



Terbit online pada laman web jurnal :
<http://ejournal.amikompurwokerto.ac.id/index.php/telematika/>

Telematika

Accredited SINTA “2” Kemenristek/BRIN, No. 85/M/KPT/2020



Website-Based Application for Flood Event Prediction Using Machine Learning Method In Cilacap District

Imam Tahyudin¹, Faiz Ichsan Jaya², Nur Faizah³

^{1,3}Department of Information System, Faculty of Computer Science

²Department of Informatics, Faculty of Computer Science

Universitas Amikom Purwokerto, Purwokerto, Central Java, Indonesia

Email: imam.tahyudin@amikompurwokerto.ac.id¹, faizichsanj@gmail.com²,

nurfaizah@amikompurwokerto.ac.id³

ARTICLE INFO

History of the article:

Received January 19, 2021

Revised January 27, 2022

Accepted February 12, 2022

Keywords:

Website,
Machine Learning,
SVM,
Random Forest

Correspondence:

E-mail:
imam.tahyudin@amikompurwokerto.ac.id

ABSTRACT

Floods are the most common natural disasters. Their intensity at a place and the number of locations of events are 40% among other natural disasters. The impact of flooding on the area in general is temporary housing in rural areas caused by flooding in addition to settlement as well as agriculture which can have an impact on the food security of the area and also a national level that is higher than the magnitude of the country. Based on data from the Badan Pusat Statistik (BPS) of Cilacap Regency, the number of flood victims in Cilacap Regency in 2018 reached 771 people and arranged for them to flee from the flood. To solve this problem, we conducted research to create a web-based application using the classification of the Support vector machine and Random Forest to predict flood events and compare the accuracy values of the two algorithms to get better prediction results.

INTRODUCTION

Rainfall is one of the elements in climate. Rainfall is also a factor that directly influences changes in weather, both changes in good weather and changes in bad weather. Rainfall is the amount of water that falls to the ground surface which is then calculated using units in millimeters above the flat ground level (M. Feridiansyah et al., 2015). The weather conditions that sometimes change quickly and extreme, will cause the community to be less responsive to the impact it causes. Extreme weather changes also affect various fields such as agriculture, aviation, animal husbandry and so on. In addition to weather, flooding is also a serious disruption to the functioning of a community, causing widespread damage to human life, both in terms of economy, society, and the environment (Rahajoeningoem and Saputra, 2017).

Based on data from the Badan Pusat Statistik (BPS) Cilacap Regency, the number of flood victims in Cilacap Regency in 2018 reached 771 people and required them to flee from the impact of the flood (BPS of Central Java Province, 2018). Therefore, the research to predict the flooding event is so important to prevent the more victims. To find out the symptoms of flooding, we obtain the data from Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). BMKG provides rainfall estimates for the community to avoid the flood that will befall the community.

The data used in this study is the rainfall data. The available data can be extracted using data mining method (Gorunescu, 2011). The task of predictive data mining is to draw conclusions from the latest data to

make predictions. Furthermore, the classification methods used were Support Vector Machine (SVM) and Random Forest (RF) methods. The SVM method has the advantage of being able to generalize or be able to classify data that does not belong to training data correctly. The level of generalization is not affected by the observed dimensions of variables and faster computing processes because only 2 involves a few observations in the formation of functions his decision (Fadli, 2010) (Siregar, 2017). Whereas Random Forest is based on decision tree techniques so as to overcome non-linear problems. To identify explanatory variables that are relevant to the response variable, the random forest produces a measure of the importance of the explanatory variable (Breiman, & Cutler, 2005) (Uriarte, & Andrés, 2006) (Adnyana, 2015). We compare the highest level of accuracy in the SVM and Random Forest models to be implemented on the website.

The previous research about Support Vector Machine (SVM) and Random Forest has been conducted in Breast Cancer Diagnosis. It aims to compare the two algorithms and find the highest accuracy value. The accuracy value generated on the Random Forest (RF) algorithm gives a much higher result by 100% while the Support Vector Machine (SVM) model in predicting training data gets an accuracy rate of 94% (Aliady, Tuasikal, and Widodo, 2018). Next, on web-based application for prediction rainfall, water discharge, and flood events using machine learning in Deli Serdang. The SVM and BP-NN algorithms were implemented for the calculation of rainfall data from Kualanamu, Tuntung and Sampali stations and the Ular River water discharge from 1 January 2016 to October 22, 2017. The accuracy of the prediction of the classification of flood days and not floods from this application is 94.4% but has not been able to predict the occurrence of floods (Fitriyaningsih, Basani, and Ginting, 2018). Furthermore, SVMs are today know as robust and efficient Machine Learning (ML) algorithms for flood prediction. SVM and SVR emerged as alternative ML methods to ANNs, with high popularity among hydrologists for flood prediction. They use the statistical learning theory of structural risk minimization (SRM), which provides a unique architecture for delivering great generalization and superior efficiency. Most importantly, SVMs are both suitable for linear and nonlinear classification, and the efficient mapping of inputs into feature spaces. Thus, they were applied in numerous flood prediction cases with promising results, excellent generalization ability, and better performance, compared to ANNs and MLRs (Mosavi et al, 2018).

Therefore, this paper presents a prediction of flooding event in Cilacap Regency using classification methods, SVM and RF. The best accuracy of both methods will be used for implementing in the application of flooding prediction. Rest of the paper is organized as: research methods. It elaborates the method of machine learning and implementation on the website using Flask. After that, the detail description of the method used. Finally, we conclude the research paper.

RESEARCH METHODS

The diagram of the research methodology is described in the figure 1.

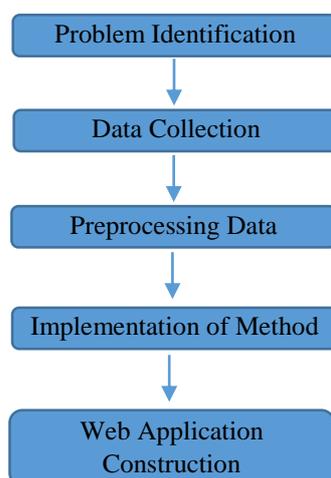


Figure 1. The Diagram of research methodology

The Figure 1 is explained detail as follow:

1. Problem identification

The problem identification process is carried out as an effort to overcome the problem and the appropriate method so that points can be determined to predict potential flood events in Cilacap. Regency.

2. Data collection

In this study the data used are rainfall data provided directly from BMKG Cilacap. The rainfall data was taken from 1980 to 2019. The variables are rainfall intensity and the category. The class variable is category.

3. Preprocessing data

This step is done to get a clean data and ready-to-use for research. Following is a detailed explanation of each stage of the preprocessing flow:

- Preparing the data – The first step that must be done is to prepare the required data. Rainfall data in Microsoft Excel (.XLS) format provided by BMKG Cilacap is then selected.
- Handling the missing value – After the data has been checked and selected, the next step is to handle the missing value. In the rainfall data provided by BMKG Cilacap, there are missing or unrecorded data. Thus, the missing data is handled by removing blank data for further analysis.
- Feature selection – The next step is feature selection. Generally, not all attributes and features will be used in analyzing and researching data. So, this step is to remove the unnecessary feature that is included on the data.

4. Implementation of SVM and RF methods

The data dividing into a training set and testing set. This data will be divided based on the amount of data. Distribution can be done using commands in Python (Sweigart, 2015) (Haverbeke, 2018). Once divided, the values in the data set are compared to feature scaling.

a. Making a model

Machine learning models are made to train data and make predictions. This model was created and tested using the Python programming language and several libraries. At this stage the results of the distribution of training data and testing data that have been made previously will be modeled using the SVM algorithm and Random Forest to see the predicted results.

b. Evaluation

This stage will be done an evaluation and measurement of the accuracy of the values generated by the model that was previously made and see the results of precision and recall.

c. Implementation

The model that has been made which has the best accuracy from both of SVM and RF is implemented in the website system to predict the results.

5. Web Application Construction

The system development method for constructing the web application is using Waterfall. According to Rosa and Saladin (2018), the SDLC model of a waterfall is often also called a sequential linear model or classic life cycle (Rosa & M. Shalahuddin, 2018). The waterfall model provides a sequential or sequential software life cycle approach starting from the analysis, design, coding, testing, and support stages. Here is figure 2 of the waterfall models:

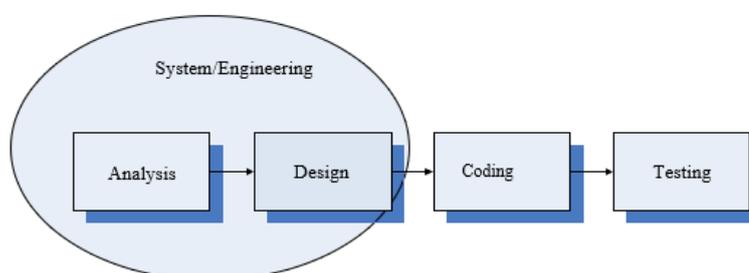


Figure 2. Illustration of a waterfall model (Rosa & M. Shalahuddin, 2018)

The following steps must be performed in the waterfall method in research:

a. Software requirements analysis

The process of gathering needs is done intensively to specify software requirements so that software can be understood as what is needed by the user.

b. Design

Software design is a multi-step process that focuses on the design of software programming including data structures, software architecture, interface representation, and coding procedures

c. Coding a program

The result of this stage is a computer program in accordance with the design that was created at the design stage.

d. Testing

Testing focuses on software in a logical and functional way and ensures that all parts have been tested.

RESULTS AND DISCUSSION

1. Data mining process

1) Problem identification

In this stage the authors conducted several preliminary studies by studying the literature relating to the author's research, namely to predict flood events and their relation to rainfall, as well as the selection of algorithms and methodologies suitable for this author's research.

2) Data collection

In this study, the authors use rainfall data provided by BMKG Cilacap by directly coming to the BMKG Cilacap office. This data is recorded as many as 5083 which consists of 7 features including station name, latitude, longitude, height, year, month, amount of rainfall. In this case the writer only uses 3 features, namely year, month, amount of rainfall.

Table 1. Category of the amount of rainfall

No	Amount of rainfall	Category
1	0 – 20 mm	Low
	21 – 50 mm	
	51 – 100 mm	
2	101 mm – 150 mm	Average
	151 mm – 200 mm	
	201 mm – 300 mm	
3	301 mm – 400 mm	High
	401 mm – 500 mm	
4	More than 500 mm	Very high

The data is not directly used but must go through processing first, namely the preprocessing stage, where this stage aims to adjust the attributes that will be used in processing the dataset.

Table 2. Daily rainfall data (millimeters) in Cilacap Regency

No	Station	Year	Month	Rainfall
1	BINANGUN	1980	January	315
2	BINANGUN	1980	February	123
3	BINANGUN	1980	March	334
4	BINANGUN	1980	April	246
5	BINANGUN	1980	May	65
...
...
...
5799	WANAREJA	2019	September	X
5800	WANAREJA	2019	October	X
5801	WANAREJA	2019	November	X
5802	WANAREJA	2019	December	X
5803	WANAREJA	2019	April	104

3) Preprocessing Data

At this stage, the data that has been collected will be processed before use. In this case, the data we have is certainly not clean and polished as desired. Usually we will have many problems with data quality (missing values, wrong values, etc.). Thus, the preprocessing stage of the data to be used is needed. There are several types of methods that can be used in this stage, the author uses Jupyter Notebook to analyze the data using the Python programming language.

4) Making a model

After dividing the data into testing and training sets. Next is implementing the model in the training data. Scikit-Learn contains the libraries needed to do the classification in this data analysis process. The author uses two algorithms in this study, namely Support Vector Machine and Random Forest. This stage is to create an SVM model, the SVM library from Scikit-learn must be entered first. Because we do the classification, the writer uses support vector classification class written with "SVC" in the library of Scikit-Learn.

The "fit" method of the SVC class is called to train its algorithm in training data, which is passed as a parameter into the "fit" method. Similar to making a Random Forest model, several libraries must be imported first to classify and predict the dataset. The "rfc" variable was created to store the "RandomForestClassifier()" function and then use the fit method to train the algorithm for training data within these parameters. After that the variable "rfc_predictions" is created to store the prediction function and display the prediction results themselves.

5) Evaluation

After building the model, the most important thing that appears is how good the model is. So evaluating the model is the most important task that illustrates how good the prediction is. The author compares the accuracy values between the models using the SVM algorithm with the Random Forest algorithm to get the highest value. If one of the models has a higher accuracy value, then the model is chosen which will be implemented in a website-based application for predicting flood events.

```
from sklearn.metrics import accuracy_score
print('Accuracy score: ', accuracy_score(y_test, y_pred)*100, '%')
Accuracy score: 98.15140845070422 %
```

Figure 3. The accuracy score using SVM algorithm

```
: from sklearn.metrics import accuracy_score
print('Accuracy score: ', accuracy_score(
    y_test, rfc_predictions)*100, '%')
Accuracy score: 1.584507042253521 %
```

Figure 4. The accuracy score using Random Forest algorithm

Based on the figure 3 and figure 4, the performance results using SVM produce an accuracy value of 98.15% with precision and recall of 98%, while the performance results using Random Forest produce an accuracy value of 1.5% with precision and recall each producing a value of 2%. Therefore, the SVM method is chosen for the constructing the web application of flood prediction.

6) Implementation

After the model is created, save the model in a pickle (.pkl) format for use on website implementation. After that copy the file into the project folder and create a new file with the name "script.py" and create 2 new folders with the name "templates" to create HTML and "static" files to store static files such as CSS and image files using Flask.

```
#importing libraries
import os
import numpy as np
import flask
import pickle
import json
import pandas as pd
```

```

from flask import Flask, render_template, request
from datetime import time

#creating instance of the class
app=Flask(__name__)

#to tell flask what url should trigger the function index()
@app.route('/')
# @app.route('/index')
def index():
    return flask.render_template('index.html')

def ValuePredictor(to_predict_list):
    to_predict = np.array(to_predict_list).reshape(1,3)
    loaded_model = pickle.load(open("model.pkl","rb"))
    result = loaded_model.predict(to_predict)
    return result[0]

@app.route('/result',methods = ['POST'])
def result():
    if request.method == 'POST':
        to_predict_list = request.form.to_dict()
        to_predict_list=list(to_predict_list.values())
        to_predict_list = list(map(int, to_predict_list))
        result = ValuePredictor(to_predict_list)
        if int(result)>500:
            prediction='Terjadi potensi banjir'
        else:
            prediction='Tidak terjadi potensi banjir'
        return render_template("result.html",prediction=prediction, result=result)

```

2. Analysis of Results

This stage requires analysis to find out all the needs to gather information to build a flood prediction application using website-based machine learning in Cilacap district. The analysis in this study is divided into two namely Functional Requirements Analysis and Non-Functional Requirements Analysis.

a. Functional Requirements Analysis

This requirements contain anything that will be processed and carried out by the system. Functional requirements in system development are as follows:

- 1) Input needs – input needed to meet the needs in the implementation of this application: the main page that serves to operate the application by inputting the date, month and year for processing the predicted flood events.
- 2) Process needs – prediction Process, used to predict flood events.
- 3) Output needs – In the form of output or prediction of flood events after the process needs to be done.
- 4) Interface needs – the website consist of three main interface namely the input, output and about page.

Figure 4. The main interface for forecasting

Figure 5. The output interface showing the result of prediction

The figure 4 is the main interface for forecasting. This page is filling out the name of station, month, and year. After that push the prediction button. The figure 5 is the output of prediction. Based on the example the flooding prediction at Binangun station on September 25th, 2009. The result obtained that no potential for flooding because the rainfall intensity is 341 mm.

CONCLUSIONS AND RECOMMENDATIONS

This research has succeeded in predicting flood events using Support Vector Machine and Random Forest Algorithms. The performance results of the data set using SVM produce an accuracy value of 98.15% with precision and recall of 98% respectively while the performance results using Random Forest produce an accuracy value of 1.5% with precision and recall each producing a value of 2%. Therefore, for constructing the web application for flood prediction is using SVM method. This research has succeeded in designing and making a website-based application to predict flood events. It would be better if more data is added such as water-flow rate and land surface in the data mining process to get better prediction results.

REFERENCES

- M. Feridiansyah et al., “Analysis Of Algorithm Prediction Of Rainfall Using Algorithm Competitive Neural Network,” vol. 2, pp. 1–8, 2015.
- T. Rahajoeningoem and I. H. Saputra, “Weather Monitoring and Flood Detection System on Android Based on Internet of Things (IoT),” Prosiding SAINTIKS FTIK UNIKOM, vol. 2, 2017.

- Gorunescu, F. 2011. "Data Mining: Concepts, Models and Techniques". Berlin: Springer. doi: [10.1007/978-3-642-19721-5](https://doi.org/10.1007/978-3-642-19721-5)
- BPS of Central Java Province, "Number of Flood Victims by Regency/City in Central Java Province, 2018," 2018.
- Zong C. Q., Xia R., Zhang J. J. (2019). Text Data Mining. Beijing: Tsinghua University Press. doi: [10.1007/978-981-16-0100-2](https://doi.org/10.1007/978-981-16-0100-2)
- A. B. Siregar, "Support Vector Machine Modeling for Classification of Monthly Rainfall in Indramayu district," 2017.
- Breiman, L., & Cutler, A. 2005. "Random Forests". (https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm)
- Díaz-Uriarte, R. & Alvarez de Andrés, S. 2006. "Gene Selection and Classification of Microarray Data Using Random Forest". *BMC Bioinformatics*, 7(1): 3. Doi: [10.1186/1471-2105-7-3](https://doi.org/10.1186/1471-2105-7-3)
- I Made Budi Adnyana, I. M. B. 2015. "The Prediction of Duration Study using Random Forest Method". *CSRID Journal*, Vol. 8 No 3, 201–208. doi: [10.1080/23249935.2020.1733132](https://doi.org/10.1080/23249935.2020.1733132)
- H. Aliady, N. J. Tuasikal, and E. Widodo, "Implementation Of Support Vector Machine (SVM) and Random Forest In Breast Cancer Diagnosis," Seminar Nasional Teknologi Informasi dan Komunikasi 2018 (SENTIKA 2018), pp. 278–285, 2018. doi: [10.1016/j.ejor.2017.12.001](https://doi.org/10.1016/j.ejor.2017.12.001)
- I. Fitriyaningsih, Y. Basani, and L. M. Ginting, "Web- Based Application Development For Predicting Rainfall, Water Discharge, And Flood Using Machine Learning Method In Deli Serdang," *Jurnal Penelitian Komunikasi dan Opini Publik*, vol. 22, pp. 132–143, 2018. doi: [10.33299/jpkop.22.2.1752](https://doi.org/10.33299/jpkop.22.2.1752)
- Mosavi, Amir, Pinar Ozturk, and Kwok-wing Chau. 2018. "Flood Prediction Using Machine Learning Models: Literature Review" *Water* 10, no. 11: 1536. doi: [10.3390/w10111536](https://doi.org/10.3390/w10111536)
- Sweigart, A. 2015. "Automate the Boring Stuff with Python". Tersedia di <https://automatetheboringstuff.com/chapter0/>
- Haverbeke, M. 2018. "Introduction: Eloquent JavaScript". Eloquent JavaScript.
- A.S., Rosa & Shalahuddin, M. 2018. "Software engineering for Structure and Oriented Object". Bandung: Informatika Bandung.