Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Mei 23, 2025 | Revised Mei 30, 2025 |
Accepted Jun 20, 2025

Hal. 878-885
Vol. 7; No. 2
Mei 2025

ANALISIS SENTIMEN LAYANAN KESEHATAN BPJS MENGGUNAKAN METODE SVM

Ratih Ayu Puspita Sari[1*], Slamet Kacung[2], Budi Santoso[3]
[1,2,3]Teknik Informatika, Universitas Dr Soetomo
email: ratihayupuspitasari158@gmail.com[1*]

Abstrak: Di era digital, teknologi komunikasi menjadi alat penting dalam mendukung penyebaran informasi dan akses layanan kesehatan secara cepat dan efisien. Mutu layanan kesehatan kini menjadi perhatian utama masyarakat, khususnya terhadap layanan BPJS. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan kesehatan BPJS melalui komentar di media sosial X dan YouTube dengan menggunakan algoritma Support Vector Machine (SVM) sebagai metode klasifikasi. Data diperoleh melalui crawling terhadap 18.582 komentar, kemudian diproses melalui tahap preprocessing dan pelabelan berbasis pendekatan lexicon-based. Hasil klasifikasi menunjukkan bahwa model SVM dengan kernel linear mampu mencapai akurasi sebesar **89,97%**. Model ini menunjukkan performa sangat baik dalam mendeteksi sentimen negatif dengan **f1-score sebesar 0,95**. Pada sentimen netral dan positif, model menunjukkan kinerja yang cukup baik dengan nilai recall masing-masing sebesar **0,51**. Penelitian ini memberikan gambaran nyata mengenai persepsi masyarakat terhadap layanan BPJS serta membuka peluang pengembangan model klasifikasi yang lebih seimbang dalam mendeteksi seluruh jenis sentimen.

Kata Kunci : BPJS, Administrasi Pelayanan, SVM, Analisis Sentimen

PENDAHULUAN

Perkembangan teknologi informasi yang sangat pesat telah menjangkau berbagai sektor, termasuk bidang kesehatan. Di era digital ini teknologi komunikasi menjadi alat penting untuk mendukung penyebaran informasi dan akses layanan kesehatan yang cepat dan efisien[1]. Saat ini mutu layanan kesehatan menjadi perhatian utama masyarakat[2]. Mutu layanan kesehatan mencerminkan upaya memenuhi kebutuhan masyarakat, didukung kebijakan pemerintah berbasis aspirasi publik. Pemerintah juga dapat menggunakan masukan ini untuk meningkatkan kinerja dan kualitas layanan kesehatan secara keseluruhan. Masukan masyarakat mencakup pendapat positif dan kritik negatif[3]. Proses meninjau pandangan masyarakat ini dikenal dengan istilah analisis sentimen[4]. Analisis sentimen atau opinion mining mengidentifikasi dan mengambil opini dari dokumen terkait topik tertentu, seperti layanan Kesehatan BPJS. BPJS atau Badan Penyelenggara Jaminan Sosial adalah lembaga yang dibentuk untuk melaksanakan program jaminan sosial di Indonesia[5]. BPJS adalah program pemerintah untuk melindungi dan menyejahterakan masyarakat. Program ini menyediakan layanan kesehatan yang adil dan merata bagi seluruh rakyat Indonesia[6].

Permasalahan dalam penelitian ini adalah persepsi masyarakat terhadap layanan kesehatan BPJS yang menjadi salah satu indikator penting dalam menilai kualitas pelayanan kesehatan di Indonesia. Persepsi ini mencakup berbagai pandangan, baik positif maupun negatif, yang dapat memengaruhi kepercayaan masyarakat terhadap BPJS sebagai penyelenggara jaminan sosial kesehatan[7]. Faktor utama yang melatarbelakangi permasalahan ini adalah kualitas layanan yang diterima masyarakat, yang sering kali dianggap belum memenuhi harapan, terutama dalam hal aksesibilitas, kecepatan, dan keadilan pelayanan. Dalam konteks era digital saat ini, analisis sentimen menjadi metode yang relevan untuk menggali opini masyarakat secara komprehensif. Oleh karena itu, penelitian ini berfokus pada pemanfaatan analisis sentimen untuk memahami persepsi masyarakat, mengidentifikasi faktor yang memengaruhi kualitas layanan, dan memberikan rekomendasi strategis untuk meningkatkan mutu layanan kesehatan BPJS secara keseluruhan.

TINJAUAN PUSTAKA

Berdasarkan studi literatur pada penelitian sebelumnya yang membahas opini masyarakat terkait layanan kesehatan BPJS melalui media sosial, penelitian [3] melakukan analisis sentimen pada komentar media sosial Instagram terkait layanan kesehatan BPJS menggunakan metode Naïve Bayes Classifier. Hasil penelitian menunjukkan bahwa metode ini memiliki performa yang baik dengan akurasi mencapai 73%. Penelitian lain [5] menggunakan analisis sentimen berdasarkan opini pengguna di Twitter dengan metode Lexicon Based dan Naïve Bayes Classifier. Penelitian ini menghasilkan akurasi sistem sebesar 71% menggunakan algoritma Naïve Bayes Classifier terhadap tweet, dengan 29% sebagai tingkat kesalahan sistem dalam membaca data.

Namun, penelitian yang memanfaatkan kombinasi metode Lexicon Based untuk pelabelan sentimen dan algoritma Support Vector Machine (SVM) untuk klasifikasi sentimen guna mengevaluasi opini masyarakat terhadap layanan BPJS menggunakan dataset Twitter dan Youtube secara mendalam masih terbatas. Oleh karena itu, Penelitian ini diharapkan dapat memberikan rekomendasi bagi pihak BPJS untuk memperbaiki kualitas layanan, serta memperkuat komunikasi dengan masyarakat guna menciptakan transparansi dan kepercayaan yang lebih baik.

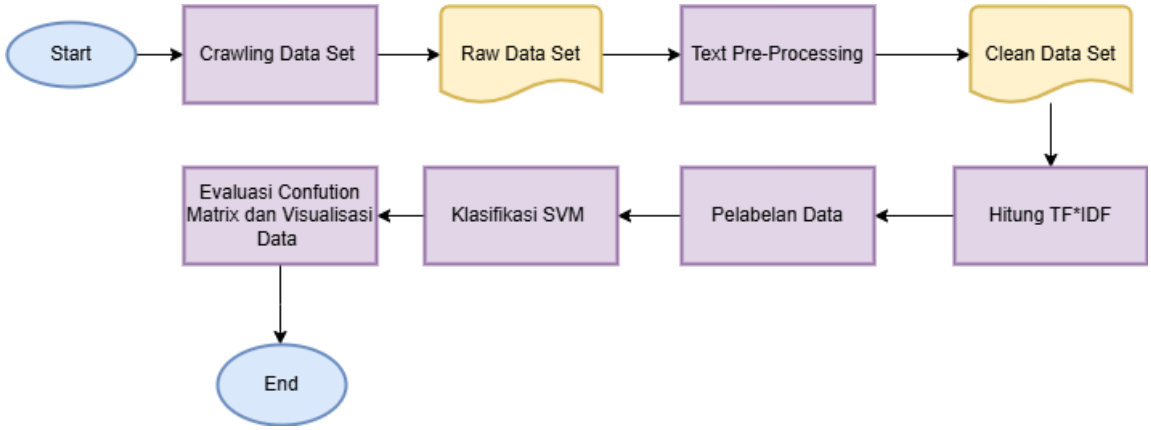
.

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Mei 23, 2025 | Revised Mei 30, 2025 |
Accepted Jun 20, 2025
Hal. 878-885
Vol. 7; No. 2
Mei 2025

METODE

Penelitian ini bertujuan untuk melakukam analisis sentimen pada komentas di media sosial X dan Youtube terkait layanan kesehatan BPJS dengan menggunakan algortima Support Vector Machine (SVM) sebagai metode klasisfikasi. Metode pelabelan dalam penelitian ini menggunakan pendekatan Lexicom Based untuk mempermudah proses identifikasi sentimen. Metode pada Penelitian ini meliputi beberapa tahapan utama, yaitu Pengumpulan data, Preprocessing Data, Implementasi metode (SVM), serta Analisis Hasil[8].

Tahap Penelitian



Gambar 1. Algoritma Proses

Pengumpulan Data

Berdasarkan Gambar 1 Tahap awal pengumpulan data dari publik berdasarkan komentar dan ulasan di media sosial Twitter dan Youtube terkait layanan kesehatan BPJS dengan menggunakan tagar dan kata kunci seperti BPJS, LayananKesehatabBPJS, dan KesehatanBPJS. Selanjutnya, data yang terkumpul akan diproses lebih lanjut pada tahap Data Preprocessing. Proses **Crawling** ini bertujuan untuk mengumpulkan halaman web yang berisi ulasan relevan dengan kriteria sebagai berikut: Dataset yang diolah adalah data teks dari repon netizen dengan bahasa Indonesia. Proses Crawling ini menggunakan tools **tweet-harvest** berbasis **Node.js** yang dijalankan melalui perintah npx, dengan bantuan **Python** untuk mengatur alur dan menyimpan hasil scraping tweet ke file CSV berdasarkan kata kunci diatas, di mana **Node.js** diinstal terlebih dahulu karena dibutuhkan oleh tweet-harvest, lalu proses scraping dilakukan dengan autentikasi token dan hasilnya digunakan untuk analisis lebih lanjut seperti analisis sentimen. Jumlah data ulasan yang diambil sebanyak 18582 data. Penelitian ini hanya akan fokus pada analisis sentimen terhadap layanan kesehatan BPJS Tingkat Pertama yaitu Administrasi Pelayanan, tanpa melibatkan topik lain yang berkaitan dengan BPJS atau jaminan sosial lainnya.

Preprocessing Data

Tahap Data Preprocessing merupakan serangkaian langkah yang diterapkan pada data mentah untuk mempersiapkan sebelum digunakan dalam analisis lanjutan atau pembuatan model. Data preprocessing bertujuan utama untuk meningkatkan kualitas data, memastikan hasil analisis lebih akurat, dan mengatasi berbagai masalah atau kekurangan yang terdapat dalam data mentah[9]. Proses preprocessing meliputi penghapusan data duplikat, perbaikan kesalahan seperti tipografi, serta pemeriksaan inkonsistensi data. Proses ini juga mencakup enrichment, yaitu menambahkan informasi relevan untuk mendukung proses Knowledge Discovery in Databases (KDD). Preprocessing data berfungsi untuk mengubah data ke dalam format yang lebih sederhana dan efisien[10]. Tahapan Text Preprocessing meliputi cleaning, case folding, normalisasi, tokenizing, stemming, dan stopword removal[11].

1. **Cleaning**
   Tahapan cleaning dalam text preprocessing adalah proses untuk membersihkan teks dari karakter atau elemen yang tidak relevan, seperti tanda baca, angka, atau simbol, guna mempersiapkan data teks menjadi lebih terstruktur dan siap untuk analisis lebih lanjut. Proses ini sering menggunakan modul seperti "re" pada Python, yang memanfaatkan ekspresi reguler untuk mendeteksi dan menghapus elemen yang tidak diperlukan[11].
2. **Case Folding**
   Tahapan case folding dalam text preprocessing adalah proses mengubah semua huruf dalam teks menjadi huruf kecil untuk memastikan konsistensi data selama analisis. Langkah ini bertujuan menghilangkan perbedaan akibat penggunaan huruf besar dan kecil, sehingga memudahkan pemrosesan teks. Case folding biasanya diterapkan menggunakan metode seperti lower(), dan hasilnya dapat disimpan dalam format terstruktur, seperti kolom baru dalam sebuah dataframe[11].
3. **Normalisasi**
   Tahapan normalisasi dalam text preprocessing adalah proses mengubah kata-kata tidak baku atau slang menjadi kata-kata baku yang sesuai dengan standar bahasa. Normalisasi bertujuan untuk meningkatkan konsistensi data sehingga mempermudah analisis. Langkah ini biasanya dilakukan dengan mencocokkan kata-kata slang dalam

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Mei 23, 2025 | Revised Mei 30, 2025 |
Accepted Jun 20, 2025
Hal. 878-885
Vol. 7; No. 2
Mei 2025

teks dengan daftar kata baku yang disimpan dalam file referensi, yang kemudian diterapkan pada teks menggunakan teknik seperti ekspresi reguler atau pengolahan dataframe[11].

4. **Tokenizing**
   Tahapan tokenizing adalah proses memecah teks menjadi unit-unit yang lebih kecil, seperti kata, frasa, atau kalimat, yang disebut token. Tujuan tokenizing adalah untuk mempermudah analisis teks dengan mengidentifikasi komponen- komponen dasar dalam teks. Fungsi tokenizing diterapkan pada data, seperti kolom hasil normalisasi, dan hasilnya disimpan dalam kolom baru, misalnya "tokenizing," untuk analisis lebih lanjut. Proses ini membantu dalam memahami struktur teks dan mempersiapkannya untuk tahap pemrosesan berikutnya[11].
5. **Stemming**
   Tahapan stemming bertujuan untuk mengubah kata-kata dalam teks menjadi bentuk dasarnya (root word). Proses ini dilakukan menggunakan modul StemmerFactory dari library Sastrawi, yang menyediakan fasilitas untuk stemming. Tahapan ini membantu menyederhanakan analisis teks dengan menghilangkan variasi kata yang memiliki makna serupa. Tahapan ini membantu menyederhanakan analisis teks dengan menghilangkan variasi kata yang memiliki makna serupa, sehingga memudahkan proses pengelompokan dan pemahaman konteks dari data teks.
6. **Stopword Removal**
   Tahapan stopword removal adalah proses menghapus kata-kata umum yang tidak memiliki nilai informasi signifikan, seperti "dan," "di," atau "yang," dari data teks. Proses ini dilakukan untuk meningkatkan efisiensi analisis dengan menyisakan kata-kata yang lebih relevan terhadap konteks. Dalam implementasinya, library seperti Sastrawi sering digunakan untuk mendapatkan daftar stopword yang dapat disesuaikan sesuai kebutuhan, sehingga proses pembersihan teks menjadi lebih efisien dan hanya berfokus pada kata-kata yang memiliki makna penting dalam analisis.

TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (Term Frequency-Inverse Document Frequency) dapat diartikan sebagai metode untuk mengukur tingkat kepentingan sebuah kata dalam suatu dokumen di antara kumpulan dokumen. Nilai TF-IDF yang tinggi mengindikasikan bahwa kata tersebut sering muncul dalam dokumen tertentu namun jarang ditemukan di dokumen lainnya, sehingga membantu mengidentifikasi kata kunci yang relevan[12]. Perhitungan TF*IDF didefinisikan sebagai berikut[13]:

Menghitung Invers Dokumen

$$idf = log\ \frac{jumlah\ seluruh\ dokumen}{jumlah\ dokumen\ yang\ berbobot\ kata} \quad (1)$$

Menghitung Frekuensi Kata

$$tf(k, d) = f(jumlah\ frekuensi\ kata\ yang\ muncul) * idf \quad (2)$$

Menghitung Nilai Tinggi Term yang Sering Muncul

$$tfidf(k, d) = tf(jumlah\ dokumen\ berbobot\ kata) * idf \quad (3)$$

Metode Analisis Sentimen

Penelitian ini menggunakan metode Support Vector Machine (SVM). Support Vector Machine (SVM) merupakan salah satu metode yang lebih baru jika dibandingkan dengan teknik lainnya, namun telah menunjukkan kinerja yang unggul di bidang seperti bioinformatika, pengenalan tulisan tangan, dan klasifikasi teks. Tujuan utama dari metode ini adalah untuk membangun Optimal Separating Hyperplane (OSH), yang menghasilkan fungsi pemisah optimal yang dapat digunakan untuk proses klasifikasi[11]. Selain itu, Metode ini bertujuan untuk meminimalkan kesalahan klasifikasi dengan cara mencari margin terbesar di antara kelas-kelas tersebut, sehingga dapat menghasilkan model yang lebih robust dan memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya[14].

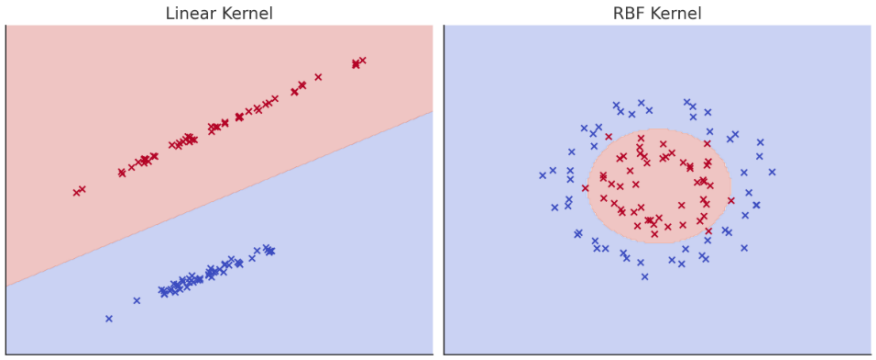
Support Vector Machine SVM dapat diartikan juga sebagai metode pembelajaran mesin yang beroperasi berdasarkan prinsip Structural Risk Minimization (SRM). Metode ini bertujuan untuk meminimalkan kesalahan klasifikasi dengan cara mencari margin terbesar di antara kelas-kelas tersebut, sehingga dapat menghasilkan model yang lebih robust dan memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya[14]. Persamaan rumus untuk tiga kernel didefinisikan sebagai berikut[15]:

Menghitung Linear Kernel

$$K(x, x') = x \cdot x' \quad (4)$$

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Mei 23, 2025 | Revised Mei 30, 2025 |
Accepted Jun 20, 2025
Hal. 878-885
Vol. 7; No. 2
Mei 2025

$Dimana:$
$\chi, \chi' = nilai\ kernel\ antara\ dua\ vektor$



Gambar 2. Visualisasi Linear Kernel dan RBF Kernel

Pelabelan Data

Pelabelan data merupakan langkah untuk menetapkan label seperti positif, negatif, atau netral pada teks guna memberikan pemahaman mendalam terhadap data. Proses ini tidak hanya mengandalkan sumber daya leksikal yang ada untuk mengidentifikasi sentimen, tetapi juga membutuhkan analisis manual atas kata-kata dan konteksnya yang dapat menunjukkan pola persetujuan maupun penolakan[16]. Pelabelan otomatis dapat dilakukan menggunakan dataset kamus emosi atau sentiment lexicon. Metode ini efektif untuk menangani jumlah teks yang besar, sehingga dapat mengurangi potensi kesalahan pada dokumen. Tujuan utama dari pelabelan data adalah untuk membentuk data latih dan data uji yang digunakan dalam menghitung akurasi hasil analisis sentimen secara otomatis. Berikut ini adalah rumus yang digunakan dalam metode Lexicon-Based[17].

$$Sentiment\ (s^i) = \Sigma_{i=1}^{n} Sentiment\ (W_i) \quad (5)$$

Dimana:
n = jumlah kata
w = kata
$si = up > 0, down < 0. unknown = 0$

Confusion Matrix

Confusion Matrix adalah metode yang umum digunakan untuk mengukur tingkat akurasi dalam proses data mining. Matrix ini memberikan informasi mengenai jumlah klasifikasi yang berhasil diprediksi dengan benar oleh suatu sistem klasifikasi[18]. Confusion Matrix disajikan dalam bentuk tabel yang terdiri dari empat kombinasi antara nilai prediksi dan nilai aktual. Confusion Matrix digunakan dalam klasifikasi biner, di mana hasil prediksi dikategorikan ke dalam dua kelas, yaitu kelas positif dan kelas negatif[19]. Berikut adalah variabel untuk menentukan hasil dari kelas positif dan negatif:

1. *Accuracy*
   Akurasi adalah salah satu metode dasar yang digunakan untuk mengukur performa dalam klasifikasi biner[19], yang dapat dirumuskan sebagai berikut:
   $$Accuracy = \frac{TN}{TN+TN+FP+FN} \quad (6)$$

2. *Precision*
   Merupakan perbandingan antara data positif yang berhasil diklasifikasikan dengan benar dan total jumlah prediksi bernilai positif[19], dengan rumus sebagai berikut:
   $$Precision = \frac{TP}{TP+FP} \quad (7)$$

3. *Recall*
   Sensitivity, atau dikenal juga sebagai True Positive Rate, merupakan istilah lain untuk recall. Sensitivity juga diartikan sebagai persentase data positif (TP) yang berhasil diprediksi dengan benar[19]. Nilai recall yang semakin tinggi menunjukkan bahwa kategori tersebut dikenali dengan baik (FN rendah). Rumusnya adalah sebagai berikut:
   $$Recall = \frac{TP}{TP+FN} \quad (8)$$

4. *F1-Score*

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Mei 23, 2025 | Revised Mei 30, 2025 |
Accepted Jun 20, 2025
Hal. 878-885
Vol. 7; No. 2
Mei 2025

Digunakan untuk mengukur keterkaitan antara nilai precision dan recall[29], yang dapat dirumuskan sebagai berikut:

$$F1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad (9)$$

Dimana:
TP (True Positive) : Merupakan jumlah data sebenarnya positif yang berhasil diprediksi sebagai positif
FP (False Positive): Merupakan jumlah data sebenarnya negatif tetapi diprediksi sebagai positif
TN (True Negative): Merupakan jumlah data sebenarnya negatif yang berhasil diprediksi sebagai negatif.
FN (False Negative): Merupakan jumlah data sebenarnya positif tetapi diprediksi sebagai negatif.

HASIL DAN PEMBAHASAN

Data dalam penelitian ini diperoleh dari media sosial Twitter dan YouTube dengan memanfaatkan teknik crawling berbasis Python, selama periode Januari 2021 hingga Januari 2025. Sebanyak 18.582 komentar berhasil dikumpulkan menggunakan kata kunci seperti "BPJS", "Kesehatan", "Administrasi", "Pelayanan", "JKN", dan "Mobile". Kata kunci ini dipilih berdasarkan topik-topik dominan yang sering muncul dalam percakapan publik mengenai layanan BPJS Kesehatan. Data yang diperoleh kemudian melalui proses preprocessing, seperti pembersihan teks, normalisasi, tokenisasi, hingga stemming dan stopword removal, guna meningkatkan kualitas input data. Setelah itu, proses pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based, yang memungkinkan klasifikasi komentar ke dalam kategori positif, negatif, dan netral.

Model klasifikasi dibangun menggunakan algoritma Support Vector Machine (SVM) dengan kernel linear, dengan komposisi data latih sebesar 80% dan data uji sebesar 20%. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 89,97%, dengan kinerja paling optimal pada klasifikasi sentimen negatif (f1-score 0,95). Model juga menunjukkan kemampuan yang memadai dalam mendeteksi sentimen netral dan positif, dengan masing-masing recall sebesar 0,51. Visualisasi melalui confusion matrix memperkuat hasil ini, di mana prediksi terhadap sentimen negatif menunjukkan tingkat ketepatan yang sangat tinggi, sementara prediksi pada kategori lain tetap menunjukkan konsistensi meskipun dalam tingkat yang lebih moderat. Hasil ini memberikan gambaran komprehensif tentang persepsi masyarakat terhadap layanan BPJS, khususnya terkait aspek administrasi dan akses digital melalui Mobile JKN.

Table 1. Hasil Komentar Twitter dan Youtube

Komentar Twitter	Komentar Youtube
Apbn untuk kesehatan itu tahun lalu cuma 187 T itu diluar masyarakat yg iuran BPJS ya.	BPJS KACAU Dokter ciptaning kritisi pak jokowi
Penanganan gangguan kesehatan mental itu emang ditanggung oleh BPJS tapi lokasinya ini yang gak banyak dan umumnya terpusat di kota-kota besar aja.	Nakes Banding Bandingkan Pelayanan Pasien BPJS dan Umum
Peserta BPJS Kes bernasib diterlantarkan. Mohn kpda Bapak Presiden Prabowo agar lebih memperhatikan peserta BPJS Kes agar diprioritaskan dlm peleyanan Kes	Viral Akibat Pelayanan Buruk, Keluarga Pasien BPJS Mengamuk
Biasa memang kalo layanan BPJS agak kurang bagus layanannya,Biasa memang kalo layanan BPJS agak kurang bagus layanannya.	KACAU! MENKES CURIGAI ORANG KAYA BIKIN BPJS BANGKRUT!
Kesel bgt bpjs kesehatan aktif gadipake tapi bayaaaar mulu,kesel bgt bpjs kesehatan aktif gadipake tapi bayaaaar mulu.	Pasien BPJS di Palopo Dicuekin Rumah Sakit, Infus dan Obat Dihentikan dan Diganti Sirup

Tabel 1 menyajikan hasil crawling data yang memuat ulasan pengguna terhadap komentar media sosial X dan Youtube. Informasi yang ditampilkan mencakup ID ulasan, nama pengguna, isi komentar. Data ini memberikan gambaran mengenai persepsi pengguna, tingkat relevansi komentar, sehingga dapat dimanfaatkan dalam analisis sentimen dan penilaian kualitas pelayanan BPJS Kesehatan.

Tahapan Preprocessing

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Mei 23, 2025 | Revised Mei 30, 2025 |
Accepted Jun 20, 2025

Hal. 878-885
Vol. 7; No. 2
Mei 2025

Tabel 2. Tahapan Preprocessing Data

Tahapan	Hasil Transformasi
Comment	Apbn untuk kesehatan itu tahun lalu cuma T itu angkanya kecil
Cleaned	Apbn untuk kesehatan itu tahun lalu cuma T itu angkanya kecil
Case Folding	apbn untuk kesehatan itu tahun lalu cuma t itu angkanya kecil
Tokenizing	['apbn', 'untuk', 'kesehatan', 'itu', 'tahun', 'lalu', 'cuma', 't', 'itu', 'angkanya', 'kecil']
Normalisasi	['apbn', 'untuk', 'kesehatan', 'itu', 'tahun', 'lalu', 'cuma', 't', 'itu', 'angkanya', 'kecil']
Stopword Removal	['apbn', 'kesehatan', 'tahun', 'lalu', 'cuma', 't', 'angkanya', 'kecil']
Stemming	['apbn', 'sehat', 'tahun', 'lalu', 'cuma', 't', 'angka', 'kecil']

Tabel 2 menunjukkan tahapan preprocessing yang diterapkan pada komentar hasil crawling, yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan dan mempersiapkan data sebelum dianalisis. Tahapan ini meliputi pembersihan data (cleansing), penyamaan huruf (case folding), tokenisasi, normalisasi, penghapusan kata tidak bermakna (stopword removal), serta proses stemming.

Pelabelan Data

```
Distribusi kelas sebelum split:
sentiment
negative    15208
positive     1688
neutral      1686
Name: count, dtype: int64

Distribusi kelas setelah split:
sentiment
negative    12166
positive     1350
neutral      1349
Name: count, dtype: int64
```

Gambar 3. Distribusi Sentimen Setelah Pelabelan

Gambar 5 menunjukkan proses pelabelan data, yaitu langkah penting dalam analisis sentimen untuk menetapkan label seperti positif, negatif, atau netral pada sebuah teks. Proses ini memanfaatkan sumber daya leksikal seperti kamus emosi (sentiment lexicon), namun tetap memerlukan analisis manual terhadap kata dan konteks untuk mengidentifikasi pola persetujuan atau penolakan dalam sebuah pernyataan.

Gambar 4. Wordcloud Sentimen

Evaluasi Model

Setelah melalui tahapan pelabelan dan pembobotan TF-IDF, proses klasifikasi dilanjutkan menggunakan metode Support Vector Machine (SVM) dengan kernel linear. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk

Jurnal Informatika Teknologi dan Sains (JINTEKS)
E-ISSN : 2686-3359
Submission Mei 23, 2025 | Revised Mei 30, 2025 |
Accepted Jun 20, 2025
Hal. 878-885
Vol. 7; No. 2
Mei 2025

pengujian. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Selain itu, visualisasi confusion matrix digunakan untuk menggambarkan distribusi prediksi antar kelas sentimen.

Kinerja Model Klasifikasi

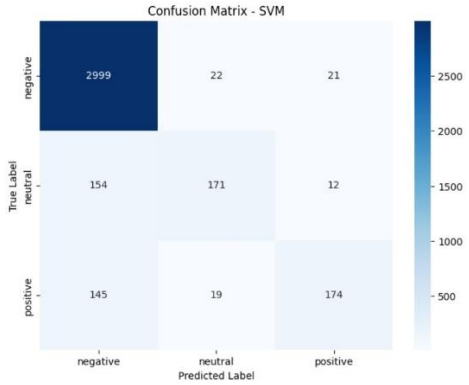
Table 3. Hasil Klasifikasi Model

Label	Precision	Recall	F1-Score	Support
Negative	0.91	0.99	0.95	3042
Neutral	0.81	0.51	0.62	337
Positive	0.84	0.51	0.64	338

Tabel 3 menampilkan hasil evaluasi model SVM dengan kernel linear untuk klasifikasi sentimen. Model ini mencapai akurasi sebesar 89,97% dan menunjukkan performa sangat baik dalam mengklasifikasikan sentimen negatif dengan f1-score 0,95. Pada kelas sentimen netral dan positif, nilai recall mencapai 0,51, sehingga kinerja model pada kelas tersebut berbeda dibandingkan kelas negatif. Secara keseluruhan, model menunjukkan performa yang baik dengan nilai rata-rata tertimbang f1-score sebesar 0,89.

Hasil model SVM dalam penelitian ini menunjukkan performa yang lebih tinggi dibandingkan dengan penelitian[20], baik dari segi akurasi maupun f1-score. Dengan workflow yang identik, mulai dari tahap pre-processing hingga klasifikasi sentimen menggunakan metode Support Vector Machine, perbedaan performa dapat dilihat dari akurasi yang dicapai. Model SVM dalam penelitian ini berhasil mencapai akurasi sebesar 89,97%, sedangkan pada penelitian [20] hanya memperoleh akurasi sebesar 73%. Salah satu faktor yang memengaruhi perbedaan ini kemungkinan terletak pada kualitas dan kelengkapan tahapan pre-processing, mengingat penelitian terdahulu menekankan bahwa tingkat akurasi sangat dipengaruhi oleh pengolahan kata-kata pada tahap awal. Selain itu, model dalam penelitian ini juga menunjukkan f1-score tinggi khususnya pada kelas sentimen negatif (0,95), yang mencerminkan keandalan model dalam mengklasifikasikan data dari platform yang sama secara lebih akurat.

Confusion Matrix



Gambar 5. Visualisasi Confusion Matrix SVM

Gambar 5 menampilkan confusion matrix dari model klasifikasi Support Vector Machine (SVM) untuk analisis sentimen dengan tiga kategori: negatif, netral, dan positif. Hasil menunjukkan bahwa model sangat akurat dalam mengklasifikasikan sentimen negatif, dengan 2.999 data berhasil diprediksi dengan benar, serta hanya sedikit kesalahan dalam memprediksi data negatif sebagai netral atau positif. Namun, akurasi model menurun pada kategori netral dan positif, terlihat dari cukup banyaknya data yang salah diklasifikasikan sebagai negative

KESIMPULAN DAN SARAN

Penelitian ini berhasil mengklasifikasikan sentimen terhadap layanan kesehatan BPJS menggunakan metode Support Vector Machine (SVM) dengan kernel linear. Model yang dikembangkan mencapai akurasi sebesar 89,97% dan menunjukkan performa sangat baik dalam mengidentifikasi sentimen negatif, dengan nilai f1-score sebesar 0,95. Sementara itu, performa model dalam mendeteksi sentimen netral dan positif menunjukkan hasil yang bervariasi, dengan nilai recall sebesar 0,51 untuk masing-masing kelas tersebut. Hasil ini mengindikasikan bahwa metode SVM efektif dalam melakukan klasifikasi sentimen, khususnya pada data dengan ciri yang tegas. Penelitian ini juga menunjukkan bahwa pendekatan lexicon-based yang digunakan dalam proses pelabelan mampu menghasilkan distribusi sentimen yang memungkinkan model untuk belajar secara optimal.

Penelitian selanjutnya disarankan untuk mengoptimalkan performa klasifikasi pada sentimen netral dan positif, misalnya melalui penyeimbangan jumlah data per kelas atau eksplorasi fitur yang lebih representatif. Penggunaan teknik pelabelan yang lebih kompleks atau gabungan pendekatan lexicon-based dan machine learning juga dapat dipertimbangkan. Selain itu, pengujian dengan model lain seperti Random Forest atau Neural Network dapat menjadi arah pengembangan untuk membandingkan efektivitas metode klasifikasi yang digunakan. Penelitian ini juga dapat diperluas dengan menganalisis sentimen terhadap aspek-aspek spesifik dalam layanan BPJS, seperti fasilitas, administrasi, atau pelayanan tenaga medis, guna memperoleh informasi yang lebih terperinci.

DAFTAR PUSTAKA

[1] M. C. Yuniar *et al.*, "Pengembangan Teknologi dalam Bidang Kesehatan," *J. Teknol. Sist. Inf. dan Apl.*, vol. 18, no. 2, pp. 49–52, 2022, [Online]. Available: https://www.e-journal.poltekkesjogja.ac.id/index.php/JTK/article/view/1143%0Ahttps://www.e-journal.poltekkesjogja.ac.id/index.php/JTK/article/download/1143/998

[2] R. Machmud, "Manajemen Mutu Pelayanan Kesehatan," *J. Kesehat. Masy. Andalas*, vol. 2, no. 2, pp. 186–190, 2008, doi: 10.24893/jkma.v2i2.31.

[3] A. Karim, S. F. C, and Mustafa, "Analisis Sentimen pada Komentar Sosial Media Instagram Layanan Kesehatan BPJS Menggunakan Naïve Bayes Classifier Instagram Layanan Kesehatan BPJS Menggunakan Naïve Bayes Classifier," *Pros. Semin. Nas. Konstelasi Ilm. Mhs. UNISSULA 7 (KIMU 7)*, vol. 7, no. Kimu 7, pp. 61–70, 2022, [Online]. Available: http://jurnal.unissula.ac.id/index.php/kimueng/article/view/20511/6627

[4] D. D. Nada, R. M. Atok, and A. P. Data, "928X Print) D480," vol. 11, no. 6, 2022, [Online]. Available: https://t.co/2nUaexGu5i

[5] A. Rosadi *et al.*, "Analisis Sentimen Berdasarkanan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naïve Bayes Classi er Twitter Text Mining," vol. 20, pp. 39–52, 2021.

[6] D. A. Aziz, A. Y. U. N. Sari, and F. Hukum, "1. Adalah dosen fakultas hukum universitas bung karno 2. Adalah mahasiswa fakultas hukum universitas bung karno," vol. 3, no. 1, 2019.

[7] M. Taufik Sugandi, Martanto, and U. Hayati, "Analisis Sentimen Komentar Pengguna Youtube terhadap Kebijakan Baru Badan Penyelenggara Jaminan Kesehatan Sosial Menggunakan Naïve Bayes," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 218–227, 2024.

[8] S. Arifin and B. A. Febryanto, "Analisis Sentimen Pengguna Aplikasi Merdeka Mengajar Menggunakan Metode Vader Lexicon Sentiment Analysis of Merdeka Mengajar Application Users Using the Vader Lexicon Method," vol. 15, no. 1, pp. 1–11, 2025.

[9] I. M. D. M. Oleh Prastyadi Wibawa Rahayu, I Gede Iwan Sudipa, Suryani Suryani, Arie Surachman, Achmad Ridwan, I Gede Mahendra Darmawiguna, Muh. Nurtanzis Sutoyo, Isnandar Slamet, Sitti Harlina, "Buku Ajar Data Mining." p. 209, 2024.

[10] M. K. Oleh Ir. T. Irfan Fajri, S.Kom., M.M.S.I, Herlina Latipa Sari, S.Kom., M.Kom, Rozzi Kesuma Dinata, S.T., M.Eng, Novia Hasdyna, S.T., M.Kom, Buyung Solihin Hasugian, S.Kom., M.Kom, Sujacka Retno, S.T., M.Kom, Sri Wahyuni, S.Kom., M.Kom, Cut Fadhilah, S.T., "Data Mining." p. 124, 2024.

[11] R. A. Oleh Moch Arifqi Ramadhan, "KLASIFIKASI TEXT SPAM MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES." p. 57, 2022.

[12] A. D. P. Oleh Randi Rian Putra, Nadya Andhika Putri, "Teknik Cosine Similarity Dan TF-IDF Dalam Analisis Data." p. 54, 2024.

[13] J. Silge and D. Robinson, *Text Mining with R Text Mining with R Revision History for the First Edition*. 2017. [Online]. Available: http://oreilly.com/catalog/errata.csp?isbn=9781491981658

[14] Y. X. Chu, X. G. Liu, and C. H. Gao, "Multiscale models on time series of silicon content in blast furnace hot metal based on Hilbert-Huang transform," *Proc. 2011 Chinese Control Decis. Conf. CCDC 2011*, pp. 842–847, 2011, doi: 10.1109/CCDC.2011.5968300.

[15] A. Kowalczyk, "Support Vector Machines - The Complete Book," p. 114, 2017, [Online]. Available: www.syncfusion.com.

[16] A. S. oleh Novi Kurnia, "BIG DATA UNTUK ILMU SOSIAL: ANTARA METODE RISET DAN REALITAS SOSIAL." p. 332, 2021.

[17] L. H. Y. A. Oleh Solimun, Adji Achmad Rinaldo Fernandes, Nurjannah, Evitari Galu Erwinda, Rindu Hardianti, "Metodologi Penelitian: Variabel Mining berbasis Big Data dalam Pemodelan Sistem untuk mengungkap Research Novelty." 202AD.

[18] A. R. Oleh I Gede Iwan Sudipa, Pratiwi, I Putu Agus Eka Darma Udayana, Ahmad Ashril Rizal, Puji Indra Kharisma, Tutuk Indriyani, Efitra, I Made Dwi Putra Asana, Anak Agung Gede Bagus Ariana, "METODE PENELITIAN BIDANG ILMU INFORMATIKA (Teori & Referensi Berbasis Studi Kasus)." p. 148, 2023.

[19] A. A. P. W. S, L. W. Santoso, G. W. N. W. A. K. W. Rahmaddeni, A. J. W. G. E. Y. Elisawati, and R. R. W. Abdurrasyid, *Machine learning*, vol. 45, no. 13. 2017. [Online]. Available: https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA

[20] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.