Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Classification of flood disaster level news articles using Machine Learning

Rahmad Santosa[1], Arna Fariza[2], Firman Arifin[3]
hadsen95@gmail.com, arna@pens.ac.id, firman@pens.ac.id
Electronic Engineering Polytechnic Institute of Surabaya

Article Information

Submitted : 29 Dec 2023
Reviewed: 3 Jan 2024
Accepted : 10 Feb 2024

Keywords

Machine Learning, NLP, Disaster, Classification

Abstract

Floods have a significant socio-economic impact on Indonesian society. Much of this information is sourced from online news articles and social media. This research investigates whether the Support Vector Machine (SVM) method can be used for flood disaster level classification (low, medium, and high). Our methodology involves preparing data extracted from textual news articles on the National Disaster Management Agency (BNPB) website on the topic of flooding. We then labeled the data according to Regulation No. 02/2012 on general guidelines for disaster assessment and used the Support Vector Machine (SVM) method. Training and testing were conducted using different datasets, followed by accuracy and error evaluation. In addition, we considered the performance comparison of SVM with other classification methods, including Decision Tree, Naive Bayes, Adaboost, Random Forest, and Xgboost. The experimental results show that SVM still does not get good accuracy results for flood disaster level classification. The SVM accuracy level result of (52%) is still low compared to Random Forest (78%), and Xgboost (68%). Further research is expected to increase the accuracy of SVM for flood level classification.

A. Introduction

Hydrometeorological disasters have a major impact on the social and economic conditions of Indonesian society. Not only that, climate change has the potential to cause losses in Indonesian society. Hydrometeorological disasters themselves, occur due to rainfall and air temperature and the influence of increased carbon dioxide gas which can cause floods, landslides, droughts, and so on [1]. Online news media has become familiar to the public to communicate, get information and disseminate information. In addition, social media is used as an active sensor both when warning of emergency events and during emergency events such as disasters such as floods, landslides, tornadoes and so on [2],[3],[4].

Text Classification is a field of Natural Processing Language (NLP) and is a branch of artificial Intelligence (AI) with one of the methods there is text mining [5],[4]. Text Classification technology aims to analyse the information in the online news editor. Therefore we conducted classification research on emergency information on hydrometeorological disasters in floods in Indonesia. The results of the low, medium, high classification referring to PERKA (Regulation of the Head of the National Disaster Management Agency) Nomor 02 of 2012 on the physical vulnerability component [6] will later be used as emergency treatment and crisis management and individual awareness.

Some related research, such as the Cross-Attention-Multi-modal (CAMM) bid method to classify disaster news, whether the news is informative or non-informative on social media (Twitter) [7],[12]. Then the classification of flood disaster messages that have eyewitnesses or not, the purpose of which is used as a social network sensor which is expected to make a positive contribution to the prevention of worse flood disaster impacts [8]. Furthermore, Machine Learning research uses SVM as a method of classifying the level of disaster areas [15]. The research aims to increase the efficiency of the SAR team in saving individuals affected by the disaster [9],[13].

In this research, we proposed a new approach to classify the level of flood disaster from the narrative text taken from the BNPB Indonesia article portal website. We used machine learning (SVM, Naive Bayes, Decision Tree, AdaBoost, XgBoost, Random Forest) and performed a comparison to classify the level of flood disaster in the article as low, medium, or high. Knowing the level of disaster is expected to increase public awareness of disasters that occur, so that they can take precautions to reduce the adverse effects that occur. And there is nothing wrong if we take care of the environment, for example to reduce the occurrence of floods, global warming is increasing in Indonesia, which later the results will be enjoyed by the younger generation and people in Indonesia and environmental governance will also be better.

B. Research Method

In this section, we will explain the classification stages using the SVM method which can be seen in Figure 1. as for the stages carried out such as crawling data (taking data on the BNPB Indonesia website), preprocessing data, and finally classifying data using SVM.

We use SVM because many researchers have used the algorithm for classification. Classification itself has 2 types of binary classification and multiclass classification which can be seen in Figure 2. Here we use multiclass classification to classify the level of flood disaster (low, medium, high).

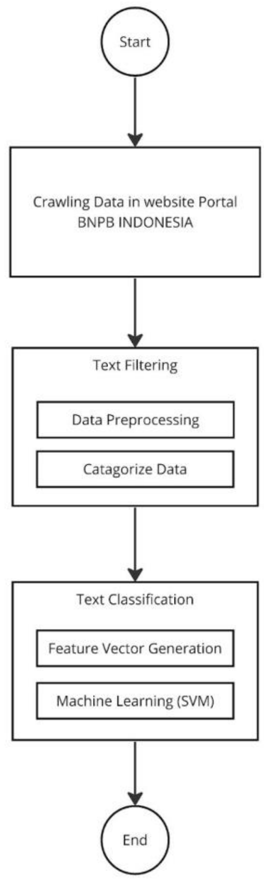

Figure 1. Stages of the research process

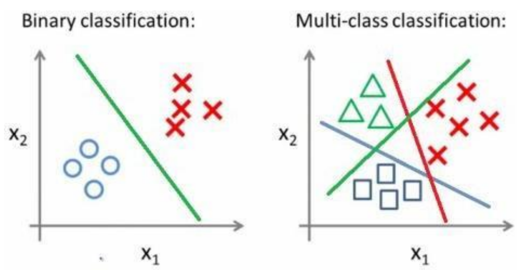

Figure 2. Binary and Multiclass Classification

1. Data Collection

Web crawling is usually referred to as web scraping, which aims to automatically collect data from the internet well in large amounts and quickly [10]. For crawling data we use python, the crawling results that we get with the crawling method get 118 data categorised as floods with column attributes title, vulnerability, details, url in Indonesian language. After the data crawling stage is complete, the text filtering stage will be carried out. The following are samples of data crawling results. For the vulnerability column, is a sentence in which the sentence contains a dictionary unigram word floods in Indonesian ('hilang', 'meninggal', 'tewas', 'mengungsi', 'luka-luka', 'terdampak', 'korban','terendam', 'rusak', 'kerugian', 'genangan', 'menggenangi', 'kerusakan', 'air', 'sawah', 'lahan', 'hancur', 'menerjang'). The results of crawling data can be seen in Figure 3.

	tanggal	judul	vulnerability	detail	url
0	01 Sep 2023 16:32 ...	Banjir Labuhanbatu Akibatkan 2 Warga Meninggal...	mengakibatkan 2 warga meninggal satu korban m...	Foto : Banjir masih melanda Kabupaten Labuhan...	https://bnpb.go.id/berita/banjir-labuhanbatu-a...
1	31 Agt 2023 21:27 ...	1.116 Rumah Terendam Akibat Banjir di Kabupate...	30 hingga 70 cm	Foto : BPBD Kabupaten Langkat melakukan asses...	https://bnpb.go.id/berita/1116-rumah-terendam-...
2	30 Agt 2023 19:38 ...	Banjir Melanda Kabupaten Labuhan Batu, 102 Kel...		Foto : Kondisi banjir yang melanda Kabupaten ...	https://bnpb.go.id/berita/banjir-melanda-kabup...
3	29 Agt 2023 17:02 ...	Sungai Alu Beutong Meluap Sebabkan Banjir di K...		Foto : Kondisi banjir yang melanda wilayah Ka...	https://bnpb.go.id/berita/sungai-alu-beutong-m...
4	28 Agt 2023 13:27 ...	Seorang Warga Meninggal Dunia Terseret Banjir ...	Seorang warga meninggal 4060 sentimeter dari ...	Foto : Kondisi banjir di Bolaang Mongondow Se...	https://bnpb.go.id/berita/seorang-warga-mening...
5	15 Agt 2023 13:32 ...	Meski telah Surut, BPBD Kabupaten Pesisir Sela...	75 cm	Foto : Banjir di wilayah Sumatra Barat telah ...	https://bnpb.go.id/berita/meski-telah-surut-bp...
6	12 Agt 2023 15:49 ...	Banjir di Aceh Selatan Berangsur Surut	cm hingga 60 cm terendam banjirPenanganan	Foto : Kondisi Banjir di Aceh Selatan (11/8) ...	https://bnpb.go.id/berita/banjir-di-aceh-selat...
7	22 Jul 2023 20:14 ...	Sebanyak 125 Jiwa Terdampak Banjir di Kota Palu	cm hingga 70 cm	Foto : Petugas BPBD Kota Palu melakukan Penan...	https://bnpb.go.id/berita/sebanyak-125-jiwa-te...
8	20 Jul 2023 21:00 ...	Arus Deras Banjir Landa Kota Jayapura		Foto : Analisis spasial inaRISK mengenai wila...	https://bnpb.go.id/berita/arus-deras-banjir-la...
9	15 Jul 2023 09:49 ...	Banjir Melanda Kabupaten Sula Sebanyak 50 Ruma...	40 hingga 50 cm	Foto : kondisi Banjir di Kabupaten Sula, Prov...	https://bnpb.go.id/berita/banjir-melanda-kabup...
10	13 Jul 2023 14:16 ...	Infrastruktur Rusak Akibat Banjir Kabupaten Ha...	rusak fasilitas rusak infrastruktur rusak sep...	Foto : Hujan intensitas tinggi dan berangsur ...	https://bnpb.go.id/berita/infrastruktur-rusak-...
11	11 Jul 2023 11:26 ...	Bolaang Mongondow Timur Dilanda Banjir dan Lon...		Foto : Tanah longsor yang terjadi di Kabupate...	https://bnpb.go.id/berita/bolaang-mongondow-ti...
12	11 Jul 2023 11:23 ...	Jembatan Penghubung di Maluku Tengah Putus Aki...	sepanjang 480 meter ini mengalami atau 150 met...	Foto : Kondisi Jembatan Kawanua yang putus ak...	https://bnpb.go.id/berita/jembatan-penghubung-...
13	10 Jul 2023 13:00 ...	Sebanyak 611 Warga Kapuas Terdampak Banjir	Ketinggian air 300 cm 100 cm 100 cm 200 cm Ket...	Foto : Permukiman warga terdampak banjir di w...	https://bnpb.go.id/berita/sebanyak-611-warga-k...
14	06 Jul 2023 10:08 ...	Tiga Hanyut dan Satu Meninggal Dunia Dalam Ban...	ditemukan dengan kondisi meninggal satu orang ...	Foto : Tim BPBD Kabupaten OKU Selatan melakuk...	https://bnpb.go.id/berita/tiga-hanyut-dan-satu...

Figure 3. Crawling Data

2. Text Filtering

In the text filtering stage, there are 2 stages, namely data preprocessing and categorising data. Data preprocessing itself is a stage in analysing data such as cleaning data, transforming, and preparing raw data so that it can be used effectively in data analysis or machine learning [11]. The following stages are carried out:

a. Data Preprocessing

The data preprocessing stage can be seen in Figure 4. Remove Missing Value is done to remove data that does not have the context of word floods obtained at the data crawling stage. Removing missing value is done in order to eliminate bias in the context of data that does not match the data.

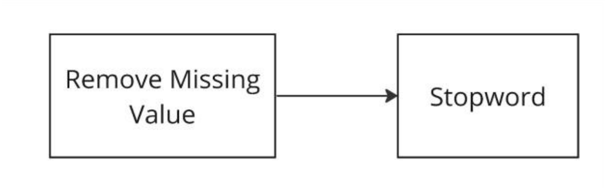

Figure 4. Stage Data Preprocessing

Stopwords are common words that often occur in text and are considered to have no significant information value in text analysis. Stopwords are done to parse noise in the text, noise here means words that have no meaning or do not contribute to the meaning of a document or sentence. An example of stopword results can be seen in Figure 5. For stopwords, we collect them manually from the news context of the crawling results. Here are the results of the stopword dictionary that we have collected ('ada', 'adalah', 'agar', 'akan', 'atau', 'dalam', 'dan', 'dari', 'bahwa', 'untuk', 'pada', 'ini', 'itu', 'dengan', 'tapi', 'jika', 'untuk', 'oleh', 'karena', 'seperti', 'sudah', 'masih', 'ketika', 'hanya', 'yang', 'di', 'oleh', 'ya', 'semua', 'kita', 'dan', 'bisa', 'menjadi', 'jika', 'sekali', 'akan', 'seperti', 'saat', 'sampai', 'terhadap', etc).

Figure 5. Remove Stopwords

b. Data Categorising

In the data categorising stage, we labelled the level of the disaster. Whether the disaster is categorised as low, medium or high level. Data labelling is done qualitatively with the guidelines of PERKA Number 02 of 2012, the following table of regulations as guidelines for labelling disaster levels can be seen in Table 1, the incident is written in Indonesian.

Table 1. Parameter Incident Level for Disaster

No	Incident	In Indonesian	*Low*	*Medium*	*High*
1	Victims Died	Korban Meninggal	< 1 orang	1 orang	> 1 orang
2	Flood height	Ketinggian Banjir	< 0.75 meter	0.75 – 1.5 meter	> 1.5 meter
3	Severe injuries	Luka Berat	< 5 orang	5 – 10 orang	> 10 orang
4	Affected house	Rumah terdampak	< 400 juta	400 – 800 juta	> 800 juta

After defining these parameters, then we perform data labeling. The following is an example of data that has been labelled manually which can be seen in Table 2. We normalized the disaster level (low, medium, high) to (0.3, 0.6, 1) in accordance with PERKA no 02 of 2012, which can be seen in Table 3.

Table 3. Normalize the Level Disaster

No	Level	*Value*
1	Low	0.3
2	Medium	0.6
3	Low	1

Table 2. Analysis Result Level Disaster

News Article in Indonesian
Pusdatinkom BNPB Abdul Muhari, Tenaga Ahli BNPB serta jajaran lainnya dijadwalkan akan melaksanakan rapat koordinasi bersama Pemerintah Provinsi Sulawesi Utara beserta Forkopimda, guna membahas percepatan penanganan darurat bencana dan antisipasi menghadapi cuaca ekstrem serta potensi bencana yang lainnya.Usai melangsungkan rapat koordinasi, Kepala BNPB beserta rombongan dan segenap Pemerintah Daerah setempat dijadwalkan meninjau lokasi terdampak dan

memberikan bantuan berupa Dana Siap Pakai (DSP) beserta bantuan lain berupa logistik dan peralatan, sesuai kebutuhan dasar warga terdampak.Adapun bantuan DSP untuk Kota Manado adalah sebesar 500 juta dan logistik senilai 250 juta. Sedangkan peralatan meliputi 2.000 selimut, 1.000 matras, 1.000 terpal, tenda ukuran 3x4 sebanyak 25 buah dan tenda ukuran 4x4 sebanyak 25 buah. Lima Korban Meninggal Dunia Bencana banjir di Kota Manado yang terjadi pada Jumat (27/1) itu dilaporkan telah merendam kurang lebih 400 rumah di 34 desa/kelurahan dan 9 kecamatan. Banjir dengan tinggi muka air yang berkisar antara 80-300 sentimeter itu telah berdampak pada 3.013 KK atau 9.382 jiwa dan merenggut satu korban jiwa.Sementara itu peristiwa tanah longsor telah berdampak pada 63 KK dan terbagi di beberapa titik di 22 desa/kelurahan dan 7 kecamatan. Petaka tersebut juga menelan empat korban jiwa, satu luka berat dan dua lainnya luka ringan. Rumah rusak ada sebanyak 53 unit termasuk 1 tempat ibadah.Di samping itu, banjir dan longsor juga memaksa 1.021 jiwa mengungsi di beberapa titik.
Analysis Result
Level High, with calculations in Indonesian: • Ketinggian Banjir: “Tinggi (kisaran antara 80-300 sentimeter)”. **High** • Meninggal: “Lima korban meninggal dunia”. **High** • Kerugian rumah terdampak: “53 unit rumah rusak”. **Medium** • Luka berat: “satu luka berat dan dua lainnya luka ringan”. **Low** Mean Score = (1+1+0.6+0.3) / 4 = 0.725, because the mean score is more than 0.6, the result is **high**

We use the average calculation method, as we think it is quite simple and can give a general idea, but it may not always accurately reflect the actual level of disaster, especially if the parameters taken into account have different weights or influences. The following equation can be seen in equation (1).

$$TB = \frac{\sum_{i=1}^{n} SK_i}{Number\ or\ parameter\ Considered} \tag{1}$$

In this formula:

- TB is Level Disaster
- Σ (sigma) is used to denote the summation of SK1 through SKn.
- "i" is the index used for summation, starting from i = 1 (beginning) and ending at i = n (end).
- SKi represents the score for the i-th parameter.
- Number_of_Parameters_Considered is the count of parameters with available information (parameters without information are excluded).

For data samples that have been labeled can be seen in table 3.

Table 3. Data Samples have been Labeled.

No	Teks Data After Preprocessing	*Value*
1	wali dewanti rumpoko mengumumkan semua korban hilang terjadinya wilayah tersebut ditemukan adapun total korban hilang mencapai tujuh orang bersyukur semua korban hilang ditemukan sebanyak tujuh orang	Low

No	Teks Data After Preprocessing	*Value*
	kata dewanti youtube bnpb sabtu dewanti menyampaikan sejumlah tempat fasilitas umum jalanan terdampak bandang sempat tetutup aksesnya adapun tinggi luapan air bervariasi	
2	ketinggian air banjir kawasan mencapai enam puluh sentimeter hingga satu meter berangsur surut malam hari seorang warga terdampak bernama cindy kawasan muara angke sering kali terjadi membuat warga terbiasa	Medium
3	hujan deras mengguyur wilayah dki jakarta mengakibatkan rt ibu kota terendam banjir senin pagi ketinggian air mulai puluh sentimeter satu meter kepala pusat data informasi kebencanaan badan penanggulangan bencana daerah bpbd dki jakarta moh insaf mengungkapkan rt terendam banjir tersebar wilayah jakarta barat jakarta timur jakarta selatan bpbd mencatat genangan sebelumnya terjadi rt tercatat ketinggian air senin pagi mencapai satu meter hingga pukul wib tinggi muka air mencapai seratus sentimeter sebut insaf berikut daftar rt terendam banjir senin pagi kelurahan cilandak timur rt terendam banjir ketinggian lima puluh sentimeter kelurahan kampung melayu rt terendam banjir ketinggian mencapai satu meter kelurahan bidara cina rt terendam banjir ketinggian mencapai sembilan puluh sentimeter kelurahan cawang rt terendam banjir ketinggian mencapai sembilan puluh sentimeter kelurahan cililitan rt terendam banjir ketinggian mencapai enam puluh sentimeter kelurahan rawa buaya rt terendam banjir ketinggian enam puluh sentimeter kelurahan tegal alur rt terendam banjir ketinggian tiga puluh lima sentimeter kelurahan kembangan utara rt terendam banjir ketinggian empat puluh sentimeter bpbd dki mengimbau masyarakat tetap berhati hati waspada potensi genangan pungkas	High
4	adzra nabila mahasiswi institut pertanian bogor ipb hilang terseret arus terperosok gorong gorong jalan dadali kota bogor selasa lalu ditemukan usai hari hilang korban ditemukan kondisi meninggal dunia minggu pagi kanal banjir barat kelurahan jembatan besi tambora jakarta barat kepala pelaksana badan penanggulangan bencana bpbd dki jakarta isnawa adji mengatakan korban terbawa jauh lokasi awal diduga lantaran debit air sungai ciliwung besar banyaknya material bambu batang pohon sampah berasal hulu hingga tersangkut pintu air manggarai menandakan volume debit air cukup tinggi kata isnawa dikonfirmasi minggu mengatakan tingginya debit air sungai diperparah intensitas curah hujan turun akhir akhir	Medium

3. Text Classification

In this text classification stage, there are 2 stages carried out, namely, Feature Vector Generation and Classification using Machine Learning. Feature vectors are used to describe and quantify the characteristics or features of data, making it easier for algorithms to analyse and make predictions.

a. Feature Vector Generation

At this stage, we transform the text data into vectors. The purpose of converting to vectors is because computers work with numbers, not text and will also facilitate classification using machine learning. To convert text data into vectors, we use TfidfVectorizer, the vector example results can be seen in Figure 6.

```
ix for the whole corpus:
```

akan	alam	banjir	bencana	besar	di	hujan	ini	kami	kerugian	kota	lalu	melanda	membahas	mencegah	mengadakan
0.00	0.33	0.17	0.33	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.20	0.00	0.30	0.00	0.00	0.38	0.00	0.00	0.38	0.38	0.38	0.00	0.00	0.00
0.00	0.00	0.19	0.00	0.29	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.36	0.00
0.36	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.36	0.00	0.36

```
or for a particular sentence:
```

akan	alam	banjir	bencana	besar	di	hujan	ini	kami	kerugian	kota	lalu	melanda	membahas	mencegah	mengadakan
0.0	0.33	0.17	0.33	0.0	0.33	0.33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 6. Convert Text to Vector

In the example above, we have a corpus consisting of sentences related to "banjir". We calculate the TF-IDF matrix for the whole corpus, which will show the TF-IDF weights for the words in each sentence in the context of "banjir". Then, we compute the TF-IDF vector for a particular sentence, i.e. "Banjir adalah bencana alam yang sering terjadi di musim hujan." This vector will show the TF-IDF weights for the words in that sentence in the context of the whole corpus. The result is two tables: one for the TF-IDF matrix of the corpus and another for the TF-IDF vector of the particular sentence. In the TF-IDF vector, the values measure how important the words in the sentence are in the context of "banjir". The higher the TF-IDF value, the more important the word is in that sentence and the context of "banjir" as a whole. Before doing TF-IDF, we also convert numbers to letters as shown in Figure 7.

Malang - Curah hujan tinggi dan tak kunjung berhenti membuat air Sungai Panguluran yang bermuara di wilayah Sitiarjo, Kecamatan Sumbermanjing Wetan, Kabupaten Malang, meluap. Air yang meluap mulai mengenangi ruas jalan dan sejumlah rumah warga setinggi 40 cm. Berdasarkan data BPBD Kabupaten Malang pukul 12.33 WIB, Jumat (7/7/2022) mencatat sejumlah rumah warga di RT 56/RW 15, Dusun Krajan Tengah atau kampung palung, Desa Sitiarjo, Kabupaten Malang, mulai tergenang air setinggi kurang lebih **40 cm**. "Ada tiga rumah di RT 56/RW 15 Kampung Palung yang tergenang air akibat luapan Sungai Panguluran. Sementara enam rumah lain juga terendam di wilayah RT18/RW15," beber Kabid Kedaruratan dan Logistik BPBD Kabupaten Malang Sadono Irawan kepada detikJatim, Jumat (7/7/2023). Sadono menambahkan ada 13 rumah warga di RT 26/RW 14 Dusun Krajan Kulon, Desa Sitiarjo, yang juga tergenang luapan air Sungai Panguluran. "Untuk akses Jalan Malang-Sendangbiru tergenang air setinggi kurang lebih 30 cm. Tepatnya di selatan jembatan Sitiarjo," kata Sadono. Selain itu, banjir juga mengenangi akses Jalan Sitiarjo-Sidodadi, Kecamatan Gedangan. Tepatnya di Dusun Krajan Kulon, Desa Sitiarjo, setinggi **1,5 meter**. "Ada beberapa titik ruas jalan yang ikut tergenang luapan air sungai. Seperti jalan kampung Dusun Krajan Tengah setinggi 70 cm," urainya. BPBD dan tim gabungan juga tengah memantau kondisi wilayah Dusun Rowoterate, Desa Sitiarjo, yang turut terdampak luapan air sungai. "Untuk Rowoterate juga ikut terdampak beberapa rumah. Tapi belum dapat data, karena akses jalan masih tergenang air," kata Sadono. Sadono mengaku, untuk saat ini masih belum disediakan pos pengungsian kepada masyarakat yang terdampak, karena perlahan banjir sudah mulai surut. "Sementara belum, karena saat ini sudah surut kembali," pungkasnya.	berdasarkan data bpbd kabupaten malang pukul wib jumat mencatat sejumlah rumah warga rt dusun krajan tengah kampung palung desa sitiarjo kabupaten malang mulai tergenang air setinggi kurang lebih **empat puluh cm** tiga rumah rt kampung palung tergenang air akibat luapan sungai panguluran sadono menambahkan tiga belas rumah warga rt dusun krajan kulon desa sitiarjo tergenang luapan air sungai panguluran akses jalan malang sendangbiru tergenang air setinggi kurang lebih tiga puluh cm tepatnya dusun krajan kulon desa sitiarjo setinggi **satu koma lima meter** beberapa titik ruas jalan ikut tergenang luapan air sungai bpbd tim gabungan tengah memantau kondisi wilayah dusun rowoterate desa sitiarjo turut terdampak luapan air sungai data akses jalan tergenang air kata sadono sadono mengaku disediakan pos pengungsian masyarakat terdampak perlahan banjir mulai surut surut pungkasnya

Figure 7. Convert Numbers to Letter

b. Classification using Machine Learning

In the classification stage, we use SVM for data classification in machine learning. Not only SVM, but we also tried to use Decision Tree, Naive Bayes, AdaBoost, and XgBoost as a comparison in a special case for the flood disaster level classification problem. The classification results can be seen in the result chapter.

C. Result and Discussion

At this stage, classification has been carried out by comparing the accuracy results of the SVM, Decision Tree, Naive Bayes, XgBoost, and AdaBoost methods. Where the training data uses 99 data, and 19 data as test data. The results show that Random Forest is superior with a score of 78% test accuracy results which can be seen in Table 4. So, for the distribution of training data can be seen in Figure 8. where for low levels there are 33 data, medium 34 data, and high 32 data.

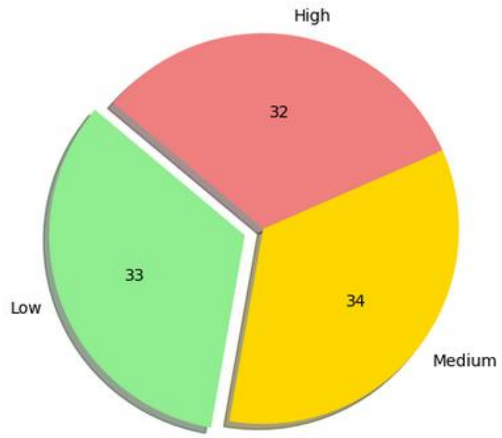

Figure 8. Data training

The results that have been displayed include the accuracy of the test data validation score, precision calculation, recall calculation, and F1-Score. Precision itself is how accurate the model is in identifying positive cases from all predicted positive cases. In this case, precision is used to identify whether the flood level is low, medium, or high. Precision is calculated by the formula (2):

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

Then recall itself measures how good the model is at identifying all true positive cases. In this case recall helps to measure how many cases the true rate can count. Recall is calculated by the formula (3):

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

While F1-Score is the harmonic mean of precision and recall, with the formula can be calculated:

$$F1 - Score = 2x(\frac{Precision\ x\ Recall}{Precision + Recall}) \qquad (3)$$

Looking at the results of accuracy, Random Forest gets a better score than other methods in the case of flood classification, with an accuracy of 78%. The following are result predict of disaster category level data that have been preprocessing data in Table 5.

Table 4. Result Data Test

No	Method	*Accuracy*	*Precision*	*Recall*	*F1-Score*
1	SVM	52	53	54	52
2	Naïve Bayes	47	50	50	46
3	XgBoost	68	64	63	60
4	Decision Tree	52	49	48	48
5	AdaBoost	57	61	60	57
6	Random Forest	78	88	76	78

Table 5. Result Data Predict

No	Teks	*Label*	*Random Forest*	*SVM*
1	banjir merendam rumah warga hingga satu koma lima meter keempat dusun diperkirakan jumlah warga terdampak sebanyak ratus enam puluh lima kepala keluarga jumlah warga terdampak diperkirakan sebanyak seribu empat ratus delapan puluh tiga jiwa pungkasnya ...	High	High	High
2	akibat kejadian lima puluh orang terdampak data awal diperoleh dinas kesehatan setempat berkoordinasi beberapa dinas terkait jumlah korban diinformasikan sebanyak nol orang terdiri nol orang meninggal nol orang hilang nol luka beratrawat inap nol luka ringanrawat jalan nol orang pengungsi	Low	Low	High
3	sejumlah kecamatan terdampak banjir terutama kecamatan sutojayan data badan penanggulangan bencana daerah bpbd ribuan warga terdampak ivong banjir menggenangi hampir seluruh kelurahan sutojayan kemudian sebagian desa sumberjo sebagian kelurahan kalipang ivong memastikan ribuan warga tiga desa tersebut dievakuasi hujan terus turun ketinggian air terus meningkat	Medium	Medium	Medium
4	akibat kejadian empat puluh tujuh kk seratus delapan puluh delapan orang terdampak data awal diperoleh dinas kesehatan setempat berkoordinasi beberapa dinas terkait jumlah korban diinformasikan sebanyak nol orang terdiri nol orang meninggal nol orang hilang nol luka beratrawat inap nol luka ringanrawat jalan nol orang pengungsi	Low	High	Low
5	wali dewanti rumpoko mengumumkan semua korban hilang terjadinya wilayah tersebut ditemukan adapun total korban hilang mencapai tujuh orang bersyukur semua korban hilang ditemukan sebanyak tujuh orang kata dewanti youtube bnpb sabtu dewanti menyampaikan sejumlah tempat fasilitas umum jalanan terdampak bandang sempat tetutup aksesnya adapun tinggi luapan air bervariasi	Low	Low	Low

It can be seen from the results of testing predictions that SVM gives some wrong answers for flood classification rates than random forest. Random forest performance and testing results are better than SVM.

D. Conclusion

In the case of flood disaster text classification SVM still does not get good accuracy compared to Random Forest. Even so there needs to be further research on this matter to improve SVM performance such as using stemming, changing water level phrases to be like floods (high, medium, low). Then use Indonesian word embedding instead of tf-idf. And we think there is a need for a disaster dictionary to make it easier to filter sentences containing damage, identification of victims, and how much loss is received. Without realizing it, we have increased the level of awareness of hydrometeorological disasters.

E. Acknowledgment

For the smooth running of this research, I would like to thank Politeknik Elektronika Negeri Surabaya for being one of the places where I studied and BNPB Indonesia for providing data on its website. We hope that this article can be a reference for users and the public to care about hydrometeorological disasters in Indonesia. Even so, this research still has many shortcomings, so that further research can be developed. And finally, hopefully the results of this research can provide insight to the readers.

F. References

[1] Ainun Rosyida, Ratih Nurmasari, and Suprapto, "Analisis perbandingan dampak kejadian bencana Hidrometeorologi dan Geologi di Indonesia dilihat dari jumlah korban dan kerusakan," *Jurnal Dialog Penanggulangan Bencana*, vol. 10, hal. 12–21, 2009.

[2] Shalunts, Gayane & Backfriend, Gerhard & Prinz, and Katja, "Sentiment Analysis of German Social Media Data for Natural Disasters," *In Information Systems for Crisis Response and Management*, 2014.

[3] Endah Nofiyanti, Erry Maricha Oki, and Nur Haryanto, "Analisis Sentimen terhadap penanggulangan bencana di Indonesia,"*Jurnal Ilmiah SINUS.* vol 19, No. 2, 2021.

[4] A. Alfarrarjeh, S. Agrawal, S. H. Kim, and C. Shahabi, "Geo-Spatial Multimedia Sentiment Analysis in Disasters," *In IEEE International Conference on Data Science and Advanced Analytics*, Tokyo, Japan, pp. 193-202, doi: 10.1109/DSAA.2017.77., 2017.

[5] Basarslan, Muhammet & Kayaalp, Fatih, "Sentiment Analysis with Machine Learning Methods on Social Media," *Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 9, pp 5-15, 2021.

[6] Badan Nasional Penanggulangan Bencana, Pedoman Umum Pengkajian Risiko Bencana, No. 02, 2012.

[7] Anuradha Khattar, and S. M. K. Quadri, "CAMM: Cross-Attention Multimodal Classification of Disaster-Related Tweets," *IEEE Access*, vol. 10, pp. 92889-92902, 2022.

[8] M. K. Delimayanti, Risna Sari, Mauldy Laya, M. R. Faisal, and Pahrul, "Pemanfaatan Metode Multiclass-SVM pada Model Klasifikasi Pesan Bencana Banjir di Twitter," *Edu Komputika Journal*, 2021.
[9] A. D. Kurniawati, P. Kristalina, Moch. Zen Samsono Hadi, and D. Widodo, "GIS-Based Disaster Management using Support Machine Mode for Hazard Level Classification in Disaster Areas," *in International Electronics Symposium (IES)*, 2023.
[10] Ryan Mitchell, *Web Scraping with Python*, USA: O'Reilly, 2018.
[11] Roy Jafari, *Hands-On Data Preprocessing in Python: Learn how to effectively prepare data for successful data analytics*, UK: Packt, 2022.
[12] A. T. Kabakus, R. Kara, "TwitterSentiDetector: a domainindependent Twitter Sentiment Analyser," *INFOR: Information Systems and Operational Research*, 2017.
[13] A. A. Anggarda, H. Purnaweni, S. Suwitri, and T. Afrizal, "Analysis of Flood Disaster Mitigation Policy in Lamongan District", *IJSSB*, vol. 5, no. 4, pp. 537–542, Dec. 2021.
[14] J. Xiong, J. Li, W. Cheng, N. Wang, and L. Guo, "A GIS-Based Support Vector Machine Model for Flash Flood Vulnerability Assessment and Mapping in China," *IJGI*, vol. 8, no. 7, p. 297, Jul. 2019.
[15] K. Saputra, M. U. H. A. Rasyid, and Muh. Z. S. Hadi, "Employing Machine Learning with Optimized SVM Parameter to Detect Water Pipe Leakage," *in 2021 International Electronics Symposium (IES)*, Surabaya, Indonesia: IEEE, Sep. 2021, pp. 298–303.