

# Analisis Sentimen Terhadap Review Google Maps Jogja City Mall Menggunakan Algoritma Support Vector Machine

Putri Marceliana Aryanto<sup>1,\*</sup>, Rodhiyah Mardhiyyah<sup>2</sup>

<sup>1</sup>Sains dan Teknologi, Informatika, Universitas Teknologi Yogyakarta, Sleman, Indonesia

<sup>2</sup>Sains dan Teknologi, Universitas Teknologi Yogyakarta, Sleman, Indonesia

Email: <sup>1,\*</sup>putrifikri017@gmail.com, <sup>2</sup>rodhiyah.office@gmail.com

Email Penulis Korespondensi: putrifikri017@gmail.com

Submitted: 09/10/2024; Accepted: 15/11/2024; Published: 15/11/2024

**Abstrak**—Analisis sentimen bekerja dengan cara menghitung jumlah ulasan berdasarkan label yang diberikan. Ulasan-ulasan yang memiliki label tertentu nantinya akan diklasifikasikan bersama dengan ulasan lain yang berlabel sama. Penelitian ini menggunakan metode Support Vector Machine (SVM) untuk melakukan analisis terhadap ulasan-ulasan pengguna Google Maps yang telah mengunjungi Jogja City Mall di Sleman, Daerah Istimewa Yogyakarta. Dalam penelitian terdahulu, SVM terbukti beberapa kali lebih unggul dalam nilai akurasi dan ketepatan dalam klasifikasi sentimen. Model SVM untuk analisis sentimen dalam penelitian ini berhasil dibuat dengan nilai akurasi mencapai 84% dan 85% dalam pengujian menggunakan K-Folds. Dengan jumlah data tes sebesar 20% dari total 1694 data, jumlah data yang berlabel positif jauh lebih banyak daripada data yang berlabel negatif. Di mana hasil ini membuktikan bahwa skor bintang 4.6 untuk Jogja City Mall sesuai dengan ulasan yang diberikan. Analisis ketidakseimbangan data ulasan serta korelasi skor bintang dengan sentimen yang diekstraksi dapat menambah pemahaman tentang perilaku konsumen di pusat perbelanjaan, memberikan implikasi praktis bagi pengelola mall dalam meningkatkan pengalaman pelanggan.

**Kata Kunci:** Analisis Sentimen; Google Maps; Jogja City Mall; Support Vector Machine

**Abstract**—Sentiment analysis works with calculating the total of reviews based on given labels. Reviews that had specific labels would be classified with other reviews that had same labels. This study used Support Vector Machine (SVM) method for analysing of reviews of Google Maps users who had visited Jogja City Mall in Sleman, Special Region of Yogyakarta. In previous studies, SVM was proved several times superior in accuracy score and the accuracy of sentiment classification. SVM model for sentiment analysis in this study was successfully created with accuracy score was 84% and 85% in K-Folds testing. With the total of testing data was 20% out of 1694 data, the total amount of positive label was so much more than the total amount of negatif label. This means the rate of 4.6 stars for Jogja City Mall is relevant with the reviews that were given. The analysis of imbalance review data and the correlation between the stars rating and sentiment that were extracted may contribute to a deeper understanding of consumer behavior in the mall, providing practical implications for mall management in improving customer experience.

**Keywords:** Analisis Sentimen; Google Maps; Jogja City Mall; Support Vector Machine

## 1. PENDAHULUAN

Google Maps adalah website atau aplikasi pemetaan yang sudah luas digunakan oleh masyarakat. Google Maps tidak hanya menampilkan jalan dan arah, tapi juga memberikan informasi suatu tempat secara detail, termasuk fitur ulasan yang memiliki banyak informasi di dalamnya[1]. Fitur ulasan dua bagian utama, yaitu rating bintang dan ulasan dalam bentuk teks. Data teks yang ada pada fitur ulasan ini dapat dianalisis untuk menemukan informasi-informasi penting. Analisis sentimen adalah metode yang digunakan untuk menganalisis data berupa teks. Model algoritma yang digunakan juga bervariasi setiap penelitiannya. Beberapa peneliti telah mengembangkan metode analisis sentimen yang lebih kompleks dan akurat. Beberapa algoritma yang sering digunakan adalah Naive Bayes, K-Nearest Neighbors, Support Vector Machine, dan sebagainya. Termasuk Long-Short Term memories (LSTM) yang berhasil memberikan akurasi yang memuaskan[2]. Analisis sentimen bekerja dengan cara menghitung jumlah ulasan berdasarkan label yang diberikan. Ulasan-ulasan Google Maps yang memiliki label tertentu nantinya akan diklasifikasikan bersama dengan ulasan lain yang berlabel sama[3].

Algoritma-algoritma di atas memiliki kekurangan dan kelebihan masing-masing. Nilai akurasi menjadi salah satu parameter yang paling sering digunakan untuk mengetahui seberapa besar ketepatan model yang dibuat. Dalam penelitian Giovanni, et. al. yang membandingkan tiga model algoritma, Support Vector Machine (SVM) memiliki nilai akurasi dan performa yang paling baik dibanding dua algoritma lainnya, yaitu Naive Bayes dan K-NN[4]. Dalam penelitian tersebut, model transformator digabungkan dengan teknik Natural Language Processing (NLP) untuk mendapatkan konteks yang lebih luas dari teks yang dijadikan obyek studi, sehingga dapat mengidentifikasi emosi dengan lebih baik. Penelitian ini menekankan pentingnya penggunaan model yang lebih canggih untuk mendapatkan hasil analisis sentimen yang lebih akurat.

Pemilihan algoritma yang tepat akan sangat mempengaruhi hasil model yang dihasilkan. Seperti dalam studi yang dilakukan oleh Pane et. al. menunjukkan bahwa penggunaan teknik deep learning dalam analisis sentimen dapat meningkatkan akurasi klasifikasi sentimen secara signifikan[5]. Tidak hanya algoritma untuk model saja, faktor lain seperti proses cleansing dan penggunaan library kata juga memiliki pengaruh yang besar terhadap model. Penelitian yang dilakukan oleh Emi Suryati, et. al. analisis sentimen yang menggabungkan ekstraksi fitur model Word2vec Text Embedding dengan SVM[6]. Klasifikasi berhasil dilakukan dengan hasil

akhir tiap studi kasus memiliki lebih banyak ulasan positif dibanding ulasan negatif. Performa Word2vec dan SVM bekerja lebih baik pada Gojek, dengan akurasi sebesar 87%, presisi 93%, dan re-call 84%. Contoh penelitian lain yang membandingkan beberapa objek yaitu penelitian yang dilakukan oleh Khofifah, et. al. terhadap ulasan Google Maps tempat wisata pantai di Kabupaten Karawang[7].

Penelitian tersebut menggunakan ulasan dari lima objek wisata pantai pada ulasan di Google Maps untuk dilakukan analisis sentimen untuk mengetahui berapa banyak ulasan positif dan negatif. Penelitian-penelitian tersebut dapat digunakan sebagai acuan dalam pemilihan algoritma yang dirasa sesuai dengan kebutuhan. Pemilihan ini dilakukan berdasarkan persentase akurasi hasil dari perbandingan dengan algoritma lain, objek penelitian yang memiliki pola data serupa atau sama, fungsionalitas algoritma, dan kesesuaiannya dengan data yang digunakan.

Analisis sentimen tidak hanya bisa membagi data ke dalam positif dan negatif, tapi juga bisa digunakan untuk mendeteksi sifat dari suatu susunan kata. Pada penelitian yang dilakukan oleh Alita, et. al., analisis sentimen dengan menggunakan Random Forest Classifier berhasil mendeteksi sarkasme pada data-data yang dikumpulkan dari X (Twitter)[8]. Selain itu, integrasi analisis sentimen dengan sistem rekomendasi dapat memberikan rekomendasi yang lebih sesuai dengan preferensi pengguna[9].

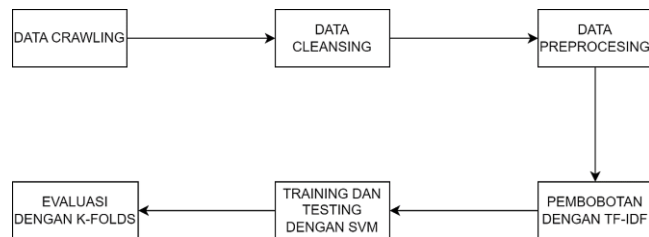
Sebagai metode pemrosesan teks, analisis sentimen bisa membagi data ke dalam kelas-kelas sesuai yang diinginkan dengan ketentuan yang dapat diatur. Dalam penelitian ini, data yang digunakan adalah data ulasan dari Jogja City Mall pada Google Maps. Total ulasan yang telah dibuat ada lebih dari 40.000 dengan rating bintang 4.6. Banyak ulasan yang hanya memberikan rating bintang, tanpa memberikan ulasan teks apapun. Hal ini dapat menimbulkan pertanyaan-pertanyaan seperti apakah rating bintang 4.6 sesuai dengan pengalaman pengunjung yang sesungguhnya.

Jumlah data yang berhasil diambil adalah sebesar 1694 yang akan diklasifikasikan ke dalam tiga kelas, yaitu positif, negatif, dan netral. Penelitian ini bertujuan untuk mengetahui jumlah ulasan pada masing-masing kelas yang akan membuktikan apakah rating bintang 4.6 benar-benar sesuai dengan ulasan yang diberikan. Algoritma yang digunakan dalam penelitian ini adalah Support Vector Machine, yang merupakan algoritma dengan kinerja rata-rata paling baik dari pada algoritma lainnya yang pernah dibandingkan[10]. Hasil penelitian ini diharapkan dapat membantu melihat kelebihan dan kekurangan dari Jogja City Mall yang berada di Kabupaten Sleman, Daerah Istimewa Yogyakarta. Data hasil penelitian bisa dimanfaatkan sebagai pertimbangan pengunjung atau pihak manajemen sebagai bahan perbaikan dan peningkatan kualitas baik fasilitas ataupun pelayanan mall.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini mengambil data ulasan pengunjung Jogja City Mall pada website Google Maps. Jumlah data yang diambil sebanyak 1694 dalam rentang 2020 sampai Maret 2024. Untuk mengolah data-data tersebut, penelitian ini dilakukan dengan beberapa tahapan, seperti pada Gambar 1.



Gambar 1. Alur Penelitian

Pada Gambar 1, terlihat proses atau tahapan yang dilakukan dalam penelitian ini. Ada enam tahapan yang dilakukan mulai dari pengambilan data atau data crawling sampai evaluasi yang bertujuan untuk mengevaluasi kinerja model dengan menggunakan K-Folds. Proses yang paling awal dilakukan adalah data crawling, di mana proses ini dilakukan untuk mengumpulkan data-data penelitian dari Google Maps dengan menggunakan browser Google Chrome dan Instant Data Scraper. Data cleansing adalah proses yang dilakukan menggunakan Google Colaboratory untuk membersihkan data penelitian dari duplikat, data kosong, dan sebagainya. Data penelitian yang telah dibersihkan kemudian dilakukan preprocessing untuk menyiapkan data untuk proses berikutnya, yaitu pembobotan dengan menggunakan TF-IDF. Setelah pembobotan selesai, model dibuat menggunakan Support Vector Machine. Setelah model selesai dibuat, evaluasi menggunakan pengujian K-Folds dilakukan untuk menguji model yang sudah dibuat. Dengan begitu, sistem yang telah dibuat dapat dilihat kinerjanya.

### 2.2 Analisis Sentimen

Analisis sentimen adalah proses untuk mengelompokkan pendapat atau opini mengenai suatu obyek ke dalam kelas positif, negatif, netral, dan kelas lainnya sesuai kebutuhan. Dalam hal ini, opini atau pendapat orang

terhadap suatu hal atau entitas tertentu seperti seperti tempat, layanan, individu, produk, masalah, dan topik dapat dianalisis dan diklasifikasikan[11]. Analisis sentimen penting untuk dilakukan dalam berbagai hal seperti penilaian produk, layanan pelanggan, dan media sosial.

### 2.3 Data Crawling

Data crawling atau web crawling merupakan proses pengumpulan data dari situs web yang kemudian akan diolah lebih lanjut. Proses ini biasa digunakan untuk berbagai tujuan seperti klasifikasi, klustering, ekstraksi informasi, dan pengambilan informasi. Crawling data secara otomatis dari internet bertujuan untuk mengindeks informasi halaman web untuk keperluan mesin pencari, pengumpulan data, dan analisis pasar[12]. Metode ini memungkinkan pengumpulan data dalam jumlah besar dengan efisiensi tinggi.

Di Indonesia, beberapa peneliti telah mengeksplorasi berbagai metode dan teknik untuk meningkatkan efektivitas dari data crawling. Misalnya, studi oleh Yuniar, et. al. yang membahas penggunaan teknik crawling berbasis API untuk mengumpulkan data dari media sosial[13]. Penelitian mereka menunjukkan bahwa metode data crawling ini tidak hanya mempercepat proses pengumpulan data tapi juga dapat meningkatkan kualitas data yang diperoleh. Selain itu, dari penelitian tersebut juga ditemukan bahwa penggunaan API memungkinkan akses ke data yang lebih terstruktur dan relevan, yang berguna dalam analisis data lebih lanjut.

Tujuan utama dari data crawling adalah mengumpulkan data dari berbagai sumber di internet dan menyatukannya dalam satu dataset. Studi oleh Karo, at. al. menyoroti pentingnya integrasi data crawling dengan teknik big data untuk menangani volume data yang lebih besar dan bervariasi[14]. Kedua peneliti tersebut telah mengembangkan sebuah framework yang memanfaatkan Hadoop untuk mengelola dan memproses data yang dikumpulkan. Hasil penelitian mereka juga menunjukkan bahwa cara ini dapat meningkatkan kecepatan dan akurasi dalam pengumpulan dan pengolahan data, yang kemudian mendukung berbagai aplikasi analitik dan penelitian yang lebih lanjut.

### 2.4 Data Preprocessing

Preprocessing data adalah proses untuk menyiapkan data yang sudah di-mining. Persiapan data dalam mengolah teks memiliki beberapa langkah, yaitu normalization, cleansing, case folding, tokenizing, filtering, dan stemming.

- a. Normalization adalah proses untuk mengubah kata-kata yang tidak baku menjadi kata-kata yang baku sehingga data serupa dan mudah untuk diproses,
- b. Cleansing adalah proses untuk membersihkan data teks dari karakter-karakter seperti tanda baca, karakter tag, dan emoji,
- c. Case folding merupakan tahap yang membuat kapitalisasi kata-kata pada data menjadi seragam, bisa menjadi uppercase atau lowercase,
- d. Tokenizing di mana pada tahap ini dilakukan proses pemisahan tiap kata dalam kalimat sehingga yang diproses berikutnya adalah kumpulan kata,
- e. Filtering adalah filterisasi terhadap kata-kata yang tidak penting atau tidak memiliki arti seperti ke, di, atau, dan sebagainya,
- f. Stemming, yaitu mengubah kata berimbuhan menjadi kata dasar.

### 2.5 Support Vector Machine

Support Vector Machine (SVM) merupakan metode prediksi yang sangat efektif untuk tugas-tugas klasifikasi dan regresi. Algoritma SVM bekerja dengan menggunakan prinsip Structural Risk Minimization (SRM) yang bertujuan untuk menemukan hyperplane optimal yang dapat memisahkan dua kelas dalam ruang input. Menurut Pratiwi dan Setiawan, SVM diterapkan dalam deteksi rumor di Twitter dan menunjukkan kinerja yang memuaskan dengan akurasi tinggi dibandingkan dengan metode lain[15]. Rumus dasar SVM untuk menemukan hyperplane optimal adalah sebagai berikut:

$$w \cdot x - b = 0 \tag{1}$$

Dalam rumus tersebut,  $w$  adalah simbol untuk bobot vektor. Sementara itu  $x$  di sini adalah input vektor, dan  $b$  sebagai bias dalam rumus dasar SVM. Keunggulan Support Vector Machine (SVM) dalam menangani data dengan kompleksitas dimensi tinggi ini membuatnya sangat efektif untuk tugas-tugas klasifikasi di mana pemilihan fitur yang tepat akan sangat mempengaruhi performa model. Arsi et. al. juga menekankan bahwa SVM dapat mengoptimalkan margin antara kelas-kelas yang berbeda, sehingga meningkatkan akurasi klasifikasi[16].

Selain itu, penelitian oleh Iskandar dan Nataliani yang membandingkan beberapa algoritma machine learning untuk klasifikasi teks juga menemukan bahwa SVM lebih sering memberikan hasil yang lebih baik dari kebanyakan algoritma lain [17]. Dari penelitian tersebut, SVM tercatat memiliki kemampuan pemrosesan yang baik pada data yang belum pernah dipelajari sebelumnya. Dalam konteks ini, SVM memanfaatkan fungsi kernel untuk memproyeksikan data ke dalam ruang berdimensi lebih tinggi agar data tersebut lebih mudah dipisahkan. Fungsi kernel yang umumnya digunakan adalah:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (2)$$

Untuk  $\phi$  adalah fungsi yang memetakan vektor input ke ruang fitur berdimensi lebih tinggi. Hal ini menunjukkan bahwa pemilihan SVM sebagai teknik klasifikasi adalah keputusan yang tepat, terutama ketika akurasi adalah prioritas utama dalam pengembangan model.

## 2.6 Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency atau TF-IDF adalah suatu metode pembobotan yang cukup sering digunakan karena memiliki nilai akurasi dan recall yang tinggi. Pembobotan TF-IDF bekerja dengan cara menghitung seberapa sering suatu kata muncul dalam dokumen. Apabila sebuah kata sering muncul dalam satu dokumen maka nilai kontribusinya semakin besar, tapi apabila kata tersebut sering muncul dalam beberapa dokumen maka kontribusinya semakin kecil. TF-IDF dibagi menjadi dua bagian, yaitu Term Frequency dan Inverse Document Frequency. TF menyatakan frekuensi kemunculan suatu term dalam dokumen, di mana semakin sering term tersebut muncul maka bobotnya semakin besar. Pada kumpulan dokumen, term dapat didistribusikan secara acak dengan IDF. Semakin sering suatu term muncul di banyak dokumen, maka nilai IDF-nya semakin kecil. Sehingga dapat dituliskan dalam rumus di bawah ini.

$$TF-IDF(tk, dj) = f(tk, dj) * \log(D/df(tk)) \quad (3)$$

Rumus di atas adalah rumus untuk pembobotan menggunakan TF-IDF. Rumus tersebut termasuk penghitungan frekuensi kemunculan kata (tk) dalam dokumen (dj). Jumlah total dokumen (D) dihitung bersama df(tk) yaitu jumlah dokumen yang mengandung kata tk.

## 2.7 Pengujian K-Folds

Pengujian K-Folds adalah metode pengujian yang menggunakan cara validasi silang yang sering digunakan dalam machine learning untuk menilai kinerja model. Metode K-Folds membagi dataset menjadi K bagian atau "folds" yang sama besar. Dalam setiap iterasi, satu fold digunakan sebagai set pengujian, sementara K-fold lainnya digunakan sebagai set pelatihan. Proses ini diulang sebanyak K kali, sehingga setiap fold berfungsi sebagai satu set pengujian. Penelitian oleh Wijianto, et. al. menunjukkan bahwa metode K-Folds dapat membuat kinerja model menjadi lebih robust dan kurang bias dibandingkan dengan metode validasi tradisional seperti hold-out [18]. Dengan memanfaatkan seluruh dataset untuk pelatihan dan pengujian, metode K-Folds ini dapat mengurangi keragaman dalam hasil evaluasi model.

Penggunaan K-Folds sangat efektif dalam mengatasi masalah overfitting, di mana model sudah cukup baik dalam memprediksi data pelatihan tetapi tidak berkinerja baik pada data uji. Menurut studi oleh Rhomadhona dan Permadi, pengujian K-Folds memberikan gambaran yang lebih akurat tentang bagaimana performa model pada data yang baru (data uji)[19]. Dalam studi kasus yang mereka lakukan, model yang divalidasi menggunakan K-Folds menunjukkan performa yang lebih stabil dan terpercaya dibandingkan dengan model yang divalidasi menggunakan metode hold-out. Penelitian tersebut membuktikan bahwa variasi hasil kinerja model berkurang secara signifikan ketika K-Folds digunakan, yang menunjukkan bahwa metode ini efektif dalam memberikan estimasi yang lebih konsisten dan dapat diandalkan.

Selain itu, pengujian K-Folds juga memaksimalkan penggunaan data latih, yang sangat penting untuk dilakukan jika data terbatas. Studi oleh Tuntun, et. al. menyoroti bahwa dalam kondisi data yang terbatas, K-Folds dapat memberikan keuntungan besar dengan memaksimalkan setiap bagian data untuk pelatihan dan pengujian[20]. Hasil penelitian mereka menunjukkan bahwa K-Folds memungkinkan model untuk belajar dari seluruh dataset, meningkatkan generalisasi model terhadap data uji. Peneliti tersebut juga mencatat bahwa K-Folds dapat diadaptasi untuk berbagai jenis model machine learning, membuat K-Folds menjadi metode validasi yang serbaguna dan cukup penting dalam penelitian dan pengembangan model machine learning. Dalam pengujian K-Folds, jumlah folds akan mempengaruhi hasil dari pengujian[21]. Oleh karena itu pemilihan jumlah folds menjadi sangat penting.

## 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, produk yang dihasilkan adalah prototipe sistem analisis sentimen dari ulasan Jogja City Mall di Google Maps menggunakan algoritma Support Vector Machine. Untuk menghasilkan prototipe ini, bahasa pemrograman yang digunakan adalah Python dengan tools Google Colab. Tahapan-tahapan untuk menghasilkan prototipe dalam penelitian ini adalah data crawling, data cleansing, data preprocessing, pembobotan menggunakan TF-IDF, pembuatan model menggunakan SVM, dan evaluasi model menggunakan K-Folds.

Pada proses data crawling, data-data ulasan dari Google Maps diambil dan dikumpulkan dalam file .xlsx menggunakan ekstensi web browser Chrome, yaitu Instant Data Scraper. Data yang berhasil di-scraper adalah

sebanyak 1694 ulasan. Data-data tersebut masih kotor dan tidak teratur. Data ulasan tersebut kemudian dibersihkan pada tahap selanjutnya.

	link	nama	guideStatus	waktuReview	isiReview
0	<a href="https://lh3.googleusercontent.com/a-/ALV-UjUGn...">https://lh3.googleusercontent.com/a-/ALV-UjUGn...</a>	Raihan Ibnu Rusydi	Local Guide · 13 ulasan · 49 foto	seminggu lalu	mall dengan arsitektur yang keren, klasik dan ...
1	<a href="https://lh3.googleusercontent.com/a-/ALV-UjU5y...">https://lh3.googleusercontent.com/a-/ALV-UjU5y...</a>	vita evayanti	Local Guide · 23 ulasan · 48 foto	sebulan lalu	Sekarang ada outlet baru Roti "O"dijogja city ...
2	<a href="https://lh3.googleusercontent.com/a-/ALV-UjXQt...">https://lh3.googleusercontent.com/a-/ALV-UjXQt...</a>	Rin	Local Guide · 546 ulasan · 2.353 foto	4 bulan lalu	Salah satu mall di jogja bagian barat yang men...
3	<a href="https://lh3.googleusercontent.com/a-/ALV-UjWnx...">https://lh3.googleusercontent.com/a-/ALV-UjWnx...</a>	Nika Woelan	Local Guide · 234 ulasan · 5.773 foto	2 bulan lalu	Yuhuuu...! Sekarang ada KKV, makin lengkap set...
4	<a href="https://lh3.googleusercontent.com/a-/ALV-UjVr5...">https://lh3.googleusercontent.com/a-/ALV-UjVr5...</a>	Princess Review	Local Guide · 46 ulasan · 85 foto	sebulan lalu	Setelah dirasain emang mallnya panas kecuali b...

Gambar 2. Tampilan Data Mentah pada Program

Gambar 2 adalah tampilan hasil kode yang untuk membersihkan data atau data cleansing. Data cleansing dilakukan agar data yang semula beragam menjadi bersih dari elemen-elemen yang tidak dibutuhkan sehingga siap digunakan untuk proses selanjutnya. Pada tahap ini data ulasan yang sudah didapat pada tahap data crawling dibersihkan dari data duplikat, kolom yang tidak terpakai atau tidak penting, data yang ber-value null, dan ulasan yang dibuat lebih dari empat tahun lalu. Pengubahan nama kolom yang tadinya berupa kode menjadi sesuai konteks juga dilakukan pada tahap ini. Selain itu, feature engineering juga dilakukan untuk memisahkan kolom waktu dibuatnya ulasan yang berupa String ke dalam dua kolom berbeda dengan bentuk String dan Integer untuk memudahkan proses filterisasi data ulasan yang dibuat lebih dari empat tahun yang lalu. Dengan begini data yang akan dipakai selanjutnya adalah data yang bersih dari duplikat dan nilai null, serta sudah yang paling update selama empat tahun terakhir. Setelah dilakukan data cleansing, jumlah data yang tersisa adalah 1361 ulasan. Beberapa data yang sudah dibersihkan dapat dilihat pada Gambar 3 di bawah.

nama	guideStatus	waktuReview	isiReview	waktuAngka	waktuString
Raihan Ibnu Rusydi	Local Guide · 13 ulasan · 49 foto	seminggu lalu	mall dengan arsitektur yang keren, klasik dan ...	1.0	minggu
vita evayanti	Local Guide · 23 ulasan · 48 foto	sebulan lalu	Sekarang ada outlet baru Roti "O"dijogja city ...	1.0	bulan
Rin	Local Guide · 546 ulasan · 2.353 foto	4 bulan lalu	Salah satu mall di jogja bagian barat yang men...	4.0	bulan
Nika Woelan	Local Guide · 234 ulasan · 5.773 foto	2 bulan lalu	Yuhuuu...! Sekarang ada KKV, makin lengkap set...	2.0	bulan
Princess Review	Local Guide · 46 ulasan · 85 foto	sebulan lalu	Setelah dirasain emang mallnya panas kecuali b...	1.0	bulan

Gambar 3. Data Bersih

Gambar 3 menampilkan hasil kode dari data yang sudah dilakukan pembersihan atau data cleansing. Sekarang data sudah tidak ada yang duplikat, tidak ada kolom kosong atau null, dan data sudah dilakukan filterisasi sehingga hanya data-data terbaru yang tidak lebih dari empat tahun saja yang akan dipakai. Data yang sudah dibersihkan kemudian diproses pada tahap selanjutnya, yaitu data preprocessing. Pada tahap preprocessing, beberapa tahap dilakukan pada data yang sudah bersih. Tahap-tahap yang dimaksud adalah normalisasi atau pengubahan kata tidak baku menjadi baku, cleansing atau pembersihan data dari karakter-karakter selain huruf seperti angka dan emoji, case folding yang menyamakan kapitalisasi data, tokenisasi yaitu pemisahan tiap kata pada data, filterisasi dari kata-kata tidak penting seperti kata hubung dan imbuhan, dan stemming yaitu pengubahan kata berimbuhan menjadi kata dasar. Semua tahap preprocessing tersebut disatukan dalam bentuk fungsi yang kemudian dipanggil untuk mengolah kolom spesifik, yaitu kolom isiReview. Hasil preprocessing disimpan dalam kolom baru yang diberi nama preprocessed\_text. Isi kolom preprocessed\_text adalah dalam bentuk List karena telah dilakukan pemisahan per kata pada tahap preprocessing tokenisasi. Pada tahap stemming, library dari Sastrawi yang menggunakan bahasa Indonesia.

```
isiReview \  
0 mall dengan arsitektur yang keren, klasik dan ...  
1 Sekarang ada outlet baru Roti "O"dijogja city ...  
2 Salah satu mall di jogja bagian barat yang men...  
3 Yuhuuu...! Sekarang ada KKV, makin lengkap set...  
4 Setelah dirasain emang mallnya panas kecuali b...  
  
preprocessed_text \  
0 [mall, arsitektur, keren, klasik, fancy, lanta...  
1 [outlet, roti, odijogja, city, mallrame, srg, ...  
2 [salah, mall, jogja, barat, tawar, desain, int...  
3 [yuhuuu, kkv, lengkap, toko, jenis, bazaar, ev...  
4 [dirasain, emang, mallnya, panas, kecuali, hal...  
  
text_string  
0 mall arsitektur keren klasik fancy lantai toil...  
1 outlet roti odijogja city mallrame srg ngantri...  
2 salah mall jogja barat tawar desain interior m...  
3 yuhuuu kkv lengkap toko jenis bazaar event ana...  
4 dirasain emang mallnya panas kecuali hall lobb...
```

Gambar 4. Data Setelah Preprocessing

Pada Gambar 4 terlihat tampilan data yang telah melalui proses preprocessing sebelumnya. Data saat ini adalah data yang sudah bersih dari elemen-elemen yang tidak diinginkan dan siap dilakukan pembobotan. Setelah preprocessing selesai dilakukan, tahap selanjutnya adalah pembobotan. Tahap pembobotan dalam penelitian ini menggunakan Term Frequency-Inverse Document Frequency atau TF-IDF. Pembobotan menggunakan TF-IDF sudah banyak dipakai dalam berbagai penelitian, di mana TF-IDF membagi bobot setiap kata berdasarkan tingkat kemunculannya sehingga kata-kata tersebut adalah kata-kata penting dan memiliki bobot kepentingan lebih dari kata-kata yang jarang muncul.

Dalam tahap pembobotan menggunakan TF-IDF ini, kolom `preprocessed_text` kembali diubah ke dalam bentuk String sehingga kolom tersebut bisa dianalisis isi dari ulasannya. Hasil perubahan bentuk data ini kemudian disimpan dalam kolom yang baru yaitu `text_string` agar tidak terjadi kesamaan nama atribut yang dapat menyebabkan kode program tidak bekerja seperti yang ada pada Gambar 5.

```
isiReview \  
0 mall dengan arsitektur yang keren, klasik dan ...  
1 Sekarang ada outlet baru Roti "O"dijogja city ...  
2 Salah satu mall di jogja bagian barat yang men...  
3 Yuhuuu...! Sekarang ada KKV, makin lengkap set...  
4 Setelah dirasain emang mallnya panas kecuali b...  
  
preprocessed_text  
0 [mall, arsitektur, keren, klasik, fancy, lanta...  
1 [outlet, roti, odijogja, city, mallrame, srg, ...  
2 [salah, mall, jogja, barat, tawar, desain, int...  
3 [yuhuuu, kkv, lengkap, toko, jenis, bazaar, ev...  
4 [dirasain, emang, mallnya, panas, kecuali, hal...
```

Gambar 5. Data Setelah Pembobotan TF-IDF

Gambar 5 adalah tampilan hasil kode dari proses pembobotan dengan TF-IDF. Perhitungan bobot TF-IDF dilakukan dengan menggunakan library yang tersedia yaitu `TfidfVectorizer` dari `sklearn`. Pembobotan hanya dilakukan pada kolom `text_string` saja yang merupakan kolom `isiReview` yang sudah di-preprocess sebelumnya. Kolom ini adalah kolom yang berisi ulasan dari pengunjung sehingga proses pembobotan hingga pembuatan model hanya akan terfokus pada kolom ini dan tidak tercampur dengan kolom lain yang tidak ada sangkut-pautnya dengan ulasan yang dianalisis.

Kemudian proses selanjutnya adalah dilakukan pelabelan secara manual ke dalam ulasan positif, ulasan negatif, dan ulasan netral. Pelabelan data sesuai dengan kata kunci yang sudah ditentukan sebelumnya, sesuai dengan label masing-masing yaitu positif dan negatif. Sementara itu, data ulasan yang tidak memiliki kata yang sama dengan daftar kata pada label positif dan negatif akan diidentifikasi sebagai ulasan yang bersifat netral. Proses pelabelan secara manual ini bertujuan untuk meningkatkan akurasi model dalam memprediksi kelas ulasan pada data uji nantinya.

Setelah pelabelan selesai dilakukan, berikutnya adalah melakukan pembuatan model dengan menggunakan Support Vector Machine (SVM). Proses pembuatan model tidak dilakukan secara manual, melainkan dengan menggunakan library yang disediakan oleh Google Colaboratory untuk klasifikasi SVM, yaitu Support Vector Classification (SVC). Penggunaan library ini dilakukan untuk mengurangi human error sehingga memungkinkan model untuk mendapatkan hasil yang maksimal dan kinerja model yang bagus. Data yang sebelumnya sudah disiapkan kemudian dibagi menjadi data latih sebesar 80% dan data uji sebesar 20%. Hasil pemodelan dapat dilihat pada Gambar 6 berikut.

Total Positive Instances: 695  
 Total Negative Instances: 71  
 Total Netral Instances: 595

**Gambar 6.** Jumlah Data Setelah Klasifikasi

Hasil klasifikasi ditunjukkan pada Gambar 6. Pada gambar tersebut, jumlah data yang masuk ke masing-masing kelas dapat terlihat. Untuk data ulasan yang berlabel positif sejumlah 695 ulasan, ulasan berlabel negatifnya sebanyak 71 ulasan, ulasan berlabel yang tidak memiliki kata-kata kunci positif ataupun negatif masuk ke kelas netral, yaitu berjumlah 595 ulasan.

Berdasarkan hasil klasifikasi yang telah dibuat, dapat terlihat adanya ketidak-seimbangan jumlah ulasan yang jarak antar datanya terlalu jauh. Antara jumlah ulasan berlabel positif dengan ulasan berlabel negatif, jarak antar datanya sangat jauh hingga ratusan. Begitupun dengan ulasan berlabel netral dan ulasan berlabel negatif. Jumlah ulasan ulasan berlabel negatif yang sangat sedikit dibanding kelas lainnya menunjukkan bahwa ulasan yang bersifat negatif tentang Jogja City Mall di Google Maps juga jumlahnya sangat sedikit.

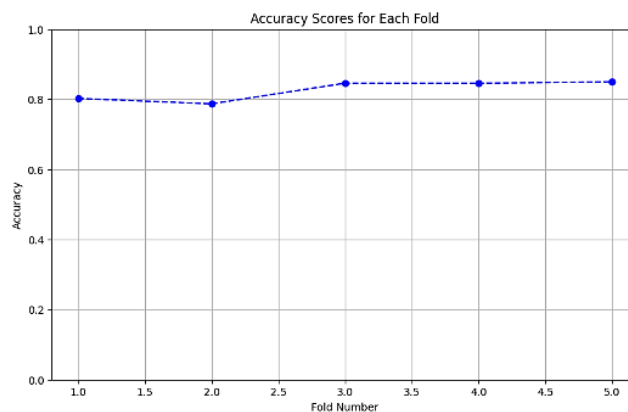
Data ulasan yang berhasil ditambah dalam penelitian ini tidak sampai dua ribu ulasan, namun jumlah ulasan untuk Jogja City Mall sebenarnya sampai angka 40.000 lebih. Hal ini terjadi karena Google Maps membatasi jumlah scroll per satu kali scroll hanya ada seribu ulasan saja yang akan ditampilkan. Sehingga menghambat proses pengambilan data dalam jumlah besar karena ulasan yang ditampilkan pada scroll yang berbeda adalah ulasan-ulasan yang kebanyakan sama atau duplikat. Oleh karena itu preprocessing untuk membersihkan data sangat penting dalam penelitian ini. Mengingat jumlah ulasan untuk Jogja City Mall pada website Google Maps sangat banyak namun rating bintang yang didapat cukup tinggi yaitu 4,6 dari lima, maka perbedaan jumlah ulasan positif dan netral dengan ulasan negatif dapat dimengerti.

	precision	recall	f1-score	support
Netral	0.78	0.95	0.85	129
negative_keywords	0.00	0.00	0.00	12
positive_keywords	0.93	0.82	0.87	132
accuracy			0.84	273
macro avg	0.57	0.59	0.57	273
weighted avg	0.82	0.84	0.82	273

**Gambar 7.** Informasi Model yang Sudah Dibuat

Gambar 7 adalah tampilan hasil laporan dari model yang telah berhasil dibuat. Dengan akurasi sebesar 84%, pemodelan menggunakan Support Vector Machine berhasil dibuat di Google Colaboratory. Rata-rata nilai presisi sebesar 82% dan rata-rata recall 84%. Namun pada Gambar 7 dapat dilihat kalau model belum bisa mengolah data ulasan-ulasan dengan label negatif karena keterbatasan jumlah data yang dipakai. Seperti pada pembahasan sebelumnya, data ulasan berlabel negatif tidak mencapai 100 ulasan, tapi data ulasan berlabel positif dan netral mencapai 500 hingga 600 ulasan.

Hasil model yang sudah terbentuk ini kemudian dilakukan pengujian dengan menggunakan K-Folds. Nilai K pada pengujian K-Folds untuk penelitian ini adalah lima folds. Hasil pengujian K-Folds pada penelitian ini menghasilkan akurasi rata-rata sebesar 83%. Nilai akurasi ini tidak jauh berbeda dengan nilai akurasi model yang sebesar 84%. Berdasarkan nilai akurasi pengujian K-Folds, dapat diartikan bahwa model memiliki tingkat akurasi yang cukup tinggi dalam memprediksi data uji. Grafik akurasi pengujian seluruh folds dari pengujian K-Folds dapat dilihat pada Gambar 8.



**Gambar 8.** Grafik Hasil Pengujian K-Folds

Pada Gambar 8 dapat terlihat ada lima nilai akurasi yang masing-masing selalu berada di atas 70%. Nilai akurasi yang paling rendah ada di 79%, sementara nilai akurasi yang paling tinggi ada di 85%. Akurasi untuk

setiap folds dengan akurasi tersebut sudah cukup bagus. Dari nilai-nilai akurasi tersebut, rata-rata akurasi dapat dicapai yaitu sebesar 84%.

Library yang digunakan untuk membangun model Support Vector Machine (SVM) adalah Support Vector Classification (SVC) dari sklearn pada Google Colaboratory. Library ini bisa bebas digunakan dan sangat mudah dalam penggunaannya. Untuk menentukan kata kunci label positif, negatif, dan netral, pelabelan dilakukan secara manual dengan menggunakan daftar kata-kata yang dinilai berbobot dan bersifat positif atau negatif. Sementara untuk data yang tidak termasuk ke dalam kelas positif ataupun negatif, akan diidentifikasi sebagai data ulasan yang bersifat netral.

Kata-kata yang diambil merupakan kata-kata yang sering muncul pada data ulasan dan dianggap penting atau dapat mewakili isi dari ulasan yang diberikan. Hal ini akan memudahkan model dalam melakukan klasifikasi agar bisa lebih mudah dimengerti oleh mesin. Kata-kata yang diambil dari data ulasan yang sudah ada di data ulasan dan dibagi hanya ke dalam dua kelas, yaitu positif dan negatif. Daftar beberapa kata-kata yang dijadikan label positif dan negatif dapat dilihat pada Tabel 1 berikut.

**Tabel 1.** Daftar Kata Kunci untuk Pelabelan

Kata Positif	Kata Negatif
Keren	Mahal
Lengkap	Kecil
Bagus	Kotor
Rame	Panas
Bersih	Jelek
Menarik	Kecewa
Rekomen	Seram

Kata-kata pada Tabel 1 tersebut dijadikan label karena muncul dalam data beberapa kali dan dibarengi dengan komentar yang menjelaskan mengenai fasilitas yang ada di Jogja City Mall. Hal ini bertujuan untuk mempermudah dalam pembagian kelas dari ulasan-ulasan lain. Misalnya untuk lima ulasan yang mengandung kata keren, lengkap, dan bagus atau salah satu dari ketiga kata itu akan diklasifikasikan ke dalam label ulasan positif. Hal ini bisa dilakukan karena pada data latih, ulasan-ulasan yang mengandung kata-kata tersebut adalah ulasan-ulasan yang bersifat positif. Begitu juga sebaliknya, ketika ada ulasan-ulasan yang mengandung kata-kata dari daftar kata kunci label, makan akan langsung diklasifikasikan ke kelas ulasan berlabel negatif. Sementara itu untuk ulasan yang tidak memiliki kata-kata dalam label maka akan ditandai sebagai ulasan yang bersifat netral karena tidak mengandung kata-kata kunci yang biasanya isi ulasannya tidak memberikan pengetahuan tentang pengalaman pengunjung selama berkunjung ke Jogja City Mall.

Pemakaian kata-kata kunci sangat membantu model dalam melakukan klasifikasi data ulasan. Pengunjung yang memberikan bintang lima terkadang hanya memberikan ulasan yang bersifat netral, sehingga tidak ada informasi yang bisa diambil untuk pengembangan pihak manajemen mall. Oleh karena itu, daftar kata-kata kunci yang digunakan juga bisa bertindak sebagai filter dari ulasan-ulasan tersebut. Dengan menggunakan kata-kata kunci tersebut klasifikasi positif dan negatif menjadi lebih terarah pada ulasan-ulasan yang memberikan informasi penting dan komentar yang membangun.

Pelabelan diterapkan pada kolom yang memuat data ulasan yang sudah dibersihkan dan dilakukan preprocessing, di mana data ulasan tersebut sudah siap untuk diproses lebih lanjut. Pemodelan menggunakan SVM kemudian dilakukan dengan menggunakan kolom hasil pembobotan TF-IDF sebagai fitur X dan kolom yang sudah diberi label sebagai target Y. Data kemudian dibagi menjadi dua bagian, yaitu data latih dan data uji. Ukuran untuk data latih adalah sebesar 80% dari total data, sementara ukuran data uji adalah sebesar 20% dari total data.

Persentase pembagian data latih dan uji tersebut sebelumnya sudah diuji coba dan dibandingkan dengan proporsi pembagian lain seperti 90:10, 70:30, 60:40, dan 50:50. Dari hasil perbandingan tersebut, didapat informasi kalau proposi pembagian data 80:20 memiliki hasil akurasi yang paling tinggi. Random state yang digunakan adalah 42. Pemodelan kemudian dilanjutkans dengan menggunakan library yang tersedia pada Google Colaboratory sehingga mendapatkan hasil prediksi dengan akurasi 84%.

Dengan akurasi tersebut, model berhasil melakukan klasifikasi data ke dalam kelas-kelas yang telah ditentukan. Hasil klasifikasi yang terbentuk untuk label positif adalah sebanyak 695 ulasan, label negatif adalah sebanyak 71 ulasan, dan label netral adalah sebanyak 595 ulasan. Dari hasil tersebut, dapat dilihat bahwa ulasan-ulasan berlabel positif memiliki jumlah yang sangat banyak, jauh lebih banyak dari ulasan berlabel negatif.

Sementara ulasan berlabel netral berada di tengah-tengah, namun masih jauh lebih banyak dari pada ulasan-ulasan yang berlabel negatif. Perbedaan jumlah label ulasan bisa ditelaah lebih jauh untuk mengetahui apakah ulasan-ulasan berlabel tersebut sudah sesuai dengan label yang didapat. Hal tersebut dapat dilakukan secara manual dengan dicek satu per satu, atau melakukan pengujian model yang sudah ada. Beberapa data ulasan yang sudah diklasifikasikan dapat dilihat pada Tabel 2 di bawah ini.

**Tabel 2.** Hasil Analisis Sentimen Beberapa Data

No.	Nama	Isi Review	Label
1	Raihan Ibnu Rusydi	mall dengan arsitektur yang keren, klasik dan fancy. tiap lantai ada toiletnya tapi harus rada effort cari kalo baru pertama kali.	Positif
2	vita evayanti	Sekarang ada outlet baru Roti "O" dijogja city mall..rame srg ngantrii.. Iq nyobain yg Igi viral cromboline " Crombo'o by Roti"O" udh 2x bli yg rasa tiramisu dan cappucino semua enak recommended dgn rasa sensasi crunchy duar,dan lumer dimulut..	Positif
3	Princess Review	Setelah dirasain emang mallnya panas kecuali bagian hall bawah setelah lobby padahal datang masih pagi banget. Kesini karena pengen coba Golden Geisha yang super rame itu sih. Toilet relatif bersih cuma aneh aja panasnya bikin gerah	Negatif
4	Nika Woelan	Yuhuuu...! Sekarang ada KKV, makin lengkap setelah ada toko sejenis. Sering ada bazaar, event dan anakku sering perform di sana, tapi ya itu sound system-nya keknya perlu diupgrade, tampil terbaru Januari 2024 sound-nya pecah. Foodcourt perlu direview dan dikurasi, soalnya ada yang rasanya biasa aja dibanding warung pinggir jalan. Lift lemot, kebersihan toilet makin menurun dibandingkan awal buka dulu. Arsitektur desainnya sih mayan, tapi kualitas pengerjaan bangunan enggak banget (jelas banget buat yg teliti melihat sambungan <sup>2</sup> , pilihan bahan, retak <sup>2</sup> , dll).	Positif
5	Harry Haryanto	Jogja city mall merupakan mall yg cukup besar dengan bangunan yg unik modern, banyak resto yg ada disana, bioskop 21, solaria, sushi station, yoshinoya dll, lokasi mudah dijangkau, parkir luas	Netral

Tabel 2 menunjukkan hasil klasifikasi menggunakan SVM yang telah dibuat. Dalam tabel tersebut, dapat dilihat lima ulasan yang diberi label masing-masing positif, negatif, dan netral. Lima data ulasan tersebut adalah ulasan-ulasan yang mendapat label sesuai dengan isi ulasannya. Klasifikasi berhasil dibuat dengan persentase akurasi cukup tinggi. Namun setelah dilakukan pengecekan ulang, masih ada beberapa data yang kurang sesuai dengan label yang didapat.

Beberapa kata juga tidak bisa diolah dengan library dari Sastrawi sehingga diperlukan proses stopwording secara manual terutama untuk kata-kata dari bahasa gaul atau bahasa asing dan bahasa daerah. Jumlah data pada ketiga kelas juga tidak seimbang sehingga perlu untuk dicek ulang. Ketidakseimbangan data bisa terjadi karena rating bintang yang tinggi yaitu 4.6 dengan total ulasan pada Google Maps sebanyak lebih dari 40.000. Hasil data yang tidak seimbang atau terlalu jauh antara label positif dan negatif dapat menguatkan validasi rating bintang yang didapat Jogja City Mall pada ulasan di Google Maps.

#### 4. KESIMPULAN

Sistem analisis sentimen yang telah dirancang dapat berjalan dengan baik dan memiliki nilai akurasi yang cukup bagus, yaitu 84% dengan akurasi pengujian K-Folds 85%. Persentase akurasi sesuai dengan harapan dan sistem bekerja dengan baik. Model yang dibuat menggunakan Support Vector Machine (SVM) juga dapat digunakan dengan baik/. Jumlah data berlabel positif diketahui lebih banyak dari jumlah data berlabel negatif dan netral, sesuai dengan skor bintang 4.6 di Google Maps. Hasil klasifikasi yang terbentuk untuk label positif adalah sebanyak 695, label negatif adalah sebanyak 71, dan label netral adalah sebanyak 595. Hasil data yang menunjukkan data berlabel positif lebih banyak dari data berlabel negatif dapat membuktikan rating bintang yang diberikan oleh pengunjung pada Google Maps dapat dipercaya atau dinyatakan valid. Sementara itu, data berlabel netral yang memiliki jumlah yang cukup banyak rata-rata memberikan rating bintang cukup baik, yaitu di rentang tiga hingga lima. Sehingga dapat disimpulkan kalau pengunjung yang memberikan ulasan yang bersifat netral kemungkinan besar sudah merasa cukup puas dengan fasilitas dan pelayanan yang didapat selama kunjungannya di Jogja City Mall. Hasil penelitian ini dapat digunakan oleh pihak manajemen Jogja City Mall guna mengetahui aspek apa saja yang perlu ditingkatkan atau dipertahankan untuk bisa menarik pengunjung lebih banyak lagi. Untuk penelitian berikutnya, bisa ditambahkan metode untuk menyeimbangkan jumlah data yang kurang seimbang tersebut apabila diperlukan. Ketidakseimbangan data pada hasil penelitian ini dapat menjadi bukti validasi tingginya rating Bintang yang didapat Jogja City Mall pada ulasan Google Maps. Semakin tinggi rating Bintang yang didapat, maka ulasan yang bersifat negatif juga semakin sedikit. Hal ini dibarengi dengan jumlah ulasan positif yang seharusnya semakin banyak.

#### REFERENCES

[1] A. Erfina dan N. R. Wardani, "ANALISIS SENTIMEN PERGURUAN TINGGI TERMEWAH DI INDONESIA MENURUT ULASAN GOOGLE MAPS MENGGUNAKAN ALGORITMA SUPPORT

- VECTOR MACHINE (SVM),” *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, vol. 5, hlm. 77–85, Jan 2022.
- [2] D. S. Amelia, “TEKS DAN ANALISIS SENTIMEN PADA CHAT GRUP WHATSAPP MENGGUNAKAN LONG SHORT TERM MEMORY (LSTM),” *Jurnal Teknologi Terkini (JTT)*, vol. 3, hlm. 1–23, Jan 2023.
- [3] J. Ipmawati, S. Saifulloh, dan K. Kusnawi, “Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, hlm. 247–256, Jan 2024, doi: 10.57152/malcom.v4i1.1066.
- [4] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, “ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI,” *Jurnal Teknoinfo*, vol. 14, no. 2, hlm. 115, Jul 2020, doi: 10.33365/jti.v14i2.679.
- [5] S. F. Pane dan J. Ramdan, “Pemodelan Machine Learning : Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter,” *Jurnal Sistem Cerdas*, vol. 5, no. 1, hlm. 12–20, Mei 2022, doi: 10.37396/jsc.v5i1.191.
- [6] E. Suryati, S. Styawati, dan A. A. Aldino, “Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding dan Algoritma Support Vector Machine (SVM),” *JURNAL TEKNOLOGI DAN SISTEM INFORMASI*, vol. 4, hlm. 96–106, Mar 2023.
- [7] W. Khofifah, D. N. Rahayu, dan A. M. Yusuf, “Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, hlm. 28–38, Jan 2022, doi: 10.35969/interkom.v16i4.192.
- [8] D. Alita dan A. R. Isnain, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *jurnal komputasi*, vol. 8, no. 2, Okt 2020, doi: 10.23960/komputasi.v8i2.2615.
- [9] Y. A. Singgalen, “Analisis Sentimen Wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel Rekomendasi Tripadvisor Menggunakan Algoritma SVM dan DT berbasis CRISP-DM,” *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 2, hlm. 367–379, Feb 2023, doi: 10.47065/josyc.v4i2.3092.
- [10] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, Tukino, dan B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter,” *Journal of Information Technology*, vol. 8, hlm. 11–18, 2023.
- [11] R. D. R. Hendrartono, I. M. Hendrati, dan W. Wardaya, “Implementasi Web Crawling Pada Website BPS Jawa Timur Pada Data Inflasi Di Jawa Timur,” *UMAT Jurnal Pengabdian Al-Shobar Untuk Masyarakat*, vol. 1, hlm. 6–13, 2024.
- [12] J. Supriyanto, D. Alita, dan A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring,” *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, hlm. 74–80, Mar 2023, doi: 10.33365/jatika.v4i1.2468.
- [13] E. Yuniar, D. S. Utsalinah, dan D. Wahyuningsih, “Implementasi Scrapping Data Untuk Sentiment Analysis Pengguna Dompot Digital dengan Menggunakan Algoritma Machine Learning,” *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 2, no. 1, hlm. 35–42, Apr 2022, doi: 10.25008/janitra.v2i1.145.
- [14] I. M. Karo Karo, J. A. Karo Karo, Y. Yuniyanto, H. Hariyanto, M. Falah, dan M. Ginting, “Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine,” *Journal of Information System Research (JOSH)*, vol. 4, no. 4, hlm. 1423–1430, Jul 2023, doi: 10.47065/josh.v4i4.3943.
- [15] Annisa Rahmaniar Dwi Pratiwi dan Erwin Budi Setiawan, “Implementation of Rumor Detection on Twitter Using the SVM Classification Method,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 5, hlm. 782–789, Okt 2020, doi: 10.29207/resti.v4i5.2031.
- [16] P. Arsi dan R. Waluyo, “Analisis Sentimen Wacana Pindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, hlm. 147, Feb 2021, doi: 10.25126/jtiik.0813944.
- [17] J. W. Iskandar dan Y. Nataliani, “Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, hlm. 1120–1126, Des 2021, doi: 10.29207/resti.v5i6.3588.
- [18] W. Wijiyanto, A. I. Pradana, S. Sopingi, dan V. Atina, “Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa,” *Jurnal Algoritma*, vol. 21, no. 1, Mei 2024, doi: 10.33364/algoritma/v.21-1.1618.
- [19] H. Rhomadhona dan J. Permadi, “Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation,” *Jurnal Sains dan Informatika*, vol. 5, no. 2, hlm. 108–117, Des 2019, doi: 10.34128/jsi.v5i2.177.
- [20] R. Tuntun, K. Kusnawati, dan K. Kusnawi, “Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, hlm. 2111, Okt 2022, doi: 10.30865/mib.v6i4.4681.

- [21] S. N. Chotimah dan A. R. Rozzaqi, “KLASIFIKASI DIAGNOSIS PENYAKIT GINJAL KRONIS DENGAN MENERAPKAN KONSEP ALGORITMA NAÏVE BAYES,” JIPETIK:Jurnal Ilmiah Penelitian Teknologi Informasi & Komputer, vol. 4, no. 1, hlm. 8–15, Jul 2023, doi: 10.26877/jipetik.v4i1.16174.