# Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naïve Bayes

# (Studi Kasus : Shopee)

**Abdul Halim Hasugian1, M.Fakhriza2, Dinda Zukhoiriyah3**

1,2,3) Ilmu Komputer, Universitas Islam Negeri Sumatera Utara

Email: 1 abdulhalimhasugian@uinsu.ac.id, 2 fakhriza@uinsu.ac.id, 3,\* dindazukhoiryah454@gmail.com

Email Penulis Korespondensi: [dindazukhoiryah454@gmail.com](mailto:dindazukhoiryah454@gmail.com)

|  |  |  |
| --- | --- | --- |
| **Article History:**  Received Jun 12th, 202x  Revised Aug 20th, 202x  Accepted Aug 26th, 202x |  | **Abstrak** |
|  | Ulasan produk sangat memberi pengaruh terhadap produsen dan konsumen karena ulasan menjadi sumber informasi tentang kualitas suatu produk. Mengolah data dalam jumlah besar dan banyak untuk setiap produk pada sebuah toko di Shopee, menganalisis serta menyimpulkan informasi secara manual membutuhkan waktu lama dan tidak efektif. Sistem analisis sentimen dibutuhkan untuk mengekstrak informasi penting dan membuat sistem yang secara objektif dapat menentukan kualitas produk dan memproses data teks besar. Sistem analisis sentimen memiliki beberapa tahap yairu crawling, preprocesssing, pembobotan kata, dan klasifikasi sentimen. Penulis mengumpulkan data sebanyak 600 data *review* menggunakan *API* Shopee untuk menghasilkan sentimen positif dan negatif. *Dataset* dibagi menjadi data latih dan data uji dengan pembagian 400 data latih dan 200 data uji dari total dataset. Klasifikasi sentimen data latih dilakukan dengan menggunakan metode Lexicon Based. Pelabelan dilakukan secara otomatis dengan cara menghitung skor setiap *term* sentimen dengan Lexicon Based. Penelitian ini menggunakan klasifikasi dengan Algoritma Naïve Bayes menggunakan pengujian confusion matrix untuk mendapatkan hasil accuracy, presision dan recall. Hasil pengujian menunjukkan nilai accuracy sebesar 99,5%, precision sebesar 99,49%, recall sebesar 100%. Dengan demikian dapat disimpulkan metode klasifikasi algoritma Naïve Bayes cukup relevan meskipun akurasi nya belum 100%.  **Kata Kunci :** Analisis Sentimen, Lexicon Based, Naïve Bayes, Review, Shopee. |
| ***Abstract*** |
| *Product reviews greatly influence producers and consumers because reviews are a source of information about the quality of a product. Processing large and large amounts of data for each product in a shop on Shopee, analyzing and summarizing information manually takes a long time and is ineffective. A sentiment analysis system is needed to extract important information and create a system that can objectively determine product quality and process large text data. The sentiment analysis system has several stages, namely crawling, preprocessing, word weighting, and sentiment classification. The author collects 600 data reviews using the Shopee API to generate positive and negative sentiments. The dataset is divided into training data and test data by dividing 400 training data and 200 test data from the total dataset. The training data sentiment classification is carried out using the Lexicon Based method. Labeling is done automatically by calculating the score for each term sentiment using Lexicon Based. This study uses classification with the Naïve Bayes Algorithm using confusion matrix testing to get accuracy, precision and recall results. The test results show an accuracy value of 99.5%, a precision of 99.49%, a recall of 100%. Thus it can be concluded that the classification method of the Naïve Bayes algorithm is quite relevant even though its accuracy is not 100%.*  ***Keyword :*** *Sentiment Analysis, Lexicon Based, Naïve Bayes, Review, Shopee* |

**1. PENDAHULUAN**

Jual beli barang dan jasa di Internet berkembang pesat selama ini. Salah satu marketplace popular untuk belanja secara online di Indonesia adalah Shopee. Shopee menawarkan pengguna dalam melakukan sebuah transaksi barang secara online. Namun tidak jarang konsumen ragu untuk membeli produk dengan harga pasar yang cukup tinggi secara online, seperti produk smartphone. Diyakini bahwa pelanggan masih kurang percaya untuk membeli smartphone sacara online[1]. Ulasan produk memberikan informasi mengenai kualitas suatu produk serta memberi berdampak besar bagi konsumen dan produsen. Ulasan produk terdiri dari peringkat bintang serta komentar ulasan yang mencakup tanggapan, pujian dan kritik, serta umpan balik atas produk yang dibeli. Kepuasan pelanggan adalah isu penting yang menjadi tujuan perusahaan. Disadari atau tidak, ulasan pelanggan yang diposting di media sosial memengaruhi pelanggan potensial dengan cara besar dan kecil [2]. Di zaman sekarang ini, banyak sentimen atau opini muncul di situs ulasan produk, yang mengarah ke banyak ulasan. Ketersediaan data *review* atau ulasan yang menjadi alasan penelitian ini menjadikan review produk untuk menganalisis opini [3].

Mengolah data dalam jumlah besar dan banyak untuk setiap produk pada sebuah toko di shopee, menganalisis dan menyimpulkan data secara manual membutuhkan waktu lama dan tidak efektif [4]. Pendapat mungkin mengandung informasi tidak lengkap dan tidak konsisten. Maka cara untuk mendapatkan informasi penting dan membuat sistem yang bisa menentukan menentukan kualitas produk untuk mengolah data teks yang sangat besar secara objektif dan otomatis adalah dengan menggunakan sistem analisis sentimen [5].

Analisis sentimen adalah metode untuk mengekstrak data opini, memahami data teks dan mengolahnya dengan otomatisasi agar dapat melihat emosi yang terkandung dalam opini tersebut. Tugas utama analisis sentimen adalah mengklasifikasikan teks kalimat atau dokumen dan kemudian ditentukan apakah pendapat tersebut dalam sentimen atau kalimat itu negatif atau positif [6].

Sistem analisis sentimen yang akan dibangun pada penelitian ini menggunakan algoritma klasifikasi Naive Bayes. Algoritma Naive Bayes berisikan asumsi yang sangat kuat (naif) bahwa setiap kondisi atau kejadian adalah independen. Keunggulan Naive Bayes merupakan proses klasifikasi data yang dapat disesuaikan dengan karakter dan kebutuhan [7].

Penelitian Darwis [8], klasifikasi menggunakan Naïve Bayesian Classifier (NBC). Sistem ini dibuat dengan memanfaatkan data yang dikumpulkan dari Internet/Twitter untuk menentukan opini netral, negatif atau positif. Penentuan ini diklasifikasikan dengan proses klasifikasi. Rating dapat memudahkan pengguna dalam melihat sentimen netral, negatif atau positif. Hasil pengujian akurasi klasifikasi Naive Bayesian sebesar 69,97%.

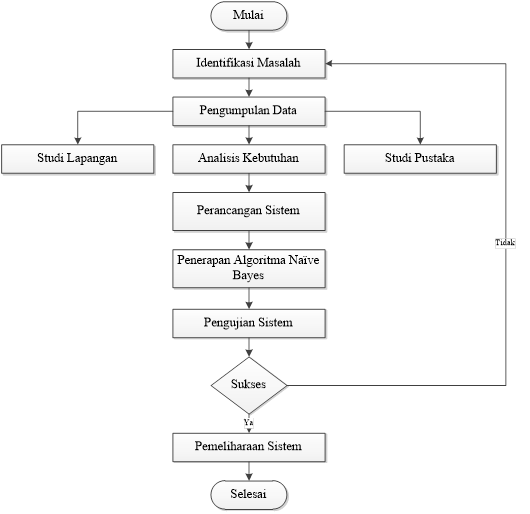
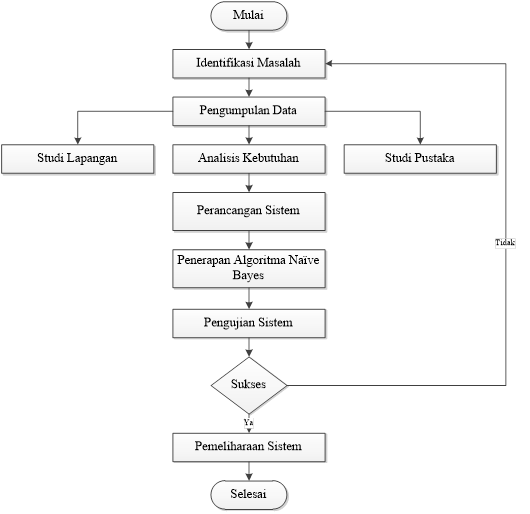
Sari,S.M [9] dalam penelitian nya menggunakan metode K-Nearest Neighbor (K-NN) pada klasifikasi sentimen terhadap New Normal. Terdapat dua data yang digunakana pada sistem yaitu data pelatihandan data pengujian*.* Pelabelan data pelatihan menggunakan metode Lexicon Based. Pengujian sistem menggunakan *Confussion Matrix*. Dari hasil masing-masing pengujian yang dilakukan diperoleh hasil dengan nilai k = 1, akurasi hasil pengujian data pelatihan adalah 94,50%, precision sebesar 100%, recall sebesar 93,89% dan f-measure sebesar 96,89%.

Pada penelitian ini produk yang akan di analisis sentimen dari ulasan pelanggan yang terdapat pada produk adalah merupakan produk Iphone *secondhand* yang dijual oleh Kinophonecell melalui marketplace Shopee. Alasan pemilihan produk tersebut adalah dikarenakan banyaknya penjualan serta banyaknya ulasan yang telah dibuat oleh pelanggan yang telah membeli produk tersebut. Berdasarkan masalah dan pemaparan di atas, penulis berencana untuk melakukan penelitian mengenai *feedback* atau *review* dari pengguna *e-commerce* untuk membantu pelaku bisnis untuk melakukan evaluasi dalam mengetahui kualitas dari suatu produk yang dibeli dengan hasil analisis berupa sentimen positif dan negatif bagi tiap produk. Dengan dibangunnya sistem analisis sentimen ini diharapkan dapat masyarakat yang mempromosikan suatu produk pada marketplace mengetahui umpan balik terhadap produknya dan mengetahui opini yang disampaikan oleh pengguna *marketplace* terhadap produk yang ditawarkan.

**2. METODOLOGI PENELITIAN**

**2.1 Tahap Persiapan**

Penelitian ini adalah penelitian dengan tujuan untuk mengembangkan suatu aplikasi berbasis web yang bisa dimanfaatkan untuk melakukan analisis sentimen menggunakan agoritma Naive Bayes. Penelitian ini menggunakan metode Research and Development (R&D). Metode Research and Development digunakan untuk membuat suatu produk kemudian menguji apakah produk tersebut efektif dalam melakukan analisis sentimen terhadap review produk yang dibuat oleh pelanggan *e-commerce* Shopee [10]. Tahapan sistem dalam melakukan analisis sentimen menggunakan Algoritma Naïve Bayes dapat dilihat dalam suatu alur seperti yang ditunjukkan pada gambar 1.



Gambar 1. Kerangka Penelitian

Masalah yang teridentifikasi pada penelitian adalah analisis sentimen terhadap review pelanggan pada produk Iphone XR 256GB/128GB/64GB di toko Kinophonecell website Shopee. Pengumpulan data dilakukan dengan 2 cara yaitu pertama cara studi literatur, mempelajari sumber-sumber dan literatur yang berkenaan pada materi skripsi. Kedua observasi yaitu mengumpulkan data review produk Iphone XR 256GB/128GB/64GB di toko Kinophonecell website Shopee.

**2.2 Tahap Perancangan**

Perancangan sistem dalam penelitian ini menggunakan *Flowchart* sistem untuk menggambarkan secara akurat semua langkah desain yang terlibat dalam pembuatan aplikasi*.*

Gambar 2. *Flowchart* Sitem

Data diambil menggunakan *API* yang telah disediakan oleh Shopee dan di crawling dengan menggunakan bahasa pemrograman PHP, kemudian data akan disimpan ke *database*. Data yang dikumpulkan sebanyak 600 dataset yang kemudian dibagi menjadi 400 data pelatihandan 200 data pengujian.

1. **HASIL DAN PEMBAHASAN**

Proses pengumpulan data dilakukan dengan menggunakan metode crawling di website Shopee dengan memanfaatkan fasilititas *Application Program Interface (API)* yang disediakan oleh Shopee dengan alamat https://shopee.co.id. Alur crawling data tersebut terlihat seperti gambar berikut :



Gambar 3. Alur Sistem *Crawling* data Review Shope

Data Latih akan dilakukan pemberian label secara otomatis dengan Lexicon Based ke dalam beberapa kelas diantaranya positif, dan negatif, sesuai dengan sentimen kata yang berada pada review produk. Penentuan sentimen data latih dilakukan dengan menghitung probabilitas kemunculan kata kunci positif dan negatif dari kamus Liu Lexicon serta penambahan kata yang berkaitan dengan topik ulasan produk.

Untuk menjelaskan proses data pelatihan dan data pengujian, diberikan 2 contoh data pelatihan dan 1 data pengujian dari seluruh data set sebanyak 600 data.

Tabel 1. Data Latih

|  |  |  |
| --- | --- | --- |
| **No** | ***Text*** | ***Kelas*** |
| 1 | Barang sesuai deskripsi ,barang sampai dengan aman dalam kondisi baik,pengemasan cepat respon seller fastrepon semoga selalu amanah,good job 👍👍👍 | Positif |
| 2 | respon penjual lama, katanya body lecet 99-98% tp ini lecet parahh disemua sisi lecetnya. kecewa bgt | Negatif |

Data uji berjumlah 200 data yang diambil dari data set, yang ditampilkan pada tabel 2 berikut :

Tabel 2. Data Uji

|  |  |  |
| --- | --- | --- |
| **No** | ***Text*** | ***Kelas*** |
| 1 | pengemasan cepat fast respon admin ramah pisan. mulus minta bh 95 d kasih 💯 % warna sesuai pesanan. tapi layar warna nya tuh kay ga cerah gitu. aku br pake ip yg ponian gini si. beda sm ip7+ aku. ga tau emng gitu ato gmn. tp overall aku suka. mudah”n awet. pertama x beli d sini ga kecewa👍 | ? |

Dari data set yang didapat, kemudian akan diolah pada proses *Pre-rocessing* data. Untuk melakukan preprocessing maka terdapat 6 tahapan yang perlu dilakukan. Alur tahap preprocessing data yang meliputi *cleaning, case folding, normalisasi, filtering, stemming, tokenizing .*

Setelah melawati 6 tahap *pre-processing,* maka hasil akhir dari tahap *tokenizing* adalah sebagai berikut :

Tabel 3. Hasil *Tokenizing*

| **Hasil *Tokenizing*** | | | |
| --- | --- | --- | --- |
| barang  sesuai  deskripsi  barang | sampai  aman  baik  ngemas | cepat  respon  seller  fastrespon | semoga  selalu  amanah  good  job |
| respon  penjual  lama | kata  body  lecet | lecet  parah  sisi | lecet  kecewa |
| ngemas  cepat  fast  respon  admin  ramah  mulus | bh  kasih  warna  sesuai  pesan  layar  warna | cerah  baru  pakai  iphone  iphone  overall | suka  mudah  awet  pertama  beli  kecewa |

Setelah tahapan *preprocessing*, kata hasil *tokenizing* diberikan bobot dengan fitur *tf (term frequency).* Pembobotan adalah konversi kata menjadi angka atau vektor [11].

Tabel 4. Hasil *Term Frequency* Data

| Data Latih Positif | | | Data Latih Negatif | | | Data Uji | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| no | Term | Frekuensi (nk) | no | Term | Frekuensi (nk) | no | Term | Frekuensi (nk) |
| 1 | barang | 2 | 1 | respon | 1 | 1 | ngemas | 1 |
| 2 | sesuai | 1 | 2 | penjual | 1 | 2 | cepat | 1 |
| 3 | deskripsi | 1 | 3 | lama | 1 | 3 | fast | 1 |
| 4 | sampai | 1 | 4 | kata | 1 | 4 | respon | 1 |
| 5 | aman | 1 | 5 | body | 1 | 5 | admin | 1 |
| 6 | baik | 1 | 6 | lecet | 3 | 6 | ramah | 1 |
| 7 | ngemas | 1 | 7 | parah | 1 | 7 | mulus | 1 |
| 8 | cepat | 1 | 8 | sisi | 1 | 8 | bh | 1 |
| 9 | respon | 1 | 9 | kecewa | 1 | 9 | kasih | 1 |
| 10 | seller | 1 |  |  |  | 10 | warna | 2 |
| 11 | fastrespon | 1 |  |  |  | 11 | sesuai | 1 |
| 12 | semoga | 1 |  |  |  | 12 | pesan | 1 |
| 13 | selalu | 1 |  |  |  | 13 | layar | 1 |
| 14 | amanah | 1 |  |  |  | 14 | cerah | 1 |
| 15 | good | 1 |  |  |  | 15 | baru | 1 |
| 16 | job | 1 |  |  |  | 16 | pakai | 1 |
|  |  |  |  |  |  | 17 | iphone | 2 |
|  |  |  |  |  |  | 18 | overall | 1 |
|  |  |  |  |  |  | 19 | suka | 1 |
|  |  |  |  |  |  | 20 | mudah | 1 |
|  |  |  |  |  |  | 21 | awet | 1 |
|  |  |  |  |  |  | 22 | pertama | 1 |
|  |  |  |  |  |  | 23 | beli | 1 |
|  |  |  |  |  |  | 24 | kecewa | 1 |

## Klasifikasi Naïve Bayes

1. Menghitung nilai priors

Nilai prioritas dapat dicari dengan menggunakan rumus

(1)

1. Menghitung nilai probabilitas setiap term

Probabilitas dari setiap term dapat dicari dengan menggunakan rumus :

(2)

Frekuensi data uji akan diklasifikasi dicari nilai probabilitasnya dengan membandingkan dan mencocokkan dengan term yang ada pada sentimen positif dan sentimen negatif data latih

* + - 1. Kelas Sentimen Positif

Berdasarkan nilai probabilitas data uji pada sentimen positif diatas, nilai untuk sentimen positif adalah sebagai berikut:

(3)

)

* + - 1. Kelas Sentimen Negatif

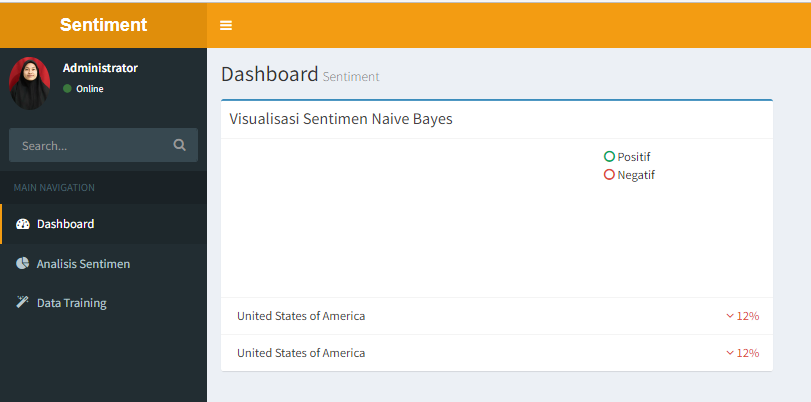
Berdasarkan nilai probabilitas data uji pada sentimen negatif diatas, nilai untuk sentimen negatif adalah sebagai berikut:

\*(0,5)

Berdasarkan perbandingan hasil perhitungan kedua sentimen, nilai sentimen positif lebih besar. Sehingga, data uji diklasifikasikan sebagai sentimen kelas **POSITIF**

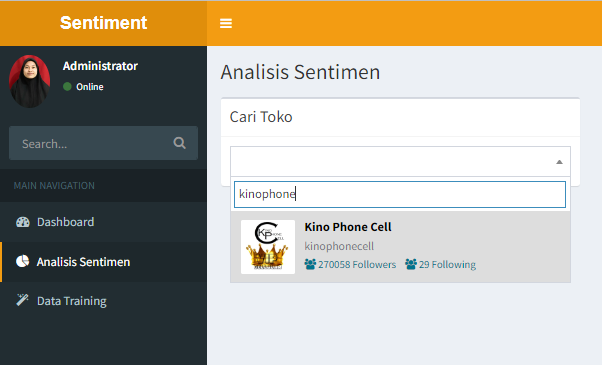
## Implementasi

Berdasarkan langkah perhitungan manual dan proses perancangan program aplikasi, langkah selanjutnya adalah melakukan langkah pengujian terhadap sistem aplikasi yang dibangun..



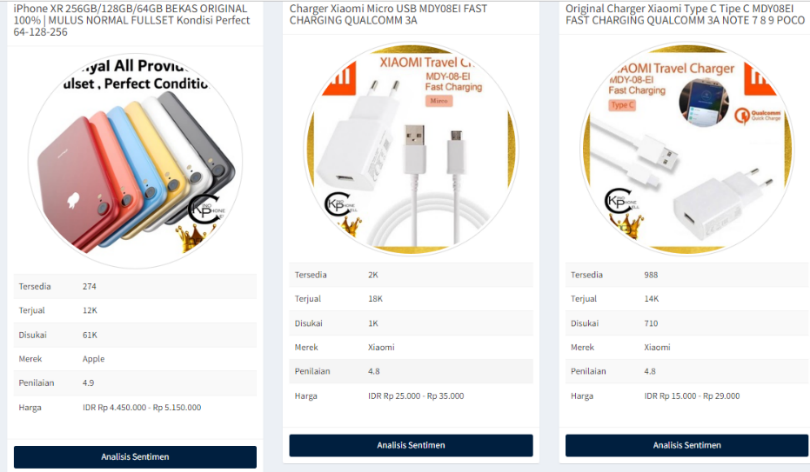
Gambar 4. Halaman Utama

Pada halaman Dashboard akan menampilkan visualisasi sentimen dengan algoritam Naïve Bayes.



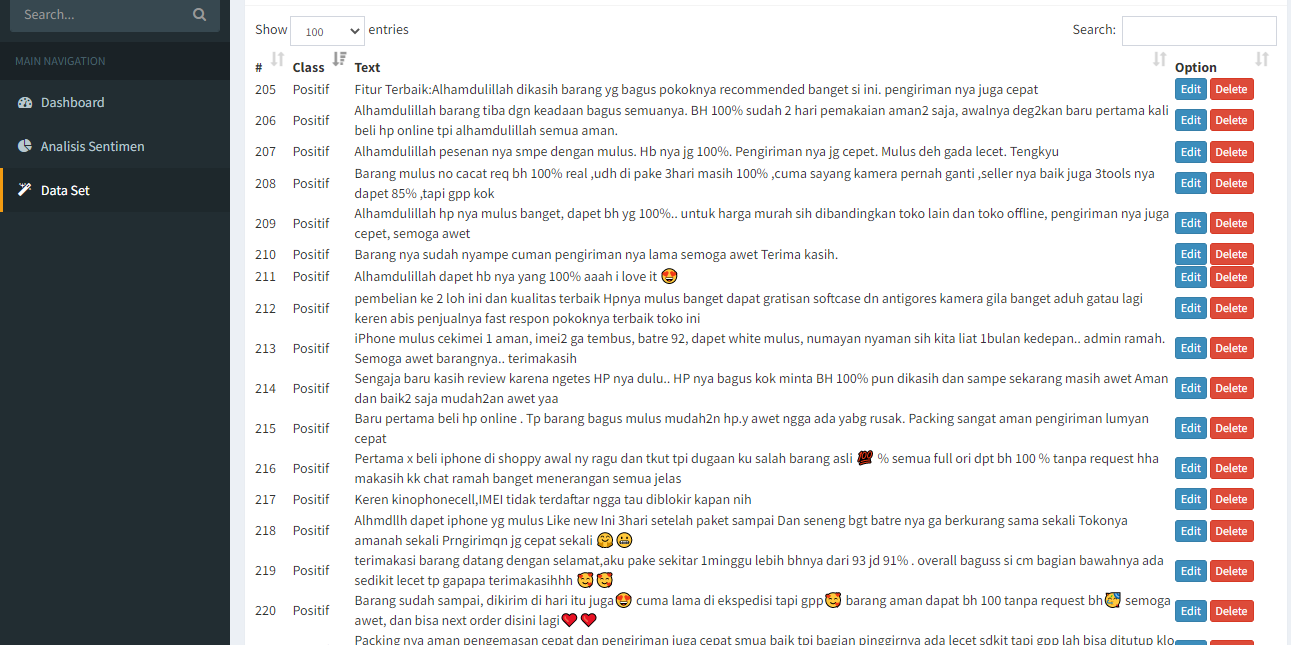
Gambar 5. Halaman Menu Cari Toko

Menu ini berfungsi untuk mencari toko yang akan dianalisis review produknya. Pada penelitian ini mengambil review pada toko Kinophonecell untuk dijadikan objek.



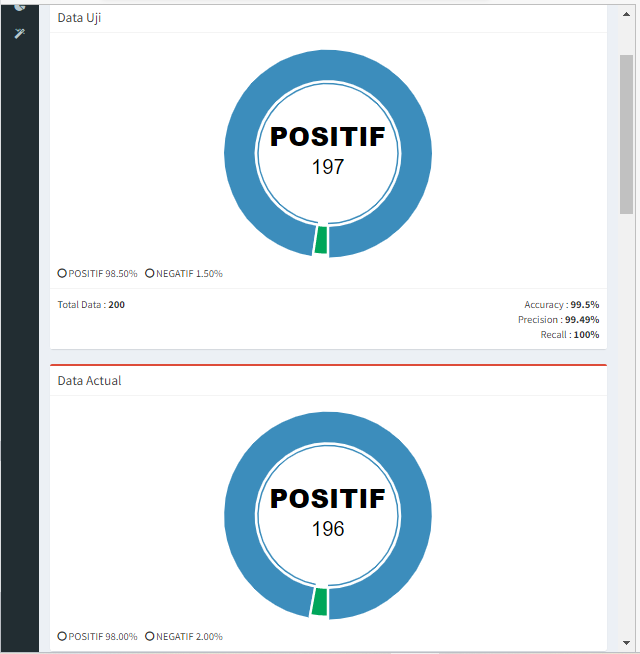
Gambar 6. Halaman Menu Produk

Pada penelitian ini, penulis membatasi produk yang akan dianalisis yaitu produk iPhone XR 256GB/128GB/64GB.

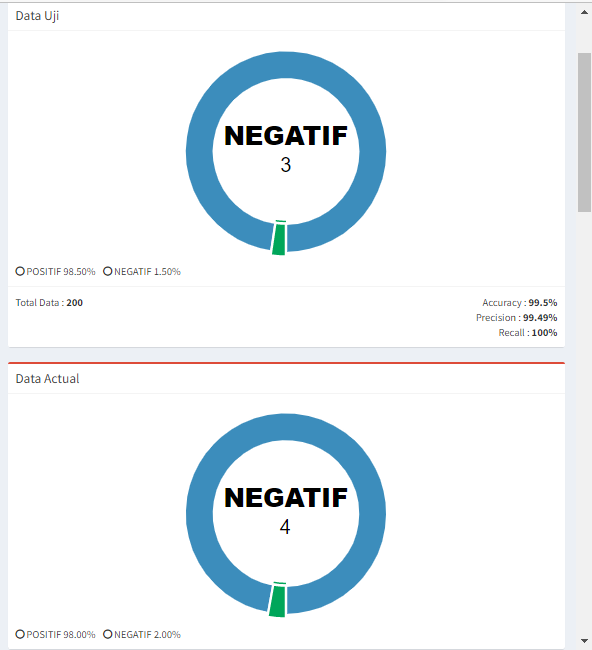


Gambar 7. Menu Data Latih

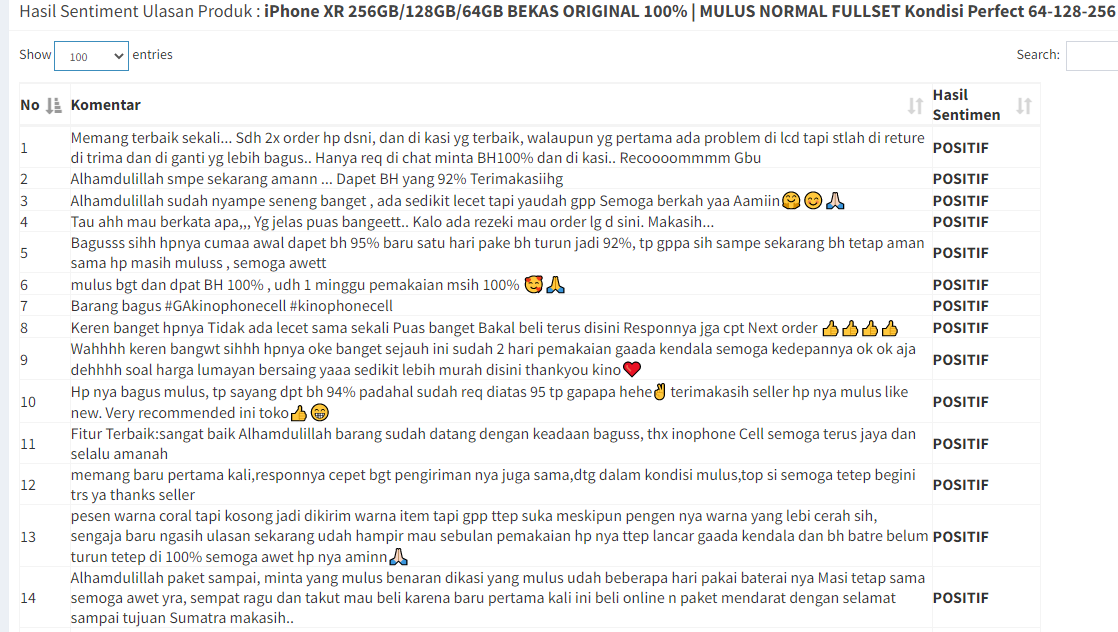
Pada gambar diatas merupakan halaman dari kumpulan data latih. Pada halaman ini terdapat data latih yang sudah dilabeli secara otomatis melalui tahap *Lexicon Based.*



Gambar 8. Tampilan Visualisasi Hasil Analisis Sentimen Positif



Gambar 9. Tampilan Visualisasi Hasil Analisis Sentimen Negatif



Gambar 10. Tampilan Hasil Analisis Data Uji

Setelah mendapatkan hasil klasifikasi, maka dilakukan pengujian berupa mengevaluasi hasil penelitian ini dengan menggunakan Confussion Matrix. Confussion Matrix digunakan untuk menghitung akurasi dan mengingat nilai untuk menentukan apakah metode yang gunakan sudah baik dalam membuat sistem klasifikasi. [12]. Pengujian data di lakukan saat membandingkan hasil perhitungan manual dan perhitungan sistem dilakukan dengan menggunakan metode Naive Bayes. Dan juga membandingkan data sentimen asli atau data actual pada data latih dengan data hasil prediksi atau data uji.

Tabel 5. Confusion Matrix Data Uji

|  |  |  |
| --- | --- | --- |
| Prediksi | Data Actual | |
| Positif | Negatif |
| Positif | 196 | 1 |
| Negatif | 0 | 3 |

Dari data pada Tabel 5. maka dapat dihitung *acuracy, precision, recall* sebagai berikut :

Tabel 6. Hasil Evaluasi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Jumlah Kelas | Total Data | *Accuracy* | *Precision* | *Recall* |
| 2 | 600 |  |  |  |

Dari hasil evaluasi tabel 6 diatas, dengan demikian dapat disimpulkan metode klasifikasi algoritma Naïve Bayes cukup relevan meskipun akurasi nya belum 100%

1. **KESIMPULAN**

Pada penelitian ini dapat ditarik kesimpulan yaitu metode Naive Bayesian dapat digunakan untuk memprediksi kategori sentimen pada review produk di toko Kinophonecell *pada platform e-commerce* Shopee dengan sistem yang telah dibangun. Analisis sentimen review pengguna Shopee dengan metode Naïve Bayes ini menghasilkan *accuracy* dengan satu kali uji dengan nilai sebesar *accuracy* sebesar 99,5%, *precision* sebesar 99,49%, *recall* sebesar 100%. Dengan demikian dapat disimpulkan metode klasifikasi algoritma Naïve Bayes cukup relevan meskipun akurasi nya belum 100%.

Hasil analisis penelitian ini menghasilkan sentimen positif lebih banyak berjumlah 197 dan jumlah sentimen negatif sebanyak 3. Jumlah data pelatihan sistem memengaruhi prediksi sistem. Selain kuantitas, kualitas data training juga penting, karena semakin tinggi data maka semakin luas perbendaharaan kata sistem sehingga semakin akurat dalam memprediksi kategori sentimen.

**UCAPAN TERIMA KASIH**

Terima kasih kepada bapak Abdul Halim Hasugian, M.Kom selaku penulis pertama yang telah membimbing serta memberi arahan dalam penyelesaian artikel ini, dan terima kasih kepada bapak Fakhriza, M.Kom selaku penulis kedua yang telah bekerja sama serta memberi arahan dalam penyelesaian artikel ini dan terima kasih kepada semua pihak yang terkait dan atas kerjasamanya hingga artikel ini diterbitkan.

**DAFTAR PUSTAKA**

[1] Iwan Setiawan, M. R. (2021). Rancang Bangun Aplikasi Sistem Niaga Unit Kegiatan Mahasiswa Politeknik Pos Indonesia Berbasis Web Framework (Sub Modul Pembelian). Jurnal Teknik Informatika, 13(2), 52–58

[2] Andika Sari, Mia., & Analysis of Factors Influencing Consumer Purchase Interest Online in Market Place. Journal of Economics & Business Volume 18 Number 2. Pages 97-108.

[3] Amrullah, A. Z., Sofyan Anas, A., & Hidayat, M. A. J. (2020). Sentiment Analysis of Movie Review Using Naive Bayes Classifier With Chi Square Feature Selection. Journal, 2(1), 40–44. <https://doi.org/10.30812/bite.v2i1.804>

[4] Yulita, W., Dwi Nugroho, E., Habib Algifari, M., Informatics Engineering Studies, P., Sumatran Technology, I., Ryacudu Canal, J., Huwi, W., Agung, J., & South, L. (2021). Sentiment Analysis of Public Opinion About the Covid-19 Vaccine Using the Naïve Bayes Classifier Algorithm. Journal of Data Mining and Information Systems, 2(2), 1–9. <https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1344>.

[5] J. Edukasi, B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” vol. 4, no. 2, pp. 113–118, 2018.

[6] Lestari, S., & Saepudin, S. (2021). Sinovac Vaccine Sentiment Analysis on Twitter Using the Naive Bayes Algorithm. Sismatic (National Seminar on Information Systems and Information Management), 163–170.

[7] Herdhianto, A., Studies, P., Informatics, T., Science, F., Technology, AND, Islam, U., & Syarif, N. (2020). Sentiment Analysis Using Naïve Bayes Classifier ( Nbc ) on Tweet About Zakat.

[8] Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Compact Techno Journal* , *15* (1), 131. https://doi.org/10.33365/jtk.v15i1.744.

[9] Sari, SM (2021). Sentiment Analysis of New Normal in the Covid-19 Era Using the K-Nearest Neighbor (K-Nn) Algorithm.

[10] Mogana, AM (2017). Research And Development Methods. *Kerangka Konsep Penelitian* , *53* (9), 1–15. http://www.elsevier.com/locate/scp.

[11] Prasetya, Y. N, Winarso, D., & Syahril. (2021). Lexicon Based Application for Sentiment Analysis on Twitter. 11(2), 97–103.

[12] Hamilton , Howard. 2017. Confusion Matrix. http://www.cs.urgina.ca/~hamilton/course/notes/confusion\_matrix

/confusion matrix.html, 24 January 2022.