Flood Prediction based on Weather Parameters in Jakarta using K-Nearest Neighbours Algorithm

Hariman Lumbantobing[1], Irma Ratna Avianti[2], Kukuh Harisapto[3], Suharjito[4]
[1,2,3,4]Universitas Bina Nusantara, Jakarta, Indonesia
Email: hariman.lumbantobing@binus.ac.id, irma.avianti@binus.ac.id, kukuh.harisapto@binus.ac.id, suharjito@binus.ac.id

ABSTRACT

Flooding is a difficult and common hazard in Indonesia, particularly in Jakarta during the rainy season. Floods have been the subject of several endeavours, ranging from discovering the causes to reducing their impacts. Floods cause significant damage to infrastructure, the social economy, and human lives. The government continues to create reliable flood risk maps and plans for long-term flood risk management. According to data from Jakarta Flood Monitoring, 12 sub-districts and 26 urban villages were hit by floods each year between 2016 and 2020, with an average flood length of nearly 2 days. The flood tendency in Jakarta decreased from 2018 to 2019, but increased in 2020. Floods are produced by a variety of reasons, including weather, geography, and human actions such as deforestation. Strong flood prediction is required for disaster management, however this might be difficult owing to changing weather conditions. This study focuses on flood prediction in Jakarta based on weather parameters utilising machine learning techniques to provide accurate and real-time predictions. K-Nearest Neighbours (KNN) is an algorithm employed to forecast the areas that will encounter the consequences of floods. The outcomes of this research with the value of k=2 to k=9 obtained the best performance values at k=7, where the level of accuracy reaches 92.25%, 88.89% precision, 92.25% recall, and F1-measure of 89.52%. The integration of machine learning algorithms which encompasses multiple weather variables provides significant utility in comprehensive flood predictions and early warning systems in flood disaster mitigation.

KEYWORDS	*Flood Prediction, Weather Parameters, Machine Learning, K-Nearest Neighbours*

Hariman Lumbantobing, Irma Ratna Avianti, Kukuh Harisapto, and Suharjito

INTRODUCTION

Floods are one of the most complicated and widespread disasters (Tayfur et al., 2018). Floods are a common occurrence in Indonesia during the rainy season, particularly in Jakarta. Numerous attempts have been undertaken to prevent and reduce the possibility of floods, starting from the identification of the causes and mitigation of their impact (Shafizadeh-Moghadam et al., 2018). Floods are the most damaging natural disasters, with widespread impacts on infrastructure, social economics, and human life (Chuang et al., 2020). Thus, the government continues to develop reliable and accurate maps of flood risk areas and prepare for future sustainable flood risk management that emphasizes protection and preventative measures (Kamal et al., 2018).

According to data collected between 2016 and 2020 in Jakarta, Indonesia from Jakarta Flood Monitoring (pantaubanjir.jakarta.go.id). The average area impacted by floods during seven years was 12 sub-districts and 26 urban villages, and the average duration of flooding was almost 2 days.
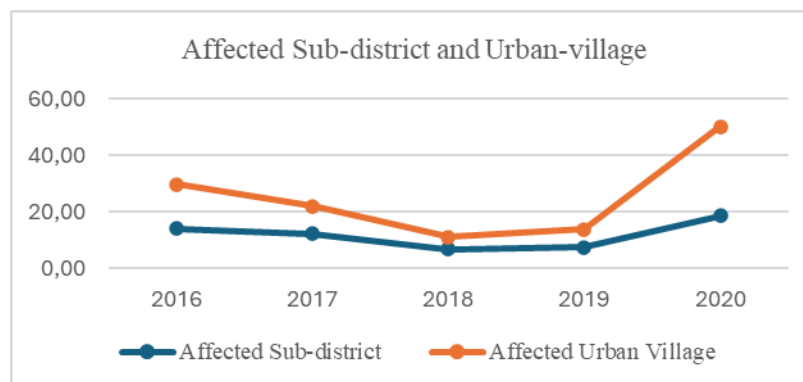

Figure 1. Average of Affected Sub-district and Urban-village by Flood in Jakarta, Indonesia

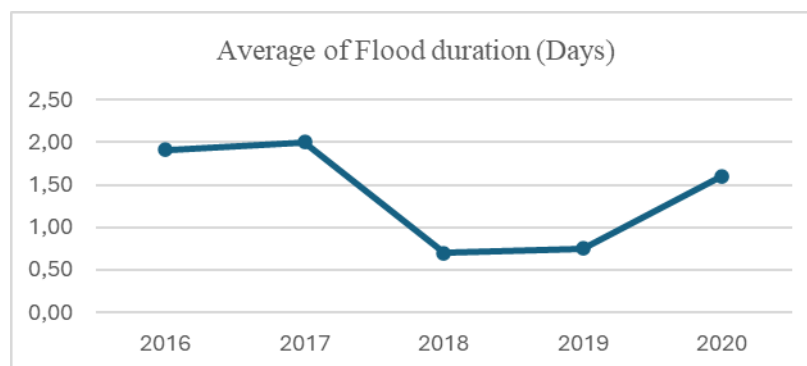**Figure 2. Average of Flood Duration Affected Areas**
(Source: pantaubanjir.jakarta.go.id)

Based on the graph above, the areas affected by flooding in the Jakarta area experienced a downward trend from 2018 to 2019 but experienced an increasing trend in 2020. Floods can be caused by a combination of several causal factors, namely high rainfall, topography, and human factors. The increase in human

activity and population is closely related to increased land use and increased land area covered. This causes the water absorption area to decrease and the abundance of water to increase along with the increase in forest conversion into agricultural areas and residential areas. This deforestation phenomenon is directly proportional to increased erosion and shallowing of rivers (Riza et al., 2020).

Floods often occur in big cities as a result of rapid urban growth which causes water catchment areas to be covered by buildings erected by humans. The occurrence of floods cannot be separated from the behavior of humans who like to throw rubbish into waterways and rivers, plus the river area has become narrow with houses built along the riverbanks (Hernawan et al., 2024). Strong, reliable, and accurate prediction models are needed in hazard assessment and extreme event management, where this contributes greatly to strategies and policies in water resource management and evacuation in the event of a disaster. Advanced prediction systems for predicting floods and other hydrological events in the short and long term are highly prioritized to reduce the damage that occurs (Harianto, 2022). However, predicting flood waiting times and locations where floods occur can be a complicated problem due to the dynamic climatic conditions in each region. Thus, current major flood prediction models mostly use specific data and have simplified assumptions (Putra et al., 2019).

In this study, the researchers conducted flood prediction to delineate the regions that would experience flooding in the Jakarta area. This was achieved through the utilization of a sophisticated computational technique that can yield accurate and real-time predictions. Machine learning methodologies have been extensively employed in all predictive analyses involving variables that have the potential to influence the outcomes of the predictions. Drawing upon a thorough examination of the existing literature about flood prediction and taking into account the available data and desired outcomes, the researchers opted for the implementation of K-Nearest Neighbors (KNN) algorithm models for flood prediction specifically in the Jakarta area situated in Indonesia.

Research conducted in Riza (2020), titled "Advancing Flood Disaster Mitigation in Indonesia using Machine Learning Methods", undertook a comprehensive literature review encompassing all publications about the utilization of machine learning in the field of flood disaster mitigation within the Indonesian context. The empirical data employed for this study encompassed flood events occurring throughout Indonesia across a span of 15 years, specifically between 2005 and 2019. This research scrutinized a wide array of factors that contribute to the occurrence of flood disasters, while also considering multiple structural and non-structural approaches to mitigating such disasters (Khosravi et al., 2018). The application of machine learning techniques for flood disaster mitigation has been explored in various research studies, utilizing flood event data from diverse regions in Indonesia and employing different algorithms. This study encompasses a broad range of factors associated with flooding, commencing with rainfall forecasting and river water level predictions, which serve as integral components of flood forecasting and early warning systems.

In the Banyuwangi region of Indonesia, ANFIS (Adaptive Neuro Fuzzy Inference System) was employed to conduct rainfall forecasts. Two neural network

methods were utilized, with the initial method displaying superior accuracy. Similarly, in the Denpasar area, a comparative analysis was conducted in 2016 to predict rainfall, employing Adaline and Multiple Linear Regression. The resulting error rates, namely MSE, were 0.025129 and 0.025953, respectively. Subsequently, the RMSE values were determined to be 0.158522 and 0.161098. In Malang City, rainfall predictions were carried out using ANN (Artificial Neural Network), yielding time-specific forecasts. The monthly, daily, and hourly error rates were determined to be 11.49%, 8.49%, and 19.32%, respectively.

In the Ciliwung region of Indonesia, the use of ANFIS, ANN, and FIS was employed to make predictions regarding river water levels. The findings demonstrated that ANFIS, utilizing three feature data, exhibited superior performance in comparison to ANN. The outcomes of ANFIS can be utilized as input for the FIS model, which subsequently enables accurate water level predictions at the Manggarai sluice gate. Within this context, BP-NN (Backpropagation-Neural Network) is employed to forecast both rainfall and water levels. Moreover, SVM is incorporated to predict floods for the upcoming six days, with the optimal combination consisting of 60 training data and 40 testing data points.

Similarly, within the Deli Serdang area of North Sumatra, BP-NN was employed to forecast rainfall and river water levels. Additionally, SVM was utilized to predict floods. In the Ular Tajur River area, BP-NN was implemented to forecast river water discharge. By utilizing a combination of training and testing data, flood predictions in a 60:40 ratio can be obtained, thereby serving as an initial warning for the subsequent six days. Furthermore, in this area, flood predictions were conducted using the K-NN algorithm and Naïve Bayes approaches, based on rainfall and water levels. These predictions demonstrated a commendable accuracy rate of 93.4%, with an associated error rate of 6.6%. Numerous other regions in Indonesia have been the subject of research endeavors aimed at predicting floods and rainfall using machine learning techniques. The outcome of this literature review research will serve as a foundation for BPBD DKI Jakarta in providing flood prediction data, which will be instrumental in making decisions regarding the opening and closing of the Manggarai floodgates. In the future, flood predictions can be facilitated through the implementation of single and double algorithms, incorporating Fuzzy Logic and ANN.

Research conducted in Sankaranarayanan (2020), titled "Flood Prediction based on Weather Parameters using Deep Learning" explored flood predictions in India. The predictions were based on various factors such as rainfall, humidity, temperature, water flow, and water level. To determine the most accurate model, the deep neural network was compared to other models including SVM, KNN, and Naïve Bayes. The research utilized a dataset consisting of flood events in India spanning from 1990 to 2002, ultimately concluding that the Deep Neural Network model was the most suitable for the Indian dataset.

In a study conducted in Gauhar (2021), titled "Prediction of Flood in Bangladesh using k-Nearest Neighbors Algorithm," flood predictions were carried out specifically in the Bangladesh region. The researchers employed the feature selection and the KNN (k-Nearest Neighbors) algorithm model. The dataset utilized consisted of 20544 data points from 32 districts in Bangladesh. The attributes used in

the study included rainfall, cloud coverage, relative humidity, minimum temperature, and wind speed. The results yielded a high testing accuracy of 94.91%, an average precision of 92%, and an average recall of 91% using the KNN model.

RESEARCH METHOD

The weather records spanning 5 years in Jakarta have been sourced from Kaggle, drawing from data compiled by various outlets located in Jakarta. Specifically, details regarding flood occurrences in specific months and years were gathered from diverse outlets such as annual flood reports, newspapers, and academic papers. These findings were subsequently generated to a dataset, accessible in, comprising a total of 90185 entries. Figure 3 explained dataset that contains 15 attributes data such as average humidity, rainfall, minimum temperature, maximum temperature, average temperature, wind direction at maximum speed, maximum wind speed, location of the station, duration of sunshine, most wind direction, station id which record the data, and the station name. These attributes data are independent variables, last column is flood as the target data, which is 1 means true and 0 means false. This flood column is dependent variable. There was 4.7% missing data should be fixed using normalization method.

Info		date	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg
6308 instances	2529	2018-06-09 00:..	27	?	?	?	?	0	6	?	?	?
15 features (4.7 % missing data)	2532	2018-06-13 00:..	27	?	?	?	?	5.3	?	?	?
No target variable.	2568	2018-07-24 00:..	27	?	?	?	?	6.7	?	?	?
No meta attributes.	2647	2018-10-17 00:..	27	?	?	?	?	8.8	?	?	?
	3248	2020-06-12 00:..	25.4	?	?	?	5.6	8.8	?	?	?
	5519	2016-10-25 00:..	24.6	28.8	24.8	97	?	?	?	?	?
	5587	2017-01-01 00:..	25	34.4	29.5	70	?	?	5	270	?
	5588	2017-01-02 00:..	25.2	34.4	29.7	64	?	?	6	230	?
	5589	2017-01-03 00:..	23	33.6	29.4	68	12	?	6	240	?
	5590	2017-01-04 00:..	22.8	30.6	26.2	84	11.8	?	3	260	?
	5591	2017-01-05 00:..	24.4	31.2	27.5	82	38.4	?	5	200	?
	5592	2017-01-06 00:..	24.6	32.6	27.7	83	12	?	4	260	?
	5593	2017-01-07 00:..	24.6	32.6	28.5	71	?	?	5	260	?
	5594	2017-01-08 00:..	24.2	32	27.9	74	0.2	?	4	310	?
	5595	2017-01-09 00:..	24	31.6	25.8	85	9.4	?	4	170	?
	5596	2017-01-10 00:..	23.7	32.6	27.7	82	4.3	?	4	360	?
	5597	2017-01-11 00:..	24	31.2	27	85	3.2	?	5	240	?
	5598	2017-01-12 00:..	25.4	32.6	27.7	78	?	?	8	250	?
	5599	2017-01-13 00:..	26	31	27.8	74	?	?	8	250	?
	5600	2017-01-14 00:..	23.9	28.2	26.6	77	3.4	?	8	250	?
	5601	2017-01-15 00:..	23.8	30.4	27.3	76	5	?	7	270	?
	5602	2017-01-16 00:..	24.2	28.2	26.5	83	0.4	?	9	250	?
	5603	2017-01-17 00:..	24	31.8	26.6	80	2.6	?	6	250	?
	5684	2017-04-08 00:..	24.6	?	26.9	78	?	?	8	240	?

Figure 3. Dataset Preparation

Given the disparate units, ranges, and magnitudes observed in the dataset's features, it became imperative to standardize or normalize the data. This study employed z-score to address this variability. This method transforms the data by centering it around a mean value of zero and adjusting its scale to achieve unit variance. The z-score normalization formula is represented as,

$$Z_{score} = \frac{(x - \mu)}{\sigma} \tag{1}$$

Here, x represents the sample data, μ signifies the mean of the training sample, and σ denotes the standard deviation of the training sample. This normalization process ensures that the features are on a comparable scale, facilitating more effective analysis and modeling.

Figure 4 represents all attributes has been normalizing using z-score whereas no more missing data in dataset. Therefore the dataset would be proceeding with k-NN algorithm and then data also be evaluated.

Info		date	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg
6308 instances (no missing data) 15 features No target variable. No meta attributes.	1	2016-01-01 00:...	0.0968801	0.0431724	0.0415606	0.103524	0.0339297	0.50000	0.027027	0.777778	0.0243902
	2	2016-01-02 00:...	0.0952381	0.0380556	0.0279898	0.118943	0.00576577	0.75000	0.0216216	0.805556	0.0243902
	3	2016-01-03 00:...	0.090312	0.0434922	0.0373198	0.101322	0.121802	0.4655172	0.0216216	0.777778	0.0243902
	4	2016-01-04 00:...	0.091954	0.0393348	0.0466497	0.103524	0.0339297	0.5689655	0.0162162	0.555556	0.0121951
	5	2016-01-05 00:...	0.0960591	0.0393348	0.0254453	0.125551	0.0339297	0.2758621	0.0162162	0.5	0.0121951
	6	2016-01-06 00:...	0.092775	0.0393348	0.0441052	0.101322	0.0136937	0.3103448	0.0216216	0.527778	0.0243902
	7	2016-01-07 00:...	0.0935961	0.0457307	0.0534351	0.0969163	0.0339297	0.6551724	0.0216216	0.166667	0.0243902
	8	2016-01-08 00:...	0.100164	0.0360637	0.0525869	0.0991189	0.0339297	0.6810345	0.0216216	0.138889	0.0243902
	9	2016-01-09 00:...	0.100164	0.0431724	0.0508906	0.0991189	0.0339297	0.8793103	0.0216216	0.305556	0.0243902
	10	2016-01-10 00:...	0.0944171	0.0425328	0.0466497	0.101322	0.0339297	0.7068966	0.0162162	0.222222	0.0243902
	11	2016-01-11 00:...	0.0952381	0.040614	0.0381679	0.10793	0.0339297	0.4224138	0.0162162	0.777778	0.0121951
	12	2016-01-12 00:...	0.0977011	0.0316597	0.0195081	0.129956	0.0414414	0.4913793	0.0162162	0.138889	0.0121951
	13	2016-01-13 00:...	0.091954	0.0383754	0.0220558	0.127753	0.130811	0.0775862	0.0162162	0.75	0.0121951
	14	2016-01-14 00:...	0.092775	0.0399744	0.0407125	0.110132	0.0158559	0.3189655	0.0162162	0.111111	0.0243902
	15	2016-01-15 00:...	0.0968801	0.0399744	0.0449534	0.0339297	0.0339297	0.4137931	0.0162162	0.194444	0.0121951
	16	2016-01-16 00:...	0.0944171	0.0393348	0.0415606	0.105727	0.00216216	0.3362069	0.0108108	0.166667	0.0121951
	17	2016-01-17 00:...	0.0960591	0.0444516	0.0373198	0.114537	0.0339297	0.2672414	0.0216216	0.444444	0.0243902
	18	2016-01-18 00:...	0.0944171	0.040614	0.0407125	0.103524	0.0162162	0.5862069	0.0216216	0.777778	0.0243902
	19	2016-01-19 00:...	0.0952381	0.0399744	0.0313825	0.121145	0.0122523	0.2672414	0.0216216	0.694444	0.0243902
	20	2016-01-20 00:...	0.0952381	0.0386952	0.0347752	0.11674	0.00324324	0.3965517	0.0216216	0.722222	0.0243902
	21	2016-01-21 00:...	0.0960591	0.0322993	0.0339271	0.114537	0.00108108	0.1379310	0.0108108	0.722222	0.0121951
	22	2016-01-22 00:...	0.0985222	0.0342181	0.0424088	0.110132	0.00036036	0.0689655	0.0216216	0.833333	0.0243902
	23	2016-01-23 00:...	0.0878489	0.029741	0.0212044	0.118943	0.0500901	0.0603448	0.0216216	0.361111	0.0243902
	24	2016-01-24 00:...	0.092775	0.0361369	0.0296862	0.125551	0.027027	0.0172414	0.0162162	0.5	0.0121951
	25	2016-01-25 00:...	0.0845649	0.0310201	0.0152672	0.136564	0.201802	0.0862069	0.0108108	0.333333	0.0121951

Figure 4. Dataset Scaling

A. Machine Learning Classifier:

The k-nearest Neighbor (k-NN) algorithm is a widely used supervised machine learning technique that leverages feature similarity to predict outcomes for new data points. This approach assigns values to predicted data points based on their resemblance to the nearest points in the training set. KNN algorithm was used for predicting floods in Jakarta that involves several steps:

1. Find the k-value, where k represents the number of nearest neighbors.
2. Determine the distances between the training data points and the data point to be classified.
3. Order the training data points by arranging them in ascending sequence according to their distance values.
4. Make predictions based on the majority of the nearest neighbors.
5. In our study, k-value adjusted from 2 to 9. We employed a uniform weight function to assign equal weights to all points within each neighborhood. To compute distances, we utilized the Minkowski distance formula,

$$\left(\sum_{i=1}^{n} |X_i - Y_i|^p \right)^{\frac{1}{p}} \tag{2}$$

Figure 1 provides a visual representation of how the k-NN algorithm operates in predicting floods. The data point under consideration is compared to its k-nearest neighbors, and based on the closest and most similar points, it is classified accordingly.

In this phase, the complete dataset was divided into two subsets: a training dataset and a testing dataset, ratio maintained by 80:20. Specifically, designating 20% of the data for testing purposes while retaining the remaining 80% for training the model. To achieve this split, we utilized a test size parameter set to 0.2,

indicating that 20% of the dataset would be allocated to the testing set. The random state was fixed at 50 to ensure reproducibility and consistency in the train-test splits across different runs. This deterministic approach guarantees same training and test subset that generated each time the dataset is partitioned, aiding in reliable model evaluation and comparison.

In assessing for effectiveness of the model, this study employed several key metrics: accuracy, precision, recall, and F1-score. True Positive (TP) signifies instances where the model correctly predicts the occurrence of floods, while True Negative (TN) represents cases where the model accurately predicts the absence of floods. False Positive (FP) occurs when the model incorrectly predicts flood occurrences, whereas False Negative (FN) indicates instances where the model inaccurately predicts the absence of floods.

The following formulas were utilized to compute these metrics:

Accuracy: This metric gauges the overall correctness of the model's predictions and is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Precision: Precision quantifies the proportion of true flood predictions among all positive predictions, and is expressed as:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall: Recall, also known as sensitivity or true positive rate, measures the ability of the model to correctly identify actual flood occurrences, and is given by:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F1-Score: The F1-score is the harmonic mean of precision and recall, offering a balanced assessment of the model's performance, and is calculated as:

$$F1 - Score = 2x + \frac{Precision\ x\ Recall}{Precision + Recall} \tag{6}$$

These metrics provide valuable insights into the model's predictive capabilities, allowing for a comprehensive evaluation of its performance across various aspects of flood prediction.

Result Analysis

The implementation of the system incorporates Orange Data Mining, as illustrated in Figure 5, representing a strategic choice in employing a versatile tool for data analysis and modeling. Following meticulous preprocessing and dataset scaling, the training data is fed into the k-NN model using an iterative function that defines k from 2 to 9. This iterative approach allows for a comprehensive

exploration of various neighborhood sizes and utilized Eucledian metric and uniform weighted each k-NN.
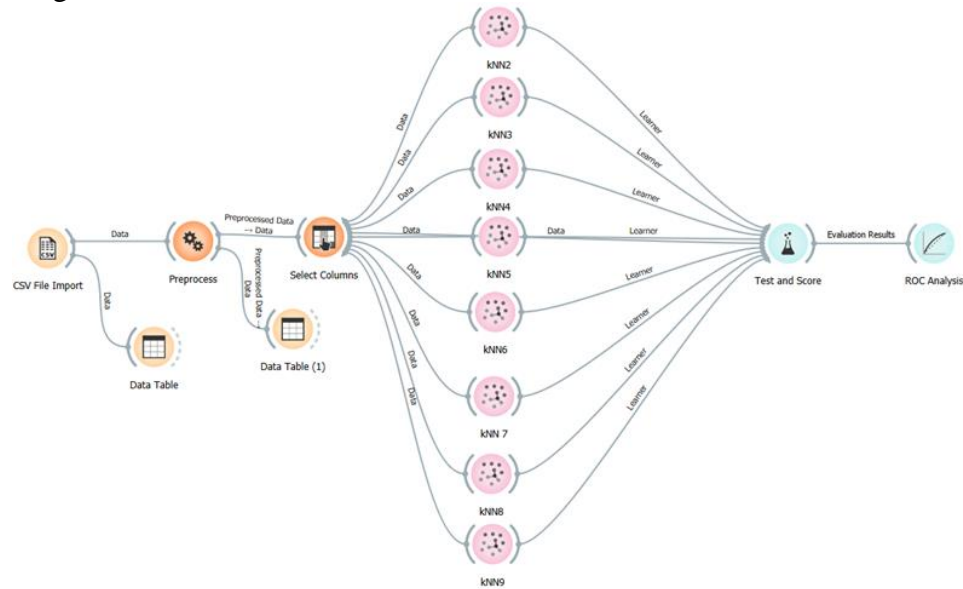

Figure 5. Implementation of k-NN Model

Test and Score with stratified number of folds value equal 2, and random sampling with several parameters such as repeat train-test 50 times and a training set size of 80%, the k-NN model undergoes rigorous evaluation to ensure robustness and generalization. This comprehensive testing strategy enables the assessment of model performance across various iterations, providing insights into its stability and consistency. By stratifying the dataset into two folds and repeating the train-test process multiple times, the model's performance is evaluated under diverse conditions, minimizing the impact of random variations and enhancing the reliability of results. Additionally, the utilization of a substantial training set size ensures ample data for model training while maintaining a sizable portion for validation, striking a balance between model complexity and data availability.

Upon executing the k-NN model with varying values of k, a detailed analysis of performance metrics is conducted, as summarized in Table 1. These metrics provide a comprehensive overview of the system's performance across different k-values. Notably, the system achieves its peak performance with an Accuracy of 92.25%, indicating a high proportion of correctly classified instances. Additionally, metrics such as F1-score, Precision, and Recall offer insights into the model's ability to balance between true positives, false positives, and false negatives, crucial for tasks where misclassifications carry significant consequences.

Table 1. Performance Measure Index

Accuracy	F1	Precision	Recall	AUC

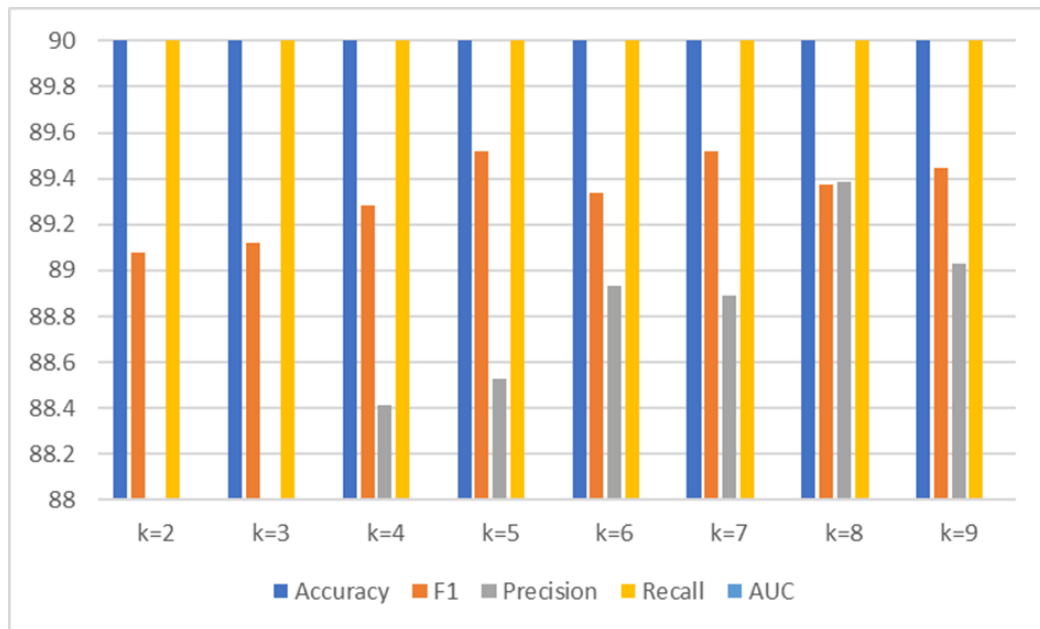|					
k=2	91.88273	89.07542	87.6295	91.88273	0.598
k=3	91.06815	89.12275	87.8290	91.06815	0.624
k=4	92.20919	89.28322	88.4142	92.20919	0.640
k=5	91.93661	89.51938	88.5256	91.93661	0.652
k=6	92.36450	89.33794	88.9306	92.36450	0.661
k=7	**92.25357**	**89.51967**	**88.89151**	**92.25357**	**0.671**
k=8	92.45325	89.37243	89.38343	92.45325	0.677
k=9	92.35182	89.44414	89.0314	92.35182	0.685

Figure 6. Performance Measurement Graph

The performance of the k-NN model across different values of k is further elucidated through Figure 6, the Performance Measurement Graph. This visual representation facilitates the identification of an optimal k-value by showcasing the relationship between k and performance metrics such as accuracy, precision, and recall. Notably, a peak in performance is observed when k = 7, suggesting an optimal balance between model complexity and predictive efficacy.

Moreover, the iterative exploration of k-values allows for the identification of trends beyond the optimal point. Beyond k = 9, a decline in both accuracy and precision values becomes apparent, indicating potential overfitting or loss of generalization beyond a certain neighborhood size. This observation underscores the importance of careful hyperparameter selection and model evaluation to ensure robust performance across varying datasets and application scenarios.

CONCLUSION

Precise flood prediction is crucial for Jakarta, enabling the nation to effectively manage the aftermath of flooding. In our study, we utilized dataset

preprocessing and scaling. Subsequently, utilizing a model of KNN for prediction of flood. By evaluating system's performance across different k values, we identified the optimal parameter. The optimal k value was determined to be 7, resulting in the highest levels of accuracy reaches 92.25%, 88.89% precision, 92.25% recall, and F1-measure of 89.52%. In the future, we anticipate that this study will significantly contribute to advancing flood prediction techniques, ultimately providing Jakarta with a strategic advantage in effectively managing future flood events.

Another limitation of the research is the lack of access to a more recent dataset for validating the machine learning model for flood prediction. This absence potentially restricts the model's ability to achieve higher accuracy. However, this limitation is not considered a major issue. The study also suggests several directions for future research. Firstly, it highlights the need to address data distribution imbalance, especially in light of advancements in neural network prediction techniques. Secondly, the incorporation of additional topographical factors, such as flood water level, could offer further insights and improve the accuracy of flood prediction models.

REFERENCES

Chuang, M.-T., Chen, T.-L., & Lin, Z.-H. (2020). A review of resilient practice based upon flood vulnerability in New Taipei City, Taiwan. *International Journal of Disaster Risk Reduction*, *46*, 101494.

Gauhar, N., Das, S., & Moury, K. S. (2021). Prediction of flood in Bangladesh using K-nearest neighbors algorithm. *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 357–361.

Harianto, D. W. I. Y. (2022). *Analisis Muka Air Banjir Sungai Segeri Pada Persilangan Jalur Ka Lintas Makassar-Parepare*. UNIVERSITAS BOSOWA.

Hernawan, A., Savandha, S. D., Karsa, A. H. A. N., Asikin, M. Z., & Fadilah, M. O. (2024). Application of Business Model Canvas in MSMEs in Karangwuni Village. *International Journal of Social Service and Research*, *4*(03), 912–917.

Kamal, A. S. M. M., Shamsudduha, M., Ahmed, B., Hassan, S. M. K., Islam, M. S., Kelman, I., & Fordham, M. (2018). Resilience to flash floods in wetland communities of northeastern Bangladesh. *International Journal of Disaster Risk Reduction*, *31*, 478–488.

Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I., & Bui, D. T. (2018). A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Science of the Total Environment*, *627*, 744–755.

Putra, F. E. K., Romadhoni, A. Z., & Moe, I. R. (2019). Evaluasi Banjir di Kecamatan Bula Kabupaten Seram Bagian Timur. *MEDIA KOMUNIKASI TEKNIK SIPIL*, *27*(2), 260–267.

Riza, H., Santoso, E. W., Tejakusuma, I. G., & Prawiradisastra, F. (2020). Advancing flood disaster mitigation in Indonesia using machine learning methods. *2020 International Conference on ICT for Smart Society (ICISS)*, 1–4.

Sankaranarayanan, S., Prabhakar, M., Satish, S., Jain, P., Ramprasad, A., & Krishnan, A. (2020). Flood prediction based on weather parameters using deep learning. *Journal of Water and Climate Change*, *11*(4), 1766–1783.

Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., & Shirzadi, A. (2018). Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *Journal of Environmental Management*, *217*, 1–11.

Tayfur, G., Singh, V. P., Moramarco, T., & Barbetta, S. (2018). Flood hydrograph prediction using machine learning methods. *Water*, *10*(8), 968.