

Implementasi Support Vector Machine pada Analisis Sentimen mengenai Bantuan Sosial di Era Pandemi Covid-19 pada Pengguna Twitter

Davriwan Dzaky Muttaqien¹, Tibyani², Pitoyo Peter Hartono³

^{1,2}Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya

³Departement Electrical and Electronics Engineering, Chukyo University

Email: ¹dzakydavriwan@student.ub.ac.id, ²tibyani@ub.ac.id, ³hartono@sist.chukyo-u.ac.jp

Abstrak

Pandemi COVID-19 menimbulkan berbagai macam permasalahan diberbagai sektor, salah satunya sektor perekonomian yang berdampak pada kesejahteraan masyarakat. Dengan tujuan untuk meningkatkan kesejahteraan masyarakat, pemerintah melalui kementerian sosial menyelenggarakan program bantuan sosial. Seiring dengan berjalannya program tersebut banyak masyarakat yang membicarakan serta memberikan opini mereka terkait program tersebut di sosial media, salah satu sosial media yang banyak digunakan masyarakat untuk membicarakan hal tersebut adalah Twitter, sehingga Twitter dapat dijadikan salah satu sumber untuk melakukan analisis sentimen terhadap bantuan sosial. Penelitian ini bertujuan untuk membangun alat analisis untuk memahami secara intuitif sentimen publik yang beredar di Twitter terkait bantuan sosial selama krisis COVID-19. Pada penelitian ini, Support Vector Machine (SVM) digunakan sebagai mesin utama untuk alat analisis yang diusulkan, karena kekuatannya dalam menangani inkonsistensi data yang secara inheren tertanam dalam tweet subjektif. Hipotesis awal adalah bahwa tidak terdapat perbedaan pada sentimen sosial di antara kelompok usia yang berbeda dalam masyarakat. Pada penelitian ini, alat analisis terdiri dari tiga SVM berbeda, masing-masing dilatih oleh tweet dari kelompok usia terkait. Dalam eksperimen, uji penilaian statistik dilakukan untuk mengevaluasi hipotesis awal, dan hasilnya digunakan lebih lanjut untuk membangun alat analisis, sementara penggunaan alat analisis ini di masa mendatang juga dijelaskan.

Kata kunci: COVID-19, Bantuan Sosial, Analisis Sentimen, P-value, Support Vector Machine, TF-IDF, Cross Validation

Abstract

The COVID-19 pandemic has caused various kinds of problems in various sectors, one of which is the economic sector that has a direct impact on people's welfare. With the aim of escalating people's welfare, the government through the Ministry of Social Affairs runs a social assistance program. Along with the running of the program, many people discussed and gave their opinions regarding the program on social media, one of the social media that is widely used by people to talk about this is Twitter, so Twitter can be used as a source to conduct sentiment analysis on social assistance. The objective of this research is to construct an analytical tool for intuitively understanding public sentiments circulating in Twitter regarding the social assistance during the COVID-19 crisis. Here, Support Vector Machine (SVM) is utilized as the main engine for the proposed analytical tool, due to its strength in dealing with data's inconsistency that is inherently embedded in subjective tweets. The initial hypothesis is that there is no difference in social sentiment among different age groups in society. Here, the analytical tool consists of three different SVM, each trained by the tweet from the associated age group. In the experiment, a statistical assessment test is executed for evaluating the initial hypothesis, and the result is utilized further to build the analytical tool, while the future usage for this analytical tool is also explained.

Keywords: COVID-19, Social Assistance, Sentiment Analysis, P-value, Support Vector Machine, TF-IDF, Cross Validation

1. PENDAHULUAN

Sejak Indonesia mengonfirmasi kasus pertama COVID-19 sudah berbagai cara telah dilakukan oleh pemerintah untuk menekan

pengaruh negatif dari pandemi COVID-19 pada berbagai macam sektor, diantaranya adalah sektor ekonomi. Dengan diberlakukannya pembatasan aktivitas masyarakat mempengaruhi aktivitas bisnis yang akhirnya berimbas pada sektor perkonomian, pertumbuhan ekonomi Indonesia minus 5,32% pada Kuartal II 2020 pada Laporan Bulan Agustus 2020 Badan Pusat Statistik (BPS).

Dengan melemahnya sektor perekonomian Indonesia, sektor ketenagakerjaan pun terkena dampak yang cukup berarti, menurut data Badan Pusat Statistik (BPS) penduduk usia kerja yang terimbas oleh pandemi COVID-19 berkisar sejumlah 29,12 juta. Oleh sebab itu, untuk meminimalisir dampak yang ditimbulkan oleh pandemi COVID-19 terhadap kesejahteraan masyarakat, pemerintah melalui Kementerian Sosial mengeluarkan Program Jaring Pengaman Sosial yang ditujukan bagi keluarga kurang mampu dan rentan terdampak COVID-19 agar beban mereka dapat diringankan selama pandemi.

Salah satu program dari Jaring Pengaman Sosial adalah bantuan sosial, bantuan sosial merupakan pengeluaran yang berbentuk transfer uang, barang atau jasa yang dikeluarkan oleh Pemerintah kepada masyarakat miskin atau tidak mampu guna melindungi masyarakat dari kemungkinan terjadinya resiko sosial, meningkatkan kemampuan ekonomi dan/atau kesejahteraan masyarakat menurut Peraturan Menteri Keuangan Nomor 254/PMK.05/2015 tentang Belanja Bantuan Sosial pada Kementerian Negara/Lembaga. Ada beberapa bentuk bantuan sosial yang sudah diberikan oleh pemerintah selama pandemi COVID-19 yang bertujuan untuk membantu meringankan beban masyarakat yang terdampak beberapa diantaranya adalah bansos sembako dan Bantuan Sosial Tunia (BST).

Dengan berjalannya program bantuan sosial oleh pemerintah ini, persoalan dan polemik mulai bermunculan, seperti data penerima bansos yang tidak valid serta tidak terintegrasi antara data pemerintah pusat, daerah dan juga kementerian sehingga menyebabkan ada bantuan yang tidak tepat sasaran, jenis bantuan yang beragam sehingga membingungkan masyarakat, anggaran bansos yang tidak mencukupi, isi paket bansos yang tidak sesuai hingga korupsi dana bansos. Dan dengan munculnya persoalan serta polemik tersebut

mulai banyak masyarakat yang membicarakan hal tersebut di berbagai macam platform, salah satu *platform* yang cukup banyak digunakan oleh masyarakat Indonesia untuk membicarakan tentang polemik bansos di masa pandemi COVID-19 ini adalah Twitter, Twitter adalah sebuah media sosial yang sangat cocok untuk membahas suatu topik karena berbentuk microblogging dan memiliki fitur *Trending* yang dapat membuat suatu topik menjadi populer sehingga menghasilkan berbagai macam opini tentang topik tersebut karena banyak dilihat dan dibahas oleh pengguna lainnya.

Oleh karena itu, tujuan dilakukannya penelitian ini adalah sebagai *Social Sensor* yang berfungsi untuk mengetahui sentimen yang beredar di tengah masyarakat tentang tweet yang berkaitan dengan bantuan sosial di era pandemi COVID-19 berdasarkan kategori rentang umur yang dikelompokkan menjadi 3 grup yaitu grup 1 yang memiliki rentang umur 20 tahun – 30 tahun, grup 2 yang memiliki rentang umur 31 tahun – 40 tahun dan grup 3 yang memiliki rentang umur 41 tahun – 50 tahun. Penelitian ini dilakukan dengan mengimplementasikan salah satu teknik analisis dari *Natural Language Processing* yaitu analisis sentimen. Analisis sentimen adalah sebuah teknik untuk mengekstrak pendapat seseorang mengenai suatu isu atau topik menjadi data yang lebih terstruktur dan menentukan opini orang tersebut menjadi 3 polaritas yaitu positif dan negatif

Metode utama yang akan digunakan untuk analisis sentimen ini adalah *Support Vector Machine*, serta digunakan 2 metode pendukung yaitu *P-Value* dan *Cross Validation*. *Support Vector Machine* adalah sebuah algoritma *Machine Learning* yang dapat digunakan untuk melakukan klasifikasi dan regresi. Pada penelitian ini algoritma *Support Vector Machine* akan dilakukan pada setiap grup sebagai representasi dari grup tersebut yang memiliki perbedaan umur, sehingga dapat diambil informasi klasifikasi sentimen yang beredar di masyarakat berdasarkan umur.

2. TINJAUAN PUSTAKA

2.1. Kajian Pustaka

Penelitian yang dikaji pertama kali merupakan penelitian yang dilakukan oleh Petrix Nomleni, yang bertujuan untuk melakukan klasifikasi informasi yang masuk pada media

center dari masyarakat yang bertujuan untuk proses pembangunan dan pelayanan yang sedang dilakukan oleh Pemerintah Kota Surabaya, *Support Vector Machine* merupakan metode yang digunakan pada penelitian ini. Penelitian kedua yang dikaji dilakukan oleh Hidayatullah Ahmad Fathan & Azhari, penelitian ini dilakukan untuk menganalisis tokoh publik yang memiliki popularitas tertinggi menjelang pemilihan umum 2014 berdasarkan beberapa lembaga survei melalui Twitter dengan menggunakan *Naive Bayes Classifier* yang digabungkan dengan TF-IDF.

Penelitian ketiga yang dikaji dilakukan oleh Rofiqoh, Perdana dan Fauzi bertujuan untuk mengetahui seperti apa tingkat kepuasan pelanggan pada suatu perusahaan yang bergerak dibidang penyedia layanan telekomunikasi seluler, analisis yang dilakukan adalah analisis sentimen pengguna Twitter untuk mengelompokkan polaritas dari data tersebut apakah termasuk positif atau negatif, pada penelitian ini metode yang digunakan adalah *Support Vector Machine* yang digabungkan dengan fitur *Lexicon Based Features*.

2.2. Analisis Sentimen

Analisis sentimen adalah suatu wujud pengimplementasian bidang *Natural Language Processing* (NLP) yang merupakan bagian dari linguistik, *Computer Science* dan *Artificial Intelligence* (AI) yang membahas tentang interaksi komputer dengan bahasa manusia. Analisis sentimen disebut juga *Opinion Mining* karena analisis sentimen adalah proses memahami, menganalisa dan mengelompokkan suatu opini yang berbentuk teks yang terdapat dalam sebuah tulisan dan sebagainya untuk mendapatkan nada emosionalnya (Positif atau negatif) (Shima, F., M., A. F. & Sigit, A., 2018).

2.3. Bantuan Sosial

Menurut Peraturan Menteri Keuangan Nomor 254/PMK.05/2015 tentang Belanja Bantuan Sosial pada Kementerian Negara/Lembaga adalah pengeluaran berupa uang, barang atau jasa yang diberikan oleh Pemerintah untuk masyarakat miskin atau masyarakat yang kurang mampu dengan tujuan untuk melindungi masyarakat dari kemungkinan terjadinya resiko sosial, meningkatkan perekonomian masyarakat dan/atau meningkatkan kesejahteraan masyarakat.

2.4. Machine Learning

Machine learning adalah sebuah aplikasi komputer yang menggunakan teknik pendekatan matematis untuk melakukan inferensi atau penyimpulan terhadap suatu data. esensi dari machine learning adalah untuk menggambarkan pola-pola data menggunakan model matematis. *Machine learning* dapat digunakan untuk banyak tujuan, beberapa diantaranya adalah memprediksi atau memperkirakan peristiwa yang tidak teramati (*unobserved event*) dan menemukan insight atau wawasan (*knowledge discovery*) (Ahmad, R., Po, A. S. & Ageng, S. R., 2020).

2.5 Support Vector Machine

Support Vector Machine (SVM) adalah sebuah teknik *Supervised Machine learning* yang berfungsi untuk menganalisa suatu data untuk klasifikasi. *Support Vector Machine* berjalan dengan cara mengkategorikan titik data yang berasal dari kumpulan data latih. Tujuan dari algoritma *Support Vector Machine* adalah untuk mencari *hyperplane* pada ruang dimensi N (jumlah fitur) yang mengklasifikasikan secara jelas titik data (Awad, M. & Khanna, R., 2015).

2.6. Text Mining

Text Mining atau Penambangan teks, juga dikenal sebagai penambangan teks atau penemuan pengetahuan dari database tekstual, mengacu pada proses pencarian pola atau pengetahuan yang penting dan bermakna dari dokumen teks, *Text Mining* adalah bidang multidisiplin yang menggabungkan penggalian informasi, analisa teks, pengelompokan, ekstraksi informasi, pengkategorian, visualisasi data, teknologi basis data, penambangan data dan machine learning (Tan, 2000).

2.7. Python

Python adalah sebuah *open source high level programming language* yang dikembangkan pada akhir tahun 1980 dan rilis pada tahun 1991 oleh Guido van Rossum. Python memiliki kemampuan untuk berjalan di hampir semua jenis arsitektur sistem sehingga Python sangat sering ditemukan di berbagai macam aplikasi, ekosistem dari Python juga sudah sangat matang karena memiliki ribuan modul pihak ketiga yang sangat berguna tersedia di Python Package Index (PyPI) dan Conda seperti Mathplotlib, Django, NumPy, Pandas dan sebagainya sehingga sangat

memudahkan pengguna untuk menggunakan Python.

2.8. P-Value

P-Value adalah angka hasil perhitungan uji statistik yang menggambarkan seberapa besar kemungkinan telah ditemukannya serangkaian pengamatan tertentu jika hipotesis nol benar. *P-Value* digunakan dalam pengujian hipotesis untuk membantu memutuskan apakah akan menolak hipotesis nol. hipotesis nol umumnya mendefinisikan bahwa tidak ada hubungan diantara variabel yang diamati atau tidak ada perbedaan di antara kelompok (Sander, G., 2016).

2.9. Pre-processing

Text Pre-processing adalah sebuah proses yang dilakukan untuk mentransformasikan sebuah teks mentah yang tidak terstruktur dan acak menjadi sebuah bentuk yang lebih terstruktur sehingga dapat di analisa, karena dengan bentuk yang lebih terstruktur algoritma *Machine Learning* dapat memproses informasi tersebut lebih optimal. *Text Pre-processing* memiliki 5 tahapan yaitu :

1. *Case Folding*
Case Folding adalah tahapan yang dilakukan dengan tujuan untuk mengubah teks menjadi bentuk yang standar.
2. *Tokenize*
Tokenize adalah proses memecah sebuah kalimat, paragraf atau dokumen menjadi bentuk yang lebih kecil yaitu karakter, kata dan sub kata, bentuk yang lebih kecil tersebut disebut token.
3. *Filtering*
Filtering atau *Stopword Removal* adalah proses menghilangkan kata yang memiliki sedikit informasi sehingga kurang penting untuk diproses dari sebuah teks seperti contoh “yang”, “dan”, “atau”, “akan” dan sebagainya.
4. *Stemming*
Stemming adalah sebuah proses yang bertujuan untuk menghilangkan infleksi pada sebuah kata, infleksi adalah penambahan imbuhan pada suatu kata sehingga terbentuk kata yang baru sehingga kata yang akan dijadikan informasi

bukanlah kata imbuhan melainkan akar kata.

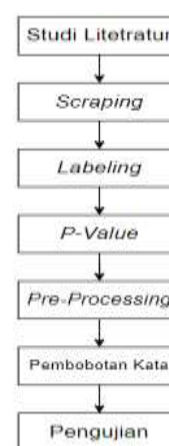
2.10. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF merupakan sebuah metode yang digunakan untuk melakukan perhitungan bobot sebuah kata, untuk memberikan bobot pada suatu kata cara yang dilakukan adalah dengan memberikan bobot hubungan suatu kata (term) terhadap suatu dokumen. Metode TF-IDF bekerja dengan cara mengkombinasikan dua buah konsep perhitungan bobot kata, yaitu *Term Frequency* dan *Inverse Document Frequency*. Frekuensi munculnya sebuah *term* pada suatu dokumen dapat menunjukkan seberapa berarti *term* tersebut di dalam sebuah dokumen (Musfiroh, et al., 2013).

2.11. Cross Validation

Cross Validation adalah metode pengambilan sampel secara ulang yang diaplikasikan untuk mengevaluasi dan membandingkan model sebuah *machine learning* yang memiliki sampel data yang sedikit atau terbatas. *Cross Validation* memiliki sebuah parameter tunggal yang disebut *K* yang menunjukkan jumlah grup yang akan dibagi menjadi beberapa sampel data tertentu (Huzain, A., Purnawansyah, Farniwati, F. & Inggrianti, P. P., 2020).

3. METODOLOGI



Gambar 1. Tahapan Penelitian

Pada Gambar 1 dapat dilihat tahapan penelitian yang dilakukan, studi literatur adalah tahapan pertama pada penelitian ini, studi literatur merupakan sebuah kegiatan yang berkaitan dengan pengumpulan informasi atau referensi yang berkaitan dengan masalah

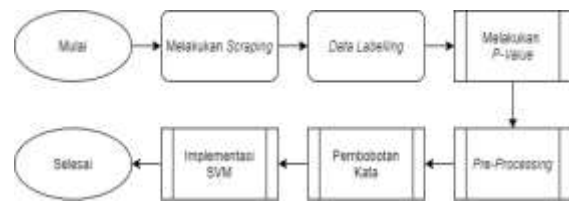
penelitian, informasi tersebut bisa didapatkan dari majalah, buku dan sebagainya. Studi literatur bertujuan untuk menambah wawasan, meningkatkan pemahaman serta mengumpulkan informasi yang berkaitan dengan penelitian. Selanjutnya adalah *scraping* atau pengumpulan data adalah sebuah aktivitas mengumpulkan data untuk tujuan penelitian, data yang akan dikumpulkan pada penelitian ini adalah kumpulan *tweet* yang berkaitan dengan topik bantuan sosial selama pandemi COVID-19 dan hasil dari kuisioner yang sudah dibagikan.

Tahap ketiga pada penelitian ini adalah *labelling*, *labelling* adalah proses memberikan label kepada *tweet* yang sudah dikumpulkan. Label yang akan diberikan berupa positif dan negatif. *Labelling* dilakukan dengan cara dilakukan dengan cara membuat sebuah kuisioner yang berisi 50 *tweet* yang sudah dikumpulkan dan diberikan kepada 3 grup responden yang sudah dikelompokkan berdasarkan rentang usia, yaitu grup 1 yang memiliki rentang usia antara 20 - 30 tahun, grup 2 yang memiliki rentang usia antara 31 - 40 tahun dan grup 3 yang memiliki rentang usia antara 41 - 50 tahun.

Selanjutnya adalah *p-value*, *p-value* adalah angka hasil perhitungan uji statistik yang menggambarkan seberapa besar kemungkinan telah ditemukannya serangkaian pengamatan tertentu jika hipotesis nol benar. Selanjutnya adalah *pre-processing*, *pre-processing* adalah sebuah proses yang dilakukan untuk mengubah sebuah teks mentah yang tidak memiliki struktur menjadi sebuah bentuk yang lebih terstruktur sehingga dapat lebih mudah di analisa dan mendapatkan hasil yang lebih optimal.

Setelah tahap *Pre-Processing* dilakukan pada data yang telah diambil, langkah selanjutnya yang diambil adalah pembobotan kata. Pembobotan kata adalah tahapan pada penelitian ini yang bertujuan untuk mengetahui seberapa besar sebuah *terms* mempengaruhi sebuah *tweet*. Metode pembobotan kata yang akan digunakan pada penelitian ini adalah metode *Term-Frequency – Inverse Document Frequency* (TF-IDF). Tahapan terakhir adalah pengujian, Tahapan pengujian adalah tahapan untuk menguji algoritma yang sudah dibuat, tahap pengujian dilakukan dengan cara mengujikan data uji setelah melakukan pelatihan pada data latih. terdapat 2 pengujian pada tahap ini yaitu pengujian menggunakan *cross validation* dan pengujian tanpa *cross validation*.

4. PERANCANGAN



Gambar 2. Diagram Alir Proses Analisis Sentimen

Pada Gambar 2 dapat dilihat tahapan untuk melakukan analisis sentimen, Langkah pertama pada proses ini yang akan dilakukan adalah melakukan *scraping tweet*, yaitu proses mengumpulkan data yang berbentuk *tweet* sesuai dengan parameter yang sudah ditentukan seperti kata kunci, tanggal *tweet* diunggah dan bahasa. Langkah selanjutnya adalah melakukan labeling pada *tweet* yang sudah dikumpulkan, labeling dilakukan dengan cara membuat sebuah kuisioner yang berisi 50 *tweet* yang sudah dikumpulkan dan diberikan kepada 3 grup responden yang sudah dikelompokkan berdasarkan rentang usia, yaitu grup 1 yang berusia diantara 20 sampai dengan 30 tahun, grup 2 yang berusia diantara 31 sampai dengan 40 tahun dan grup 3 yang berusia diantara 41 sampai 50 tahun. proses selanjutnya yaitu *P-Value*.

Tahap *P-Value* dilakukan untuk mengetahui apakah grup yang sudah ditentukan berkorelasi atau berasal dari populasi yang sama, sehingga dapat ditentukan apakah SVM akan dilakukan pada hanya pada salah satu grup atau pada setiap grup Tahap selanjutnya adalah *Pre-Processing*, pada tahap *Pre-Processing* terdapat 5 tahapan yang akan dilakukan pada data tersebut, tahapan tersebut adalah *case folding*, *tokenize*, *normalization*, *filtering* dan *stemming*. Tahapan ini bertujuan untuk mengubah data yang pada penelitian ini adalah *tweet* kedalam bentuk yang terstruktur. Contoh *P-Value* dapat dilihat pada Tabel 1.

Tabel 1. Contoh Pre-processing

Sebelum	Sesudah
@ekowboy2 Jika Pemerintah/Negara Benar-Benar Ingin Memberantas Korupsi Sampai Ke Akar-Akarnya, @KPK_RI Bersinergi Dengan @bpkri Telusuri	'ekowboy pemerintah negara benar benar berantas korupsi akar kpkri sinergi bpkri telusur kabupaten kota fakta lapangan ngeri indikasi korupsi paket bansos covid

Juga Sampai Ke
Bawah
Kabupaten/Kota,
Fakta Di Lapangan
Mungkin Lebih
Ngeri Lagi Indikasi
Korupsi Paket
Bansos Covid-19
Kemensos, Main
MarkUp/Jmlh Item.

kemensos main
menaikkan jumlah
barang'

Langkah selanjutnya adalah membobotkan seluruh kata yang terdapat pada setiap tweet menggunakan TF-IDF, Selanjutnya hasil dari TF-IDF akan digunakan untuk melakukan SVM, pada tahap ini data akan dipecah menjadi 2 yaitu data uji dan data latih, setelah itu langkah yang terakhir pada analisis sentimen ini adalah mengimplementasikan metode *Cross Validation*, metode ini digunakan untuk mengetahui bagaimana kemampuan model dari *machine learning*.

5. IMPLEMENTASI

Seluruh proses implementasi dilakukan pada IDE Jupyter dan menggunakan bahasa pemrograman Python. Langkah pertama yang dilakukan adalah *scraping* atau mengambil data, untuk melakukan *scraping*, *tools* yang digunakan adalah *library* Python bernama Twint, setelah itu seluruh *tweet* dimuat kedalam kuisioner dah disebar sebagai proses *labelling*. Setelah proses *labelling* selesai proses selanjutnya adalah *p-value*, proses ini dilakukan menggunakan *library* bernama *statsmodels*.

Setelah itu akan dilakukan *pre-processing* kepada semua *tweet* yang sudah dikumpulkan *library* yang dibutuhkan untuk melakukan *pre-processing* adalah *pandas*, *numpy*, *RE (Regular Expression)* dan *NLTK (Natural Language Toolkit)* dan untuk melakukan proses *stemming* dibutuhkan *library* bernama *Sastrawi*.

Tahap selanjutnya adalah pembobotan kata, *tools* yang dibutuhkan untuk melakukan proses ini adalah *library* bernama *sklearn* dan kelas didalamnya yang bernama *TfidfVectorizer*. Langkah terakhir yang dilakukan adalah pengujian, terdapat 2 langkah pada tahap pengujian, Langkah pertama yang dilakukan adalah mengimpor fungsi 'train_test_split' pada *library* *scikit-learn* yang berfungsi untuk membagi test dan train data, lalu mengimpor fungsi 'SVC' dari *library* *scikit-learn*, setelah itu

membuat sebuah variabel yang berfungsi untuk menginstansiasi kelas 'SVC' dengan menggunakan parameter yang ditentukan.

Lalu langkah yang dilakukan untuk pengujian menggunakan *Cross Validation* adalah mengimpor kelas 'cross_val_score' pada *library* *scikit-learn*, setelah itu membuat sebuah variabel dengan nama 'cv_result' yang berfungsi untuk menginstansiasi kelas 'cross_val_score' dengan menggunakan parameter yang ditentukan.

6. PENGUJIAN

6.1. Pengujian P-Value

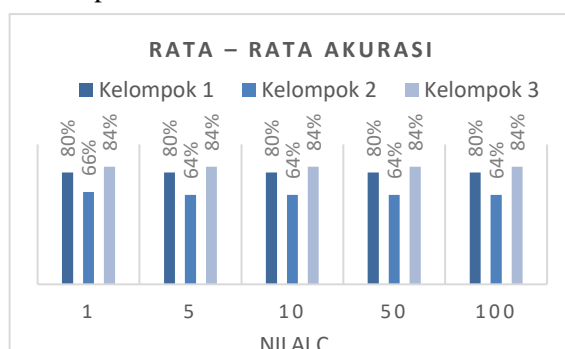
Pengujian yang pertama dilakukan adalah *p-value*, pengujian ini dilakukan sebelum mengimplementasikan SVM, hasil dari pengujian *p-value* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Perhitungan P-value

Pengujian	Hasil
Kelompok1 – Kelompok 2	0.000048
Kelompok 1 – Kelompok 3	0.004812
Kelompok 2 – Kelompok 3	0.087135

6.2. Pengujian Nilai Parameter C

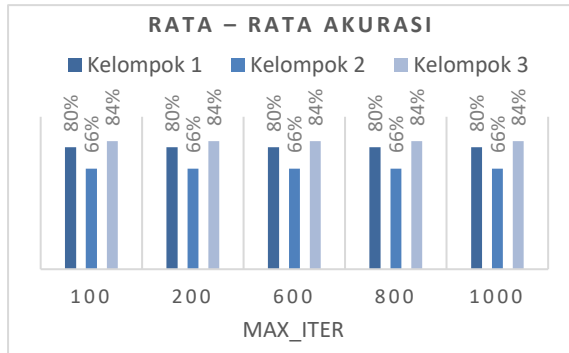
Pada pengujian nilai parameter C, nilai yang akan diuji adalah 1, 5, 10, 50 dan 100 serta nilai iterasi maksimum yang akan digunakan adalah 1000, hasil pengujian dari tiap kelompok dapat dilihat pada Gambar 3.



Gambar 3. Grafik Rata-Rata Akurasi Parameter C

6.3. Pengujian Nilai Iterasi Maksimum

Pada pengujian nilai iterasi maksimum, nilai iterasi yang akan digunakan adalah 100, 200, 600, 800 dan 1000 serta nilai C yang akan digunakan adalah 1, hasil pengujian dari tiap kelompok dapat dilihat pada Gambar 4.



Gambar 4. Grafik Rata-Rata Akurasi Iterasi Maksimum

6.4. Evaluasi Performa

Pada pengujian evaluasi performa, parameter yang akan digunakan berupa parameter terbaik yang sudah didapatkan dari pengujian sebelumnya, dan menggunakan Nilai K = 5 untuk pengujian menggunakan Cross Validation, hasil pengujian dari tiap kelompok dapat dilihat pada Tabel berikut.

Tabel 3. Evaluasi Perfroma Cross Validation Kelompok 1

	Percobaan Ke-					Rata-Rata
	1	2	3	4	5	
Akurasi	88%	88%	88%	75%	75%	83%
Precision	44%	44%	44%	38%	38%	41%
Recall	50%	50%	50%	50%	50%	50%
F-Measure	47%	47%	47%	43%	43%	45%

Tabel 4. Evaluasi Perfroma Cross Validation Kelompok 2

	Percobaan Ke-					Rata-Rata
	1	2	3	4	5	
Akurasi	75%	75%	75%	63%	63%	70%
Precision	38%	38%	38%	31%	31%	35%
Recall	50%	50%	50%	50%	50%	50%
F-Measure	43%	43%	43%	39%	39%	41%

Tabel 5. Evaluasi Perfroma Cross Validation Kelompok 3

	Percobaan Ke-					Rata-Rata
	1	2	3	4	5	
Akurasi	88%	88%	88%	88%	75%	85%
Precision	44%	44%	44%	44%	38%	43%
Recall	50%	50%	50%	50%	50%	50%
F-Measure	47%	47%	47%	47%	43%	46%

Untuk hasil evaluasi performa tiap kelompok dapat dilihat pada Tabel berikut.

Tabel 6. Evaluasi Performa Kelompok 1

Nilai Asli	Nilai Prediksi	
	Positif	Negatif
Positif	7	3
Negatif	0	0

Tabel 7. Confusion Matrix Kelompok 1

Akurasi	Precision	Recall	F-Measure
70%	70%	100%	82%

Tabel 8. Evaluasi Performa Kelompok 2

Nilai Asli	Nilai Prediksi	
	Positif	Negatif
Positif	6	4
Negatif	0	0

Tabel 9. Confusion Matrix Kelompok 2

Akurasi	Precision	Recall	F-Measure
60%	60%	100%	75%

Tabel 10. Evaluasi Performa Kelompok 3

Nilai Asli	Nilai Prediksi	
	Positif	Negatif
Positif	8	2
Negatif	0	0

Tabel 11. Confusion Matrix Kelompok 3

Akurasi	Precision	Recall	F-Measure
---------	-----------	--------	-----------

80%	80%	100%	88%
-----	-----	------	-----

7. KESIMPULAN DAN SARAN

7.1. Kesimpulan

1. Untuk mengimplementasikan metode *Support Vector Machine* pada analisis sentimen tentang bantuan sosial pada masa pandemi COVID-19 terdapat beberapa tahapan yang harus dilakukan, yaitu *Scraping, Labelling, Probability Test, Pre-Processing*, pembobotan kata dan evaluasi performa.
2. Untuk mendapatkan parameter terbaik dilakukan pengujian parameter C dan iterasi maksimum. Pada pengujian Nilai parameter C, kelompok 1 mendapatkan hasil yang sama pada setiap nilai parameter yang diujikan yaitu 80%, kelompok 2 mendapatkan hasil terbaik pada nilai parameter C 1 yaitu 66% dan kelompok mendapatkan hasil yang sama pada setiap parameter yang diujikan yaitu 84%. Pada pengujian iterasi maksimum, kelompok 1 mendapat hasil yang sama pada setiap nilai parameter yang diujikan yaitu 80%, kelompok 2 mendapatkan hasil yang sama pada setiap parameter yang diujikan yaitu 66% dan kelompok 3 mendapatkan hasil yang sama pada setiap parameter yang diujikan yaitu 84%.
3. Pada pengujian evaluasi performa menggunakan parameter terbaik dari masing – masing kelompok, kelompok 1 mendapatkan nilai akurasi sebesar 70%, *precision* sebesar 70%, *recall* sebesar 100% dan *f-measure* sebesar 82.4%, kelompok 2 mendapatkan nilai akurasi sebesar 60%, *precision* sebesar 60%, *recall* sebesar 100% dan *f-measure* sebesar 74.9% dan kelompok 3 mendapatkan nilai akurasi sebesar 80%, *Precision* sebesar 80%, *recall* sebesar 100% dan *f-measure* sebesar 88%.
4. Dari hasil pengujian, SVM yang dilatih menggunakan data dari kelompok 3 atau kelompok umur 41 – 50 tahun menghasilkan performa terbaik diantara kelompok lainnya, dan dapat dikatakan bahwa kelompok 3 merupakan kelompok yang paling konsisten, sehingga kelompok 3 merupakan kelompok yang *reliable* untuk dijadikan responden untuk penelitian analisis sentimen.

7.2. Saran

Pada subbab ini terdapat saran yang penulis dapatkan dari kekurangan yang ada pada penelitian yang telah dilakukan, sehingga dapat dipertimbangkan pada penelitian yang akan dilakukan kedepannya yang pertama yaitu menambahkan jumlah dataset agar hasil dapat lebih optimal, lalu menambahkan jumlah kelompok untuk mengetahui sentimen dari kelompok umur lainnya dan yang terakhir menambahkan fitur seperti *emoji weighting*. Dan dari hasil evaluasi performa, bahwa kelompok 3 sangat baik untuk dijadikan target responden karena memiliki data yang konsisten.

8. DAFTAR PUSTAKA

- Akbar, T. M., Kurniati, A. P. & Bijaksana, M. A., 2012. ANALISIS PERBANDINGAN METODE PEMBOBOTAN KATA TF.IDF DAN TF.RF TERHADAP PERFORMANSI KATEGORISASI TEKS. s.l.:Telkom University.
- Awad, M. & Khanna, R., 2015. Support Vector Machine for Classification, s.l.: ResearchGate.
- Buntoro, G. A., 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal*, 2(1), pp. 32-41.
- Evgeniou, T. & Pontil, M., 2001. SUPPORT VECTOR MACHINES: THEORY AND APPLICATIONS. Cambridge, ResearchGate.
- Hidayatullah, A. F. & Azhari, S., 2014. ANALISIS SENTIMEN DAN KLASIFIKASI KATEGORI TERHADAP TOKOH PUBLIK PADA TWITTER. Seminar Nasional Informatika 2014, pp. 115-122.
- Iping, B., 2020. PERLINDUNGAN SOSIAL MELALUI KEBIJAKAN PROGRAM BANTUAN LANGSUNG TUNAI (BLT) DI ERA PANDEMI COVID-19: TINJAUAN PERSPEKTIF EKONOMI DAN SOSIAL. 2 July.
- Muhammad, J. & Shahid, K., 2018. Normalization of Unstructured and Informal Text in. *International Journal of Advanced Computer Science and Applications*, 9(10).
- Murfi, D. r. n. H., n.d. Support Vector Machine, Depok: Intelligent Data Analysis (IDA) Group.

- Musfiroh, N., Hamdani & Inda, F. A., 2013. PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING. JURNAL INFORMATIKA MULAWARMAN, 8(3), pp. 110-113.
- Nomleni, P., 2015. SENTIMENT ANALYSIS MENGGUNAKAN SUPPORT VECTOR MACHINE(SVM), Surabaya: Institut Teknologi Sepuluh Nopember.
- Nugroho, A. S., Witarto, A. B. & Handoko, D., 2003. Support Vector Machine - Teori dan Aplikasinya dalam Bioinformatika, s.l.: IlmuKomputer.Com.
- Rofiqoh, U., Perdana, R. S. & Fauzi, M. A., 2017. Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. Pengembangan Teknologi Informasi dan Ilmu Komputer, Volume 1, pp. 1725-1732.
- Santoso, V. I., Virginia, G. & Lukito, Y., 2017. PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE. JURNAL TRANSFORMATIKA, 20 January.
- Siregar, A. M., 2017. PERBANDINGAN PEMBOBOTAN KATA DALAM SISTEM TEMU BALIK INFORMASI. Jurnal Teknologi & Ilmu Komputer, 2(2).
- Tan, A.-H., 2000. Text Mining: The State Of The Art And The Challenges, Singapore: Kent Ridge Digital Labs.
- Sander, G., 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol, Band 31, pp. 337-350.
- Huzain, A., Purnawansyah, Farniwati, F. & Inggrianti, P. P., 2020. Performa Klasifikasi K-NN dan Cross-validation pada Data Pasien Pengidap Penyakit Jantung. ILKOM Jurnal Ilmiah, 12(2), pp. 81-86.
- Shima, F., M., A. F. & Sigit, A., 2018. Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2(8), pp. 2766-2770.