

PERBANDINGAN KINERJA SVM DAN LOGISTIC REGRESSION UNTUK KLASIFIKASI PRODUK E-COMMERCE

PENULIS

¹⁾Firtsly Shures Jeffryson, ²⁾Sari Ningsih.

ABSTRAK

Pengelompokan produk pada E-commerce dapat dilakukan dengan memanfaatkan informasi yang terdapat pada nama produk. Namun, perbedaan penulisan nama produk dapat menyebabkan proses klasifikasi menjadi lebih sulit. Penelitian ini bertujuan untuk menerapkan serta membandingkan kinerja Support Vector Machine (SVM) dan Logistic Regression dalam mengklasifikasikan produk Tokopedia berdasarkan nama produk. Data yang digunakan sebanyak 3.000 produk yang terbagi dalam tiga kategori, yaitu Elektronik, Fashion, dan Makanan. Data nama produk diproses melalui Text Preprocessing, kemudian diberi bobot menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Hasil pembobotan digunakan sebagai fitur pada SVM dengan Linear Kernel dan Logistic Regression. Pengujian dilakukan dengan pembagian data 80:20 dan 70:30, kemudian dievaluasi menggunakan Accuracy, Precision, Recall, dan F1-Score. Hasil pengujian menunjukkan bahwa pada pembagian data 80:20, kedua algoritma memperoleh Accuracy sebesar 97,40%. Pada pembagian 70:30, SVM memperoleh Accuracy sebesar 96,18%, sedangkan Logistic Regression memperoleh 96,88%. Berdasarkan hasil tersebut, kedua algoritma memiliki kinerja yang sama pada skenario 80:20, sedangkan Logistic Regression memberikan hasil lebih tinggi pada skenario 70:30.

Kata Kunci

Klasifikasi Teks; Kategori Produk; TF-IDF; Support Vector Machine; Logistic Regression;

AFILIASI

Prodi, Fakultas

^{1,2)} Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika

Nama Institusi

^{1,2)} Universitas Nasional

Alamat Institusi

^{1,2)} Jl.Sawo Manila No.61, RT.14/RW.7, Pejaten Bar.,Ps.Minggu, Jakarta Selatan, Daerah Khusus Ibukota Jakarta.

KORESPONDENSI

Penulis

Sari Ningsih

Email

lectures.sariningsih@gmail.com ; andinihumaira0708@gmail.com

LICENSE



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

I. PENDAHULUAN

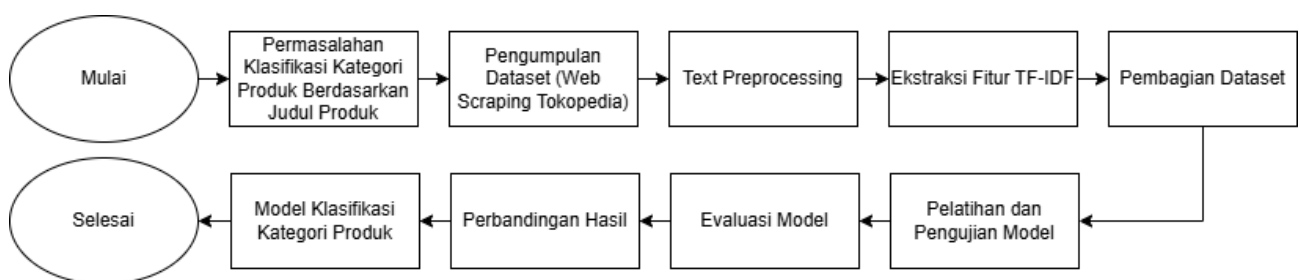
Perkembangan *E-commerce* telah menghasilkan jumlah data produk yang besar dan beragam, salah satunya berupa nama atau judul produk yang memiliki karakteristik teks pendek dan tidak terstruktur. Perbedaan penggunaan kata, singkatan, istilah, serta gaya penulisan oleh penjual menyebabkan data tersebut perlu diolah sebelum digunakan dalam klasifikasi. Klasifikasi teks dapat membantu mengelompokkan produk secara otomatis berdasarkan informasi yang terkandung dalam teks [11][12]. Dalam pengolahan teks, *Natural Language Processing* (NLP) dan *Text Preprocessing* dapat digunakan untuk membersihkan dan menormalisasi data sebelum dilakukan ekstraksi fitur [4][7]. Selanjutnya, metode TF-IDF dapat digunakan untuk mengubah teks menjadi representasi numerik yang dapat dimanfaatkan sebagai fitur dalam proses klasifikasi [1][2][11].

Beberapa penelitian menunjukkan bahwa SVM dan *Logistic Regression* dapat digunakan untuk klasifikasi data *E-commerce*. SVM memiliki kemampuan yang baik dalam menangani data teks dengan jumlah fitur yang tinggi dan representasi yang *sparse*, sedangkan *Logistic Regression* dapat digunakan untuk menentukan probabilitas suatu data terhadap kelas tertentu [5][11][12]. Penelitian Aditama et al. menunjukkan bahwa SVM dengan TF-IDF mampu memberikan performa tinggi dalam klasifikasi produk *marketplace* [1,3], sementara Patra et al. menunjukkan bahwa *Logistic Regression* juga dapat diterapkan pada klasifikasi produk *E-commerce* [12]. Namun, performa algoritma dapat berbeda bergantung pada karakteristik dataset dan skenario pembagian data yang digunakan. Oleh karena itu, perbandingan SVM dan *Logistic Regression* pada dataset yang sama diperlukan untuk mengetahui metode yang memberikan kinerja lebih baik [6][9][17].

Penelitian ini berfokus pada klasifikasi 3.000 nama produk Tokopedia ke dalam tiga kategori, yaitu Elektronik, Fashion, dan Makanan, dengan masing-masing kategori terdiri atas 1.000 data. Nama produk digunakan sebagai fitur teks yang diproses dan direpresentasikan menggunakan TF-IDF, kemudian diklasifikasikan menggunakan SVM dengan *Linear Kernel* dan *Logistic Regression*. Pengujian dilakukan dengan dua tipe pembagian data, yaitu 80:20 dan 70:30, sedangkan evaluasi model menggunakan *Confusion Matrix*. Penelitian ini dapat merepresentasikan nama produk serta membandingkan kinerja SVM dan *Logistic Regression* dalam mengklasifikasikan kategori produk Tokopedia secara otomatis. Hasil evaluasi performa kedua algoritma ditampilkan melalui aplikasi berbasis Streamlit, sehingga proses analisis dan perbandingan akurasi dapat disajikan secara interaktif dan mudah dipahami.

II. METODE PENELITIAN

Digram alur kerangka pemikiran dari setiap tahapan pada penelitian ini sebagai berikut: Gambar 1 menggambarkan tahapan diagram alur kerangka pemikiran.



Gambar 1. Diagram Alur Kerangka Pemikiran

2.1 Tahapan Pendahuluan

Tahapan ini merupakan langkah awal penelitian yang diawali dengan permasalahan klasifikasi kategori produk Tokopedia berdasarkan judul produk yang masih berbentuk teks tidak terstruktur. Data dikumpulkan melalui *web scraping*, kemudian diolah menggunakan pendekatan NLP untuk menghasilkan teks yang lebih terstruktur [4][7]. Selanjutnya, teks diubah menjadi representasi numerik menggunakan TF-IDF sebagai fitur masukan algoritma klasifikasi [1][5][11]. *Dataset* kemudian dibagi menjadi data latih dan data uji dengan skenario 80:20 dan 70:30, selanjutnya dilakukan klasifikasi menggunakan SVM dan *Logistic Regression* [11][12]. Kinerja kedua algoritma dievaluasi menggunakan *Confusion Matrix* dan *Classification Report* kemudian hasilnya dibandingkan untuk menentukan algoritma yang memberikan performa terbaik dalam klasifikasi kategori produk Tokopedia [1][6][17].

2.2 Pengumpulan Data

Proses pengumpulan data dimulai dengan menentukan kategori produk yang akan dijadikan objek penelitian. Setelah kategori ditentukan, tahap berikutnya adalah melakukan proses *web scraping* pada halaman produk Tokopedia menggunakan *Web Scraper* dan *Instant Data Scraper*.

Tabel 1. Dataset

id	nama_produk	harga	kota	toko	kategori
1	Ac portable 1 pk promo midea ac portabel 1pk mrf2-09cm1 mrf09cm1	Rp4.399.000	Kab. Bogor	Bogor Solusi Elektronik	Elektronik
2	?COD?Tixx 40L Pendingin Udara Air Portable Kipas Angin Dingin Super Dingin Hemat Energi 3 Kecepatan Mudah Dibawa Cocok untuk Rumah Kantor Kafe Usaha Kualitas Tinggi Tahan Lama Pendingin Ruangan Terbaik Hemat Daya Ringkas Praktis	Rp719.200	Tangerang	Tixx MegaElectron	Elektronik
3	TITON Freeze 1 Mini Kipas Refrigeration Fan Speed Digital Display Adjustable Turbo Kipas Tangan Handheld	Rp182.775	Jakarta Pusat	TITON Indonesia	Elektronik
...	...	...	...	...	...
2998	[EDISI TERBATAS x TAHILALATS] Crystalin Air Minum Mineral 600ml - [1 Botol]	Rp4.250	Jakarta Barat	OT-Store_NEW	Makanan
2999	Vit air mineral btl 1500 gr	Rp4.910	Denpasar	Primo Supermarket level 21	Makanan
3000	Nestle pure life 600 ml - air mineral	Rp3.730	Bandung	Hyfresh Bandung Indah Plaza	Makanan

Berdasarkan Tabel 1, dataset terdiri atas 3.000 data produk dengan enam atribut, yaitu id, nama_produk, harga, kota, toko, dan kategori. Adapun penjelasan dari setiap atribut pada dataset:

1. *id* merupakan kode identitas unik yang digunakan untuk membedakan setiap data produk.
2. *nama_produk* berisi informasi judul produk.
3. *harga* menunjukkan nilai harga dari masing-masing produk.
4. *kota* menunjukkan lokasi asal penjual produk.
5. *toko* berisi informasi mengenai nama toko atau penjual produk.
6. *kategori* merupakan label kelas yang digunakan sebagai target klasifikasi, yaitu Elektronik, Fashion, dan Makanan.

2.3 Text Preprocessing

Tahap *text preprocessing* dilakukan untuk menyiapkan data nama produk hasil *web scraping* agar lebih seragam dan siap digunakan dalam ekstraksi fitur TF-IDF. Proses ini terdiri atas lima tahapan, yang diimplementasikan menggunakan Python pada Google Colab dengan bantuan pustaka re, NLTK, dan Sastrawi. Setiap tahapan menghasilkan data yang lebih terstruktur hingga diperoleh teks akhir yang siap digunakan dalam proses klasifikasi. Berdasarkan fitur terpilih, dibangun dua model menggunakan SVM dan *Logistic Regression*, lalu dievaluasi dengan metrik akurasi, presisi, recall, dan F1-score. Model kemudian diintegrasikan ke dalam aplikasi berbasis Streamlit.[11],[12].[13].

2.4 Ekstraksi Fitur Pembobotan Kata (TF-IDF)

Setelah melalui *text preprocessing*, sebanyak 2.877 data judul produk dari kategori Elektronik, Fashion, dan Makanan ditransformasikan ke dalam bentuk numerik menggunakan metode TF-IDF, yang memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan keseluruhan dataset [1][5][11]. Setiap judul produk diperlakukan sebagai satu dokumen, kemudian proses ekstraksi fitur dilakukan menghasilkan matriks numerik sebagai representasi fitur teks. Matriks TF-IDF digunakan sebagai masukan dalam proses pelatihan dan pengujian model SVM dan *Logistic Regression* [11][12].

2.5 Streamlit

Pada tahap ini, Streamlit digunakan sebagai media untuk menyajikan hasil analisis dan prediksi dari model yang telah diperoleh. Model dengan performa terbaik, yaitu SVM atau *Logistic Regression*, diintegrasikan ke dalam Streamlit untuk menampilkan hasil klasifikasi secara interaktif. Informasi yang disajikan meliputi nilai akurasi, hasil prediksi, dan visualisasi data dengan tampilan yang informatif serta mudah dipahami. visualisasi hasil penelitian melalui antarmuka Streamlit. [14],[15].

2.6 Proses Evaluasi

Tahap ini merupakan proses evaluasi terhadap hasil analisis dan prediksi yang telah ditampilkan menggunakan Streamlit. [16], Pengujian dilakukan untuk memastikan bahwa model prediksi, seperti SVM atau *Logistic Regression*, memberikan hasil yang akurat dan konsisten. Selanjutnya, hasil pengujian dianalisis untuk menilai tingkat keakuratan dan efektivitas model dalam memprediksi risiko penyakit jantung. Berdasarkan analisis tersebut, disusun kesimpulan mengenai performa model serta rekomendasi untuk pengembangan penelitian di masa mendatang[18],[19],[20].

2.7 Pengolahan Data

Langkah sistematis yang dilakukan untuk mempersiapkan, mengolah, dan menganalisis data menggunakan algoritma SVM dan algoritma *Logistic Regression*.

1. Algoritma SVM adalah algoritma *supervised learning* yang digunakan untuk tugas klasifikasi dan regresi, baik pada data linear maupun non-linear. Prinsip kerjanya adalah menentukan hyperplane terbaik yang memisahkan antar kelas dengan memaksimalkan margin antara decision boundary dan titik data terdekat. [7],[8],[9],[10].
2. Algoritma *Logistic Regression* adalah metode machine learning yang digunakan untuk klasifikasi dengan mengestimasi probabilitas suatu data masuk ke kelas tertentu melalui fungsi logistik[17]. Dalam penelitian ini, *Logistic Regression* digunakan sebagai metode pembandingan terhadap SVM dalam mengklasifikasikan kategori produk Tokopedia. Proses klasifikasi dilakukan dengan memanfaatkan kombinasi linear antara fitur, bobot (*weight*), dan bias untuk menghasilkan nilai *logit* sebagai dasar penentuan kelas.

$$z = (w \cdot x) + b \quad (1)$$

$$p = \frac{1}{1+e^{-z}} \quad (2)$$

2.8 Confusion matrix

adalah tabel untuk menilai kinerja model klasifikasi dengan cara membandingkan hasil prediksi model terhadap label sebenarnya dalam dataset[23]. Matriks ini memberikan informasi tentang jumlah prediksi yang benar maupun salah, serta jenis kesalahan yang terjadi dalam proses klasifikasi[24],[25].

Tabel 2. Confusion matrix

Prediksi Sesuai	TP	FP
Prediksi Tidak Sesuai	FN	TN

Metrik tabel 2. ini memberikan gambaran seberapa baik model dalam mengklasifikasikan data dengan benar serta seberapa efektif dalam menangani kesalahan prediksi.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \times 100\% \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

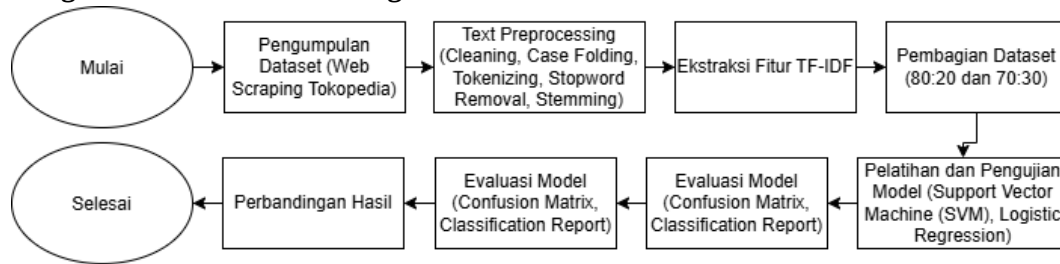
$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

III. HASIL DAN PEMBAHASAN

3.1 Perancangan Sistem

Proses implementasi pada algoritma SVM dan algoritma *Logistic Regression* menggunakan python untuk proses prediksi. Tahapan implementasi meliputi:

3.1.1 Rancang Model Machine Learning



Gambar 2. Diagram Alur Tahapan Penelitian

Pada gambar 2. menggambarkan alur penelitian yang dimulai dari pengumpulan dataset produk Tokopedia melalui web scraping, kemudian dilakukan text preprocessing. Data yang telah diproses selanjutnya diekstraksi menggunakan TF-IDF dan dibagi menjadi dua skenario, yaitu 80:20 dan 70:30, untuk proses pelatihan dan pengujian model SVM dan Logistic Regression. Kedua model dievaluasi menggunakan confusion matrix dan classification report, kemudian hasil evaluasi dibandingkan untuk mengetahui performa masing-masing algoritma hingga diperoleh kesimpulan penelitian.

3.1.2 Dataset

Tabel 3. Data Set

id	nama_produk	harga	kota	toko	kategori
1	AC PORTABLE 1 PK PROMO MIDEA AC PORTABEL 1PK MRF2-09CRN1 MRF09CRN1	Rp4.399.000	Kab. Bogor	Bogor Solusi Elektronik	Elektronik
2	?COD?Tixx 40L Pendingin Udara Air Portable Kipas Angin Dingin Super Dingin Hemat Energi 3 Kecepatan Mudah Dibawa Cocok untuk Rumah Kantor Kafe Usaha Kualitas Tinggi Tahan Lama Pendingin Ruang an Terbaik Hemat Daya Ringkas Praktis	Rp719.200	Tangerang	Tixx MegaElectron	Elektronik
3	TITON Freeze 1 Mini Kipas Refrigeration Fan Speed Digital Display Adjustable Turbo Kipas Tangan Handheld	Rp182.775	Jakarta Pusat	TITON Indonesia	Elektronik
...	...	...	...	...	...
2998	[EDISI TERBATAS x TAHILALATS] Crystalin Air Minum Mineral 600ml - [1 Botol]	Rp4.250	Jakarta Barat	OT-Store_NEW	Makanan
2999	VIT AIR MINERAL BTL 1500 GR	Rp4.910	Denpasar	PrimoSupermarket level 21	Makanan
3000	NESTLE PURE LIFE 600 ML - AIR MINERAL	Rp3.730	Bandung	Hyfresh Bandung Indah Plaza	Makanan

Tabel 3. menampilkan *dataset* terdiri atas 3.000 data produk dengan enam atribut, yaitu id, nama_produk, harga, kota, toko, dan kategori. Adapun penjelasan dari setiap atribut pada *dataset* adalah sebagai berikut: Id, digunakan untuk membedakan setiap data produk, nama_produk, judul produk. Harga, nilai harga dari masing-masing produk, Kota, menunjukkan lokasi asal penjual produk, Toko, berisi informasi mengenai nama toko atau penjual produk dan Kategori, label kelas yang digunakan sebagai target klasifikasi, yaitu Elektronik, Fashion, dan Makanan.

Tabel 4. Distribusi Dataset Berdasarkan Kategori Produk

No	Kategori	Jumlah Data
1	Elektronik	1000
2	Fashion	1000
3	Maknan	1000
Total	3 Kategori	3000 Data

Berikut tabel 4. adalah 3 kategori teridiri dari elektronik, fashion dan makanan masing-masing terdiri dari 1000 dataset.

Tabel 5. Dataset Hasil Scraping

id	nama_produk	harga	kota	toko	kategori
1	AC PORTABLE 1 PK PROMO MIDEA AC PORTABEL 1PK MRF2-09CRN1 MRF09CRN1	Rp4.399.000	Kab. Bogor	Bogor Solusi Elektronik	Elektronik
.....					
1003	Jas Hujan Setelan Mesi Jas Hujan Pria dan Wanita 66003 - Ralen Star	Rp49.900	Kab. Bandung Barat	Ralen Star Official Shop	Fashion
.....					
2001	Swallow Globe Agar Agar Merah 7 Gr	Rp5.980	Padang	Foodmart Basko Padang	Makanan

Berikut tabel 5. adalah hasil *scraping* memuat beberapa informasi mengenai produk, seperti nama produk, harga, lokasi penjual, nama toko, dan kategori.

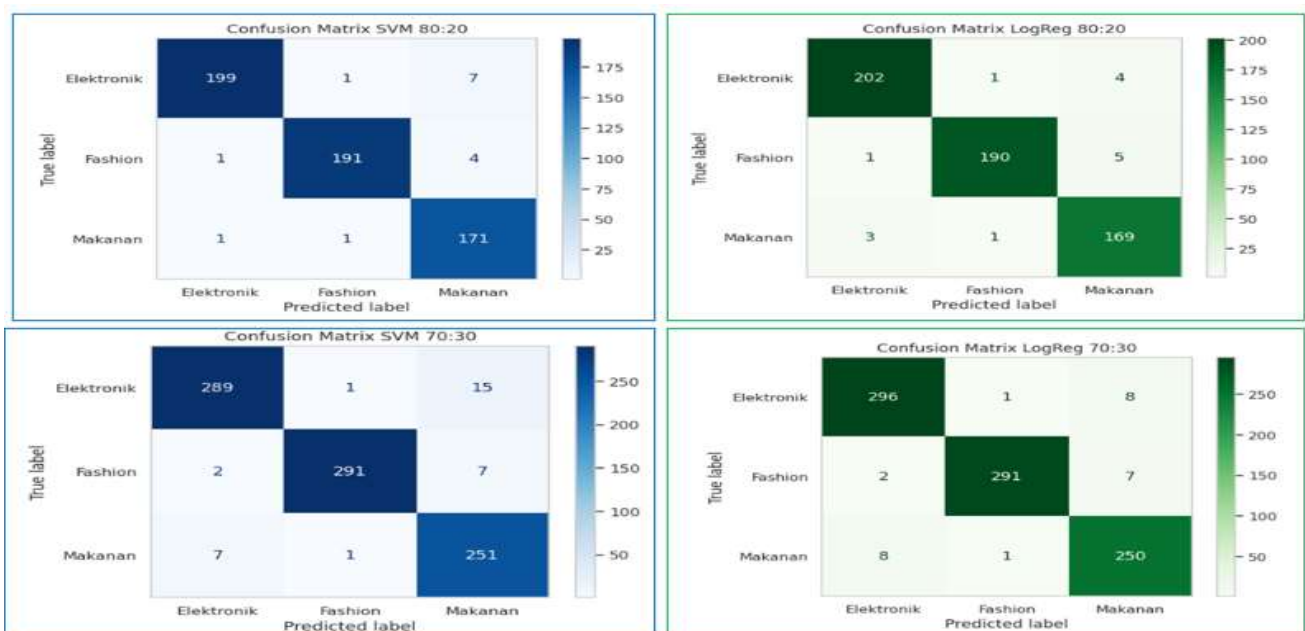
3.1.3 Hasil

PERBANDINGAN HASIL EVALUASI MODEL						
	Algoritma	Skenario	Accuracy	Precision	Recall	F1-Score
0	SVM	80:20	97.40%	97.31%	97.48%	97.36%
1	Logistic Regression	80:20	97.40%	97.32%	97.40%	97.35%
2	SVM	70:30	96.18%	96.08%	96.22%	96.12%
3	Logistic Regression	70:30	96.88%	96.80%	96.86%	96.82%

Gambar 3. Hasil Akurasi Algoritma SVM dan Logistic Regression

Pada Gambar 3. dapat dilihat bahwa hasil *classification report* menunjukkan bahwa model *Logistic Regression* memiliki performa yang tinggi dengan akurasi 97%. *Logistic Regression* memberikan hasil prediksi yang sangat akurat, sedangkan model SVM menunjukkan performa yang sangat baik dengan akurasi 96%.

3.2 Implementasi Algoritma



Gambar 4. Hasil Confusion matrix model Algoritma SVM dan Logistic Regression

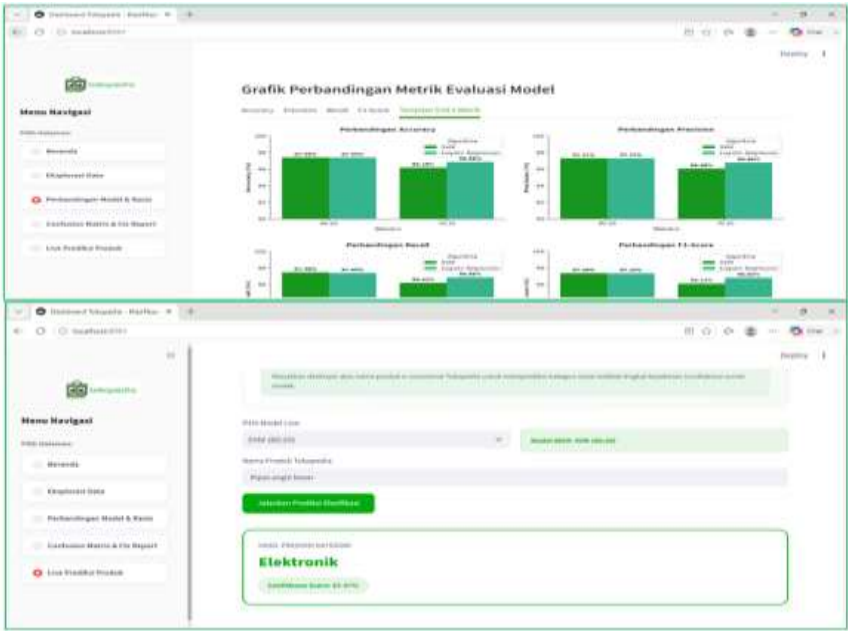
Gambar 4. Pada confusion matrix terlihat bahwa model SVM dan *Logistic Regression* mampu melakukan klasifikasi kategori produk Tokopedia dengan tingkat ketepatan yang tinggi pada tiga kategori, yaitu Elektronik, Fashion, dan Makanan. Pada skenario 80:20, kedua model menunjukkan jumlah prediksi benar yang relatif sama, yaitu SVM sebanyak 561 data dan *Logistic Regression* sebanyak 561 data, dengan kesalahan klasifikasi yang relatif kecil pada masing-masing kategori. Sementara itu, pada skenario 70:30, *Logistic Regression* menghasilkan jumlah prediksi benar sebanyak 837 data, lebih tinggi dibandingkan SVM sebanyak 831 data, yang menunjukkan bahwa *Logistic Regression* memiliki performa sedikit lebih baik pada skenario tersebut.

Tabel 6. Perhitungan

Algoritma SVM	Algoritma Logistic Regression
$Accuracy = \frac{TP_{Elektronik} + TP_{Fashion} + TP_{Makanan}}{Total\ Data\ Uji}$	$Accuracy = \frac{TP_{Elektronik} + TP_{Fashion} + TP_{Makanan}}{Total\ Data\ Uji}$
$Accuracy = \frac{289 + 291 + 251}{289 + 291 + 251 + 2 + 7 + 1 + 15 + 1 + 7} \times 100\%$ $= 96.18\%$	$Accuracy = \frac{296 + 291 + 250}{296 + 291 + 250 + 2 + 8 + 1 + 1 + 8 + 7} \times 100\%$ $= 96.88\%$
$Precision = 0.9608$	$Precision = 0.9680$
$Recall = 0.9622$	$Recall = 0.9682$
$F1\ Score = 0.9612$	$F1\ Score = 0.9682$
$Accuracy = \frac{\sum_{i=1}^n I(\hat{y}_i = y_i)}{n} = \frac{831}{863} = 0.96$	$Accuracy = \frac{\sum_{i=1}^n I(\hat{y}_i = y_i)}{n} = \frac{837}{863} = 0.97$

Berdasarkan hasil perhitungan table 6. akurasi model adalah 0,97 atau 97%. Nilai ini diperoleh dari 837 prediksi benar dari total 863 data uji. Dengan tingkat kesalahan yang sangat kecil (hanya 1 data), model dapat dikatakan memiliki performa yang sangat baik dalam melakukan klasifikasi.

3.3 Tampilan Aplikasi



Gambar 5. Tampilan Streamlit

Gambar 5. menyajikan perbandingan performa SVM dan *Logistic Regression* berdasarkan metrik pada skenario pembagian data 80:20 dan 70:30, serta menampilkan fitur *live prediction* yang memungkinkan pengguna memilih model, memasukkan nama atau deskripsi produk, dan memperoleh hasil klasifikasi secara langsung. Pada pengujian menggunakan produk “Kipas angin besar”, model SVM dengan skenario pembagian data 80:20 berhasil mengklasifikasikan produk ke dalam kategori Elektronik dengan *confidence score* sebesar 97,37%, sehingga menunjukkan bahwa model yang telah dibangun dapat diimplementasikan secara interaktif untuk mendukung proses klasifikasi kategori produk.

3.4 Hasil Perbandingan

Kedua algoritma memiliki kemampuan yang baik dalam melakukan klasifikasi kategori Elektronik, Fashion, dan Makanan. Pembagian data 80:20 menghasilkan kinerja yang lebih tinggi untuk kedua algoritma dibandingkan skenario 70:30. Pada skenario 70:30, Logistic Regression memberikan hasil accuracy yang sedikit lebih tinggi daripada SVM. Hasil pengujian, skenario 80:20 memberikan hasil klasifikasi terbaik dengan nilai accuracy 97,40%, sedangkan algoritma yang memberikan nilai accuracy lebih tinggi pada skenario 70:30 adalah *Logistic Regression* dengan nilai 96,88%.

IV. SIMPULAN

Penerapan *Text Preprocessing* yang meliputi *Cleaning*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming*, dilanjutkan dengan pembobotan TF-IDF, efektif digunakan sebagai tahapan dalam klasifikasi nama produk Tokopedia ke dalam kategori Elektronik, Fashion, dan Makanan menggunakan SVM dan *Logistic Regression*. Pada skenario pembagian data 80:20, kedua algoritma menghasilkan kinerja yang sama dengan *Accuracy* sebesar 97,40% serta *Precision*, *Recall*, dan *F1-Score* sebesar 97%, sedangkan pada skenario 70:30, *Logistic Regression* memperoleh *Accuracy* sebesar 96,88%, lebih tinggi 0,70 poin persentase dibandingkan SVM sebesar 96,18%. Dengan demikian, kedua metode menunjukkan kinerja klasifikasi yang sangat baik, namun tidak terdapat algoritma yang secara konsisten unggul pada seluruh skenario, dengan skenario pembagian data 80:20 memberikan hasil terbaik bagi kedua algoritma.

REFERENSI

- [1]. Aditama, F. S., Krismawati, D., & Pramana, S. (2024). Multiclass Classification of Marketplace Products With Machine Learning. *Media Statistika*, 17(1), 25–35. <https://doi.org/10.14710/medstat.17.1.25-35>
- [2]. Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54. <https://doi.org/10.57152/malcom.v2i1.195>
- [3]. Akbar, I., Marwondo, M., & Nugraha, N. (2023). Optimasi Algoritma Support Vector Machine untuk Analisis Klasifikasi Teks Pemintaan Informasi di Platform Online Shop. *Jurnal Accounting Information System (AIMS)*, 6(2), 119–126. <https://doi.org/10.32627/aims.v6i2.819>
- [4]. Alaiya, A., Nurdin, N., & Agusniar, C. (2025). Sentiment Analysis of E-Commerce Product Reviews on Tokopedia Using Support Vector Machine. *Journal of Applied Informatics and Computing*, 9(5), 2869–2878. <https://doi.org/10.30871/jaic.v9i5.10977>
- [5]. Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 6(2), 129. <https://doi.org/10.30998/string.v6i2.10133>
- [6]. Aulya, F. U., & Kusnawi, K. (2025). Evaluating Classification Models for Predicting Product Success in Indonesian E-Commerce. *Jurnal Teknik Informatika (Jutif)*, 6(4), 2723–2739. <https://doi.org/10.52436/1.jutif.2025.6.4.5071>
- [7]. Dharrao, D., Deokate, S., Bongale, A. M., & Urolagin, S. (2023). E-Commerce Product Review Classification based on Supervised Machine Learning Techniques. *2023 9th International Conference on Advanced Computing and Communication Systems, ICACCS 2023*, 1934–1939. <https://doi.org/10.1109/ICACCS57279.2023.10112717>
- [8]. Fiddin, F., Hidayat, T., & Djamaludin, D. (2025). Analisis Sentimen Ulasan Produk Sayur di Tokopedia Menggunakan Model Support Vector Machine Dengan Representasi TF-IDF. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 24(2), 162–172. <https://doi.org/10.53513/jis.v24i2.11479>
- [9]. Indra Dikusuma, & Yunus Widjaja. (2025). Perbandingan Naïve Bayes dan Logistic Regression

- untuk Analisis Sentimen Ulasan Produk Skincare di Tokopedia. *Jurnal Informatika Dan Teknologi Komputer (JITEK)*, 5(3), 242–251. <https://doi.org/10.55606/jitek.v5i3.8420>
- [10].Indriani, N., Rainarli, E., & Dewi, K. E. (2017). JURNAL INFOTEL Informatika-Telekomunikasi-Elektronika Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen. *Jurnal Infotel*, 9(4), 416–421.
- [11].Pakpahan, D., Siallagan, V., & Siregar, S. (2023). Classification of E-Commerce Product Descriptions with The Tf-Idf and Svm Methods. *Sinkron*, 8(4), 2130–2137. <https://doi.org/10.33395/sinkron.v8i4.12779>
- [12].Patra, A., Vivek, V., Shambhavi, B. R., Sindhu, K., & Balaji, S. (2021). Product Classification in E-Commerce Sites. *Advances in Intelligent Systems and Computing*, 1299 AISC, 485–495. https://doi.org/10.1007/978-981-33-4299-6_40
- [13].Raharko, N. I., & Yamasari, Y. (2024). Klasifikasi Emosi Ulasan Produk E- Commerce Menggunakan Support Vector Machine. *Journal of Informatics and Computer Science (JINACS)*, 6(03), 606–616. <https://doi.org/10.26740/jinacs.v6n03.p606-616>
- [14].Riki, Eliesse Dameria, S., Hermawan, A., Junaedi, & Kurnia, Y. (2025). Optimizing Sentiment Classification of E-Commerce Product Reviews: A Comparative Study of Naïve Bayes and SVM with SMO. *JUITA: Jurnal Informatika*, 13(3), 287–295. <https://doi.org/10.30595/juita.v13i3.26642>
- [15].Ryzal Nur Alvandy, & Arita Witianti. (2025). Analisis Sentimen Ulasan E-Commerce Menggunakan Metode SVM. *Elkom: Jurnal Elektronik Dan Komputer*, 18(2), 164–175. <https://doi.org/10.51903/elkom.v18i2.3253>
- [16].Sumual, I. M., Supriadi, J., Effendy, E., & Wijaya, A. (2024). KLASIFIKASI KATEGORI PRODUK TERLARIS PADA E-COMMERCE MENGGUNAKAN ALGORITMA NAIVE BAYES. *Jurnal Sistem Informasi (JUSIN)*, 5(2), 51–60. <https://doi.org/10.32546/jusin.v5i2.2893>
- [17].Zainottah, M. Z., Rengga, R. S., Yustian, Y. S., & Isa, I. R. (2025). Critical Sentiment Analysis of Tokopedia Electronic Products Using SVM-Logistic & TF-IDF Ensemble Methods. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(3), 2476–2482. <https://doi.org/10.59934/jaiea.v4i3.1194>