas for data processing, matplotlib.pyplot and seaborn for data visualization, as well as other supporting libraries. The source code used for this import process is shown in Figure 3.



```
# ==========================
# 1. IMPORT LIBRARY
# ==========================
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import cross_val_score
from sklearn.neighbors import KNeighborsClassifier
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
import warnings
warnings.filterwarnings('ignore')
```

Figure 3. Import Library

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Data Retrieval and Cleaning

The dataset was retrieved from an Excel file named "Cleaned Dataset.xlsx", which contains multiple worksheets, each representing a different area. Each worksheet was read individually using the pandas.read_excel() function. This process is illustrated in Figure 4.

```
# ============================
# 2. LOAD FILE DAN SHEET
# ============================
file_path = "/content/Cleaned Dataset.xlsx"
xls = pd.ExcelFile(file_path)
sheet_names = xls.sheet_names
```

Figure 4. Load file and Sheet

Subsequently, missing values (NaN) were removed using the dropna() function to ensure high data quality. In addition, the dataset was visualized using a frame plot to generate a structured graphical output, as illustrated in Figure 5.

```
# ============================
# 3. INISIASI FRAME PLOT
# ============================
fig, axs = plt.subplots(2, 3, figsize=(18, 10))
axs = axs.flatten()
best_k_list = []

# ============================
# 4. LOOP UNTUK TIAP SHEET
# ============================
for i, sheet in enumerate(sheet_names):
    df = pd.read_excel(file_path, sheet_name=sheet)
    df = df.dropna()
```

Figure 5. Plot Frame

Separating Features and Target Variables

The dataset columns were divided into two parts: the features (X), which consist of columns from the fourth up to the penultimate column, and the target (Y), which is the final column containing the prediction labels. The corresponding source code is shown in Figure 6

```
# Asumsikan fitur = semua kolom
X = df.iloc[:, 3:-1]
y = df.iloc[:, -1]
```

Figure 6. Column Feature Assumptions And Target Data

Determining the Optimal K Value

An iterative process was conducted by testing various K values ranging from 1 to 30, with 31 set as the endpoint or upper limit. For each iteration, a KNN model was constructed using a specific number of neighbors. The model's accuracy was evaluated using 5-fold cross-validation. The K value that produced the highest average accuracy was selected as the optimal K for each region. The source code for this process is shown in Figure 7.

```
# Cari nilai K optimal
error_rate = []
cv_scores = []
k_range = range(1, 31)

for k in k_range:
    knn = KNeighborsClassifier(n_neighbors=k)
    scores = cross_val_score(knn, X_train, y_train, cv=5, scoring='a
    error_rate.append(1 - scores.mean())
    cv_scores.append(scores.mean())
best_k = k_range[cv_scores.index(max(cv_scores))]
best_k_list.append(best_k)
```

Figure 7. Source code for finding the optimal k value

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Training the K-NN Model, Prediction, and Classification

After splitting the dataset, the testing phase involved training the K-NN algorithm for flood prediction classification. The process incorporated normalization techniques such as Z-score, Min-Max Scaling, and evaluation using a confusion matrix. To enhance model performance, training was conducted using cross-validation with x_train and y_train data. These three processing steps aimed to optimize the model for future flood event prediction. The corresponding source code is shown in Figure 8.

```
# Buat dan latih model KNN dengan k terbaik
k_optimal = best_k_list[i]
knn_model = KNeighborsClassifier(n_neighbors=k_optimal)
knn_model.fit(X_train, y_train)

# Prediksi
y_pred = knn_model.predict(X_test)

# Classification Report
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Figure 8. Training the K-NN Model

Confusion Matrix and Regional Data Report

The final step involved visualizing the data using a confusion matrix derived from the K-NN model's prediction results for all regions. This evaluation was conducted to compare the classification performance by measuring the percentage match between the actual values (y_test) and the predicted values (y_pred). The heatmap also displays numerical values within each box, representing the types of outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Figure 9 presents the prediction results for the Central Jakarta region.

Figure 9. Prediction results for the Central Jakarta area

Subsequently, the prediction results for the North Jakarta region are presented in Figure 10.

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

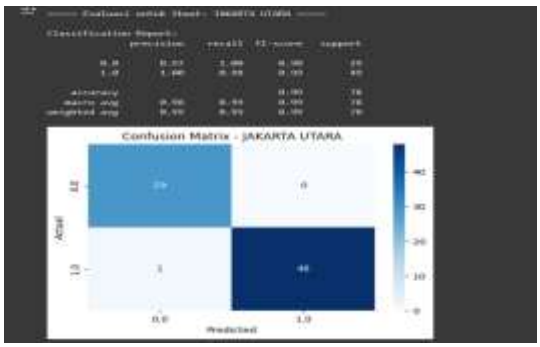


Figure 10. Prediction results for the North Jakarta area

Next, the prediction results for the West Jakarta region are shown in Figure 11.

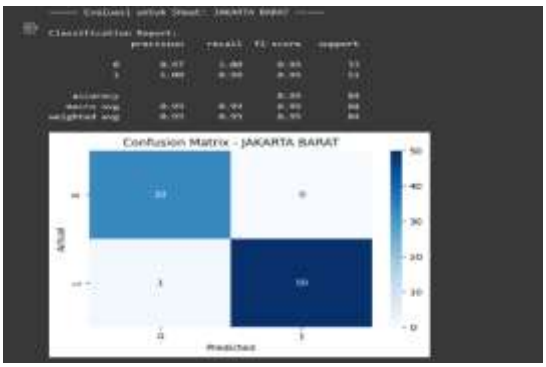


Figure 11 Prediction results in the West Jakarta area

Next, the prediction results for the South Jakarta region are presented in Figure 12.

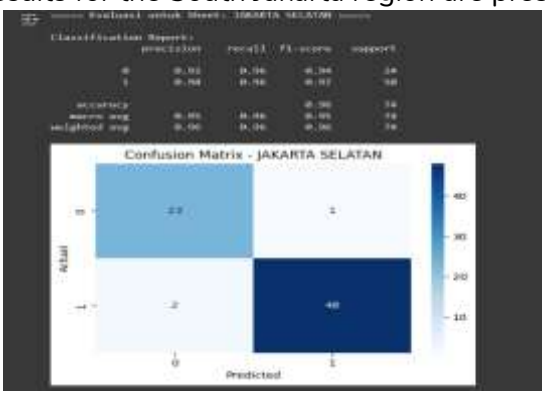


Figure 12 Prediction results in the South Jakarta area

Next, the prediction results for the East Jakarta region are displayed in Figure 13.

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

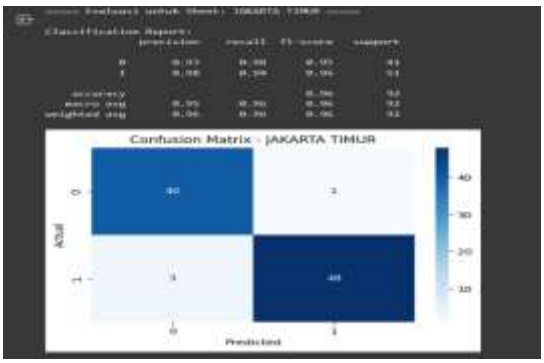


Figure 13 Prediction results in the East Jakarta area

Subsequently, the prediction results for the Thousand Islands region are illustrated in Figure 14.

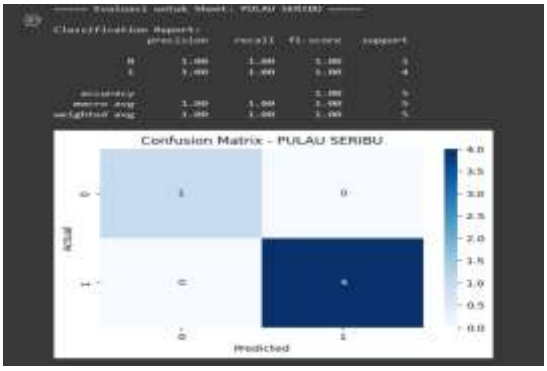


Figure 14. Prediction results in the Thousand Islands area of Jakarta

Classification and Prediction Results

The author presents findings on the classification and prediction of potential flood occurrences in the Jakarta region by applying the K-Nearest Neighbors (KNN) algorithm. This study was conducted using historical data from 2013 to 2024, which had been preprocessed through normalization using both Z-Score and Min-Max Scaling methods. The resulting model was then employed to estimate flood risk levels for the years 2026 through 2028.

Accordingly, the analysis of water level data and graphical visualizations across Jakarta and surrounding areas produced optimal predictive assumptions for the three-year forecast period (2026, 2027, and 2028). These predictions were based on the normalized classification outputs and optimal bias-calibrated results. The complete classification and prediction results are summarized in Table 3.

Table 3. Classification and Prediction Results

Region / RW	Water Level 2025 (cm)	Z-Score	Min-Max	Prediction 2026	Prediction 2027	Prediction 2028	Accuracy	Precision	Recall	F1-Score
Thousand Islands (A)	30	–1.26	0.00	Low	Moderate	Moderate	1.00	1.00	1.00	1.00
South Jakarta (B)	40	–0.63	0.25	High	Moderate	High	0.96	0.95	0.96	0.95
West Jakarta (C)	50	0.00	0.50	High	High	High	0.99	0.99	0.99	0.99
East Jakarta (D)	60	0.63	0.75	Moderate	Moderate	High	0.96	0.95	0.96	0.96
North Jakarta (E)	70	1.26	1.00	High	High	High	0.99	0.98	0.99	0.99
Central Jakarta (F)	80	1.90	1.25	High	High	High	0.97	0.95	0.97	0.96

The prediction results indicate that regions expected to experience consistently high flood risk over the next three years include West Jakarta, North Jakarta, and Central Jakarta. Meanwhile, South Jakarta and East Jakarta fall into the moderate risk category, with potential

increases to high risk in certain years. The Thousand Islands region is predicted to remain within the low to moderate risk category, mainly due to data bias resulting from a smaller sample size. Nevertheless, the normalization process has produced a significant comparative calculation, allowing the model to generate low predicted values despite achieving high and optimal evaluation metrics.

CONCLUSION

Based on the findings of this study, which aimed to analyze the potential application of the K-Nearest Neighbors (KNN) method in constructing a flood risk classification model using water level data and historical flood events in the Jakarta Province, several important conclusions can be drawn. The KNN algorithm proved to be effective in classifying flood risk into three categories—low, moderate, and high—with high levels of accuracy, ranging from 96% to 99%, along with consistently strong performance in precision, recall, and F1-score. The normalization techniques using Z-Score and Min-Max Scaling played a significant role in enhancing the model's accuracy by adjusting feature scale uniformity. The optimal K values were determined through cross-validation, where K=3 yielded the best classification results, while K=1 was more effective in normalization for larger samples. The prediction results for the years 2026 to 2028 reveal a significant increase in flood risk for certain regions, particularly Central Jakarta, West Jakarta, and North Jakarta, based on the normalized water level data. South Jakarta and East Jakarta are projected to experience moderate to high risk, whereas the Thousand Islands region is expected to remain within the low to moderate risk range.

REFERENCE

Angreni, D. S., & Pusadan, M. Y. (2024). Klasifikasi curah hujan menggunakan algoritma K-Nearest Neighbor (KNN) di Sulawesi Tengah. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 9(4), 2316–2324.

Cumel, D. Z., & Rahmaddeni, S. (2022). Perbandingan metode data mining untuk prediksi banjir dengan algoritma Naïve Bayes dan KNN. *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian kepada Masyarakat*, 40–48. https://journal.irpi.or.id/index.php/sentimas/article/view/353 https://journal.irpi.or.id/index.php/sentimas/article/download/353/132

Hermawan, E., Panjaitan, S. D., & Ripanti, E. F. (2024). Sistem prediksi banjir rob Kota Pontianak berbasis machine learning menggunakan framework Streamlit. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 10(3), 351–361.

Rafi Nahjan, M., Heryana, N., & Voutama, A. (2023). Implementasi RapidMiner dengan metode clustering K-Means untuk analisa penjualan pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104. https://doi.org/10.36040/jati.v7i1.6094

Restu, A. K. A., Siswa, T. A. Y., & Pranoto, W. J. (2024). Model optimasi KNN-PSORF dalam menangani high dimensional data banjir Kota Samarinda. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 7(3), 1289–1299. https://doi.org/10.32493/jtsi.v7i3.41587

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan machine learning dalam berbagai bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. https://doi.org/10.31294/ijcit.v5i1.7951

Sandiwarno, S. (2024). Penerapan machine learning untuk prediksi bencana banjir. *Jurnal Sistem Informasi Bisnis*, 14(1), 62–76. https://doi.org/10.21456/vol14iss1pp62-76

Wardani, S. D. K., Ariyanto, A. S., Umroh, M., & Rolliawati, D. (2023). Perbandingan hasil metode clustering K-Means, DB Scanner & Hierarchical untuk analisa segmentasi pasar. *JIKO (Jurnal Informatika dan Komputer)*, 7(2), 191. https://doi.org/10.26798/jiko.v7i2.796

Whendasmoro, R. G., & Joseph, J. (2022). Analisis penerapan normalisasi data dengan menggunakan Z-Score pada kinerja algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 872. https://doi.org/10.30865/jurikom.v9i4.4526

Yudono, M. A. S., & Akbar, J. (2024). Water level classification for early flood detection using KNN method. *Fidelity: Jurnal Teknik Elektro*, 6(2), 49–57. https://doi.org/10.52005/fidelity.v6i2.227