



Pemanfaatan Metode Multiclass-SVM pada Model Klasifikasi Pesan Bencana Banjir di Twitter

Mera Kartika Delimayanti^{1)✉}, Risna Sari¹⁾, Mauldy Laya¹⁾, Muhammad Reza Faisal²⁾, dan Pahru²⁾

¹⁾Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta, Depok, Indonesia

²⁾Program Studi Ilmu Komputer, Universitas Lambung Mangkurat, Banjarmasin, Indonesia

Info Artikel

Sejarah Artikel:

Diterima: Juni 2021

Direvisi: Juni 2021

Disetujui: Juni 2021

Keywords:

Klasifikasi, Twitter, Bencana Banjir, Multiclass, Support Vector Machine

Abstrak

Musibah bencana alam banjir merupakan salah satu musibah alam yang sangat umum kejadian di Indonesia baik di pulau Jawa ataupun di luar Pulau Jawa. Sumber informasi dari media social Twitter dapat digunakan sebagai *Social Network Sensor* (SNS) untuk bencana banjir. Penelitian ini bertujuan untuk melakukan klasifikasi pesan twitter khususnya tentang bencana banjir sebagai hasil dari algoritma *machine learning*. Pada klasifikasi pesan bencana banjir dilakukan serangkaian proses *data preprocessing* dan dilanjutkan dengan ekstraksi fitur dan pembobotan kata dari data twitter. Berbagai teknik yang dilakukan pada proses *data preprocessing* berpengaruh terhadap tingkat akurasi hasil klasifikasi yang dihasilkan. Selanjutnya, penelitian ini menyampaikan bahwa algoritma *Support Vector Machine* (SVM) adalah sebuah algoritma dapat dimanfaatkan untuk klasifikasi data ke dalam tiga (3) kelas saksi mata, kelas non saksi mata dan kelas tidak diketahui. Total jumlah data yang telah diolah sebanyak 3000 data untuk 3 kelas label data. Pendekatan metode *One Versus One* (OVO) untuk dua kelas label data dan *One Versus All* (OVA) untuk lebih dari dua kelas label pada algoritma SVM. Pada penelitian ini telah diterapkan algoritma *Support Vector Machine* untuk *multiclass* dengan metode OVA. Hasil eksperimen telah menunjukkan pendekatan OVA pada algoritma SVM dengan kernel RBF menghasilkan evaluasi performansi yang paling tinggi. Hasil evaluasi tersebut adalah nilai akurasi hingga 87.03%. Sedangkan hasil riset sebelumnya dengan algoritma SVM tetapi pendekatan yang berbeda hanya mencapai nilai akurasi hingga 77.87%.

Abstract

Among Indonesia's most frequent natural disasters is flooding, both on and off the island of Java. Twitter information sources can be used as a Social Network Sensor (SNS) for flood disasters. In attempt to decide the evaluation performance of the classification model, flood-related tweets were classified. A series of data preprocessing steps are followed by feature extraction and word weighting from Twitter data in the classification of twitter contained flood disaster messages. The technique used to preprocess data affects the level of accuracy performance. Furthermore, the Support Vector Machine (SVM) algorithm was executed to classify data into three (3) classes: eyewitness, non-eyewitness, and unknown. The total amount of data processed is 3000 for three class labels. The SVM algorithm has two methods: One Versus One (OVO) for two-class labels and One Versus All (OVA) for more than two class labels. The multiclass SVM with the OVA method was used in this study. When compared to previous research and methods that achieved 77.87 percent accuracy, the OVA method with the RBF kernel had a classification accuracy value of 87.03 percent.

PENDAHULUAN

Bencana alam banjir merupakan bencana alam yang sering terjadi di kepulauan negara Indonesia. Berdasarkan data Badan Nasional Penanggulangan Bencana bahwa sejak awal tahun hingga bulan Mei tahun 2020, telah terjadi 1296 bencana alam dengan kejadian bencana alam banjir mencapai 495 kasus (BNPB, 2020). Bencana alam banjir dapat mengakibatkan kehidupan efek negatif di masyarakat sebagai contoh terjadinya kehancuran pada fasilitas umum jalan dan jembatan. Selain itu, menyebabkan banyaknya kerusakan pada rumah dan alat transportasi seperti mobil dan sepeda motor. Kerusakan akibat bencana banjir adalah bersifat masif sehingga masyarakat cenderung untuk mencari atau memberi informasi dengan kaitan bencana banjir. Pada beberapa kejadian bencana alam, media sosial telah dimanfaatkan sebagai media penyampaian informasi yang lebih cepat daripada penyampaian informasi melalui televisi dan radio. Informasi adalah berupa data yang diambil dari tweet atau media sosial dalam bentuk twitter.

Media sosial memberikan data yang dimanfaatkan sebagai sumber informasi untuk menanggapi terhadap kejadian bencana alam di suatu wilayah. Penelitian yang telah dilakukan oleh Wu, telah memberikan konsep baru dengan menggunakan media sosial sebagai *social network* (Wu, 2016). Pemanfaatan media sosial dapat sebagai sumber informasi dari warga atau masyarakat yang dalam hal ini sebagai saksi mata atau pengamat terhadap kejadian bencana alam. Informasi yang berupa laporan saksi mata lebih disukai dibandingkan informasi yang berasal dari non saksi mata / informasi berantai. (misalnya berasal dari pesan di luar area bencana). Sebagaimana kita ketahui, informasi dari saksi mata atau sumber berita merupakan informasi yang dicari oleh pewawancara atau lembaga penegak hukum karena lebih dapat dipercaya. Penelitian telah dilakukan dengan melakukan klasifikasi pada data media sosial menjadi tiga kelas yakni kelas (i) Saksi mata, kelas (ii) non-saksi mata, dan kelas (iii) tidak diketahui (Zahra et al., 2020). Penelitian tentang teknik klasifikasi pesan di twitter telah dilakukan oleh Zahra dkk. Penelitian tersebut adalah melakukan klasifikasi pesan twitter untuk bencana alam ke dalam tiga (3) kelas. (Zahra et al., 2020).

Algoritma *Support Vector Machine* (SVM) digunakan untuk mengolah informasi dari Twitter seperti yang telah dilakukan oleh penelitian sebelumnya (Annisa Rahmaniari Dwi Pratiwi & Erwin Budi Setiawan, 2020; Dhina Nur Fitriana & Yuliant Sibaroni, 2020; Fitriyyah et al., 2019). Data twitter telah dimanfaatkan

untuk penelitian sentiment analisis pada sebuah topik tertentu (Dhina Nur Fitriana & Yuliant Sibaroni, 2020), deteksi rumor (Annisa Rahmaniari Dwi Pratiwi & Erwin Budi Setiawan, 2020) dan ujaran kebencian (Oryza Habibie Rahman et al., 2021). Algoritma SVM sangat umum digunakan pada studi kasus klasifikasi dan regresi karena algoritma SVM menghasilkan klasifikasi dengan nilai akurasi lebih tinggi untuk data twitter dibandingkan yang metode lainnya (Pratama & Murfi, 2014; Rachmatika & Bisri, 2020). Selain itu, SVM digunakan untuk klasifikasi analisis sentimen komentar pada media sosial Instagram (Luqyana et al., 2018). Penelitian tersebut telah dilaksanakan dengan teknik ekstraksi fitur pada Instagram dengan teknik TF-IDF dan menggunakan model klasifikasi algoritma SVM. Hasil pengujian pada penelitian tersebut didapatkan hasil dengan akurasi tertinggi adalah 90%. Oleh sebab itu, dari hasil pada berbagai penelitian, algoritma SVM dapat dijadikan alternatif dalam melakukan teknik klasifikasi pesan saksi mata pada saat terjadi bencana banjir. Algoritma SVM yang dimanfaatkan pada penelitian ini khususnya dengan multiclass SVM karena untuk melakukan klasifikasi lebih dari 2 (dua) kelas. Data sentiment analisis dengan klasifikasi lebih dari 2 kelas menunjukkan algoritma multiclass SVM memberikan hasil klasifikasi lebih tinggi dengan Teknik unigram (Dhina Nur Fitriana & Yuliant Sibaroni, 2020). Hasil perbaikan metode pada algoritma SVM dilakukan pada penelitian ini. Sebelumnya telah dilakukan penelitian dengan algoritma SVM secara umum (Delimayanti et al., 2020). Berbagai penyesuaian pada metode yang telah digunakan sebelumnya telah menunjukkan hasil akurasi yang lebih baik dengan algoritma SVM *multiclass*. Harapannya adalah hasil dari penelitian ini dapat digunakan untuk prediksi kejadian bencana alam dengan melihat asal informasi dari ketiga kelas tersebut. Berikutnya penelitian dapat memberikan *early warning* pada kejadian bencana alam dari klasifikasi informasi pesan twitter untuk bencana alam banjir kejadiannya sangat sering di Indonesia. Hasil pengolahan dari data SNS khususnya data Twitter dapat memprediksi kejadian bencana sesuai dengan *class* yang telah ditentukan. Oleh sebab itu, penelitian ini dapat memberikan kontribusi positif bagi pencegahan terjadinya dampak terjadinya bencana alam yang lebih buruk bagi masyarakat.

METODE PENELITIAN

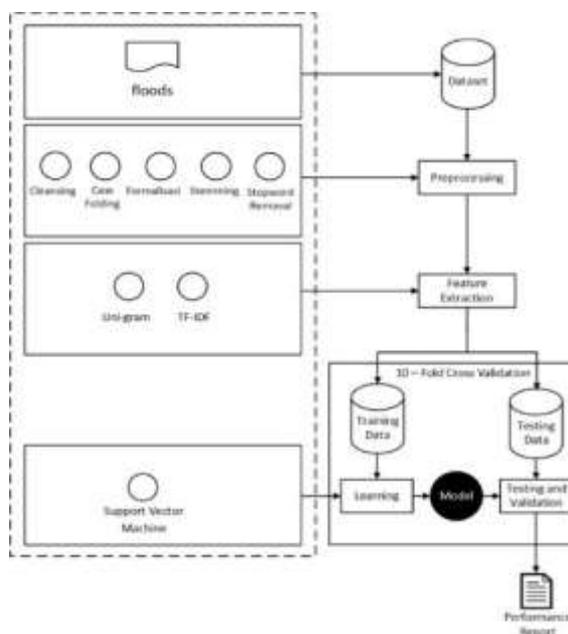
Salah satu definisi dari *Social Network Sensor* (SNS) ialah sebuah teknik baru dari sensor

fisik. Pada umumnya, sensor fisik dapat mengukur dan menghasilkan data fisis maupun data kimia seperti temperatur, kelembaban, dan cahaya. Selain itu, sensor sosial adalah sebuah sensor yang mendapatkan datanya dari media sosial salah satunya berupa pesan tweeter, data pengguna, lokasi pengguna dan lainnya (Wu, 2016). Media sosial telah memberikan dampak pada percepatan penyebaran berita atau kabar. Bahwasannya, pengguna media sosial mengabarkan apa yang dilihat oleh pengguna sesaat setelah kejadian berlangsung. Oleh sebab itu, beberapa peneliti telah banyak melakukan penelitian di bidang pemanfaatan media sosial sebagai sensor atau *Social Network Sensor* (Christakis & Fowler, 2010)(Kryvasheyev et al., 2015) dan penelitian pada penggunaan media sosial untuk peringatan dini dan pemantauan pasca kejadian bencana alam. Dengan kata lain adalah penggunaan media sosial sebagai sensor (SNS) (Hernandez-Suarez et al., 2019). Beberapa jenis bencana alam yang menggunakan media sosial sebagai sensor adalah kebakaran hutan, gempa bumi, angin puting beliung dan banjir (Zahra et al., 2020). Beberapa penelitian dengan menggunakan algoritma dan teknik pengolahan data soial media sebagai sensor adalah penelitian di bidang *text mining*.

Teknik penambangan data berupa teks oleh komputer untuk mendapatkan informasi baru yang tersirat atau implisit secara otomatis dari berbagai sumber teks yang berbeda adalah merupakan definisi dari *Text mining* (Feldman & Sanger, 2007). Text Mining merupakan salah satu cabang penelitian dari data mining. Oleh sebab itu, tingkat aritektur yang digunakan pada text mining adalah sama dengan data mining. (Falahah & Dwiki Adriadi Nur, 2015). Pada Gambar 1 menunjukkan langkah-langkah penelitian yang berupa data *preprocessing*. Pada tahapan data *preprocessing* meliputi pembersihan data /*data cleansing*, *case folding* atau perubahan bentuk teks, formalisasi, *stemming* atau perubahan ke dalam kata dasar, dan *stopword removal*. Langkah selanjutnya adalah dengan melakukan pemilihan fitur atau *feature selection*. Langkah *preprocessing* merupakan langkah penting karena dapat mempengaruhi hasil evaluasi algoritma data mining yang dipilih(Pradana & Hayaty, 2019). Metode

algoritma yang dipilih pada penelitian ini tidak sama dengan penelitian sebelumnya yakni penggunaan algoritma SVM dengan dengan berbagai parameter yang diujicobakan pada algoritma tersebut(Delimayanti et al., 2020).

Penelitian yang dilakukan menggunakan data pesan twitter khususnya untuk pesan bencana alam banjir memanfaatkan *library* dari Rtweet dengan akun pribadi. Tujuannya agar mendapatkan data dengan class sebagai saksi mata, non saksi mata atau tidak tahu. Data yang masuk selanjutnya dilakukan pengolahan data / *pre-processing* yang diperlukan. Total data yang digunakan adalah 3000 data berasal dari data twitter tentang bencana banjir di Indonesia pada bulan November 2019 – April 2020 dalam Bahasa Indonesia. Jumlah data awal adalah 3300 data dan setelah melewati *pre-processing* menjadi 3000 data seperti tercantum pada Tabel 1. pengolahan dari data tweet didapatkan dengan memanfaatkan *Twitter Scraper package*. Package ini termasuk salah satu *package* pada bahasa Python. Tabel 2 memaparkan hasil pengolahan data berupa informasi dari dataset yang digunakan.



Gambar 1. Tahapan proses penelitian

Tabel 1. Total Data dalam Dataset

No	Data	saksi mata	non-saksi mata	tidak tahu	Jumlah
1	Data-1	4	107	43	195
2	Data-2	269	306	180	755
3	Data-3	60	102	61	223
4	Data-4	11	14	12	37
5	Data-5	9	52	0	61
6	Data-6	18	33	7	58
7	Data-7	245	127	161	533
8	Data-8	113	102	45	260
9	Data-9	101	67	95	263
10	Data-10	76	49	335	460
11	Data-11	88	82	84	254
12	Data-12	68	67	66	201
	Total	1103	1108	1089	3300

Tim peneliti melakukan pembersihan data atau data *pre-processing* dengan maksud untuk menyesuaikan data yang ada sesuai format yang diperlukan pada penelitian. Berikut ini adalah tahapan dalam *pre-processing* yakni:

A. Pelabelan

Data hasil dari proses scrapping twitter diberikan label sesuai isinya. Data dibagi menjadi tiga (3) kelas yakni (i) kelas Saksi mata, (ii) kelas non-saksi mata, dan (iii) kelas tidak diketahui. Teknik pelabelan data dilakukan secara manual dengan melihat karakteristik isi pesan twitter. Tabel 2 menunjukkan contoh data dan pelabelan data.

Tabel 2. Contoh Data dan Pelabelan Data

Data asal	Label
Banjir woiiii... Hujan tak berhenti dari semalam.... pic.twitter.com/kdYZ0Z3KTO	Saksi mata
Hujan lebat gais. banjir disini. huhu	Saksi mata
Rumah aku x banjir	Saksi mata
Kali Jantung Meluap, Ratusan Rumah di Depok Terendam Banjir:	Non-saksi mata
Hujan deras membuat sejumlah rumah di Depok Terendam Banjir 10/10 ep 13 14 banjir air mata, ga nyangka, baper jadi satu, alur nya ga ketebak pula	Tidak Tahu

B. *Cleansing* dan *Remove Duplicate*

Proses dilanjutkan dengan pembersihan / *Cleansing* yakni salah satu proses pembersihan dengan menghapus karakter yang tidak dibutuhkan pada waktu proses klasifikasi sehingga hanya karakter alfabet saja yang

tertinggal. Pada proses *cleansing* dilakukan penghapusan username, hashtag, URL, RT, simbol/karakter seperti ("+=!&?*^~#-_), dan angka. *Remove Duplicate* adalah penghapusan pada data yang sama akibat pada saat proses scrapping data twitter. Tabel 3 menunjukkan hasil data setelah dilakukan tahapan *cleansing* dan *remove duplicate*.

Tabel 3. Hasil Data Setelah Tahap *Cleansing* dan *Remove Duplicate*

Data setelah Proses	Label
Banjir woiiii Hujan tak berhenti dari semalam	Saksi mata
Hujan lebat gais banjir disini huhu	Saksi mata
Rumah aku x banjir	Saksi mata
Kali Jantung Meluap Ratusan Rumah di Depok Terendam Banjir	Non-saksi mata
Hujan deras membuat sejumlah rumah di Depok Terendam Banjir ep banjir air mata ga nyangka baper jadi satu alur nya ga ketebak pula	Tidak Tahu

C. *Case Folding*

Case folding adalah proses untuk mengubah isi text pada twitter menjadi huruf kecil. Selain itu, dilakukan penghapusan karakter selain huruf. Hasil proses *case folding* dapat dilihat di tabel 4.

Tabel 4. Hasil Data Setelah Tahap *Case Folding*

No	Data setelah Proses <i>Case Folding</i>
1	banjir woiiii hujan tak berhenti dari semalam
2	hujan lebat gais banjir disini huhu
3	rumah aku x banjir
4	kali jantung meluap ratusan rumah di depok terendam banjir hujan deras membuat sejumlah rumah di depok terendam banjir
5	ep banjir air mata ga nyangka baper jadi satu alur nya ga ketebak pula

D. *Formalisasi*

Fungsi utama pada tahapan ini adalah mendeteksi dan sekaligus memperbaiki data yang rusak atau yang tidak akurat. Tujuan dari proses ini untuk menambah nilai akurasi pada proses klasifikasi. Proses formalisasi termasuk didalamnya adalah menghilangkan istilah bahasa gaul atau 'alay' yang menyebabkan penggunaan Bahasa Indonesia yang tidak baku sesuai dengan KBBI. Tabel 5 menunjukkan contoh kamus slang yang dipakai pada proses formalisasi.

Tabel 5. Contoh Kamus Slang

No	Slang	Formal
1	banjirrrr	banjir
2	dpn	depan
3	gais	teman
4	smpai	sampai
5	waduh	aduh
...
887	rmh	rumah

E. *Stemming*

Proses ini untuk melakukan transformasi kata dengan menggunakan kata dasarnya (*root word*) dan menghilangkan semua imbuhan kata. Hasil data yang muncul adalah berupa kata dasar. Hasil proses *stemming* data ditunjukkan di Tabel 6.

Tabel 6. Hasil Proses *Stemming* Data

No	Data Setelah Proses <i>Stemming</i>
1	banjir woiiii hujan tak henti dari malam
2	hujan lebat teman banjir sini huhu
3	rumah saya x banjir
4	kali jantung luap ratus rumah di depok rendam banjir hujan deras buat sejumlah rumah di depok rendam banjir
5	episode banjir air mata tidak nyangka baper jadi satu alur nya tidak tebak pula

F. *Stopword Removal*

Stopword Removal atau filter kata digunakan untuk menghapus kata yang terdapat didalam tweet tetapi kata tersebut tidak memiliki arti atau tidak baku (*stopword*). Terdapat 2(dua) kamus *stopword* yang telah dicoba pada tahap penelitian ini yakni kamus id.stopwords.02.01.2016.txt (Devid Haryalesmana & Martijn Wieriks, 2016). Satu lagi adalah kamus yang dibuat berdasarkan data banjir yang digunakan. Tabel 7 menunjukkan hasil data setelah tahapan *Stopword removal*.

Tabel 7. Hasil Proses Tahap *Stopword Removal*

Data Setelah Proses <i>Stopword Removal</i>	Label
banjir hujan tak henti dari malam	Saksi mata
hujan lebat teman banjir sini	Saksi mata
rumah saya banjir	Saksi mata
jantung luap ratus rumah depok rendam banjir hujan deras buat sejumlah rumah depok rendam banjir	Non - Saksi mata
episode banjir air mata tidak nyangka baper jadi satu alur tidak tebak pula	Tidak tahu

Tahapan setelah data *pre-processing* adalah tahap ekstrasi fitur dengan pembobotan TF-IDF

dan Unigram. Proses berikutnya adalah pemisahan data menjadi data testing dan training secara bergantian dengan teknik *k-fold cross validation*. Model klasifikasi dilakukan dengan menggunakan algoritma *Support Vector Machine* (SVM) dengan pendekatan SVM *multiclass*. Terdapat dua parameter yakni parameter *One Versus One* (OVO) dan *One Versus All* (OVA).

Penjelasan untuk proses pembobotan dengan TF-IDF adalah sebuah teknik perhitungan bobot setiap kata hasil ekstraksi dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Penjelasan dari metode ini adalah *Term Frequency* (TF) merupakan proses untuk menghitung jumlah kemunculan term dalam satu dokumen. Sedangkan *Inverse Document Frequency* (IDF) digunakan untuk menghitung term yang muncul di berbagai dokumen (komentar) yang dianggap sebagai term umum atau dinilai tidak penting (Akbari et al., 2017). Berikut ini adalah beberapa langkah pada teknik pembobotan TF-IDF yakni:

1. Lakukan perhitungan *Term Frequency* (tft,d)
2. Lakukan perhitungan (Wtf t,d) / *Weighting Term Frequency*

$$Wtf_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d} & , \text{ if } tf_{t,d} > 0 \\ 0 & , \text{ if } tf_{t,d} = 0 \end{cases} \quad (1)$$

3. Lakukan perhitungan DF / *Document Frequency*
4. Lakukan perhitungan bobot IDF / *Inverse Document Frequency*

$$idf_t = \log_{10} \frac{N}{df} \quad (2)$$

5. Dan lakukan perhitungan nilai bobot TF-IDF

$$Wt_{t,d} = Wtft_{t,d} \times idft \quad (3)$$

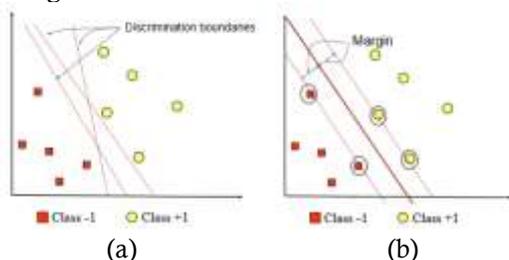
Proses selanjutnya adalah proses ekstraksi fitur. Tujuannya adalah mengubah data yang tak terstruktur menjadi data terstruktur dengan memanfaatkan proses *text mining*. Salah satu teknik yang digunakan adalah Unigram, yakni ekstraksi fitur ke kata tunggal. Tabel 8 menunjukkan cara kerja unigram dalam melakukan pemisahan dokumen pesan twitter menjadi kata tunggal dengan cara ekstraksi fitur.

Tabel 8. Hasil Dari Proses Ekstraksi Fitur Unigram

rumah saya banjir	“rumah”, “saya”, “banjir”
jantung luap ratus rumah	“jantung”, “luap”, “ratus”,
depok rendam banjir	“rumah”, “depok”,
hujan deras buat	“rendam”, “banjir”,
sejumlah rumah depok	“hujan”, “deras”, “buat”,
rendam banjir	“rendam”, “banjir”.

Proses selanjutnya adalah membuat model untuk melakukan pengujian dengan melakukan pemisahan atau *splitting* data dengan teknik *k-fold cross validation*. Teknik ini bertujuan untuk membagi data menjadi sejumlah k bagian yang mempunyai ukuran sama. Setiap sub sampel akan mengalami sebagai data training dan data tesing sebanyak k kali. Teknik k-fold cross validation memiliki keuntungan lebih dari pengulangan sampel acak sebagai data training dan data testing dan validasi dilakukan setidaknya sekali (Raju et al., 2018).

Selanjutnya adalah algoritma yang digunakan untuk proses klasifikasi ditetapkan menggunakan SVM. Algoritma ini termasuk dalam teknik yang baik untuk kasus klasifikasi maupun regresi untuk prediksi data. Teknik SVM dapat digunakan untuk memberikan solusi pada kasus linear maupun non-linear dengan memasukkan konsep kernel pada 2-dimensi maupun dimensi tinggi. Teknik SVM ialah sebuah teknik yang menggunakan *hyperplane* sebagai pemisah data secara linier. Untuk menanggulangi pengolahan data secara non linier dapat menggunakan metode *kernel trick*. Cara kerja algoritma SVM adalah dengan menemukan *hyperplane* yang dapat memisahkan set antara data dari dua kelas yang berbeda seperti pada gambar 2.



Gambar 2. Diagram proses SVM

Secara umum, algoritma SVM hanya dimanfaatkan untuk melakukan klasifikasi ke dalam dua kelas. Selanjutnya, telah dikembangkan untuk klasifikasi lebih dari dua kelas atau *multiclass*. Pada dasarnya terdapat dua pendekatan untuk kasus *multiclass* yang sering

dimanfaatkan pada algoritma SVM yakni teknik OVO dan OVA. Teknik OVO adalah salah satu teknik klasifikasi dengan membandingkan satu kelas dengan satu kelas lainnya. Sedangkan, teknik algoritma OVA adalah dengan membandingkan sebuah kelas dengan semua kelas selain dirinya masih dalam satu kesatuan. Penelitian yang telah dilakukan dengan memanfaatkan algoritma SVM dengan pendekatan teknik OVA untuk kasus *multiclass*. Penyetingan lain yang dimanfaatkan adalah dengan melakukan setting parameter kernel *Radial Basis Function* (RBF) pada *package kernlab*.

Seperti penelitian yang telah dilakukan sebelumnya, metode evaluasi performansi berupa *Confusion matrix*. Metode evaluasi *Confusion matrix* merupakan salah satu teknik evaluasi untuk mengukur kinerja model klasifikasi yang mana termasuk dalam *supervised learning* pada kecerdasan buatan. Perhitungan *Confusion Matrix* atau disebut dengan *Error Matrix* adalah dengan melakukan perbandingan hasil klasifikasi oleh model dengan hasil klasifikasi yang seharusnya. *Confusion matrix* berupa tabel matriks yang memberikan gambaran hasil kinerja model klasifikasi pada data testing dengan nilai sebenarnya yang diketahui (Annisa Rahmaniari Dwi Pratiwi & Erwin Budi Setiawan, 2020). Selanjutnya untuk menghitung akurasi dari evaluasi kasus klasifikasi adalah dengan rumus 4.

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (4)$$

		Aktual	
		1	0
Prediksi	1	True Positive (TP)	False Positive (FP)
	0	False Negative (FN)	True Negative (TN)

Gambar 3. Confusion matrix

HASIL DAN PEMBAHASAN

Penelitian yang telah dilakukan dengan klasifikasi pesan tweet terkait bencana alam dengan algoritma SVM dan kernel berbasis RBF dan type spoc-svc. Algoritma yang digunakan dengan SVM dengan pendekatan OVO dan OVA seperti yang telah dilakukan juga pada penelitian sebelumnya (Dhina Nur Fitriana & Yuliant Sibaroni, 2020). Metode OVO dimanfaatkan untuk klasifikasi dalam dua kelas, sedangkan metode OVA lebih utama untuk metode lebih dari dua kelas atau dalam hal ini adalah 3(tiga) kelas. Nilai akurasi dengan algoritma SVM dan pendekatan OVO mencapai nilai akurasi tertinggi hingga 77.90% seperti ditunjukkan pada tabel 9 (Delimayanti et al., 2020). Sedangkan pada

penelitian ini menggunakan algoritma SVM type spoc-svc dengan kernel RBF untuk tiga kelas label yakni label kelas saksi mata (*eyewitness*), kelas non-saksi mata (*non-eyewitness*) and kelas tidak diketahui (*None*). Setiap kelas mempunyai sensitifitas yang berbeda. Teknik pemilihan kernel RBF memberikan hasil akurasi yang lebih baik dibandingkan dengan hasil menggunakan kernel Polynomial. Teknik ini berhubungan dengan perhitungan hasil kali dalam ruang fitur antara dua argument vector. Dalam beberapa penelitian, kernel RBF yang menggunakan Gaussian dapat memberikan hasil akurasi lebih baik dibandingkan kernel lainnya (Prayoginingsih & Kusumawardani, 2018) (Nguyen et al., 2020).

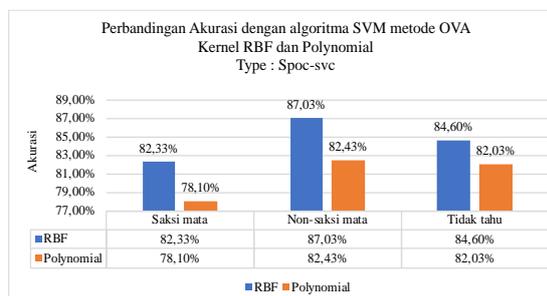
Tabel 9. Hasil akurasi teknik OVO

One Versus One (OVO) : Kernlab_spoc-svc				
Label	Type	Kernel	Parame ter	Sensitifitas Akurasi
Saksi mata	spoc-svc	RBF	default	79,40%
Non-Saksi mata			77,90%	
Tidak tahu	spoc-svc	Polynomial	default	71,60%
Saksi mata			72,77%	
Non-Saksi mata				76,50%
Tidak tahu				75,30%
				66,50%

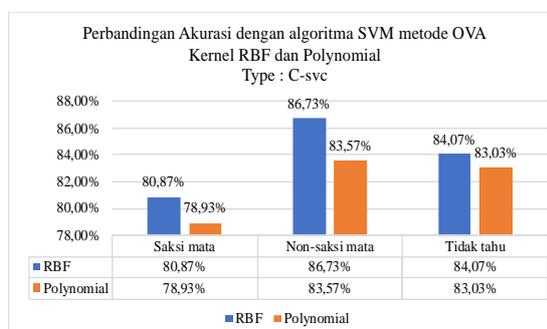
Berikutnya adalah percobaan dengan algoritma SVM untuk pendekatan multiclass *One Versus All* (OVA) dengan type kernel RBF dan Polynomial. Penelitian dilakukan dengan membedakan type spoc-svc dan type C-svc pada metode SVM. Spoc dan C merupakan parameter regularisasi yang membantu implementasi penalti pada kesalahan klasifikasi pada saat memisahkan class. Dengan demikian membantu dalam meningkatkan akurasi pada proses klasifikasi. Nilai akurasi yang didapatkan dengan SVM metode OVA dengan type Spoc-svc dan kernel RBF menunjukkan hasil 87.03%, yang merupakan nilai yang terbaik dibandingkan SVM dengan metode OVO maupun OVA dengan berbagai type kernel.

Type *Spoc-svc* – *Crammer-Singer native multiclass* yang lebih tepat penggunaan untuk kasus klasifikasi multi class seperti pada kasus penelitian ini. Penggunaan type C-svc lebih tepat adalah untuk klasifikasi 2 class seperti pada penelitian Fuchida yakni untuk menggolongkan jenis nyamuk maupun serangga lainnya (Fuchida et al., 2017). Perbandingan nilai akurasi hasil klasifikasi dengan metode OVA dengan type Spoc-svc dan type C-svc dan kernel RBF dan

Polynomial ditunjukkan pada gambar 4 dan Gambar 5.



Gambar 4. Perbandingan akurasi dengan algoritma SVM metode OVA kernel RBF dan polynomial untuk type spoc-svc



Gambar 5. Perbandingan akurasi dengan algoritma SVM metode OVA kernel RBF dan polynomial untuk type c-svc

Gambar 3 telah memberikan informasi bahwa nilai akurasi mencapai 87.03% dan merupakan nilai akurasi tertinggi untuk kelas non-saksi mata. Hasil tersebut didapatkan dengan algoritma SVM dan setting kernel RBF dengan tipe Spoc-svc. Hal ini terlihat jelas bahwa pendekatan multiclass OVA pada algoritma SVM dapat memperbaiki nilai akurasi hasil pengolahan data twitter untuk bencana alam banjir ke dalam tiga (3) kelas informasi. Hasil ini memperbaiki hasil riset sebelumnya dengan algoritma SVM namun dengan pendekatan OVO karena untuk klasifikasi ke dalam dua (2) kelas. Nilai akurasi yang didapat mencapai 77.87% (Delimayanti et al., 2020).

SIMPULAN

Pada penelitian ini telah dilakukan klasifikasi pesan data media sosial twitter yang berisi pesan twitter bencana banjir dengan menggunakan algoritma SVM dan metode multiclass OVA. Pada penelitian sebelumnya telah dilakukan dengan menggunakan berbagai teknik feature

selection dan algoritma SVM sehingga mendapatkan hasil akurasi yang lebih rendah dari hasil saat ini (Delimayanti et al., 2020).

Sedangkan hasil penelitian saat ini menunjukkan bahwa nilai akurasi tertinggi didapatkan dengan metode OVA yang merupakan metode yang tepat untuk multi class seperti pada kasus penelitian ini yakni ada tiga (3) kelas yakni kelas saksi mata, kelas non-saksi mata dan kelas tidak tahu. Selain itu type pada algoritma SVM adalah menggunakan type Spocsvc untuk kasus multiclass dengan kernel RBF yang berbasis pada Gaussian, dengan nilai akurasi tertinggi adalah 87.03%.

Hasil penelitian dapat dikembangkan untuk pemantauan dan peringatan bencana alam banjir dalam bentuk sistem aplikasi berbasis web. Selain itu, pemantauan bencana alam dapat dimanfaatkan pada jenis bencana alam lain seperti kebakaran hutan yang banyak terjadi di luar pulau jawa.

DAFTAR PUSTAKA

- Akbari, M., Novianty, A., & Setianingsih, C. (2017, August). *Analisis Sentimen Menggunakan Metode Learning Vector Quantization Sentiment Analysis Using Learning Vector Quantization Method*.
- Annisa Rahmaniari Dwi Pratiwi & Erwin Budi Setiawan. (2020). Implementation of Rumor Detection on Twitter Using the SVM Classification Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 782–789. <https://doi.org/10.29207/resti.v4i5.2031>
- BNPB. (2020). *Infografis Update Data Bencana Alam di Indonesia* [Infografis]. Badan Nasional Penanggulangan Bencana. <https://bnpb.go.id/infografis/update-bencana-indonesia-tahun-2020>
- Christakis, N. A., & Fowler, J. H. (2010). Social Network Sensors for Early Detection of Contagious Outbreaks. *PLoS ONE*, 5(9), e12948. <https://doi.org/10.1371/journal.pone.012948>
- Delimayanti, M. K., Sari, Risna, Laya, Mauldy, Faisal, M. Reza, Pahrul, & Naryanto, R. Fitri. (2020, October 7). The Effect of Pre-Processing on the Classification of Twitter's Flood Disaster Messages Using Support Vector Machine Algorithm. *International Conference on Applied Engineering (ICAE) 2020*. International Conference on Applied Engineering (ICAE), Batam, Indonesia.
- Devid Haryalesmana & Martijn Wieriks. (2016, February 1). *Kamus id stopword Bahasa Indonesia* [Dataset GItHub]. <https://github.com/masdevid/ID-Stopwords>
- Dhina Nur Fitriana & Yuliant Sibaroni. (2020). Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 846–853. <https://doi.org/10.29207/resti.v4i5.2231>
- Falahah, & Dwiki Adriadi Nur, D. (2015). PENGEMBANGAN APLIKASI SENTIMENT ANALYSIS MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus Sentiment Analysis dari media Twitter). *Seminar Nasional Sistem Informasi Indonesia, November, 2–3*.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press. <http://www.books24x7.com/marc.asp?bookid=23164>
- Fitriyiyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 5(3), 279. <https://doi.org/10.26418/jp.v5i3.34368>
- Fuchida, M., Pathmakumar, T., Mohan, R., Tan, N., & Nakamura, A. (2017). Vision-Based Perception and Classification of Mosquitoes Using Support Vector Machine. *Applied Sciences*, 7(1), 51. <https://doi.org/10.3390/app7010051>
- Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Perez-Meana, H., Portillo-Portillo, J., Sanchez, V., & García Villalba, L. (2019). Using Twitter

- Data to Monitor Natural Disaster Social Dynamics: A Recurrent Neural Network Approach with Word Embeddings and Kernel Density Estimation. *Sensors*, 19(7), 1746.
<https://doi.org/10.3390/s19071746>
- Kryvasheyev, Y., Chen, H., Moro, E., Van Hentenryck, P., & Cebrian, M. (2015). Performance of Social Network Sensors during Hurricane Sandy. *PLOS ONE*, 10(2), e0117288.
<https://doi.org/10.1371/journal.pone.0117288>
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 2(11), 4704–4713.
- Nguyen, H., Bui, X.-N., Choi, Y., Lee, C. W., & Armaghani, D. J. (2020). A Novel Combination of Whale Optimization Algorithm and Support Vector Machine with Different Kernel Functions for Prediction of Blasting-Induced Fly-Rock in Quarry Mines. *Natural Resources Research*.
<https://doi.org/10.1007/s11053-020-09710-7>
- Oryza Habibie Rahman, Gunawan Abdillah, & Agus Komarudin. (2021). Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 17–23.
<https://doi.org/10.29207/resti.v5i1.2700>
- Pradana, A. W., & Hayaty, M. (2019). The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(4), 375–380.
<https://doi.org/10.22219/kinetik.v4i4.912>
- Pratama, M. L., & Murfi, H. (2014). *STUDI KOMPARASI METODE MULTICLASS SUPPORT VECTOR MACHINE UNTUK MASALAH ANALISIS SENTIMEN PADA TWITTER*. 20.
- Prayoginingsih, S., & Kusumawardani, R. P. (2018). Klasifikasi Data Twitter Pelanggan Berdasarkan Kategori myTelkomsel Menggunakan Metode Support Vector Machine (SVM). *Sisfo*, 07(02).
<https://doi.org/10.24089/j.sisfo.2018.01.002>
- Rachmatika, R., & Bisri, A. (2020). Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 6(3), 6.
- Raju, K. S., Murty, M. R., Rao, M. V., & Satapathy, S. C. (2018). Support Vector Machine with K-fold Cross Validation Model for Software Fault Prediction. *International Journal of Pure and Applied Mathematics*, 118(20), 321–334.
- Wu, C.-H. (2016). SOCIAL SENSOR: AN ANALYSIS TOOL FOR SOCIAL MEDIA. *International Journal of Electronic Commerce Studies*, 7(1), 77–94.
<https://doi.org/10.7903/ijecs.1411>
- Zahra, K., Imran, M., & Ostermann, F. O. (2020). Automatic identification of eyewitness messages on twitter during disasters. *Information Processing and Management*, 57(1), 102107.
<https://doi.org/10.1016/j.ipm.2019.102107>