

PENGARUH RFE TERHADAP *LOGISTIC REGRESSION* DAN *SUPPORT VECTOR MACHINE* PADA ANALISIS SENTIMEN HOTEL SHANGRI-LA SURABAYA

Fakhri Maulana Herza, Basuki Rahmat, Muhammad Muharrom AL Haromainy

Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur

Jalan Raya Rungkut Madya No. 1 Surabaya, Indonesia

fakhrimaulanaherza@gmail.com

ABSTRAK

Analisis sentimen merupakan alat penting dalam industri pariwisata untuk memahami respon dan pengalaman tamu terhadap layanan hotel, di mana ulasan tamu sebelumnya berperan krusial dalam membentuk persepsi calon tamu terhadap kualitas fasilitas dan daya tarik hotel. Tantangan utama dalam analisis sentimen adalah memilih fitur yang paling relevan untuk meningkatkan kinerja model prediksi, karena tidak semua kata atau fitur dalam ulasan memiliki kontribusi signifikan dalam membedakan sentimen positif dan negatif. Dalam konteks ini, metode *Recursive Feature Elimination* (RFE) memiliki potensi untuk mengoptimalkan pemilihan fitur dengan mengeliminasi fitur yang kurang informatif, sehingga dapat meningkatkan akurasi model *Logistic Regression* dan *Support Vector Machine* (SVM). Penelitian ini fokus pada pengaruh penerapan RFE terhadap kinerja kedua model tersebut dalam analisis sentimen ulasan tamu di Hotel Shangri-La Surabaya, dengan data sebanyak 3719 ulasan. Hasil pengujian menunjukkan bahwa pada model *Logistic Regression* yang menggunakan RFE, terdapat peningkatan performa yang signifikan dalam presisi, sensitivitas, *F1 Score*, dan akurasi, dengan rata-rata peningkatan sebesar 9%. Pada model SVM, peningkatan performa bahkan lebih signifikan dengan rata-rata peningkatan sebesar 14%, yang menunjukkan bahwa penerapan RFE secara efektif meningkatkan kualitas prediksi kedua model dalam konteks analisis sentimen ulasan hotel.

Kata kunci : Analisis Sentimen, Sektor Pariwisata, *Google Maps*, *Recursive Feature Elimination*, *Support Vector Machine*, *Logistic Regression*.

1. PENDAHULUAN

Industri pariwisata, khususnya perhotelan, adalah sumber lapangan kerja yang potensial. Kemudahan akses internet membuat calon tamu semakin bergantung pada ulasan online untuk memutuskan pilihan menginap. Penelitian telah membuktikan bahwa ulasan ini sangat berpengaruh terhadap keputusan pembelian [2].

Analisis sentimen telah menjadi fokus berbagai penelitian sebelumnya. Dalam studi yang dilakukan pada tahun 2021 [7], sentimen masyarakat terkait vaksin COVID-19 dianalisis menggunakan algoritma SVM. Penelitian ini memanfaatkan 2634 komentar YouTube yang telah diolah menggunakan teknik TF-IDF untuk mengonversi teks menjadi format yang dapat dipahami komputer, dan hasilnya menunjukkan bahwa SVM mampu mengklasifikasikan sentimen dengan tingkat akurasi sebesar 77,75% [14]. Penelitian lain yang dilakukan [6] menunjukkan bahwa metode *Recursive Feature Elimination* (RFE) efektif dalam meningkatkan akurasi klasifikasi sentimen. Pada analisis sentimen terkait *e-wallet* di *Twitter*, penggunaan RFE berhasil meningkatkan akurasi algoritma SVM dari 74% menjadi 81%. Penelitian lain mengenai analisis sentimen PSBB di DKI Jakarta dan dampaknya terhadap IHSG [1] menemukan bahwa *Logistic Regression* merupakan metode paling efektif, [8] dengan akurasi mencapai 75% dalam mengklasifikasikan sentimen, diikuti oleh *random forest* (60%), KNN (70%), dan Naive Bayes (70%) [1].

Dalam penelitian ini, penulis akan menganalisis 3.719 data ulasan yang diambil dari *Google Maps*

Hotel Shangri-La Surabaya. Ulasan tersebut akan dikategorikan menjadi dua jenis sentimen, yaitu positif dan negatif. Untuk klasifikasi, algoritma SVM dan *Logistic Regression* akan digunakan, dengan data teks yang terlebih dahulu diubah menjadi representasi numerik menggunakan teknik *Word Embedding* TF-IDF. *Word Embedding* ini akan mengubah setiap kata menjadi vektor numerik yang dapat diproses oleh mesin [3].

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk melakukan klasifikasi data. [9] SVM bekerja dengan memetakan data ke ruang dimensi yang lebih tinggi, kemudian mencari hyperplane yang optimal untuk memisahkan data menjadi beberapa kelas. Keunggulan SVM terletak pada kemampuannya mencapai akurasi tinggi, bahkan dengan jumlah data yang terbatas. [10]

Logistic Regression adalah sebuah metode statistik yang digunakan untuk memprediksi kemungkinan terjadinya suatu peristiwa berdasarkan sejumlah faktor yang saling berkaitan. [12] Misalnya, kita bisa memprediksi apakah seseorang akan membeli produk tertentu berdasarkan usia, jenis kelamin, dan pendapatannya. Model ini menghasilkan nilai probabilitas antara 0 dan 1. RFE adalah teknik yang digunakan untuk memilih fitur-fitur paling penting dari data yang kita miliki. Dengan kata lain, RFE membantu kita menentukan faktor mana yang paling berpengaruh terhadap hasil yang ingin kita prediksi. Dengan menggabungkan kedua metode ini, kita bisa membangun model prediksi yang lebih akurat dan efisien. [13]

2. TINJAUAN PUSTAKA

Berikut ini penjelasan mengenai berbagai literatur yang relevan dengan topik penelitian yang diangkat. Hal ini dilakukan untuk mendapatkan gambaran secara menyeluruh tentang penelitian yang dilakukan.

2.1. Shangri-La

Shangri-La Surabaya, yang beralamat di Jalan Mayjen Sungkono, Surabaya, adalah destinasi ideal bagi mereka yang mencari kemewahan dan kenyamanan. Hotel bintang lima ini menyuguhkan kamar-kamar yang didesain dengan perpaduan gaya kontemporer dan sentuhan khas Jawa. Sebagai anggota Horizon Club, tamu akan dimanjakan dengan berbagai fasilitas istimewa, mulai dari check-out lambat hingga layanan koktail. Dengan fasilitas dan layanan kelas dunia, Shangri-La Surabaya menjadi salah satu hotel termahal namun paling diminati di Surabaya. [2]

2.2. Google Maps

Google Maps adalah aplikasi peta digital yang paling populer di dunia. Dengan *Google Maps*, kita bisa dengan mudah mencari lokasi, merencanakan perjalanan, dan menemukan tempat-tempat baru. Fitur-fitur seperti tampilan jalan 360 derajat, informasi lalu lintas *real-time*, dan navigasi yang akurat membuat *Google Maps* menjadi alat yang sangat berguna bagi pengendara. Selain itu, fitur ulasan yang menggabungkan pendapat dari berbagai sumber membantu kita membuat keputusan yang lebih baik sebelum mengunjungi suatu tempat. [5] [14]

2.3. Analisis Sentimen

Analisis sentimen adalah metode yang menggunakan komputer untuk memahami perasaan yang terkandung dalam teks. [4] Dengan kata lain, kita bisa mengetahui apakah seseorang memiliki pendapat yang positif atau negatif tentang suatu hal, misalnya produk atau isu tertentu. Penelitian ini bertujuan untuk mengukur opini publik secara keseluruhan, apakah mayoritas orang memiliki pandangan yang baik atau buruk. Analisis sentimen menjadi semakin populer sejak tahun 2002 dan sangat berguna dalam pengambilan keputusan. [12]

2.4. Preprocessing

Data teks mentah yang diambil langsung dari sumbernya seringkali tidak teratur dan sulit dipahami oleh komputer. Sebelum menganalisis data ini, kita perlu melakukan beberapa langkah persiapan, yang disebut praproses. Salah satu teknik yang sering digunakan adalah TF-IDF, yaitu memberikan bobot pada kata-kata penting dalam teks. Secara umum, praproses teks melibatkan beberapa tahap, seperti mengubah semua huruf menjadi huruf kecil, menghapus tanda baca dan karakter yang tidak perlu, membagi teks menjadi kata-kata individu, menghilangkan kata-kata umum seperti "dan", "atau", dan menyederhanakan bentuk kata. [13]

2.5 TF-IDF

TF-IDF adalah cara untuk mengukur seberapa penting suatu kata dalam sebuah teks. [4] Metode ini memberikan skor pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan seberapa jarang kata tersebut muncul di dokumen lain. Kata-kata yang memiliki skor TF-IDF tinggi dianggap lebih khas dan membedakan satu dokumen dengan dokumen lainnya.

2.6 Recursive Feature Elimination

Recursive Feature Elimination (RFE) adalah metode yang digunakan untuk memilih fitur-fitur penting dalam data. Bayangkan kita memiliki banyak sekali fitur, tapi tidak semuanya berguna untuk memprediksi sesuatu. RFE bekerja dengan cara secara bertahap menghilangkan fitur yang kurang penting, sehingga kita hanya menggunakan fitur-fitur yang paling berpengaruh. Proses ini diulang terus menerus sampai kita mendapatkan kumpulan fitur terbaik. Dengan menggunakan RFE, model kita akan menjadi lebih sederhana, lebih cepat, dan lebih mudah dipahami. [9]

2.7 Support Vector Machine

Support Vector Machine (SVM) adalah metode klasifikasi dalam data mining yang memanfaatkan pemetaan *non-linier* untuk mengubah data ke dimensi yang lebih tinggi, sehingga memungkinkan pemisahan data dari dua kategori dengan menggunakan *hyperplane*. Tujuan dari algoritma SVM adalah untuk menemukan *hyperplane* dalam ruang N-dimensi yang dapat dengan jelas memisahkan data. Dimensi *hyperplane* berkaitan dengan jumlah fitur; misalnya, jika ada dua fitur input, *hyperplane* akan berupa garis yang memisahkan dua kelompok data tersebut. [10]

2.8 Logistic Regression

Logistic Regression adalah sebuah alat statistik yang digunakan untuk memprediksi kemungkinan terjadinya suatu peristiwa. Peristiwa yang dimaksud di sini adalah peristiwa yang memiliki dua kemungkinan hasil, seperti "ya" atau "tidak", "benar" atau "salah", atau "1" atau "0". Contohnya, kita ingin memprediksi apakah seseorang akan membeli produk tertentu atau tidak berdasarkan beberapa faktor seperti usia, pendapatan, dan jenis kelamin. [12]

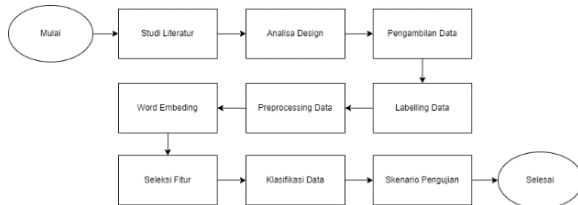
3. METODE PENELITIAN

Metode penelitian ini mencakup tahapan penelitian, studi literatur, pengambilan data, preprocessing data, hingga implementasi rfe dan evaluasi model klasifikasi yaitu *Support Vector Machine* dan *Logistik Regression*.

3.1. Tahapan Penelitian

Pada tahap penelitian dimulai dari studi literatur untuk mencari permasalahan permasalahan yang ada di dunia teknologi yang akan dipecahkan pada penelitian ini. Kemudian dilanjutkan dengan proses

analisa dan desain, pengambilan data, *labelling* data, *preprocessing* data, *word embedding*, seleksi fitur *Recursive Feature Elimination*, klasifikasi data, dan skenario pengujian. Detail mengenai masing-masing tahapan pelaksanaan penelitian akan dijelaskan satu persatu pada sub-bab berikutnya.

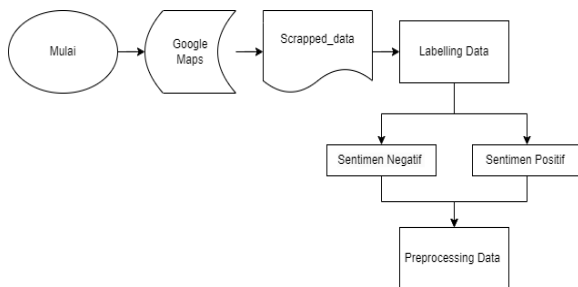


Gambar 1. Tahapan Penelitian

3.2. Studi Literatur

Tahap pertama yang dilakukan penulis pada pelaksanaan penelitian ini ialah dengan mengumpulkan sumber bacaan / literatur. Fungsi tahap ini ialah untuk mencari dan mempelajari permasalahan di dunia teknologi. Selain itu, studi literatur ini juga bermanfaat bagi penulis untuk mendapat penjelasan terkait konsep-konsep dan definisi dari sebuah metode dan istilah-istilah asing. Literatur didapatkan penulis dari banyak sumber, namun mayoritas berasal dari dokumentasi penelitian sebelumnya seperti jurnal, artikel prosiding, artikel website, skripsi terdahulu, dan lain-lain. Dari tahapan ini penulis juga bisa menemukan gap dari penelitian sebelumnya dan mendapatkan beberapa opsi cara dalam melakukan perancangan sistem. Berdasarkan uraian diatas, dapat disimpulkan bahwa tahapan studi literatur ini termasuk tahapan yang cukup penting pada proses penelitian.

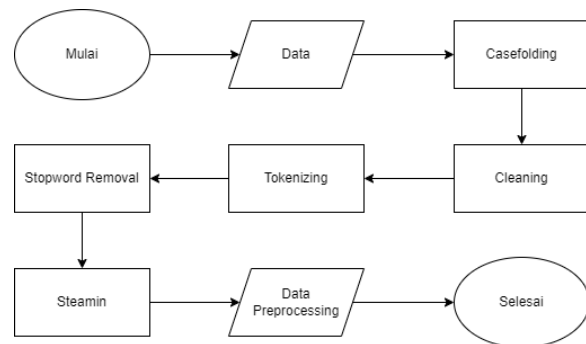
3.3. Pengambilan Data



Gambar 2. Pengambilan Data

Langkah pertama dalam perancangan sistem ini adalah tahap pengambilan data, di mana sistem akan melakukan *scrapping* terhadap 3719 data yang selanjutnya akan diproses untuk diklasifikasi. Dataset diambil menggunakan salah satu ekstensi *Google*, yaitu *Instant Data Scraper*. *Instant Data Scraper* adalah ekstensi *Google* yang dapat mengumpulkan berbagai jenis data, termasuk kolom ulasan di *Google Maps*.

3.4. Preprocessing Data



Gambar 3. Preprocessing Data

Tahap berikutnya setelah akuisisi dan pelabelan data adalah tahap praproses. Secara singkat, tahap ini bertujuan untuk mempersiapkan data agar siap digunakan dalam langkah-langkah selanjutnya. Beberapa proses yang dilakukan selama praproses data meliputi *case-folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*.

3.5. TF-IDF

Penelitian dengan topik klasifikasi tentunya membutuhkan fitur yang akan menjadi acuan dalam proses klasifikasi. Di dalam penelitian dengan topik analisis sentimen, dikenal istilah *word embedding* sebagai proses ekstraksi fiturnya. *Word Embedding* merupakan sebuah proses pemberian value berupa angka pada dataset yang berbentuk teks. Salah satu metode *word embedding* ialah TF-IDF. Metode TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya. Namun, sebelum menerapkan metode TF-IDF, perlu dilakukan proses *label encoding* untuk mengonversi kelas target menjadi format numerik.

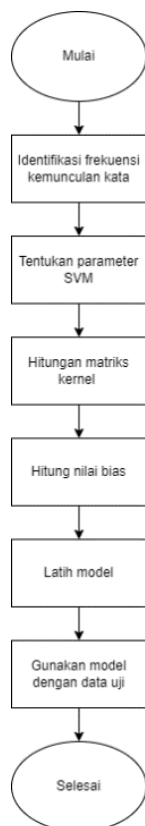
3.6. Recursive Feature Elimination



Gambar 4. Diagram Alur RFE

Dataset awal yang kami miliki merupakan kumpulan data yang cukup kompleks dengan sepuluh fitur pada setiap entri, di mana masing-masing fitur menyimpan nilai numerik hasil dari proses tokenisasi. Data ini memberikan gambaran rinci mengenai berbagai parameter yang terkait dengan studi kami. Namun, seiring dengan meningkatnya jumlah fitur, muncul tantangan dalam hal efisiensi pemrosesan dan kompleksitas model. Untuk mengatasi masalah ini dan meningkatkan kualitas model, kami menerapkan teknik *Recursive Feature Elimination* (RFE). RFE adalah metode yang dirancang untuk memilih fitur-fitur yang paling relevan dengan mengeliminasi fitur-fitur yang memiliki kontribusi kecil terhadap performa model.

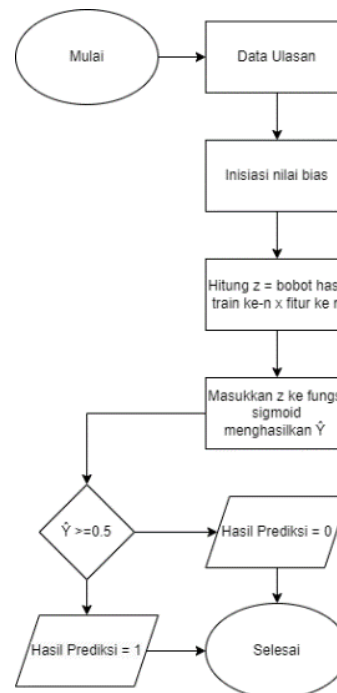
3.7. Support Vector Machine



Gambar 5. Diagram Alur SVM

Diagram alur di atas menggambarkan proses klasifikasi teks menggunakan algoritma *Support Vector Machine* (SVM). Proses dimulai dengan mengubah teks menjadi representasi numerik berdasarkan frekuensi kemunculan kata. Selanjutnya, parameter SVM seperti jenis kernel dan nilai C ditentukan. Model kemudian dilatih untuk menemukan batas optimal yang memisahkan data ke dalam kelas-kelas yang berbeda. Setelah pelatihan, model diuji menggunakan data baru untuk mengevaluasi performanya. Langkah-langkah ini melibatkan perhitungan matriks kernel dan nilai bias untuk membangun model yang akurat.

3.8. Logistic Regression



Gambar 6. Diagram Alur Logistic Regression

Diagram alur ini menggambarkan proses klasifikasi sentimen teks menggunakan model regresi logistik. Dimulai dengan input berupa data ulasan, model kemudian menginisialisasi nilai bias sebagai titik awal perhitungan. Selanjutnya, setiap kata dalam ulasan dikalikan dengan bobotnya masing-masing yang telah diperoleh dari tahap pelatihan untuk menghasilkan nilai z . Nilai z ini kemudian diubah menjadi probabilitas antara 0 dan 1 melalui fungsi sigmoid. Jika probabilitas ini lebih besar sama dengan 0.5, maka ulasan dikategorikan sebagai positif, sebaliknya jika kurang dari 0.5, maka ulasan dikategorikan sebagai negatif.

4. HASIL DAN PEMBAHASAN

Pada bab ini akan membahas terkait hasil implementasi program dari *Recursive Feature Elimination* pada analisis sentimen hotel shangri-la surabaya menggunakan *Support Vector Machine* dan *Logistic Regression* Untuk selanjutnya akan dijelaskan terkait hasil evaluasi program tersebut.

4.1. Akusisi dan Labeling data

Di dalam tahap akusisi dan *labeling* data, penelitian ini menggunakan data berupa teks yang merupakan ulasan di kolom komentar *Google Maps* mengenai ulasan terkait Hotel Shangri-La Surabaya. Total dari hasil pengumpulan data berjumlah sebanyak 3719 data ulasan dari *Google Maps*. Pada proses pengumpulan data, peneliti menggunakan proses *scrapping* data. Hasil dari pengumpulan data disimpan ke file dengan format csv.

Gambar 7. Hasil Scrapping Data

Gambar 8. Hasil Pelabelan Sentimen

Dalam penelitian ini, tahapan selanjutnya yaitu masing-masing token yang ada di dalam dataset akan dihitung bobotnya, berbagai metode digunakan untuk representasi fitur dan pembobotan, seperti TF-IDF untuk penyematan kata. Sementara TF-IDF digunakan untuk kolom 'Ulasan', Label encoding diterapkan ke kolom 'Sentimen' untuk mengubah kelas sentimen menjadi kode numerik, dengan sentimen positif direpresentasikan sebagai 1 dan sentimen negatif sebagai 0. Proses Label Encoding ini sangat penting untuk mengubah data kategoris menjadi format yang dapat digunakan oleh algoritma pembelajaran mesin secara efektif. Pustaka sklearn dimanfaatkan untuk perhitungan TF-IDF dan Label Encoding.

11616

Berikut adalah contoh fitur sebelum diproses RFE

index	0	1	2	3	4
0	-1.289606683811866	1.3349467283836431	0.5549671649988613	-0.8976956308910022	1.2235796597592736
1	3.247142959358156	-1.2421533507916103	-1.0381069864610342	2.245119719314721	-0.9888152681895256
2	-1.1729501349033876	-0.7675070368738105	0.1023821983476529	-0.8690341731454873	-0.45242426480904147
3	-1.0241759577669223	-0.8212863840728593	-0.07893638138549375	-0.054102549093622	-0.1591833461675044
4	1.28974171233791	1.2239202016169175	-0.3926127308442393	0.029722460805411043	-0.43706646689361503
5	-0.612059201193422	-1.2338365862541953	1.1893913786129298	0.2516813846568609	-0.38791257361803495
6	1.62357889686317	0.040798218223812355	-0.0023434031495201645	1.4449692665975475	-0.5948203104540761
7	0.783613132092555	-0.7260041032460013	-0.822002520114908	0.6723052416315588	-1.5939125266840073
8	-0.7737794828025139	0.8090542890877785	1.0347991305262363	-0.8125570246340255	0.554254476028521
9	1.3193795562733612	0.656044545371783	-1.9798213656006851	-0.12201334360238117	-1.1590465425293566
5	0.15892505899675513	0.04132683187700316	-0.6237160349163927	1.3435465066220513	0.9355457608405814
6	-0.41546622397609134	1.59310622159029255	-0.977284832042989	-0.027555936515566763	-2.3180684424156364
7	0.09100067597177355	-2.143966831621994	0.09643037952850442	-0.6298673491373752	0.9815029943245259
8	0.787449688312832	0.9454973048593947	-0.30639542708651957	0.2850347420993615	-0.8804559477927234
9	-1.0395160970964155	2.5518400945723556	0.3984108241501123	0.4799328107989457	1.2168246158219649
0	0.7569939341062406	0.2760790054815457	0.3181423146963727	0.29915097295744597	-1.22934703194412736
1	0.11649247877136737	-1.0749895671304055	1.8458013203719315	2.93032028549251	-1.9550706461392329
2	0.03097994714371364	-0.623914201844913	0.36939486101994184	-0.38789382493422025	-0.8816263038636215
3	-0.1801201976171233	-0.32883082658724	-0.0053305733261247	-1.0844706235787835	1.237681775250514
4	-1.2089086712174377	-1.4195415457708802	1.434445169634489	0.726823960180868	1.6117350836760291

Gambar 11. Data Sebelum Proses RFE

Berikut adalah contoh fitur setelah diproses RFE:

	0	1	2	3	4
0	-1.289607	0.554967	-0.897696	0.158925	0.935546
1	3.247143	-1.038107	2.245120	-0.415466	-2.318068
2	-1.172090	0.102382	-0.869034	0.091001	0.981503
3	-1.024176	-0.079396	-0.055410	0.787450	-0.880456
4	1.298974	-0.392613	0.029722	-1.039516	1.216825
...	...	...	...	...	...
3714	-1.867374	0.488520	-1.037448	0.494042	0.706731
3715	-2.033016	-0.798473	-1.199226	0.467717	0.941643
3716	0.771047	-0.944956	-0.213047	-0.849031	1.291867
3717	1.755849	-0.718598	1.001514	-0.438363	-0.728781
3718	0.370457	0.443160	-0.347937	-0.654889	1.227030

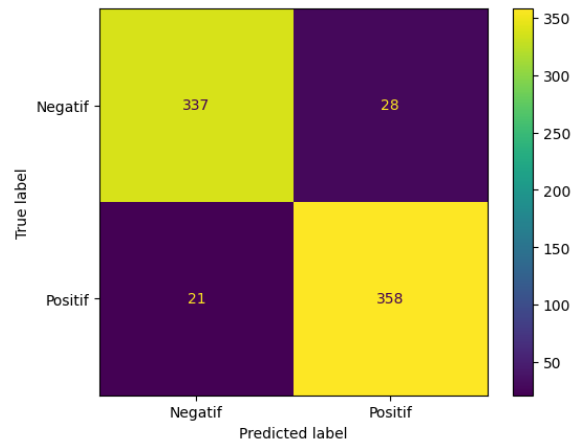
Gambar 12. Data Setelah Proses RFE

Dengan menerapkan teknik Recursive Feature Elimination (RFE), kami berhasil mengurangi dimensi dataset dari sepuluh fitur menjadi lima fitur. Proses ini dilakukan dengan secara bertahap menghapus fitur-fitur yang kurang berkontribusi terhadap model. Hasilnya, model yang dihasilkan diharapkan lebih efisien dan akurat karena hanya berfokus pada fitur-fitur yang paling relevan. Pembulatan nilai fitur juga dilakukan untuk menyederhanakan data dan memudahkan analisis lebih lanjut.

4.5. Hasil Evaluasi Model

Bagian selanjutnya setelah melakukan evaluasi model, kemudian dilakukan analisa hasil.

Berikut adalah contoh perhitungan evaluasi model SVM dengan data train 80% data uji 20%:



Gambar 13. Confusion Matrix

1. Perhitungan Presisi

$$\text{Negatif} = \frac{TN}{TN + FN} = \frac{337}{337 + 28} = \frac{337}{365} = 0.92$$

$$\text{Positif} = \frac{TP}{TP + FP} = \frac{358}{358 + 21} = \frac{358}{379} = 0.94$$

2. Perhitungan Sensitivitas

$$\text{Negatif} = \frac{TN}{TN + FP} = \frac{337}{337 + 21} = \frac{337}{358} = 0.94$$

$$\text{Positif} = \frac{TP}{TP + FN} = \frac{358}{358 + 28} = \frac{358}{386} = 0.92$$

3. Perhitungan F1 Score

$$\text{Negatif} = 2 \times \frac{\text{Presisi} \times \text{Sensitivitas}}{\text{Presisi} + \text{Sensitivitas}} = 2 \times \frac{0.92 \times 0.94}{0.92 + 0.94} = 0.93$$

$$\text{Positif} = 2 \times \frac{\text{Presisi} \times \text{Sensitivitas}}{\text{Presisi} + \text{Sensitivitas}} = 2 \times \frac{0.98 \times 0.92}{0.98 + 0.92} = 0.95$$

4. Perhitungan Akurasi

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{358 + 337}{358 + 337 + 21 + 28} = 0.93$$

Tabel 1. Hasil Evaluasi Keseluruhan

Model	Optimasi	Train	Test	Presisi	Sensitivitas	F1 score	Akurasi
SVM	RFE	90	10	0.91	0.91	0.91	0.91
		80	20	0.95	0.93	0.93	0.93
		70	30	0.92	0.92	0.93	0.92
	- RFE	90	10	0.79	0.81	0.93	0.92
		80	20	0.75	0.76	0.74	0.92
		70	30	0.76	0.75	0.75	0.91
LR	RFE	90	10	0.92	0.92	0.91	0.92
		80	20	0.92	0.92	0.92	0.92
		70	30	0.92	0.92	0.91	0.92
	- RFE	90	10	0.74	0.91	0.79	0.94
		80	20	0.73	0.88	0.78	0.94
		70	30	0.73	0.88	0.78	0.94

Tabel di atas menyajikan hasil evaluasi kinerja model Support Vector Machine (SVM) dan Logistic Regression (LR) dengan dan tanpa penggunaan

metode Recursive Feature Elimination (RFE) pada pembagian data pelatihan (train) dan pengujian (test) yang berbeda (90-10, 80-20, dan 70-30). Evaluasi

dilakukan dengan mengukur metrik performa seperti presisi, sensitivitas, F1 score, dan akurasi.

Pada model SVM dengan RFE, hasil menunjukkan bahwa model berkinerja konsisten dengan akurasi yang tinggi di semua pembagian data. Pada split 80-20, model mencapai akurasi tertinggi sebesar 93% dengan presisi 0.95, sensitivitas 0.93, dan F1 score 0.93, mengindikasikan bahwa optimasi dengan RFE mampu meningkatkan performa model. Di sisi lain, SVM tanpa RFE menghasilkan performa yang lebih rendah, dengan akurasi maksimal 92% pada split 90-10, dan presisi yang turun menjadi 0.75-0.76 pada split 80-20 dan 70-30.

Untuk model Logistic Regression dengan RFE, performanya juga sangat konsisten dengan akurasi sebesar 92% di semua pembagian data. Presisi, sensitivitas, dan F1 score juga tetap tinggi, menunjukkan stabilitas model dalam menangani variasi data. Namun, Logistic Regression tanpa RFE mengalami penurunan performa, terutama pada presisi yang hanya mencapai 0.73-0.74, meskipun akurasi masih terjaga pada 94%. Hal ini mengindikasikan bahwa meskipun akurasi tinggi, tanpa RFE, model Logistic Regression kurang presisi dalam membedakan antara kelas positif dan negatif.

Secara keseluruhan, penggunaan RFE memberikan peningkatan kinerja signifikan pada kedua model, terutama dalam hal presisi dan konsistensi hasil klasifikasi.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa penggunaan metode Recursive Feature Elimination (RFE) pada model SVM dan Logistic Regression secara signifikan meningkatkan kinerja keduanya. Model SVM yang menggunakan RFE menunjukkan peningkatan yang jelas dalam presisi, sensitivitas, dan F1 Score dibandingkan dengan SVM tanpa RFE. Peningkatan ini juga terlihat pada model Logistic Regression dengan RFE. Pada model SVM, penerapan RFE berhasil meningkatkan presisi dari 80% menjadi 93% untuk proporsi data 90:10, sementara sensitivitas dan F1 Score juga mengalami peningkatan yang signifikan, dengan akurasi tetap stabil di angka 92%. Hal ini menunjukkan bahwa RFE tidak hanya meningkatkan kemampuan deteksi model tetapi juga menjaga akurasi keseluruhan. Pada model Logistic Regression, RFE menghasilkan peningkatan performa yang konsisten di semua proporsi data yang diuji, dengan presisi, sensitivitas, dan F1 Score meningkat, meskipun akurasi model tanpa RFE sedikit lebih tinggi. Namun, peningkatan pada metrik lainnya menunjukkan bahwa RFE memberikan keseimbangan performa yang lebih baik. Berdasarkan analisis, sebagian besar ulasan masyarakat tentang Hotel Shangri-La Surabaya menunjukkan sentimen positif, yang berpotensi memberikan dampak positif bagi hotel tersebut karena memperkuat keyakinan masyarakat terhadap kualitas hotel dan mendorong peningkatan layanan serta fasilitasnya. Untuk pengembangan

selanjutnya, disarankan untuk menambah jumlah dataset yang digunakan agar dapat meningkatkan akurasi dan prediksi dalam proses analisis sentimen. Selain itu, penelitian di masa mendatang diharapkan dapat mempertimbangkan aspek-aspek lain sehingga kelebihan dan kelemahan dari aplikasi yang digunakan dapat lebih teridentifikasi secara mendalam.

DAFTAR PUSTAKA

- [1] Abdulloh, F. F., & Pambudi, I. R. (2021, 3 13). CSRID Journal. *Analisis sentimen pengguna youtube terhadap program vaksin covid-19*, 13(3), 141 <https://www.doi.org/10.22303/csrid>
- [2] *About Shangri-La Surabaya*. (n.d.). Shangri-La Hotel. Retrieved February 11, 2024, from <https://www.shangrila.com/surabaya/shangrila/about/>
- [3] Dewi, N. R., Puspaningrum, E. Y., & Maulana, H. (2022, April). Analisis sentimen tweet vaksinasi covid-19 menggunakan rnn dengan metode tf-idf dan word2vec. *Jurnal Informatika dan Sistem Informasi*, 3(1), 56-65
- [4] Edwar, & Samsudin, M. (2023, Agustus). Informatics for Educators And Professionals : Journal of Informatics. *Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini pilkada dki 2017)*, 8(1), 11-18. E-ISSN: 2548-3412
- [5] *Google Maps Adalah Layanan Pemetaan Wilayah Melalui Web, Ketahui Manfaatnya*. (2020, November 28). Merdeka.com. Retrieved February11,2024,from<https://www.merdeka.com/jatim/google-maps-adalah-layananpemetaan-wilayah-melalui-web-ketahui-selengkapnya-kln.html>
- [6] Hendriyanto, H. D., Ridha, A. A., & Enri, U. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm. *Journal of Information Technology and Computer Science(intecomS)*, 5 <https://doi.org/10.31539/intecomS.v5i1.3708>
- [7] Laia, M. L., & Setyawan, Y. (2020, Juli). Jurnal Statistika Industri dan Komputasi. *Perbandingan hasil klasifikasi curah hujan menggunakan metode svm dan nbc*, 5(2), 51-61. <https://doi.org/10.34151/statistika.v5i02>
- [8] Pangaribuan, J. J., Tanjaya, H., & Kenichi. (2021, Juli 2). Information system development. *Mendeteksi penyakit jantung menggunakan machine learning dengan algoritma logistic regression*, 6(2), 1-10.
- [9] Perbandingan kinerja algoritma Naïve Bayes dan k-nn Pendekatan Lexicon pada Analisis Sentimen di Media Twitter. (2019). *Prosiding Seminar Nasional Fisika Universitas Riau IV*, 3011(1), 3011–3016. <https://repository.uin-suska.ac.id/26321/>

- [10] Rahman, M. A. (2019). Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network. *Journal of intelligent systems and computation*, 1(2), 50-57. <https://doi.org/10.52985/insyst.v1i2.83>
- [11] Sabrani, A., Wedashwara, I. W., & Bimantoro, F. (2020). Metode Multinomial Naïve Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (jtika)*, 2(1), 89 <https://doi.org/10.29303/jtika.v2i1.87>
- [12] Saraswita, E. F., Rini, D. P., & Abdiansah. (2021, Oktober). Jurnal Media Informatika Budidarma. *Analisis Sentimen E-Wallet di Twitter Menggunakan Support Vector Machine dan Recursive Feature Elimination*, 5(4), 1195-1200. DOI 10.30865/mib.v5i4.3118
- [13] Syahadati, A., Lengkong, N. C., & Safitri, O. (2021). Jurnal Teknoinfo. *Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg*, 15(1), 20-25. doi : 10.33365/jti
- [14] Yutika, C. H., Adiwijaya, & Faraby, S. A. (2021, April). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422-430. DOI 10.30865/mib.v5i2.2845