

Implementasi Algoritma Support Vector Machine dan Random Forest Terhadap Analisis Sentimen Masyarakat dalam Penggunaan Aplikasi Tiket.com, Traveloka, dan Agoda pada Google Playstore

Calleb Bhaskoro Prabowo^{*1}, Teguh Iman Hermanto², Imam Ma'ruf³

^{1,2,3})Program Studi Teknik Informatika, Fakultas Teknik, STT Wastukencana Purwakarta

Email: ^{*1}callebbhaskoro23@wastukencana.ac.id, ²teguhiman@wastukencana.ac.id,

³imam.ma@wastukencana.ac.id

(Naskah masuk: 05 Juli 2023, diterima untuk diterbitkan: 21 Desember 2023)

Abstrak: Internet hadir sebagai elemen penting dalam menyokong perkembangan teknologi dan informasi dalam segala sektor aktifitas manusia. Pada sektor perdagangan dan pariwisata contohnya, aplikasi Tiket.com, Traveloka, dan Agoda menjadi aplikasi yang paling diminati masyarakat Indonesia saat ini. Ulasan atau review yang disematkan oleh para pengguna merupakan hal penting bagi pihak perusahaan untuk mengetahui kepuasan pelanggan yang nantinya digunakan untuk meningkatkan kualitas dalam segi pelayanan. Proses menganalisis ulasan komentar memiliki beberapa tahapan karena memiliki data yang jumlahnya tidak sedikit. Penggunaan suatu metode membantu dalam melakukan proses klasifikasi komentar yang bersifat positif atau negatif. Ulasan pengguna aplikasi yang diproses diambil dari platform layanan penyedia aplikasi Google Playstore, lalu menarik data yang diinginkan dengan memasukan library Google Scraper pada Python. Data yang sudah ditarik selanjutnya diberi label untuk memisahkan ulasan yang bersifat positif dan negatif hal ini bertujuan untuk mempermudah proses klasifikasi dengan menggunakan metode Support Vector Machine (SVM) dan Random Forest. Hasil yang didapatkan nantinya merupakan tingkat akurasi dari dua metode yang digunakan berdasarkan pengolahan data yang sudah dilakukan pada masing-masing ulasan yang menjadi data set pada setiap aplikasi. Support Vector Machine memiliki akurasi yang lebih baik dibandingkan dengan Random Forest dengan rincian 85,5%, 87%, dan 88,7% pada urutan aplikasi Tiket.com, Traveloka, dan Agoda. Sedangkan Random Forest memiliki akurasi 84.7%, 84.7%, dan 88.2% dengan urutan aplikasi yang sama.

Kata Kunci – Tiket.com; Traveloka; Agoda; Support Vector Machine; Random Forest

Implementation of the Support Vector Machine and Random Forest Algorithms on Community Sentiment Analysis in Using the Tiket.com, Traveloka, and Agoda Applications on Google Playstore

Abstract: The internet is present as an important element in supporting the development of technology and information in all sectors of human activity. In the trade and tourism sector, for example, the Tiket.com, Traveloka and Agoda applications are the applications that are most popular with Indonesian people at the moment. Reviews posted by users are important for companies to know customer satisfaction which will later be used to improve quality in terms of service. The process of analyzing comment reviews has several stages because it contains quite a lot of data. The use of a method helps in the process of classifying comments that are positive or negative. Processed application user reviews are taken from the Google Playstore application provider service platform, then extract the desired data by inputting the Google Scraper library in Python. The data that has been drawn is then labeled to separate positive and negative reviews. This aims to make the classification process easier using the Support Vector Machine (SVM) and Random Forest methods. The results obtained will be the level of accuracy of the two methods used based on data processing that has been carried out in each review which becomes the data set for each application. Support Vector Machine has better accuracy compared to Random Forest with details of 85.5%, 87%, and 88.7% for the Tiket.com, Traveloka, and Agoda applications. Meanwhile, Random Forest has an accuracy of 84.7%, 84.7%, and 88.2% with the same application sequence.

Keywords – Tiket.com; Traveloka; Agoda; Support Vector Machine; Random Forest

1. PENDAHULUAN

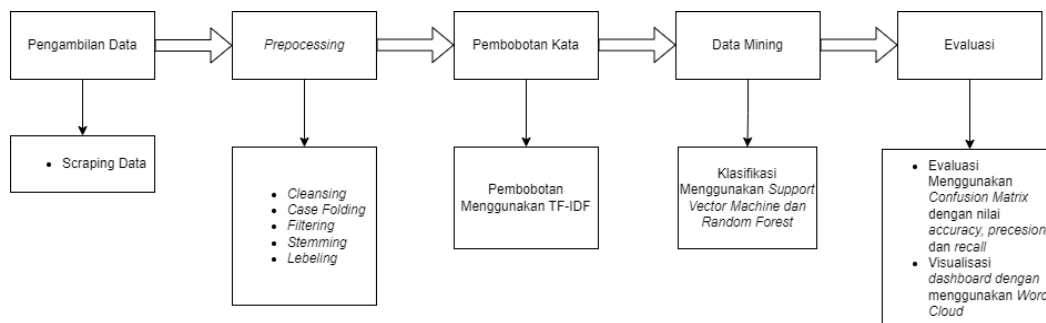
Perkembangan pariwisata saat ini sudah semakin pesat, baik dari segi pelayanan maupun teknologinya. Kebutuhan pariwisata khususnya di bidang perjalanan semakin lama semakin cepat pelayanannya, maka teknologi pun menyesuaikan untuk memenuhi kebutuhan tersebut[1]. Tidak dapat dipungkiri bahwa internet berkembang pesat di sektor pariwisata dan perjalanan sehingga memberi dampak tentang bagaimana bisnis terhubung dengan konsumen secara efektif. Sehingga banyak bermunculan Online Travel Agent (OTA) yang menyediakan layanan akomodasi dan tiket secara daring, membuat konsumen saat ini untuk melakukan perencanaan dan pemesanan kamar hotel atau tiket akomodasi dengan mudah, cepat, dan praktis menggunakan OTA contohnya seperti aplikasi Traveloka dan Tiket.com[2].

Namun terkadang sebuah sistem juga mempunyai sebuah kelemahan, seperti ada bug dalam aplikasi, server yang kadang mengalami gangguan, dan masih banyak lagi, untuk itu perlu adanya peningkatan dan perbaikan yang perlu dilakukan pada aplikasi dengan mengacu pada review yang dituliskan oleh pengguna aplikasi. Hal ini diharapkan ada umpan balik dari pengguna. Oleh karena itu dibutuhkan suatu metode yang dapat mensortir dan memantau ulasan tersebut secara cepat dan otomatis dalam mengkategorikan ulasan-ulasan tersebut baik yang positif maupun yang bersifat negatif[3].

Klasifikasi data review pengguna berdasarkan jenis sentimen akan mempermudah pihak development dalam mendapatkan informasi persepsi dari pengguna aplikasi. Hal ini diharapkan agar aplikasi tersebut dapat melakukan banyak perbaikan guna meningkatkan kualitas pelayanan terhadap pengguna aplikasi. Proses pengklasifikasian ini dilakukan dengan menggunakan dua metode yaitu Support Vector Machine dan Random Forest.

2. METODE PENELITIAN

Pada proses penelitian ini data akan melalui beberapa tahapan pengolahan. Hal ini bertujuan agar data tersusun rapih dan tertata ketika dimasukan kedalam proses pengklasifikasian, supaya hasil yang didapat nantinya lebih maksimal. Tahapan penelitian tersebut dapat dilihat pada kerangka penelitian yang terdapat pada gambar 1 dibawah ini.



Gambar 1. Kerangka Penelitian

2.1. Pengambilan Data

Data yang diambil untuk melakukan penelitian ini berasal dari kolom komentar aplikasi yang ada pada Google Playstore. Tahapan pengambilan data ini menggunakan Scraping data melalui Python. Jumlah data yang diambil sejumlah 2.000 data setiap aplikasi, untuk total keseluruhan data yaitu 6.000 data yang akan diambil. Scraping data ini dilakukan dengan mengunjungi situs Google Playstore dan mencari aplikasi yang dituju, lalu selanjutnya menarik komentar dengan menggunakan google play scraper dan di import ke dalam excel.

2.2. Preprocessing

Pada tahapan ini data yang sudah dikumpulkan diproses dengan tujuan untuk menghilangkan data-data yang tidak digunakan. Tujuan lain dari preprocessing ini agar data yang akan di proses nantinya menjadi lebih rapih, terstruktur dan lebih mudah untuk dilakukan pengolahan data. Pada proses ini dilakukan beberapa tahapan seperti case folding, cleansing, tokenization, filtering, stemming dan labeling.

2.3. Pembobotan Kata Menggunakan TF-IDF

Proses selanjutnya yaitu pembobotan kata menggunakan TF-IDF (term frequency-inversed document frequency) atau jumlah kemunculan kata. Ini dilakukan untuk menghitung berapa jumlah kata yang keluar pada komentar. Perhitungan bobot tiap term bertujuan untuk mengetahui ketersediaan dan kemiripan suatu term didalam komentar, semakin banyak kata yang sering muncul pada komentar, semakin tinggi pula bobot term tersebut.

$$W_{ij} = tf_{ij} \times \log\left(\frac{D}{df_j}\right) \quad (1)$$

Keterangan:

w_{ij} : bobot term terhadap dokumen

tf_{ij} : jumlah kemunculan term dalam dokumen

D : total jumlah dokumen

df_j : jumlah dokumen yang terdapat term

2.4. Klasifikasi Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan metode machine learning yang memprediksi kelas berdasarkan pola dari hasil proses training yang diciptakan[4].

Meskipun waktu pelatihan SVM kebanyakan lambat, tetapi metode ini sangat akurat karena kemampuannya untuk menangani model-model nonlinear yang kompleks. SVM kurang rentan terhadap overfitting dibandingkan metode lainnya. SVM dapat digunakan untuk prediksi dan klasifikasi. Contoh penerapannya antara lain deteksi tulisan tangan, pengenalan obyek, identifikasi suara dan lain-lain[5].

2.5. Klasifikasi Random Forest

Random forest merupakan algoritma esamble learning yang menggunakan serta membangun struktur tree secara bertahap. Dalam penggunaan pohon keputusan dibangun berdasarkan memilih ataupun mengambil data secara acak. Penentuan kelas dalam suatu data menggunakan sistem voting berdasarkan hasil dari decision tree[6].

Metode random forest adalah pengembangan dari metode CART, yaitu dengan menerapkan metode bootstrap aggregating (bagging) dan random feature selection. Dalam random forest, banyak pohon ditumbuhkan sehingga terbentuk hutan (forest), kemudian analisis dilakukan pada kumpulan pohon tersebut[6].

2.6. Evaluasi

Proses terakhir ini merupakan tahap evaluasi, dimana model yang dipakai untuk proses evaluasi ini menggunakan confusion matrix. Tahap ini bertujuan untuk mengevaluasi hasil kinerja klasifikasi yang telah diperoleh dengan menghitung nilai accuracy, precision, dan recall. Serta langkah yang terakhir yaitu memvisualisasikan hasil klasifikasi dengan menggunakan dashboard word cloud.

3. HASIL DAN PEMBAHASAN

3.1. Scraping Data

Data yang diambil pada ulasan Google Playstore diambil menggunakan Google Scraper dengan menggunakan library pada Python. Setelah memasukan library dan jumlah data yang diinginkan maka akan muncul hasil dari data yang sudah di proses. Pada gambar 2 dibawah menunjukkan bahwa data yang diambil terdiri dari username, content, tanggal komentar masuk dan lain-lain. Data yang didapat setelah melalui proses Scraping sejumlah 6.000 data, masing-masing aplikasi menarik 2.000 data.

	reviewId	userName	userImage	content	score	thumbsupCount	reviewCreatedVersion	at
0	1a6dc8e7-b484-444b-8fb3-2e278f6861fa	Rheza Wibowo	lh.googleusercontent.com/a-/AD_cm...	Tinggal melakukan pembayaran kenapa tidak bisa...	1	12	4.59.0	2023-06-07 05:38:00
1	7dfe44fb-6e3d-4a7d-a9f1-265d3d05f30c	yugi engineer	lh.googleusercontent.com/a-/AD_cm...	Sangat mengecewakan saya sudah download aplikasi...	1	19	4.59.0	2023-06-05 11:16:58
2	ea3fb617-94ba-4ebd-a096-a734ca2036da	rahmat taufik	lh.googleusercontent.com/a/AAcHTL...	Aplikasi mengecewakan, saya sudah lakukan pemb...	1	39	4.56.0	2023-05-21 03:49:17
3	545dd3e3-f902-4220-b06c-3c4db3556c82	All Fals	lh.googleusercontent.com/a-/AD_cm...	... aku udah pesan tiket pesawat, Karna kondisi...	1	31	4.56.0	2023-05-25 06:31:30
4	d1536372-5d14-413d-8ece-f76cc2f28e23	Taufik Saetika	lh.googleusercontent.com/a-/AD_cm...	Mudah dan simple harga yg dibayarkan sesuai de...	5	3	4.56.0	2023-06-02 09:20:35

Gambar 2. Hasil Scraping

3.2. Preprocessing Data

Proses preprocessing merupakan proses dimana data yang sudah diambil diolah menggunakan bahasa pemrograman Python, karena banyak noise yang dapat menyebabkan proses klasifikasi menghasilkan hasil yang kurang maksimal. Salah satu contoh noise yang dapat kita jumpai ialah banyaknya ditemukan kata singkatan yang tidak masuk ke dalam Kamus Besar Bahasa Indonesia(KBBI).

3.3. Pembobotan Kata Menggunakan TF-IDF

Penggunaan metode TF-IDF ini untuk mengubah data test menjadi vector namun harus diperhatikan juga bahwa kata tersebut bersifat informatif atau tidak .Proses TF-IDF ini menggunakan library yang di import pada Sklearn.

3.4. Pengklasifikasian Menggunakan Support Vector Machine dan Random Forest

Data yang sudah melalui tahapan pembobotan kata selanjutnya dimasukan kedalam pemodelan algoritma Support Vector Machine dan Random Forest. Pada proses ini akan dilakukan pemisahan data testing dan data training yang akan diuji dengan perbandingan 80:20. Hal ini bertujuan untuk melihat berapa tingkat akurasi, precision, dan recall dari masing-masing algoritma.

3.5. Hasil dan Visualisasi

3.5.1. Hasil Perbandingan Algoritma

Data training dan testing yang sudah dilakukan bertujuan untuk melihat nilai yang dihasilkan, serta membandingkan persentase nilai algoritma dari setiap aplikasi. Hasil dari pengujian setiap model algoritma pada setiap aplikasi dapat dilihat pada tabel 1 dibawah ini.

Tabel 1 Tabel Perbandingan Algoritma

Algoritma	Akurasi	Precision	Recall
SVM (Tiket.com)	85.5%	84.4	97.5
Random Forest (Tiket.com)	84.7%	85.4	94.7
SVM (Traveloka)	87%	82.3	95.6
Random Forest (Traveloka)	84.7%	82.4	89.9
SVM (Agoda)	88.7%	88.5	99
Random Forest (Agoda)	88.2%	88.6	98.4

Tabel diatas menunjukan bahwa performa model Support Vector Machine memiliki nilai yang lebih baik, karena tingkat akurasi dan recall yang dimiliki lebih tinggi dibanding dengan Random Forest.

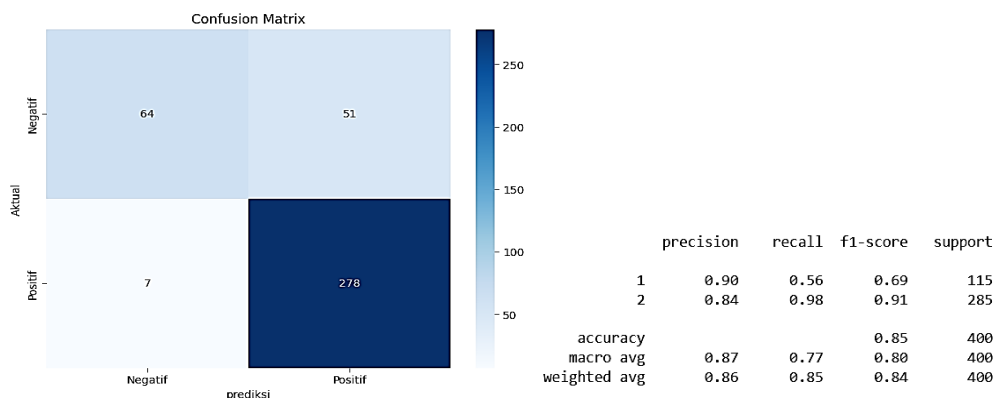
3.5.2. Hasil Confusion Matrix Support Vector Machine dan Random Forest

Untuk memastikan hasil dari pemodelan yang sudah dihasilkan, selanjutnya data dimasukan kedalam Confusion Matrix untuk diuji kembali tingkat akurasi, precision dan recall pada setiap aplikasi yang sudah dihasilkan sebelumnya. Pengujian ini juga bisa dilakukan secara manual untuk menyocokkan hasil yang didapat dengan menggunakan rumus yang ada dibawah ini.

$$\text{Rumus Akurasi} : A = \frac{TP+TN}{\text{Jumlah Data}} \quad (2)$$

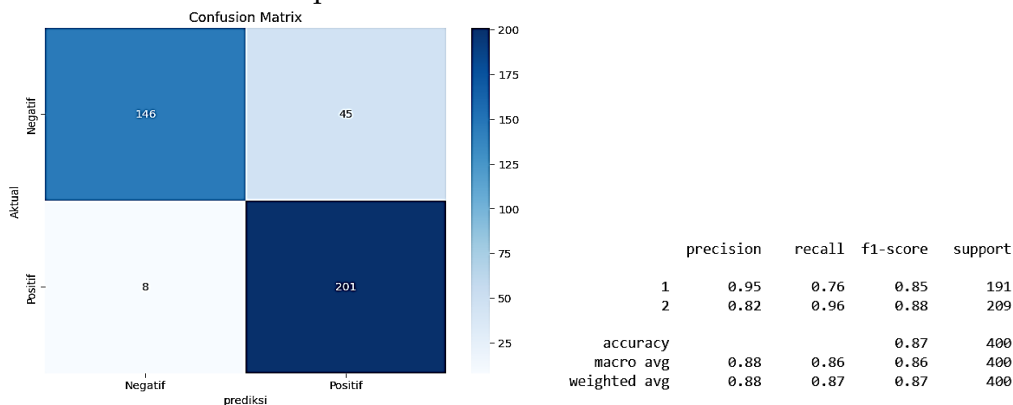
$$\text{Rumus Precision} : \text{All Precision} = \frac{\text{Precision A+B}}{\text{Jumlah Kelas}} \quad (3)$$

$$\text{Rumus Recall} : \text{Recall} = \frac{\text{Recall A+B}}{\text{Jumlah Kelas}} \quad (4)$$



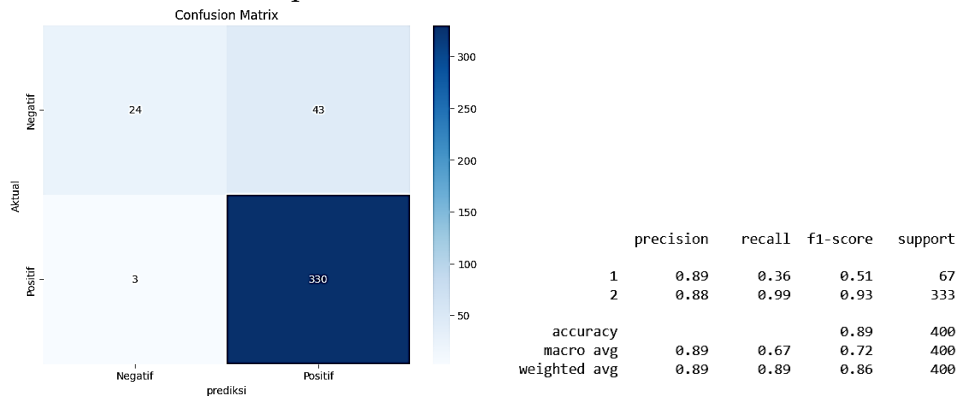
Gambar 3. Confusion Matrix Support Vector Machine Tiket.com

Pada pengujian confusion matrix yang dilakukan pada aplikasi Tiket.com dengan algoritma Support Vector Machine mendapatkan hasil akurasi sebesar 85%.



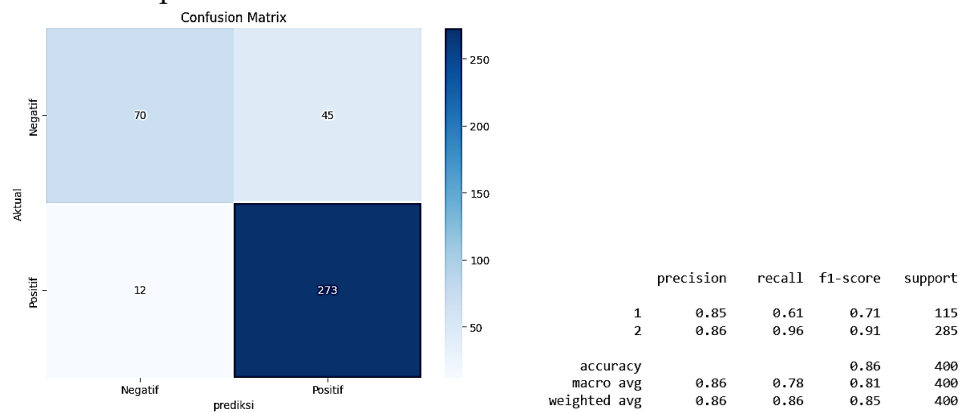
Gambar 4. Confusion Matrix Support Vector Machine Traveloka

Pada pengujian confusion matrix yang dilakukan pada aplikasi Traveloka dengan algoritma Support Vector Machine mendapatkan hasil akurasi sebesar 87%.



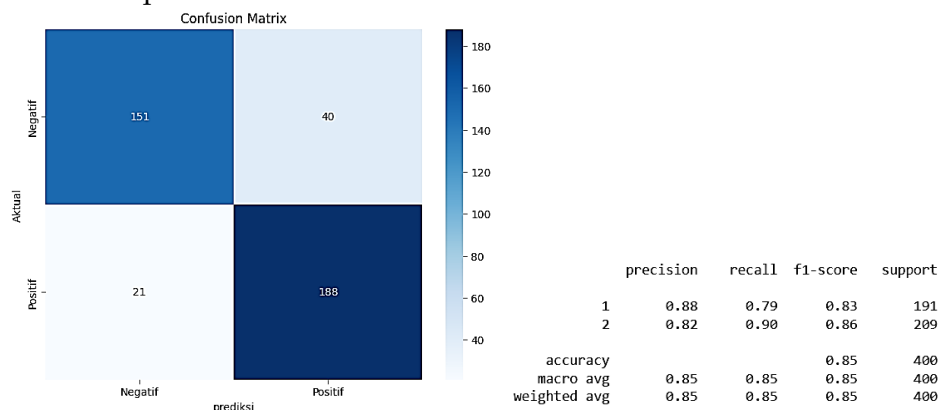
Gambar 5. Confusion Matrix Support Vector Machine Agoda

Pada pengujian confusion matrix yang dilakukan pada aplikasi Agoda dengan algoritma Support Vector Machine mendapatkan hasil akurasi sebesar 89%.



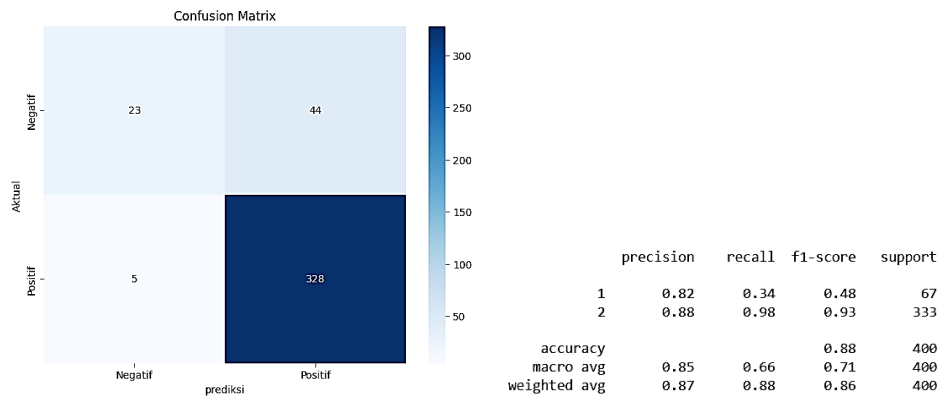
Gambar 6. Confusion Matrix Random Forest Tiket.com

Pada pengujian confusion matrix yang dilakukan pada aplikasi Tiket.com dengan algoritma Random Forest mendapatkan hasil akurasi sebesar 86%.



Gambar 7. Confusion Matrix Random Forest Traveloka

Pada pengujian confusion matrix yang dilakukan pada aplikasi Tiket.com dengan algoritma Random Forest mendapatkan hasil akurasi sebesar 86%.



Gambar 8. Confusion Matrix Random Forest Agoda

Pada pengujian confusion matrix yang dilakukan pada aplikasi Agoda dengan algoritma Random Forest mendapatkan hasil akurasi sebesar 88%.

3.5.3. Word Cloud

Tahap Word Cloud berfungsi untuk menampilkan kata-kata yang sering muncul pada kolom komentar Google Playstore setiap aplikasi baik kata positif dan negatif.



Gambar 9. Word Cloud positif Tiket.com



Gambar 10. Word Cloud negatif Tiket.com



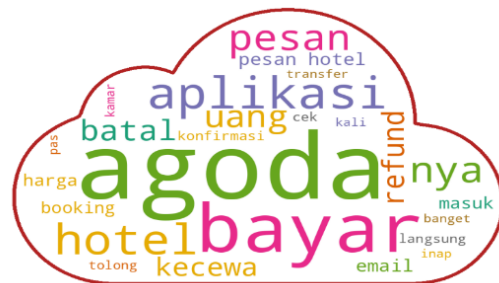
Gambar 12. Word Cloud positif Traveloka



Gambar 11. Word Cloud negatif Traveloka



Gambar 14. Word Cloud positif Agoda



Gambar 13. Word Cloud negatif Agoda

3.5.4. Visualiasi Power BI

Selain menggunakan word cloud visualisasi juga dilakukan menggunakan Power BI. Agar data bisa dilihat secara rapih dan tersusun, serta kita juga bisa melihat detail data baik dari keseluruhan, peraplikasi, dan juga persentimen.



Gambar 15. Visualisasi Power BI

4. KESIMPULAN

1. Dari hasil penelitian yang telah dilakukan terhadap aplikasi Tiket.com, Traveloka, dan Agoda. Data yang diambil dari Google Play dengan menggunakan Google Scraper.
2. Data sudah melalui beberapa tahapan proses baik proses preprocessing, pembobotan kata serta pengukuran akurasi menggunakan algoritma Support Vector Machine dan Random Forest.
3. Hasil akhir pengukuran menunjukan bahwa metode Support Random Machine lebih unggul dibandingkan dengan Random Forest. Dengan rincian akurasi Support Vector Machine pada Tiket.com sebesar 85.5% ,dan 84.7% pada Random Forest, pada aplikasi Traveloka yang menggunakan Support Vector Machine sebesar 87% dan 84.7% untuk Random Forest, yang terakhir aplikasi Agoda menggunakan Support Vector Machine 88.7% dan 88.2% pada Random Forest. Hasil ini didapat dengan menggunakan perbandingan data sebesar 80:20. Pada objek penelitian ini metode Support Vector Machine menunjukan keunggulan performa dibanding dengan metode Random Forest.

DAFTAR PUSTAKA

- [1] N. Luh, P. Riska, R. Paramita, W. Suardana, and I. M. Sendra, "Efektivitas Promosi Tiket.Com Terhadap Keputusan Wisatawan Domestik Dalam Pembelian Tiket Pesawat Menuju Bali", [Online]. Available: <https://www.similiarweb.com>
- [2] R. Islami, S. S. Hilabi, and A. Hananto, "Analisis User Experience Aplikasi Traveloka dan Tiket.Com Menggunakan Metode User Experience Quesionnaire," *remik*, vol. 7, no. 1, pp. 497–505, Jan. 2023, doi: [10.33395/remik.v7i1.12106](https://doi.org/10.33395/remik.v7i1.12106).

- [3] A. A. Carin, R. . Sund, and B. K. Lahkar, "Implementasi Naïve Bayes Classifier Dan Asosiasi Untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee Pada Situs Google Play Tugas," J. Control. Release, vol. 11, no. 2, pp. 430–439, 2018.
- [4] V. Fitriyana et al., "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," 2023.
- [5] R. I. Nurachim, "Pemilihan Model Prediksi Indeks Harga Saham Yang Dikembangkan Berdasarkan Algoritma Support Vector Machine(Svm) Atau Multilayer Perceptron(Mlp) Studi Kasus : Saham Pt Telekomunikasi Indonesia Tbk," 2019.
- [6] Y. Yuliani, "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst," Infotek J. Inform. dan Teknol., vol. 5, no. 2, pp. 298–306, Jul. 2022, doi: [10.29408/jit.v5i2.5896](https://doi.org/10.29408/jit.v5i2.5896).