

ANALISIS SENTIMEN MENGGUNAKAN METODE *NAÏVE BAYES* TERHADAP PRODUK PT.IMIN TECHNOLOGY BERDASARKAN ULASAN DARI TOKOPEDIA

Syahjuddin Azra¹, Rizky Tahara Shita^{2*}

^{1,2*}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹syahjudinazra@gmail.com, ²rizky.taharashita@budiluhur.ac.id
(* : corresponding author)

Abstrak-Dalam melakukan perdagangan online seringkali dijumpai permasalahan pada kualitas barang yang dijual, Sehingga diperlukan solusi untuk mengatasi permasalahan pada kualitas barang tersebut. Ulasan pelanggan menjadi sumber informasi yang berharga bagi penjual, sehingga penting untuk memahami sentimen pelanggan terkait produk tertentu. Data penelitian ini diambil dari ulasan pelanggan yang berhubungan dengan produk yang dijual di platform Tokopedia dengan nama toko Imin Official Store. Data ulasan tersebut kemudian diproses dan dianalisis menggunakan metode *Naive Bayes*. Adapun tahapan yang dilakukan dalam penelitian analisis sentimen ini yaitu *crawling*, *pre-processing*, pembuatan model dan mengelompokkan data menjadi sentimen negatif dan positif. Selanjutnya, akan diuji dengan pengujian matrix confusion dengan parameter akurasi, recall, dan ketepatan. Pengujian ini dilakukan dengan 227 data ulasan. Akurasi data uji dan latih masing-masing 20% dan 80%. Dari pengujian yang dilakukan memperoleh hasil akurasi sebesar 86,957%, presisi 90,476% dan *recall* 95%. Penelitian ini dapat membantu penjual perangkat Android POS dalam meningkatkan kualitas produk mereka dan merespons umpan balik pelanggan secara efektif.

Kata Kunci: Android POS, Analisis Sentimen, Metode *Naive Bayes*.

SENTIMENT ANALYSIS USING THE NAÏVE BAYES METHOD ON PT.IMIN TECHNOLOGY PRODUCTS BASED ON REVIEWS FROM TOKOPEDIA

Abstract- *In conducting online trading, problems are often encountered with the quality of goods sold, so a solution is needed to overcome problems with the quality of these goods. Customer reviews are a valuable source of information for sellers, so it's important to understand customer sentiment regarding a particular product. The research data is taken from customer reviews related to products sold on the Tokopedia platform under the store name Imin Official Store. The review data was then processed and analyzed using the Naive Bayes method. The stages carried out in this sentiment analysis research are crawling, pre-processing, making models and classifying data into negative and positive sentiments. Next, it will be tested by testing matrix confusion with parameters of accuracy, recall, and accuracy. This test was conducted with 227 review data. The accuracy of the test and train data is 20% and 80%, respectively. From the tests carried out obtained results of accuracy of 86.957%, precision of 90.476% and recall of 95%. This research can help Android POS device sellers improve their product quality and respond effectively to customer feedback.*

Keywords: *Android POS, Sentiment Analysis, Naive Bayes Method.*

1. PENDAHULUAN

Dalam era digital saat ini, banyak konsumen mengandalkan platform *online* seperti Tokopedia untuk mencari dan membeli berbagai produk. Di antara perusahaan teknologi terbesar di Indonesia, PT. Imin Technology memiliki beragam produk yang ditawarkan kepada konsumen melalui platform ini. Dalam konteks persaingan yang semakin ketat di industri teknologi, penting bagi PT. Imin Technology untuk memahami pandangan dan sentimen pelanggan terhadap produk-produk yang ditawarkan. Karena memiliki peran yang krusial dalam efisiensi dan kualitas layanan yang ditawarkan oleh perusahaan ritel, restoran, atau bisnis kecil lainnya.

Dalam melakukan kegiatan perdagangan secara *online*, Seringkali dijumpai permasalahan kualitas produk yang kurang memenuhi ekspektasi dari pelanggan sehingga berdampak menurunnya intensitas penjualan produk. Dalam hal ini platform *e-commerce* seperti Tokopedia menjadi tempat utama bagi pelanggan untuk berbagi pengalaman mereka melalui ulasan produk Imin *Official Store*. Oleh karena itu, menganalisis sentimen pelanggan berdasarkan ulasan di Tokopedia dapat memberikan wawasan yang berharga tentang persepsi pelanggan terhadap kualitas produk yang ditawarkan di pasar.

Dalam konteks ini, metode klasifikasi sentimen seperti metode *Naive Bayes* dapat digunakan untuk menganalisis dan mengkategorikan ulasan pelanggan sebagai sentimen positif atau negatif. Metode *Naive Bayes* telah terbukti efektif dalam berbagai aplikasi analisis sentimen, termasuk dalam analisis ulasan produk di platform *e-commerce*. Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen pelanggan terkait kualitas produk berdasarkan ulasan yang ditemukan di Tokopedia.

Melalui penelitian ini, diharapkan penjual dapat memperoleh wawasan yang lebih mendalam tentang kekuatan dan kelemahan produk mereka. Dengan demikian, mereka dapat meningkatkan kualitas perangkat mereka dan merespon masukan pelanggan secara lebih efektif. Penelitian ini juga dapat membantu pembeli membuat keputusan yang lebih akurat dan informasi saat memilih produk yang sesuai dengan kebutuhan mereka.

[1] Analisis Sentimen, juga dikenal sebagai Mining Opini, adalah deteksi sikap terhadap sesuatu atau orang. Ini dapat digunakan untuk mengetahui persentase sentimen positif dan negatif terhadap seseorang, industri, lembaga, produk, atau situasi tertentu. [2] Text mining adalah proses penggalian informasi yang digunakan untuk mendapatkan informasi bermanfaat dari sekumpulan dokumen. Tujuan utama proses text mining adalah untuk membantu proses penemuan pengetahuan tentang koleksi dokumen yang terdiri dari kategorisasi.

Berdasarkan penelitian sebelumnya [3] Penelitian ini menggunakan data dari Tripadvisor yaitu situs web yang menawarkan berbagai informasi tentang penginapan, restoran, objek wisata, dan hal lainnya. Hasil menunjukkan opini. Dataset telah dikategorikan ke dalam kategori positif 19, dan kategori negatif. Secara keseluruhan, seratus ulasan terdiri dari lima puluh kategori negatif dan lima puluh kategori positif. [4] Selanjutnya, hasil pengujian yang dilakukan dengan menggunakan metode support vector machine menunjukkan bahwa nilai akurasi kernel linear adalah 98,67%, akurasi 98%, recall 99%, dan F1-Score adalah 98%; nilai akurasi kernel RBF adalah 98,34%, akurasi 97%, recall 98%, dan F1-Score adalah 98%. Berdasarkan hasil e-test, dapat disimpulkan bahwa sentimen masyarakat pengguna Twitter terhadap program kartu pra kerja di masa pandemi lebih condong ke netral sebesar 98,34%. [5] Hasil perbandingan metode naive bayes dan support vector machine menunjukkan bahwa SVM memiliki akurasi sebesar 73,65%. [6] Selanjutnya, pengujian yang menggunakan algoritma kualifikasi dilakukan untuk mengevaluasi perasaan aplikasi ruang guru. Hasil pengujian menunjukkan bahwa algoritma PSO berbasis SVM, dengan nilai akurasi 78,55% dan AUC 0,853, adalah algoritma yang paling efektif dan terbaik untuk mengklasifikasikan komentar positif dan negatif yang terkait dengan aplikasi ruang guru. [7] Selain itu, algoritma naive bayes digunakan untuk menganalisis data Twitter BMKG nasional. Ini menghasilkan uji akurasi metode naive bayes untuk klasifikasi sebesar 69,97%. [8] Menurut penelitian yang melakukan analisis review restoran Singapura menggunakan metode Naive Bayes, hasilnya menunjukkan bahwa kepuasan pengunjung restoran dengan algoritma Naive Bayes memberikan nilai akurasi sebesar 73%. [9] Metode K-Nearest Neighbor juga digunakan untuk mengevaluasi pembatasan masyarakat. Metode ini menghasilkan nilai ketepatan 83,74%, ketepatan 77,78%, dan nilai recall 84,00% dari 244 data dengan sentimen negatif dan 175 data dengan sentimen positif. [10] Selain itu, metode *naive bayes* digunakan untuk mengukur pendapat masyarakat tentang vaksin COVID-19. Sebagian besar tweet menunjukkan sikap positif (60,3 %), sedangkan sikap netral (34,4 %) lebih banyak daripada sikap menentang (5,4 %). Hasilnya menunjukkan akurasi 0,93%.

2. METODE PENELITIAN

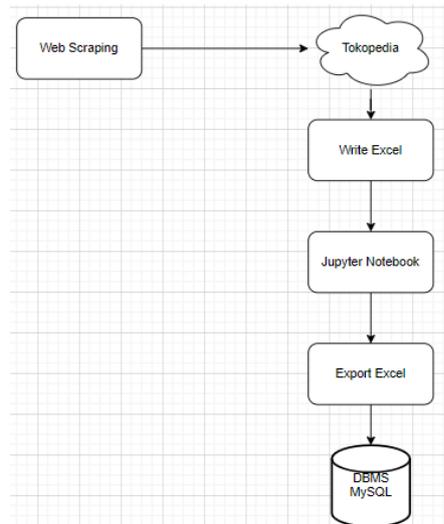
2.1 Data Penelitian

Data persiapan adalah langkah pertama dalam penelitian ini. Data ini dikumpulkan dari ulasan toko resmi Imin di Tokopedia, yang *disrape* melalui web dari 1 Juli 2023 hingga 3 Juli 2023, menghasilkan 227 set ulasan.

2.2 Penerapan Metode

Dengan menggunakan metode *Naive Bayes* untuk membuat sistem analisis sentimen pelanggan yang terkait dengan produk PT Imin Technology, ada beberapa fase yang menjadi perhatian khusus selama proses perancangan sistem dari awal hingga akhir. Fase pengumpulan data merupakan fase terpenting dalam penelitian ini, Yang dilakukan dengan menggunakan teknik *crawling* dan *labelling*.

2.2.1 Metode Pengumpulan Data

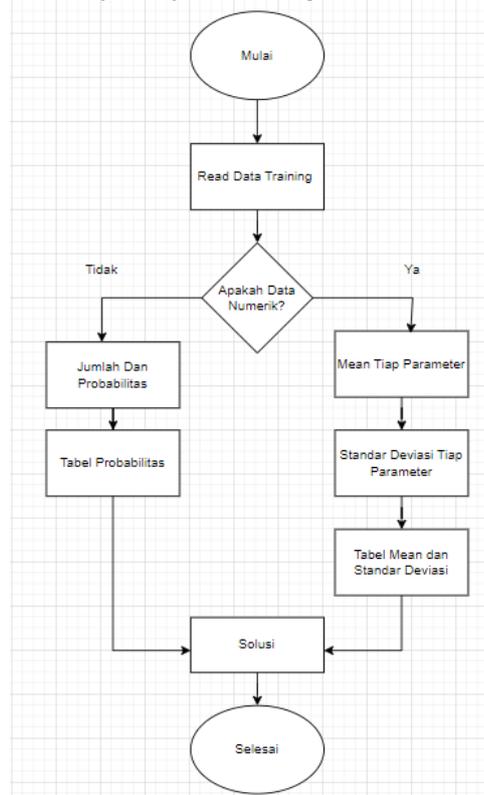


Gambar 1. Metode Pengumpulan Data

Penelitian ini meliputi pengumpulan data dari ulasan toko Imin Official Store di Tokopedia. Pengumpulan data tersebut menggunakan Web Scraping, Data yang dihasilkan akan langsung dimasukkan kedalam file excel dengan data yang terkumpul sebanyak 227 data, Selanjutnya akan dilakukan uji data training dan data uji.

2.2.2 Metode Algoritma Naïve Bayes

Adapun tahapan yang dilakukan dalam penerapan naïve bayes untuk analisis sentimen, yaitu:

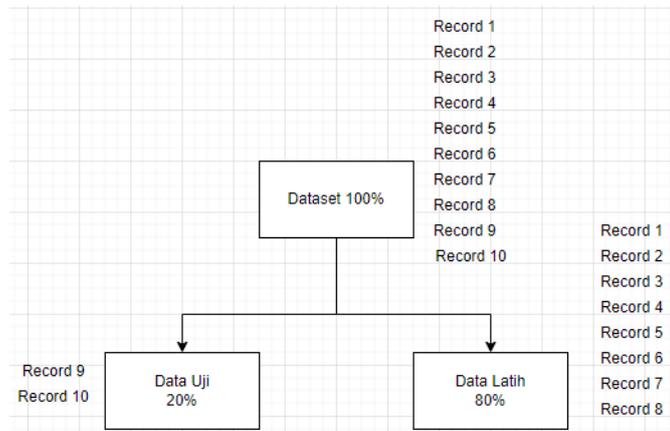


Gambar 2. Algoritma Naïve Bayes

Tahapan dilakukan dengan membaca data latih kemudian menghitung jumlah data numerik sebagai nilai probabilitas dan terakhir mendapatkan hasil dari probabilitas tersebut.

2.2.3 Pembagian Data

Dataset akan dibagi menjadi dua bagian selama pengujian untuk mengetahui hasil akurasi nantinya dalam penelitian ini, dibagi menjadi 80:20, yaitu 20% data uji dan 80% data latih. Tahap pembagian data dapat dilihat pada gambar 3, Sebagai berikut:



Gambar 3. Pembagian Data

1. Data Uji

Data untuk menguji tingkat keakuratan algoritma klasifikasi yang didasarkan pada model latih.

2. Data Latih

Kumpulan data yang berasal dari data yang telah melalui tahap *preprocessing*.

2.3 Rancangan Pengujian

2.3.1 Metode Pengujian

Metode simulasi yang dimaksud kemungkinan lebih mengarah pada pendekatan berbasis probabilitas dalam klasifikasi. Dalam metode ini, data pelatihan berisi atribut dan label kelas yang diketahui. Data tersebut digunakan untuk melatih model yang mampu mengenali pola dalam data dan memahami hubungan antara atribut dan kelas.

2.3.2 Preprocessing

Dalam bahasa tertentu tidak teratur karena banyaknya bahasa yang digunakan oleh orang di seluruh dunia berbeda-beda. Data teks sering ditarik di jaringan media sosial saat melakukan sentimen analisis. Menulis teks seperti ini biasanya dilakukan dalam bahasa informal dan tidak teratur. Sebelum data diproses untuk analisis sentimen, teks sebelumnya harus diproses agar lebih gramatikal dan teratur [11]. Tahapan preprocessing adalah upaya untuk mencapai tujuan mengubah data teks yang tidak terstruktur menjadi bentuk terstruktur.

1. Case Folding

Mengubah semua huruf teks menjadi huruf kecil yang sama. Karakter dari huruf a hingga z dihilangkan dan dianggap sebagai pembatas.

2. Normalisasi/Cleaning

Normalisasi adalah mengubah kata-kata yang disingkat ke dalam kamus bahasa. Pembersihan adalah proses menghilangkan angka, URL, penyebutan, tagar, tanda baca, dan tanda baca.

3. Tokenizing

Tokenizing adalah cara untuk membagi teks menjadi kata-kata.

4. Filtering

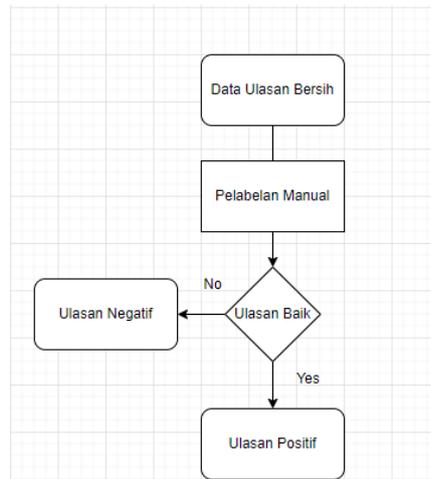
Filtering adalah proses memilih kata yang penting pada teks yang dihasilkan dengan menggunakan algoritma daftar berhenti, yang menghilangkan istilah yang dianggap tidak penting, atau algoritma daftar kata, yang menyimpan istilah dasar.

5. Stemming

Berasal dari kata yang dibubuhkan dan diubah menjadi ekspresi esensial.

2.3.3 Tahap Pelabelan

Dilakukan selama sesi pelabelan atau kelas sesuai dengan karakteristik yang ada dalam kalimat. Deklarasi kelas positif memiliki informasi tentang fakta, sedangkan deklarasi kelas negatif adalah yang bertentangan dengan deklarasi kelas positif. Tahap Pelabelan dapat dilihat pada gambar 4 sebagai berikut:



Gambar 4. Pelabelan

2.4 Algoritma Naïve Bayes

Pada abad ke-18, Reverend Thomas Bayes mengembangkan metode Naive Bayes, yang umumnya digunakan untuk klasifikasi dengan menggunakan pendekatan peluang atau probabilitas. Metode ini menggunakan pengalaman masa lalu untuk memprediksi kemungkinan di masa depan. Naïve Bayes Classifier (NBC) mengaitkan nilai target ke data baru dengan mencari kemungkinan nilai terbesar dari semua anggota kumpulan domain V . V merupakan kumpulan emosi. NBC menghitung probabilitas V_{map} (kelas emosi dengan probabilitas tertinggi) dengan memanfaatkan Teorema Bayes, dengan asumsi bahwa setiap atribut x_i (seperti kata-kata dalam permintaan data) adalah independen satu sama lain ketika diberikan nilai target V . Persamaan pada V dapat dilihat pada persamaan 1, Sebagai berikut:

$$VMAP = \frac{\text{armax} P(X_1, X_2, X_3, \dots, X_n | V_j) p(V_j)}{P(X_1, X_2, X_3, \dots, X_n)} \quad (1)$$

Persamaan dapat ditulis untuk $P(x_1, x_2, x_3 \dots x_n)$ nilainya konstan untuk semua kategori (V_j).

$$VMAP = \text{armax} P(x_1, x_2, x_3, \dots, x_n | V_j) p(V_j) \quad (2)$$

Persamaan diatas dapat disederhanakan menjadi berikut:

$$VMAP = \text{armax}_{ni=1} P(x_i | V_j) P(V_j) \quad (3)$$

Keterangan:

V_j = 1, 2, 3, ... n, kategori komentar menunjukkan sentimen positif, negatif, dan netral.

$P(x_i | V_j)$ = Kemungkinan nilai x_i pada V_j

$P(V_j)$ = Kemungkinan nilai pada (V_j)

Nilai $P(V_j)$ dan $P(x_i | V_j)$ dihitung pada saat pelatihan dimana persamaanya adalah sebagai berikut:

$$P(V_j) = \frac{|docs_j|}{|contoh|} \quad (4)$$

$$P(x_i|V_j) = \frac{nk + 1}{n + |kosakata|} \quad (5)$$

Keterangan:

- |docs_j| = Hasil berkas pada setiap kategori j
 |contoh| = Hasil berkas dari semua kategori yang ada
 nk = Hasil frekuensi untuk kemunculan kata
 n = Hasil frekuensi untuk kemunculan setiap kata dari kategori
 |kosakata| = Hasil semua kategori dari semua kata

2.5 Confusion Matrix

Perhitungan Confusion Matrix menggunakan pengujian data untuk menemukan nilai akurasi, Recall, dan Precision. Nilai-nilai ini dihitung dengan membandingkan data sebelum dan sesudah masuk sistem. Jadi, hasilnya adalah sebagai berikut:

Tabel 1. Confusion Matrix

	Aktual Positif	Aktual Negatif
Prediksi Positif	True Positive (TP)	False Positive (FP)
Prediksi Negatif	False Negative (FN)	True Negative (TN)

Berdasarkan confusion matrix di atas, berikut adalah pengukuran performa yang umum digunakan dalam evaluasi model klasifikasi:

1. Accuracy: Seberapa sering model memprediksi dengan benar (TP dan TN) dari total jumlah prediksi (TP, TN, FP, FN).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \times 100\% \quad (6)$$

2. Precision: Seberapa sering model memprediksi positif ketika benar-benar positif (TP) dari total jumlah prediksi positif (TP dan FP).

$$Precision = TP / (TP + FP) \times 100\% \quad (7)$$

3. Recall (Sensitivity atau True Positive Rate): Seberapa sering model memprediksi positif ketika seharusnya positif (TP) dari total jumlah aktual positif (TP dan FN).

$$Recall = TP / (TP + FN) \times 100\% \quad (8)$$

3. HASIL DAN PEMBAHASAN

Proses Labeling dan preprocessing dilakukan pada data yang dikumpulkan melalui teknik web scraping, yang kemudian dikelompokkan menjadi dua kategori: positif dan negatif. Dari 227 data yang dikumpulkan, 200 dikategorikan sebagai data positif dan 27 dikategorikan sebagai data negatif. Selanjutnya, perhitungan probabilitas dilakukan menggunakan persamaan 1.

3.1 Jumlah Dataset

Setelah melakukan proses Crawling dan Labelling, Didapati dataset yang digunakan berjumlah 227 dataset, Diantaranya 200 positif dan 27 Negatif.

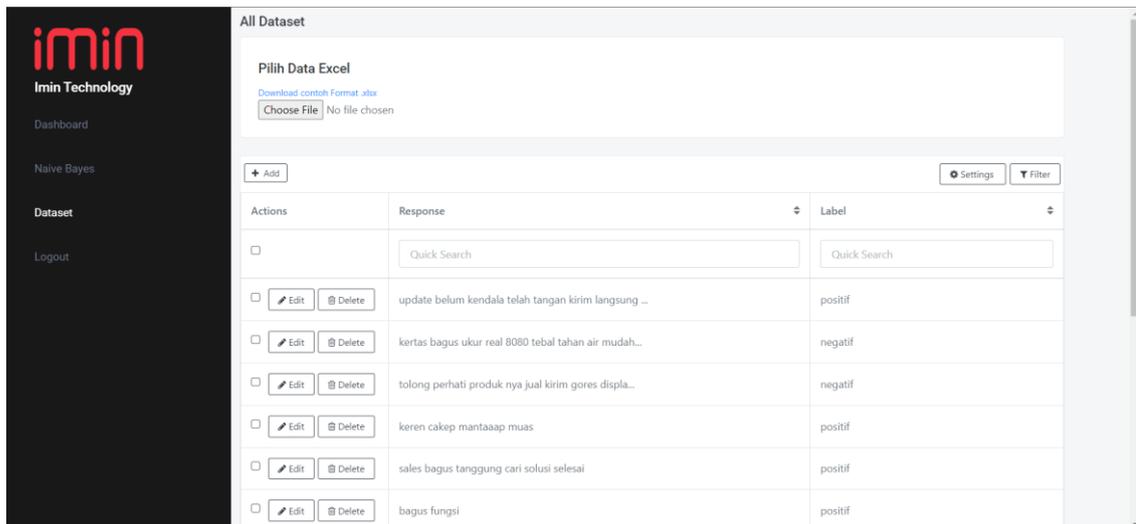
Tabel 2. Jumlah Dataset

Keterangan	Jumlah
Positif	200
Negatif	27

Total Dataset 227

3.2 Dataset

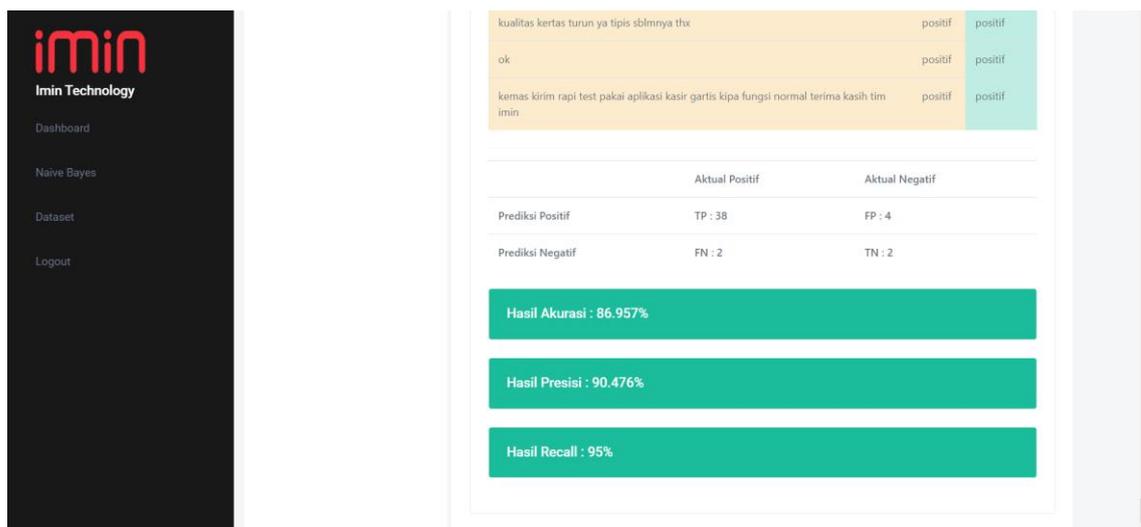
Halaman berikut merupakan halaman untuk menampilkan seluruh dataset yang telah dimasukkan kedalam sistem analisis sentiment juga dapat melakukan penambahan data secara manual. Halaman dataset dapat dilihat pada gambar 5, Sebagai berikut:



Gambar 5. Dataset

3.3 Uji Akurasi

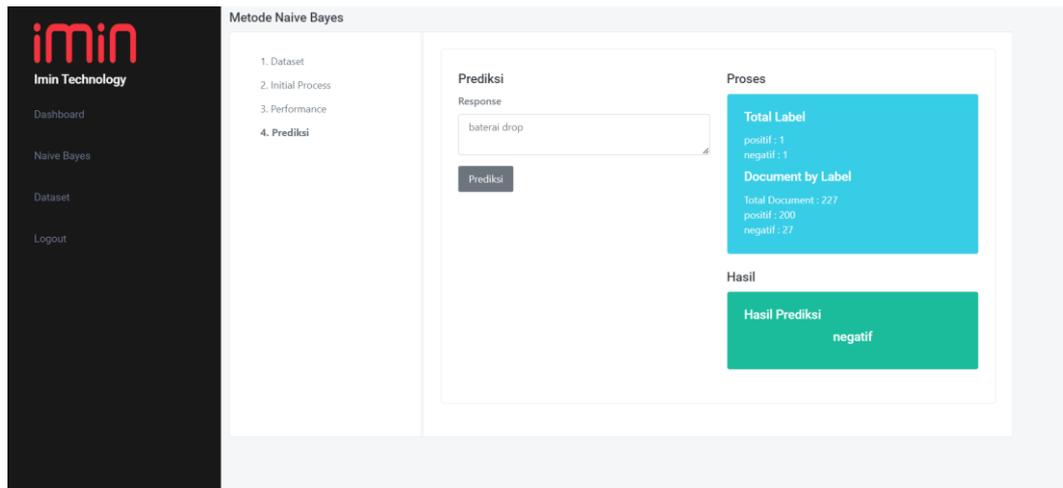
Proses evaluasi dan penganalisaan digunakan untuk mencapai tujuan ini. Proses ini juga mencakup menentukan tingkat akurasi sistem yang dikembangkan. Untuk melakukan prediksi pada data uji, algoritma Naive Bayes Classifier (NBC) digunakan. Selama pengujian, data uji dan data latih dibagi menjadi 20% dan 80%, masing-masing, untuk menentukan nilai akurasi. Halaman uji akurasi dapat dilihat pada gambar 6, Sebagai berikut:



Gambar 6. Uji Akurasi

3.4 Pengujian Prediksi

Pengujian prediksi dilakukan dengan memasukkan kata yang akan diuji ke dalam area teks yang tersedia. Hasilnya dapat dilihat pada gambar 7 berikut:



Gambar 7. Pengujian Prediksi

4. KESIMPULAN

Pada hasil penelitian yang dilakukan mengenai topik Analisis Sentimen Pelanggan Terhadap Produk PT. Imin Technology Berdasarkan Ulasan Pada Tokopedia Menggunakan Metode Naive Bayes, Menghasilkan kesimpulan sebagai berikut, Dari hasil *web scraping* yang dilakukan didapati 227 dataset. Uji Akurasi sentimen pelanggan terhadap produk PT. Imin Technology berada di angka 86.987%, *Recall* 95% dan Presisi 90.476%. Tahap utama yang dilakukan pada penelitian ini adalah fase *preprocessing* yang meliputi *Crawling*, *Labelling*, *Modelling* dan Klasifikasi. Penelitian ini juga memberikan hasil akurasi dari penerapan aplikasi metode *naive bayes* menggunakan ulasan pelanggan terkait produk yang dijual, Sehingga penjual memiliki gambaran untuk meningkatkan kualitas produk untuk kedepanya.

DAFTAR PUSTAKA

- [1] Mulyani, E. D. S., Hidayatuloh, A. T., & Agustawan, T. (2022). "Analisis Sentimen Ulasan Produk Pada Top Brand Produk Masker Di Tokopedia Menggunakan Naive Bayes." In Seminar Nasional CORIS 2022..
- [2] Wardhani, L. K. (2020). "Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naive Bayes Classifier (NBC) Pada Tweet Hashtag# 2019gantipresiden." KILAT, 9(1), pp. 103-114.
- [3] Sari, R., & Hayuningtyas, R. Y. (2019). "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website." Indonesian Journal on Software Engineering (IJSE), 5(2), pp. 51-60.
- [4] Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine." Jurnal Informatika: Jurnal Pengembangan IT, 6(3), pp. 150-155.
- [5] Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). "Perbandingan metode naive bayes dan support vector machine pada analisis sentimen twitter." SMATIKA JURNAL: STIKI Informatika Jurnal, 10(02), pp. 71-76.
- [6] Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi." Jurnal Teknoinfo, 14(2), pp. 115-123.
- [7] Darwis, D., Siskawati, N., & Abidin, Z. (2021). "Penerapan analisis sentimen pada pengguna twitter menggunakan metode K-Nearest Neighbor." Jurnal Tekno Kompak, 15(1), pp. 131-145.
- [8] Permadi, V. A. (2020). "Analisis sentimen menggunakan algoritma Naive Bayes terhadap review restoran di Singapura." Jurnal Buana Informatika, 11(2), pp. 141-151.
- [9] Anwari, V. B., & Yuliazmi, Y. (2022). "Implementasi Algoritma K-Nearest Neighbors Pada Analisis Sentimen Masyarakat Terhadap Penerapan Pemberlakuan Pembatasan Kegiatan Masyarakat." SKANIKA, 5(1), pp. 72-81.
- [10] Yulita, W. (2021). "Analisis sentimen terhadap opini masyarakat tentang vaksin COVID-19 menggunakan algoritma naive bayes classifier." Jurnal Data Mining dan Sistem Informasi, 2(2), pp. 1-9.
- [11] Waworundeng, J. M. S., Sandag, G. A., Sahulata, R. A., & Rellely, G. D. (2022). "Sentiment Analysis of Online Lectures Tweets using Naive Bayes Classifier." CogITo Smart Journal, 8(2), pp. 371-384.