

Sentimen Analysis pada Opini Masyarakat terhadap Pelayanan Publik Polres Ponorogo menggunakan Metode *Support Vector Machine*

Fawwaz Kumudani Widyadhana¹, Nanang Yudi Setiawan², Bayu Rahayudi³

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹fawwazdani@student.ub.ac.id, ²nanang@ub.ac.id, ³ubay1@ub.ac.id

Abstrak

Penelitian ini menganalisis sentimen masyarakat terhadap pelayanan publik Polres Ponorogo, Indonesia, menggunakan metode Support Vector Machine (SVM) untuk mengidentifikasi komentar negatif atau positif. Data yang digunakan berasal dari media sosial Twitter, Google Review, dan sistem Binmas Online System (BOS V2). Teknik Text Mining digunakan untuk memproses data, sementara SVM digunakan untuk melatih model dengan parameter terbaik. Preprocessing data melibatkan tokenisasi, pembersihan data, case folding, filterisasi, dan stemming. Pembobotan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengidentifikasi kata-kata kunci yang mempengaruhi sentimen. Untuk menangani ketidakseimbangan data, digunakan teknik undersampling dan k-fold cross validation digunakan untuk menguji model. Evaluasi model dilakukan dengan confusion matrix dan classification report. Hasil analisis sentimen menunjukkan akurasi rata-rata 93%, presisi 92%, recall 93%, dan F-Measure 93%. Penelitian ini memberikan kontribusi dalam menganalisis sentimen masyarakat terhadap pelayanan publik Polres Ponorogo, serta mengidentifikasi faktor-faktor yang mendasari sentimen negatif atau positif dengan menggunakan *root cause analysis*. Penelitian ini juga mengimplementasikan dashboard menggunakan perangkat lunak Tableau memungkinkan pemangku kepentingan untuk memahami hasil analisis secara visual dan interaktif. Serta model analisis sentimen yang dikembangkan menggunakan SVM dapat digunakan untuk mengklasifikasikan komentar masyarakat sebagai sentimen negatif atau positif secara otomatis, sehingga membantu Polres dalam memantau dan menanggapi sentimen masyarakat dengan lebih efektif.

Kata kunci: Analisis sentimen, pelayanan publik, media sosial, Support Vector Machine (SVM), opini masyarakat, root cause analysis, tableau, Polres Ponorogo.

Abstract

This study analyzes public sentiment towards the Ponorogo Police Public Service, Indonesia, using the Support Vector Machine (SVM) method to identify negative or positive comments. The data used comes from social media Twitter, Google Review, and the Binmas Online System (BOS V2). Text Mining techniques are used to process data, while SVM is used to train models with the best parameters. Data preprocessing involves tokenization, data cleaning, case folding, filtering, and stemming. The weighting uses the Term Frequency-Inverse Document Frequency (TF-IDF) method to identify key words that affect sentiment. To handle data imbalance, undersampling technique is used and k-fold cross validation is used to test and evaluate the model. Model evaluation is done by using the confusion matrix and classification report. Sentiment analysis results show an average accuracy of 93%, 92% precision, 93% recall, and 93% F-Measure. This study contributes to analyzing public sentiment towards the Ponorogo Polres public service, as well as identifying the factors underlying negative or positive sentiment using root cause analysis. This study also implements a dashboard using Tableau software enabling stakeholders to understand the results of the analysis visually and interactively. As well as the sentiment analysis model developed using SVM can be used to automatically classify public comments as negative or positive sentiments, thus helping Polres in monitoring and responding to public sentiments more effectively.

Keywords: Sentiment analysis, public service, social media, Support Vector Machine (SVM), public opinion, root cause analysis, tableau, Polres Ponorogo.

1. PENDAHULUAN

Pertumbuhan Teknologi Informasi telah membawa dampak yang signifikan bagi masyarakat. Keberadaan media sosial telah mengubah perilaku masyarakat secara menyeluruh, Termasuk budaya, etika, dan norma-norma yang ada. Di Indonesia, dengan populasi yang besar dan keragaman suku, ras, dan agama, terdapat potensi besar untuk perubahan sosial. Hampir semua kelompok dan usia masyarakat di Indonesia memiliki dan menggunakan media sosial sebagai sarana untuk mendapatkan dan menyebarkan informasi kepada publik (Cahyono, 2020).

Dalam menghadapi perbaikan kondisi sosial masyarakat yang semakin baik, penting untuk meningkatkan kualitas pelayanan publik. Setiap kelalaian dalam pelayanan publik dapat ditanggapi oleh masyarakat melalui gerakan atau tuntutan di media cetak maupun elektronik. Pelayanan publik yang ditujukan kepada masyarakat sangat mempengaruhi persepsi masyarakat terhadap sebuah instansi. Kualitas pelayanan yang baik akan mencerminkan bahwa instansi pemerintah tersebut baik di mata masyarakat. Sebaliknya, pelayanan publik yang kurang memuaskan tidak akan memberikan kepuasan kepada masyarakat. Oleh karena itu, penting bagi instansi pemerintah untuk memberikan pelayanan yang optimal dengan mengevaluasi berbagai aspek kualitas pelayanan yang ada, sebagai upaya dalam memenuhi kepuasan masyarakat terhadap pelayanan yang diberikan (Rini et al., 2021).

Polri, sebagai institusi negara Indonesia, memiliki peran penting dalam menjaga ketertiban dan keamanan masyarakat, menegakkan hukum, serta memberikan perlindungan, pengayoman, dan pelayanan kepada warga negara demi menjaga stabilitas dalam negeri (Peraturan Kepala Kepolisian Negara Republik Indonesia Nomor 16 Tahun 2010 Tentang Tata Cara Pelayanan Informasi Publik Di Lingkungan Kepolisian Negara Republik Indonesia, 2010).

Sebelumnya, sekelompok peneliti yang terdiri dari Syahputra, Dhimas Wida Rahayudi, Bayu Muflikhah, dan Lailil telah melakukan studi yang mengimplementasikan algoritma Support Vector Machine (SVM) dalam analisis sentimen terhadap Opini Masyarakat terhadap Pelayanan Publik Polres Ponorogo. Penelitian ini difokuskan pada analisis sentimen terkait

Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat. SVM digunakan dengan parameter-parameter terbaik yang dihasilkan, dan mencapai tingkat akurasi rata-rata sebesar 78%, presisi sebesar 79,53%, Recall sebesar 76,13%, dan F-Measure sebesar 77,36%. Tujuan dari analisis sentimen ini adalah untuk menentukan apakah sebuah komentar bersifat negatif atau positif. Dalam konteks ini, analisis sentimen digunakan untuk mengidentifikasi komentar-komentar publik terkait kebijakan penegakan ketat komunitas Twitter. Hasil analisis tersebut memberikan pemahaman bahwa opini publik dapat digunakan sebagai evaluasi terhadap kebijakan yang diimplementasikan. Berdasarkan penelitian-penelitian sebelumnya, metode Support Vector Machine (SVM) dipilih untuk melakukan analisis Opini Masyarakat terhadap Pelayanan Publik Polres Ponorogo (Syahputra et al., 2022).

2. DASAR TEORI

2.1 Google Review

Salah satu fitur yang disediakan oleh Google adalah *Google Review*, yang memungkinkan pengguna untuk memberikan ulasan tentang tempat-tempat yang tercantum dalam peta Google. Fitur ini umumnya muncul saat pengguna mencari suatu lokasi atau setelah mereka mengunjungi suatu tempat. Misalnya, setelah mengunjungi restoran tertentu, pengguna mungkin akan menerima pemberitahuan di layar ponsel mereka untuk memberikan ulasan tentang restoran tersebut. Ulasan Google juga dapat ditemukan ketika pengguna mencari suatu tempat. Ketika seseorang mencari destinasi yang diinginkan, ulasan dalam bentuk bintang dan ulasan tulisan akan muncul. Melalui ulasan tersebut, pengguna dapat melihat sejauh mana orang-orang yang telah mengunjungi destinasi tersebut puas dengan layanan atau pengalaman yang mereka terima (Haq & Rachmat, 2020).

2.2 BOS V2 (Binmas Online System)

Binmas Online System merupakan sebuah aplikasi yang dirancang khusus untuk digunakan oleh personel Bhabinkamtibmas, Satpam, Badan Usaha Jasa Pengamanan, dan juga masyarakat umum. Aplikasi ini bertujuan untuk mendukung Korbinmas Baharkam Polri dalam menjalankan fungsi pembinaan masyarakat. Tujuan utama dari aplikasi ini adalah untuk menciptakan masyarakat desa yang aman, tenteram, dan

damai. Melalui Binmas Online System, personel yang bertugas dapat melakukan berbagai kegiatan pembinaan masyarakat dengan lebih efektif dan efisien. Aplikasi ini memberikan akses dan fasilitas yang memudahkan dalam berkomunikasi, berbagi informasi, serta melakukan kegiatan pembinaan dengan masyarakat. Dengan adanya Binmas *Online System*, diharapkan dapat terjalin hubungan yang lebih baik antara pihak kepolisian dan masyarakat, sehingga tercipta lingkungan yang aman dan harmonis. (Kepolisian Republik Indonesia, 2022)

2.3 Kepolisian Resort Ponorogo

Kepolisian Resort adalah salah satu unit organisasi kepolisian yang ada di Indonesia. Kepolisian Resort merupakan bagian dari Kepolisian Daerah (Polda) yang memiliki wilayah kerja di sebuah daerah tertentu. Kepolisian Resort Ponorogo adalah Kepolisian Resort yang berada di Ponorogo, sebuah kota di Provinsi Jawa Timur, Indonesia.

Tugas utama Kepolisian Resort Ponorogo adalah menjaga keamanan, ketertiban, dan kenyamanan masyarakat di wilayah Ponorogo. Kepolisian Resort Ponorogo juga bertanggung jawab dalam memberikan pelayanan kepada masyarakat, melakukan penyidikan tindak pidana, serta memberikan bantuan dan pemeliharaan ketertiban umum.

2.4 Text Mining

Text Mining adalah sebuah proses penggalian data yang dilakukan pada teks, di mana data teks tersebut biasanya diperoleh dari dokumen. Tujuan dari *Text Mining* adalah untuk menemukan kata-kata yang dapat mewakili isi dari dokumen tersebut, sehingga memungkinkan dilakukannya analisis terhadap hubungan antar dokumen tersebut. Proses *Text Mining* melibatkan penggunaan teknik-teknik komputasional dan algoritma untuk melakukan pemrosesan dan analisis terhadap teks-teks yang ada. Dengan demikian, *Text Mining* dapat membantu mengidentifikasi pola, tema, atau informasi yang tersembunyi dalam dokumen-dokumen teks, sehingga memungkinkan pengambilan keputusan yang lebih baik dan pemahaman yang lebih dalam terhadap data teks yang ada. (Aditya, 2015)

2.5 Pre Processing

Tahap selanjutnya adalah tahap *preprocessing*,

di mana data mentah atau data yang diperoleh dari hasil crawling disiapkan agar memiliki format yang teratur dan mudah diproses dalam metode *Text Mining*. Tahap ini melibatkan menghilangkan data yang tidak diperlukan atau tidak sesuai, serta mengubahnya menjadi bentuk lain yang lebih mudah dikenali atau diterapkan dalam metode *Text Mining*. Dengan melakukan *preprocessing*, data yang awalnya beragam dalam bentuk dapat diolah menjadi data yang lebih terstruktur dan siap untuk dilakukan analisis menggunakan metode *Text Mining* (Nofiyani, 2022).

Mengenai tahapan-tahapan *pre processing* sebagai berikut:

- Tokenisasi: Memotong teks menjadi unit-unit kecil yang disebut token, seperti kata-kata atau frasa, untuk analisis selanjutnya (Bagus Sasmita et al., 2022).
- Cleaning: Menghapus bagian teks yang tidak relevan atau tidak berpengaruh, seperti username, URL, atau hashtag, agar data teks lebih fokus dan relevan (Nandini, 2018).
- Case Folding: Mengubah semua huruf dalam teks menjadi huruf kecil untuk menghilangkan variabilitas dalam penulisan huruf dan mempermudah analisis teks (Nandini, 2018).
- Filterisasi: Menghapus kata-kata yang dianggap tidak penting, seperti kata-kata penghubung atau kata-kata umum, agar fokus pada kata-kata kunci yang relevan dalam analisis teks (Rausanfitra, 2018).
- Stemming: Mengubah kata-kata menjadi bentuk dasarnya dengan menghapus awalan dan akhiran kata, menggunakan algoritma stemming seperti yang disediakan oleh pustaka Sastrawi (Rosid et al., 2020).

2.6 Analisis Sentimen

Tujuan dari analisis sentimen ini adalah untuk memahami opini atau perasaan pengguna media sosial terkait dengan topik yang sedang ramai dibicarakan. Dengan menganalisis sentimen, kita dapat mengetahui tanggapan positif dan negatif yang sering diberikan oleh pengguna media sosial terhadap topik tersebut. Dalam riset atau penelitian yang berfokus pada analisis sentimen, kita dapat memperoleh wawasan mengenai respons dan pendapat pengguna terkait dengan topik yang sedang populer dibicarakan (Bagus Sasmita et al., 2022).

2.7 Support Vector Machine

Support Vectors Machines telah menjadi alat yang mapan dalam pembelajaran mesin. Mereka bekerja dengan baik dalam praktik dan sekarang telah digunakan di berbagai aplikasi mulai dari mengenali angka tulisan tangan, hingga identifikasi wajah, kategorisasi teks, bioinformatika, dan pemasaran basis data. Dalam buku ini kami memberikan gambaran pendahuluan tentang subjek ini. Kita mulai dengan *Support Vector Machine* sederhana untuk melakukan klasifikasi biner sebelum mempertimbangkan klasifikasi multi-kelas dan belajar dengan adanya noise. Kami menunjukkan bahwa kerangka kerja ini dapat diperluas ke banyak skenario lain seperti prediksi dengan keluaran bernilai nyata, deteksi kebaruan, dan penanganan struktur keluaran kompleks seperti pohon parse. Terakhir, kami memberikan gambaran umum tentang jenis kernel utama yang digunakan dalam praktik dan cara mempelajari dan membuat prediksi dari berbagai jenis data input (Campbell & Ying, 2011).

2.8 Tableau

Tableau adalah sebuah perangkat lunak aplikasi *Business Intelligence* (BI) yang memberikan solusi visualisasi data yang interaktif. Dengan menggunakan Tableau, pengguna dapat menganalisis data secara efektif dan membantu dalam pengambilan keputusan di perusahaan. Tableau memiliki beberapa produk yang ditawarkan, antara lain Tableau Desktop (tersedia dalam edisi professional dan personal), Tableau Server, Tableau Online (untuk mendukung penggunaan oleh ribuan pengguna), dan Tableau Public. Produk-produk tersebut menyediakan beragam fitur dan fungsionalitas untuk memenuhi kebutuhan analisis data dan visualisasi yang berbeda (Silvana et al., 2017).

2.9 Penanganan Imbalance Data

Penanganan ketidakseimbangan data bertujuan untuk mengatasi distribusi yang tidak seimbang antara kelas mayoritas dan kelas minoritas dalam sebuah dataset. Ketidakseimbangan data dapat menyebabkan model Machine Learning salah dalam melakukan klasifikasi. Oleh karena itu, penanganan ketidakseimbangan data dilakukan untuk mencapai keseimbangan distribusi antara kelas mayoritas dan kelas minoritas. Dalam literatur, terdapat berbagai metode, metrik

kinerja, dan dataset yang digunakan dalam klasifikasi untuk mengatasi ketidakseimbangan data. Hasil penelitian menunjukkan variasi performa yang dihasilkan, dan terdapat beberapa teknik yang dapat digunakan untuk menangani ketidakseimbangan data. Teknik-teknik tersebut bekerja dengan cara memanipulasi distribusi data kelas minoritas dan kelas mayoritas. (Mutmainah, 2021)

2.10 Pembobotan Term Frequency-Inverse Document Frequency

Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan salah satu teknik dalam ekstraksi fitur yang digunakan untuk memberikan nilai pada setiap kata dalam dokumen latihan (data latih). Tujuan dari pembobotan ini adalah untuk menentukan seberapa penting suatu kata dalam mewakili sebuah kalimat atau dokumen. Pada pembobotan TF-IDF, skor diberikan berdasarkan frekuensi kemunculan kata dalam dokumen. Nilai TF-IDF dapat dihitung menggunakan persamaan matematis (1), (2), dan (3) (Syahputra et al., 2022).

$$W_{t,d} = \begin{cases} 1 + \log_{10} t_{f_{t,d}}, & \text{if } t_{f_{t,d}} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$idf_t = \log_{10} \left(\frac{N}{df_t} \right) \quad (2)$$

$$W_{t,d} = W_{t,d} * idf_t \quad (3)$$

Keterangan:

- $W_{t,d}$ = bobot kata dalam setiap dokumen
- $t_{f_{t,d}}$ = jumlah kemunculan kata t dalam dokumen d
- N = jumlah seluruh dokumen
- df = jumlah dokumen yang mengandung Term
- idf_t = bobot *Inverse* dari nilai df
- $W_{t,d}$ = bobot TF-IDF

2.11 Root Cause Analysis

Implementasi root cause analysis dengan forum group discussion melibatkan langkah-langkah yang sistematis untuk mengidentifikasi dan menganalisis akar penyebab suatu masalah. Dalam diskusi tersebut, anggota kelompok secara terperinci mengidentifikasi gejala, dampak, dan hasil terkait dengan masalah yang sedang dibahas. Metode analisis seperti "5 Whys" atau "Fishbone Diagram" digunakan untuk menggali penyebab-penyebab yang mungkin. Akar penyebab dievaluasi dan diprioritaskan berdasarkan tingkat pengaruh dan

kepentingannya. Solusi yang spesifik dan efektif dikembangkan untuk mengatasi akar penyebab tersebut, dengan langkah-langkah implementasi dan tanggung jawab yang ditentukan. Pemantauan dan evaluasi dilakukan untuk memastikan keberhasilan implementasi solusi, dengan perbaikan atau penyesuaian jika diperlukan. Komunikasi yang efektif dan kolaboratif antara anggota kelompok menjadi kunci dalam proses ini, dengan fokus pada tujuan mencari akar penyebab masalah.

3. METODOLOGI PENELITIAN

3.1 Tipe Penelitian

Penelitian ini menggunakan tipe penelitian non-implementatif dengan pendekatan analitik. Pendekatan analitik bertujuan untuk mengidentifikasi dan menganalisis tingkat hubungan antara bagian-bagian yang ada dalam topik penelitian dalam kondisi tertentu yang sedang diteliti. Pendekatan ini akan menghasilkan analisis yang relevan terhadap topik tersebut.

3.2 Strategi Penelitian

Tahap awal penelitian ini melibatkan pengumpulan data ulasan dari google *review* dan Twitter Polres Ponorogo. Data ulasan tersebut akan digunakan sebagai basis data untuk penelitian. Setelah data ulasan terkumpul, langkah selanjutnya adalah melakukan pelabelan data, di mana setiap ulasan akan dikategorikan sebagai positif atau negatif.

Setelah dilakukan pelabelan, data ulasan akan melalui serangkaian tahapan *preprocessing*. Tahapan ini mencakup *Case folding*, *remove punctuation*, *Tokenizing*, *stemming*, *stopwords removal*, dan *remove whitespace*. Tujuan dari tahap *preprocessing* ini adalah untuk membersihkan data dan mengubahnya ke dalam format yang lebih terstruktur dan mudah diproses.

Selanjutnya, karena data ulasan cenderung tidak seimbang (*imbalance*), dilakukan *undersampling* untuk mengatasi masalah tersebut. Setelah itu, data akan diolah menggunakan metode pembobotan TF-IDF. Pembobotan ini bertujuan untuk memberikan nilai bobot pada setiap kata dalam data latih.

Setelah tahap pembobotan, data akan masuk ke tahap klasifikasi menggunakan metode *Support Vector Machine* (SVM). Model hasil klasifikasi akan disimpan dan dimuat dalam

tampilan *dashboard* menggunakan perangkat lunak Tableau.

Selanjutnya, dilakukan Implementasi root cause analysis dengan forum group discussion melibatkan langkah-langkah yang sistematis untuk mengidentifikasi dan menganalisis akar penyebab suatu masalah.

Terakhir, dilakukan pengujian untuk melihat performa sistem yang telah dibuat. Pengujian ini akan mengevaluasi hasil klasifikasi dan mengukur seberapa baik sistem dapat mengidentifikasi sentimen dari ulasan.



Gambar 3.1 Diagram Alir Penelitian

3.3 Lokasi Penelitian

Lokasi penelitian yang digunakan adalah Polres Ponorogo. Karena tempat tersebut merupakan penunjang sarana untuk mengetahui opinion mining untuk mengetahui persepsi masyarakat terhadap pelayanan publik.

3.4 Metode Pengambilan Data

Proses pengumpulan data dari objek yang akan diproses dilakukan dalam beberapa tahap. Tujuan utama dari pengumpulan data ini adalah untuk mengumpulkan data yang akan dikaji dalam penelitian. Berikut adalah tahapan yang dilakukan untuk mengumpulkan data dari sumber-sumber yang disebutkan:

1. Pengumpulan Data dari Twitter: Data diperoleh dari situs Twitter dengan menggunakan metode manual. Penulis melakukan pencarian menggunakan kata kunci "Polres Ponorogo" dan hashtag #polresponorogo. Dalam pencarian ini, penulis mencari dokumen-dokumen yang

berbahasa Indonesia terkait topik yang dibahas. Data tersebut kemudian dikumpulkan dan digunakan dalam penelitian.

2. Pengumpulan Data dari Google Review: Pengambilan data dilakukan dari Google Review dengan menggunakan API Google Maps. API ini memungkinkan penulis untuk mengakses ulasan dan feedback dari pengguna terkait lokasi atau institusi yang spesifik, dalam hal ini Polres Ponorogo. Data yang diperoleh dari Google Review digunakan sebagai sumber informasi dalam penelitian.
3. Pengumpulan Data dari Data BOS V2 (Binmas Online System) Ponorogo: Data BOS V2 (Binmas Online System) Ponorogo diperoleh melalui website resmi <https://bos.polri.go.id/>. Pihak Humas Polres Ponorogo menyediakan data ini sebagai sumber informasi yang relevan dengan penelitian. Penulis mengakses dan mengumpulkan data dari sistem ini untuk digunakan dalam penelitian.

Melalui tahapan-tahapan tersebut, penulis berhasil mengumpulkan data dari berbagai sumber yang relevan dengan topik penelitian. Data yang diperoleh dari Twitter, Google Review, dan Data BOS V2 Ponorogo akan dikaji dan dianalisis sesuai dengan tujuan penelitian yang telah ditetapkan.

3.5 Evaluasi

Evaluasi teknik dalam penelitian ini bertujuan untuk menghasilkan rata-rata akurasi dari percobaan yang dilakukan dengan metode 5-fold cross validation. Proses ini melibatkan pembagian data menjadi kelompok-kelompok yang saling tumpang tindih untuk digunakan sebagai data latih dan data uji secara bergantian. Hal ini dilakukan untuk memastikan bahwa data latih dan uji yang digunakan mencakup variasi yang representatif.

Selain itu, penelitian ini juga menggunakan matriks kebingungan (confusion matrix) untuk mengevaluasi hasil klasifikasi. Matriks kebingungan ini memberikan informasi tentang seberapa baik hasil klasifikasi dapat membedakan kelas-kelas yang ada. Terdapat empat nilai yang dihasilkan dari matriks kebingungan, yaitu True Positive, False Positive, False Negative, dan True Negative. Dengan menggunakan matriks kebingungan ini, kita dapat menghitung berbagai metrik evaluasi

penting, seperti precision (presisi), recall (sensitivitas), dan f-measure.

Dengan demikian, evaluasi teknik dalam penelitian ini mencakup perhitungan rata-rata akurasi menggunakan 5-fold cross validation, serta analisis menggunakan matriks kebingungan untuk mendapatkan nilai-nilai precision, recall, dan f-measure.

4. PERANCANGAN DAN IMPLEMENTASI

4.1 Deskripsi Umum Penelitian

Dalam melakukan analisis sentiment terhadap skenario Polisi Resort Ponorogo yang sedang terjadi di Indonesia, dibangun sebuah sistem yang nantinya akan melakukan analisis terhadap opini tentang Polisi Resort Ponorogo yang muncul dari masyarakat menggunakan Text Mining dan akan diklasifikasi menggunakan metode *Support Vector Machine*. Nantinya pengklasifikasian akan dibagi menjadi 2 kelas yaitu sentimen positive, sentimen negative.

Selanjutnya akan diberikan visualisasi dashboard mengenai sentiment masyarakat beserta model prediksinya menggunakan streamlit dan juga analisis *root cause* mengenai opini publik Polisi Resort Ponorogo tersebut.

4.2 Pengumpulan Data

Pada penelitian ini data bersumber dari ulasan masyarakat Ponorogo yang terdapat pada *Google Review*. Data dikumpulkan dalam Bahasa Indonesia. Data yang diekstrak dari *Google Review* antara lain nama pemberi ulasan, tanggal ulasan diberikan dan teks ulasan. Teknik pengumpulan data yang digunakan adalah teknik *web scraping* dengan memanfaatkan API *Apify.com*. Proses pengumpulan data menghasilkan sebanyak 600 data ulasan *Google Review* Polres Ponorogo dengan periode waktu dari tahun 2013 hingga 2022 untuk *preview dashboard* dan 1284 untuk tes model sentiment yang berasal dari *google review* Polres Ponorogo, Madiun, Batu, dan Malang.

4.3 Penanganan Imbalance Data

Under sampling adalah metode penyeimbangan dataset dengan mengurangi jumlah sampel di kelas mayoritas (yaitu kelas dengan sampel lebih banyak). Idenya adalah untuk menurunkan sampel kelas mayoritas agar lebih seimbang dengan kelas minoritas.

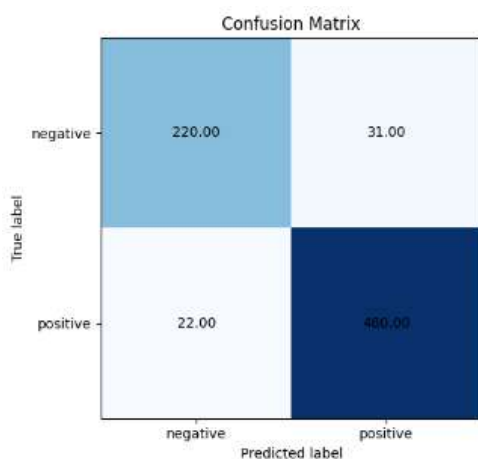
Salah satu cara untuk mengimplementasikan under sampling adalah dengan memilih secara acak subset dari sampel kelas mayoritas dan menghapus sisanya. Ini dapat dilakukan dengan menggunakan fungsi sampling acak di Python, seperti `random.sample` dari modul `random`.

Pada kasus undersampling data disini saya menggunakan rasio 2:1 untuk sentiment positif dan negatif dikarenakan data positif jauh lebih banyak daripada data negative yang jika disamakan dengan rasio 1:1 *sample* data akan terlalu kecil untuk diklasifikasi.

4.4 Skenario Pengujian

Skenario pengujian akan dilakukan dengan cara melihat pengaruh jumlah yang ada pada data. Pengujian dilakukan sebanyak 5 kali dengan perbandingan jumlah data latih dan data uji yang berbeda-beda. Pengujian tersebut akan menggunakan pengaturan dataset yang ada pada Cross Validation untuk pengujian 5-fold, pengaturan tersebut akan menjadikan seluruh dataset akan menjadi data uji sekaligus akan menjadi data latih. Proses pengujian akan menggunakan data sebanyak 753 yang terdiri dari kelas kategori *positive* dan *negative* dengan pembagian data uji sebesar 20% dan data latih sebesar 80%. Dari pengujian tersebut akan didapatkan hasil akurasi, precision, recall dan f-measure yang nantinya akan dilakukan perbandingan.

5. PENGUJIAN DAN ANALISIS



Gambar 5.1 Confusion Matrix

Berdasarkan Gambar 5.1 diatas, total confusion matrix yang diberikan, terdapat 480 data yang diprediksi benar sebagai kelas positif (True Positive), 220 data diprediksi benar sebagai kelas negatif (True Negative), 31 data

diprediksi salah sebagai kelas positif (False Positive), dan 22 data diprediksi salah sebagai kelas negatif (False Negative). Selain itu, juga terdapat average confusion matrix yang menunjukkan hasil TP, FP, TN, dan FN yang dihitung secara rata-rata.

Tabel 5.1 Hasil Pengujian K-Fold Cross Validation

Fold ke-	Presisi	Recall	F1-Score	Accuracy
1	0.90	0.90	0.90	0.90
2	0.91	0.91	0.91	0.91
3	0.87	0.87	0.87	0.87
4	0.91	0.91	0.91	0.91
5	0.89	0.89	0.89	0.89

Tabel 5.2 Classification Report on Average 5 Fold

Label	Presisi	Recall	F1-Score	Accuracy
Positif	0.939	0.956	0.947	0.924
Negatif	0.909	0.876	0.892	0.921

Berdasarkan data Tabel 5.2 diatas, dapat disimpulkan bahwa klasifikasi positif memiliki tingkat presisi yang tinggi sebesar 93.9%. Hal ini menunjukkan bahwa sebagian besar prediksi positif yang dilakukan benar-benar merupakan kasus yang positif. Selain itu, recall atau tingkat keberhasilan dalam mengidentifikasi kasus positif juga cukup tinggi dengan nilai 0.956. Hal ini menunjukkan bahwa sebagian besar kasus positif berhasil diidentifikasi dengan baik. F1-Score yang mencapai 0.947 menunjukkan keseimbangan yang baik antara presisi dan recall. Dengan akurasi sebesar 92.4%, dapat disimpulkan bahwa model secara keseluruhan memiliki kinerja yang baik dalam melakukan klasifikasi positif.

Tingkat presisi sebesar 90.9% menunjukkan bahwa sebagian besar prediksi negatif yang dilakukan oleh model benar-benar merupakan kasus negatif. Recall sebesar 87.6% menunjukkan bahwa sebagian besar kasus negatif berhasil diidentifikasi dengan baik. F1-Score sebesar 0.892 menunjukkan keseimbangan yang baik antara presisi dan recall dalam klasifikasi negatif. Dengan akurasi sebesar 92.1%, dengan pengujian menggunakan Cross Validation 5-fold, dapat disimpulkan bahwa model juga memiliki kinerja yang baik dalam melakukan klasifikasi negatif.

Secara keseluruhan, model memiliki kinerja yang baik dengan tingkat presisi, recall, dan F1-

Score yang tinggi untuk kedua klasifikasi. Dengan akurasi yang cukup tinggi, model ini dapat diandalkan dalam melakukan klasifikasi positif dan negatif.

6. KESIMPULAN DAN SARAN

6.1 Kesimpulan

Setelah melakukan pembahasan penelitian yang telah dilakukan dan berdasarkan rumusan masalah yang telah didefinisikan sebelumnya dapat ditarik 3 kesimpulan sebagai Berikut:

1. Pada model label berdasarkan total confusion matrix yang diberikan, dapat disimpulkan bahwa model memiliki kinerja yang baik dalam memprediksi kedua kelas, yaitu kelas positif dan kelas negatif. Dari 480 data yang merupakan kelas positif, sebanyak 480 data diprediksi dengan benar sebagai kelas positif (True Positive), sedangkan hanya 22 data yang diprediksi salah sebagai kelas negatif (False Negative). Hal ini menunjukkan bahwa model memiliki tingkat akurasi, presisi, dan Recall yang tinggi untuk kelas positif. Selain itu, dari 220 data yang merupakan kelas negatif, sebanyak 220 data diprediksi dengan benar sebagai kelas negatif (True Negative), dan hanya 31 data yang diprediksi salah sebagai kelas positif (False Positive). Ini menunjukkan bahwa model juga memiliki tingkat akurasi, presisi, dan Recall yang baik untuk kelas negatif. Dengan nilai akurasi sebesar 0.924, dapat disimpulkan bahwa model secara keseluruhan memiliki kinerja yang baik dalam memprediksi kelas pada dataset yang diberikan.
2. Sedangkan pada model aspek berdasarkan analisis menggunakan confusion matrix, model menunjukkan kinerja yang baik dalam memprediksi kelas "Keamanan" dengan presisi, Recall, dan F1-Score mencapai 1.000. Untuk kelas "Lingkungan Fisik", model memiliki Recall yang tinggi (1.000), tetapi presisi yang rendah (0.417), menghasilkan F1-Score sebesar 0.588. Pada kelas "Pelayanan", model memiliki presisi tinggi (0.983), Recall tinggi (0.974), dan F1-Score sebesar 0.979. Namun, pada kelas "Umum", model memiliki presisi rendah (0.182) dengan Recall tinggi (1.000), menghasilkan F1-Score sebesar 0.308. Secara keseluruhan, mean Accuracy

sebesar 0.840 menunjukkan bahwa model memiliki performa yang baik dalam memprediksi secara keseluruhan pada dataset tersebut.

3. Untuk meningkatkan efisiensi dan pelayanan dalam institusi atau organisasi, terutama dalam penegakan hukum dan pelayanan administrasi, beberapa solusi yang dapat diimplementasikan meliputi mengidentifikasi dan mengatasi bottleneck dalam proses kerja, memperbaiki koordinasi antara departemen terkait, melaksanakan pelatihan untuk meningkatkan keterampilan komunikasi interpersonal, membangun budaya yang mendorong profesionalisme dan kesantunan, meninjau serta memperbaiki prosedur pengurusan administrasi, memperketat pengawasan dan memperkuat sistem pengendalian internal, menyediakan informasi yang jelas dan mudah diakses, menggunakan teknologi untuk meningkatkan efisiensi layanan, melatih petugas dalam keterampilan komunikasi dan keramahan, meningkatkan koordinasi antara petugas, dan menciptakan lingkungan fisik yang nyaman dengan fasilitas penunjang yang memadai.
4. Untuk hasil analisis *root cause* Dalam rangka meningkatkan efisiensi dan pengelolaan waktu, perlu dilakukan identifikasi dan penanganan hambatan atau bottleneck dalam proses, koordinasi yang baik antara departemen terlibat, pelatihan petugas untuk keterampilan komunikasi interpersonal, membangun budaya profesionalisme dan kesantunan, serta memperbaiki prosedur pengurusan administrasi. Dalam hal keamanan, diperlukan pengawasan yang ketat, peningkatan kehadiran polisi, pelatihan petugas keamanan, dan saluran komunikasi yang kuat dengan masyarakat. Implementasi teknologi, penyediaan informasi yang jelas, dan evaluasi terhadap layanan dan fasilitas juga penting untuk memastikan pelayanan yang efisien dan responsif terhadap kebutuhan masyarakat.

6.2 Saran

Berikut beberapa saran yang dapat dilakukan pada penelitian dengan topik serupa berikutnya :

1. Dapat melakukan scrape secara real-time, Apify.com menyediakan beberapa layanan

yang gratis, tetapi untuk mengakses beberapa fitur yang lebih canggih dan berperforma tinggi, pengguna harus membayar biaya bulanan. Biaya yang ditetapkan tergantung pada jumlah permintaan API, jumlah pemrosesan dan volume penyimpanan yang diinginkan. Ini mungkin menjadi kendala bagi pengguna yang memerlukan banyak data dan informasi untuk dikumpulkan. Selain itu, Meskipun Apify.com dapat melakukan scrape secara terjadwal, namun tidak mampu melakukan scraping secara real-time. Hal ini bisa menjadi kendala bagi pengguna yang memerlukan informasi dalam waktu singkat.

2. Analisis sentimen dengan SVM adalah salah satu teknik untuk memprediksi sentimen pada sebuah teks, baik itu positif, negatif, atau netral. Namun, teknik pelabelan manual yang biasanya digunakan dalam analisis sentimen dengan SVM memiliki beberapa kelemahan. Salah satu kelemahannya adalah adanya subjektivitas dalam menentukan sentimen suatu teks, karena pelabelan manual tergantung pada pandangan subjektif labeler. Hal ini dapat menghasilkan hasil analisis yang kurang akurat dan konsisten. Selain itu, pelabelan manual juga membutuhkan waktu dan biaya yang cukup besar, terutama jika jumlah data yang akan dianalisis sangat banyak. Untuk mengatasi kelemahan ini, terdapat beberapa saran yang dapat dilakukan, seperti menggunakan teknik active learning, crowdsourcing, atau transfer learning. Dalam teknik active learning, SVM dilatih dengan sejumlah data awal dan diprediksi oleh model untuk menentukan data mana yang harus dilabeli oleh labeler manusia. Sedangkan, teknik crowdsourcing memanfaatkan kerja sama dari sekelompok orang untuk melakukan pelabelan data secara bersama-sama untuk meminimalkan tingkat subjektivitas dalam pelabelan manual. Terakhir, teknik transfer learning memanfaatkan model yang telah dilatih pada data yang sudah diberi label, kemudian model tersebut digunakan untuk menganalisis data yang belum diberi label. Semua teknik tersebut memiliki kelebihan dan kekurangan masing-masing, sehingga perlu dipilih dengan hati-hati sesuai dengan kebutuhan analisis sentimen yang akan dilakukan.

3. Dashboard analisis sentimen sebaiknya dibuat dalam bentuk real-time agar sistem dapat melakukan analisis dan menampilkan ulasan baru dari pelanggan secara cepat. Hal ini juga akan membantu dalam meningkatkan efektivitas penggunaan dashboard untuk keperluan evaluasi secara rutin. Dengan dashboard real-time, pengguna dapat melihat ulasan-ulasan terbaru segera setelah diterima dan dapat mengambil tindakan yang tepat dengan cepat. Selain itu, dashboard real-time juga dapat membantu meningkatkan efisiensi dalam pengumpulan data, analisis, dan evaluasi untuk pengambilan keputusan yang lebih baik. Namun, perlu diingat bahwa penggunaan dashboard analisis sentimen real-time dapat memerlukan biaya dan sumber daya yang cukup besar, sehingga perlu direncanakan dengan baik sebelum diimplementasikan.

7. DAFTAR PUSTAKA

- Aditya, B. R. (2015). Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining. *JURNAL INFOTEL - Informatika Telekomunikasi Elektronik*, 7(2), 93. <https://doi.org/10.20895/infotel.v7i2.35>
- Bagus Sasmita, A., Rahayudi, B., & Muflikhah, L. (2022). Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes. 6(3), 1208–1214. <http://j-ptiik.ub.ac.id>
- Cahyono, A. S. (2020). Pengaruh Media Sosial Terhadap Perubahan Sosial Masyarakat di Indonesia. *Publiciana*, 5(2), 202–225. <https://journal.unita.ac.id/index.php/publiciana/article/download/79/73/>
- Haq, F. U., & Rachmat, H. (2020). Penggunaan Google Review Sebagai Penilaian Kepuasan Pengunjung Dalam Pariwisata. *Journal of Sustainable Tourism Research*, 2(1), 10–12.
- Peraturan Kepala Kepolisian Negara Republik Indonesia Nomor 16 Tahun 2010 Tentang Tata Cara Pelayanan Informasi Publik di Lingkungan Kepolisian Negara Republik Indonesia, (2010).
- Mutmainah, S. (2021). Penanganan Imbalance Data Pada Klasifikasi. *SNATi*, 1, 10–16.

- Nandini, R. A. (2018). Analisis Sentimen Impor Beras 2018 Pada Twitter Menggunakan Metode Support Vector Machine dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(4), 3396–3406.
- Nofiyani, W. (2022). *Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining*. 6, 1621–1629. <https://doi.org/10.30865/mib.v6i3.4332>
- Rausanfita, A. (2018). Analisis Sentimen Twitter Menggunakan Ensemble Feature dan Metode Extreme Learning Machine (ELM) (Studi Kasus: Samsung Indonesia). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(12), 6409–6417. <http://j-ptiik.ub.ac.id>
- Rini, A., Hernimawati, & Pebriana, M. (2021). PENINGKATAN KUALITAS PELAYANAN KEPOLISIAN DALAM UPAYA MENANGANI PENGADUAN MASYARAKAT PADA POLSEKLIMA PULUH KOTA PEKANBARU. *JIANA: Jurnal Ilmu Administrasi Negara*.
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1). <https://doi.org/10.1088/1757-899X/874/1/012017>
- Silvana, M., Akbar, R., & Tifani, R. (2017). Penerapan Dashboard System di Perpustakaan Universitas Andalas Menggunakan Tableau Public. *Seminar Nasional Sains Dan Teknologi 2017, November*, 1–6.
- Syahputra, D. W., Rahayudi, B., & Muflikhah, L. (2022). Analisis Sentimen Twitter terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat menggunakan Metode Support Vector Machine. *Fakultas Ilmu Komputer, Universitas Brawijaya*, 6(3), 1067–1072.