Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13973
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Model Random Forest and Support Vector Machine for Flood Classification in Indonesia

Sintia Eka Purwati[1)*], Yoga Pristyanto[2)]
[1)2)]Universitas Amikom Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
[1)]sintiaekapurwati@students.amikom.ac.id, [2)]yoga.pristyanto@amikom.ac.id

Submitted :Aug 3, 2024 | **Accepted** : Aug 27, 2024 | **Published** : Oct 2, 2024

Abstract: People, especially those living in lowland areas and along rivers. This flood phenomenon significantly affects various aspects, both in terms of economics, environment, and public safety. Flooding is a disaster that often causes problems for most people, especially those living in lowland areas and on riverbanks. This flood phenomenon significantly affects various aspects, such as the economy, environment, and community safety. This research compares the Random Forest and Support Vector Machine (SVM) methods for flood classification in Jakarta. The data used is flood data from 2016 – 2020 in Jakarta, obtained from Kaggle. Model performance evaluation is carried out using accuracy, precision, recall, and F1- Score metrics. The analysis results show that both models accurately classification floods, with Random Forest showing a more stable performance than SVM.

Keywords: Flood Classification, Flood Prediction, Jakarta Flood, Model Evaluation, Random Forest, Support Vector Machine (SVM)

INTRODUCTION

Floods are one of the disasters that often cause problems for most people, especially those living in lowland areas and along rivers. This flood phenomenon significantly affects various aspects, both in terms of economics, environment, and public safety (Hasanah et al., 2021). One of the factors causing flooding is high rainfall. When heavy rain occurs, the water flowing through river channels can exceed its capacity, causing the river to overflow and inundate the surrounding lowlands.

Flooding has been a severe problem in the Special Region of Jakarta for years. During the rainy season, various areas in Jakarta are often hit by floods, which cause material loss and loss of life and hamper community activities (Triyanto et al., 2021). This problem is increasingly complex and requires a comprehensive and sustainable solution. Floods in Jakarta are not just an inevitable natural phenomenon. This problem is the result of a combination of various interrelated factors.

One of the leading causes of flooding in Jakarta is geographical factors. Jakarta is located in the lowlands with an average height of only 5 meters above sea level, so it is very vulnerable to waterlogging, especially during high rainfall. Apart from that, the condition of rivers in Jakarta could be better due to sedimentation and accumulation of rubbish, hampering water flow and thereby increasing the risk of flooding (Wasis et al., 2024). Therefore, routine cleaning and river rehabilitation efforts are crucial in reducing the impact of flooding in Jakarta and implementing flood control technology, such as infiltration wells, to manage water flow in low-lying areas effectively.

Various studies have examined methods for predicting and managing flood risk in Jakarta. (Hamami & Dahlan, 2022) implemented a Random Forest model with oversampling techniques to classify weather in DKI Jakarta Province, which showed the algorithm had an accuracy of 82%. (Primajaya & Sari, 2018) studied the use of the random forest algorithm to predict rainfall, showing that the algorithm has an accuracy of 83%, which can be applied in predicting floods in Jakarta. (Sandiwarno, 2024) compared the Decision Tree, Random Forest, and Naïve Bayes algorithms to predict flooding in Dayeuhkolot Village, finding that each algorithm had certain advantages and limitations. In addition, (Arya Darmawan et al., 2023) uses an Ensemble Machine Learning Technique that combines BP-NN (Back Propagation Neural Network) and SVM (Support Vector Machine) to predict floods, showing

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13973
e-ISSN : 2541-2019
p-ISSN : 2541-044X

that this approach can improve prediction accuracy compared to a single algorithm. (Fitriyaningsih & Basani, 2019) examined the application of various Machine Learning algorithms to predict flood disasters, emphasizing the importance of using advanced technologies in disaster mitigation.

Therefore, an in-depth analysis of the crucial factors affecting flooding in Jakarta is needed to design more efficient strategies. Additionally, analysis is needed to identify priority areas to focus on optimal resource allocation and prioritize preventive actions in areas that most need flood prevention. This study uses the Random Forest and Support Vector Machine (SVM) methods to determine the most effective method and achieve the best performance classification floods in Indonesia.

METHOD

This research was carried out using data acquisition, pre-processing, classification, and evaluation stages. Each stage is explained in Figure 1.

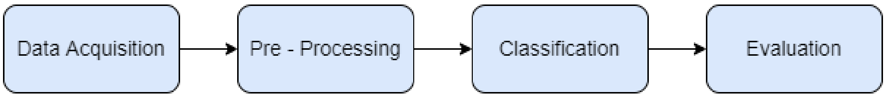

Figure 1. Research Stages

Data Acquisition

The data used in the research is flood data in Jakarta during the 2016 - 2020 period obtained from Kaggle. This dataset can be accessed via the following link. https://www.kaggle.com/datasets/christopherrichardc/climate-and-flood-jakarta/
The data includes essential parameters that influence the occurrence of flooding in Jakarta. Such as rainfall levels, river water levels, wind speed, temperature, and the area affected by flooding. These attributes are critical for analyzing factors that contribute to flood events. Table 1 contains the dataset attributes used.

Table I. Dataset Attributes

Attribute	Description
Date	Date, to determine the time pattern of flooding
Tn	Minimum temperature in Celsius
Tx	Maximum temperature in Celsius
Tavg	Average temperature in Celsius
RH_avg	Average air humidity in percentage
RR	Rainfall in millimeters recorded in a certain period
ss	Duration of sunlight in hours
ff_x	Maximum wind speed in meters per second
ddd_x	Maximum recorded wind direction in degrees
ff_avg	Average wind speed in meters per second
ddd_car	Main wind direction in degrees
station_id	ID of the station that recorded the data
station_name	The name of the station that recorded the data
region_name	The name of the region where the data is recorded
flood	Flood indicator (0 = no flood, 1 = flood)

Pre-processing

The pre-processing stage is a critical step in data analysis, which includes several key processes that ensure the data is clean, accurate, and ready for meaningful analysis. This stage includes data selection, transformation, conversion, and handling of missing values.

First, data selection is the process of determining which data is relevant and should be used for further analysis. This step involves identifying and extracting the most relevant variables or data that suit the research objectives, ensuring that only data that is necessary and useful for the analysis is included. This helps simplify the analysis process and avoid unnecessary complications from extraneous data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13973
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Second, data transformation involves modifying data representations to simplify, enhance, or make them more suitable for analysis. This includes normalizing values, creating new variables from existing data, or converting categorical data to a numeric format. The aim of this step is to improve the quality and understanding of the information, making it easier to work with and interpret.

Third, data conversion is a step that involves changing data into a consistent and appropriate analysis format. This includes changing data types, such as converting text data to numeric form or ensuring all date and time values are standardized to a common format. This step is important to ensure that all data points can be processed uniformly by analysis tools and algorithms, thereby reducing errors and inconsistencies.

Finally, handling missing values is an important data cleansing step to address incomplete, missing, or invalid data entries. This can involve techniques such as deleting or imputing missing data points, or replacing them with statistical measures such as the mean, median, or mode. Handling missing values is critical to maintaining the integrity of the data set and ensuring that analysis results are reliable and not affected by data gaps or inaccuracies (Nurmasani & Pristyanto, 2021) (Putriana et al., 2024).

Classification

This research uses two models: Random Forest and Support Vector Machine (SVM). Random Forest is an ensemble algorithm used to classify weather and rainfall data. This algorithm builds several decision trees during training and outputting a class of modes (Dwiasnati & Devianto, 2021). Random Forest's advantages in handling large and complex data make it the right choice for weather data classification.

Support Vector Machine (SVM) will be used in this research for classification and regression. SVM was chosen because of its high ability to produce accurate predictions by maximizing the margin of separation between data classes. SVM works by finding the best hyperplane that separates classes in feature space and has advantages in handling high-dimensional data and cases where the number of features is greater than the number of samples (Rahmat et al., 2023).

Evaluation

The final stage of this research is to evaluate the algorithm using historical data on flood events. The results of a data classification process can be categorized into four types: false positive (FP), which is the number of negative records classified as positive. False negative (FN) is the number of positive records classified as negative, True negative (TN) is the number of negative records classified as negative. True Positives (TP) is the number of positive records classified as positive. Table 2 below is a confusion matrix table (Widjiyati, 2021).

Table II. Confusion Matrix

Actual	**Prediction**	
	TRUE	**FALSE**
TRUE	TP	FN
FALSE	FP	TN

Evaluation is carried out using several metrics, namely accuracy, precision, recall, and F1-Score. The following is the formula for each evaluation matrix used.

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Accuracy measures the proportion of correct predictions out of all predictions (Widjiyati, 2021).

Precision

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13973
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Precision measures how many positive predictions are correct out of all positive predictions (Azimah & Rizky Nova Wardani, 2022).

Recall

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Recall measures how many true positive cases were detected out of all true positive cases (Utiarahman & Pratama, 2024).

F1-Score

$$F1 - Score = 2\ X\ \frac{Precision\ x\ Recall}{Precision+Recall} \quad (4)$$

F1-Score is the average of precision and recall which provides a balanced picture between the two (Handayani & Fauzan, 2024).

RESULT

Data Acquisition

The research uses data originating from Kaggle, namely Jakarta flood data in the 2016 - 2020 period.

Table III. Dataset Kaggle

date	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car	station_id	station_name	region_name	flood
01/01/2016	26.0	34.8	28.6	81.0	5.8	5.0	280.0	2.0	S	96733	Stasiun	Klimatologi	Banten	Jakarta
02/01/2016	25.6	33.2	27.0	88.0	1.6	8.7	4.0	290.0	2.0	W	96733	Stasiun	Klimatologi	Banten
03/01/2016	24.4	34.9	28.1	80.0	33.8	5.4	4.0	280.0	2.0	SW	96733	Stasiun	Klimatologi	Banten
04/01/2016	24.8	33.6	29.2	81.0	6.6	3.0	200.0	1.0	S	96733	Stasiun	Klimatologi	Banten	Jakarta
05/01/2016	25.8	33.6	26.7	91.0	3.2	3.0	180.0	1.0	S	96733	Stasiun	Klimatologi	Banten	Jakarta
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
27/12/2018	23.8	32.0	28.0	70.0	12.0	180.0	5.0	W	96747	Halim	Perdana	Kusuma	Jakarta	Jakarta
28/12/2018	24.0	33.4	28.5	69.0	14.0	250.0	3.0	SE	96747	Halim	Perdana	Kusuma	Jakarta	Jakarta
29/12/2018	25.2	33.4	28.7	70.0	14.0	120.0	5.0	SW	96747	Halim	Perdana	Kusuma	Jakarta	Jakarta
30/12/2018	24.0	34.4	30.0	64.0	14.0	240.0	5.0	W	96747	Halim	Perdana	Kusuma	Jakarta	Jakarta
31/12/2018	25.4	32.8	28.2	69.0	9.9	14.0	180.0	5.0	SE	96747	Halim	Perdana	Kusuma	Jakarta

Pre-processing

At the data selection stage, the dataset consists of 6308 data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13973
e-ISSN : 2541-2019
p-ISSN : 2541-044X

```
date             object
Tn              float64
Tx              float64
Tavg            float64
RH_avg          float64
RR              float64
ss              float64
ff_x            float64
ddd_x           float64
ff_avg          float64
ddd_car          object
station_id        int64
station_name     object
region_name      object
flood             int64
dtype: object
```

Figure 2. Dataset Variables

Information :
Float = Decimal number.
Object = Data type for strings or various other types of data.
Int = Data type for round numbers (integer) without a decimal component.

Data transformation is carried out using several processes, namely creating new columns from date data to year, month and day as well as deleting columns that are not relevant for further analysis.

```
Informasi Data setelah Transformasi:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6308 entries, 0 to 6307
Data columns (total 14 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Tn            5996 non-null   float64
 1   Tx            6095 non-null   float64
 2   Tavg          6262 non-null   float64
 3   RH_avg        6256 non-null   float64
 4   RR            3993 non-null   float64
 5   ss            5049 non-null   float64
 6   ff_avg        6215 non-null   float64
 7   ddd_car       6207 non-null   object
 8   station_name  6308 non-null   object
 9   region_name   6308 non-null   object
 10  flood         6308 non-null   int64
 11  year          6308 non-null   int32
 12  month         6308 non-null   int32
 13  day           6308 non-null   int32
dtypes: float64(7), int32(3), int64(1), object(3)
memory usage: 616.1+ KB
None
```

Figure 3. Data Transformation

The data conversion process involves replacing variables with a categorical data type into a numeric data type. This process is essential to ensure that the machine-learning model can work effectively because most machine-learning algorithms require input in the form of numerical data. The variables station_id, ddd_x, and ff_x are not used in the analysis because they are irrelevant variables.

```
     Tn    Tx  Tavg  RH_avg    RR   ss  ff_avg  ddd_car  station_name  \
0  26.0  34.8  28.6    81.0   NaN  5.8     2.0      0.0             0
1  25.6  33.2  27.0    88.0   1.6  8.7     2.0      1.0             0
2  24.4  34.9  28.1    80.0  33.8  5.4     2.0      2.0             0
3  24.8  33.6  29.2    81.0   NaN  6.6     1.0      0.0             0
4  25.8  33.6  26.7    91.0   NaN  3.2     1.0      0.0             0

   region_name  flood  year  month  day
0            0      0  2016      1    1
1            0      1  2016      1    2
2            0      1  2016      1    3
3            0      0  2016      1    4
4            0      0  2016      1    5
```

Figure 4. Data Conversion

The data used has 15 variables, and there are several Missing Values , such as empty data and unavailable values, which are then filled in with the average value for each month. This process involves

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13973
e-ISSN : 2541-2019
p-ISSN : 2541-044X

calculating the average of each variable that has missing values for each month and then replacing the missing values with the average value. So there are no more missing values in the data.

```
     Tn         Tx  Tavg  RH_avg         RR    ss  ff_avg  ddd_car  \
0  26.0  34.800000  28.6    81.0   8.309639   5.8     2.0      0.0
1  25.6  33.200000  27.0    88.0   1.600000   8.7     2.0      1.0
2  24.4  34.900000  28.1    80.0  33.800000   5.4     2.0      2.0
3  24.8  33.600000  29.2    81.0   8.309639   6.6     1.0      0.0
4  25.8  33.600000  26.7    91.0   8.309639   3.2     1.0      0.0
5  25.0  33.600000  28.9    80.0   3.800000   3.6     2.0      3.0
6  25.2  35.600000  30.0    78.0   8.309639   7.6     2.0      3.0
7  26.8  32.526316  29.9    79.0   8.309639   7.9     2.0      4.0
8  26.8  34.800000  29.7    79.0   8.309639  10.2     2.0      3.0
9  25.4  34.600000  29.2    80.0   8.309639   8.2     2.0      0.0

   station_name  region_name  flood  year  month  day
0             0            0      0  2016      1    1
1             0            0      1  2016      1    2
2             0            0      1  2016      1    3
3             0            0      0  2016      1    4
4             0            0      0  2016      1    5
5             0            0      0  2016      1    6
6             0            0      0  2016      1    7
7             0            0      0  2016      1    8
8             0            0      0  2016      1    9
9             0            0      0  2016      1   10
```

Figure 5. Data Cleaning

Classification and Evaluation

Model performance evaluation uses historical data with the primary metrics of accuracy, precision, recall, and F1-score. The following are the evaluation results of the two models.

Table IV. Random Forest Model Evaluation

Kelas	Precision	Recall	F1-Score	Support
0	0.92	0.99	0.96	1154
1	0.50	0.06	0.10	108
accuracy	0.91			
macro avg	0.71	0.53	0.53	1262
weighted avg	0.88	0.91	0.88	1262

Based on the evaluation results, the Random Forest model has an accuracy of 91%. With Precision, Recall, and F1 Score values close to 1.00, it shows that the model works well in predicting floods.

Table V. SVM Model Evaluation

Kelas	Precision	Recall	F1-Score	Support
0	0.91	1.00	0.96	1154
1	1.00	0.00	0.00	108
accuracy	0.91			
macro avg	0.96	0.50	0.48	1262
weighted avg	0.92	0.91	0.87	1262

The results of testing the Support Vector Machine (SVM) model show that the model has an accuracy of 91%. However, the Precision, Recall, and F1 Score values for certain classes indicate an imbalance in predictions.

DISCUSSIONS

In selecting the best model for classification, two important metrics to consider are accuracy and precision values. accuracy measures the proportion of correct predictions made by the model, while precision focuses on the model's ability to correctly identify positive classes among all positive predictions generated. High accuracy shows that the model is able to make predictions correctly, while high precision shows that the model can specifically differentiate positive and negative classes well.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.13973 p-ISSN : 2541-044X

Based on the results of the evaluation carried out, the Random Forest and Support Vector Machine (SVM) models have high accuracy values, which means they are both able to handle data well. However, there is a significant difference in performance stability between the two models. Random Forest shows more stable performance than Support Machine Learning (SVM) models in several evaluation metrics such as recall and F1-score.

The advantage of Random Forest comes from an ensemble approach that combines many decision trees, making it more resistant to overfitting and noise in the data. In contrast, Support Vector Machine (SVM) is more sensitive to parameter selection and requires more careful tuning to achieve optimal performance. Random forest is more computationally efficient, especially on large datasets. Taking into account the stability of more consistent and efficient performance, Random Forest is considered a more reliable and versatile model than Support Vector Machine (SVM).

CONCLUSION

Based on the results of the analysis, this research shows that rainfall is the main factor causing flooding in Jakarta. Random Forest and Support Vector Machine (SVM) models are applied to predict flood events based on available data. The evaluation results show that both models have a high level of accuracy in classification floods, with an accuracy of 91%. However, the Random Forest model not only provides high accuracy, but also shows more stable performance on various evaluation metrics, such as precision, recall, and F1 score. This shows that although both models are effective in classification floods, Random Forest provides more reliable results in various data conditions. For future research, it is recommended to include additional variables such as drainage capacity, land use patterns, and elevation data to improve the accuracy and robustness of the model. In addition, extending the data collection period and using more comprehensive data preprocessing techniques is expected to further improve model performance. Exploration of other advanced models such as Gradient Boosting or Neural Networks is also recommended to provide a more comprehensive comparison and potential improvement in flood classification performance.

REFERENCES

Azimah, F., & Rizky Nova Wardani, K. (2022). Sistem Pendeteksi Gejala Awal Covid-19 dengan Penggunaan Metode Al Project Cycle. *Journal Locus Penelitian Dan Pengabdian*, *1*(6), 405–418. https://doi.org/10.36418/locus.v1i6.135

Darmawan, M. B. A., Dewanta, A., & Astuti, S. (2023). Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot. *TELKA - Telekomunikasi Elektronika Komputasi Dan Kontrol*, *9*(1), 52–61. https://doi.org/10.15575/telka.v9n1.52-61

Dwiasnati, S., & Devianto, Y. (2021). Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir. *Prosiding SISFOTEK*, 202–207. http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/283

Fitriyaningsih, I., & Basani, Y. (2019). Flood Prediction with Ensemble Machine Learning using BP-NN and SVM. *Jurnal Teknologi Dan Sistem Komputer*, *7*(3), 93–97. https://doi.org/10.14710/jtsiskom.7.3.2019.93-97

Hamami, F., & Dahlan, I. A. (2022). Klasifikasi Cuaca Provinsi Dki Jakarta Menggunakan Algoritma Random Forest Dengan Teknik Oversampling. *Jurnal Teknoinfo*, *16*(1), 87. https://doi.org/10.33365/jti.v16i1.1533

Handayani, P., Fauzan, A. C., & Harliana. (2024). Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random. *Klik : Kajian Ilmiah Informatika Dan Komputer*, *4*(6), 3064–3072. https://doi.org/10.30865/klik.v4i6.1909

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, *5*(2), 103–108. https://doi.org/10.30871/jaic.v5i2.3200

Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class. *Pseudocode*, *8*(1), 21–26. https://doi.org/10.33369/pseudocode.8.1.21-26

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 4, October 2024
DOI : https://doi.org/10.33395/sinkron.v8i4.13973
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Primajaya, A., & Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, *1*(1), 27. https://doi.org/10.24014/ijaidm.v1i1.4903

Putriana, P., Suarna, N., & Prihartono, W. (2024). Analisis Clustering Prestasi Atlet Pada Berbagai Cabang Olahraga Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3435–3442. https://doi.org/10.36040/jati.v7i6.8211

Rahmat, A. ., Ladjamuddin, M. ., & Awaludin, T. . (2023). Perbandingan Algoritma Decision Tree, Random Forest Dan Naive Bayes Pada Prediksi Penilaian Kepuasan Penumpang Maskapai Pesawat Menggunakan Dataset Kaggle. *Jurnal Rekayasa Informasi*, *12*(2), 150–159. www.kaggle.com,

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

Sasoko, W. H., Pujiharto, E. W., Haris, R., Kania, A. Y., Kusrini, & Kusnawi. (2024). Prediksi Banjir Di Dki Jakarta Dengan Menggunakan Algoritma K-Means Dan Random Forest. *Jurnal Informatika Dan Teknologi Komputer*, *5*(1), 43–49. https://ejurnalunsam.id/index.php/jicom/

Triyanto, S., Sunyoto, A., & Arief, M. R. (2021). Analisis Klasifikasi Bencana Banjir Berdasarkan Curah Hujan Menggunakan Algoritma Naïve Bayes. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, *5*(2), 109–117. https://doi.org/10.35145/joisie.v5i2.1785

Utiarahman, S. A., & Pratama, A. M. M. (2024). Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(6), 3137–3146. https://doi.org/10.30865/klik.v4i6.1871

Widjiyati, N. (2021). Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval. *Jurnal Janitra Informatika Dan Sistem Informasi*, *1*(1), 1–7. https://doi.org/10.25008/janitra.v1i1.118