T e c h n o v a s i a
Volume **1**, Number **1**, February **2025**
IAM International Journal ⓒ 2025
pp. **10–17**

FLOOD PREDICTION WITH NAÏVE BAYES METHOD

Martin Saputra

Faculty of Computer Science
Esa Unggul University
Jl Arjuna Utara, Jakarta Barat, Indonesia
*Corresponding author: martinsaputra05192021@gmail.com

Received December 2024; accepted February 2025

ABSTRACT

The problem of traffic jams in the Special Region of Jakarta (DKJ) includes various kinds of traffic problems in Jakarta, including traffic jams and flooding. According to the Journal of Sustainable Development Education and Environment, land is lacking for housing, industrial development, and urbanization. The development of an urban area can be influenced by the rapid rate of urbanization, which requires a lot of land, while land in urban areas is relatively minimal. In the literature journal related to the research that will be studied using the Naïve Bayes Algorithm and K-Means Clustering, it can be concluded that it can be used to predict the 2025 flood. The above research results are provided in the form of diagrams, codes, and dashboards, which have a value of results. From the results of clustering using the K-Means method, a water level prediction of 0.96 x 100% = 96% of the data in clustering is very accurate, with an average accuracy of 97% when there is a flood. From the research conducted by the researcher, it can be concluded that data mining is processed through an algorithm. The use of the algorithm affects the data processing results. The researcher used the Naïve Bayes based on the Reference Journal related to the best algorithm in data mining processing. The results of data mining processing are in the form of a curve that can be presented to readers to estimate whether a flood disaster will occur.

Keyword : K-Means, Naïve Bayes, Flood, DKJ, Prediction.

1 Introduction

The problem of traffic congestion in the Special Region of Jakarta (DKJ) includes various types of traffic congestion problems in Jakarta, including traffic jams and flooding in Jakarta. According to the journal of sustainable development education and environment, the lack of land for housing and industrial development and urbanization. The development of an urban area can be influenced by the rapid rate of urbanization, which requires a lot of land [1], while land in urban areas is quite minimal.

According to the journal "Analysis of the Implementation of 30% Green Open Space in DKJ (Special Region of Jakarta)" only around 10% while green open space is very important for water absorption areas in the capital city [4], In addition to the lack of green open space spatial planning, slum housing in DKJ (Special Region of Jakarta) can have an impact on triggering flooding, in a journal note related to slum settlement land entitled "The Effect Of Slum Settlement Relocation On Environmental Health Quality (Case Study Of Kampung Pulo, East Jakarta)" [5].

One of the areas called Kampung Pulo, as many as 927 families live there, but has been relocated by the Regional Government (PEMDA), as many as 518 heads of families. Kampung Pulo is one of the residential areas that often experiences flooding. Many slum housing is built next to the river, causing flooding.

A total of 92 points were flooded in 2015, according to BMKG (Meteorology, Climatology and Geophysics Agency) in the journal “Analysis Of Flood Disaster Preparedness In Jakarta” [6] with a height of 10-80 cm spread across several locations, 28 points in West Jakarta, 17 points in North Jakarta, 35 points in Central Jakarta, 8 points in East Jakarta and 5 points in South Jakarta.

Through statistical data from 2018-2020, several areas were inundated by quite high floods, but in 2021-2024 the flood problem in Indonesia has not occurred again, however, from the intensity of sea water levels, it is possible that flooding could occur again, even having a greater impact on the intensity of water levels in DKJ (Special Region of Jakarta).

In preventing a flood in preparedness can estimate when a flood will occur in DKJ (Special Region of Jakarta) using the SVM machine learning method in the journal "Application of Machine Learning for Flood Disaster Prediction" discusses using the SVM method can estimate when a flood will occur The results of the study obtained an average accuracy value of 85% with the best combination of features, namely MFCC (Mel-Frequency Cepstrum Coefficient) and pitch [7].

In a research report above, it is expected as follows 1. Predicting flood height data in 2025 from 2018-2020 data, 2. Grouping flood inundation points from 2018-2020, RT & RW, 3. Making flood report results in 2025 and finding solutions before a flood occurs, 4. Development of the data mining results dashboard in the form of a dashboard using Power BI (Business Intelligence).

2 Methodology

In this study, the main objective is to predict flood events in Jakarta in 2025 using data mining processing methods. The applied data mining processing consists of several stages, each of which has an important role in achieving the desired results. This process involves collecting, processing, and predicting based on the collected flood data. The following are the stages of the methodology used in this study.

2.1. Data Collection

In the first stage, the data used in this study was obtained through data collection. The data required is raw data covering flood events in Jakarta over the past few years. Based on previous research [30], data collection is the stage where unprocessed data is collected for further analysis purposes.

The data used in this study was taken from the Kaggle site, with flood data in the Special Region of Jakarta (DKJ). The collected data is then stored on a shared drive that can be accessed via the following link: [klik]

2.2. Data Processing

After the data is collected, the next step is data processing. In this study, researchers used two main methods in data processing, namely Naïve Bayes for classification and K-Means Clustering for data grouping.

- **Naïve Bayes** is used to classify data into categories such as water depth, flood-affected areas, and sub-districts.
- **K-Means Clustering** is applied to group data based on water depth and affected locations, in order to determine the pattern of flood distribution in Jakarta.

Definition 2.2.1. This data processing system is stable if and only if the results of the data classification and grouping process can provide consistent predictions of flood events.
Lemma 2.2.2. If the data used for processing through the Naïve Bayes and K-Means Clustering algorithms can be classified well, then flood predictions can be made accurately.
Theorem 2.2.3. Consider a system that uses Jakarta flood data. If the Naïve Bayes algorithm is applied to classify the data and K-Means Clustering is used to group the data based on water depth and location, then this system can predict flood events with a high degree of accuracy.
Corollary 2.2.4. If the data used for prediction is clean and relevant data, then the accuracy of the prediction of the Jakarta flood event in 2025 will be very high.

Proof: Suppose we use data that has been classified and grouped according to relevant parameters such as water depth and affected area. Based on the results of the analysis, we can prove that this model is able to produce predictions that are close to the reality of flood events.

2.3. Data Rule

In data processing, there are several rules that need to be followed so that the processed data can provide maximum results. The rules applied are as follows:

- **Processed data must be raw data**
  The data used in processing is raw data that has not been processed. This aims to ensure that the analysis is carried out based on undistorted data.

- **The processed data must include data from the last 10 years**
  The data used in this study must include data from the last 10 years in order to provide a representative picture of flood events in the long term.

- **Data must be clean (cleansing)**
  The data used must be free from missing values or blanks. Clean data ensures that the analysis is not compromised by invalid data.

- **Data must be clean (cleansing)**
  The data used must be free from missing values or blanks. Clean data ensures that the analysis is not compromised by invalid data.

2.4. Data Prediction

After the data has been successfully processed, the next stage is to predict the occurrence of flooding in Jakarta in 2025. The stages carried out in data prediction are as follows:

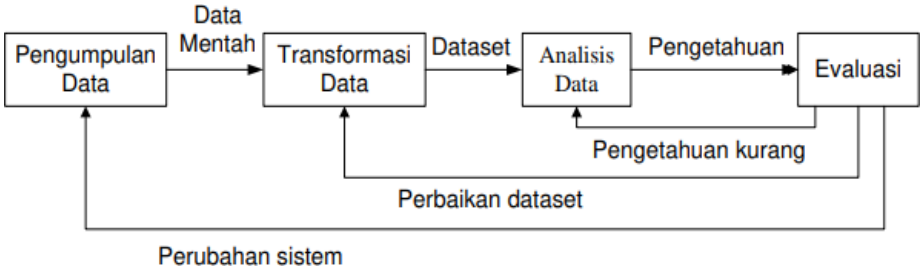

Figure 1. Stages in Predicting Data

- **Data Collection**
  Data collection is the first step in predicting flood events. Data collected in the previous stage will be used as input for further analysis stages.

- **Data Transformation**
  At this stage, the collected data will be processed and filtered. This process includes separating irrelevant data and cleaning data to ensure that the data used can provide accurate results.

- **Data Analysis**
  The data analysis stage aims to determine the most relevant variables in predicting flood events. At this stage, researchers will evaluate the data and select the columns that can provide the greatest contribution to the prediction analysis.

- **Evaluation**
  The evaluation stage is the final stage in the prediction process. At this stage, the results of the data analysis will be evaluated to ensure whether the predictions produced are in accordance with reality or not. This evaluation is very important to validate the results and assess the accuracy of the predictions.

3 Results and Discussion

3.1. Data Mining Results

The results of the data mining process carried out in this study aim to provide information to the wider community regarding flood predictions in the Special Region of Jakarta (DKJ) in 2025. Some of the main results of the data mining analysis used in this study are as follows:

3.1.1. Trend of Number of Affected Per Year

Figure 2. Trend of Number of Affected RW Per Year

Based on data mining, the number of affected RW showed a significant increase in 2019-2020 with moderate flood intensity. However, the prediction results for 2025 show a potential decrease in the number of affected RW.

Figure 3. Trend in the Number of Affected Households Per Year

Data mining analysis based on the number of affected households per year also shows a similar pattern. In 2019- 2020, there was an increase in the number of affected households with moderate flood intensity, while the prediction for 2025 shows a possible decrease in the number of affected households.

Figure 4. Trend in the Number of Affected Households Per Year

Based on the number of affected households (KK), the pattern that emerged was an increase in 2019-2020, but the prediction for 2025 showed a significant downward trend.

Figure 5. Trend in the Number of Affected People Per Year

Data mining shows that the number of affected people increased in 2019-2020, but is predicted to decrease in 2025.

3.1.2. Clustering of Water Levels Based on Clusters

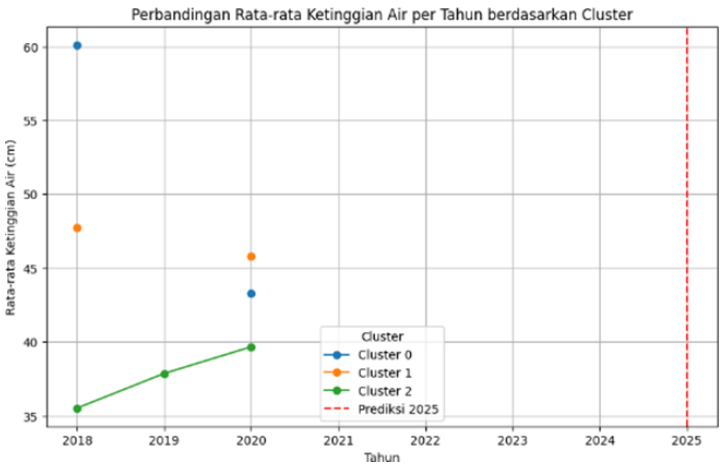

Figure 6. Comparison of Average Water Levels Per Year based on Clusters

Using the clustering method, analysis was conducted on the categories of RT, RW, Jiwa, and KK with different levels of impact. The results of the analysis show that the jiwa category has the highest

level of impact with an increase in flood intensity in 2019-2020, while the prediction for 2025 shows a decrease in flood intensity.

3.2. Water Level Prediction

Classification Using the Naïve Bayes Method

From the classification results using the Naïve Bayes method, the water level prediction is 43.39 cm.

Clustering Using the K-Means Method

With the K-Means clustering method, the water level prediction results show an accuracy level of 96%, with an average accuracy reaching 97%. This shows that the method used has a very high level of reliability in predicting water levels in the analyzed period.

4. Conclusions

From the research conducted by the researcher, it can be concluded that data mining is processed through an algorithm, the use of the algorithm affects the results of the data processing, the researcher uses the K-Means algorithm [2] and Naïve Bayes [3] based on the Reference Journal related to the best algorithm in data mining processing, the results of data mining processing are in the form of a curve that can be presented to readers to estimate that a flood disaster will occur in the future.

In the development of a system for processing Data Mining related to predicting floods in DKJ (Special Region of Jakarta) it will be presented in the form of a dashboard, the Dashboard supports displaying data so that the public can see, the management of the Dashboard stage requires a more detailed discussion in the latest journal by the researcher.

The researcher understands that there are still shortcomings in writing scientific journal reports related to flood predictions in DKJ (Special Region of Jakarta), the researcher hopes that the journal can be developed again as long as the researcher's name is listed as the latest scientific journal that can be presented to the wider community.

REFERENCES

[1] J. Pendidikan *et al.*, "ANALISIS PENYEBAB BANJIR DI DKI JAKARTA", doi: 10.21009/PLPB.221.05.

[2] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, pp. 574–583, 2022, doi: 10.5281/zenodo.7052276.

[3] Y. S. Sari, "Penerapan Metode Naïve Bayes Untuk Mengetahui Kualitas Air Di Jakarta," *Jurnal Ilmiah FIFO*, vol. 13, no. 2, p. 222, Nov. 2021, doi: 10.22441/fifo.2021.v13i2.010.

[4] P. Prakoso and H. Herdiansyah, "ANALISIS IMPLEMENTASI 30% RUANG TERBUKA HIJAU DI DKI JAKARTA," *MAJALAH ILMIAH GLOBE*, vol. 21, no. 1, p. 17, Apr. 2019, doi: 10.24895/mig.2019.21-1.869.

[5] A. Taryana, M. Rifa, E. Mahmudi, and H. Bekti, "ANALISIS KESIAPSIAGAAN BENCANA BANJIR DI JAKARTA," 2022.

[6] S. Sandiwarno, "Penerapan Machine Learning Untuk Prediksi Bencana Banjir," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 1, pp. 62–76, Jan. 2024, doi: 10.21456/vol14iss1pp62-76.

[7] M. Safii, B. Efendi Damanik, and G. Artikel, "Algoritma Naïve Bayes Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar Naïve Bayes Algorithm For Predicting Sales at the Pematang Siantar VJCakes Store Article Info ABSTRAK," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 4, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i4.1674.

[8] K. Steven, S. Hariyanto, R. Arijanto, and A. H. Wijaya, "PENERAPAN BUSINESS INTELLIGENCE UNTUK MENGANALISIS DATA PADA PT. SURYAPLAS INTITAMA MENGGUNAKAN MICROSOFT POWER BI," 2021. [Online]. Available: https://jurnal.buddhidharma.ac.id/index.php/algor/index

[9] R. Maharani, A. Hutagaol, V. Tesalonika Lana, Z. A. Dzunnurain, and R. Kurniawan, "Penerapan Machine Learning dalam Prediksi Klasifikasi Big Data Kedalaman Gempa Bumi di Indonesia

Tahun 2015-2024," *Seminar Nasional Sains Data*, vol. 2024.

[10] Aisyah, N. Fahira, and R. Nooraeni, "Optimasi Parameter ST-DBSCAN dengan KNN dan Algoritma Genetika Studi Kasus: Data Bencana Alam di Pulau Jawa 2021."

[11] S. M. Natzir, "Perbandingan Kinerja Model Pembelajaran Mesin dalam Prediksi Banjir menggunakan KNN, Naive Bayes, dan Random Forest," p. 59, doi: 10.52972/hoaq.vol14no1.p59-64.

[12] A. Zumarniansyah, R. Pebrianto, ; Normah, and W. Gata, "TWITTER SENTIMENT ANALYSIS OF POST NATURAL DISASTERS USING COMPARATIVE CLASSIFICATION ALGORITHM SUPPORT VECTOR MACHINE AND NAÏVE BAYES." [Online]. Available: www.nusamandiri.ac.id

[13] A. Frenica and S. Soim, "Implementasi Algoritma Support Vector Machine (SVM) untuk Deteksi Banjir," vol. 8, no. 2, p. 2023.

[14] I. Sugiyarto, R. Irawan, and D. Rosiyadi, "Kota Jakarta Timur, telp (021) 8462039 dari Universitas Nusa Mandiri; Nusa Mandiri," RW, 2021. [Online]. Available: http://ejurnal.ubharajaya.ac.id/index.php/JSRCS

[15] D. Susanti and T. Wahyuni, "ANALISIS POTENSI BENCANA ALAM TANAH LONGSOR KABUPATEN MAJALENGKA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER," *INFOTECH journal*, vol. 9, no. 2, pp. 299–306, Jul. 2023, doi: 10.31949/infotech.v9i2.5645.

[16] "garuda2346869".

[17] M. A. Wicaksono, C. Rudianto, and P. F. Tanaem, "Rancang Bangun Sistem Informasi Arsip Surat Menggunakan Metode Prototype," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 2, Aug. 2021, doi: 10.28932/jutisi.v7i2.3664.

[18] N. Wijaya, "Pendekatan Multidisiplin Ilmu dalam Manajemen Bencana IMPLEMENTATION OF NAIVE BAYES CLASSIFICATION ALGORITHM FOR OCCUPANCY HOUSE DATA STATUS FUND ASSISTANCE REHABILITATION AND RECONSTRUCTION POST OF ERUPTION DISASTERS MERAPI MOUNTAIN 2010."

[19] A. M. Maksun, Y. A. Sari, and B. Rahayudi, "Analisis Sentimen pada Twitter Bencana Alam di Kalimantan Selatan menggunakan Metode Naïve Bayes," 2021. [Online]. Available: http://j-ptiik.ub.ac.id

[20] "1169-2975-1-SM".

[21] "KINERJA ALGORITMA SUPPORT VECTOR MACHINE."

[22] K. Redaksi, A. Af Supianto, B. Riset dan Inovasi Nasional BRIN, E. Rosana Widasari, and I. Wahyuni, "UB Oicial BITS Webmail UB News Dewan Editor", doi: 10.25126/jtiik.2023105.

[23] "222-Article Text-1187-1268-10-20200819".

[24] M. Fadawkas Oemarki, M. Dimas Mufti Baskara, and I. Ernawati, "PERBANDINGAN AKURASI METODE SUPPORT VECTOR MACHINE DAN K- NEAREST NEIGHBOUR DALAM PREDIKSI CURAH HUJAN POTENSI BANJIR," 2024. [Online]. Available: https://dataonline.bmkg.go.id/home.

[25] H. Tantyoko, D. Kartika Sari, and A. R. Wijaya, "PREDIKSI POTENSIAL GEMPA BUMI INDONESIA MENGGUNAKAN METODE RANDOM FOREST DAN FEATURE SELECTION," 2023. [Online]. Available: http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexHenriTantyoko|http://jo m.fti.budiluhur.ac.id/index.php/IDEALIS/index|

[26] G.-H. Di, S. Utara, M. Sinambela □, and Y. S. Suharini, "PENDEKATAN AI DAN DATA SAINS DALAM BENCANA," vol. 4, no. 1, 2024, doi:10.46880/tamika.Vol4No1.pp152-158.

[27] S. Untuk, P. Bencana, B. Di, L. Basah, and M. R. Faisal, "PEMANFAATAN MACHINE LEARNING SEBAGAI SENSOR BERBASIS MEDIA," 2023. [Online]. Available: https://www.researchgate.net/publication/372677757

[28] V. J. Rivaldo, T. A. Y. Siswa, and W. J. Pranoto, "Perbaikan Akurasi Naïve Bayes dengan Chi-

Square dan SMOTE Dalam Mengatasi High Dimensional dan Imbalanced Data Banjir," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, p. 1656, Jul. 2024, doi: 10.30865/mib.v8i3.7886.

[29] D. Firdaus, "Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer," 2017.