

Analisis Sentimen Program Jaminan Kesehatan Nasional Menggunakan Multiclass Support Vector Machine

Ni Gusti Ayu Dasriani^{1*}, Lalu Ahmad Gede Pariandi², I Made Yadi Dharma³

¹Ilmu Komputer, Universitas Bumigora, ayu.areyu@universitasbumigora.ac.id

²Ilmu Komputer, Universitas Bumigora, laluyandi80@gmail.com

³Ilmu Komputer, Universitas Bumigora, yadi_dharma@universitasbumigora.ac.id

Keywords :

Sentiment Analysis,
JKN Program,
Classification,
Multiclass SVM,
Confusion Matrix,

ABSTRACT

Optimizing the implementation on National Health Insurance which requires the use of BPJS participant cards in various public services is one of the government policies that is widely discussed and has garnered many opinions in the community. Public opinion is expressed through social media, one of which is through Twitter. The aim of this research is to classify public opinion regarding the new regulations of the National Health Insurance Program as a form of government policy to implement Presidential Instruction Number 1 of 2022 using Twitter data. Public opinion as many as 1.179 tweets were labeled positive, negative and neutral sentiments, then TD-IDF weighting was carried out and analyzed using the multiclass SVM algorithm with the One Against All approach. The results of the analysis showed that Multiclass SVM with a linear kernel was able to classify with an accuracy level of 81% where the classification of positive sentiment was 17 (7.6%), negative sentiment was 115 (48.7%) and neutral sentiment are 104 (44.1%). This shows that public sentiment is dominated by negative sentiment or disagreement with the new regulations of the National Health Insurance Program.

Kata Kunci :

Analisis Sentimen,
Program JKN,
Klasifikasi,
Multiclass SVM,
Confusion Matrix,

ABSTRAK

Optimalisasi pelaksanaan Jaminan Kesehatan Nasional yang mewajibkan penggunaan kartu peserta BPJS dalam berbagai pelayanan publik merupakan salah satu kebijakan pemerintah yang banyak diperbincangkan dan menuai banyak opini di masyarakat. Opini masyarakat tersebut dituangkan melalui media sosial, salah satunya melalui *Twitter*. Tujuan penelitian ini adalah untuk mengklasifikasikan opini masyarakat mengenai aturan baru Program Jaminan Kesehatan Nasional sebagai bentuk kebijakan pemerintah dalam rangka menjalankan Instruksi Presiden Nomor 1 tahun 2022 dengan menggunakan data *Twitter*. Opini masyarakat sebanyak 1.179 tweet diberi label sentimen positif, negatif dan netral, kemudian dilakukan pembobotan TF-IDF dan dianalisis menggunakan algoritma *Multiclass SVM* dengan pendekatan *One Against All*. Hasil analisis menunjukkan bahwa *Multiclass SVM* dengan kernel linear mampu melakukan klasifikasi dengan tingkat akurasi sebesar 81% dimana klasifikasi sentimen positif sebanyak 17 (7.6%), sentimen negatif sebanyak 115 (48.7%) dan sentimen netral sebanyak 104 (44.1%). Hal ini menunjukkan bahwa sentimen masyarakat didominasi oleh sentimen negatif atau ketidaksetujuan terhadap aturan baru Program Jaminan Kesehatan Nasional.

Korespondensi Penulis:

Ni Gusti Ayu Dasriani,
Universitas Bumigora, Jl. Ismail Marzuki No. 22 Mataram
Telepon : +6287753939344
Email: ayu.areyu@universitasbumigora.ac.id

Submitted : 12-08-2024; Accepted : 05-11-2024;

Published : 17-11-2024

Copyright (c) 2024 The Author (s) This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

1. PENDAHULUAN

Media sosial merupakan media alternatif untuk berkomunikasi dan berekspresi, bahkan saat ini media sosial telah menjadi alat utama dalam kebebasan berekspresi [1]. Salah satu media sosial yang banyak digunakan adalah *Twitter*, dimana pengguna dapat dengan mudah menyampaikan kritik, saran, opini dan komentar terhadap suatu produk, suatu peristiwa atau kebijakan baru pemerintah sehingga secara tidak langsung akan memberikan pengaruh pada reputasi suatu produk atau peristiwa yang terjadi. Bahkan informasi yang terkandung dalam *tweet* juga sangat berharga sebagai data pendukung dalam pengambilan kebijakan [2][3]. *Tweet* yang masih tersusun secara acak

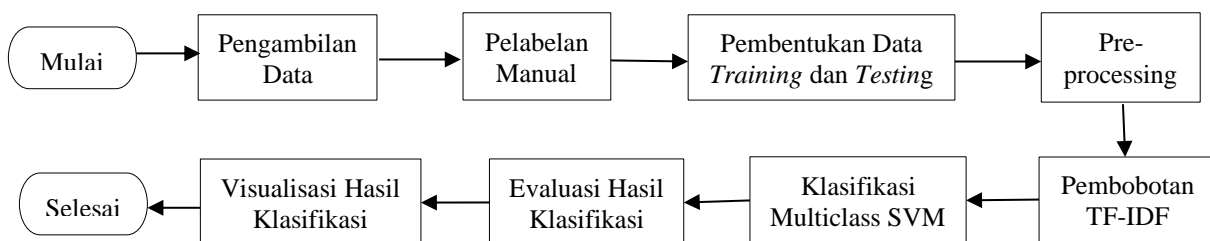
menyebabkan kesulitan bagi pihak-pihak tertentu dalam mengetahui opini dan komentar yang diberikan masyarakat [2]. Informasi berupa *tweet twitter* dapat dianalisis menggunakan analisis sentimen [4]. *Tweet* dievaluasi dan dianalisis untuk dapat diketahui apakah *tweet* mengarah pada opini positif atau negatif [5] dan netral [2]

Instruksi Presiden Republik Indonesia (Inpres) Nomor 1 Tahun 2022 mengatur tentang persyaratan administrasi berbagai pelayanan publik seperti pengurusan jual beli tanah, pengurusan Surat Izin Mengemudi (SIM), STNK, SKCK, ibadah haji dan Umrah, pengajuan Kredit Usaha Rakyat (KUR), dan pelayanan pendidikan formal dan non formal yang harus didaftarkan sebagai peserta BPJS Kesehatan serta status aktif [6][7]. Alasan dan pertimbangan mendasar dikeluarkannya Inpres Nomor 1 ini berkaitan dengan optimalisasi penyelenggaraan program Jaminan Kesehatan Nasional, peningkatan akses terhadap pelayanan kesehatan yang berkualitas dan menjamin keberlangsungan pelaksanaan program Jaminan Kesehatan Nasional [7]. Optimalisasi pelaksanaan Jaminan Kesehatan Nasional (JKN) ini merupakan salah satu kebijakan pemerintah yang sedang ramai diperbincangkan dan menimbulkan banyak polemik di masyarakat [8], karena pengurusan suatu layanan publik hanya dapat dilakukan jika telah menjadi peserta aktif JKN [6]. Evaluasi dan analisis *tweet* publik terkait aturan baru yang mengatur penggunaan JKN sebagai syarat layanan publik sampai saat ini belum dilakukan, sehingga belum dapat diketahui apakah opini mengarah pada opini positif, negatif [5] atau netral [2] dan belum dapat digunakan sebagai data pendukung dalam pengambilan keputusan dari penerapan aturan baru tersebut.

Penelitian terdahulu oleh M. Al Khadafi, Kurnia Paranitha Kartika, and Filda Febrinita (2022), melakukan analisis sentimen pada twitter menggunakan metode *Naïve Bayes Classifier* dan *Lexicon Based* yang mengklasifikasikan *tweet* positif dan negatif mengenai persyaratan BPJS Kesehatan dalam kegiatan jual beli tanah ke dalam dua kelas yaitu *cyberbullying* atau *non cyberbullying* [8]. Penelitian lain oleh Ridho Darman (2023) yang melakukan analisis sentimen pada twitter menggunakan metode *Lexicon Based* untuk mengklasifikasikan *tweet* dalam tiga kelas yaitu positif, negatif dan netral mengenai kebijakan kartu peserta BPJS Kesehatan yang menjadi syarat dalam pelayanan pendaftaran peralihan hak atas tanah [10]. Kedua penelitian yang dilakukan sebelumnya [8][10] hanya membahas persyaratan kartu peserta BPJS Kesehatan untuk pelayanan dan pengurusan jual beli tanah. Penelitian ini berbeda dengan penelitian sebelumnya dimana penelitian ini bertujuan melakukan analisis sentimen untuk melihat tingkat akurasi atau kinerja *multiclass SVM* dalam melakukan klasifikasi terhadap data *tweets* terkait opini masyarakat pengguna twitter tentang aturan baru Program JKN yang mensyaratkan penggunaan kartu peserta BPJS dalam berbagai layanan publik sebagai bentuk kebijakan pemerintah dari implementasi Inpres Nomor 1 tahun 2022. Berdasarkan tinjauan dari penelitian sebelumnya menunjukkan bahwa penelitian ini memiliki kebaharuan yang dapat dilihat dari metode klasifikasi yang digunakan yaitu *multiclass SVM* sedangkan penelitian terdahulu menggunakan metode *naive bayes classifier* dan *lexicon based* terhadap aturan baru program JKN.

2. METODE PENELITIAN

Sistem klasifikasi yang dibangun pada penelitian ini dilakukan dalam beberapa tahapan prosedur sehingga proses penelitian dapat berjalan secara sistematis sebagaimana terlihat pada Gambar 1.



Gambar 1. Diagram Alir Proses Analisis Sentimen

Secara garis besar sistem yang dibangun adalah sistem yang dapat mengklasifikasikan sentimen masyarakat terhadap aturan baru program Jaminan Kesehatan Nasional (JKN) yang mewajibkan keanggotaan aktif sebagai syarat pengurusan berbagai jenis layanan publik. Analisis dilakukan untuk mengklasifikasikan *tweet* yang mengandung sentimen positif, negatif atau netral.

2.1 Pengambilan Data

Data yang digunakan dalam penelitian ini didapatkan dari *twitter* dengan menggunakan *add-ons* dari *google spreadsheet* yaitu *tweet archiver*. Data yang diambil adalah *tweets* berbahasa Indonesia tentang kebijakan pemerintah dalam upaya optimalisasi pelaksanaan JKN dengan kata kunci “bpjs jadi syarat”, “JKN”, “Inpres No 1 tahun 2022” dan “Jaminan Kesehatan Nasional”. Data awal yang dapat diambil dari *twitter* berjumlah 1.664 data *tweets*. Selanjutnya data yang telah didapat akan diproses ke tahap verifikasi kualitas data yang bertujuan untuk menghindari *tweet* ganda dan *tweet* yang tidak berhubungan dengan fokus penelitian yaitu dengan melakukan penghapusan data ganda dan data yang tidak relevan. Sehingga setelah dilakukan proses verifikasi kualitas data jumlah data menjadi 1.179 data *tweets*.

2.2 Pelabelan Sentimen

Proses pelabelan dilakukan karena SVM termasuk *machine learning (supervised learning)* yang membutuhkan data latih untuk membuat dan mendapatkan suatu pola atau model terbaik [11]. Pada penelitian ini pelabelan data dilakukan dengan memberikan label pada setiap *tweets* pengguna *twitter* secara manual oleh seorang pakar bahasa untuk menentukan *tweet* termasuk kedalam sentimen positif, negatif atau netral.

2.3 Pembentukan Data Training dan Data Testing

Proses *training* dijalankan untuk mempelajari data teks sehingga ketika menerima teks baru, sistem dapat memprediksi teks baru berdasarkan teks yang dipelajari. Proses ini membagi dataset menjadi data *training* dan data *testing*. Data *testing* digunakan untuk melakukan pengujian prediktif pada sistem dan memastikan keakuratan model [12]. Pembentukan data *training* dan data *testing* dilakukan menggunakan *google colab* dengan memanfaatkan *library train test split* dan *random_split* yang masing-masing berfungsi memecah data menjadi dua bagian dengan presentase yang telah ditentukan serta mengacak dan menjaga data agar tetap sama ketika sistem dijalankan kembali. Pada penelitian ini data *tweet* yang digunakan untuk tahapan klasifikasi sebanyak 1179 data dimana data yang digunakan sebagai data *training* sebesar 80% dari total data *tweet* yaitu 943 data dan data *testing* sebesar 20% dari total data *tweet* yaitu 236 data. Hasil data *training* dan data *testing* disimpan dalam format *Comma Separated Values (CSV)*.

2.4 Preprocessing

Algoritma pembelajaran mesin yang diawasi atau tidak diawasi memerlukan *preprocessing* yang harus dilakukan pada data sebelum dapat dianalisis [12]. Tahap *preprocessing* merupakan tahapan yang fokus pada pembersihan data dan bertujuan untuk memastikan bahwa data telah siap untuk digunakan termasuk mengatasi informasi yang hilang atau tidak lengkap serta memastikan tidak terdapat *noise* yang dapat mengganggu proses klasifikasi [11][9].

Tahapan *preprocessing* pada penelitian ini terdiri atas beberapa proses yang dimulai dari: *data cleaning, case folding, tokenizing, stemming, stopword removal* [3]. *Data cleaning* bertujuan untuk membersihkan data dari simbol-simbol yang terdapat pada sebuah *tweet* sehingga dapat mengurangi *noise* dan mempermudah proses klasifikasi [13]. *Case folding* yaitu mengubah semua kata yang terdapat pada data *tweet* menjadi huruf kecil (*lowercase*) yang bertujuan untuk menghindari permasalahan *font* tidak beraturan antara *uppercase* dan *lowercase* [13]. *Tokenizing* dilakukan dengan cara menguraikan sebuah kalimat berdasarkan tiap kata yang menyusun kalimat tersebut [8]. *Stemming* dilakukan dengan mengubah setiap kata yang berimbuhan menjadi kata dasar [14] menggunakan *library* sastrawi. *Stopword removal* yaitu proses penyaringan untuk menghilangkan kata yang tidak penting dan tidak diperlukan dalam proses klasifikasi sentimen [11][13] dengan menggunakan *library Natural Language Toolkit (NLTK)*.

2.5 Pembobotan Term Frequency-Inverse Document Frequency (TF-IDF)

Salah satu faktor penting dalam klasifikasi kata adalah proses representasi kata. Hal ini dapat dilakukan dengan mengubah data teks menjadi vektor yang dapat diproses oleh mesin [9]. Dalam penelitian ini, digunakan TF-IDF untuk representasi kata. Pembobotan TF-IDF merupakan suatu proses pemberian bobot pada setiap kata yang bertujuan untuk mengetahui seberapa penting suatu kata mewakili sebuah kalimat berdasarkan frekwensi dari kata yang muncul dalam suatu dokumen [4].

$$W_{td} = tf_{td} \times idf_t \quad (1)$$

$$idf_t = \log\left(\frac{D}{df_t}\right) \quad (2)$$

Disini, W_{dt} = bobot dari kata t dalam setiap dokumen d , tf_{td} = banyaknya kemunculan kata t yang dicari dalam dokumen d , idf_t = bobot invers df , t = kata t dari kata kunci, d = dokumen d , D = banyaknya semua dokumen, df_t = banyaknya dokumen yang mengandung kata t .

2.6 Multiclass Support Vector Machine (SVM)

Metode *Support Vector Machine (SVM)* merupakan metode klasifikasi *supervised learning* yang menangani kasus klasifikasi biner [15]. SVM bekerja dengan suatu pemetaan non linier yang digunakan untuk mengubah data latih menjadi suatu pola sehingga dapat dicari *hyperplane* pemisah yang paling optimal [9][15] dan SVM menggunakan konsep kernel dengan dimensi tinggi [12]. Metode *Multiclass SVM* menangani klasifikasi non biner seperti klasifikasi positif, negatif dan netral dengan pendekatan *One Against All (OAA)* atau *One Against One (OAO)*. Pendekatan OAA menyelesaikan masalah *multiclass* dengan N *decision boundary* dari hasil pencarian *hyperplane* setiap kelas ke- i dengan kelas sisa lainnya. Sedangkan OAO menyelesaikan masalah dengan $N(N-1)/2$ *decision boundary* dari hasil pencarian *hyperplane* dari setiap kelas dengan setiap satu kelas lainnya [16][15].

Pada penelitian ini untuk mencari setiap *hyperplane* dari 3 (tiga) kelas yang digunakan yaitu kelas positif, negatif dan netral digunakan metode *multiclass SVM* dengan pendekatan OAA. Hal ini disebabkan karena pendekatan

OAA memiliki performa yang lebih baik dibandingkan dengan pendekatan OAO [16] dan lebih sesuai untuk jumlah label yang relatif sedikit [17]. Data latih dinyatakan dengan (x_i, y_i) dimana atribut set untuk data latih kelas ke- i adalah $x_i = \{x_1, x_2, \dots, x_{iq}\}$ sedangkan untuk label kelas dinyatakan dengan $y_i \in \{-1, 1\}$. Untuk suatu *hyperplane* pemisah yang optimal didefinisikan dengan persamaan (3).

$$w \cdot xi + b = 0 \tag{3}$$

Data xi terbagi kedalam dua kelas yaitu kelas negatif pada persamaan (4) dan kelas positif pada persamaan (5)

$$w \cdot xi + b = -1 \tag{4}$$

$$w \cdot xi + b = +1 \tag{5}$$

Notasi perhitungan SVM terdiri dari variable xi untuk data input, yi untuk data label, w untuk bobot vektor dan b untuk skalar bias. Variabel w dan b merupakan parameter yang dicari nilainya, jika label data $yi = -1$, maka *hyperplane* dapat dirumuskan dengan persamaan (6) dan jika $yi = +1$, maka *hyperplane* dapat dirumuskan dengan persamaan (7)

$$w \cdot xi + b \leq -1 \tag{6}$$

$$w \cdot xi + b \geq +1 \tag{7}$$

Untuk nilai xi dan yi , didapatkan dengan mengubah setiap *tweet* menjadi nilai *support vektor* = $\begin{bmatrix} x \\ y \end{bmatrix}$ dan nilai *support vector* tersebut disubstitusikan ke persamaan (8)

$$\phi \begin{bmatrix} x \\ y \end{bmatrix} = \left\{ \begin{array}{l} \sqrt{x_n^2 + y_n^2} > 2 \text{ maka } \begin{bmatrix} \sqrt{x_n^2 + y_n^2} - x + |x - y| \\ \sqrt{x_n^2 + y_n^2} - y + |x - y| \end{bmatrix} \\ \sqrt{x_n^2 + y_n^2} \leq 2 \text{ maka } \begin{bmatrix} x \\ y \end{bmatrix} \end{array} \right\} \tag{8}$$

Nilai x diperoleh dengan persamaan (9) dan nilai y diperoleh dengan menggunakan persamaan (10).

$$\sum_{i=1, j=1}^n x_i x_j^T, (i, j = 1, \dots, n) \tag{9}$$

$$\sum_{i=1, j=1}^n y_i y_j^T, (i, j = 1, \dots, n) \tag{10}$$

Hasil dari perhitungan nilai x dan y diberi nilai bias = 1 untuk mendapatkan jarak tegak lurus optimal. Proses pencarian nilai fungsi dengan persamaan (11) dan pencarian parameter ai berdasarkan kernel linier yang didefinisikan pada persamaan (12).

$$\sum_{i=1, j=1}^n a_i a_j^T S_j \tag{11}$$

$$\sum_{i=1, j=1}^n a_i a_j^T S_j = y_i \tag{12}$$

Proses eliminasi dengan metode gauss jordan berdasarkan nilai ai ($ai \geq 0$) yang diperoleh dan disubstitusikan ke persamaan (13).

$$\bar{W} = \sum_{i=1}^n ai Si \tag{13}$$

Mencari nilai w dan b yang disebut *hyperplane* dengan persamaan (14) yang digunakan dalam proses klasifikasi.

$$y = wx + b \tag{14}$$

Konsep pendekatan OAA ini dimisalkan ada tiga kelas, yaitu kelas 1, 2 dan 3. Jika kelas 1 yang diuji maka kelas 1 diberi label +1, sedangkan kelas lainnya yaitu kelas 2 dan 3 diberi label -1. Hal ini juga diberlakukan untuk kelas 2 dan kelas 3 yang diuji. Berdasarkan proses tersebut di dapat masing-masing *hyperlane* dari ketiga kelas. Kelas dari data uji ditentukan berdasarkan nilai *hyperlane* terbesar menggunakan persamaan (15).

$$\text{kelas } x = \arg \max l = 1..k \left((w^{(l)})^T \cdot \phi(x) + b^{(l)} \right) \tag{15}$$

2.7 Evaluasi

Pengujian dan evaluasi dilakukan untuk mengetahui kinerja model klasifikasi yang telah dibuat [18] sehingga dapat ditentukan pengembangan model selanjutnya [9]. Pada penelitian ini, metode evaluasi yang digunakan adalah *confusion matrix* seperti tabel 1, dimana cara kerjanya yaitu melakukan perbandingan hasil klasifikasi dari sistem dengan hasil klasifikasi yang sebenarnya [18][15]. Setiap baris pada tabel 1 merepresentasikan kelas yang sebenarnya dan setiap kolom merepresentasikan kelas dari hasil prediksi atau klasifikasi, yang mana TP adalah jumlah data positif yang terklasifikasi dengan benar, TN adalah jumlah data negatif yang terklasifikasi dengan benar, TNt adalah jumlah data netral yang terklasifikasi dengan benar, FP adalah jumlah data positif namun terklasifikasi salah oleh sistem, FN adalah jumlah data negatif namun terklasifikasi salah oleh sistem dan FNt adalah jumlah data netral namun terklasifikasi salah oleh sistem.

Tabel 1. *Confusion Matrix*

Aktual	Klasifikasi		
	Positif	Negatif	Netral
Positif	True Positif (TP)	False Negatif (FN)	False Neutral (FNt)
Negatif	False Positif (FP)	True Negatif (TN)	False Neutral (FNt)
Netral	False Positif (FP)	False Negatif (FN)	True Neutral (TNt)

Berdasarkan proses *confusion matrix*, maka akan didapatkan nilai dari *precision*, *recall*, *f1-score* untuk ketiga sentimen dan keseluruhan *accuracy* yang dihasilkan [13]. *Precision* merupakan rasio prediksi atau klasifikasi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif [13] yang ditulis dalam persamaan (16).

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{16}$$

Recall merupakan rasio prediksi atau klasifikasi benar positif dibandingkan dengan keseluruhan data yang benar positif [13] dalam bentuk persamaan (17)

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{17}$$

F1-Score merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan [13] dalam persamaan (18)

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall+Precision} \tag{18}$$

Accuracy merupakan rasio prediksi benar (positif, negatif dan netral) dengan keseluruhan data [15] dalam persamaan (19)

$$Accuracy = \frac{TP+TN+TNt}{TP+TN+TNt+FP+FN+FNt} \times 100\% \tag{19}$$

2.8 Visualisasi

Pada penelitian ini, ketiga sentimen hasil klasifikasi yaitu sentimen positif, negatif dan netral divisualisasikan dalam bentuk *wordcloud*, *bar chart* dan *pie chart* dengan menggunakan *library python* yaitu *matplotlib* dan *wordcloud* dengan tujuan agar mudah dibaca dan dipahami oleh masyarakat sehingga informasi yang disampaikan dapat tersampaikan dengan baik.

3. HASIL DAN ANALISIS

Dalam penelitian ini dilakukan analisis sentimen masyarakat terhadap pemberlakuan aturan baru program JKN yang diklasifikasikan kedalam tiga kelas yaitu sentimen positif, negatif dan netral. Beberapa penelitian sebelumnya menunjukkan bahwa permasalahan pemberlakuan aturan baru program JKN belum pernah diteliti sebelumnya, sehingga penelitian ini merupakan penelitian baru yang dapat memberikan wawasan berharga dan bermanfaat.

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari media *twitter* dengan memanfaatkan *add-ons* dari *google spreadsheet* yaitu *tweet archiver* menggunakan keywords pencarian bpjs jadi syarat, JKN, Inpres No 1 tahun 2022 dan Jaminan Kesehatan Nasional. Data *tweet* yang diperoleh dari *twitter* berjumlah 1.664 *tweets*. Gambar 2 menunjukkan contoh hasil pengambilan data menggunakan salah satu keywords yaitu bpjs jadi syarat dan terlihat bahwa terdapat beberapa data ganda serta data yang tidak relevan seperti hanya berisikan simbol, tidak sesuai dengan topik pembahasan, tidak berbahasa Indonesia. *Tweet* yang ganda dihapus dengan menggunakan fitur *remove duplicate* dari *google spreadsheet* dan dilanjutkan dengan penghapusan secara manual untuk memastikan tidak ada data ganda

yang tertinggal. Jumlah data setelah proses verifikasi yaitu 1.179 *tweets* yang disimpan dalam format *Comma Separated Value* (.csv).

Twitter Query: bpjs jadi syarat -filter:retweets -filter:replies							
1	Date	Screen Name	Full Name	Tweet Text	Tweet ID	Link(s)	Media
3	28/02/2022 23:18	@Lamz48	Freelance	Boro-boro Mikiran M	1498316770301468	https://twitter.com/historyofpersib/stati	
4	28/02/2022 21:34	@kompascom	Kompas.com	Kegiatan jual-beli tan	1498290429501997	http://dlvr.it/SKpCkc	
5	28/02/2022 20:36	@kompascom	Kompas.com	Kegiatan jual-beli tan	1498275839540035	http://dlvr.it/SKp2CQ	
6	28/02/2022 20:03	@nomorsatukaltim	nomorsatu_kaltim	Kantor BPJS Kesehat	1498267568880062	https://nomorsatukaltim.com/index-ber	
7	28/02/2022 20:00	@officialnova_id	NOVA	Penting buat jual be	1498266743394816	https://nova.grid.id/amp/053160984/jad	
8	28/02/2022 19:38	@FaktaJombang	Fakta Jombang	BPJS Jadi Syarat Mut	1498261249544495	https://faktajombang.com/bpjs-jadi-syar	
9	28/02/2022 18:13	@IDCorner	IDCorner ID	Kapan BPJS Kesehat	1498239918618910	https://idcorner.co.id/2022/02/28/kapar	
10	28/02/2022 18:05	@penabicara17	penabicara	BPJS Kesehatan Jadi	1498237948013154	https://www.penabicara.com/nusantara	
11	28/02/2022 17:47	@Kimberley20101	Kimberley	Jangan jadikan kesel	1498233384346480	https://twitter.com/pic.twitter.com/o0	
12	28/02/2022 17:22	@CeritaParapuan	cerita.parapuan	BPJS Kesehatan Jadi	1498226991497023	http://dlvr.it/SKnV0C	

Gambar 2. Hasil Pengambilan Data

3.2 Pelabelan Manual

Proses pelabelan data dilakukan secara manual oleh pakar bahasa dengan memberikan label positif, negatif dan netral untuk setiap data *tweet* seperti yang ditunjukkan pada tabel 2.

Tabel 2. Pelabelan Manual

Tweet	Label
Optimalisasi JKN Jamin Perlindungan Kesehatan Masyarakat!! #AyoDukungKebangkitanRI	Positif
TOLAK BPJS KESEHATAN JADI SYARAT UMROH/HAJI GAK RELEVAN & MEMBERATKAN!! #MKPengkhiranDemokrasi #MKPengkhiranDemokrasi	Negatif
Mulai 1 Maret, Syarat Pengurusan Jual Beli Tanah Harus Terdaftar Jadi Peserta BPJS Kesehatan	Netral

3.3 Pembentukan Data Training dan Data Testing

Data *tweet* yang digunakan pada proses klasifikasi sebanyak 1.179 data dan dilakukan pembagian data menjadi dua bagian. Gambar 3 menunjukkan proses pembagian data menjadi dua bagian yaitu data *training* sebesar 80% (943 data *tweets*) dan data *testing* 20% (236 data *tweets*) dengan memanfaatkan *library train test split* serta menggunakan *random_state* untuk mengacak dan menjaga data agar tetap sama ketika sistem dijalankan kembali. Nilai *random state* yang digunakan adalah nilai tertinggi dari rentan nilai *random state* yaitu 42.

```
#PembentukanData_train_Test
import pandas as pd
from sklearn.model_selection import train_test_split

[ ] data = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/file_for_colab/jkn_tweet.csv")
data.head()

[ ] x = data['Tweet']
y = data['Sentimen']
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2,random_state=42,stratify=y)
```

Gambar 3. Pembentukan Data Training dan Data Testing

3.4 Preprocessing

Tahap persiapan data atau *preprocessing* bertujuan untuk memastikan data yang akan digunakan sudah siap dan menghilangkan *noise* yang akan mengganggu proses klasifikasi [12].

3.4.1 Data Cleaning

Data dibersihkan dari simbol atau bagian-bagian yang tidak diperlukan yang umumnya terkandung di dalam sebuah *tweet* untuk mengurangi *noise* sehingga mempermudah proses klasifikasi. Tabel 3 menunjukkan hasil proses *data cleaning* dimana terlihat pada hasil *data cleaning* adanya pembersihan *username* maupun simbol-simbol seperti hilangnya *username*, *hashtag*, tanda baca, angka, *emoticon* dan tanda *retweet*

Tabel 3. Data Cleaning pada Data Tweets

Tweet	Hasil Data Cleaning
BPJS Jadi Syarat Urus SIM, STNK, Naik Haji, hingga Jual Beli Tanah Umatnya dituduh radikal, duitnya ngiler Hajaaaaalll 🗨️ 👍 🙌 #CabutAturanBPJS #CabutAturanBPJS	BPJS Jadi Syarat Urus SIM STNK Naik Haji hingga Jual Beli Tanah Umatnya dituduh radikal duitnya ngiler Hajaaaaalll

3.4.2 Case Folding

Proses ini bertujuan menghindari kendala font tidak beraturan antara uppercase dan lowercase yang dilakukan dengan merubah semua kata pada data tweet menjadi huruf kecil (lowercase). Hasil data setelah melalui proses case folding ditunjukkan pada tabel 4 dimana pada kolom hasil case folding semua kata pada data tweet hasil data cleaning telah berubah menjadi huruf kecil.

Tabel 4. Case Folding pada Data Tweets

Hasil Data Cleaning	Hasil Case Folding
BPJS Jadi Syarat Urus SIM STNK Naik Haji hingga Jual Beli Tanah Umatnya dituduh radikal duitnya ngiler Hajaaaaalll	bpjs jadi syarat urus sim stnk naik haji hingga jual beli tanah umatnya dituduh radikal duitnya ngiler hajaaaaalll

3.4.3 Tokenizing

Proses menguraikan sebuah kalimat berdasarkan tiap kata yang menyusun kalimat seperti ditunjukkan pada tabel 5. Pada kolom hasil tokenizing terlihat bahwa kalimat hasil case folding telah diurai menjadi per kata yang ditunjukkan dengan adanya penggunaan tanda baca koma sebagai pemisah antar kata.

Tabel 5. Tokenizing pada Data Tweets

Hasil Case Folding	Hasil Tokenizing
bpjs jadi syarat urus sim stnk naik haji hingga jual beli tanah umatnya dituduh radikal duitnya ngiler hajaaaaalll	bpjs,jadi,syarat,urus,sim,stnk,naik,haji,hingga,jual,beli,tanah,umatnya,dituduh,radikal,duitnya,ngiler,hajaaaaalll

3.4.4 Stemming

Kata-kata pada data tweet yang memiliki imbuhan dihilangkan sehingga hanya menyisakan kata dasar, seperti ditunjukkan pada tabel 6. Contoh yang dapat dilihat yaitu kata hasil tokenizing ‘umatnya’ dengan imbuhan ‘nya’ berubah menjadi kata ‘umat’ setelah melalui proses stemming.

Tabel 6. Stemming pada Data Tweets

Hasil Tokenizing	Hasil Stemming
bpjs,jadi,syarat,urus,sim,stnk,naik,haji,hingga,jual,beli,tanah,umatnya,dituduh,radikal,duitnya,ngiler,hajaaaaalll	bpjs,jadi,syarat,urus,sim,stnk,naik,haji,hingga,jual,beli,tanah,umat,tuduh,radikal,duit,ngiler,hajaaaaalll

3.4.5 Stopword Removal

Menghilangkan kata yang tidak penting dan tidak diperlukan dalam proses klasifikasi sentimen dengan memanfaatkan library NLTK dan hasil proses stopwords removal pada tabel 7. Pada tabel 7 terlihat bahwa terdapat kata-kata pada hasil stemming seperti ‘jadi’, ‘naik’, ‘hingga’ yang hilang setelah melalui proses stopwords removal.

Tabel 7. Stopword Removal pada Data Tweets

Hasil Stemming	Stopword Removal
bpjs,jadi,syarat,urus,sim,stnk,naik,haji,hingga,jual,beli,tanah,umat,tuduh,radikal,duit,ngiler,hajaaaaalll	bpjs,syarat,urus,sim,stnk,haji,jual,beli,tanah,umat,tuduh,radikal,duit,ngiler,hajaaalll

3.5 Pembobotan TF-IDF

Gambar 4 menunjukkan proses pembobotan terhadap setiap kata/term pada dataset dengan tujuan untuk merubah kata-kata menjadi sebuah vektor yang merepresentasikan nilai dari banyaknya kata pada dokumen dimana data yang digunakan pada proses pembobotan ini adalah data yang sudah melalui tahap preprocessing

```

Proses tf-idf_n_SVM_tweet jkn.ipynb
File Edit View Insert Runtime Tools Help

+ Code + Text

[ ] from sklearn.feature_extraction.text import TfidfVectorizer

[ ] dttrain = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/file_for_colab/jkn_train_preproc.csv")
    dttest = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/file_for_colab/jkn_test_preproc.csv")

[ ] x_train = dttrain['tweettrain']
    x_test = dttest['tweettest']
    ytrain = dttrain['sentimen']
    ytest = dttest['sentimen']

[ ] vectorizer = TfidfVectorizer(use_idf=True)
    xtrain = vectorizer.fit_transform(x_train)
    xtest = vectorizer.transform(x_test)
    
```

Gambar 4. Proses Pembobotan TF-IDF

3.6 Klasifikasi Multiclass Support Vector Machine (SVM)

Multiclass SVM menggunakan pendekatan OAA bekerja dengan mencari setiap *hyperplane* atau garis pemisah yang mengklasifikasi kelas sesuai dengan banyak kelas yang digunakan. Pencarian *hyperplane* dilakukan dengan memberikan label 1 pada kelas yang dicari dan label -1 pada kelas lain. Tabel 8 menunjukkan data *sample* yang digunakan pada proses klasifikasi dengan T^{uji} sebagai data *testing* dan proses diawali dengan tahap pencarian *hyperplane* kelas pertama yaitu kelas positif. Pencarian *hyperplane* menggunakan persamaan (6) dan (7) dimana T^{uji} tidak diproses karena T^{uji} digunakan sebagai data *testing*.

Tabel 8. Data *Sample* untuk Klasifikasi

Data	Tweets	Label
T1	jokowi terbit inpers jkn warga jamin sehat masyarakat taat prokes	1
T2	warga protes bpjs syarat umroh jual beli tanah sim	-1
T3	kartu bpjs sehat syarat jual beli tanah maret cek berita lengkap	-1
T^{uji}	Syarat jual beli tanah bpjs sehat aktif	0

Pencarian nilai *support vector* pada setiap dokumen diawali dengan mencari nilai x menggunakan persamaan (9) dan nilai y menggunakan persamaan (10). Nilai x dan y di substitusikan ke persamaan (8) sehingga didapatkan hasil berupa nilai *support vector* dan untuk mendapatkan jarak tegak lurus yang optimal nilai *support vector* diberikan nilai bias 1 dengan hasil pada tabel 9.

Tabel 9. Hasil *Support Vector*

Tweets	S1	S2	S3
<i>Support Vector</i>	3.187 [6.132]	1.055 [1]	1.510 [1]
Bias	1	1	1

Dilakukan perhitungan menggunakan persamaan (11) dari nilai *support vector* dan hasil perhitungan disubstitusikan menggunakan persamaan (12) sehingga didapatkan tiga bentuk persamaan yang mana masing-masing bentuk tersebut digunakan untuk mendapatkan nilai a_1, a_2, a_3 dengan metode gauss Jordan dan nilai a_i yang nol atau positif di substitusikan ke persamaan (13). Mencari nilai w dan b dilakukan dengan menggunakan persamaan (14) sehingga didapatkan nilai *hyperplane* untuk ketiga kelas pada tabel 10.

Tabel 10. Hasil *Hyperplane*

Kelas	Pasangan Kelas	<i>Hyperplane</i> (w^i dan b^i)
1	Positif = 1 Bukan Positif = -1	$w^1 = \begin{bmatrix} 1,055 \\ 3,317 \end{bmatrix}$ dan $b^1 = 3,317$
2	Negatif = 1 Bukan Negatif = -1	$w^1 = \begin{bmatrix} 18,832 \\ 12,469 \end{bmatrix}$ dan $b^1 = 12,469$
3	Netral = 1 Bukan Netral = -1	$w^1 = \begin{bmatrix} 34,400 \\ 32,605 \end{bmatrix}$ dan $b^1 = 32,605$

Pengujian model dilakukan dengan menggunakan data *testing* yang sudah ditentukan yaitu T^{uji} bersama dengan ketiga data training sebelumnya. Data *testing* diberi label 0 dan proses yang dilakukan untuk mencari nilai *support vector* T^{uji} sama persis dengan yang telah dilakukan sebelumnya. Hasil *support vector* untuk data testing T^{uji} adalah $\begin{bmatrix} 0,643 \\ 0 \end{bmatrix}$ dan nilai tersebut disubstitusikan ke persamaan (15) dengan nilai w^1 dan b^1 didapatkan dari hasil pelatihannya sebelumnya pada tabel 10. Nilai *hyperplane* terbesar yang didapatkan adalah 54,740 yang merupakan nilai *hyperplane* kelas ketiga yaitu kelas netral. Sehingga data T^{uji} termasuk kedalam kelas sentiment netral.

3.7 Evaluasi

Pengujian atau evaluasi hasil dilakukan untuk mengecek kinerja *multiclass SVM* dalam melakukan proses analisis sentimen dengan menggunakan *confusion matrix* yang mengelompokkan hasil klasifikasi kedalam tiga sentimen yaitu sentimen positif, negatif dan netral seperti ditunjukkan pada tabel 11.

Tabel 11. *Confusion Matrix*

Aktual	Klasifikasi		
	Positif	Negatif	Netral
Positif	17	4	6
Negatif	0	93	17
Netral	0	18	81

Berdasarkan hasil klasifikasi sentimen pada tabel 11 setiap nilai TP, FP, TN dan FN digunakan untuk mendapatkan nilai *accuracy, precision, recall* dan *f1-score* pada sentimen positif, negatif dan netral menggunakan persamaan (16-19). Hasil pengujian ditunjukkan pada gambar 5.

negatif atau tidak setuju terhadap aturan baru program JKN yang mensyaratkan penggunaan kartu peserta BPJS dalam berbagai layanan publik. Untuk penelitian selanjutnya perlu adanya proses non manual terkait dengan penanganan data ganda yang tidak dapat diatasi dengan *remove duplicate*, menambah jumlah data dan metode lain seperti adaboost sehingga dapat meningkatkan kinerja dari metode yang digunakan.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada semua pihak yang telah membantu dan memberikan kontribusi dalam penelitian serta penulisan artikel sehingga dapat diselesaikan dengan tepat waktu. Ucapan terima kasih juga disampaikan kepada semua pihak yang telah membantu dalam publikasi artikel ini.

REFERENSI

- [1] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [2] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine," *J. Inform. Polinema*, vol. 4, no. 2, pp. 93–99, 2018, doi: 10.33795/jip.v4i2.152.
- [3] P. M. Nirmala Dharmapatni and N. L. P. Merawati, "Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan," *J. Bumigora Inf. Technol.*, vol. 2, no. 2, pp. 105–112, 2020, doi: 10.30812/bite.v2i2.904.
- [4] H. Y. Pradana, I. Slamet, and E. Zukhronah, "Analisis Sentimen Kinerja Pemerintahan Menggunakan Algoritma NBC, KNN, dan SVM," in *Prosiding Simposium Nasional Multidisiplin (SinaMu)*, 2022, p. 114. doi: 10.31000/sinamu.v4i1.7869.
- [5] S. Rahmawati and M. Habibi, "Public Sentiments Analysis about Indonesian Social Insurance Administration Organization on Twitter," *IJID (International J. Informatics Dev.)*, vol. 9, no. 2, pp. 87–93, 2020, doi: 10.14421/ijid.2020.09205.
- [6] Inpres No 1, "Instruksi Presiden RI Nomor 1 Tahun 2022," 2022.
- [7] A. Syaid, "Pemberlakuan Inpres No. 1 Tahun 2022 Tentang Optimalisasi Pelaksanaan Program Jaminan Kesehatan Nasional Dalam Rangka Menyehatkan Bpjs Kesehatan," *J. Huk. to-ra Huk. Untuk Mengatur dan Melindungi Masy.*, vol. 9, no. 1, pp. 38–57, 2023, doi: 10.55809/tora.v9i1.165.
- [8] M. Al Khadafi, Kurnia Paranitha Kartika, and Filda Febrinita, "Penerapan Metode Naïve Bayes Classifier Dan Lexicon Based Untuk Analisis Sentimen Cyberbullying Pada Bpjs," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 725–733, 2022, doi: 10.36040/jati.v6i2.5633.
- [9] E. Hokijuliandy, H. Napitupulu, and Firdaniza, "Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia's National Health Insurance Mobile Application," *Mathematics*, vol. 11, no. 17, p. 3765, 2023, doi: 10.3390/math11173765.
- [10] R. Darman, "Analisis Sentimen Respons Twitter Terhadap Persyaratan Badan Penyelenggara Jaminan Sosial (BPJS) Di Kantor Pertanahan," *J. Widya Bhumi*, vol. 3, no. 2, pp. 113–136, 2023.
- [11] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] A. F. Nuraini, R. D. Pertiwi, M. Z. Subarkah, and K. Ferawati, "Analisis Sentimen Pelaksanaan Vaksinasi Covid-19 secara Massal pada Media Sosial Twitter," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 1385–1394, 2022, doi: 10.34123/semnasoffstat.v2022i1.1564.
- [13] A. Syarif, A. I. Akaf, R. Prabowo, and K. Muludi, "Analisis Sentimen Opini Masyarakat Terhadap Pelayanan BPJS Kesehatan Provinsi Lampung Berbasis Twitter," *J. Pepadun*, vol. 3, no. 3, pp. 380–388, 2022, doi: 10.23960/pepadun.v3i3.136.
- [14] Wulandari Putri Dewi, Enri Ultach, and Primajaya Aji, "Analisis Sentimen Terhadap Kenaikan Iuran Bpjs Kesehatan Pada Twitter Menggunakan Naïve Bayes Classifier," *J. Ilmu Komput. Dan Teknol.*, vol. 1, no. 2, pp. 18–22, 2020.
- [15] D. N. Fitriana and Y. Sibaroni, "Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM)," *RESTI J.*, vol. 4, no. 5, pp. 846–853, 2020.
- [16] D. N. Fitriana and Y. Sibaroni, "Klasifikasi Data Tweet dengan Menggunakan Metode Klasifikasi Multi-Class Support Vector Machine: Studi Kasus PT. KAI," *e-Proceeding Eng.*, vol. 7, no. 2, pp. 8493–8505, 2020, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/12746>
- [17] L. R. Halim and A. Suryadibrata, "Cyberbullying Sentiment Analysis with Word2Vec and One-Against-All Support Vector Machine," *IJNMT (International J. New Media Technol.)*, vol. 8, no. 1, pp. 57–64, 2021, doi: 10.31937/ijnmt.v8i1.2047.
- [18] M. K. Delimayanti, R. Sari, M. Laya, M. R. Faisal, and P. Pahrul, "Pemanfaatan Metode Multiclass-SVM

pada Model Klasifikasi Pesan Bencana Banjir di Twitter,” *Edu Komputika J.*, vol. 8, no. 1, pp. 39–47, 2021, doi: 10.15294/edukomputika.v8i1.47858.