



Analisis Sentimen terhadap Ulasan *Cashless* Menggunakan Metode *Knowledge Discovery Databases* Berbasis Web

Annisa Farah^{1*}, Edy Rahman Syahputra²

¹⁻² Universitas Harapan Medan, Indonesia

Jl. HM. Joni No.70, Teladan Bar., Kec. Medan Kota, Kota Medan, Sumatera Utara

Korespondensi penulis: ydeaja@yahoo.com

Abstract. *Sentiment analysis is the process of automatically extracting, processing, and understanding unstructured text data to obtain sentiment information contained in an opinion. This study conducted a sentiment analysis of cashless reviews using a web-based knowledge discovery database method. The knowledge discovery database method aims to extract hidden knowledge or information in previously unknown data, through the stages of data selection, preprocessing, transformation, data mining, and evaluation of 300 training data and 250 test data. Based on the original labels, there were 46 positive comments and 204 negative comments, while the prediction results showed 34 positive comments and 216 negative comments. The evaluation process produced an accuracy rate of 83.2% for positive labels and 83.2% for negative labels. The precision and recall values differed for each label, namely: positive labels had a precision of 55.88% and negative labels of 87.50%; positive labels had a recall of 41.30% and negative labels of 92.65%. Based on these scores, negative reviews were more dominant than positive reviews regarding the use of cashless. Positive sentiment focused on the ease of transactions through cashless payments, while negative sentiment focused on technical issues during payments, refunds, and frozen PayLater accounts. Therefore, e-commerce platforms implementing cashless payments need to provide better education and ensure transparency of digital processes to ensure users feel safer and more comfortable with cashless transactions.*

Keywords: *Cashless, Databases, Knowledge Discovery Methods, Review Comments, Websites.*

Abstrak. Analisis sentimen atau *sentiment analysis* merupakan proses mengekstraksi, mengolah, dan memahami data berupa teks yang tidak terstruktur secara otomatis untuk memperoleh informasi sentimen yang terdapat pada suatu pendapat atau opini. Penelitian ini melakukan analisis sentimen terhadap ulasan *cashless* menggunakan metode *knowledge discovery databases* berbasis web. Metode *knowledge discovery databases* bertujuan untuk mengekstraksi pengetahuan atau informasi tersembunyi dalam data yang sebelumnya tidak diketahui, melalui tahapan *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation* terhadap 300 data latih serta 250 data uji. Berdasarkan label asli, terdapat 46 komentar positif dan 204 komentar negatif, sedangkan hasil prediksi menunjukkan 34 komentar positif dan 216 komentar negatif. Proses evaluasi menghasilkan tingkat akurasi sebesar 83,2% untuk label positif dan 83,2% untuk label negatif. Nilai *precision* dan *recall* berbeda untuk masing-masing label, yaitu: label positif memiliki *precision* sebesar 55,88% dan label negatif sebesar 87,50%; label positif memiliki *recall* sebesar 41,30% dan label negatif sebesar 92,65%. Berdasarkan skor tersebut, ulasan negatif lebih dominan dibandingkan ulasan positif terhadap penggunaan *cashless*. Sentimen positif berfokus pada kemudahan transaksi melalui pembayaran *cashless*, sedangkan sentimen negatif berfokus pada masalah teknis saat pembayaran, pengembalian dana, dan pembekuan akun *paylater*. Oleh karena itu, pihak e-commerce yang menerapkan *cashless* perlu memberikan edukasi yang lebih baik serta memastikan transparansi proses digital, agar pengguna merasa lebih aman dan nyaman dalam bertransaksi menggunakan *cashless*.

Kata kunci: *Cashless, Ulasan Komentar, Metode Knowledge Discovery, Databases, Website.*

1. LATAR BELAKANG

Perkembangan teknologi digital dunia telah berkembang pesat belakangan ini, terkhusus pada bidang e-commerce. Adanya perkembangan teknologi ini membuat perubahan yang membawa alternatif dan kemudahan dalam bertransaksi antara penjual (seller) dan pembeli (buyer). Perubahan teknologi menciptakan kebiasaan baru yaitu sebuah fenomena yang terletak pada perubahan metode yang digunakan dalam transaksi konvensional atau tunai

menuju transaksi online atau cashless yang merupakan fenomena masyarakat bertransaksi tanpa uang tunai (Marsela et al., 2022).

Seiring dalam perkembangan dunia, saat ini e-commerce termasuk suatu kebutuhan untuk meningkatkan serta memenangkan persaingan bisnis dan penjualan produk . Pada proses penggunaan e-commerce kegiatan jual beli maupun pemasaran lebih efisien dimana penggunaan e-commerce tersebut akan memperlihatkan adanya kemudahan bertransaksi (Fitriyani Yapan, n.d.).

Cashless termasuk dalam suatu kegiatan transaksi yang mengganti fisik uang berupa lembaran dan logam dengan uang elektronik (e-money) berupa kartu debit atau kredit dan internet banking atau mobile banking sebagai alat transaksi pembayaran (Wulandari et al., 2023). Beberapa penyedia layanan cashless memberikan promosi berupa potongan harga, gratis ongkir, cashback atau poin loyalitas. Hal ini tentu menarik keinginan pembeli dalam melakukan transaksi pembelian memakai sistem cashless (Aghitsni & Busyra, 2022).

Namun, di balik pertumbuhan pesat penggunaan sistem cashless terdapat pengguna yang mengalami kegagalan transaksi namun saldonya tetap terpotong dan tidak terdaftar di mutasi sehingga tidak bisa membuat bukti dan itu terjadi lebih dari sekali. selain itu, para konsumen juga kerap melaporkan berbagai permasalahan teknis seperti error atau gangguan saat melakukan pembayaran yang menghambat mereka dalam melakukan transaksi, metode pembayaran yang tidak memadai untuk melakukan pengiriman gratis, keamanan pada sistem pembayaran, tidak bisa melakukan pengajuan pengembalian dana, dan juga sering lupa antara pin dan kata sandi (Hasibuan et al., 2024).

Analisis sentimen atau sentiment analysis adalah salah satu proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis untuk mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini yang memiliki makna atau emosional pesan sehingga dapat di kategorikan positif atau negatif (Arsi & Waluyo, 2021).

Penelitian ini bertujuan mengetahui efektifitas metode Knowledge Discovery Databases dalam mengklasifikasikan opini pengguna cashless pada aplikasi e-commerce yang terdapat di google playstore lewat analisis sentimen.

2. KAJIAN TEORITIS

Analisis Sentimen

Analisis sentimen merupakan proses komputasi untuk melakukan klasifikasi terhadap data tekstual sesuai dengan sentimen yang dimiliki. Selain alasan ketersediaan data teks yang

melimpah saat ini, penelitian analisis sentimen semakin populer karena alasan kebutuhan dari banyak pihak mengenai pendapat publik terhadap suatu topik tertentu. Proses analisis sentimen dipengaruhi oleh dataset yang digunakan. Untuk dataset yang terdiri dari kumpulan kalimat yang cukup panjang membutuhkan penanganan yang berbeda (Widayat, 2021).

Cashless

Sistem pembayaran *cashless payment* merupakan transaksi pembayaran digital yang saat ini banyak digunakan oleh banyak orang. Dalam khusus ini dapat disimpulkan bahwa pembayaran tidak lagi menggunakan uang tunai atau uang fisik, selain untuk memudahkan transaksi *cashless* juga lebih dipercaya keamanannya. Penggunaan *cashless payment* tidak hanya mempengaruhi kemudahan dalam transaksi, tetapi juga berpotensi untuk memengaruhi minat beli konsumen (Tazkia et al., 2024).

Knowledge Discovery Databases

Definisi *Knowledge Discovery in Database* (KDD) adalah metode dan cara mendapatkan sebuah informasi melalui basis data yang telah tersedia. Berikut tahap penyelesaiannya, yaitu :

- *Data Selection*
Data selection adalah pengambilan data dari sebuah dataset yang nantinya akan diolah.
- *Data Preprocessing*
Data preprocessing adalah tahap untuk melakukan sebuah proses awal dalam pengolahan data. Pada tahap ini data yang akan diolah bertujuan untuk menghindarkan dari data yang mengganggu (*noise*) atau data yang tidak konsisten.
- *Transformation*
Transformation adalah tahap membuat data yang akan diproses menjadi sesuai dengan model ataupun algoritma yang ingin digunakan dalam tahap pemrosesan data.
- *Data Mining*
Proses pencarian dan penggalian pengetahuan hingga diperoleh suatu model yang dapat digunakan.
- *Evaluation*
Merepresentasikan hasil model yang telah diperoleh serta menguji akurasi dan kesesuaian terhadap data-data yang bersinggungan (Alghifari & Juardi, 2021).

Web Scraping

Web Scraping merupakan suatu teknik yang digunakan untuk mendapatkan suatu data atau informasi pada *website* tertentu. Informasi yang didapat berupa, teks, tautan, video, audio

ataupun dokumen. Penelitian Teknik *Web Scraping* dilakukan dengan batasan-batasan yang ditetapkan (Kusumo, 2022).

Google Colab

Google Colab (Google Colaboratory) adalah *platform* berbasis *cloud computing* yang disediakan oleh *Google*. Ini memungkinkan pengguna untuk mengeksekusi kode *Python* langsung di *browser* web tanpa perlu menginstal (Yanuar, 2024).

Website

Web didefinisikan sebagai sekelompok halaman yang berguna untuk memunculkan informasi, *text*, *silent picture*, animasi, suara, dan gabungan statis dan dinamis berbentuk pola bangunan yang saling berhubungan ke masing-masing jaringan halaman *web* (*hyperlink*) (Lim & Silalahi, 2023).

Data Mining

Data Mining adalah Serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data dengan melakukan penggalian pola pola dari data dengantujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data (Ordila et al., 2020).

Naive Bayes

Algoritma *Naive Bayes* merupakan penggabungan probabilitas dan kondisional, untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data *training*. Penggunaan *Naive Bayes* ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil saja untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Rahayu et al., 2021).

Pada proses data uji dilakukan perhitungan *prior* dan perhitungan log probabilitas. Perhitungan *prior* menggunakan rumus:

$$\log P(W_j) = \frac{N_{wj}}{N_c} \quad (1)$$

Keterangan:

$\log P(W_j)$: log *prior*

N_{wj} : Banyak data yang termasuk kelas w_j

N_c : Banyak semua data latih

Perhitungan log probabilitas setiap kelas positif atau negatif menggunakan rumus:

$$\log P(X_i|W_j) = \log \frac{N_{xi,wj}+1}{N_{wj}+|v|} \quad (2)$$

Keterangan :

$\log$	: Logaritma dari bilangan kurang dari 1 selalu negatif.
$(X_i W_j)$	: Probabilitas
N_{x_i,w_j}	: Banyak fitur muncul dalam kelas w_j (Positif atau negatif)
N_{w_j}	: Banyak data yg termasuk kelas w_j (positif atau negatif)
1	: Nilai <i>add one</i> agar nilai <i>probabilitas</i> tidak 0
$ v $	: Jumlah seluruh data kelas w_j (positif dan negatif)

Pada perhitungan probabilitas log diperoleh jumlah dari total log probabilitas kelas positif atau negatif, dimana nilai tertinggi dari hasil log probabilitas kelas tersebut maka dapat di kategorikan menjadi label prediksi sentimen positif atau negatif .

Confusion Matrix

Confusion Matrix ini biasa digunakan untuk mengukur kinerja dari machine learning dan dapat digunakan sebagai alat visual untuk membantu dalam mengevaluasi *accuracy*, *precision* dan *recall*. Model klasifikasi Algoritma *Naive Bayes* (Indransyah et al., 2022).

Accuracy merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. *Accuracy* menjawab pertanyaan “Berapa persen tagihan yang benar diprediksi Prioritas dan Tidak Prioritas” dengan menggunakan persamaan $\frac{TP+TN}{TP+TN+FP+FN}$

Precision merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Precision* menjawab pertanyaan “Berapa persen tagihan yang benar Prioritas dari keseluruhan tagihan yang diprediksi” dengan menggunakan persamaan. $\frac{TP}{TP+FP}$

Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. *Recall* menjawab pertanyaan “Berapa persen tagihan yang diprediksi Prioritas dibandingkan keseluruhan tagihan yang sebenarnya Prioritas” dengan menggunakan persamaan. $\frac{TP}{TP+FN}$

3. METODE PENELITIAN

Bahan dan Alat Penelitian

Bahan penelitian dari Analisis Sentimen Terhadap Ulasan *Cashless* Menggunakan Metode *Knowledge Discovery Databases* Berbasis WEB membutuhkan beberapa perangkat agar sesuai kebutuhan sistem. Kebutuhan sistem ini terbagi menjadi dua bagian yaitu:

Tabel 1 Perangkat Keras

Perangkat Keras (<i>Hardware</i>)
Processor Core i3
Memory 4 GB

Tabel 2 Perangkat Lunak

Perangkat Lunak (<i>Software</i>)
Googel Colab
Xampp
Personal Hypertext Processor (PHP)
Notepad++
Google Chrome Notepad++

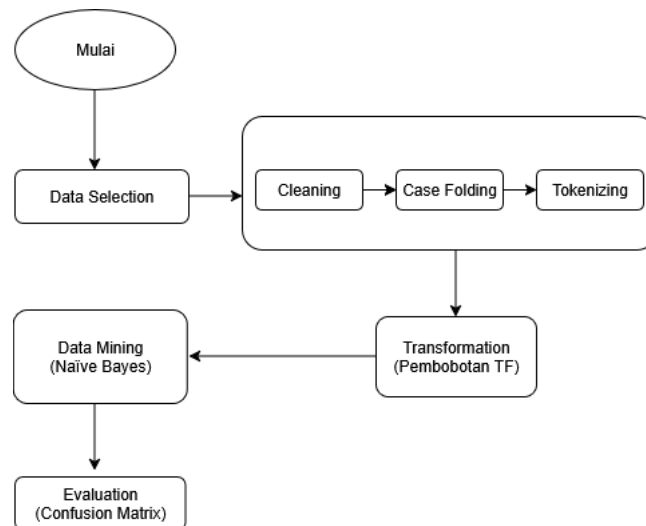
Prosedur Penelitian

Prosedur penelitian dilakukan untuk mendapatkan penjelasan berupa langkah-langkah dalam menganalisis terhadap ulasan cashless menggunakan metode KDD berbasis Website. Pada tahap pertama prosedur penelitian terdapat identifikasi masalah, indentifikasi masalah tersebut telah di jelaskan pada bagian latar belakang dari penelitian.

Metode Pengumpulan Data

Pada proses *Scraping* Data, penelitan ini dilakukan dengan pengumpulan data dengan *web scraping* pada *play store* dengan menggunakan python yang diolah dari *google colab* dengan kata kunci yang berhubungan degan *cashless*. Terdapat dua jenis yang diperoleh dari proses *scraping* yaitu, data latih sebanyak 300 ulasan dan data uji sebanyak 250 ulasan.

Penelitian ini menggunakan metode KDD karena memiliki keunggulan proses pengidentifikasian pola yang terorganisir dari sekumpulan data yang kompleks, sehingga datanya mudah untuk dipahami. Di bawah ini merupakan alur penelitian metode *Knowledge Discovery Databases*.



Penjelasan mengenai alur metode Knowledge Discovery Databases sebagai berikut:

- *Data Selection*

Pada tahap ini akan dilakukan pengumpulan, seleksi dan pelebelan data. Data dikumpulkan dari ulasan aplikasi *e-commerce* di situs *google playstore* menggunakan *tools google colab* sebagai proses *scraping* untuk memperoleh data ulasan yang berhubungan dengan penggunaan *cashless*.

- *Preprocessing*

Pada tahap ini data mentah yang telah dikumpulkan akan diolah menjadi data yang dapat digunakan pada tahap selanjutnya. Adapun 3 tahapan yang termasuk kedalam *preprocessing* yaitu:

- *Cleaning*, merupakan proses untuk menghilangkan atribut yang tidak berpengaruh terhadap klasifikasi yaitu tanda baca, karakter kosong, dan emoji.
- *Case Folding*, merupakan proses untuk mengubah semua huruf yang ada pada dokumen menjadi huruf kecil.
- *Tokenizing*, merupakan proses untuk memisahkan kalimat menjadi beberapa bagian kata.

- *Transformation*

Pada tahap ini pembobotan kata menggunakan TF (*Term Frequency*) pada jumlah term positif dan term negatif.

- *Data Mining*

Pada tahap ini akan dilakukan proses klasifikasi data latih dan data uji untuk memperoleh sentimen positif atau negatif pada data ulasan dengan menggunakan algoritma *Naive Bayes*.

- *Evaluation*

Pada tahap ini menggunakan *confusion matrix* untuk mengetahui performansi hasil evaluasi berupa nilai *accuracy*, *precision* dan *recall*.

Analisis Kebutuhan

Analisis kebutuhan mencakup aspek fungsional dan non-fungsional. Secara fungsional, sistem memungkinkan pengguna login dan logout, mengimpor serta menyimpan data ulasan dari file CSV, melakukan pembersihan data, case folding, dan tokenisasi. Sementara itu, dari sisi non-fungsional, sistem diharapkan memiliki performa tinggi dalam menampilkan hasil analisis sentimen secara cepat serta mendukung penggunaan multibahasa, khususnya Bahasa Indonesia dan Inggris.

Pembuatan Perancangan

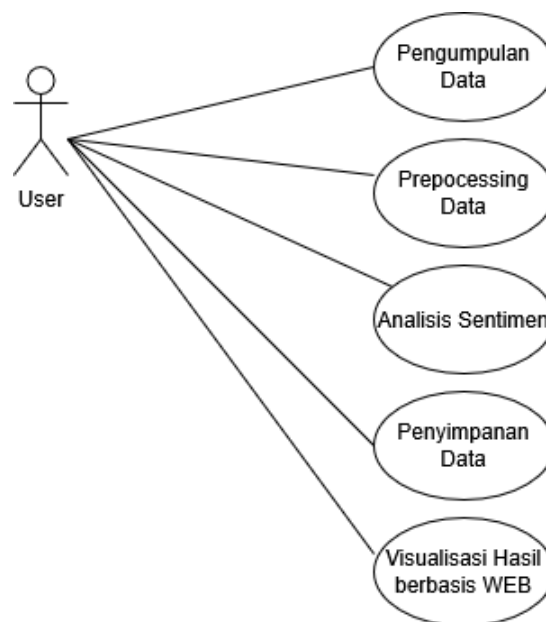
Analisis Sentimen berbasis *website* ini dirancang dengan visualisasi yang menampilkan hasil klasifikasi sentimen yang telah melewati proses *scraping*. Sehingga data ulasan yang ditampilkan dengan persentase mengikuti hasil analisis.

Perancangan Sistem

Perancangan sistem dilakukan dengan menggunakan UML (*Unified Modelling Language*) dengan tahapan perancangan *use case diagram*, perancangan *class diagram*, perancangan *sequence diagram* dan perancangan *activity diagram*.

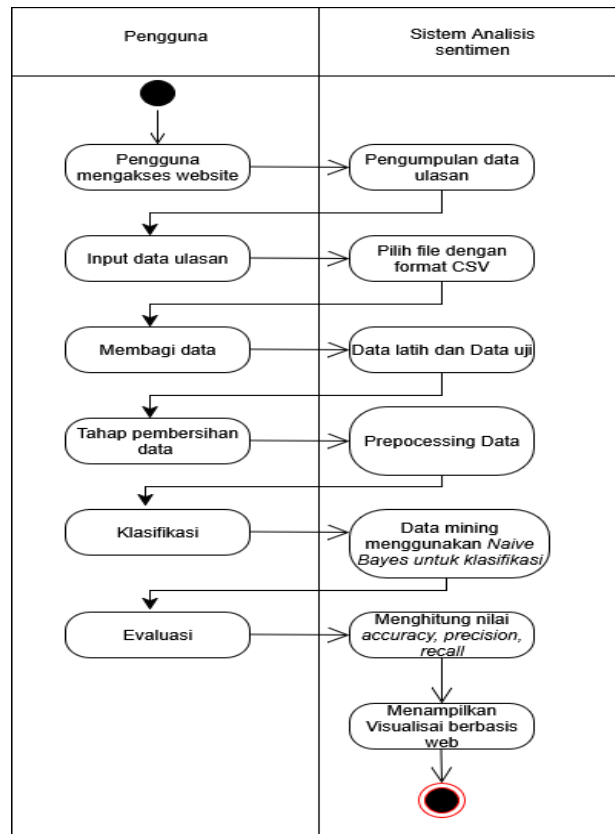
Adapun perancangan sistem yang digunakan pada penelitian ini yaitu:

- Perancangan *Use Case*



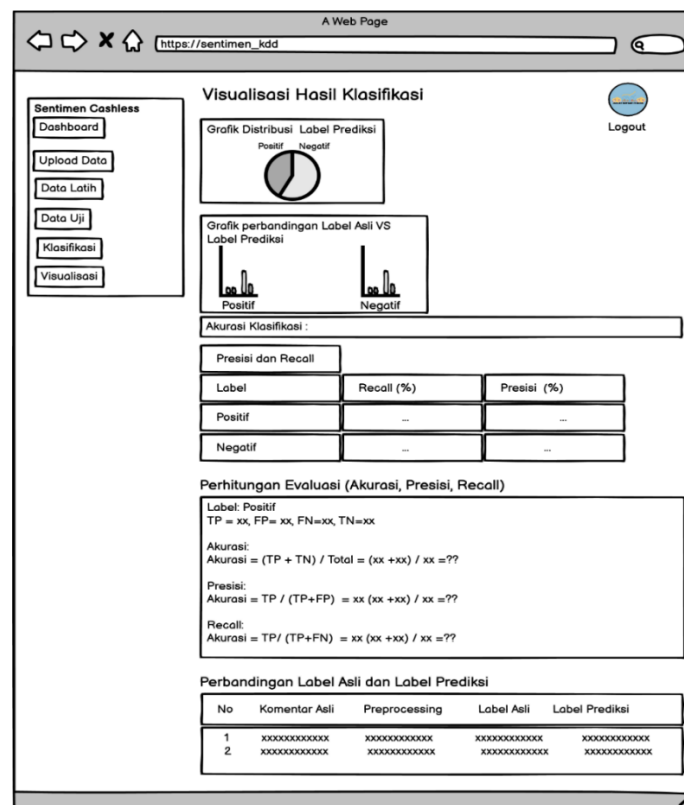
Gambar 1. Perancangan Sistem *Use Case*

- Perancangan *Activity Diagram*



Gambar 2. Perancangan Sistem Activity Diagram

- Tampilan Perancangan Website

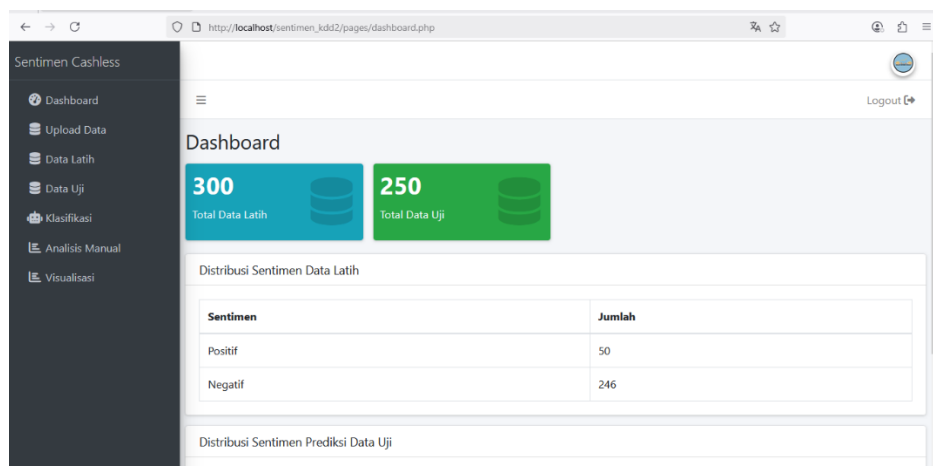


Gambar 3. Halaman Visualisasi

Pada halaman visualisasi klasifikasi menampilkan kolom yang terdiri dari:

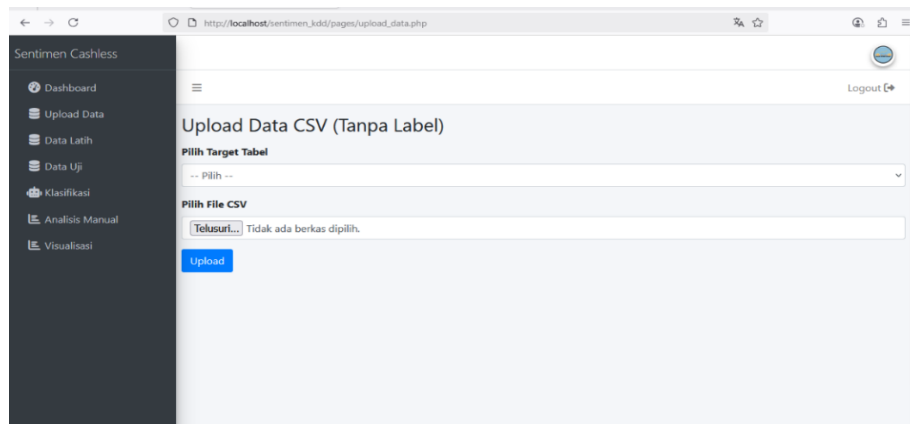
- Kolom “Grafik Distribusi Label Prediksi” untuk menampilkan grafik *pie chart* label prediksi pada data uji sentimen positif dan negatif.
- Kolom “*Bar chart*” untuk menampilkan perbandingan label asli dan prediksi sentimen positif dan negatif.
- Kolom “*Accuracy, precision* dan *recall*” untuk menampilkan hasil persentase dari klasifikasi data latih dan data uji.
- Kolom “Perhitungan Evaluasi *Accuracy, precision* dan *recall*” untuk menampilkan hasil evaluasi label positif dan negatif.
- Kolom “Perbandingan Label Asli dan Label Prediksi” untuk menampilkan hasil perbandingan komentar yang telah melewati proses *preprocessing* beserta sentimen label asli dan label prediksi.

4. HASIL DAN PEMBAHASAN



Gambar 4. Dashboard

Pada Halaman *Dashboard* terdapat total Data Latih 300 komentar dan Data uji 250 komentar. Distribusi sentimen data latih sentimen positif berjumlah 50 sedangkan sentimen negatif sebanyak 246. Lalu distribusi sentimen prediksi data uji sentimen positif 34 dan negatif 216. Setelah melewati proses metode *knowledge discovery databases* dengan tahapan *data selection, preprocessing, transformation, data mining* dan *evaluation*.



Gambar 5. Upload Data

Pada halaman *upload data*, mengunggah data yang telah melewati proses pengumpulan data melalui *scraping* pada *google colab* dengan format file CSV kemudian dipilih untuk ditempatkan di data latih atau data uji.

No	Komentar	Sentimen	Aksi
1	saya atas nama nagato*****389 member lama & masih aktif di shopee , bisa dicek riwayat transaksi saya pribadi sudah berjuta* , tidak pernah ada kendala dalam pembayaran & bayar tagihan selalu lebih awal sebelum jatuh tempo sampai selesai kontrak dan program SPINJAM adalah fitur tidak berguna , gilanya lebih dari 4x pengajuan saya ditolak , kuali riwayat transaksi saya bermasalah oke saya pasti terima ! fitur cuma variasi mah minding dihapus atuh !	negatif	Edit
2	Tolong ya, klo ada voucher itu sebelum kita claim (rupanya sebenarnya dibeli) itu dikasih tau lah, klo vouchernya sudah habis. Masa iya sudah habis tetap bisa kita click, ya tersedot lah akhirnya uang kita di shopeepay, merugikan... Tolong ya, shopee sistemnya di perjelas!!! Agar tidak merugikan pengguna seperti ini. Saya sudah 2x terjebak seperti ini!!!	negatif	Edit
3	sejauh ini belanja di shopee puas, pesan saya jaga kerahasiaan identitas pembeli agar jgn dimanfaatkan identitasnya oleh org tdk bertanggung jawab, fasilitasi pembatalan pesanan & pengembalian yg mudah. Tolong sediakan sistem pembayaran menggunakan LinkAja..agar pilihan pembayaran banyak pilihan&memudahkan pembayaran, saya harap Sophie amanah selalu yaa. Bintang akan saya tambah jika sistem pembayaran LinkAja dnt disediakan terimakasih	positif	Edit

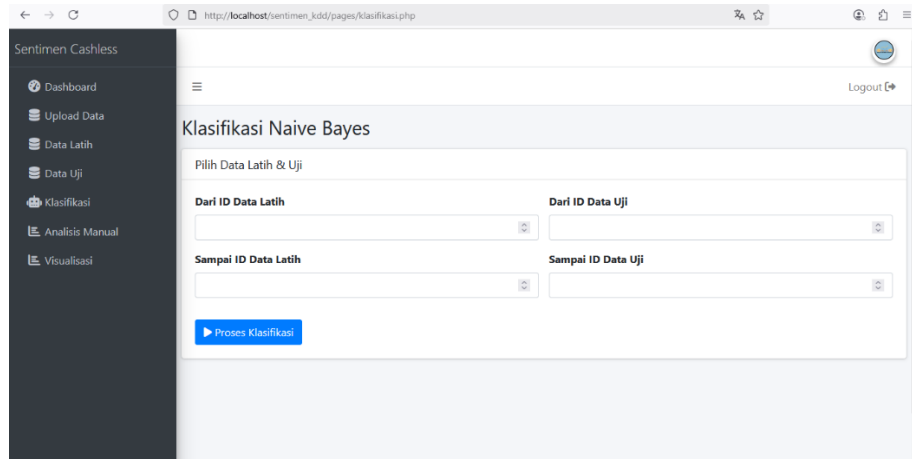
Gambar 6. Data Latih

Pada data latih memberi label sentimen termasuk kategori sentimen positif atau negatif.

No	Komentar	Label Asli	Label Prediksi	Aksi
1	Keren udah 2 kali problem masalah pembayaran. Sekali delay topup shopeepay baru masuk besoknya, sekali saldo kepotong tapi pembayaran gagal. Aplikasi jg lemot banget padahal internet cepet. Jadi paham kenapa kompetitor biaya pemeliharaan sistem nya gede tapi mereka sangat stabil dan aman.	negatif	positif	Edit
2	Ada perubahan yang saya suka di shopee, yaitu GRATIS ONGKIRNYA murah meriah, good job! Masukan saya, tolong tambahkan pembayaran e-wallet pas checkout (DANA/GOPAY)!! Soalnya kalo pake shopeepay, biaya layanannya mahal banget Rp.1900, kalo e-wallet cuma Rp.1000!! Kalo gak mau atau takut, tolonglah dimurahin seperti e-wallet lainnya!! Kalo ini terjadi, percayalah shopee "PALING TERBAIK" gak ada lawan!! rekomendasi e-commerce belanja online seindonesia :3	positif	negatif	Edit
3	☆☆☆☆☆ Aplikasi belanja terbaik dan terpercaya! Shopee sangat membantu dalam belanja online. Banyak pilihan produk, harga terjangkau, promo menarik, dan fitur gratis ongkir yang memudahkan. Tampilan aplikasinya user-friendly, mudah digunakan bahkan untuk pemula. Proses pembayaran juga aman dan pilihan ekspedisi lengkap. Fitur chat dengan penjual juga sangat membantu. Semoga semakin banyak orang yang menggunakan Shopee. Terima kasih Shopee!	positif	positif	Edit

Gambar 7. Data Uji

Pada data uji terdapat label asli dan label prediksi, label asli merupakan pelabelan yang di isi melalui ikon aksi dengan memilih komentar termasuk kedalam kategori sentimen positif atau negatif. Setelah itu label prediksi akan terisi secara otomatis dengan melakukan proses klasifikasi terlebih dahulu.



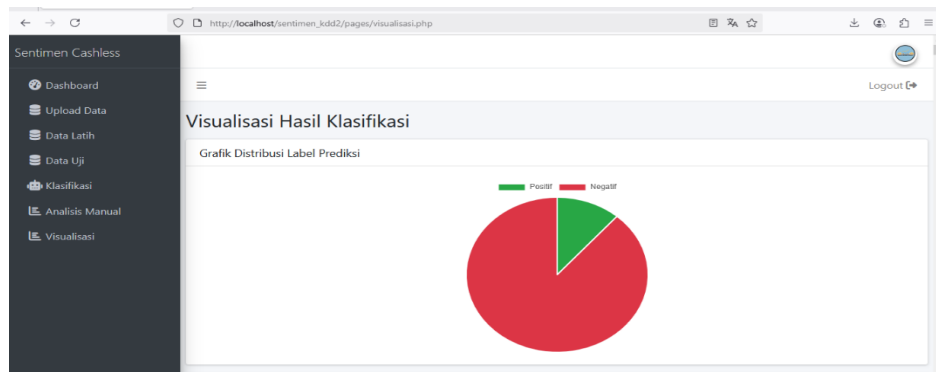
Gambar 8. Klasifikasi

Pada proses klasifikasi memilih ID data latih 1 sampai ID data latih ke 300 dan ID data uji 1 sampai ID data uji ke 250 untuk mendapatkan hasil label prediksi yang terdapat pada halaman data uji sebelumnya.



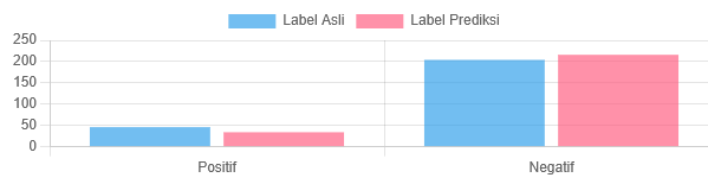
Gambar 9. Analisis Manual

Pada analisis manual menampilkan keseluruhan perhitungan dari ulasan komentar data uji, sehingga hasil perhitungannya sesuai dengan sentimen label prediksi.



Gambar 10. Visualisasi *Pie Chart*

Perbandingan Label Asli vs Label Prediksi



Gambar 11. Visualisasi *Bar Chart*

Akurasi: 83.2%

Presisi dan Recall

Label	Presisi (%)	Recall (%)
Positif	55.88	41.3
Negatif	87.5	92.65

Gambar 12. Visualisasi *Accuracy*, *Tabel precision dan recall*

Perhitungan Evaluasi (Akurasi, Presisi, Recall)

Label: Positif

TP = 19, FP = 15, FN = 27, TN = 189

Akurasi:

Akurasi = $(TP + TN) / \text{Total} = (19 + 189) / 250 = 83.2\%$

Presisi:

Presisi = $TP / (TP + FP) = 19 / (19 + 15) = 55.88\%$

Recall:

Recall = $TP / (TP + FN) = 19 / (19 + 27) = 41.3\%$

Label: Negatif

TP = 189, FP = 27, FN = 15, TN = 19

Akurasi:

Akurasi = $(TP + TN) / \text{Total} = (189 + 19) / 250 = 83.2\%$

Presisi:

Presisi = $TP / (TP + FP) = 189 / (189 + 27) = 87.5\%$

Recall:

Recall = $TP / (TP + FN) = 189 / (189 + 15) = 92.65\%$

Gambar 13. Visualisasi Perhitungan Evaluasi

Perbandingan Label Asli dan Label Prediksi

No	Komentar Asli	Preprocessing	Label Asli	Label Prediksi
1	Keren udah 2 kali problem masalah pembayaran. Sekali delay topup shopeepay baru masuk besoknya, sekali saldo kepotong tapi pembayaran gagal. Aplikasi jg lemot banget padahal internet cepet. Jadi paham kenapa kompetitor biaya pemeliharaan sistem nya gede tapi mereka sangat stabil dan aman.	keren udah kali problem masalah pembayaran sekali delay topup shopeepay baru masuk besoknya sekali saldo kepotong tapi pembayaran gagal aplikasi jg lemot banget padahal internet cepet jadi paham kenapa kompetitor biaya pemeliharaan sistem nya gede tapi mereka sangat stabil dan aman	negatif	positif
249	baru ini saya kasih centang bintang 3 dikarenakan aplikasi ini tidak memberikan solusi bagi yang mempunyai kesalahan double pembayaran, cara untuk pengendalian dana sulit sudah lapor cs malah chatbot. deskripsi di aplikasi Tokopedia tidak tersedia tapi pemotongan saldo dari ATM sudah ada tolong bapak/ibu aplikasi nya diperbaiki lagi.	baru ini saya kasih centang bintang dikarenakan aplikasi ini tidak memberikan solusi bagi yang mempunyai kesalahan double pembayaran cara untuk pengendalian dana sulit sudah lapor cs malah chatbot deskripsi di aplikasi tokopedia tidak tersedia tapi pemotongan saldo dari atm sudah ada tolong bapakibu aplikasi nya diperbaiki lagi	negatif	negatif
250	saya pengguna lama toped,semakin kesini semakin gk nyamam pakai toped,dr penjualnya yg sering dpt yg abal" tidak memproses pesanan,buka keranjang / prodak sering error rumah toped.sistem pembayaran tidak seperti dulu. semoga segera ada perbaikan dr toped.🙏	saya pengguna lama topedsemakin kesini semakin gk nyamam pakai topeddr penjualnya yg sering dpt yg abal tidak memproses pesananbuka keranjang prodak sering error rumah topedsistem pembayaran tidak seperti dulu semoga segera ada perbaikan dr toped	negatif	negatif

Gambar 14. Visualisasi Perbandingan Label asli dan Label Prediksi

Pada halaman visualisasi menampilkan hasil berupa grafik pie chart distribusi label prediksi dengan jumlah positif 34 dan negatif 216, lalu pada bar chart pebandingan label asli positif 46, label asli negatif 204 dan label prediksi 34 positif, 216 negatif. *Accuracy* 83,2%, *precision* label positif 55,88% dan label negatif 87,5. Serta *Recall* positif 41,3% dan recall negatif 92,65% lewat perhitungan yang telah dilakukan pada tabel perhitungan evaluasi. Halaman visualisasi juga menampilkan keseluruhan ulasan komentar data uji sebelum melewati proses *preprocessing* dan setelah melewati proses *preprocessing* dengan label asli dan label prediksi setiap komentar.

Diperoleh hasil sebagai berikut pada tabel dibawah ini:

Label asli sentimen Positif	Label asli sentimen Negatif
46	204
Label Prediksi sentimen Positif	Label Prediksi sentimen Negatif
34	216

Perbandingan ini menunjukkan bahwa meskipun terdapat sedikit perbedaan antara pelabelan data asli dan hasil prediksi, model sudah cukup mampu mengklasifikasikan data secara mendekati distribusi sebenarnya.

Label Positif	Label Negatif
<i>Accuracy</i> = 83,2%	<i>Accuracy</i> = 83,2%
<i>Precision</i> = 55,88%	<i>Precision</i> = 87,5%
<i>Recall</i> = 41,3%	<i>Recall</i> = 92,65%

Namun, jika dilihat dari metrik lainnya: *Precision* untuk label positif sebesar 55,88% dan untuk label negatif 87,5% menandakan bahwa model cukup tepat dalam memprediksi

kedua jenis sentimen, terutama dalam menghindari *false positives* pada label negatif. *Recall* untuk label positif hanya mencapai 41,3%, sedangkan *recall* label negatif mencapai 92,65%. Ini menunjukkan bahwa model lebih baik dalam mengenali komentar negatif dibandingkan komentar positif. Rendahnya *recall* pada label positif menunjukkan bahwa masih ada cukup banyak komentar positif yang gagal dikenali dengan benar oleh model. Halaman visualisasi juga menampilkan keseluruhan ulasan komentar data uji sebelum melewati proses *preprocessing* dan setelah melewati proses *preprocessing* dengan label asli dan label prediksi setiap komentar.

Rendahnya *recall* pada label positif menunjukkan bahwa masih ada cukup banyak komentar positif yang gagal dikenali dengan benar oleh model. Selain hasil prediksi, visualisasi juga menampilkan keseluruhan data komentar dari data uji baik sebelum maupun setelah melewati tahapan *preprocessing*. Hal ini memberikan gambaran yang jelas mengenai dampak dari tahapan *preprocessing* seperti *cleaning*, *case folding* dan *tokenizing* terhadap teks ulasan

5. KESIMPULAN DAN SARAN

Berdasarkan dari hasil penelitian terhadap analisis sentimen ulasan *cashless* menggunakan metode *knowledge discovery databases* dapat disimpulkan sebagai berikut:

- Lewat proses *scraping* pada *google colab* terhadap aplikasi *e-commerce* yang ada di *google playstore* menerapkan penggunaan *cashless* diperoleh data latih sebanyak 300 ulasan komentar dan data uji 250 ulasan komentar.
- Data yang telah di peroleh tersebut melewati tahapan seperti *preprocessing*, *transformation*, *data mining (naïve bayes)* dan *evaluation* sehingga menghasilkan tingkat *accuracy* sebesar 83,2 % pada label positif dan negatif sedangkan untuk *precision* dan *recall* memiliki hasil yang berbeda antar label positif dan negatif yaitu: label positif *precision* 55,88% , label negatif 87,5%. Label positif *recall* 41,3% , label negatif 92,65%.
- Berdasarkan dari perolehan skor diatas tersebut ulasan negatif lebih dominan dibanding ulasan positif terhadap penggunaan *cashless*. Sentimen positif berfokus pada pembayaran *cashless* yang mempermudah transaksi, sedangkan sentimen negatif berfokus pada masalah teknis pada saat pembayaran, pengembalian dana dan pembekuan akun *paylater*.

Sistem berbasis *website* mempermudah pengelolaan data ulasan. Fitur *website* dimulai dari halaman *login*, *dashboard*, halaman *upload* data, data latih, data uji, klasifikasi, analisis manual dan visualisasi berfungsi dan mudah dikelola

DAFTAR REFERENSI

- Aghitsni, W. I., & Busyra, N. (2022). JIMEA | Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi). *Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi)*, 6(3), 38–51. <https://doi.org/10.31955/mea.v6i3.2271>
- Alghifari, F., & Juardi, D. (2021). Penerapan data mining pada penjualan makanan dan minuman menggunakan metode algoritma naïve Bayes. *Jurnal Ilmiah Informatika*, 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>
- Arsi, P., & Waluyo, R. (2021). Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma support vector machine (SVM). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Fitriyani Yapan, U. (n.d.). Penerapan e-commerce sebagai media penjualan online (Studi kasus pada Toko Sinar Terang Bandar Lampung). *Z.A. Pagar Alam*, 7, 40115.
- Hasibuan, M. H., Rahmani, N. A. B., & Aslami, N. (2024). Analisis penggunaan fitur pembayaran online e-commerce dan top up e-wallet terhadap kenyamanan bertransaksi menggunakan BSI Mobile (Studi kasus mahasiswa Universitas Negeri Sumatera Utara). *Jesya*, 7(2), 2121–2133. <https://doi.org/10.36778/jesya.v7i2.1772>
- Indransyah, R., Chrisnanto, Y. H., Sabrina, P. N., & Kom, S. (2022). Klasifikasi sentimen pergelaran MotoGP di Indonesia menggunakan algoritma correlated naïve Bayes classifier. *INFOTECH Journal*, 8(2), 60–66. <https://doi.org/10.31949/infotech.v8i2.3103>
- Kusumo, S. (2022). Penerapan web scraping deskripsi produk menggunakan Selenium Python dan framework Laravel. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(4), 3426–3435. <https://doi.org/10.35957/jatisi.v9i4.2727>
- Lim, N. E., & Silalahi, M. (2023). Rancang bangun sistem e-administrasi berbasis CodeIgniter framework di KP2A Batam. *Computer and Science Industrial Engineering (COMASIE)*, 8(1), 37–46. <https://doi.org/10.33884/comasiejournal.v8i1.6639>
- Marsela, A. D., Nathanael, J., & Marchelyta, N. (2022). Penggunaan e-wallet sebagai kemajuan teknologi digital dalam menentukan preferensi masyarakat di Surabaya. *Prosiding Seminar Nasional Ilmu Ilmu Sosial*, 1, 784–790.
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). Penerapan data mining untuk pengelompokan data rekam medis pasien berdasarkan jenis penyakit dengan algoritma clustering (Studi kasus: Poli Klinik PT. Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153. <https://doi.org/10.33060/jik/2020/vol9.iss2.181>
- Rahayu, W. I., Prianto, C., & Novia, E. A. (2021). Perbandingan algoritma K-means dan naïve Bayes untuk memprediksi prioritas pembayaran tagihan rumah sakit berdasarkan tingkat kepentingan pada PT. Pertamina (Persero). *Jurnal Teknik Informatika*, 13(2), 1–8.
- Tazkia, S. R., Ardi, H. A., Ekonomi, F., & Riau, U. M. (2024). Pengaruh penggunaan cashless payment terhadap kemudahan transaksi konsumen Cafe Monocsky Pekanbaru untuk memanfaatkan kemajuan teknologi baru. *Kemudahan Transaksi Merupakan Situasi*, 4(1), 20–25.
- Wahidna, F. J., & Nerisafitra, P. (2023). Analisis sentimen pengguna sistem pay later menggunakan support vector machine metode pembobotan lexicon. *Journal of Informatics and Computer Science (JINACS)*, 4, 334–343. <https://doi.org/10.26740/jinacs.v4n03.p334-343>

- Widayat, W. (2021). Analisis sentimen movie review menggunakan Word2Vec dan metode LSTM deep learning. *Jurnal Media Informatika Budidarma*, 5(3), 1018. <https://doi.org/10.30865/mib.v5i3.3111>
- Wulandari, A., Kustina, L., & Nurastuti, P. (2023). Analisis faktor-faktor yang mempengaruhi cashless society. *Jurnal Investasi*, 9(2), 101–107. <https://doi.org/10.31943/investasi.v9i2.271>
- Yanuar, R. A. A. (2024). Jurnal Teknik Informatika, Vol. 16, No. 2, April 2024. *Jurnal Teknik Informatika*, 16(2), 1–7. <https://ejurnal.ulbi.ac.id/index.php/informatika/article/view/3533>