

PENERAPAN METODE NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN APLIKASI GOPAY

Riskawati, Fatihanursari, Iin, Ade Rizki Rinaldi

Teknik Informatika, STMIK IKMI Cirebon

Jalan Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon

riskawati335@gmail.com, fatihanursari@gmail.com, iin@gmail.com, aderizkirinaldi@gmail.com

ABSTRAK

Perkembangan teknologi telah membawa dampak besar dalam bidang pembayaran digital atau E-Wallet. Aplikasi pembayaran digital seperti GoPay telah muncul sebagai solusi praktis untuk melakukan berbagai transaksi. Meskipun GoPay memberikan kemudahan, respon atau komentar pengguna terhadap aplikasi ini tidak selalu positif. Beberapa orang mungkin memberikan komentar yang negatif atau netral terkait dengan aplikasi Go-Pay. Namun, menganalisis secara manual semua komentar pengguna tidak mudah, mengingat jumlahnya yang sangat besar sehingga dapat menghambat proses umpan balik. Penelitian ini bertujuan untuk menganalisis sentimen aplikasi GoPay dengan algoritma Naïve Bayes. Dalam penelitian ini metode yang akan digunakan yaitu Knowledge Discovery in Database (KDD). Dari hasil klasifikasi analisis sentimen menggunakan algoritma Naïve Bayes Classifier serta evaluasi data dengan confusion Matrix menghasilkan akurasi sebesar 85%, presisi sebesar 84%, recall sebesar 96% dan f1-score sebesar 90%. Ini menunjukkan bahwa algoritma Naïve Bayes Classifier mampu mengklasifikasikan sentimen pengguna terhadap aplikasi GoPay dengan tingkat akurasi yang memadai. Hasil analisis sentimen dapat memberikan masukan yang berharga kepada penyedia aplikasi GoPay untuk memperbaiki dan meningkatkan layanan mereka berdasarkan umpan balik pengguna. Selain itu, hasil penelitian akan menjadi dasar untuk penelitian lebih lanjut dalam analisis sentimen dalam konteks aplikasi seluler dan layanan keuangan digital.

Kata Kunci : Analisis Sentimen, Naïve Bayes Classifier, GoPay

1. PENDAHULUAN

Perkembangan teknologi telah membawa dampak besar dengan memudahkan proses transaksi elektronik. Inovasi-inovasi terbaru, seperti dompet digital atau e-wallet, telah muncul sebagai hasil dari perkembangan ini [1]. Kemudahan bertransaksi membuat masyarakat beralih kepada metode pembayaran digital. Kehadiran teknologi tidak dapat disangkal telah berdampak pada berbagai aspek kehidupan manusia, termasuk dalam konteks jual beli dan penggunaan uang tunai. Di Indonesia, telah muncul berbagai layanan keuangan berbasis digital, seperti e-wallet. Sejumlah layanan e-wallet di Indonesia tersedia dalam bentuk mobile wallet atau aplikasi [2]. Salah satu aplikasi e-wallet yang berkembang di masyarakat yaitu GoPay. Go-Pay merupakan bagian dari aplikasi Gojek dan salah satu fintech yang paling diminati [3]. Meskipun Go-Pay memberikan kemudahan, respon atau komentar pengguna terhadap aplikasi ini tidak selalu positif. Beberapa orang mungkin memberikan komentar yang negatif atau netral terkait dengan Go-Pay. Namun, menganalisis secara manual semua pendapat atau komentar pengguna menjadi tugas yang tidak mudah, mengingat jumlah pendapat atau komentar yang sangat besar sehingga dapat menghambat proses umpan balik. Maka dari itu penelitian ini juga memiliki tujuan yaitu menerapkan metode klasifikasi Naive Bayes untuk mengkategorikan ulasan pengguna ke dalam kelas sentimen yang sesuai (positif dan negatif). Dalam penelitian ini, metode yang digunakan adalah Knowledge Discovery in

Database (KDD). Tahapannya yaitu seleksi data, prapemrosesan, transformasi, data mining dan evaluasi [4]. Teknik klasifikasi yang digunakan pada penelitian analisis sentimen ini yaitu algoritma Naïve Bayes Classifier dan bahasa pemrograman yang digunakan yaitu Python serta Tools yang digunakan yaitu Google Collab. Keunggulan menggunakan Naive Bayes terletak pada fakta bahwa metode ini membutuhkan jumlah data pelatihan yang relatif kecil untuk menentukan parameter yang diperlukan dalam proses klasifikasi [5]. Hasil analisis sentimen dapat memberikan masukan yang berharga kepada penyedia aplikasi GoPay untuk memperbaiki dan meningkatkan layanan mereka berdasarkan umpan balik pengguna. Selain itu, hasil penelitian akan menjadi dasar untuk penelitian lebih lanjut dalam analisis sentimen dan metode analisis data lainnya dalam konteks aplikasi seluler dan layanan keuangan digital. Ini akan mengembangkan bidang ilmu informatika yang lebih luas

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Paper [6] membahas mengenai penelitian yang berjudul "Analisis Kepuasan Data Pengguna Layanan E-Wallet Gopay Menggunakan Metode Naïve Bayes Classifier Algorithm". Penelitian ini berdasarkan dataset kepuasan pengguna E-Wallet GoPay di Kecamatan Denpasar Timur, khususnya terhadap pelaku UMKM di bidang kuliner, ditemukan 100 data ulasan. Dari data tersebut, 91 dianggap sebagai "Puas", sedangkan 9 lainnya dianggap "Tidak Puas".

Hasil analisis menunjukkan bahwa hasil probabilitas kelas “Tidak Puas” mencapai 21,00% dan kelas “Puas” mencapai 79,00% dalam aplikasi Orange. Dengan demikian dapat disimpulkan bahwa pengguna E-Wallet GoPay di Kecamatan Denpasar Timur secara umum puas dengan layanan GoPay. Hasil evaluasi model Naïve Bayes Classifier dalam mengklasifikasikan kepuasan pengguna E-Wallet GoPay menunjukkan bahwa model ini memiliki performa yang baik dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan.

Paper [7] membahas mengenai penelitian yang berjudul “Analisis Sentimen Twitter Aplikasi Pedulilindungi Menggunakan Naïve Bayes Dan Support Vector Machine”. Penelitian ini bertujuan untuk membandingkan tingkat akurasi antara Naïve Bayes Classifier dan Support Vector Machine, yang diimplementasikan dengan bahasa pemrograman Python. Hasil penelitian menunjukkan bahwa metode Naïve Bayes menghasilkan tingkat akurasi sebesar 90%, dengan nilai presisi negatif sekitar 10%, nilai presisi positif sekitar 99%, nilai recall negatif sekitar 68%, dan recall positif sekitar 99%. Di sisi lain, metode Support Vector Machine menghasilkan tingkat akurasi 91% dengan nilai presisi negatif 16%, nilai presisi positif sekitar 99%, nilai recall negative 74%, dan nilai recall positif 91%. Dengan demikian dapat disimpulkan bahwa dalam penelitian ini, Support Vector Machine mengungguli Naïve Bayes Classifier dengan perbedaan akurasi 1%.

Paper [8] membahas mengenai penelitian yang berjudul “Analisis Sentimen Tweet Vaksin Covid-19 Menggunakan Recurrent Neural Network Dan Naïve Bayes”. Penelitian ini Menggunakan Recurrent Neural Network Dan Naïve Bayes”. Penelitian ini dilakukan untuk memeriksa performa beberapa metode RNN dan NB. Menggunakan teknik pembobotan TF-IDF. Dataset yang digunakan terdiri dari 5000 tweet mengenai vaksin COVID-19, dengan rincian 3800 tweet memiliki sentimen positif, 800 tweet dengan sentimen negatif dan 400 tweet dengan sentimen netral. Hasil penelitian menunjukkan bahwa metode RNN (TF-IDF) mencapai akurasi tertinggi, yakni sebesar 97,77%, dibandingkan dengan Naïve Bayes (TF-IDF) yang memiliki akurasi sekitar 80%.

Paper [9] membahas mengenai penelitian tentang “Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine”. Hasil penelitian menunjukkan bahwa metode klasifikasi SVM, dengan perbandingan antara berbagai jenis kernel, memberikan hasil yang cukup baik. Ketika menggunakan kernel linear, tingkat akurasi mencapai 89,17%, yang memungkinkan sistem untuk mengklasifikasikan dengan benar sebanyak 1109 ulasan. Sementara itu, dengan penggunaan kernel polynomial, tingkat akurasi mencapai 84,38%, dan sistem yang dapat mengklasifikasikan dengan benar sebanyak 1021 ulasan. Penelitian ini juga mengkonfirmasi bahwa data Twitter untuk ulasan terkait GoPay dapat

dipisahkan secara cukup baik secara linear. Ini sesuai dengan hipotesis dalam hal ini, komentar positif mencapai 72%, sementara komentar negatif hanya 17% dalam kernel linear SVM.

Paper [10] membahas mengenai penelitian yang berjudul “Sentimen Analisis Terhadap Kebijakan Pembelajaran Jarak Jauh Selama Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes”. Hasil analisis data dari total 991 komentar tweet, terdapat 32 komentar yang berhasil diidentifikasi sebagai positif. 52 komentar yang diidentifikasi negatif, dan 894 komentar yang diidentifikasi sebagai netral. Ini menghasilkan persentase rata-rata recall, precision, dan F1-Measure yang berkisar antara 93,7% hingga 97,4%. Proses pengujian menggunakan algoritma Naïve Bayes dengan bantuan alat RapidMiner menghasilkan tingkat akurasi sebesar 98,7%.

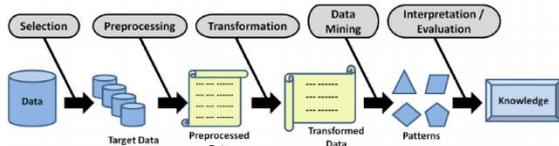
Paper [11] membahas mengenai penelitian yang berjudul “Analisis Sentimen Dompot Elektronik Pada Media Social Twitter Menggunakan Naïve Bayes Classifier”. Hasil pengujian mengungkapkan dalam hal sentimen positif, dompet elektronik GoPay menunjukkan nilai yang lebih tinggi, mencapai 46,6%, sementara Dana mencapai 37,5%, dan OVO mencapai 16,67%. Sebaliknya, dalam hal sentimen negatif, dompet elektronik OVO memiliki nilai yang lebih tinggi mencapai 63,33%, sementara GoPay mencapai 53,33%, dan Dana hanya sebesar 30,00%. Nilai sentimen netral pada dompet elektronik Dana lebih tinggi yaitu 32,50%, dibandingkan dengan OVO yang mencapai 20,00%, dan GoPay tidak memiliki nilai sentiment netral, yaitu sebesar 0,00%.

2.2. Naïve Bayes

Klasifikasi Naive Bayes Classifier adalah metode yang sering digunakan dalam data mining karena kemudahannya dalam penggunaan dan pemrosesan yang cepat. Metode ini diimplementasikan dengan struktur yang sederhana, tetapi tetap efektif dengan tingkat efektivitas yang tinggi. Klasifikasi Naive Bayes juga menunjukkan akurasi yang tinggi dan kecepatan yang baik saat digunakan untuk dataset berukuran besar [2].

3. METODE PENELITIAN

Metode penelitian yang digunakan penelitian ini adalah Knowledge Discovery from Data (KDD) yang terdiri dari lima tahapan, yaitu seleksi data, preprocessing, transformasi, data mining dan evaluasi. Metode KDD merupakan suatu metode penggalian data untuk memperoleh informasi dan pola dari data yang telah dipilih sebelumnya [12]. Untuk memberikan pemahaman yang lebih jelas tentang alur penelitian, sebuah ilustrasi grafis disertakan dalam gambar berikut :



Gambar 3. 1 Alur penelitian (Sumber: Francesco Gullo, 2015)

3.1. Pengumpulan Data

Data dalam penelitian ini diambil dari ulasan aplikasi pada Google Play Store dengan metode web scraping. Data yang dihasilkan berjumlah 1000 data ulasan rentang waktu dari 16 Agustus 2023 sampai 26 November 2023.

3.2. Pelabelan

Setelah melakukan pengumpulan data, selanjutnya akan dilabeli berdasarkan rating. Dataset yang dikumpulkan kemudian diberi label dengan dua kategori untuk setiap ulasan yaitu positif negatif [13]. Sentimen yang memiliki rating satu sampai rating tiga akan dilabeli sebagai sentimen positif dan sentimen yang memiliki rating empat sampai rating lima akan dilabeli sebagai sentimen positif.

3.3. Preprocessing

Tahap preprocessing adalah tahap di mana semua komentar dalam dataset diolah untuk menghilangkan kata-kata yang tidak memiliki dampak pada klasifikasi yang akan dilakukan. Proses ini melibatkan beberapa langkah, diantaranya cleansing, case folding, stopword removal, tokenizing, stemming, dan penghapusan kata yang kurang dari 3 huruf.

3.4. Transformation

Transformasi data merupakan suatu upaya yang dilakukan dengan tujuan utama mengubah skalapengukuran data asli ke dalam format lain agar data tersebut sesuai dengan asumsi-asumsi yang mendasari analisis berbagai teknik data mining [14]. Dalam tahap ini akan dilakukan pembobotan kata menggunakan metode TF-IDF.

3.5. Data Mining

Data mining adalah istilah yang digunakan untuk mengeksplorasi pengetahuan dalam big data. Data mining adalah istilah yang digunakan oleh untuk mengeksplorasi pengetahuan dalam big data. Data mining menggunakan metode statistic, matematika, kecerdasan buatan dan pembelajaran mesin untuk membantu mengidentifikasi informasi berguna dalam kumpulan big data. Pada penelitian ini diterapkan algoritma Naïve Bayes untuk mengklasifikasikan sentimen [15]. Sebelum melakukan data mining data akan dibagi terlebih dahulu. Pembagian data training dan data testing dengan pembagian rasio 70% data training dan 30% data testing. Pada tahap ini akan dilakukan analisis sentimen menggunakan algoritma Naïve Bayes Classifier dan dalam tahap ini algoritma

yang akan digunakan untuk menghasilkan nilai akurasi, presisi, recall dan f1-score Classifier.

3.6. Evaluasi Model

Evaluasi kinerja model klasifikasi yang dihasilkan, dilakukan dengan menggunakan hasil pengujian objek prediksi yang salah dan benar. Model klasifikasi menentukan akurasi. Akurasi menentukan presentase keakuratan rekaman yang dihasilkan dengan benar. Hasilperhitungan ditabulasikan dalam bentuk matrix konfusi dengan jumlah nilai true positif (TP), false negatif (FN), true negatif (TN), dan false positif (PF) [16]. Dalam penelitian ini confusion matrix digunakan sebagai alat evaluasi kinerja model klasifikasi yang memberikan gambaran tentang sejauh mana model mampu melakukan klasifikasi yang benar atau keliru.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Pengambilan data dalam penelitian ini menggunakan metode scraping pada ulasan Google Play Store dan dengan bantuan bahasa Pemrograman Python pada Google Colaboratory. Data yang dihasilkan berjumlah 1000 data ulasan rentang waktu dari 16 Agustus 2023 sampai 26 November 2023.

reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	replyIsLike	appVersion
422e15da-2037-40a6-427a-c5b888103	Dan Bahari	https://play.googleusercontent.com/ia-ALiK-U...	Sebelumnya sudah lama saya ganti gppp ini...	5	144	1.12.1	2023-11-24 12:09:39	None	NaN	1.12.1
a85d594e-2204-4419-5073-c969707026c	Pangji Suprman	https://play.googleusercontent.com/ia-ALiK-U...	GoPay, dompet digital yang tak pernah mengece...	5	45	1.13.0	2023-11-26 15:54:33	None	NaN	1.13.0
8116ac11c-2a2b-428b-8a0c-4a1543b96044	Yargheli William	https://play.googleusercontent.com/ia-ALiK-U...	Aplikasi GoPay ini sangat user-friendly meny...	5	79	1.12.1	2023-11-23 15:16:15	None	NaN	1.12.1
78c3104-d483-45e5-418a-6a843000443	Pati Rahmawati	https://play.googleusercontent.com/ia-ALiK-U...	Aplikasinya sangat membantu, bisa langsung ber...	5	189	1.12.1	2023-11-24 07:30:29	None	NaN	1.12.1
57f1a05-c548-4a13-4414-114e31300292	Puji Rini	https://play.googleusercontent.com/ia-ALiK-U...	Aplikasi nya bagus banget, lumayan menarik. F...	5	12	1.12.1	2023-11-24 08:25:41	None	NaN	1.12.1

Gambar 4. 1 Hasil scraping data

Setelah melakukan scraping, data awalnya tidak berurutan kemudian diurutkan berdasarkan kronologi menggunakan excel dan hanya di ambil kolom ulasan, score (rating).

	content	score	at
0	Ulnya bagus, tapi fungsionalitas ngga dicek QA...	2	8/16/2023 16:42
1	Sampai sejauh ini g ada kendala. Seneng aplikasi...	5	8/27/2023 12:25
2	Keren banget aplikasi ini, semua fungsi e bank...	5	8/31/2023 23:08
3	Aplikasinya bagus banget dan ngebantu banget, ...	5	9/5/2023 10:36
4	Fitur untuk top upnya kurang lengkap. Kalau bi...	2	9/12/2023 05:06

Gambar 4. 2 Data berurutan

4.2. Pelabelan

Setelah melakukan scraping data, berikutnya adalah memberi label ke setiap sentimen berdasarkan rating yang di dapatkan dari hasil scraping data. Sebelumnya akan dilakukan penghapusan kolom at (tanggal) sehingga yang tersisa hanya kolom ulasan dan score (rating). Ulasan dengan rating satu sampai tiga dilabeli dengan sentimen positif dan ulasan rating empat sampai lima dilabeli dengan sentimen positif.

	ulasan	score	Label
0	Ulnya bagus, tapi fungsionalitas ngga dicek QA...	2	Negatif
1	Sampai sejauh ini g ada kendala. Seneng aplika...	5	Positif
2	Keren banget aplikasi ini, semua fungsi e bank...	5	Positif
3	Aplikasinya bagus banget dan ngebanu banget, ...	5	Positif
4	Fitur untuk top upnya kurang lengkap. Kalau bi...	2	Negatif

Gambar 4. 3 Hasil pelabelan data

4.3. Preprocessing

Tahap preprocessing adalah tahap di mana semua komentar dalam dataset diolah untuk menghilangkan kata-kata yang tidak memiliki dampak pada klasifikasi yang akan dilakukan. Proses ini melibatkan beberapa langkah, diantaranya cleansing, case folding, stopword removal, tokenizing, stemming, dan penghapusan kata yang kurang dari 3 huruf.

4.3.1. Cleansing

Proses cleansing atau pembersihan data merupakan proses pembersihan data yang bertujuan untuk menghilangkan data yang tidak bernilai, data yang salah, data yang tidak relevan, data duplikat, dan data tidak konsisten karena keberadaannya bisa mengurangi kualitas atau keakuratan hasil. Pembersihan data juga mempengaruhi kinerja sistem data mining, karena mengurangi jumlah dan kompleksitas data yang diproses [17].

4.3.2. Case Folding

Case folding merupakan proses mengubah huruf kapital pada dataset agar menjadi huruf kecil.

Tabel 4. 1 Hasil cleansing dan case folding

Sebelum	Sesudah
Aplikasinya bagus banget dan ngebanu banget, kalau mau beli pulsa atau paket data jadi lebih gampang, bisa bayar listrik juga, mau transfer uang juga mudah dan ga ada biaya transfernya juga, mantap pokoknya. fitur didalamnya juga simpel tapi tetep modern,.. Pokoknya rekomended banget deh gopay iniâ€i, •	aplikasinya bagus banget dan ngebanu banget kalau mau beli pulsa atau paket data jadi lebih gampang bisa bayar listrik juga mau transfer uang juga mudah dan ga ada biaya transfernya juga mantap pokoknya fitur didalamnya juga simpel tapi tetep modern pokoknya rekomended banget deh gopay ini

4.3.3. Stopword removal

Stopword removal bertujuan untuk menghapus seluruh kata yang masih umum pada data set yang dianggap tidak relevan dan tidak akan mempengaruhi sentimen dalam kalimat akan dihapus. Kata-kata yang dianggap tidak relevan disini mencakup kata-kata sambung seperti di, ke, ini, dan dari. Penghapusan kata sambung ini bertujuan untuk menyusutkan kalimat ulasan sehingga lebih ringkas.

Tabel 4. 2 Hasil stopword removal

Sebelum	Sesudah
aplikasinya bagus banget dan ngebanu banget kalau mau beli pulsa atau paket data jadi lebih gampang bisa bayar listrik juga mau transfer uang juga mudah dan ga ada biaya transfernya juga mantap pokoknyafitur didalamnya juga simpel tapi tetep modern pokoknya rekomended banget deh gopay ini	aplikasinya bagus banget ngebanu banget beli pulsa paket data gampang bayar listrik transfer uang mudah ga biaya transfernya mantap pokoknyafitur didalamnya simpel tetep modern pokoknya rekomended banget deh gopay

4.3.4. Tokenisasi

Tokenisasi adalah proses di mana kalimat atau teks dalam data akan dipotong-potong menjadi kata-kata atau token.

Tabel 4. 3 Hasil tokenisasi

Sebelum	Sesudah
aplikasinya bagus banget ngebanu banget beli pulsa paket data gampang bayar listrik transfer uang mudah ga biaya transfernya mantap pokoknyafitur didalamnya simpel tetep modern pokoknya rekomended banget deh gopay	['aplikasinya', 'bagus', 'banget', 'ngebanu', 'banget', 'beli', 'pulsa', 'paket', 'data', 'gampang', 'bayar', 'listrik', 'transfer', 'uang', 'mudah', 'ga', 'biaya', 'transfernya', 'mantap', 'pokoknyafitur', 'didalamnya', 'simpel', 'tetep', 'modern', 'pokoknya', 'rekomended', 'banget', 'deh', 'gopay']

4.3.5. Stemming

Stemming merupakan proses mengubah kata-kata ke bentuk dasarnya, sehingga kata-kata yang memiliki imbuhan yang ada di awalan, sisipan, akhiran dan kombinasi imbuhan dari sebuah kata akan di reduksi menjadi kata dasar.

Tabel 4. 4 Hasil stemming

Sebelum	Sesudah
['aplikasinya', 'bagus', 'banget', 'ngebanu', 'banget', 'beli', 'pulsa', 'paket', 'data', 'gampang', 'bayar', 'listrik', 'transfer', 'uang', 'mudah', 'ga', 'biaya', 'transfernya', 'mantap', 'pokoknyafitur', 'didalamnya', 'simpel', 'tetep', 'modern', 'pokoknya', 'rekomended', 'banget', 'deh', 'gopay']	['aplikasi', 'bagus', 'banget', 'ngebanu', 'banget', 'beli', 'pulsa', 'paket', 'data', 'gampang', 'bayar', 'listrik', 'transfer', 'uang', 'mudah', 'ga', 'biaya', 'transfer', 'mantap', 'pokoknyafitur', 'dalam', 'simpel', 'tetep', 'modern', 'pokok', 'rekomended', 'banget', 'deh', 'gopay']

4.3.6. Hanya yang Lebih dari Tiga Huruf yang Dipakai

Selanjutnya, kata-kata yang memiliki panjang kurang dari tiga huruf akan dihapus untuk membuat kalimat ulasan lebih ringkas dan meysisakan kata-kata yang relevan dalam analisis sentimen.

Tabel 4. 5 Hasil penghapusan kata kurang tiga huruf

Sebelum	Sesudah
['aplikasi', 'bagus', 'banget', 'ngebantu', 'banget', 'beli', 'pulsa', 'paket', 'data', 'gampang', 'bayar', 'listrik', 'transfer', 'uang', 'mudah', 'ga', 'biaya', 'transfer', 'mantap', 'pokoknyafitur', 'dalam', 'simpl', 'tetep', 'modern', 'pokok', 'rekomended', 'banget', 'deh', 'gopay']	aplikasi bagus banget ngebantu banget beli pulsa paket data gampang bayar listrik transfer uang mudah biaya transfer mantap pokoknyafitur dalam simpl tetep modern pokok rekomended banget gopay

4.3.7. Visualisasi Word Cloud

Dalam ulasan positif dan negatif, tentunya terdapat kata-kata yang paling sering muncul. Dalam analisis 1000 data ulasan, dilakukan visualisasi word cloud untuk menemukan tren kata pada setiap label.



Gambar 4. 4 Visualisasi word cloud positif



Gambar 4. 5 Visualisai word cloud negatif

Gambar 4.4 merupakan hasil dari visualisasi word cloud pada kelas sentimen positif. Kata yang paling banyak muncul diantaranya, bayar, transaksi, aplikasi, gopay, bantu, mudah, banget, dan seterusnya. Dan Gambar 4.5 merupakan hasil dari visualisasi word cloud dari kelas sentimen negatif. Kata yang paling sering muncul diantaranya bayar, saldo, masuk, akun, limit, pakai, dan seterusnya.

4.4. Tranformation

Proses tranformasi melibatkan penggunaan algoritma pembobotan kata Term Frequency-Inverse Document Frequency (TF-IDF). Setiap kata yang ada dalam ulasan akan mendapatkan nilai bobot yang sesuai dengan perhitungan yang dilakukan dengan menggunakan algoritma TF-IDF ini. Tujuannya untuk mengukur seberapa penting suatu kata dalam sebuah dokumen. Setelah melakukan proses TF-IDF, selanjutnya, data set akan dibagi menjadi dua terlebih dahulu yaitu data training dan data testing. Data training atau data latih digunakan untuk melatih model sedangkan data testing atau data pengujian

akan digunakan untuk menguji model yang telah dilatih untuk dapat mengetahui seberapa baik modelyang telah dibuat.

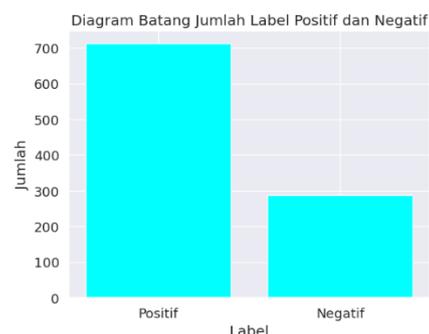
4.5. Data Mining

Pada tahap ini dilakukan klasifikasi sentimen pada data ulasan aplikasi GoPay. Data ini melalui prapemrosesan dan transformasi menggunakan algoritma Naïve Bayes untuk membangun model dari hasil, termasuk label positif dan negatif dari data ulasan. Berdasarkan hasil klasifikasi Naive Bayes, dilakukan perhitungan presisi, presisi, dan recall dengan menggunakan pemrograman Python menggunakan Google Collaboratory tools. Sebelum klasifikasi, data yang diperoleh dari preprocessing yang diberi label sesuai review pengguna dilakukan dengan membagi 1000 data validasi menjadi 70% pelatihan dan 30% pengujian. Setelah model diperoleh dari data pelatihan, selanjutnya algoritma Naïve Bayes digunakan untuk mengklasifikasikan data uji untuk mengetahui hasil probabilitas sentimen.

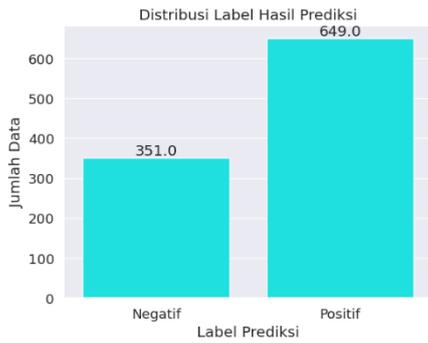
Accuracy: 0.85	precision	recall	f1-score	support
0	0.85	0.55	0.67	85
1	0.84	0.96	0.90	215
accuracy			0.85	300
macro avg	0.85	0.76	0.79	300
weighted avg	0.85	0.85	0.84	300

Gambar 4. 6 Hasil klasifikasi

Dari Gambar diatas dapat diketahui bahwa hasil klasifikasi menggunakan algoritma Naïve Bayes adalah 85 sentimen negatif dan 215 sentimen positif. Sentimen menunjukkan akurasi 85%, presisi 84%, recall 96% dan nilai f1-score 90%. Hasil kelasifikasi sentimen berdasarkan rating menghasilkan 712 sentimen yang berlabel positif dan 288 sentimen yang berlabel negatif dari jumlah 1000 data.



Gambar 4. 7 Grafik sentimen berdasarkan rating



Gambar 4. 8 Grafik hasil prediksi naïve bayes

Sedangkan setelah di analisis menggunakan Naïve Bayes menghasilkan 649 sentimen yang berlabel positif dan 351 sentimen yang berlabel negatif dari 1000 data. Berikut ini merupakan perbandingan dari data awal dan data prediksi Naïve Bayes dari beberapa data.

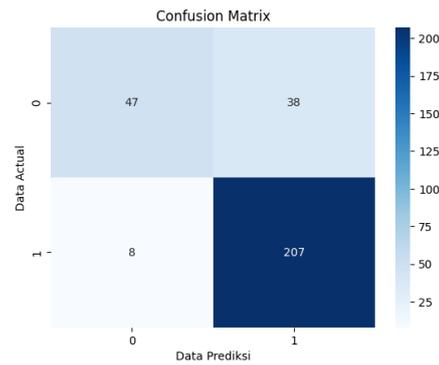
Tabel 4. 6 Perbandingan data awal dan data prediksi

No.	Ulasan	Sentimen Berdasarkan Rating	Prediksi Sentimen menggunakan NBC
1.	kasih bintang kadang transfer kadang kadang mohon baik mudah transfer	Negatif	Positif
2.	gabisa pake driver alfa kalo kena admin	Negatif	Positif
3.	upgrade beli pulsadan transfer yaaa upgrade daftar mati	Positif	Negatif

Dari Tabel 4.6 dapat dilihat bahwa sentimen berdasarkan rating yang berlabel negatif, setelah di analisa menggunakan Naïve Bayes menjadi Positif begitupun ada yang sebaliknya.

4.6. Evaluasi

Dalam penelitian ini confusion matrix digunakan sebagai alat evaluasi kinerja model. *Confusion matrix* adalah pengujian yang dapat digunakan untuk menghitung tingkat kinerja atau keakuratan suatu proses klasifikasi. *Confusion Matrix* dapat digunakan untuk menganalisis seberapa baik pengklasifikasi dapat mengenali kumpulan dapat mengenali record data dari kelas berbeda [18]



Gambar 4. 9 Confusion matrix

Dari Gambar 4.9 dapat ditarik beberapa kesimpulan sebagai berikut :

- True Negatif (TN) = data berlabel negatif yang diprediksi benar sebagai label negatif sebanyak 47
- True Positif (TP) = Data berlabel positif yang diprediksi benar sebagai label positif sebanyak 207
- False Negatif (FN) = Data berlabel positif yang diprediksi salah sebagai label negatif sebanyak 8
- False Positif (FP) = Data berlabel negatif yang diprediksi salah sebagai label positif sebanyak 38

Berdasarkan hasil klasifikasi pada dataset uji yang berjumlah 300, ditemukan bahwa 254 ulasan diprediksi benar, sedangkan 46 ulasan diprediksi salah. Performa ini sudah bisa dikatakan sangat baik.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh kesimpulan bahwa algoritma Naïve Bayes telah berhasil diimplementasikan pada analisis sentimen ulasan aplikasi GoPay di situs *Google Play Store* yang memberikan hasil klasifikasi berupa sentimen positif dan negatif.

1. Dari data penelitian yang diambil melalui proses web scraping terhadap 1000 ulasan aplikasi GoPay dari 16 Agustus sampai 26 November 2023, dapat ditarik kesimpulan bahwa analisis sentimen pengguna dapat dikategorikan menjadi positif dan negatif. Hasil menunjukkan bahwa dari total ulasan, 712 berlabel positif dan 288 berlabel negatif, berdasarkan rating yang diberikan oleh pengguna. Setelah dilakukan analisis menggunakan metode Naïve Bayes, terdapat 649 sentimen positif dan 351 sentimen negatif. Hal ini mengindikasikan bahwa mayoritas pengguna memberikan ulasan positif, meskipun sebagian juga menyampaikan ulasan negatif.
2. Dari segi performa, hasil perhitungan menunjukkan tingkat akurasi sebesar 85%. Selain itu, terdapat nilai presisi sebesar 84%, recall sebesar 96% dan f1-score sebesar 90%.

Oleh karena itu, dapat disimpulkan bahwa model analisis sentimen ini mampu memberikan prediksi dengan akurasi yang tinggi, meskipun terdapat variasi dalam nilai presisi, recall, dan f1-score.

Peneliti menyadari bahwa penelitian mengenai analisis sentimen ini masih memiliki kelemahan dan pembatasan tertentu. Oleh karena itu, peneliti mempertimbangkan beberapa saran untuk pengembangan hasil di masa depan dalam penelitian ini.

1. Menggunakan kumpulan data dengan jumlah yang lebih besar guna meningkatkan variasi pola yang diperoleh oleh model.
2. Pengujian lebih lanjut dapat dilakukan menggunakan berbagai algoritma klasifikasi lainnya, seperti SVM, Decision Tree, K-NN dan sebagainya.
3. Mengkaji analisis sentimen pada tingkat aspek memberikan wawasan lebih mendalam mengenai objek penelitian dibandingkan dengan melakukan analisis sentimen pada tingkat lainnya.

DAFTAR PUSTAKA

- [1] N. W. A. Pohan, B. Indra, and R. R. Suryono, "Borrower sentiment on P2P lending in Indonesia based on Google Playstore reviews," ... *Int. Conf. ...*, 2020, [Online]. Available: <https://www.atlantispress.com/proceedings/siconian-19/125939928>
- [2] S. A. Aputra, D. Rosiyadi, W. Gata, and H. Syepri, "Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019, [Online]. Available: <http://jurnal.iaii.or.id/index.php/RESTI/article/view/1118>
- [3] K. Septiani, "Perbandingan Analisis Sentimen Terhadap Pembayaran Digital 'Go-Pay' Dan 'Ovo' Di Media Sosial Twitter Menggunakan Metode Naive Bayes Dan Word Cloud," *J. Ilmu Data*, vol. 2, no. 10, pp. 1–10, 2022, [Online]. Available: <http://ilmudata.org/index.php/ilmudata/article/view/233>
- [4] S. A. R. Rizaldi, S. Alam, and I. Kurniawan, "Analisis Sentimen Pengguna Aplikasi JMO (Jamsostek Mobile) Pada Google Play Store Menggunakan Metode Naive Bayes," *STORAGE –Jurnal Ilm. Tek. dan Ilmu Kompute*, vol. 2, no. 3, pp. 109–117, 2023, [Online]. Available: <https://journal.literasisains.id/index.php/storage/article/view/2334>
- [5] A. Nurian, "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI GOOGLE PLAY MENGGUNAKAN NAÏVE BAYES," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 1, pp. 829–835, 2023, [Online]. Available: <http://journal.eng.unila.ac.id/index.php/jitet/article/view/3348>
- [6] I. G. I. Sudipa, I. Asana, K. J. Atmaja, P. P. Santika, and D. Setiawan, "Analisis Data Kepuasan Pengguna Layanan E-Wallet Gopay Menggunakan Metode Naïve Bayes Classifier Algorithm," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 3, pp. 726–735, 2023, [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/219>
- [7] I. Yunanto and S. Yulianto, "Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes and Support Vector Machine," *J. Tek. Inform.*, vol. 3, no. 4, pp. 807–814, 2022, doi: 10.20884/1.jutif.2022.3.4.292.
- [8] Merinda Lestandy, Abdurrahim Abdurrahim, and Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.
- [9] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *urnal Tek. Univ. Muhammadiyah Ponorogo*, vol. 3, no. 2, pp. 52–63, 2019, [Online]. Available: <http://studentjournal.umpo.ac.id/index.php/komputek/article/view/270>
- [10] D. Nababan, "Sentimen Analisis Terhadap Kebijakan Pembelajaran Jarak Jauh Selama Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes," *J. Tek. Inform.*, vol. 14, no. 1, pp. 51–56, 2021, doi: 10.15408/jti.v14i1.17002.
- [11] A. P. M. Wino, S. Susanti, E. Erlin, and H. Herwin, "Analisis Sentimen Dompot Elektronik Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier," *IT Journal Research and Development (ITJRD)*, 2020.
- [12] M. K. Khoirul Insan, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373.
- [13] Irvandi, B. Irawan, and O. Nurdiawan, "Naive Bayes Dan Wordcloud Untuk Analisis Sentimen Wisata Halal Pulau Lombok," *INFOTECH J.*, vol. 9, no. 1, pp. 236–242, 2023, doi: 10.31949/infotech.v9i1.5322.
- [14] O. Nurdiawan, A. Irma Purnamasari, and I.

- Ali, "Analisa Penjualan Mobil Dengan Menggunakan Algoritma K-Means Di PT. Mulya Putra Kencana," *J. Data Sci. dan Inform.*, vol. 1, no. 2, pp. 32–35, 2021.
- [15] M. Walid and F. Halimiyah, "Klasifikasi Kemandirian Siswa SMA/MA Double Track Menggunakan Metode Naive Bayes," *J. ICT Inf. Commun. Technol.*, vol. 22, pp. 190–197, 2022.
- [16] Syafi'i, O. Nurdiawan, and G. Dwilestari, "Penerapan Machine Learning Untuk Menentukan Kelayakan Kredit Menggunakan Metode Support Vektor Machine," *J. Sist. Inf. dan Manaj.*, vol. 10, no. 2, pp. 1–6, 2022.
- [17] Agung Nugraha, Odi Nurdiawan, and Gifthera Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 1–7, 2022.
- [18] S. Proboningrum and Acihmah Sidauruk, "Sistem Pendukung Keputusan Pemilihan Supplier Kain Dengan Metode Moora," *JSiI (Jurnal Sist. Informasi)*, vol. 8, no. 1, pp. 43–48, 2021, doi: 10.30656/jsii.v8i1.3073.