JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i1.5746

PERBANDINGAN NAIVE BAYES, SUPPORT VECTOR MACHINE, LOGISTIC REGRESSION DAN RANDOM FOREST DALAM MENGANALISIS SENTIMEN MENGENAI TIKTOKSHOP

Octavia Salwa Dzaky Fadhillah[1], Jajam Haerul Jaman[2], Carudin[3]

[1,2,3] Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361; Telp. (0267) 641177

Received: 18 Desember 2024
Accepted: 14 Januari 2025
Published: 20 Januari 2025

Keywords:
Sentiment Analysis;
Confusion Matrix;
KDD;
Support Vector Machine;
Tiktokshop.

Corespondent Email:
octaviasalwa051002@gmail.com

Abstrak. Pertumbuhan *e-commerce* yang pesat di Indonesia dan ramainya pembicaraan salah satu platform yaitu Tiktokshop, mendorong pentingnya analisis sentimen untuk memahami tanggapan publik. Penelitian ini bertujuan menganalisis sentimen pengguna terhadap Tiktokshop melalui tweet di platform X, menggunakan algoritma *Naive Bayes*, *Support Vector Machine* (SVM), *Logistic Regression*, dan *Random Forest*. Data diambil melalui web scraping dan diproses menggunakan metodologi *Knowledge Discovery in Database* (KDD). Tahapan KDD meliputi *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining*, *Evaluation*, dan *Knowledge Presentation*. Label sentimen ditentukan dengan pendekatan lexicon, sehingga didapatkan 521 data label negatif dan 502 data label positif. Pengujian performa algoritma klasifikasi menggunakan *Confusion Matrix* dan *Classification Report*. Pengujian tersebut menghasilkan nilai akurasi tertinggi pada SVM sebesar 81%, diikuti *Random Forest* dengan 80%, *Logistic Regression* dengan 79%, dan *Naive Bayes* sebesar 75%. Visualisasi *word cloud* menunjukkan kata-kata dominan untuk sentimen positif seperti ’beli’, ’checkout’, ’barang’, ’murah’, dan ’suka’, sedangkan untuk sentimen negatif yaitu ’belanja’, ’live’, ’habis’ dan ’astaga’. Hasil penelitian ini diharapkan membantu perusahaan dalam mengevaluasi layanan dan strategi pemasaran Tiktokshop.

Abstract. *The rapid growth of e-commerce in Indonesia and the talk of one of the platforms, Tiktokshop, encourages the importance of sentiment analysis to understand public responses. This research aims to analyze user sentiment towards Tiktokshop through tweets on platform X, using Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and Random Forest algorithms. Data was collected through web scraping and processed using the Knowledge Discovery in Database (KDD) methodology. KDD stages include Data Selection, Preprocessing, Transformation, Data Mining, Evaluation, and Knowledge Presentation. Sentiment labels were determined using a lexicon approach, resulting in 521 negative label data and 502 positive label data. Classification algorithm performance testing using Confusion Matrix and Classification Report. The test resulted in the highest accuracy value in SVM of 81%, followed by Random Forest with 80%, Logistic Regression with 79%, and Naive Bayes with 75%. Word cloud visualization shows dominant words for positive sentiments such as 'buy', 'checkout', 'goods', 'cheap', and 'like', while for negative sentiments, they are 'shopping', 'live', 'run out' and 'golly'. The results of this study are expected to help companies evaluate Tiktokshop's services and marketing strategies.*

1. PENDAHULUAN

Kemajuan teknologi telah mendorong perkembangan pesat sektor e-commerce, termasuk di Indonesia yang diproyeksikan memiliki tingkat pertumbuhan tertinggi di dunia pada tahun 2024 sebesar 30,5% [1]. Salah satu platform yang menjadi sorotan adalah Tiktokshop, fitur belanja daring pada aplikasi Tiktok yang menggabungkan interaksi sosial dan transaksi online. Popularitas fitur ini meningkat pesat karena memungkinkan pembelian langsung melalui platform sekaligus menjaga interaksi antar pengguna. Setelah sempat ditutup akibat kebijakan Peraturan Menteri Perdagangan (Permendag) Nomor 31 Tahun 2023 yang melarang media sosial bertindak sebagai marketplace, Tiktokshop dibuka kembali pada Desember 2023, memicu diskusi luas di media sosial. Oleh karena itu, analisis sentimen terhadap opini pengguna menjadi langkah strategis untuk mengevaluasi dampak pembukaan kembali Tiktokshop.

Penelitian terkait analisis sentimen telah banyak dilakukan menggunakan algoritma seperti *Naive Bayes*, *Support Vector Machine* (SVM), *Logistic Regression*, dan *Random Forest*. Namun, sebagian besar penelitian tersebut tidak secara spesifik menargetkan fenomena Tiktokshop sebagai subjek utama. Selain itu, studi komparatif yang mengevaluasi performa berbagai metode algoritma untuk analisis sentimen Tiktokshop masih jarang ditemukan. Gap inilah yang menjadi dasar kebutuhan penelitian ini, yaitu memberikan kontribusi baru dengan melakukan analisis sentimen terhadap Tiktokshop menggunakan berbagai algoritma dan membandingkan performa masing-masing algoritma. Tujuan dari penelitian ini adalah untuk mengevaluasi opini publik terhadap pembukaan kembali Tiktokshop dengan menggunakan analisis sentimen berbasis algoritma *Naive Bayes*, *Support Vector Machine* (SVM), *Logistic Regression*, dan *Random Forest*. Penelitian ini juga bertujuan untuk mengidentifikasi metode yang paling efektif dalam memahami sentimen publik terkait perubahan layanan pada platform Tiktokshop. Dengan demikian, penelitian ini diharapkan dapat memberikan pandangan strategis bagi pengambil keputusan dalam mengelola fitur-fitur *e-commerce* berbasis media sosial.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan proses evaluasi opini, sikap, dan ekspresi terhadap suatu topik dengan mengklasifikasikan teks ke dalam kategori sentimen positif, negatif, atau netral. Proses ini melibatkan tahapan mulai dari pengumpulan data melalui media sosial atau sumber lain, pra-pemrosesan untuk membersihkan data, transformasi teks menjadi representasi numerik seperti TF-IDF, pemodelan menggunakan algoritma *machine learning* seperti *Naive Bayes* atau SVM, hingga klasifikasi dan visualisasi hasil. Pendekatan yang digunakan meliputi *supervised learning*, berbasis lexicon, atau *hybrid*. Pendekatan *supervised* bergantung pada data latih, sedangkan pendekatan berbasis lexicon menggunakan kamus sentimen untuk menentukan polaritas teks [2]. Kombinasi keduanya, atau pendekatan *hybrid*, menawarkan keunggulan seperti sensitivitas rendah terhadap perubahan domain dan kemampuan analisis pada tingkat konsep [3]. Dengan manfaatnya dalam memahami tren dan opini publik, analisis sentimen menjadi alat penting dalam mendukung pengambilan keputusan strategis di berbagai bidang, termasuk *e-commerce*.

2.2. Algoritma *Naive Bayes*

Algoritma *Naive Bayes* merupakan algoritma teknik klasifikasi yang menggunakan metode probabilitas dan statistic[4]. Menurut IEEE *International Conference on Data Mining* di Hongkong, metode klasifikasi pada algoritma ini cukup populer hingga masuk ke dalam kategori sepuluh algoritma teratas dalam data mining. Model *Multinomial Naive Bayes* akan melakukan klasifikasi pada dokumen dengan mempertimbangkan jumlah kemunculan term. Setiap dokumen akan diwakilkan oleh jumlah kemunculan term-term yang ada di dalamnya. Model ini menggunakan informasi tersebut untuk menentukan kelas berdasarkan distribusi term yang telah dilatih pada dataset. Berikut ini merupakan persamaan algoritma *Multinomial Naive Bayes*.

$$P(H \mid X) = \frac{p(H)\, P(X|H)}{P(X)} \quad (1)$$

2.3. Algoritma *Support Vector Machine*

Algoritma Support Vector Machine merupakan teknik machine learning yang cukup populer digunakan untuk mengklasifikasikan teks dikarenakan performa yang baik pada domain. SVM memiliki kemampuan untuk mengidentifikasi hyperplane secara terpisah di antara dua kelas yang berbeda membantu memaksimalkan jarak antara data yang paling dekat dengan *hyperplane*[5]. Berikut merupakan persamaan rumus *kernel linear* yang digunakan untuk menentukan nilai accuracy nya.

$$K(x_i x) = x_i^T x \quad (2)$$

2.4. Algoritma *Logistic Regression*

Algoritma *Logistic Regression* merupakan tipe analisis regresi yang bertujuan untuk menggambarkan hubungan antara variabel dependen dan variabel independen, mengaitkan satu atau lebih variabel bebas dengan variabel terikat dalam bentuk kategori seperti 0 dan 1, benar atau salah. Variabel bebasnya bersifat kategori. Inilah yang membedakan regresi logistik dari regresi berganda atau regresi linear lainnya[6]. Pada persamaan (3) merupakan persamaan *Logistic Regression*, dan persamaan untuk mencari peluang atau nilai p (Y=1) dapat digunakan rumus (4).

$$Ln(\frac{p}{1-p}) = B_0 + B_1 X \quad (3)$$

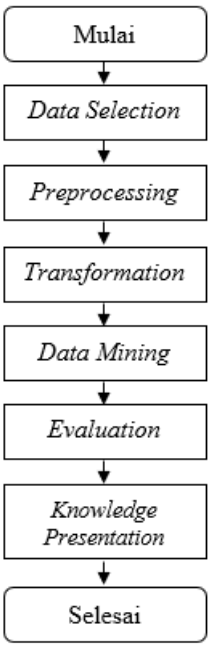
$$p = \frac{e(B_0 + B_1 X)}{1 - e(B_0 + B_1 X)} \quad (4)$$

2.5. Algoritma *Random Forest*

Algoritma *Random Forest* merupakan algoritma teknik pohon keputusan dari pengembangan metode CART (*Classification and Regression Trees*)[7]. Perbedaan antara metode random forest dan CART adalah penerapan teknik *bootstrap aggregating* (*bagging*) pada random forest dan juga seleksi fitur random yang dikenal *random feature selection*. *Random forest* terdiri dari 3 tahapan utama: pertama, melakukan *bootstrap sampling* untuk membangun pohon prediksi; kedua, masing-masing pohon keputusan akan melakukan prediksi menggunakan prediktor acak; dan ketiga, *random forest* akan menggabungkan hasil prediksi dari masing-masing pohon keputusan dengan cara *majority vote* untuk melakukan klasifikasi atau menghitung rata-rata regresi[8].

3. METODE PENELITIAN

Metodologi yang digunakan dalam penelitian ini yaitu metodologi penelitian *Knowledge Discovery in Database* (KDD). Metodologi KDD ini memiliki 6 tahapan pada prosesnya, yaitu *Data Selection, Preprocessing, Transformation, Data Mining, Evaluation* dan *Knowledge Presentation*.



Gambar 1 Alur Metodologi KDD

Objek dalam penelitian ini yaitu sentimen mengenai Tiktokshop berdasarkan *tweet* di platform X. Data yang digunakan berupa hasil *web scraping* dari X menggunakan Bahasa Pemrograman Python yang akan diklasifikasikan ke dalam label positif, serta negatif. Data yang diambil sejak Desember 2023 hingga Juni 2024 menggunakan Bahasa Indonesia.

Data hasil *web scraping* tersebut akan diolah menggunakan KDD sesuai dengan tahapan-tahapannya. Rancangan penelitian berdasarkan tahapan-tahapan metodologi penelitian KDD diatas, yaitu:

3.1. *Data Selection*

Tahap ini adalah tahap pengambilan data dan pemilihan atribut data yang diperlukan untuk mendapatkan pandangan pengguna mengenai Tiktokshop melalui analisis sentimen. Data yang diambil merupakan tweet pada platform X sejak bulan Desember 2023 hingga bulan Juni 2024 dengan kata kunci 'Tiktokshop', setelah itu baru dilakukan pemilihan atribut yang diperlukan untuk analisis sentimen. Data diambil menggunakan teknik web scraping pada platform X yang dilakukan pada Google Colab menggunakan Bahasa Pemrograman Python.

3.2. *Preprocessing*

Tahap *preprocessing* adalah tahap membersihkan data yang telah dikumpulkan. Tujuannya supaya data yang lebih bersih dan terstruktur sehingga siap untuk dilakukan analisis. Tahap ini memiliki beberapa tahapan kecil lagi, yaitu *Cleaning, Case Folding, Tokenizing, Normalizing, Filtering,* dan *Stemming*[9].

1. *Cleaning*

Tahap ini merupakan salah satu proses untuk mengurangi *noise* pada data, seperti URL, HTML, emoji, angka, username, dan symbol dari teks.

2. *Case Folding*

Tahap ini merupakan tahap untuk memastikan konsistensi data dengan mengubah semua huruf di dalam teks menjadi huruf kecil.

3. *Tokenizing*

Tahap ini merupakan tahap pemisahan setiap kata dalam suatu kalimat utuh untuk mempermudah analisis teks dan proses selanjutnya.

4. *Normalizing*

Tahap ini merupakan tahap mengubah kata yang tidak baku menjadi baku. Tujuannya untuk mengurangi variasi kata sehingga meningkatkan accuracy model.

5. *Filtering*

Tahap ini merupakan tahap penghapusan kata-kata umum yang tidak membawa informasi penting sehingga dapat lebih fokus menganalisis pada kata-kata yang lebih bermakna.

6. *Stemming*

Tahap ini merupakan tahap mengubah suatu kata imbuhan menjadi bentuk dasar sehingga mengurangi keragaman dalam teks yang bisa mengganggu analisis atau pemodelan data dan analisis teks menjadi lebih efektif dan akurat.

3.3. *Transformation*

Tahap *transformation* adalah tahap penyesuaian data dengan mengonversi teks menjadi representasi numerik, yang disebut pembobotan kata. Pada tahap ini juga dilakukan labeling dan pembagian data sehingga data telah siap untuk dilakukan pemodelan.

1. Labeling

Tahap ini dilakukan dengan pendekatan lexicon VADER (*Valence Aware Dictionary and Sentiment Reasoner*) untuk menentukan label sentimen dari setiap data, baik sentimen positif ataupun negatif[10].

2. Pembagian Data

Tahap ini menggunakan TF-IDF (*Term Frequency Inverse Document Frequency*) dan Prinsip Pareto (*Pareto principle*) yang mengatakan bahwa perbandingan rasio yang biasa digunakan yaitu 80%:20% untuk data latih dan data uji.

3. Pembobotan Kata

Tahap ini dilakukan juga oleh TF-IDF (*Term Frequency Inverse Document Frequency*) sehingga teks yang berhasil diubah menjadi representasi numerik, siap untuk dilakukan pemodelan pada masing-masing algoritma.

3.4. *Data Mining*

Pada tahap ini, data yang telah dibagi akan dilakukan pemodelan klasifikasi. Klasifikasi pada penelitian ini menggunakan penerapan algoritma *Naive Bayes, Support Vector Machine (SVM), Logistic Regression dan Random Forest* untuk membandingkan kinerja terbaik dalam klasifikasi sentimen. Proses klasifikasi ini masih menggunakan library Scikit-learn.

1. Algoritma *Naive Bayes*

Penerapan Algoritma *Naive Bayes* ini, menggunakan fungsi *MultinomialNB*() pada library sklearn.naive_bayes.

2. Algoritma *Support Vector Machine*

Penerapan Algoritma SVM ini, menggunakan fungsi SVC pada library sklearn.svm.

3. Algoritma *Logistic Regression*

Penerapan Algoritma *Logistic Regression* ini, menggunakan fungsi LogisticRegression pada library sklearn.linear_model.

4. Algoritma *Random Forest*

Penerapan Algoritma *Random Forest* ini, kelas yang digunakan yaitu RandomForestClassifier pada library sklearn.ensemble.

3.5. *Evaluation*

Tahap *evaluation* adalah tahap menentukan model yang paling efektif dalam mengklasifikasi. Evaluasi klasifikasi ini menggunakan *Confusion Matrix* untuk membandingkan antara data prediksi dengan data sebenarnya dan *Classification Report* digunakan untuk menganalisis nilai hasil pengujian klasifikasi atau menentukan model terbaik dengan menghitung nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari setiap model[11].

3.6. *Knowledge Presentation*

Tahap *Knowledge Presentation* yaitu tahap akhir yang menyajikan hasil analisis sentimen melalui visualisasi data. Visualisasi data ini menggunakan library *Word Cloud* untuk mendapatkan informasi yang dapat dijadikan sebagai bahan evaluasi oleh perusahaan[12].

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini yaitu bagaimana melakukan analisis sentimen terkait Tiktokshop pada platform X untuk mengidentifikasi, menganalisis, dan menginterpretasikan dari *tweet* pengguna, dan membandingkan hasil *accuracy* algoritma *Naive Bayes*, *Support Vector Machine* (SVM), *Logistic Regression dan Random Forest* dengan menggunakan *Confusion Matrix* dan *Classification Report* untuk mengetahui algoritma mana yang lebih baik. Klasifikasi data akan dibagi dua kategori, yaitu positif dan negatif. Untuk selengkapnya, hasil penelitian ini dibagi menjadi beberapa tahap yaitu sebagai berikut.

4.1. Data Selection

Tahap ini adalah tahap pengumpulan data dan pemilihan data. Pengumpulan data penelitian ini menggunakan teknik *web scraping* pada platform X. Proses pengambilan data dilakukan dari bulan Desember 2023 hingga bulan Juni 2024, dengan kata kunci "Tiktokshop". Data hasil *scraping* yang didapatkan terdiri atas 3300 data dengan 15 atribut, namun hanya kolom *full_text* yang digunakan karena berisi informasi inti berupa *tweet* pengguna terkait Tiktokshop.

Gambar 2 Data Hasil *Web Scraping*

4.2. Preprocessing

Data *full_text* masih berupa teks yang tidak terstruktur karena tingginya tingkat *noise* di dalam teks. Oleh karena itu, data akan dibersihkan dan disesuaikan terlebih dahulu pada tahap *preprocessing* ini sehingga lebih bersih dan terstruktur sebelum diolah pada tahap berikutnya.

Preprocessing ini memiliki beberapa tahapan kecil lagi, yaitu *cleaning, case folding, tokenizing, normalization, filtering,* dan *stemming*. Setelah dilakukan *preprocessing*, didapatkan sekitar 1.023 data yang siap diolah pada tahap berikutnya.

Tabel 1 Hasil *Cleaning*

full_text	*cleaning*
@_7ri1ngzz kmrn gua beliin cowo gua bagus di tiktokshop	kmrn gua beliin cowo gua bagus di tiktokshop

Tabel 2 Hasil *Case Folding*

cleaning	*case folding*
kmrn gua beliin cowo gua bagus di tiktokshop	kmrn gua beliin cowo gua bagus di tiktokshop

Tabel 3 Hasil *Tokenizing*

case folding	*tokenizing*
kmrn gua beliin cowo gua bagus di tiktokshop	['kmrn', 'gua', 'beliin', 'cowo', 'gua', 'bagus', 'di', 'tiktokshop']

Tabel 4 Hasil *Normalizing*

tokenizing	*normalizing*
['kmrn', 'gua', 'beliin', 'cowo', 'gua', 'bagus', 'di', 'tiktokshop']	['kemarin', 'saya', 'belikan', 'laki-laki', 'saya', 'bagus', 'di', 'tiktokshop']

Tabel 5 Hasil *Filtering*

normalizing	*filtering*
['kemarin', 'saya', 'belikan', 'laki-laki', 'saya', 'bagus', 'di', 'tiktokshop']	['kemarin', 'belikan', 'laki-laki', 'bagus', 'tiktokshop']

Tabel 6 Hasil *Stemming*

filtering	*stemming*
['kemarin', 'belikan', 'laki-laki', 'bagus', 'tiktokshop']	['kemarin', 'belikan', 'laki', 'bagus', 'tiktokshop']

4.3. Transformation

Tahap *transformation* adalah tahap penyesuaian data dengan mengonversi teks menjadi representasi numerik, yang disebut pembobotan kata. Pada tahap ini juga dilakukan labeling dan pembagian data sehingga data telah siap untuk dilakukan pemodelan. Hasil tahap *transformation* yaitu sebagai berikut.

Tabel 7 Hasil Pelabelan

Label	Jumlah
Positif	502
Negatif	521
Total	1023

Tabel 8 Hasil Pembagian Data

Rasio / Data	Latih	Uji
80 : 20	818	205

4.4. Data Mining

Tahap *data mining* adalah tahap pemodelan klasifikasi berdasarkan algoritma. Data yang digunakan yaitu data latih dan data uji.

1. Algoritma *Naive Bayes*

Pada pemodelan algoritma *naive bayes* ini, data diolah kedalam model *Multinomial Naive Bayes* untuk melakukan perhitungan probabilitas.

2. Algoritma *Support Vector Machine*

Pada pemodelan algoritma SVM ini, data diolah kedalam model SVC untuk melakukan pemisahan 2 kelas dengan garis lurus pemisah menggunakan perhitungan keputusan model (*hyperlane*).

3. Algoritma *Logistic Regression*

Pada pemodelan algoritma *logistic regression* ini, data diolah kedalam model *LogisticRegression* untuk untuk menggambarkan hubungan antara variabel dependen dan variabel independen, mengaitkan satu atau lebih variabel bebas (dalam bentuk kategori) dengan variabel terikat menggunakan perhitungan bobot dari setiap fitur dalam model.

4. Algoritma *Random Forest*

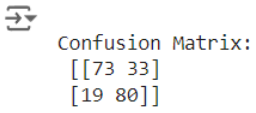
Pada pemodelan algoritma *random forest* ini, data diolah kedalam model *RandomForestClassifier* untuk mendapatkan hasil akhir pohon keputusan.

4.5. Evaluation

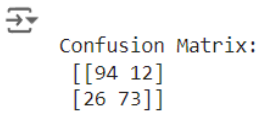
Tahap evaluasi ini bertujuan untuk menentukan model yang paling efektif dalam mengklasifikasi. Melakukan pengujian performa klasifikasi ini dengan metode *confusion matrix* dan *classification report*.

1. Confusion Matrix

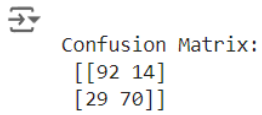
Confusion Matrix dilakukan untuk membandingkan antara data prediksi dengan data sebenarnya, kemudian hasil perhitungan direpresentasikan dengan tabel matrik.



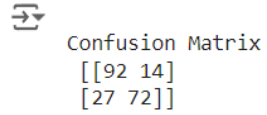
Gambar 3 *Confusion Matrix Naive Bayes*



Gambar 4 *Confusion Matrix* SVM



Gambar 5 *Confusion Matrix Logistic Regression*



Gambar 6 *Confusion Matrix Random Forest*

2. Classification Report

Classification Report digunakan untuk menganalisis nilai hasil pengujian klasifikasi atau *confusion matrix*, serta menentukan model terbaik dengan menghitung nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dari setiap model.

Tabel 9 Hasil Perhitungan *Classification Report*

/	*NB*	*SVM*	*LR*	*RF*
Accuracy	0,75	**0,81**	0,79	0,80
Precision	0,71	**0,86**	0,83	0,84
Recall	**0,81**	0,74	0,71	0,73
F1-Score	0,78	**0,79**	0,77	0,78

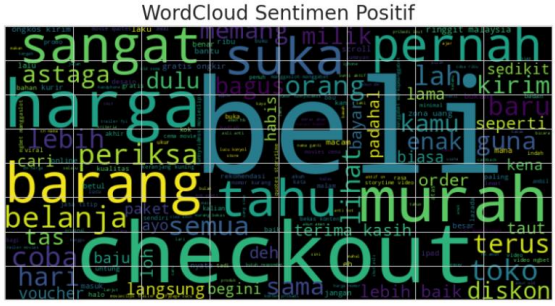
Keterangan:
NB = *Naive Bayes*
SVM = *Support Vector Machine*
LR = *Logistic Regression*
RF = *Random Forest*

Berdasarkan tabel hasil perhitungan *classification report* diatas, didapatkan informasi urutan algoritma-algoritma unggul yang berbeda berdasarkan perhitungan masing-masing alat ukurnya.

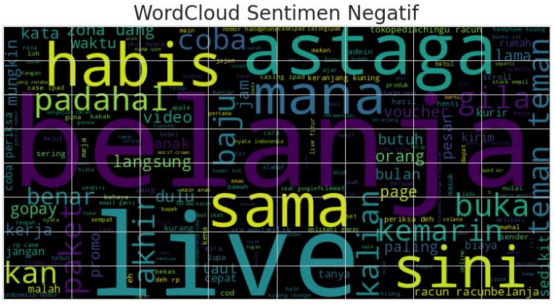
- ***Accuracy*** **:** *SVM, RF, LR,* dan *NB.*
- ***Precision*** **:** *SVM, RF, LR,* dan *NB.*
- ***Recall*** **:** *NB, SVM, RF,* dan *LR.*
- ***F1*****-Score :** *SVM, NB, RF,* dan *LR.*

4.6. Knowledge Presentation

Tahap *Knowledge Presentation* ini menggunakan *Word Cloud*. *Word Cloud* merupakan library Python yang visualisasi data berbentuk teks berdasarkan kata paling banyak muncul dalam dokumen. Visualisasi data ini dibagi menjadi 2 berdasarkan label, yaitu positif dan negatif.



Gambar 7 *Word Cloud* Sentimen Positif



Gambar 8 *Word Cloud* Sentimen Negatif

Berdasarkan gambar diatas, diketahui kata yang paling banyak muncul pada label sentimen positif yaitu ’beli’, ’checkout’, ’barang’, ’murah’, ’suka’, dan lain-lain. Sedangkan pada label sentimen negatif, kata yang paling banyak muncul adalah ’belanja’, ’*live*’, ’habis’, ’astaga’, dan lain-lain.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, beberapa hal dapat disimpulkan sebagai berikut:

a. Analisis sentimen mengenai Tiktokshop pada platform media sosial X menggunakan algoritma *Naive Bayes*, *Support Vector Machine*, *Logistic Regression* dan *Random Forest*. Analisis ini menggunakan *tools* Google Colab dengan Bahasa Pemrograman Python mulai dari pengumpulan data hingga visualisasi data. Pengumpulan data dilakukan dengan teknik *web scraping* pada platform X dengan kata kunci ’Tiktokshop’ dan rentang waktu dari bulan Desember 2023 hingga bulan Juni 2024. Jumlah data yang berhasil dikumpulkan dan dibersihkan yaitu 1.023 data, lalu data yang berlabel sentimen negatif ada 521 data dan 502 data berlabel sentimen positif. Selanjutnya pemodelan data sesuai algoritma masing-masing dengan menggunakan data yang telah dibagi 80% untuk data latih dan 20% untuk data uji. Pengujian performa dari masing-masing algoritma dilakukan untuk mengetahui algoritma mana yang lebih baik pada penelitian ini dengan membandingkan nilai *Accuracy*. Visualisasi data berdasarkan kata yang paling sering muncul, pada sentimen positif yaitu kata ’beli’, ’checkout’, ’barang’, ’murah’, dan ’suka’, sedangkan pada sentimen negatif yaitu kata ’belanja’, ’*live*’, ’habis’ dan ’astaga’.
b. Perbandingan algoritma *Naive Bayes*, *Support Vector Machine*, *Logistic Regression* dan *Random Forest* dalam penelitian ini menggunakan nilai *accuracy* dari hasil pengujian performa. Berdasarkan hasil pengujian performa dari masing-masing algoritma diatas, didapatkan nilai *Accuracy* dari yang terbesar yaitu algoritma SVM sebesar 0,81, algoritma *Random Forest* sebesar 0,80, algoritma *Logistic Regression* sebesar 0,79 dan algoritma *Naive Bayes* sebesar 0,75.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan dan bantuan terhadap penelitian ini, terutama dosen pembimbing saya yang telah membimbing saya dalam melakukan penelitian.

DAFTAR PUSTAKA

[1] E. F. Santika, "ECDB: Proyeksi Pertumbuhan e-commerce Indonesia Tertinggi Sedunia pada 2024," Katadata.co.id; Databoks. Available: https://databoks.katadata.co.id/datapublish/2024/04/29/ecdb-proyeksi-pertumbuhan-e-commerce-indonesia-tertinggi-sedunia-pada-2024.

[2] A. Syakur, "Implementasi Metode Lexicon Base Untuk Analisis Sentimen Kebijakan Pemerintah Dalam Pencegahan Penyebaran Virus Corona COVID-19 Pada Twitter," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 3, pp. 247-260, 2021.

[3] D. R. Ramadhanty, "Implementasi algoritma support vector machine pada analisis sentimen data twitter (Studi kasus: ulasan tentang indohome)," *Universitas Islam Indonesia*, vol. 16, pp. 1-83, 2021. Available: https://dspace.uii.ac.id/handle/123456789/36015.

[4] A. Sitanggang, Y. Umaidah, and R. I. Adam, "Analisis sentimen masyarakat terhadap program makan siang gratis pada media sosial X menggunakan algoritma Naïve Bayes," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, pp. 1-10, 2023, pISSN: 2303-0577, eISSN: 2830-7062. doi: 10.23960/jitet.v12i3.4902. Available: http://dx.doi.org/10.23960/jitet.v12i3.4902.

[5] A. Maulana, N. Afifah, I. K. No, N. A. Mubarrak, N. K. R. Fauzan, N. A. Dwintara, and B. P. Zen, "Comparison of Logistic Regression, MultinomialNB, SVM, and K-NN Methods on Sentiment Analysis of Gojek App Reviews on the Google Play Store," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 6, pp. 1487-1494, 2023.

[6] A. Maulana, N. Afifah, I. K. No, N. A. Mubarrak, N. K. R. Fauzan, N. A. Dwintara, and B. P. Zen, "Comparison of Logistic Regression, MultinomialNB, SVM, and K-NN Methods on Sentiment Analysis of Gojek App Reviews on the Google Play Store," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 6, pp. 1487-1494, 2023.

[7] A. Syah, F. Nurdiyansyah, and A. Y. Rahman, "Analisis sentimen aplikasi Shopee, Tokopedia, Lazada dan Blibli menggunakan leksikon dan Random Forest," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, pp. S1-10, 2023, pISSN: 2303-0577, eISSN: 2830-7062. doi: 10.23960/jitet.v12i3S1.5155. Available: http://dx.doi.org/10.23960/jitet.v12i3S1.5155.

[8] F. Azimah and K. R. N. Wardani, "Klasifikasi Deteksi Gejala Awal COVID-19 Dengan Metode Logistic Regression, Random Forest Classifier dan Support Vector Machine," *Jurnal Locus: Penelitian dan Pengabdian*, vol. 1, no. 9, 2022.

[9] S. Nadhifah, F. N. Aini, H. H. Kusumawardhani, and M. Y. Febrianto, "Analisis sentimen ulasan aplikasi Gopay pada Google Play Store menggunakan algoritma Support Vector Machine," *Jurnal Surya Informatika*, vol. 14, no. 1, pp. 1-6, 2024.

[10] A. Syah, F. Nurdiyansyah, and A. Y. Rahman, "Analisis sentimen aplikasi Shopee, Tokopedia, Lazada dan Blibli menggunakan leksikon dan Random Forest," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, pp. S1-10, 2023, pISSN: 2303-0577, eISSN: 2830-7062. doi: 10.23960/jitet.v12i3S1.5155. Available: http://dx.doi.org/10.23960/jitet.v12i3S1.5155.

[11] M. S. Alrajak, I. Ernawati, and I. Nurlaili, "Analisis sentimen terhadap pelayanan PT PLN di Jakarta pada Twitter dengan algoritma k-nearest neighbor (k-NN)," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, vol. 1, no. 2, pp. 110-122, 2020.

[12] E. Suryati, A. Aldino, N. Penulis Korespondensi, and E. Suryati, "Analisis sentimen transportasi online menggunakan ekstraksi fitur model Word2Vec text embedding dan algoritma Support Vector Machine (SVM)," *J. Teknol. Sist. Inf.*, vol. 4, no. 1, pp. 96-106, 2023. doi: 10.33365/jtsi.v4i1.2445. Available: https://doi.org/10.33365/jtsi.v4i1.2445.