

Sentimen Analisis Opini Masyarakat Jakarta Pada Kinerja Pemerintah Jakarta Terhadap Isu Tenggelamnya Jakarta Menggunakan Algoritma *Support Vector Machine*

Adithya Kharisma¹, Iin Ernawati²

Program Studi S1 Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450
adithyak@upnvj.ac.id¹, iinernawati@upnvj.ac.id²

Abstrak. Isu tenggelamnya Jakarta menimbulkan berbagai opini masyarakat terutama masyarakat Jakarta. Opini tersebut menimbulkan pro dan kontra, Opini masyarakat dapat dijadikan sebagai ukuran untuk melihat kepuasan masyarakat terhadap kinerja pemerintah Jakarta, hal ini menimbulkan permasalahan yaitu bagaimana membuat opini menjadi sebuah data bermanfaat untuk model SVM, oleh karena itu dilakukan analisis sentimen untuk melihat kepuasan masyarakat terhadap kinerja pemerintah. Analisis sentimen yang dilakukan memiliki 2 label yaitu label positif dan negatif. kemudian diolah menggunakan *Natural Language Processing* atau *text mining* dan dibuat menjadi 3 buah dataset yaitu dataset awal, dataset *undersampling* dan dataset *oversampling*. Kemudian ketiga dataset tersebut dijadikan data untuk pembuatan model SVM dengan parameter *cost* dan kernel yang berbeda, setelah pembuatan model kemudian dilakukan evaluasi *confussion matrix*. Hasil evaluasi tertinggi dari 3 model dataset tersebut adalah dataset *oversampling* dengan nilai rata-rata akurasi sebesar 93%, *precision* sebesar 96%, *recall* sebesar 90%, *specificity* sebesar 97% dan *f1-score* sebesar 93%.

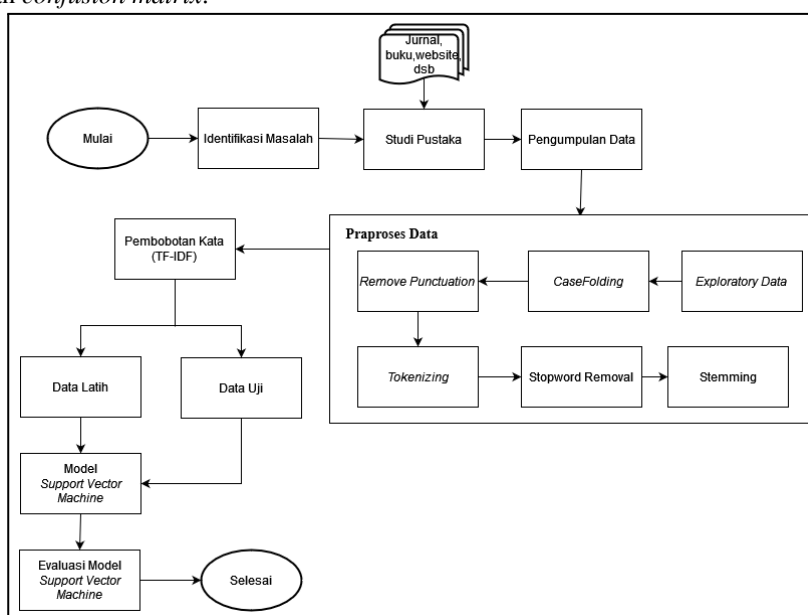
Kata Kunci: Sentimen Analisis, Isu Jakarta Tenggelam, SVM, *Text Mining*, *Confussion Matrix*.

1 Pendahuluan

Zaman kontemporer dimana teknologi internet mengalami perkembangan yang pesat, informasi berjalan secara cepat dan membuat berbagai pengaruh positif ataupun negatif kepada penggunanya. Salah satu pengaruh dari berkembangnya teknologi internet adalah manusia bisa saling terhubung kepada manusia lain tanpa mempedulikan jarak. Hal ini membuat masyarakat terutama di Indonesia bisa saling berbagi informasi. Informasi tersebut sangatlah beragam mulai dari informasi pribadi hingga informasi publik seperti berita. Pesatnya perkembangan teknologi informasi yang modern ini harus dimanfaatkan dengan baik, sehingga pengetahuan umum dapat bertambah seperti memperluas wawasan atau menyebarkan manfaat yang positif sehingga bisa menjadi hal optimis. Penggunaan media sosial ini juga dapat dimanfaatkan oleh pemerintah Indonesia dan masyarakat untuk melihat perkembangan Indonesia.

Dari penyebaran berita di Indonesia yang dibagikan di media sosial terutama berita terhadap isu tenggelamnya Jakarta menimbulkan berbagai macam opini masyarakat Indonesia terutama di Jakarta, hal ini berkaitan dengan kinerja pemerintah Indonesia atau pemerintah daerah jakarta mengenai isu ini. Dari berita isu tersebut timbul banyaknya opini masyarakat Jakarta yang bisa dijadikan sebagai data yang bermanfaat untuk evaluasi kinerja pemerintah. Dari hal tersebut menimbulkan permasalahan yaitu bagaimana caranya untuk mengolah opini masyarakat yang berbentuk *natural language* menjadi sebuah data yang bermanfaat untuk melihat sentimen yang ada di masyarakat jakarta dan apakah model SVM menghasilkan hasil evaluasi yang bagus. Pada analisis sentimen, ulasan pengguna dikelompokkan menjadi dua label. yaitu sentimen positif dan negatif [1]. Tujuan utama yang terdapat pada sentimen analisis adalah untuk mengidentifikasi aspek positif atau negatif pada teks yang terdapat pada dokumen, sehingga dapat menentukan kalimat atau pendapat tertentu positif, negatif, atau netral [2]. Tujuan dari penelitian ini yaitu melakukan pembentukan model SVM dari sentimen yang ada kemudian mengukur hasil evaluasi *confussion matrix*.

Dalam penelitian terdahulu yang dilakukan oleh M. Reza F menggunakan SVM menghasilkan akurasi 97% dengan dataset ulasan aplikasi PPID[2]. Penelitian lainnya yang dilakukan M. Fadli dan Kemas menghasilkan akurasi SVM sebesar 94,36% yang dilakukan pada analisis sentimen kinerja PSSI[3]. Penelitian yang dilakukan oleh Tri rivanie dkk menghasilkan akurasi SVM sebesar 72,57% pada dataset opini kinerja menteri kesehatan selama pandemi covid-19[4] dan penelitian yang dilakukan oleh Imam N. F. Dkk menghasilkan akurasi SVM 70.39% dengan dataset kuesioner kepuasan mahasiswa terhadap layanan dan fasilitas kampus[5], oleh karena itu peneliti melakukan riset dengan melakukan sentimen analisis terhadap opini masyarakat Jakarta melalui survey singkat berupa kuesioner yang dibagikan di beberapa media sosial mengenai kinerja pemerintah Jakarta terhadap isu tenggelamnya Jakarta untuk diketahui evaluasi dari model SVM yang dibuat. Opini masyarakat tersebut akan menjadi data yang akan diolah menggunakan NLP (*Natural Language Processing*) seperti melakukan *case folding* untuk mengubah text menjadi lebih konsisten, melakukan *tokenizing*, *filtering*, dan *stemming* pada data. label yang diberikan pada data, ada dua yaitu label positif dan negatif. Setelah itu dilangsungkan pemodelan *machine learning* menggunakan *text transformation* dan Algoritma SVM. Algoritma tersebut kemudian di evaluasi menggunakan *confusion matrix*.



2. Metode Penelitian

Gambar. 1. Tahapan Metode Penelitian

2.1 Identifikasi Masalah

Langkah awal dengan mengidentifikasi dan mengartikulasikan masalah. Hal ini dilakukan untuk menemukan permasalahan yang relevan di lingkungan penelitian. Pada tahap ini peneliti menemukan masalah yaitu berita mengenai isu tenggelamnya Jakarta semakin populer maka peneliti mencoba untuk melakukan sentimen analisis opini masyarakat Jakarta terhadap kinerja pemerintah Jakarta menggunakan algoritma *support vector machine* kemudian melihat hasil evaluasi dari pemodelan SVM menggunakan *confusion matrix*.

2.2 Studi Pustaka

Pada tahap ini dilakukan studi literatur terhadap penelitian terdahulu untuk dijadikan sebagai referensi selama penelitian berlangsung. Sumber referensi yang digunakan antara lain jurnal, *youtube*, *website* dan penelitian terdahulu terkait sentimen analisis bahasa Indonesia dengan algoritma *support vector machine*.

2.3 Pengumpulan Data

Data dikumpulkan dengan melakukan survey singkat yaitu peneliti membuat sebuah survey singkat mengenai opini masyarakat Jakarta terhadap kinerja pemerintah Jakarta mengenai isu tenggelamnya Jakarta. Survey ini dibuat menggunakan google form kemudian survey tersebut dibagikan melalui beberapa media sosial seperti *whatsapp*, *discord*, *line* dan lain-lain. Pada penelitian ini pengumpulan data dilakukan pada bulan Januari tahun 2023 dan mendapatkan sebanyak 253 data, kemudian data dilakukan pemberian sentimen negatif atau positif oleh 3 orang anotorator. Hasil dari tahapan ini adalah dataset dalam bentuk format file CSV.

2.4 Praproses Data

Praproses dilakukan untuk mengolah sebuah data kotor menjadi data yang bersih untuk pembuatan model[6]. Pada tahapan ini yaitu.

2.4.1 *Exploratory Data*

Pada tahapan ini peneliti melakukan *exploratory data* atau penjelajahan pada data untuk memahami lebih dalam mengenai dataset yang diteliti, seperti melihat jumlah responden dan persebaran data responden, melihat *imbalance data* pada dataset dan memperbaiki data yang *imbalance* menggunakan *random undersampling* dan *random oversampling* dan info atau deskripsi dataset sehingga penulis bisa menganalisis dan melakukan tahapan praproses selanjutnya dengan benar.

2.4.2 *Case Folding*

Tahapan yang bertujuan untuk mengubah seluruh data teks menjadi huruf kecil atau *lower case* agar data yang dipakai konsisten [6].

2.4.3 *Remove Punctuation*

Setelah melakukan *case folding* pada data kemudian tahapan selanjutnya adalah melakukan *remove punctuation*, pada tahapan ini data dibersihkan dari simbol yang tidak diperlukan seperti tanda tanya, tanda seru, tanda “@”, tanda titik dan lainnya[7].

2.4.4 *Tokenizing*

Tokenizing merupakan proses pemisahan teks berupa kalimat menjadi bagian teks kata atau token. Pada tahapan ini kalimat atau opini yang telah di bersihkan dari simbol kemudian diubah menjadi sebuah token kecil atau diubah dari sebuah kalimat menjadi kata per kata[7].

2.4.5 *Stopword Removal*

Tujuan dari tahapan ini untuk mengurangi jumlah fitur kata dengan memilih kata-kata yang relevan dari hasil *tokenizing* dan menghilangkan kata yang kurang diperlukan, seperti "yang mana", "kecuali", "tetapi", "dan", "bisa", dan sebagainya. Pada tahapan ini dihapuskan kata yang tidak memiliki pengaruh seperti kata penghubung, kata “adalah”, “yaitu”, dan lainnya, hal ini dilakukan agar program bisa lebih fokus terhadap kata yang penting[8].

2.4.6 *Stemming*

Metode untuk mengubah kata berimbuhan ke dalam bentuk dasar. Pada tahapan ini dihapuskan imbuhan dalam bahasa Indonesia menjadi sebuah kata dasar. Hal ini dilakukan agar mempermudah dalam pembobotan kata [9].

2.5 Pembobotan data

Salah satu pembobotan kata yaitu TF-IDF menggunakan vector space model. TF-IDF merupakan teknik yang menggabungkan Term Frequency (TF) atau jumlah kemunculan *term* kata dan Inverse Document Frequency yang merupakan metode pengukuran kata penting dalam sebuah dokumen dengan menghitung kemunculan sebuah istilah yang ada pada semua dokumen. Sedikitnya istilah yang ada di semua dokumen maka nilai idf nya akan besar. Rumus TF-IDF dapat dilihat sebagai berikut:

$$w_{t,d} = w_{tf} \times idf_t \quad (1)$$

Keterangan:

$w_{t,d}$ = Pembobotan TF-IDF

w_{tf} = Bobot frekuensi kata

idf_t = Bobot kata penting

2.6 Pembagian Data Latih dan Data Uji

Tahapan ini membagi data menjadi data latih dan data uji pada dataset. Data yang digunakan untuk melatih model disebut data latih dengan tujuan agar model dapat melakukan prediksi terhadap data tes atau data uji, sementara data uji berfungsi untuk menguji model yang telah dibuat.

2.7 Pembuatan Model *Support Vector Machine*

Tahapan ini menggunakan algoritma klasifikasi SVM untuk pembelajaran mesin dengan bertujuan mencari garis batas maksimum, yang disebut Maximum Marginal Hyperplane (MMH), untuk memisahkan kelas-kelas yang ada. Support Vector Machine mencoba mencari garis batas dengan jarak terpendek ke salah satu sisi batas dengan jarak yang sama ke batas bagian lainnya, dengan asumsi kedua batas sejajar dengan garis batas (hyperplane). Posisi hyperplane diantara dua kelas, Hal ini berarti jarak pembatas hyperplane dengan objek data berbeda dengan kelas luar yang berdekatan. Data terluar yang paling dekat dengan pembatas hyperplane disebut support vector. Objek yang disebut support vector paling sulit untuk diklasifikasikan karena posisi yang hampir tumpang tindih (overlap) dengan kelas lain. karena sifat support vector yang kritis, hanya support vector inilah yang diperhitungkan untuk menemukan hyperplane yang paling optimal oleh SVM [20]. Cara mendapatkan garis hyperplane yang paling optimal ketika memisahkan data dari dua kelas, perhitungan margin hyperplane digunakan dan titik maksimum ditemukan. Saat mendapatkan hyperplane di SVM, Dapat menggunakan persamaan [10]. Berikut rumus minimalkan margin:

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2 + \dots + w_n^2) \quad (2)$$

Dengan Keterangan:

w = Parameter pembatas *hyperplane*

Proses SVM dapat bekerja dengan memetakan data ke dalam fitur ruang dimensi tinggi sehingga dapat diklasifikasikan pada data. Ketika garis pemisah ditemukan di antara kategori, data dapat diubah sehingga pembagi batas dapat didefinisikan sebagai hyperplane. Properti data baru dapat digunakan untuk menentukan kelompok mana data baru yang harus dikumpulkan. Selain linear, SVM dapat memanipulasi kernel non-linear melalui penggunaan trik kernel. Menggunakan trik kernel, input diberikan ke ruang fitur dimensi tertinggi dan hasil input dibedakan secara linear dan membentuk hyperplane yang tepat [11].

2.9 Evaluasi Model *Support Vector Machine*

Evaluasi pada tahapan ini dilakukan menggunakan *Confussion Matrix* hal ini agar model bisa diukur nilai akurasi, *recall*, *precision*, *specificity* dan *F1-Score* dari model yang dibuat. Pada evaluasi *confusion matrix* menggunakan 4 hasil dari nilai prediksi dan nilai aktual[12].

Tabel. 1. Evaluasi *Confusion Matrix*.

	Aktual Positif	Aktual Negatif
Prediksi Positif	True Positive (TP)	False Positive (FP)
Prediksi Negatif	False Negative (FN)	True Negative (TN)

Keterangan:

True Negative = Nilai aktual negatif dan nilai hasil prediksi negatif

False Negative = Nilai aktual positif dan hasil nilai prediksi negatif

False Positive = Nilai aktual negatif dan hasil prediksi positif

True Positive = Nilai aktual dan prediksi positif

Berikut merupakan rumus untuk menghitung akurasi, *recall*, *precision*, *specificity*, *F1-Score*:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Akurasi berfungsi untuk mengukur seberapa akurat model dalam melakukan keseluruhan prediksi model.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Recall berfungsi untuk mengukur rasio benar positif dari keseluruhan data yang bernilai benar positif.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Precision berfungsi untuk mengukur nilai yang diprediksi positif dan benar masuk kedalam kategori positif.

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

Specificity berfungsi untuk melihat proporsi benar dalam prediksi negatif dibandingkan dengan jumlah keseluruhan data negatif.

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (7)$$

F1 Score berfungsi untuk mengukur bobot dari rata-rata *recall* dan *precision*

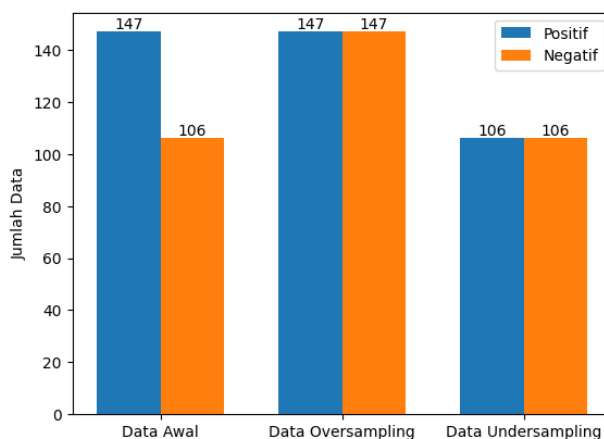
3 Hasil Penelitian

Data yang digunakan dalam penelitian ini adalah data kuesioner yang diperoleh dari google form yang dibagikan ke masyarakat jakarta di beberapa sosial media seperti aplikasi line, whatsapp, dan discord. Data utama pada penelitian ini yaitu opini atau sentimen masyarakat Jakarta pada kinerja pemerintah jakarta terhadap isu tenggelamnya jakarta. Data yang diperoleh dari penelitian ini memiliki kolom Timestamp (Stempel Waktu), Nama, Gender, Asal Kota, kolom pertanyaan "Apakah kamu mengetahui mengenai isu bahwa jakarta akan tenggelam?", kolom kepuasan "Seberapa kamu merasa senang dengan kinerja pemerintah selama ini?", kolom kepuasan "Seberapa kamu merasa puas dengan kinerja pemerintah dalam mengatasi isu Jakarta yang akan tenggelam?" dan kolom opini atau sentimen dengan nama kolom "Berikan pendapat anda atau opini anda mengenai kinerja pemerintah dalam mengatasi Jakarta yang akan tenggelam". Data berjumlah 253 baris yang dikumpulkan dari 2 Januari 2023 hingga 29 Januari 2023.

3.1 Praproses Data

Setelah data didapatkan, dilakukan praproses data dengan beberapa langkah-langkah. Untuk pengerjaan praproses

data ini berupa rentetan proses yang berurutan seperti melakukan *balancing* pada data dengan *random oversampling* dan *random undersampling*, kemudian melihat jumlah data sentimen positif dan negatif pada setiap dataset yang ditunjukkan pada gambar 2 berikut:



Gambar. 2. Perbandingan Jumlah Data

Dataset yang telah dilakukan *balancing* kemudian akan diproses lebih lanjut seperti melakukan *case folding*, *remove punctuation*, *tokenizing*, *stopword removal* dan *stemming* berikut merupakan data opini yang ada pada kuesioner yang ditunjukkan pada tabel 2:

Tabel. 2. Data Opini pada Kuesioner.

No	Opini Masyarakat	Label Sentimen
1	Menurut saya upaya pemerintah dalam mengatasi masalah tersebut sudah cukup baik	Sentimen Positif
2	Menurut saya upaya pemerintah untuk menangani isu tenggelamnya jakarta ini sangat kurang, baik itu secara praktik, sosialisasi, maupun penanganannya	Sentimen Negatif
...
252	Kinerja pemerintah dalam upaya mengatasi isu jakarta yang akan tenggelam sudah cukup baik	Sentimen Positif
253	Upaya yang dilakukan pemerintah untuk mencegah tenggelamnya jakarta sudah maksimal	Sentimen Positif

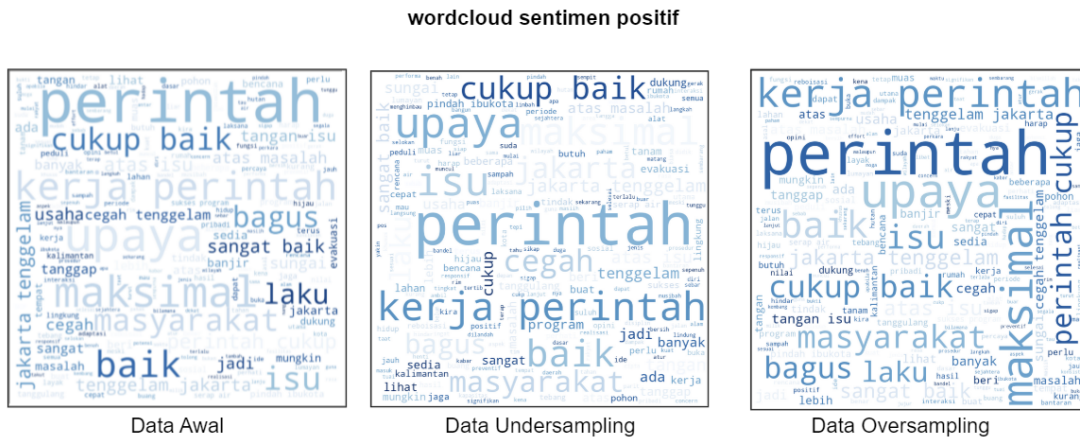
Data pada tabel 2 diatas kemudian dilakukan praproses lebih lanjut sehingga menghasilkan kalimat seperti yang ditunjukkan pada tabel 3:

Tabel. 3. Praproses Data.

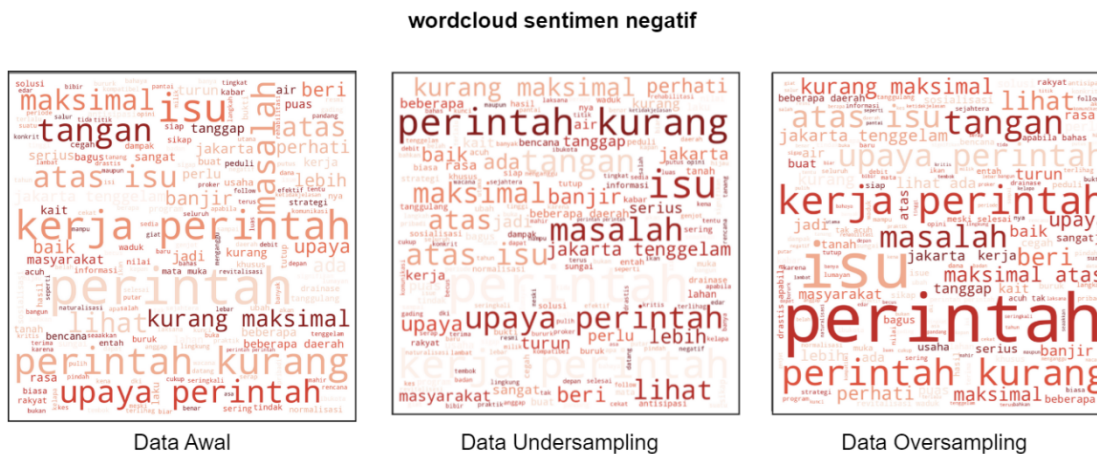
No	Kalimat Sebelum Praproses	Kalimat setelah Praproses
1	Semua upaya telah dilaksanakan oleh pemerintah secara maksimal, hanya saja banyak masyarakat yang belum paham betul, apa penyebab munculnya isu jakarta tenggelam	semua upaya laksana perintah maksimal banyak masyarakat paham betul apa sebab muncul isu jakarta tenggelam

No	Kalimat Sebelum Praproses	Kalimat setelah Praproses
2	Menurut saya, kinerja pemerintah untuk mencegah terjadinya isu Jakarta telah tenggelam sudah baik, namun hanya alam yang dapat mendukung semua kinerja pemerintah dalam mensukseskan program-program yang telah dibuat.	kerja perintah cegah jadi isu jakarta tenggelam baik alam dukung semua kerja perintah sukses program program buat
...
252	kinerja pemerintah dalam upaya mengatasi isu jakarta yang akan tenggelam sudah cukup baik	kerja perintah upaya atas isu jakarta tenggelam cukup baik
253	upaya yang dilakukan pemerintah untuk mencegah tenggelamnya jakarta sudah maksimal	upaya laku perintah cegah tenggelam jakarta maksimal

Berikut merupakan visualiasi *wordcloud* pada dataset yang telah dilakukan praproses dimulai dari tahapan *explaratory data* hingga ke tahapan *stemming*. Sehingga data yang dihasilkan sudah bersih.



Gambar. 3. Wordcloud Sentimen Positif pada Dataset



Gambar. 4. Wordcloud Sentimen Negatif pada Dataset

3.2 Pembobotan Kata

Setelah dilakukan praproses kemudian dilakukan pembobotan kata, hal ini dilakukan agar data dapat dilatih menggunakan algoritma SVM. Pada pembobotan kata menggunakan text transformation dengan menggunakan perhitungan TF IDF yaitu mengubah kata menjadi sebuah vektor dengan rumus Term Frequency dan Inverse Document Frequency. Berikut tabel 4 yang berisi contoh teks untuk dilakukan TF-IDF:

Tabel. 4. Dokumen Teks TF-IDF.

Doc	Teks	Sentimen
D1	semua upaya laksana perintah maksimal banyak masyarakat paham betul apa sebab muncul isu jakarta tenggelam	Positif
D2	kerja perintah cegah jadi isu jakarta tenggelam baik alam dukung semua kerja perintah sukses program program buat	Positif
D3	kurang puas beberapa daerah seringkali banjir drainase buruk	Negatif

Berikut tabel 5 yaitu contoh hasil perhitungan TF-IDF dari tabel 4:

Tabel. 5. Contoh hitung TF-IDF.

Token	Term Frekuensi			DF	IDF	TF-IDF		
	D1	D2	D3			D1	D2	D3
Alam	-	1	-	1	0.48	0.00	0.48	0.00
Apa	1	-	-	1	0.48	0.48	0.00	0.00
...
Perintah	1	2	-	2	0.18	0.18	0.36	0.00

3.3 Proses Klasifikasi

Data yang sudah ditransformasi menggunakan TF-IDF dan sudah memiliki bobot kata kemudian akan dilakukan proses klasifikasi menggunakan algoritma *machine learning* yaitu algoritma *support vector machine*.

3.3.1 Pembagian Data Latih dan Data Uji

Pada tahapan ini dilakukan pembagian data latih dan data uji dengan 2 perbandingan yaitu perbandingan 80:20 dan perbandingan 70:30, berikut merupakan tabel 6 yang berisi pembagian data latih dan data uji pada setiap dataset:

Tabel. 6. Perbandingan Data Latih dan Uji.

Jenis Data	Perbandingan 80:20		Perbandingan 70:30		Total Data
	Data Latih	Data Uji	Data Latih	Data Uji	
Data Awal	202	51	177	76	253
Data Undersampling	169	43	148	64	212
Data Oversampling	235	59	205	89	294

3.3.1 Pembuatan Model SVM

Setelah dilakukan pembagian data, tahapan selanjutnya adalah melatih data tersebut menggunakan *machine learning* dengan algoritma SVM. Pada pembuatan model klasifikasi SVM ini akan menggunakan kernel yang

berbeda. Kernel yang digunakan pada penelitian ini ada 3 jenis yaitu kernel *linear*, kernel *polynomial* dan kernel RBF (*Radial Basis Function*). Selain kernel parameter yang digunakan yaitu *max_iter*, *cost* (c) dan γ (y). Pada model *support vector machine* parameter γ yang digunakan yaitu “*scale*”, *max_iter* yang digunakan adalah 1000 kali dan juga dilakukan perubahan nilai *cost*(c) dengan nilai (1.0), (5.0) dan (10.0), perubahan nilai ini dilakukan untuk mengetahui pengaruh dari parameter *cost* pada model yang dibuat. Parameter *Cost* atau biasa disebut sebagai C adalah parameter yang bekerja sebagai pengoptimalan SVM untuk menghindari misklasifikasi di setiap sampel dalam dataset training. dan nilai γ (y) yang berfungsi untuk menentukan seberapa banyak lengkungan yang diharapkan dalam batasan keputusan (decision boundary) [13].

3.4 Evaluasi Model

Setelah dilakukan pemodelan machine learning kemudian melakukan evaluasi model untuk mengetahui akurasi dari model yang telah dibuat. Evaluasi model machine learning pada penelitian ini menggunakan confusion matrix. Pada confusion matrix akan menghasilkan nilai TP (True Positive), TN (True Negative), FP (False Positive) dan FN (False Negative). Nilai tersebut dibutuhkan untuk menghitung hasil akurasi, precision, recall, specificity dan F1-Score pada sebuah model machine learning. Berikut merupakan hasil evaluasi confusion matrix:

Tabel. 6. Hasil Evaluasi.

No	Perbandingan Dataset	Jenis Dataset	Nilai Rata-Rata				
			Akurasi	Precision	Recall	Specificity	F1-Score
1	80% Data Latih dan 20% Data Uji	Dataset Awal	0,8147	0.8532	0.8333	0.7883	0.8415
		Dataset <i>Undersampling</i>	0.7260	0.7517	0.6560	0.7928	0.7005
		Dataset <i>Oversampling</i>	0.9359	0.9676	0.9003	0.9703	0.9326
2	70% Data Latih dan 30% Data Uji	Dataset Awal	0.8026	0.8103	0.8661	0.7152	0.8359
		Dataset <i>Undersampling</i>	0.7534	0.8162	0.6562	0.8506	0.7269
		Dataset <i>Oversampling</i>	0.8750	0.9043	0.8358	0.9135	0.8686

Pada tabel 6 dapat diketahui dataset *oversampling* dengan perbandingan 80:20 menghasilkan nilai evaluasi *confusion matrix* yang lebih tinggi dibandingkan dataset yang lainnya. Hasil evaluasi tersebut lalu diubah menjadi persen.

4 Kesimpulan dan Saran

4.1 Kesimpulan

1. Untuk membangun model *machine learning* pada data *natural language* dapat dilakukan yaitu dengan melakukan praproses terlebih dahulu seperti *balancing* pada data kemudian melakukan *case folding*, *remove punctuation*, *tokenizing*, *stopword removal*, *steming* setelah itu dilakukan pembobotan kata dengan TF-IDF dan membagi data latih dan data uji menjadi 80:20 dan 70:30, kemudian dilakukan pemodelan SVM dengan parameter *cost* dan kernel yang berbeda dan di evaluasi *confusion matrix* dengan melihat akurasi, *recall*, *precision*, *specificity* dan *F1 Score* dari dataset original, dataset *undersampling* dan dataset *oversampling* untuk mencari nilai rata-rata hasil evaluasi yang terbaik diantara ketiga jenis dataset.

2. Dari hasil evaluasi algoritma SVM pada dataset sentimen masyarakat Jakarta mengenai kinerja pemerintah Jakarta pada isu tenggelamnya Jakarta menghasilkan hasil evaluasi yang bagus yaitu terdapat pada perbandingan 80% data latih dan 20% data uji menggunakan dataset *oversampling* dengan nilai rata-rata akurasi sebesar 93%, *precision* sebesar 96%, *recall* sebesar 90%, *specificity* sebesar 97% dan *f1-score* sebesar 93%. Sementara dataset

original menghasilkan nilai rata-rata akurasi terbaik pada perbandingan 80:20 dengan akurasi sebesar 81%, *precision* sebesar 85%, *recall* sebesar 83%, *specificity* sebesar 78% dan *f1-score* sebesar 84%. Untuk dataset *undersampling* menghasilkan nilai rata-rata akurasi terbaik pada perbandingan 70:30 dengan akurasi sebesar 75%, *precision* sebesar 81%, *recall* sebesar 65%, *specificity* sebesar 85% dan *f1-score* sebesar 72%.

4.2 Saran

Saran yang dapat diberikan yaitu Menambah jumlah data sehingga data menjadi lebih banyak, menggunakan Parameter yang lebih variatif, sehingga memungkinkan untuk mencapai nilai akurasi terbaik dan menggunakan teknik smote, adasyn dan teknik yang lain dalam memperbaiki data yang tidak seimbang.

5 Referensi

- [1] Sulaeman, A. F., Supianto, A. A., & Bachtiar, A. F. (2019).” Analisis Sentimen Opini Mahasiswa Terhadap Saran Evaluasi Kinerja Dosen Menggunakan TF-IDF dan Support Vector Machine.”. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol.3 No.6, hlm. 5647-5655..
- [2] Fahlevvi, M. R., (2022). “Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine”, Jurnal Teknologi dan Komunikasi Pemerintahan Vol 4, No. 1, hlm 1-13.
- [3] Asshiddiqi, M. F., & Lhaksana, K. M. (2020). “Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI”. e-Proceeding of Engineering : Vol.7, No.3 Desember 2020 hlm 9936.
- [4] Rivanie, T., Pebrianto, R., Hidayat, T., Bayhaqy, A., Gata, W., & Novitasari, B. (2021). “Analisis Sentimen Terhadap Kinerja Menteri Kesehatan Indonesia Selama Pandemi Covid-19”. Jurnal Informatika (Vol. 21, Issue 1).
- [5] Fakhri, I. N., Jondri., & Umbara, R. F., (2019). “Analisis Sentimen pada Kuisisioner Kepuasan Terhadap Layanan dan Fasilitas Kampus Universitas Dengan Menggunakan Klasifikasi Support Vector Machine (SVM)”. e-Proceeding of Engineering : Vol.6, No.2 Agustus 2019.
- [6] Darwis, D., Shintya Pratiwi, E., Ferico, A., & Pasaribu, O. (2020). “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia”. In Jurnal Ilmiah Edutic (Vol. 7, Issue 1).
- [7] Almuayqil, S. N., Humayun, M., Jhanjhi, N. Z., Almufareh, M. F., & Khan, N. A. (2022). “Enhancing Sentiment Analysis via Random Majority Under-Sampling with Reduced Time Complexity for Classifying Tweet Reviews”. Electronics (Switzerland), 11(21).
- [8] Jumeilah, F. S., (2017). “Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian”, Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) Vol. 1, No.1 hlm 19-25.
- [9] Hendrawan, A., Winarti, T., & Indriyawati, H, (2020). “Pengembangan Stemming Untuk Artikel Berbahasa Indonesia”. Laporan Penelitian Universitas Semarang.
- [10] Husada, H. C., & Paramita, A. S. (2021).”Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)”. Teknika, Volume 10(1) hlm 18–26.
- [11] Sudhir, P., & Suresh, V. D. (2021). Comparative study of various approaches, applications and classifiers for sentiment analysis. Global Transitions Proceedings, 2(2), 205–211. <https://doi.org/10.1016/j.gltp.2021.08.004>.
- [12] Pangestu, S. Y., Astuti, Y., & Farida, L. D. (2019). “Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia”. Jurnal Mantik Penusa, 3(1), hlm 236–241.
- [13] Chairunissa, C., (2022). “Klasifikasi Sentimen Ulasan Pengguna Aplikasi Peduli Lindungi Di Google Play Menggunakan Algoritma Support Vector Machine Dengan Seleksi Fitur Chi-Square”. Skripsi Universitas Pembangunan Nasional Veteran Jakarta.