

Sentiment Analysis Reviews on Google Maps Using the Support Vector Machine Method with Selection Features of Mutual Information and Chi Square for Recommendations for Tourist Destinations in Yogyakarta City

Analisis Sentimen Ulasan di *Google Maps* Menggunakan Metode *Support Vector Machine* dengan Seleksi Fitur *Mutual Information* dan *Chi Square* Untuk Rekomendasi Destinasi Tempat Pariwisata di Kota Yogyakarta

Yuniar Rahmatika¹, Rifki Indra Perwira², Mangaras Yanu Florestiyanto³

^{1,2,3} Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

^{1*}123160168@student.upnyk.ac.id,² rifki_ip@yahoo.com,³mangaras.yanu@upnyk.ac.id

*: Penulis korespondensi (corresponding author)

Informasi Artikel

Received: Juni 2021

Revised: -

Accepted: -

Published: -

Abstract

Purpose: Implement the Support Vector Machine method to classify the assessment of tourism places into positive sentiments or negative sentiments based on review data on Google Maps, knowing the comparison of accuracy between the Support Vector Machine method with the selection of the Mutual Information feature and the Support Vector Machine method with the Chi Square feature selection, as well as providing information or recommendations regarding tourism destinations in Yogyakarta City based on positive sentiment analysis from the data. reviews on Google Maps. Design/methodology/approach: Applying Chi Square and Mutual Information feature selection to the Support Vector Machine method.

Findings/result: From the results of testing the classification model of the two methods, it shows that the Mutual Information feature selection has a better accuracy result of 90.3% with a precision of 90.2% and recall of 90.3% compared to the Chi Square feature selection which only produces an accuracy of 90.19% with a precision of 89.93.% and 90.19% recall.

Originality/value/state of the art: In this study, sentiment analysis will be carried out using data from the assessment of tourism places in the review feature on Google Maps by comparing the Chi Square and Mutual Information feature selection so that it can be seen which feature selection has better performance when applied to the Support Vector Machine method.

Abstrak

Keywords: Sentiment Analysis, Feature Selection, Support Vector Machine, Chi Square, Mutual Information

Kata kunci: Analisis Sentimen, Seleksi Fitur, Support Vector Machine, Chi Square, Mutual Information

Tujuan: Mengimplementasikan metode *Support Vector Machine* untuk mengklasifikasikan penilaian tempat pariwisata ke dalam sentimen positif atau sentimen negatif berdasarkan dari data ulasan di *Google Maps*, mengetahui perbandingan akurasi antara metode *Support Vector Machine* dengan seleksi fitur *Mutual Information* dan metode *Support Vector Machine* dengan seleksi fitur *Chi Square*, serta memberikan informasi atau rekomendasi mengenai destinasi tempat pariwisata di Kota Yogyakarta berdasarkan analisis sentimen positif dari data ulasan di *Google Maps*.

Perancangan/metode/pendekatan: Menerapkan seleksi fitur *Chi Square* dan *Mutual Information* pada metode *Support Vector Machine*.

Hasil: Dari hasil pengujian model klasifikasi dari kedua metode tersebut, memperlihatkan bahwa seleksi fitur *Mutual Information* memiliki hasil akurasi lebih baik sebesar 90.3% dengan presisi sebesar 90.2% dan *recall* sebesar 90.3% dibandingkan seleksi fitur *Chi Square* yang hanya menghasilkan akurasi sebesar 90.19% dengan presisi sebesar 89.93% dan *recall* sebesar 90.19%.

Keaslian/state of the art: Pada penelitian ini akan melakukan analisis sentimen dengan menggunakan data dari penilaian tempat pariwisata yang ada pada fitur ulasan di *Google Maps* dengan membandingkan seleksi fitur *Chi Square* dan *Mutual Information* sehingga dapat diketahui seleksi fitur mana yang memiliki performansi yang lebih baik ketika diterapkan pada metode *Support Vector Machine*.

1. Pendahuluan

Analisis sentimen adalah mengekstraksi pendapat, sentimen, evaluasi, dan emosi orang tentang suatu topik tertentu yang tertulis menggunakan teknik pemrosesan bahasa alami [1]. Analisis sentimen menggunakan algoritma *machine learning* untuk proses klasifikasinya. Algoritma *machine learning* yang sering kali digunakan antara lain *Support Vector Machine*, *Naïve Bayes*, *Decision Tree* dan *K-Nearest Neighbor*. Dalam beberapa penelitian sebelumnya, yang dilakukan oleh [2] tentang Algoritma *Naïve Bayes*, *Decision Tree*, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah, terbukti bahwa algoritma SVM lebih tinggi dibandingkan dengan algoritma *Naïve Bayes* dan *Decision Tree* dengan hasil akurasi sebesar 89.86% sedangkan algoritma *Naïve Bayes* hanya memperoleh hasil akurasi sebesar 77.29% dan algoritma *Decision Tree* sebesar 89.02%. Pada penelitian yang dilakukan oleh [3] tentang *Twitter Sentiment Analysis of the Relocation of Indonesia's Capital City*, terbukti bahwa algoritma SVM mendapatkan hasil akurasi, presisi, *recall* dan *f-measure* paling unggul sebesar 97.72%, 96.01%, 99.18%, dan 97.57% dibandingkan dengan algoritma *Naïve Bayes*, *Logistic Regression* dan *K-Nearest Neighbor*. Jadi, pada penelitian ini akan menggunakan algoritma

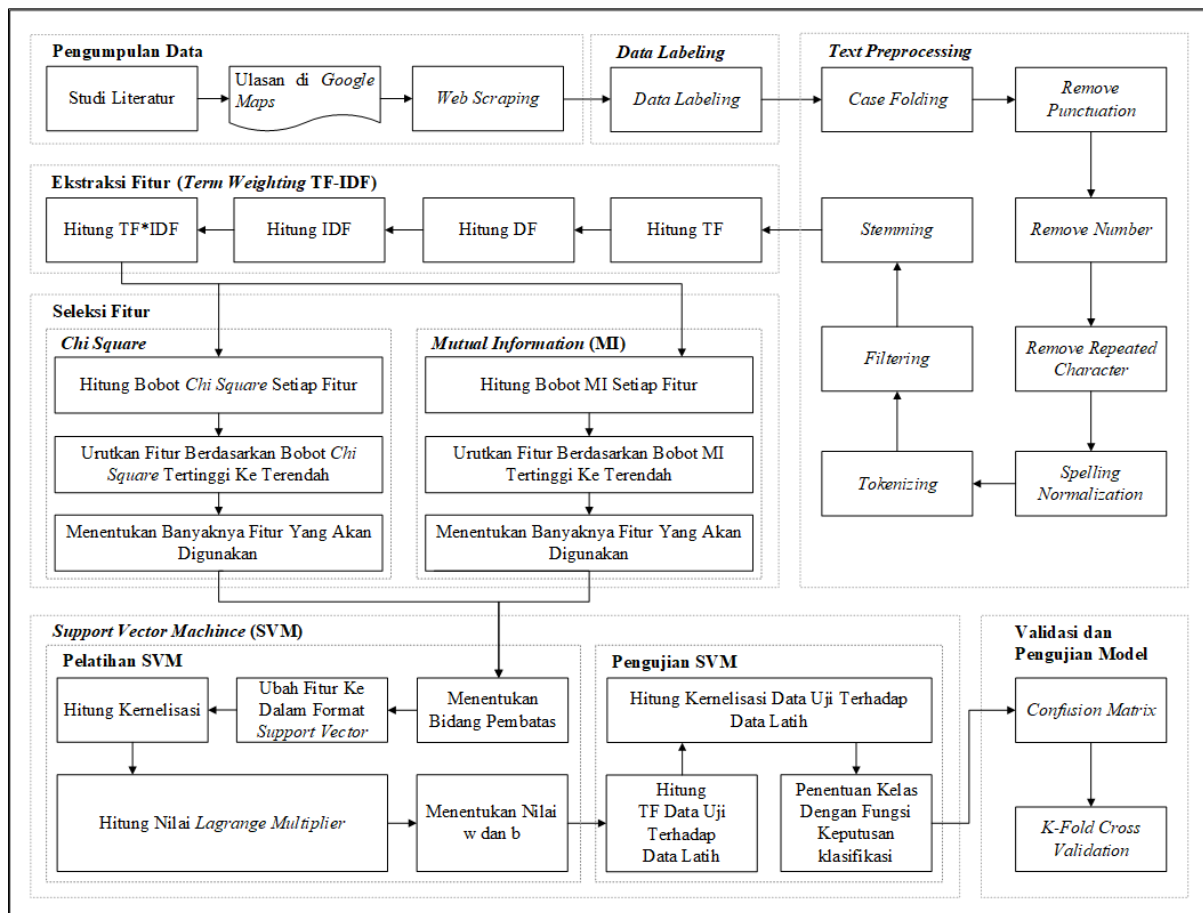
Support Vector Machine untuk klasifikasi analisis sentimen penilaian positif dan negatif pada fitur ulasan di *Google Maps* mengenai tempat pariwisata di Kota Yogyakarta. Akan tetapi, *Support Vector Machine* memiliki kekurangan dalam pemilihan fitur yang sesuai. Pemilihan fitur ini dapat mempengaruhi hasil akurasi karena tidak semua fitur sama pentingnya. Maka diperlukan seleksi fitur untuk menghilangkan data yang berlebihan, tidak relevan, dan menghilangkan data yang *noise* sehingga dapat meningkatkan hasil akurasi pada algoritma *Support Vector Machine*.

Beberapa seleksi fitur yang dapat digunakan antara lain *Mutual Information* dan *Chi Square*. Seleksi fitur *Mutual Information* merupakan sebuah metode pengukuran kuantitatif dari jumlah ketergantungan (informasi) antara dua variabel acak [4]. Berdasarkan penelitian yang dilakukan oleh [5] tentang Klasifikasi Berita Bahasa Indonesia menggunakan *Mutual Information* dan *Support Vector Machine*, terbukti bahwa metode klasifikasi *Support Vector Machine* yang menggunakan seleksi fitur *Mutual Information* memiliki hasil akurasi lebih baik dibandingkan dengan metode *Support Vector Machine* tanpa seleksi fitur *Mutual Information* dengan hasil akurasi sebesar 93.98%. Pada penelitian ini, penggunaan seleksi fitur *Mutual Information* dapat menyeleksi fitur-fitur yang tidak menggambarkan sebuah kelas, sehingga dapat meningkatkan performansi metode klasifikasi *Support Vector Machine* agar lebih cepat dan efektif karena tidak perlu menggunakan semua fitur yang ada. Seleksi fitur *Chi Square* merupakan sebuah metode yang menggunakan teori statistika untuk menguji independensi sebuah term dengan kategorinya [6]. Dalam penelitian yang dilakukan oleh [7] tentang *Improve the Accuracy of Support Vector Machine Using Chi Square Statistic and Term Frequency Inverse Document Frequency on Movie Review Sentiment Analysis*, membahas mengenai pengaruh penerapan seleksi fitur *Chi Square* dan pembobotan TF-IDF untuk meningkatkan akurasi metode klasifikasi *Support Vector Machine* dalam analisis sentimen ulasan film pada situs *Internet Movie Database (IMDB)*. Dari hasil penelitian menunjukkan bahwa penerapan seleksi fitur *Chi Square* dan Pembobotan TF-IDF dapat menaikkan hasil akurasi menjadi 80.2% dari hasil akurasi sebelumnya sebesar 68.7% yang hanya menggunakan metode *Support Vector Machine*. Berdasarkan beberapa penelitian sebelumnya, dapat ditarik kesimpulan bahwa penggunaan seleksi fitur *Mutual Information* dan *Chi Square* dapat mengatasi permasalahan pemilihan fitur pada metode *Support Vector Machine*. Keduanya, terbukti dapat meningkatkan tingkat akurasi metode *Support Vector Machine*. Akan tetapi, belum diketahui perbandingan tingkat akurasi dari seleksi fitur *Mutual Information* dan *Chi Square* pada metode *Support Vector Machine* untuk permasalahan yang sama.

Maka pada penelitian ini, akan membandingkan hasil akurasi dari metode klasifikasi *Support Vector Machine* menggunakan seleksi fitur *Mutual Information* dan metode klasifikasi *Support Vector Machine* menggunakan seleksi fitur *Chi Square* dalam menganalisis sentimen tempat pariwisata di Kota Yogyakarta pada ulasan *google maps*. Sehingga, dapat diketahui seleksi fitur mana yang memiliki hasil akurasi paling tinggi jika diterapkan pada metode klasifikasi *Support Vector Machine* untuk permasalahan yang sama. Kemudian, model dari metode klasifikasi *Support Vector Machine* dengan seleksi fitur yang memiliki akurasi paling tinggi akan digunakan untuk menentukan daftar rekomendasi destinasi tempat pariwisata di Kota Yogyakarta.

2. Metode Penelitian

Metode penelitian merupakan cara ilmiah yang digunakan untuk mencapai tujuan dan menentukan jawaban dari permasalahan yang akan diselesaikan. Metodologi penelitian yang digunakan pada penelitian ini adalah metodologi penelitian kuantitatif. Metodologi penelitian kuantitatif merupakan metodologi penelitian yang digunakan untuk menjawab permasalahan yang berkaitan dengan data berupa angka dan data kuantitatif secara lebih terperinci menggunakan analisa statistik, meneliti populasi dan sampel tertentu, instrumen penelitian, teknik pengumpulan data, dan analisis data dalam suatu proposal atau laporan penelitian sesuai dengan kaidah penulisan karya ilmiah yang disyaratkan [8]. Tahapan metodologi penelitian yang akan dilakukan antara lain pengumpulan data, *data labelling*, *text preprocessing*, ekstraksi fitur (*term weighting* TFIDF), seleksi fitur, *support vector machine*, validasi dan pengujian model. Adapun detail dari setiap tahapan metodologi penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

2.1. Pengumpulan Data

Pengumpulan data bertujuan untuk mendapatkan informasi mengenai apa saja yang dibutuhkan untuk membantu proses penelitian ini. Data yang dikumpulkan pada penelitian ini merupakan data yang dapat dipertanggung jawabkan keasliannya. Pada penelitian ini proses pengumpulan data dibagi menjadi dua tahapan yaitu studi literatur dan *web scraping*.

2.1.1 Studi Literatur

Studi literatur merupakan proses mengumpulkan jurnal, buku, artikel terkait mengenai penelitian ini. Studi literatur juga bertujuan untuk memberikan informasi mengenai metode yang digunakan dalam penelitian ini, sehingga dapat menambah pemahaman dan pengembangan dalam penelitian ini. Studi literatur yang digunakan dalam penelitian ini diambil dari berbagai sumber yang dapat dipercaya dan dipertanggung jawabkan.

2.1.2 Web Scraping

Web scraping merupakan sebuah proses mengekstraksi data dari website dan menyimpannya ke dalam sebuah *file* atau ke dalam *database* yang dilakukan melalui kode program. *Web scraping* dapat mempermudah pengambilan data karena tidak perlu menyalin data satu persatu secara manual sehingga lebih cepat. *Web scraping* pada penelitian ini dilakukan dengan mengambil data yang terdapat didalam fitur ulasan di *google maps* dengan teknik *web scraper* yang dilakukan menggunakan bahasa pemrograman *python* dengan *library selenium* dan dengan bantuan *chromedriver* untuk pengambilan data latihnya, sedangkan untuk penambahan data sentimen baru menggunakan bahasa pemrograman *php* dengan *library curl*. Data latih yang diambil sebanyak 100 komentar untuk setiap pariwisata. Data daftar pariwisata di Kota Yogyakarta yang akan dianalisis sentimen diambil dari situs *visitingjogja.com* (situs resmi Dinas Pariwisata Daerah Istimewa Yogyakarta) yaitu sebanyak 27 pariwisata. Data latih yang diambil digunakan untuk pembuatan model dan sekaligus digunakan sebagai data uji untuk menguji model yang telah dibuat. Berikut ini merupakan contoh data ulasan yang telah diambil dapat dilihat pada **Tabel 1**.

Tabel 1. Contoh Data Ulasan

Ulasan
Sebenarnya tempatnya banyak barang,, barang sejarah, tapi sayang seperti tidak terawat banyak sampah daun2 dimana,, mana, tidak ada pemandu juga untuk menjelaskan satu sama lain.
Semakin usang, kotor, dan tidak terawat,, pemerintah harus lebih memperhatikan pemeliharaan dan penataan ruangnya,,
sayang pas kesana kondisinya kotor banyak daun2 kering.. Terus harga tiketnya juga gak sesuai dengan yang tertulis di depan. MAHAL SEKALI!!!!
Tujuan wisata yg tdk boleh dilewatkan saaaat berwisata ke jogja? disini kita bisa tahu dan belajar sejarah dan adat serta budaya jogjakarta, pemandunya ramah, penjelasannya sangat jelas dan informatif
Bersih, Bagus tempat wisatanya. Pelayanan pemandunya sangat Mantap, Ramah2 Orang jualan di Luar
Tempatnya bersih,,, ketika datang langsung disambut dan dipandu mengelilingi keraton sehingga dapat menambah ilmu dan wawasan,,,
Destinasi wisata wajib kalau anda datang ke Jogja suasana dan bangunan yang klasik membuat anda seakan akan berada di zaman jawa kuno, indah banget pokoknya, pemandunya juga ramah dan sopan

2.2. Data Labelling

Data labeling merupakan tahapan memberikan label pada dataset secara manual. Semua dataset yang digunakan akan diberi label berdasarkan kelas sentimen positif atau negatif. Jumlah dataset yang akan diberi label sebanyak 2700 data ulasan. Proses *labeling* dataset ini akan dilakukan oleh 14 orang responden yang terpilih, masing-masing responden akan melabeli minimal sebanyak 100 data ulasan. Sehingga diperoleh sebanyak 2366 data ulasan merupakan kelas positif dan 334 data ulasan merupakan kelas negatif. Berikut ini merupakan contoh data ulasan yang sudah diberi label dapat dilihat pada **Tabel 2**.

Tabel 2. Contoh Data Ulasan Yang Sudah Diberi Label

Ulasan	Kelas
Sebenarnyaaa tempatnya banyak barang,, barang sejarah, tapi sayang seperti tidak terawat banyak sampah daun2 dimana,, mana, tidak ada pemandu juga untuk menjelaskan satu sama lain.	negatif
Semakin usang, kotor, dan tidak terawat,, pemerintah harus lebih memperhatikan pemeliharaan dan penataan ruangnya,,	negatif
sayang pas kesana kondisinya kotor banyak daun2 kering.. Terus harga tiketnya juga gak sesuai dengan yang tertulis di depan. MAHAL SEKALI!!!!	negatif
Tujuan wisata yg tdk boleh dilewatkan saaaat berwisata ke jogja? disini kita bisa tahu dan belajar sejarah dan adat serta budaya jogjakarta, pemandunya ramah, penjelasannya sangat jelas dan informatif	positif
Bersih, Bagus tempat wisatanya. Pelayanan pemandunya sangat Mantap, Ramah2 Orang jualan di Luar	positif
Tempatnya bersih,, ketika datang langsung disambut dan dipandu mengelilingi keraton sehingga dapat menambah ilmu dan wawasan,,	positif
Destinasi wisata wajib kalau anda datang ke Jogja suasana dan bangunan yang klasik membuat anda seakan akan berada di zaman jawa kuno, indah banget pokoknya, pemandunya juga ramah dan sopan	positif

2.3. Text Preprocessing

Text preprocessing merupakan proses awal dalam *text mining* sebelum masuk ke proses klasifikasi. *Text preprocessing* merupakan proses yang bertujuan untuk mengubah dokumen teks semi-terstruktur atau tidak terstruktur menjadi dokumen teks terstruktur, dengan cara menyamakan struktur serta mengurangi dimensi yang melebar. Pada penelitian ini *text preprocessing* dibagi menjadi delapan tahapan antara lain *case folding*, *remove punctuation*, *remove number*, *remove repeated characters*, *spelling normalization*, *tokenizing*, *filtering* dan *stemming*.

2.3.1 Case Folding

Case folding merupakan proses mengubah huruf *uppercase* (huruf kapital) menjadi huruf *lowercase* (huruf kecil). *Case folding* diperlukan karena tidak semua data dalam media sosial konsisten untuk menggunakan huruf kecil. Berikut ini merupakan contoh hasil *case folding* yang dapat dilihat pada **Tabel 3**.

Tabel 3. Contoh Hasil *Case Folding*

No	Ulasan
D1	sebenarnyaaa tempatnya banyak barang,, barang sejarah, tapi sayang seperti tidak terawat banyak sampah daun2 dimana,, mana, tidak ada pemandu juga untuk menjelaskan satu sama lain.

2.3.2 Remove Punctuation

Remove punctuation merupakan proses untuk menghilangkan tanda baca pada teks dokumen yang bertujuan untuk membersihkan elemen yang tidak berpengaruh terhadap hasil klasifikasi. Elemen tanda baca yang dihapus antara lain tanda seru, titik, titik koma, *slash*, tanda petik satu, tanda petik dua (kutipan), strip, tanda tanya, tanda kurung, dan sebagainya. Berikut ini merupakan contoh hasil *remove punctuation* yang dapat dilihat pada **Tabel 4**.

Tabel 4. Contoh Hasil *Remove Punctuation*

No	Ulasan
D1	sebenarnyaaa tempatnya banyak barang barang sejarah tapi sayang seperti tidak terawat banyak sampah daun2 dimana mana tidak ada pemandu juga untuk menjelaskan satu sama lain

2.3.3 Remove Number

Remove number adalah proses penghapusan karakter angka di dalam teks dokumen. *Remove number* dilakukan karena angka dianggap delimiter atau tidak berpengaruh terhadap hasil klasifikasi bahkan dapat mengganggu proses klasifikasi. Berikut ini merupakan contoh hasil *remove number* yang dapat dilihat pada **Tabel 5**.

Tabel 5. Contoh Hasil *Remove Number*

No	Ulasan
D1	sebenarnyaaa tempatnya banyak barang barang sejarah tapi sayang seperti tidak terawat banyak sampah daun dimana mana tidak ada pemandu juga untuk menjelaskan satu sama lain

2.3.4 Remove Repeated Characters

Remove repeated characters adalah proses penghapus karakter yang berulang didalam teks dokumen. Karakter berulang yang dihapus adalah huruf yang memiliki lebih dari dua karakter, dan akan disisakan dua karakter saja. Berikut ini merupakan contoh hasil *remove repeated characters* yang dapat dilihat pada **Tabel 6**.

Tabel 6. Contoh Hasil *Remove Repeated Characters*

No	Ulasan
D1	sebenarnya tempatnya banyak barang barang sejarah tapi sayang seperti tidak terawat banyak sampah daun dimana mana tidak ada pemandu juga untuk menjelaskan satu sama lain

2.3.5 Spelling Normalization

Spelling normalization merupakan proses perbaikan kata yang salah eja maupun singkatan (kata tidak baku) menjadi bentuk kata aslinya (kata baku). *Spelling normalization* dilakukan untuk menghindari dimensi kata yang melebar, karena kesalahan ketik pada penulisan atau singkatan, padahal kata tersebut memiliki arti yang sama. Jika tidak dilakukan *spelling normalization* maka kata tersebut akan dianggap sebuah entitas yang berbeda pada penyusunan matriks. Pada penelitian ini menggunakan daftar kamus *spelling normalization* dari penelitian [9] dan ditambah secara manual dari pengamatan data latih yang digunakan pada penelitian ini. Berikut ini merupakan contoh hasil *spelling normalization* yang dapat dilihat pada **Tabel 7**.

Tabel 7. Contoh Hasil *Spelling Normalization*

No	Ulasan
D1	sebenarnya tempatnya banyak barang barang sejarah tapi sayang seperti tidak terawat banyak sampah daun dimana mana tidak ada pemandu juga untuk menjelaskan satu sama lain

2.3.6 Tokenizing

Tokenizing adalah proses pemotongan string input menjadi beberapa kata berdasarkan tiap kata yang menyusunnya. Pada proses ini, data teks dokumen akan dipisah menjadi tiap kata berdasarkan spasi atau karakter lain sehingga menghasilkan token. Berikut ini merupakan contoh hasil tokenizing yang dapat dilihat pada **Tabel 8**.

Tabel 8. Contoh Hasil *Tokenizing*

No	Ulasan
D1	['sebenarnya', 'tempatnya', 'banyak', 'barang', 'barang', 'sejarah', 'tapi', 'sayang', 'seperti', 'tidak', 'terawat', 'banyak', 'sampah', 'daun', 'dimana', 'mana', 'tidak', 'ada', 'pemandu', 'juga', 'untuk', 'menjelaskan', 'satu', 'sama', 'lain']

2.3.7 Filtering

Filtering adalah proses mengambil kata-kata penting yang dihasilkan oleh token. Pada penelitian ini akan menggunakan algoritma *stopword removal* untuk membuang katakata yang tidak penting yang didapat dari hasil token seperti “yang”, “di”, “dari”, “untuk”, “ke” yang

memiliki frekuensi kemunculan yang tinggi dalam sebuah teks dokumen. Pada penelitian ini menggunakan kamus *filtering* atau *stopword removal* dari penelitian [10]. Berikut ini merupakan contoh hasil *filtering* yang dapat dilihat pada **Tabel 9**.

Tabel 9. Contoh Hasil *Filtering*

No	Ulasan
D1	['tempatnya', 'barang', 'barang', 'sejarah', 'sayang', 'terawat', 'sampah', 'daun', 'dimana', 'pemandu']

2.3.8 Stemming

Stemming adalah tahapan untuk mengubah kata atau token hasil *filtering* menjadi kata dasarnya dengan menghilangkan imbuhan pada kata tersebut. Algoritma *stemming* yang digunakan pada penelitian adalah algoritma *stemming* Nazief dan Adriani. Berikut ini merupakan contoh hasil *stemming* yang dapat dilihat pada **Tabel 10**.

Tabel 10. Contoh Hasil *Stemming*

No	Ulasan
D1	['tempat', 'barang', 'barang', 'sejarah', 'sayang', 'awat', 'sampah', 'daun', 'mana', 'pandu']

2.4. Ekstraksi Fitur (*Term Weighting TF-IDF*)

Pada tahapan ini akan dilakukan pembobotan fitur atau kata dengan metode TF-IDF. Tahapan ini dilakukan setelah dilakukannya *text preprocessing*. Pada tahap ini, setiap fitur akan dihitung frekuensi kemunculannya dalam dokumen yang disebut *term frequency* (TF). Kemudian, akan dihitung nilai *inverse document frequency* (IDF), dan yang terakhir akan dihitung nilai *term frequency-invers document frequency* (TF-IDF). TF-IDF merupakan penggabungan dari perhitungan TF murni dengan IDF. Berikut merupakan persamaan menghitung nilai IDF yang dapat dilihat pada **persamaan (1)**:

$$IDF_j = \log \left(\frac{D}{df_j} \right) \quad (1)$$

Berikut ini merupakan contoh hasil perhitungan nilai IDF yang dapat dilihat pada **Tabel 11**.

Tabel 11. Contoh Hasil Perhitungan Nilai IDF

No	Fitur	TF							DF	IDF
		D1	D2	D3	D4	D5	D6	D7		
1	tempat	1	0	0	0	0	1	0	2	0.544068
2	barang	2	0	0	0	0	0	0	2	0.544068
3	sejarah	1	0	0	1	0	0	0	2	0.544068
...
61	sopan	0	0	0	0	0	0	1	1	0.845098

Berikut merupakan persamaan menghitung nilai TF-IDF yang dapat dilihat pada **persamaan (2)** dan **persamaan (3)**:

$$W_{ij} = tf_{ij} \times idf_j \quad (2)$$

$$W_{ij} = tf_{ij} \times \log \left(\frac{D}{df_i} \right) \quad (3)$$

Berikut ini merupakan contoh hasil perhitungan nilai TF-IDF yang dapat dilihat pada **Tabel 12**.

Tabel 12. Contoh Hasil Perhitungan Nilai TF-IDF

No	Fitur	TF-IDF						
		D1	D2	D3	D4	D5	D6	D7
1	tempat	0.544068	0	0	0	0	0.544068	0
2	barang	0.544068	0	0	0	0	0	0

3	sejarah	0.544068	0	0	0.544068	0	0	0
...
61	sopan	0	0	0	0	0	0	0.845098

2.5. Seleksi Fitur

Seleksi fitur merupakan tahapan pemilihan fitur-fitur yang memiliki pengaruh paling tinggi terhadap pembangunan model klasifikasi metode SVM dan menghilangkan fitur yang kurang berpengaruh atau tidak berpotensi menjadi indeks. Penerapan tahapan seleksi fitur ini dapat mempercepat proses klasifikasi pada metode SVM karena fitur yang digunakan lebih sedikit. Tahapan proses seleksi fitur antara lain menghitung nilai bobot setiap fitur, lalu mengurutkan fitur dari bobot tertinggi ke terendah, kemudian menentukan berapa banyak fitur yang akan digunakan untuk menghilangkan fitur pengganggu dalam klasifikasi. Pada penelitian ini setiap fitur akan dihitung bobot nilainya menggunakan metode seleksi fitur *mutual information* dan *chi square*. Kedua metode tersebut kemudian akan dibandingkan hasil akurasi, sehingga dapat diketahui metode seleksi fitur mana yang lebih baik dalam meningkatkan hasil akurasi dari klasifikasi metode SVM.

2.5.1 Chi Square

Chi Square merupakan seleksi fitur yang menguji keterikatan hubungan atau pengaruh sebuah *term* dengan kategorinya. Pada metode seleksi fitur *chi square* pertama menempatkan fitur atau *term* yang akan dihitung nilai jumlah kemunculan katanya menggunakan tabel kontingensi untuk kelas positif dan kelas negatif. Kemudian hasil nilai *chi square* fitur dari kedua kelas tersebut akan dibandingkan, dan akan dipilih nilai *chi square* maksimal dari fitur tersebut. Kemudian setelah semua fitur telah diketahui nilai bobot *chi square*-nya, maka semua fitur akan diurutkan berdasarkan nilai bobot *chi square* tertinggi ke terendah. Kemudian, menentukan banyaknya fitur yang akan digunakan untuk proses klasifikasi. Berikut ini merupakan contoh menghitung nilai bobot *chi square* dari fitur “tempat”, Pertama, menghitung nilai kemunculan fitur “tempat” untuk kelas positif dan negatif, dengan menempatkan nilai jumlah kemunculan *term* atau kata menggunakan tabel kontingensi seperti berikut ini :

Tabel 13. Kontingensi Nilai Jumlah Kemunculan *Term* Tempat Untuk Kelas Positif.

c = Positif t = “tempat”	ec = 1	ec = 0
et = 1	1	1
et = 0	3	2

Tabel 14. Kontingensi Nilai Jumlah Kemunculan *Term* Tempat Untuk Kelas Negatif.

c = Negatif t = “tempat”	ec = 1	ec = 0
et = 1	1	1
et = 0	2	3

Kedua, menghitung nilai bobot *chi square* fitur “tempat” dengan menggunakan **persamaan (4)** [11] pada kelas positif dan negatif sebagai berikut :

$$X^2(t, c) = \frac{N \times (A \times D - C \times B)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (4)$$

$$X^2(\text{tempat}, \text{positif}) = \frac{7 \times (1 \times 2 - 3 \times 1)^2}{(1 + 3)(1 + 2)(1 + 1)(3 + 2)}$$

$$\begin{aligned}
 &= \frac{7 \times (2 - 3)^2}{(4)(3)(2)(5)} \\
 &= \frac{7}{120} \\
 &= 0.058333 \\
 X^2(\text{tempat, negatif}) &= \frac{7 \times (1 \times 3 - 2 \times 1)^2}{(1 + 2)(1 + 3)(1 + 1)(2 + 3)} \\
 &= \frac{7 \times (3 - 2)^2}{(3)(4)(2)(5)} \\
 &= \frac{7}{120} \\
 &= 0.058333
 \end{aligned}$$

$$X^2(\text{tempat}) = \text{Max}(X^2(\text{tempat, positif}), X^2(\text{tempat, negatif})) = 0.058333$$

Dari perhitungan di atas, diperoleh nilai bobot *chi square* fitur “tempat” adalah 0.058333. Berikut ini merupakan hasil perhitungan nilai bobot *chi square* untuk semua fitur yang dapat dilihat pada **Tabel 15**.

Tabel 15. Hasil Perhitungan Nilai Bobot *Chi Square* Untuk Semua Fitur

No	Fitur	Kelas = Positif				Kelas = Negatif				Bobot <i>Chi Square</i>		Max Bobot <i>Chi Square</i>
		A	B	C	D	A	B	C	D	Positif	Negatif	
1	tempat	1	1	3	2	1	1	2	3	0.058333	0.058333	0.058333333
2	barang	0	1	4	2	1	0	2	4	1.555556	1.555556	1.55555556
3	sejarah	1	1	3	2	1	1	2	3	0.058333	0.058333	0.058333333
...
61	sopan	1	0	3	3	0	1	3	3	0.875	0.875	0.875

Setelah semua fitur diketahui hasil bobot *chi square*-nya, langkah selanjutnya mengurutkan fitur berdasarkan bobot *chi square* tertinggi ke terendah. Berikut ini merupakan hasil mengurutkan fitur berdasarkan bobot *chi square* tertinggi ke terendah yang dapat dilihat pada **Tabel 16**.

Tabel 16. Hasil Mengurutkan Bobot *Chi Square* Tertinggi Ke Terendah

No	Fitur	Max Bobot <i>Chi Square</i>
1	sayang	7
2	wisata	3.9375
3	ramah	3.9375
...
61	sejarah	0.058333333

Setelah semua fitur diurutkan berdasarkan bobot *chi square*-nya, langkah selanjutnya menentukan banyaknya fitur yang akan digunakan. Dimisalkan hanya menggunakan 40 fitur dengan bobot *chi square* tertinggi, maka perubahan tabel TF-IDF setelah melalui proses seleksi fitur *chi square* dapat dilihat pada **Tabel 17**.

Tabel 17. Hasil Perubahan Tabel TF-IDF Setelah Seleksi Fitur *Chi Square*

No	Fitur	TF-IDF						
		D1	D2	D3	D4	D5	D6	D7
1	sayang	0.367977	0.367977	0.367977	0	0	0	0
2	wisata	0	0	0	0.243038	0.243038	0	0.243038
3	ramah	0	0	0	0.367977	0.367977	0	0.367977
...
40	langsung	0	0	0	0	0	0.845098	0

2.5.2 Mutual Information

Mutual Information merupakan seleksi fitur yang mengukur atau menghitung jumlah ketergantungan (informasi) antara dua variabel acak. Pada metode seleksi fitur *mutual information* pertama menempatkan fitur atau *term* yang akan dihitung nilai jumlah kemunculan katanya menggunakan tabel *confusion matrix* untuk kelas positif dan kelas negatif. Kemudian hasil nilai *mutual information* fitur dari kedua kelas tersebut akan dibandingkan, dan akan dipilih nilai *mutual information* terbesar dari fitur tersebut. Kemudian setelah semua fitur telah diketahui nilai bobot *mutual information*-nya, maka semua fitur akan diurutkan berdasarkan nilai bobot *mutual Information* yang paling tinggi ke yang paling rendah. Kemudian, menentukan banyaknya fitur yang akan digunakan untuk proses klasifikasi. Berikut ini merupakan contoh menghitung nilai bobot *mutual information* (MI) dari fitur “tempat”, Pertama, menghitung nilai kemunculan fitur “tempat” untuk kelas positif dan negatif, dengan menempatkan nilai jumlah kemunculan *term* atau kata menggunakan tabel *confusion matrix* seperti berikut ini :

Tabel 18. *Confusion Matrix* Nilai Jumlah Kemunculan *Term* Tempat Untuk Kelas Positif.

c = Positif t = “tempat”	ec = 1	ec = 0
et = 1	1	1
et = 0	3	2

Tabel 19. *Confusion Matrix* Nilai Jumlah Kemunculan *Term* Tempat Untuk Kelas Negatif.

c = Negatif t = “tempat”	ec = 1	ec = 0
et = 1	1	1
et = 0	2	3

Kedua, menghitung nilai bobot MI fitur “tempat” dengan menggunakan **persamaan (5)** [5] pada kelas positif dan negatif sebagai berikut :

$$I(U, C) = \frac{N_{11}}{N} \log_2 \frac{N \cdot N_{11}}{N_1 \cdot N_1} + \frac{N_{01}}{N} \log_2 \frac{N \cdot N_{01}}{N_0 \cdot N_1} + \frac{N_{10}}{N} \log_2 \frac{N \cdot N_{10}}{N_1 \cdot N_0} + \frac{N_{00}}{N} \log_2 \frac{N \cdot N_{00}}{N_0 \cdot N_0} \quad (5)$$

1. Fitur “tempat” kelas positif

$$\begin{aligned} I(\text{tempat, positif}) &= \frac{1}{7} \log_2 \frac{7 \times 1}{(1+1) \times (3+1)} + \frac{3}{7} \log_2 \frac{7 \times 3}{(3+2) \times (3+1)} + \frac{1}{7} \log_2 \frac{7 \times 1}{(1+1) \times (1+2)} + \frac{2}{7} \log_2 \frac{7 \times 2}{(3+2) \times (1+2)} \\ &= \frac{1}{7} \log_2 \frac{7}{2 \times 4} + \frac{3}{7} \log_2 \frac{21}{5 \times 4} + \frac{1}{7} \log_2 \frac{7}{2 \times 3} + \frac{2}{7} \log_2 \frac{14}{5 \times 3} \\ &= \frac{1}{7} \log_2 \frac{7}{8} + \frac{3}{7} \log_2 \frac{21}{20} + \frac{1}{7} \log_2 \frac{7}{6} + \frac{2}{7} \log_2 \frac{14}{15} \\ &= 0.005978 \end{aligned}$$

2. Fitur “tempat” kelas negatif

$$\begin{aligned} I(\text{tempat, negatif}) &= \frac{1}{7} \log_2 \frac{7 \times 1}{(1+1) \times (2+1)} + \frac{2}{7} \log_2 \frac{7 \times 2}{(2+3) \times (2+1)} + \frac{1}{7} \log_2 \frac{7 \times 1}{(1+1) \times (1+3)} + \frac{3}{7} \log_2 \frac{7 \times 3}{(2+3) \times (1+3)} \\ &= \frac{1}{7} \log_2 \frac{7}{2 \times 3} + \frac{2}{7} \log_2 \frac{14}{5 \times 3} + \frac{1}{7} \log_2 \frac{7}{2 \times 4} + \frac{3}{7} \log_2 \frac{21}{5 \times 4} \\ &= \frac{1}{7} \log_2 \frac{7}{6} + \frac{2}{7} \log_2 \frac{14}{15} + \frac{1}{7} \log_2 \frac{7}{8} + \frac{3}{7} \log_2 \frac{21}{20} \end{aligned}$$

$$= 0.005978$$

Dari perhitungan di atas, diperoleh nilai bobot *mutual information* fitur “tempat” pada kelas positif 0.005978 dan pada kelas negatif 0.005978. Maka hasil nilai bobot *mutual information* terbesar dari fitur “tempat” kelas positif dan negatif adalah 0.005978. Berikut ini merupakan hasil perhitungan nilai bobot *mutual information* untuk semua fitur pada kelas positif dan negatif yang dapat dilihat pada **Tabel 20**.

Tabel 20. Hasil Perhitungan Nilai Bobot *Mutual Information* Untuk Semua Fitur

No	Fitur	Kelas = positif				Kelas = negatif				Bobot MI		Max Bobot MI
		N11	N10	N01	N00	N11	N10	N01	N00	Positif	Negatif	
1	tempat	1	1	3	2	1	1	2	3	0.005978	0.005978	0.005978
2	barang	0	1	4	2	1	0	2	4	0.198117	0.198117	0.198117
3	sejarah	1	1	3	2	1	1	2	3	0.005978	0.005978	0.005978
...
61	sopan	1	0	3	3	0	1	3	3	0.005978	0.005978	0.005978

Setelah semua fitur diketahui hasil bobot *mutual information*-nya, langkah selanjutnya mengurutkan fitur berdasarkan bobot *mutual information* tertinggi ke terendah. Berikut ini merupakan hasil mengurutkan fitur berdasarkan bobot *mutual information* tertinggi ke terendah yang dapat dilihat pada **Tabel 21**.

Tabel 21. Hasil Mengurutkan Bobot *Mutual Information* Tertinggi Ke Terendah

No	Fitur	Bobot MI
1	sayang	0.985228
2	wisata	0.521641
3	ramah	0.521641
...
61	sejarah	0.005978

Setelah semua fitur diurutkan berdasarkan bobot *mutual information*-nya, langkah selanjutnya menentukan banyaknya fitur yang akan digunakan. Dimisalkan hanya menggunakan 40 fitur dengan bobot *Mutual Information* tertinggi, maka perubahan tabel TF-IDF setelah melalui proses seleksi fitur *mutual information* dapat dilihat pada **Tabel 22**.

Tabel 22. Hasil Perubahan Tabel TF-IDF Setelah Seleksi Fitur *Mutual Information*

No	Fitur	TF-IDF						
		D1	D2	D3	D4	D5	D6	D7
1	sayang	0.367977	0.367977	0.367977	0	0	0	0
2	wisata	0	0	0	0.243038	0.243038	0	0.243038
3	ramah	0	0	0	0.367977	0.367977	0	0.367977
...
40	langsung	0	0	0	0	0	0.845098	0

2.6. Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan *machine learning* (*supervised learning*) yang memprediksi kelas berdasarkan model atau pola dari hasil proses *training* [12]. Pada SVM dokumen yang akan diklasifikasi diubah terlebih dahulu menjadi *support vector*. Biasanya teks dokumen diubah menjadi *vector* multidimensional tf-idf [13]. Klasifikasi pada SVM dilakukan dengan mencari *hyperplane* atau kernel maksimal yang memisahkan dua buah kelas. Dua buah kelas tersebut antara lain kelas sentimen positif (berlabel +1) dengan dan kelas sentimen negatif (berlabel -1). Pada SVM dua buah kelas dipisahkan oleh bidang pembatas yang sejajar atau kernel, sehingga diperoleh [14]:

1. Diasumsikan kedua kelas -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* atau bidang pembatas berdimensi d , yang didefinisikan :

$$w \cdot x + b = 0 \quad (6)$$

2. *Pattern* x_i yang termasuk ke dalam kelas -1 (kelas negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan sebagai berikut :

$$w \cdot x_i + b \leq -1 \quad (7)$$

3. *Pattern* x_i yang termasuk ke dalam kelas 1 (kelas positif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan sebagai berikut :

$$w \cdot x_i + b \geq 1 \quad (8)$$

4. Fungsi untuk mencari nilai *Lagrange Multiplier* α sebagai berikut :

$$\max_{\alpha} L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (9)$$

dengan syarat : $\alpha_i \geq 0$ ($i = 1, 2, \dots, n$) dan $\sum_{i=1}^n \alpha_i y_i = 0$

Setelah diperoleh nilai dari fungsi di atas, maka selanjutnya mencari nilai w , nilai bias, dan mencari fungsi keputusan klasifikasi *sign* ($f(x)$) dengan persamaan atau *formula* sebagai berikut [15] :

5. Persamaan untuk mencari nilai w sebagai berikut :

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (10)$$

6. Persamaan untuk mencari nilai bias sebagai berikut :

$$b = -\frac{1}{2} (x_i^+ \cdot w + x_i^- \cdot w) \quad (11)$$

7. Persamaan untuk mencari fungsi keputusan klasifikasi *sign*($f(x)$) sebagai berikut :

$$f(x) = \text{sign} \left[\sum_{i=1}^m \alpha_i y_i K(x, x_i) + b \right] \quad (12)$$

Fungsi *sign*() merupakan fungsi normalisasi, jika nilai x di dalam fungsi *sign*() lebih dari 0 atau > 0 maka fungsi tersebut memberikan nilai 1 (kelas positif). Jika nilai x dalam fungsi *sign*() kurang dari 0 atau < 0 maka fungsi tersebut memberikan nilai -1(kelas negatif).

Berikut ini merupakan beberapa fungsi kernel yang sering digunakan dalam metode *Support Vector Machine* [16] :

1. Linear Kernel

$$K(x_i, x_j) = x_i^T x_j \quad (13)$$

Dengan x_i merupakan data latih dan x_j adalah data uji.

2. *Polynomial Kernel*

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^p, \quad \gamma > 0 \quad (14)$$

Dengan x_i merupakan data latih, x_j adalah data uji, p adalah derajat *polynomial*, γ adalah *scalling* parameter dari jarak *Euclidean* dan r adalah nilai konstanta.

3. *Radial Basis Function* (RBF) Kernel

$$K(x_i, x_j) = \exp \left(-\gamma \|x_i - x_j\|^2 \right), \quad \gamma > 0 \quad (15)$$

Dengan x_i merupakan data latih, x_j adalah data uji, dan γ adalah *scalling* parameter dari jarak *Euclidean*.

4. *Sigmoid* (Tangen Hiperbolik) Kernel

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (16)$$

Dengan x_i merupakan data latih, x_j adalah data uji, γ adalah *scaling* parameter dari jarak *Euclidean* dan r adalah nilai konstanta.

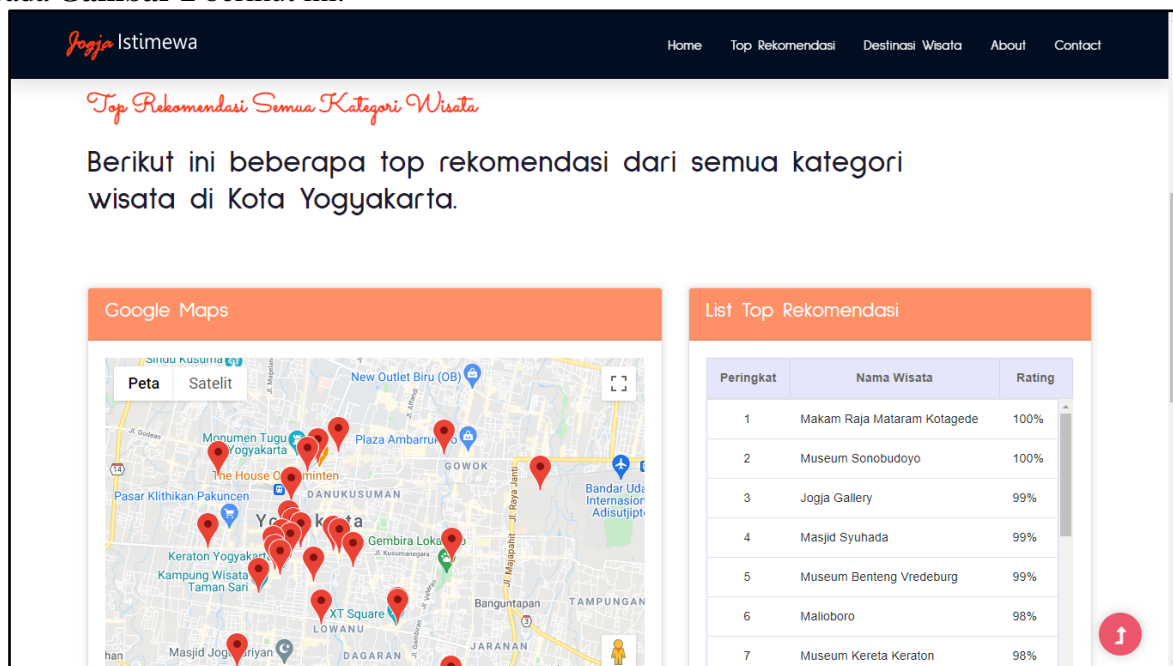
Pada tahap klasifikasi dengan metode *support vector machine* pada penelitian ini akan dibagi menjadi dua bagian yaitu klasifikasi metode *support vector machine* dengan seleksi fitur *chi square* dan klasifikasi metode *support vector machine* dengan seleksi fitur *mutual information*. Pada setiap bagian akan dibagi lagi menjadi dua tahapan, yaitu tahap pelatihan metode *support vector machine* dan tahap pengujian metode *support vector machine*.

3. Hasil dan Pembahasan

Pada bagian ini akan menjelaskan mengenai hasil, pengujian dan pembahasan dari implementasi perancangan sistem analisis sentimen. Hasil dari implementasi perancangan analisis sentimen kemudian akan melalui tahap pengujian model dan pengujian sistem. Pada pengujian model akan menggunakan *confusion matrix* dan *k-fold cross validation*. Setelah melalui tahap pengujian maka akan diperoleh pembahasan mengenai apakah sistem yang telah dibuat sudah menjawab rumusan masalah pada penelitian ini.

3.1. Hasil Penelitian

Pada bagian ini akan membahas mengenai hasil implementasi analisis sentimen ulasan di *google maps* untuk rekomendasi destinasi tempat pariwisata di Kota Yogyakarta yang dapat dilihat pada **Gambar 2** berikut ini.



Gambar 2. Implementasi *User Interface* Pada Halaman Top Rekomendasi

Gambar diatas merupakan hasil top rekomendasi destinasi tempat pariwisata yang ada di Kota Yogyakarta berdasarkan hasil pembobotan sentimen positif masing-masing wisata yang telah diurutkan dari peringkat tertinggi ke terendah.

3.2. Pengujian Model

Pada bagian ini akan menjelaskan mengenai hasil pengujian pada penelitian yang telah dilakukan. Pengujian model merupakan pengujian untuk melihat performansi dari model yang telah dibangun oleh metode yang digunakan pada penelitian ini. Pada penelitian ini menggunakan pengujian model *confusion matrix* dan *k-fold cross validation*.

3.2.1 Confusion matrix

Confusion matrix merupakan pengujian model untuk mengukur perbandingan performansi dari metode *support vector machine* dengan seleksi fitur *chi square* dan metode *support vector machine* dengan seleksi fitur *mutual information*. *Confusion matrix* memiliki empat kemungkinan yang merepresentasikan hasil dari proses klasifikasi, diantaranya *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) yang akan menghasilkan nilai akurasi, presisi, dan *recall* dari model yang dibangun. Pada penelitian ini sebanyak 2700 data yang digunakan akan dibagi menjadi data latih dan data uji secara bergantian untuk setiap iterasinya, dimana sebanyak 385 atau 386 data digunakan sebagai data uji dan 2315 atau 2314 data sisanya digunakan sebagai data latih. Berikut ini merupakan hasil perhitungan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) dari metode *support vector machine* (SVM) dengan seleksi fitur *chi square* yang dapat dilihat pada **Tabel 23**.

Tabel 23. *Confusion Matrix* Metode SVM dengan seleksi fitur *Chi Square*

K-Fold	Confusion Matrix			
	TP	FN	FP	TN
1	338	3	34	11
2	352	5	19	10
3	311	2	57	16
4	338	2	36	10
5	328	0	41	17
6	341	0	30	14
7	344	2	34	5

Berikut ini merupakan hasil perhitungan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) dari metode *support vector machine* (SVM) dengan seleksi fitur *mutual information* yang dapat dilihat pada **Tabel 4.2**.

Tabel 24. *Confusion Matrix* Metode SVM dengan seleksi fitur *Mutual Information*

K-Fold	Confusion Matrix			
	TP	FN	FP	TN
1	338	3	34	11
2	352	5	18	11
3	311	2	56	17
4	338	2	36	10
5	328	0	41	17
6	341	0	30	14
7	345	1	34	5

3.2.2 K-fold cross validation

K-fold cross validation merupakan validasi untuk mengukur rata-rata keberhasilan dari model yang dibangun. Pada penelitian ini akan menggunakan iterasi $k = 7$ dimana sebanyak 2700 dataset yang digunakan akan dibagi menjadi 7 bagian. Pada iterasi pertama dataset ke-1 sampai ke-385 akan digunakan sebagai data uji sedangkan dataset ke-386 sampai 2700 akan digunakan sebagai data latih. Pembagian dataset menjadi data latih dan data uji ini akan dilakukan secara berulang sampai iterasi ke tujuh (sampai semua data pernah menjadi data uji dan data latih).

Hasil dari *k-fold cross validation* ini akan memvalidasi perbandingan dari tingkat akurasi, presisi dan *recall* dari model klasifikasi yang dibangun oleh metode *support vector machine* dengan seleksi fitur *chi square* dan metode *support vector machine* dengan seleksi fitur *mutual information* untuk setiap iterasinya, sehingga dapat diketahui seleksi fitur mana yang lebih baik dalam meningkatkan performansi metode *support vector machine*. Berikut ini merupakan hasil perhitungan akurasi, presisi dan *recall* untuk setiap *k-fold cross validation* dari metode *support vector machine* (SVM) dengan seleksi fitur *chi square* yang dapat dilihat pada **Tabel 4.3**.

Tabel 25. Pengujian *K-Fold Cross Validation* Metode SVM dengan seleksi fitur *Chi Square*

<i>K-Fold</i>	Akurasi	Presisi	<i>Recall</i>
1	90.41%	89.43%	90.41%
2	93.78%	92.76%	93.78%
3	84.72%	85.34%	84.72%
4	90.16%	89.54%	90.16%
5	89.38%	90.56%	89.38%
6	92.21%	92.84%	92.21%
7	90.65%	89.02%	90.65%
Rata-Rata	90.19%	89.93%	90.19%

Berikut ini merupakan hasil perhitungan akurasi, presisi dan *recall* untuk setiap *k-fold cross validation* dari metode *support vector machine* (SVM) dengan seleksi fitur *mutual information* yang dapat dilihat pada **Tabel 4.4**.

Tabel 26. Pengujian *K-Fold Cross Validation* Metode SVM dengan seleksi fitur *Mutual Information*

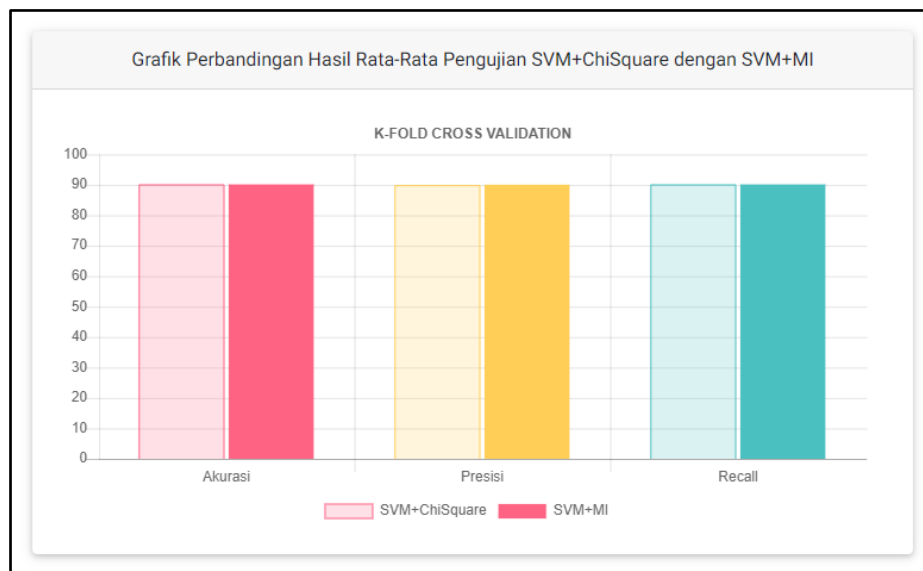
<i>K-Fold</i>	Akurasi	Presisi	<i>Recall</i>
1	90.41%	89.43%	90.41%
2	94.04%	93.15%	94.04%
3	84.97%	85.64%	84.97%
4	90.16%	89.54%	90.16%
5	89.38%	90.56%	89.38%
6	92.21%	92.84%	92.21%
7	90.91%	90.25%	90.91%
Rata-Rata	90.3%	90.2%	90.3%

Berikut ini merupakan perbandingan hasil perhitungan rata-rata akurasi, presisi dan *recall* untuk setiap *k-fold cross validation* dari metode *support vector machine* (SVM) dengan seleksi fitur *chi square* dan metode *support vector machine* (SVM) dengan seleksi fitur *mutual information* yang dapat dilihat pada **Tabel 4.5**.

Tabel 4.5. Perbandingan Hasil Rata-Rata Pengujian *K-Fold Cross Validation*

Akurasi		Presisi		<i>Recall</i>	
SVM+ <i>Chi Square</i>	SVM+MI	SVM+ <i>Chi Square</i>	SVM+MI	SVM+ <i>Chi Square</i>	SVM+MI
90.19%	90.3%	89.93%	90.2%	90.19%	90.3%

Berikut merupakan grafik perbandingan hasil rata-rata pengujian model antara metode *support vector machine* dengan seleksi fitur *chi square* dan metode *support vector machine* dengan seleksi fitur *mutual information*, dengan menggunakan *confusion matrix* dan validasi *k-fold cross validation* yang dapat dilihat pada **Gambar 3**.



Gambar 3. Grafik Perbandingan Hasil Rata-Rata Pengujian Model

3.3. Pembahasan

Berdasarkan hasil dari implementasi dan pengujian pada penelitian ini, didapatkan bahwa analisis sentimen ulasan dari *google maps* telah menghasilkan rekomendasi tempat pariwisata yang ada di Kota Yogyakarta, dengan jumlah tempat pariwisata sebanyak 27 tempat pariwisata, yang telah diurutkan dari tempat wisata dengan peringkat tertinggi ke terendah berdasarkan dari perhitungan pembobotan banyaknya jumlah sentimen positif pada masing-masing tempat pariwisata dari hasil klasifikasi.

Dari pengujian model yang telah dilakukan, tingkat performansi dari model klasifikasi *support vector machine* dengan seleksi fitur *mutual information* yang telah dibangun menghasilkan rata-rata akurasi sebesar 90.3%, rata-rata presisi sebesar 90.2% dan rata-rata *recall* sebesar 90.3%. Sedangkan, tingkat performansi dari model klasifikasi *support vector machine* dengan seleksi fitur *chi square* yang telah dibangun menghasilkan rata-rata akurasi sebesar 90.19%, rata-rata presisi sebesar 89.93%, dan rata-rata *recall* sebesar 90.19%. Dari hasil pengujian model, dapat dilihat bahwa penerapan seleksi fitur *mutual information* pada metode *support vector machine* memiliki tingkat akurasi yang lebih unggul sebesar 0.11% dibandingkan dengan penerapan seleksi fitur *chi square* pada metode *support vector machine* ketika menggunakan jumlah fitur sebanyak 2400 fitur.

Dari tingkat performansi yang dihasilkan maka model klasifikasi yang dibangun oleh metode *support vector machine* dengan seleksi fitur *mutual information* yang digunakan untuk analisis sentimen ulasan di *google maps* sehingga

4. Kesimpulan dan Saran

4.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka diperoleh hasil kesimpulan sebagai berikut :

1. Pada penelitian ini berhasil membangun sistem analisis sentimen ulasan di *Google Maps* menggunakan metode *Support Vector Machine* dengan membandingkan seleksi fitur

Mutual Information dan *Chi Square* sehingga diperoleh rekomendasi destinasi tempat pariwisata di Kota Yogyakarta dengan total tempat pariwisata sebanyak 27 yang telah diurutkan berdasarkan pembobotan jumlah sentimen positif wisata tersebut.

2. Hasil pengujian model klasifikasi dari metode *Support Vector Machine* dengan seleksi fitur *Mutual Information* dan metode *Support Vector Machine* dengan seleksi fitur *Chi Square* dimana menggunakan 2700 data ulasan yang dibagi menjadi data latih dan data uji menggunakan metode *k-fold cross validation* dengan $k=7$, memperlihatkan bahwa seleksi fitur *Mutual Information* memiliki hasil akurasi lebih baik sebesar 90.3% dengan presisi sebesar 90.2% dan *recall* sebesar 90.3% dibandingkan seleksi fitur *Chi Square* yang hanya menghasilkan akurasi sebesar 90.19% dengan presisi sebesar 89.93% dan *recall* sebesar 90.19%.

4.2 Saran

Adapun beberapa saran untuk penelitian selanjutnya berdasarkan kekurangan dan keterbatasan pada penelitian ini, antara lain :

1. Data tempat pariwisata yang akan dianalisis diperluas tidak hanya di Kota Yogyakarta saja.
2. Dapat menambahkan klasifikasi untuk kategori dari ulasan tidak hanya mengklasifikasikan sentimen pada ulasan saja. Misalnya mengkategorikan ulasan berdasarkan topik yang diulas seperti kebersihan, aksesibilitas, daya tarik, fasilitas dan sebagainya.
3. Dapat menambahkan data training ulasan tempat pariwisata dari sumber lain tidak hanya dari *google maps* saja, sehingga model klasifikasi yang dibangun dapat melakukan klasifikasi ulasan tempat wisata dengan lebih baik.

Daftar Pustaka

1. Sentiaji, A. R., & Bachtiar, A. M. (2014). Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik. *Bandung: Universitas Komputer Indonesia*.
2. Nurajijah, N., & Riana, D. (2019). Algoritma naïve Bayes, decision tree, dan SVM untuk klasifikasi persetujuan pembiayaan nasabah koperasi syariah. *Jurnal Teknologi dan Sistem Komputer*, 7(2), 77–82.
3. Sutoyo, E., & Almaarif, A. (2020). Twitter sentiment analysis of the relocation of Indonesia's capital city. *Bulletin of Electrical Engineering and Informatics*, 9(4), 1620–1630.
4. Sulaiman, M. A., & Labadin, J. (2015). Feature selection based on mutual information. In *2015 9th International Conference on IT in Asia (CITA)* (pp. 1–6). IEEE.
5. Irham, L. G., Adiwijaya, A., & Wisesty, U. N. (2019). Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 3(4), 284–292.
6. Ling, J., Kencana, I. P. E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92–99.

7. Ukhti Ikhsani Larasati, I. U., Much Aziz Muslim, I. U., Riza Arifudin, I. U., & Alamsyah, I. U. (2019). Improve the Accuracy of Support Vector Machine Using Chi Square Statistic and Term Frequency Inverse Document Frequency on Movie Review Sentiment Analysis. *Scientific Journal of Informatics*, 6(1), 138–149.
8. Harahap, N., & Lubis, S. D. (2019). Metodologi Penelitian Kuantitatif.
9. Noehilasari, S. A. A. (2014). PERIODISASI DAN PROSES PEMBENTUKAN KOSAKATA BAHASA GAUL TAHUN 1990-2012. *undefined*. Retrieved from <https://www.semanticscholar.org/search?q=PERIODISASI%20DAN%20PROSES%20PEMBENTUKAN%20KOSAKATA%20BAHASA%20GAUL%20TAHUN%201990-2012&sort=relevance>
10. Tala, F. (2003). A study of stemming effects on information retrieval in Bahasa Indonesia.
11. Thaseen, I. S., & Kumar, C. A. (2017). Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *Journal of King Saud University-Computer and Information Sciences*, 29(4), 462–472.
12. Maulana, R., & Redjeki, S. (2017). ANALISIS SENTIMEN PENGGUNA TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE BERBASIS CLOUD COMPUTING. *Jurnal TAM (Technology Acceptance Model)*, 6, 23–28.
13. Ipmawati, J. (2016). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *IJNS-Indonesian Journal on Networking and Security*, 6(1).
14. Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support vector machine. *Proceeding Indones. Sci. Meeting Cent. Japan*.
15. Sembiring, K. (2007). Modul SVM.
16. Durgesh, K. S., & Lekha, B. (2010). Data classification using support vector machine. *Journal of theoretical and applied information technology*, 12(1), 1–7.