

**REKOMENDASI PRODUK *E-COMMERCE* BERBASIS
KLASIFIKASI MENGGUNAKAN ALGORITMA
*MACHINE LEARNING***

SKRIPSI



**Oleh:
Intensif Nehe
210210072**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2025**

**REKOMENDASI PRODUK *E-COMMERCE* BERBASIS
KLASIFIKASI MENGGUNAKAN ALGORITMA
*MACHINE LEARNING***

SKRIPSI

**Untuk memenuhi salah satu syarat
Memperoleh gelar sarjana**



**Oleh:
Intensif Nehe
210210072**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2025**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama : Intensif Nehe

NPM : 210210072

Fakultas : Teknik Dan Komputer

Menyatakan bahwa “Skripsi” yang saya buat dengan judul:

REKOMENDASI PRODUK *E-COMMERCE* BERBASIS KLASIFIKASI MENGGUNAKAN ALGORITMA *MACHINE LEARNING* adalah karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya di dalam naskah skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata didalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur **PLAGIASI**, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 24 Juli 2025



INTENSIF NEHE
210210072

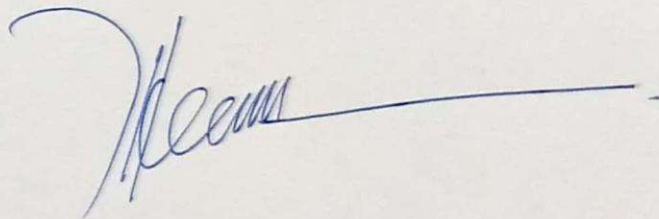
**REKOMENDASI PRODUK *E-COMMERCE* BERBASIS
KLASIFIKASI MENGGUNAKAN ALGORITMA
*MACHINE LEARNING***

SKRIPSI

**Untuk memenuhi salah satu syarat
Memperoleh gelar sarjana**

**Oleh:
Intensif Nehe
210210072**

**Telah disetujui oleh pembimbing pada tanggal
Seperti yang tertera dibawah ini
Batam, 24 Juli 2025**

A handwritten signature in blue ink, appearing to read 'Hotma', followed by a long horizontal line extending to the right.

**Hotma Pangaribuan, S.kom.,M.SI.
Pembimbing**

ABSTRAK

Perkembangan e-commerce yang sangat pesat telah mendorong meningkatnya jumlah ulasan produk dari konsumen. Namun, tidak semua ulasan yang tersedia bersifat informatif dan relevan untuk dijadikan dasar pengambilan keputusan. Kondisi ini menimbulkan kebutuhan akan sistem yang mampu menyaring serta mengklasifikasikan ulasan sehingga hanya informasi relevan yang dapat digunakan dalam mendukung rekomendasi produk. Penelitian ini bertujuan membangun sistem klasifikasi ulasan produk pada platform Tokopedia menggunakan algoritma Naive Bayes, khususnya varian Multinomial, untuk memprediksi relevansi ulasan. Dataset yang digunakan berjumlah 40.607 baris ulasan yang diperoleh dari Kaggle, dengan atribut utama berupa teks ulasan dan rating. Tahapan penelitian meliputi preprocessing data (cleaning, tokenisasi, stopwords removal, dan labeling), pembentukan fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dengan 5.000 kata fitur, pembagian data menjadi 80% untuk pelatihan dan 20% untuk pengujian, serta evaluasi model melalui confusion matrix, akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model Multinomial Naive Bayes mampu mencapai akurasi sebesar 93,71%. Untuk kelas relevan, diperoleh precision 0,94, recall 0,99, dan F1-score 0,96. Meskipun performa klasifikasi pada kelas tidak relevan relatif rendah akibat distribusi data yang tidak seimbang, model tetap efektif dalam mengidentifikasi ulasan relevan yang mendominasi dataset. Temuan ini membuktikan bahwa Naive Bayes dapat menjadi metode sederhana namun kuat dalam mendukung sistem rekomendasi produk berbasis ulasan konsumen. Implementasi model ini diharapkan mampu meningkatkan kepercayaan pengguna terhadap rekomendasi e-commerce, mempermudah konsumen menemukan produk sesuai preferensi, serta memberikan manfaat praktis bagi pengembang platform dalam mengoptimalkan pengalaman belanja online.

Kata Kunci : *Naive Bayes*, Ulasan Produk, Tokopedia, Klasifikasi

ABSTRACT

The rapid development of e-commerce has driven an increase in the number of product reviews from consumers. However, not all available reviews are informative and relevant to use as a basis for decision-making. This creates a need for a system that can filter and classify reviews so that only relevant information can be used to support product recommendations. This study aims to build a product review classification system on the Tokopedia platform using the Naive Bayes algorithm, especially the Multinomial variant, to predict the relevance of reviews. The dataset used amounted to 40,607 review lines obtained from Kaggle, with the main attributes being review text and ratings. The research stages include data preprocessing (cleaning, tokenization, stopword removal, and labeling), feature formation using the Term Frequency-Inverse Document Frequency (TF-IDF) method with 5,000 feature words, data division into 80% for training and 20% for testing, and model evaluation through confusion matrix, accuracy, precision, recall, and F1-score. The results showed that the Multinomial Naive Bayes model was able to achieve an accuracy of 93.71%. For the relevant class, a precision of 0.94, a recall of 0.99, and an F1-score of 0.96 were obtained. Although the classification performance of irrelevant classes is relatively low due to unbalanced data distribution, the model remains effective in identifying relevant reviews that dominate the dataset. These findings prove that Naive Bayes can be a simple yet powerful method in supporting a consumer review-based product recommendation system. The implementation of this model is expected to increase user trust in e-commerce recommendations, make it easier for consumers to find products according to their preferences, and provide practical benefits for platform developers in optimizing the online shopping experience.

Keywords : Naive Bayes, Product Reviews, Tokopedia, Classification

KATA PENGANTAR

Segala puji dan Syukur kita panjatkan kepada Tuhan Yang Maha Esa yang telah melimpah segala Rahmat dan karuniaNya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi Strata Satu (S1) pada program studi Teknik Informatika Universitas Putera Batam.

Penulis menyadari bahwa Skripsi ini masih jauh dari kata sempurna. Karena itu kritik dan saran akan senantiasa penulis dengan senang hati. Dengan segala keterbatasan, penulis juga menyadari bahwa skripsi ini tak akan terwujud tanpa bantuan, bimbingan dan dorongan dari berbagai pihak. Untuk itu dengan segala kerendahan hati, penulis menyampaikan ucapan terimakasih kepada

1. Rektor Universitas Putera Batam Ibu Dr. Nur Elfi Husda,.S.kom.,M.SI.
2. Dekan Fakultas Teknik dan Komputer Universitas Putera Batam Bapak
3. Ketua Program Studi Teknik Informatika Universitas Putera Batam Bapak Andi Maslan,.S.T.,M.SI.,P.HD.
4. Bapak Hotma Pangaribuan, S.kom.,M.SI. selaku Pembimbing Skripsi pada Program Studi Teknik Informatika Universitas Putera Batam.
5. Dosen dan Staff Universitas Putera Batam yang telah memberikan ilmu pengetahuan, bimbingan serta pelayanan yang baik kepada penulis
6. Orang Tua yang saya sayangi A/I Hemi Nehe, yang selalu memberikan dukungan doa yang terbaik buat penulis selama ini.
7. Teman-teman seperjuangan angkatan tahun 2021 Universitas Putera Batam yang selama ini membantu penulis untuk mengerjakan tugas akhir ini.

Semoga Tuhan Yang Maha Esa membalas kebaikan dan mencurahkan berkat kepada semua orang yang telah membantu penulis dalam menyelesaikan skripsi ini.

DAFTAR ISI

SKRIPSI.....	i
SKRIPSI.....	ii
SURAT PERNYATAAN ORISINALITAS	iii
SKRIPSI.....	Error! Bookmark not defined.
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR RUMUS.....	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi masalah	4
1.3 Batasan Penelitian	5
1.4 Rumusan Masalah	5
1.5 Tujuan Penelitian.....	6
1.6 Manfaat Penelitian	6
1.6.1 Manfaat Teoritis	6
1.6.2 Manfaat Praktis	6
BAB II TINJAUAN PUSTAKA	8
2.1 Teori Dasar	8
2.2 <i>Machine learning</i>	11
2.2.1 <i>Support Vector Machine (SVM)</i>	13
2.2.2 <i>Logistic Regression</i>	14
2.2.3 <i>Decesion Tree</i>	14
2.2.4 <i>Random Forest</i>	14
2.2.5 <i>Naive Bayes</i>	15
2.2.6 Klasifikasi	16
2.3 <i>Software pendukung</i>	19

2.4	Penelitian terdahulu.....	21
2.5	Kerangka pikir.....	23
BAB III METODE PENELITIAN.....		23
3.1	Desain Penelitian.....	23
3.2	Metode Pengumpulan Data.....	24
3.3	Metode Perancangan	25
3.3.1	Pengumpulan Data	25
3.3.3	Ekstraksi Fitur	26
3.3.4	Treaning dan Testing.....	27
3.3.5	Klasifikasi <i>Naive Bayes</i>	28
3.3.6	Evaluasi Model.....	28
3.4	Jadwal Peneltian.....	29
BAB IV HASIL DAN PEMBAHASAN.....		30
4.1	Hasil Penelitian	30
4.1.1	Deskripsi Dataset	30
4.1.2	Hasil Preprocessing Data	32
4.1.3	Pembentukan Fitur (TF-IDF).....	37
4.1.4	Pelatihan dan Pengujian Model <i>Naive Bayes</i>	39
4.1.5	Evaluasi Kinerja Model.....	43
4.2	Pembahasan.....	48
4.2.1.	Mengatasi Kurangnya Data historis Produk	48
4.2.2.	Kepercayaan Berbelanja Melalui E-Commerce.....	49
4.2.3.	Rekomendasi Produk Berdasarkan Preferensi Pengguna	50
4.2.4.	<i>Model Naive Bayes</i> Dalam Klasifikasi Ulasan Produk.....	51
BAB V PENUTUP.....		53
5.1	KESIMPULAN	53
5.2	SARAN	54
DAFTAR PUSTAKA.....		55
LAMPIRAN I		58
LAMPIRAN II		68
LAMPIRAN III.....		69

DAFTAR GAMBAR

Gambar 1. 1	Jumlah Pengguna E-Commerce di Indonesia Periode 2020-2029 ...	2
Gambar 2. 1	Jupyter Notebook.....	20
Gambar 2. 2	Microsoft Excel	21
Gambar 2. 3	Kerangka pikir	24
Gambar 3. 1	Desain Penelitian	23
Gambar 3. 2	Perencanaan Model.....	25
Gambar 4. 1	Dataset	31
Gambar 4. 2	Kolom Text Dan Rating.....	32
Gambar 4. 3	Hasil Cleaning Data	33
Gambar 4. 4	Hasil Tokenization	34
Gambar 4. 5	Hasil Stopword Removal.....	35
Gambar 4. 6	Label Relevansi Ulasan Produk.....	36
Gambar 4. 7	Pembentukan Fitur TF-IDF	38
Gambar 4. 8	Pembagian Data Training Dan Testing	40
Gambar 4. 9	Naïve Bayes	41
Gambar 4. 10	prediksi naive bayes.....	42
Gambar 4. 11	Evaluasi Model	43
Gambar 4. 12	Hasil Akurasi	43
Gambar 4. 13	Confusion Matrix Model Klasifikasi Naive Bayes.....	44
Gambar 4. 14	Hasil Classification Report	46

DAFTAR RUMUS

Rumus 1 Naive Bayes	16
Rumus 2 Akurasi	18
Rumus 3 presisi	19
Rumus 4 Recall	19
Rumus 5 F1_Score	19
Rumus 6 TF (Term Frequency)	27
Rumus 7 IDF (Inverse Docement Frequency)	27
Rumus 8 TF-IDF Score	27