



KLASIFIKASI DAN PREDIKSI ULASAN APLIKASI DANA PADA *GOOGLE PLAY STORE* MENGGUNAKAN ALGORITMA *NAÏVE BAYES*

Cucu Hayati^{1*}, Tukino², Shofa Shofiah Hilabi³, April Lia Hananto⁴

^{1,2,3,4}Sistem Informasi, Universitas Buana Perjuangan Karawang

email: si21.cucuhayati@mhs.ubpkarawang.ac.id^{1*}

Abstrak: Pada era digital saat ini, kemajuan teknologi yang pesat telah mendorong masyarakat beralih ke transaksi digital melalui *financial technology (Fintech)*, salah satu inovasi *fintech* yaitu aplikasi DANA. Aplikasi DANA sangat populer dengan lebih dari 100 juta unduhan dan 7 juta ulasan di *Google Play Store*. Meningkatnya jumlah pengguna menimbulkan kekhawatiran tentang kualitas layanan, menjadikan ulasan sebagai sumber informasi penting. Saat ini, ulasan hanya dikelompokkan berdasarkan rating tanpa klasifikasi rinci, yang menyulitkan pengembang dalam mengidentifikasi masalah. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasi dan prediksi ulasan aplikasi DANA menggunakan Algoritma *Naïve Bayes*. Metode yang digunakan meliputi pengumpulan 1.500 data ulasan dari *Google Play Store*, *manual labelling* berdasarkan lima kategori (transaksi, keamanan, performa aplikasi, pelayanan, serta aktivasi dan verifikasi), *preprocessing*, pembobotan kata, model *Naïve Bayes*, dan evaluasi. Berdasarkan hasil analisis, tingkat akurasi yang diperoleh adalah sebesar 87%, dengan presisi mencapai 87%, *recall* sebesar 84%, dan *f1-score* sebesar 85%. Maka, dapat disimpulkan bahwa model pada penelitian ini mampu dalam mengklasifikasikan seluruh kategori dan dapat memberikan prediksi yang akurat, meskipun terdapat perbedaan nilai presisi, *recall*, dan *f1-score*. Diharapkan penelitian ini dapat bermanfaat bagi pengembang aplikasi DANA, dan juga dapat memberikan informasi mengenai efisiensi dan efektivitas algoritma *Naïve Bayes* dalam klasifikasi dan prediksi.

Kata Kunci : Dana, Klasifikasi, *Naïve Bayes*, Prediksi, Ulasan

PENDAHULUAN

Pada era digital saat ini, kemajuan teknologi yang pesat telah mencakup berbagai aspek kehidupan, salah satunya pada bidang sistem pembayaran atau transaksi [1]. Sehingga, aktivitas transaksi dalam penjualan pun semakin didominasi oleh teknologi digital yang dapat memberikan kemudahan dan efisiensi [2]. Dengan demikian, hal tersebut mendorong masyarakat beralih ke transaksi digital melalui *financial technology (Fintech)* [3].

Kategori layanan keuangan yang dikenal sebagai *Fintech* telah berkembang pesat, layanan pembayaran yang digunakan dalam bentuk e-wallet atau dompet digital merupakan salah satu yang paling populer di industri *fintech* [4]. DANA, Gopay, Flip, dan OVO merupakan beberapa contoh layanan dompet digital berbasis *fintech* yang banyak digunakan oleh masyarakat Indonesia.

Dari beberapa contoh aplikasi dompet digital tersebut, DANA merupakan aplikasi keuangan yang paling banyak diunduh di Layanan *Google Play Store* dengan fungsi sebagai pembayaran tagihan, transfer dana, investasi, dan layanan keuangan lainnya [5]. Dalam kehidupan sehari-hari, banyak pengguna yang kini mengandalkan aplikasi DANA untuk berbagai transaksi tanpa uang tunai. Mulai dari membayar belanjaan harian, hingga menambahkan saldo ke akun investasi, semuanya bisa dilakukan secara praktis melalui aplikasi tersebut.

DANA Dompet Digital Indonesia

PT Espay Debit Indonesia Koe

Dompet digital (e-Wallet) untuk berbagai transaksi non-tunai dan non-kartu



Gambar 1. Aplikasi DANA Dompet Digital Indonesia

DANA Dompet Digital Indonesia merupakan aplikasi dompet digital yang diluncurkan pada tahun 2018 dan dikembangkan oleh PT Espay Debit Indonesia Koe (EDIK), dengan fokus pada teknologi finansial dan menyediakan layanan pembayaran digital tanpa uang tunai atau kartu [6]. Hingga kini, Aplikasi DANA Dompet Digital Indonesia telah sangat populer dengan lebih dari 100 juta unduhan. Rating mencapai 4,7 dari 5,0, aplikasi ini juga telah mendapatkan ulasan dari sebanyak 7,88 juta pengguna di *Google Play Store*.

Google Play Store merupakan platform utama untuk mengunduh aplikasi dan produk digital di perangkat Android. *Google Play Store* memiliki fitur yang memungkinkan pengguna untuk memberikan rating atau ulasan tentang aplikasi yang telah mereka gunakan [7]. Ulasan aplikasi ini, salah satu sarana bagi pengguna untuk memberikan pendapat, kritik, saran, dalam bentuk rating atau juga teks berdasarkan pengalaman mereka selama menggunakan aplikasi, yang nantinya ulasan tersebut menjadi sumber informasi yang sangat berharga, karena dapat mencerminkan kualitas layanan dan kepuasan pengguna terhadap aplikasi. Dengan memanfaatkan ulasan secara optimal, pengembang dapat mengevaluasi dan meningkatkan layanan yang diberikan [8].



Namun, saat ini ulasan pengguna di Google Play Store hanya dikelompokkan berdasarkan rating bintang, tanpa adanya klasifikasi yang lebih rinci berdasarkan isi atau topik ulasan. Hal ini menimbulkan kesenjangan informasi, karena menyulitkan pengembang dalam memahami secara spesifik permasalahan atau keunggulan yang dirasakan pengguna. Kekhawatiran terhadap kualitas layanan seiring bertambahnya jumlah pengguna menjadikan analisis ulasan sebagai langkah penting untuk peningkatan mutu aplikasi.

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk mengklasifikasi dan memprediksi ulasan pengguna terhadap aplikasi DANA menggunakan algoritma *Naïve Bayes*. Ulasan akan diberi label berdasarkan beberapa kategori utama seperti, transaksi, keamanan, performa aplikasi, pelayanan, serta aktivasi dan verifikasi. Dengan adanya penelitian ini, diharapkan dapat memberikan gambaran yang lebih terstruktur tentang pengalaman pengguna, sehingga dapat dijadikan sebagai dasar evaluasi dalam upaya meningkatkan kualitas layanan aplikasi DANA. Dan juga diharapkan dapat memberikan informasi mengenai efisiensi dan efektivitas algoritma *Naïve Bayes* dalam klasifikasi dan prediksi.

TINJAUAN PUSTAKA

Penelitian sebelumnya, yang dilakukan oleh Gery Riansyah Ramadhan dan Castaka Agus Sugianto [7]. Penelitian tersebut membahas analisis ulasan sentimen aplikasi DANA menggunakan metode klasifikasi *Naïve Bayes*, dengan pendekatan Knowledge Discovery from Data (KDD). Tahapan penelitiannya meliputi, pengumpulan data (*scraping*), *labeling*, *preprocessing*, pembobotan kata, klasifikasi, dan evaluasi. Hasil penelitian ini menunjukkan bahwa akurasi mencapai 86%, dengan presisi dan *f1-score* juga masing-masing sebesar 86% dan 92%, serta tingkat *recall* yang sangat tinggi, yaitu 98%. Penelitian ini menunjukkan bahwa model analisis yang digunakan memiliki kemampuan untuk memberikan prediksi yang akurat. Meskipun terdapat perbedaan dalam nilai presisi, *recall*, dan *f1-score*, hasilnya tetap menunjukkan kinerja baik.

Selanjutnya penelitian dilakukan oleh Sheva Aditya [9], Penelitian ini membahas penerapan Algoritma TF-IDF dan *Naïve Bayes* terhadap ulasan pengguna aplikasi dompet digital Flip yang terdapat di *Google Play Store*. Penelitian ini memiliki tujuan untuk mengelompokkan ulasan tersebut berdasarkan tiga aspek penting: kecepatan, keamanan, dan biaya. Tahapan penelitiannya meliputi, pengumpulan data (*scraping*), *labeling manual*, *preprocessing*, pembobotan kata, model *Naïve Bayes*, dan evaluasi. Untuk hasil penelitiannya menunjukkan tingkat akurasi dengan rata-rata mencapai 84%. Dan penelitian tersebut dapat membantu pengembang Flip dalam memperbaiki aspek layanan yang paling banyak dikeluhkan pengguna.

Berdasarkan uraian diatas, dapat disimpulkan bahwa algoritma *Naïve Bayes* mampu secara efektif dalam mengolah data ulasan aplikasi, dengan menerapkan teknik pengolahan data teks seperti TF-IDF, data yang awalnya tidak terstruktur bisa diubah menjadi informasi yang lebih terarah dan bermakna. Kedua penelitian tersebut juga menunjukkan bahwa penggunaan algoritma yang tepat dan tahapan *preprocessing* yang baik sangat mempengaruhi hasil akhir klasifikasi. Oleh karena itu, penelitian ini difokuskan pada penerapan metode *Naïve Bayes* untuk melakukan klasifikasi dan prediksi terhadap ulasan pengguna aplikasi DANA, serta fitur TF-IDF untuk meningkatkan efisiensi proses klasifikasi. Dengan membagi kelas atau kategori, menjadi (1) Transaksi, (2) Keamanan, (3) Performa Aplikasi, (4) Pelayanan, serta (5) Aktivasi dan Verifikasi.

Data mining merupakan proses sistematis yang bertujuan menemukan pola atau informasi tersembunyi dalam kumpulan data yang besar dan kompleks. Proses ini dilakukan secara otomatis dengan menggunakan metode komputasi untuk menganalisis dan mengekstraksi pola atau korelasi dari data. Data mining menghasilkan wawasan berharga yang sulit ditemukan melalui metode manual, membantu dalam eksplorasi pengetahuan yang lebih mendalam [10].

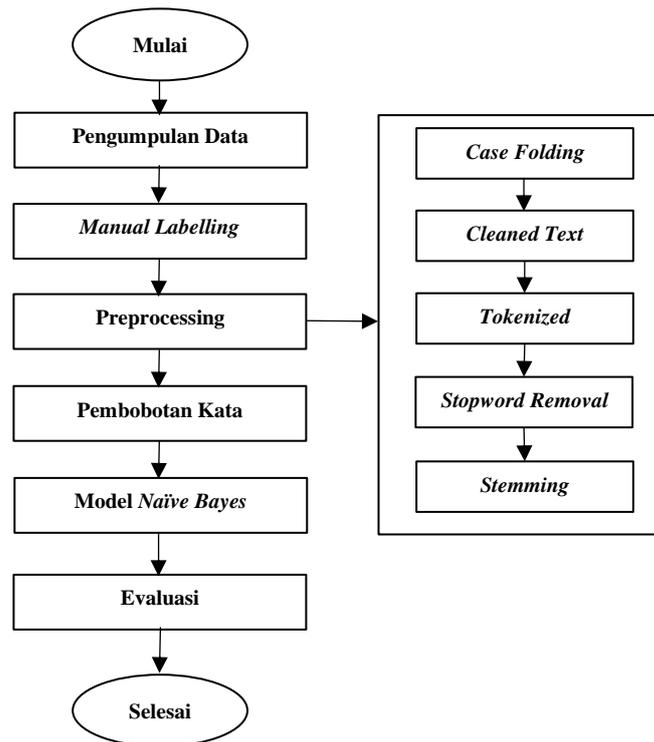
Pada data mining, klasifikasi merupakan suatu proses untuk membangun model atau fungsi yang mampu menggambarkan dan membedakan berbagai kategori atau kelas data. Sasaran utamanya adalah menentukan kelas dari objek yang belum diketahui labelnya [11].

Prediksi adalah proses memperkirakan atau menebak suatu nilai, kejadian, atau hasil yang belum terjadi berdasarkan data atau informasi yang telah tersedia sebelumnya [12]. Pada data mining, proses prediksi dilakukan dengan menggunakan model atau algoritma yang telah dilatih sebelumnya untuk mengenali pola dalam data.

Untuk melakukan proses klasifikasi dan prediksi ini, digunakan algoritma *Naïve Bayes*. *Naïve Bayes* merupakan salah satu model yang banyak digunakan karena mudah diterapkan dan cukup efisien. Setiap atribut dalam data dianggap memberi pengaruh yang sama dalam proses pengambilan keputusan. Selain itu, algoritma ini mampu menghitung nilai rata-rata dan sebaran data yang dibutuhkan untuk klasifikasi dan prediksi, meskipun hanya menggunakan sedikit data pelatihan [13].

METODE

Dalam penelitian ini, objek yang digunakan adalah ulasan aplikasi DANA Dompet Digital Indonesia yang diperoleh melalui layanan *Google Play Store* dengan menggunakan teknik *scraping*. Ulasan tersebut kemudian dikelompokkan ke dalam lima kategori yang telah ditentukan sebelumnya. Kategorinya, meliputi: (1) Transaksi, (2) Keamanan, (3) Performa Aplikasi, (4) Pelayanan, serta (5) Aktivasi dan Verifikasi. Adapun tahapan-tahapan penelitian ini akan diuraikan pada alur berikut.



Gambar 2. Alur Penelitian

Pengumpulan data

Pada tahap ini, data diperoleh dari platform *Google Play Store* melalui teknik *web scraping*. Teknik ini adalah metode otomatis yang memanfaatkan program atau skrip komputer untuk mengekstraksi data dari situs web [14]. Teknik *web scraping* pada penelitian ini diterapkan menggunakan *Google Colab* dengan bahasa pemrograman Python serta library *google-play-scraper* yang perlu diinstal terlebih dahulu ke dalam *Python*. Ulasan yang berhasil diambil dari proses *scraping* disimpan dalam bentuk dataframe, lalu dikonversi ke dalam format CSV untuk memudahkan proses pengolahan data.

Manual Labelling

Tahap selanjutnya adalah *manual labelling* atau pelabelan *manual*, di mana setiap ulasan akan dikelompokkan ke dalam 5 kategori berikut:

Transaksi

Kategori transaksi ini berfokus pada ulasan yang membahas seputar proses pembayaran, pengiriman uang, pengisian saldo.

Keamanan

Ulasan pada kategori keamanan yaitu ulasan yang membahas aspek keamanan data, keamanan transaksi, perlindungan akun, privasi, serta kasus penipuan.

Performa Aplikasi

Kategori yang berisi ulasan tentang kecepatan, kestabilan, *bug*, *error*, atau kemudahan penggunaan aplikasi.

Pelayanan

Kategori yang berisi ulasan mengenai *respons* dari *customer service*, pengalaman pengguna dengan dukungan teknis, serta kualitas layanan yang diberikan.

Aktivasi dan Verifikasi

Ulasan yang berkaitan dengan proses pendaftaran akun, verifikasi identitas, serta kendala selama proses aktivasi maupun verifikasi.

Manual labelling atau pelabelan manual yaitu pemberian label atau kategori secara manual oleh pelabel, berdasarkan sudut pandang masing-masing serta kriteria tertentu yang telah ditetapkan sebelumnya [15]. Proses ini dilakukan tanpa bantuan otomatisasi atau algoritma, melainkan melalui pengamatan dan interpretasi pelabel terhadap setiap ulasan.



Preprocessing

Preprocessing merupakan tahapan untuk menyiapkan atau membersihkan data agar siap untuk diproses pada tahapan selanjutnya, hal tersebut dikarenakan data ulasan yang diambil dari Layanan *Google Play Store* tidak terstruktur [16]. Tahapan pertama dalam proses *preprocessing* data pada penelitian ini adalah mengkonversi seluruh huruf kapital menjadi huruf kecil (*Case Folding*) [17]. Tahap menghapus karakter yang tidak bermakna dari data, seperti tanda baca, angka, maupun simbol yang tidak memiliki nilai informasi (*Cleaning text*) [18]. Selanjutnya, tahap pemisahan teks menjadi komponen-komponen yang lebih kecil seperti kata atau frasa, yang disebut sebagai token (*Tokenized*) [19]. Pada tahap *stopword removal*, kata-kata yang tidak relevan atau terlalu umum akan dihilangkan [20], contoh *stopword* dalam bahasa Indonesia yaitu “dan”, “yang”, “dari”. Dan terakhir, tahap *stemming*, bertujuan untuk mengganti kata imbuhan dengan kata dasar, sehingga kata tersebut dapat diubah dari kata turunan menjadi bentuk kata dasarnya (kata baku) [21].

Pembobotan Kata

Teknik terminologi pembobotan frekuensi, yang dikenal sebagai *Term Frequency-Inverse Document Frequency* (TF-IDF), digunakan untuk mengkonversi data berbentuk teks menjadi data berbentuk angka atau numerik. Tujuan dari metode ini yaitu memberikan bobot pada kata-kata atau fitur tertentu dalam sebuah dataset. TF-IDF merupakan metode yang memberikan nilai bobot pada kata-kata berdasarkan berapa kali kata-kata tersebut muncul dalam dokumen. Suatu term, yang dapat berupa kata, frasa, atau elemen indeks dalam naskah, ditunjukkan dengan *Term Frequency* (TF). Semakin sering sebuah istilah digunakan, semakin tinggi bobotnya [22].

Model Naïve Bayes

Dalam penelitian ini, pemodelan yang diterapkan adalah *Multinomial Naïve Bayes*. Model ini dipilih karena kesederhanaannya serta popularitasnya yang tinggi. Model ini memungkinkan setiap atribut berkontribusi secara setara pada keputusan akhir. Salah satu keunggulan dari metode *Naïve Bayes* adalah kemampuannya untuk memperkirakan parameter rata-rata dan variasi variabel yang diperlukan untuk klasifikasi, hanya dengan memerlukan jumlah data latihan yang relatif kecil. Teorema bayes mempunyai rumus:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Keterangan :

Pada persamaan ini, $P(c|x)$ merepresentasikan probabilitas posterior untuk kelas c berdasarkan fitur x . Sementara itu, $P(x|c)$ menyatakan probabilitas kemunculan fitur x dalam kelas c . Adapun $P(c)$ menggambarkan probabilitas prior untuk kelas c , dan $P(x)$ menunjukkan probabilitas keseluruhan dari fitur x .

Sebelum tahap pemodelan dilakukan, dataset perlu dipisahkan menjadi dua bagian. Bagian pertama, yaitu data latih, digunakan untuk melatih model dalam mengenali pola klasifikasi, sementara bagian kedua, yaitu data uji, dimanfaatkan untuk menilai seberapa baik kinerja model tersebut [23]. Akurasi model yang lebih baik dapat dicapai dengan jumlah data latih yang lebih besar. Pemodelan ini bertujuan untuk melatih model agar dapat mengklasifikasikan dan memprediksi data.

Evaluasi

Langkah berikutnya adalah evaluasi, tujuan dari tahap evaluasi adalah untuk mengidentifikasi dan menetapkan hasil dari penerapan model yang telah dilakukan pada tahap sebelumnya [24]. Dalam evaluasi model, terdapat berbagai metrik yang digunakan, termasuk akurasi, presisi, *recall*, dan *f1-score*. Akurasi mengacu pada persentase prediksi yang benar dari keseluruhan data. Presisi menunjukkan seberapa tepat model dalam memprediksi suatu kelas tertentu, *recall* mengukur kemampuan model dalam menemukan seluruh data yang termasuk dalam suatu kelas tertentu, dan *f1-score* menggambarkan keseimbangan antara presisi dan *recall*. Selain itu, dalam tahap evaluasi *confusion matrix* juga digunakan untuk memberikan gambaran yang lebih mendetail tentang performa model. *Confusion matrix* merupakan suatu alat yang dimanfaatkan untuk mengevaluasi kinerja model dengan cara menunjukkan label hasil prediksi model dan label manual. *Matrix* ini menunjukkan seberapa akurat model memprediksi kelas yang benar, serta menggambarkan kesalahan yang dilakukan model dalam proses klasifikasi dan prediksinya.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Dalam penelitian ini, data diambil dari ulasan pengguna aplikasi DANA yang tersedia di *Google Play Store*. Pengumpulan data dilakukan melalui metode web scraping yang menggunakan Python, dengan dukungan platform *Google Colab*. Hasil *scraping* ulasan pengguna aplikasi DANA tersebut memperoleh 1.500 ulasan dari bulan September 2024 - Desember 2024.



Tabel 1. *Dataset*

No.	Teks Ulasan	Tanggal
1	Tolong bantuannya, saya tidak bisa cek riwayat transaksi. Menu Aktivitas tidak berfungsi sama sekali, muter2 loading di tunggu berjam jam tetep tidak terbuka. Upaya yang sudah saya lakukan bersihkan chace, data sudah.	2024-11-28 00:12:02
2	Yang mau instal aplikasi ini mending mikir" lagi, apalagi ini aplikasi buat transaksi. Saya pakai ini berjalan 1 tahun normal" saja. Baru hari ini tadi tiba" saldo saya hilang. Ada notifikasi pembayaran tanpa konfirmasi pin sama sekali.	2024-12-10 09:51:25
3	HAI min aku mau kasih tahu ada masalah di bagian verifikasi wajah sudah berapa kali coba di pencahayaan yang terang sampai di bawah lampu tapi tetep saja hasil enggak bisa dengan alasan pencahayaan kurang tolong di perbaiki.	2024-12-11 11:52:36
4	semudah itu aplikasi kalian down hanya karena banyak yang sedang transaksi, cuma aplikasi keuangan biasa bisa-bisanya down tolong tingkatkan lagi fungsi aplikasi kalian	2024-11-28 04:25:57
5	Kenapa aplikasi begini ya.. Jelek banget, saya yang dari kemarin punya keluhan tidak di respon dengan baik malah di persulit, padahal nomor handphone selalu aktif...	2024-12-02 10:10:52
6	Terulang lagi, tahun kemarin kirim saldo tidak terkirim ndak ada pengembalian atau pemberitahuan lebih lanjut. Ehh sekarangg transfer dua kali ke no rek bank, yang pertama tulisannya sukses tapi belum masuk ke rek tujuan	2024-11-22 17:00:10

Manual Labelling

Pada *manual labelling*, data yang diperoleh dikategorikan kedalam 5 kategori meliputi: (1) Transaksi, (2) Keamanan, (3) Performa Aplikasi, (4) Pelayanan, serta (5) Aktivasi dan Verifikasi. Dalam tahapan ini pelabel menggunakan microsoft excel untuk menilai ulasan termasuk ke dalam kategori mana. Hasil data ulasan yang dilabeli sebanyak 1.499 ulasan.

Tabel 2. Hasil *Manual Labelling*

No.	Teks Ulasan	Label Manual
1	Tolong bantuannya, saya tidak bisa cek riwayat transaksi. Menu Aktivitas tidak berfungsi sama sekali, muter2 loading di tunggu berjam jam tetep tidak terbuka. Upaya yang sudah saya lakukan bersihkan chace, data sudah.	Performa Aplikasi
2	Yang mau instal aplikasi ini mending mikir" lagi, apalagi ini aplikasi buat transaksi. Saya pakai ini berjalan 1 tahun normal" saja. Baru hari ini tadi tiba" saldo saya hilang. Ada notifikasi pembayaran tanpa konfirmasi pin sama sekali.	Keamanan
3	HAI min aku mau kasih tahu ada masalah di bagian verifikasi wajah sudah berapa kali coba di pencahayaan yang terang sampai di bawah lampu tapi tetep saja hasil enggak bisa dengan alasan pencahayaan kurang tolong di perbaiki.	Aktivasi dan Verifikasi
4	semudah itu aplikasi kalian down hanya karena banyak yang sedang transaksi, cuma aplikasi keuangan biasa bisa-bisanya down tolong tingkatkan lagi fungsi aplikasi kalian	Performa Aplikasi
5	Kenapa aplikasi begini ya.. Jelek banget, saya yang dari kemarin punya keluhan tidak di respon dengan baik malah di persulit, padahal nomor handphone selalu aktif...	Pelayanan
6	Terulang lagi, tahun kemarin kirim saldo tidak terkirim ndak ada pengembalian atau pemberitahuan lebih lanjut. Ehh sekarangg transfer dua kali ke no rek bank, yang pertama tulisannya sukses tapi belum masuk ke rek tujuan	Transaksi

Preprocessing

Setelah pengumpulan proses *manual labelling*, *preprocessing* adalah langkah berikutnya yang dilakukan. *Preprocessing* dilakukan dengan tujuan menghilangkan data yang tidak diperlukan, menghasilkan dokumen yang lebih rinci dan terorganisir, dan menyiapkan data agar siap untuk proses selanjutnya. Proses ini sangat penting dikarenakan data ulasan yang dikumpulkan dari Layanan *Google Play Store* seringkali tidak teratur. Proses *preprocessing* ini terdiri dari beberapa langkah :



Case Folding

Tahapan pertama dalam proses preprocessing data pada penelitian ini adalah mengkonversi seluruh huruf kapital menjadi huruf kecil.

Tabel 3. Contoh Hasil Case Folding

Teks Ulasan	Case Folding
Tolong bantuannya, saya tidak bisa cek riwayat transaksi. Menu Aktivitas tidak berfungsi sama sekali, muter2 loading di tunggu berjam jam tetep tidak terbuka. Upaya yang sudah saya lakukan bersihkan chace, data sudah.	tolong bantuannya, saya tidak bisa cek riwayat transaksi. menu aktivitas tidak berfungsi sama sekali, muter2 loading di tunggu berjam jam tetep tidak terbuka. upaya yang sudah saya lakukan bersihkan chace, data sudah.

Clean Text

Proses menghapus angka, tanda baca seperti titik, koma, tanda seru, tanda tanya, simbol-simbol, dan lainnya yang tidak memiliki nilai informasi. Pada teks dibawah ini terdapat koma, angka, dan titik.

Tabel 4. Contoh Hasil Clean Text

Case Folding	Cleaning Text
Tolong bantuannya, saya tidak bisa cek riwayat transaksi. Menu Aktivitas tidak berfungsi sama sekali, muter2 loading di tunggu berjam jam tetep tidak terbuka. Upaya yang sudah saya lakukan bersihkan chace, data sudah.	tolong bantuannya saya tidak bisa cek riwayat transaksi menu aktivitas tidak berfungsi sama sekali muter loading di tunggu berjam jam tetep tidak terbuka upaya yang sudah saya lakukan bersihkan chace data sudah

Tokenized

Proses pemisahan teks menjadi komponen-komponen yang lebih kecil, seperti kata atau frasa.

Tabel 5. Contoh Hasil Tokenized

Cleaning Text	Tokenized
tolong bantuannya saya tidak bisa cek riwayat transaksi menu aktivitas tidak berfungsi sama sekali muter loading di tunggu berjam jam tetep tidak terbuka upaya yang sudah saya lakukan bersihkan chace data sudah	['tolong', 'bantuannya', 'saya', 'tidak', 'bisa', 'cek', 'riwayat', 'transaksi', 'menu', 'aktivitas', 'tidak', 'berfungsi', 'sama', 'sekali', 'muter', 'loading', 'di', 'tunggu', 'berjam', 'jam', 'tetep', 'tidak', 'terbuka', 'upaya', 'yang', 'sudah', 'saya', 'lakukan', 'bersihkan', 'chace', 'data', 'sudah']

Stopword Removal

Proses menghapus kata-kata umum atau kata-kata yang sering muncul namun tidak memberikan makna yang signifikan.

Tabel 6. Contoh Hasil Stopword Removal

Tokenizes	Stopword Removal
['tolong', 'bantuannya', 'saya', 'tidak', 'bisa', 'cek', 'riwayat', 'transaksi', 'menu', 'aktivitas', 'tidak', 'berfungsi', 'sama', 'sekali', 'muter', 'loading', 'di', 'tunggu', 'berjam', 'jam', 'tetep', 'tidak', 'terbuka', 'upaya', 'yang', 'sudah', 'saya', 'lakukan', 'bersihkan', 'chace', 'data', 'sudah']	tolong bantuannya cek riwayat transaksi menu aktivitas berfungsi muter loading tunggu berjam jam tetep terbuka upaya lakukan bersihkan chace data

Stemming

Pada tahap stemming, bertujuan untuk mengganti kata imbuhan dengan kata dasar, sehingga kata tersebut dapat diubah dari kata turunan menjadi bentuk kata dasarnya (kata baku).

Tabel 7. Contoh Hasil Stemming

Stopword Removal	Stemming
tolong bantuannya cek riwayat transaksi menu aktivitas berfungsi muter loading tunggu berjam jam tetep terbuka upaya lakukan bersihkan chace data	tolong bantu cek riwayat transaksi menu aktivitas fungsi muter loading tunggu jam jam tetep buka upaya laku bersih chace data



Pembobotan Kata

Selanjutnya pembobotan kata, metode *Term Frequency-Inverse Document Frequency* (TF-IDF) diterapkan untuk mengonversi teks yang telah melalui tahap *preprocessing* menjadi bentuk numerik, setiap kata dalam ulasan akan diberi bobot nilai berdasarkan perhitungan yang menggunakan algoritma TF-IDF ini.

Tabel 8. *Matrix* Hasil Pembobotan TF-IDF

No. Dok	dana	aplikasi	transaksi	...	nomor	premium	bagus
1	0.028	0.000	0.042	...	0.000	0.000	0.061
2	0.000	0.000	0.136	...	0.000	0.000	0.000
3	0.000	0.000	0.053	...	0.000	0.000	0.000
4	0.000	0.000	0.000	...	0.000	0.000	0.000
5	0.088	0.000	0.000	...	0.000	0.000	0.000
...
1495	0.000	0.179	0.000	...	0.000	0.000	0.000
1496	0.044	0.101	0.065	...	0.000	0.000	0.000
1497	0.117	0.000	0.000	...	0.000	0.129	0.000
1498	0.000	0.000	0.088	...	0.000	0.000	0.000
1499	0.000	0.000	0.000	...	0.000	0.000	0.145

Berdasarkan hasil Pembobotan TF-IDF diatas, teks ulasan telah dikonversi kedalam bentuk numerik. Sebagai contoh, pada Dokumen ke-1 menyatakan bahwa kata “transaksi” dengan bobot 0.42, kata “bagus” dengan bobot 0,061 artinya kata tersebut muncul dan cukup penting dalam dokumen ke-1. Lalu ada juga kata “premium” memiliki bobot 0.000, berarti kata tersebut tidak muncul dalam dokumen ke-1. Sama halnya dengan dokumen-dokumen lain, semakin tinggi bobotnya maka menandakan semakin penting kata tersebut dalam dokumen tertentu.

Model Naïve Bayes

Setelah data dikonversi menggunakan metode TF-IDF, langkah berikutnya adalah membagi data menjadi dua bagian, yaitu data latih dan data uji, dengan rasio 80:20. Penelitian ini menerapkan model *Multinomial Naïve Bayes*, karena algoritma ini mampu memberikan hasil yang akurat untuk melakukan klasifikasi dan prediksi berbasis data teks, terutama yang telah dikonversi ke dalam bentuk terstruktur seperti TF-IDF. Selain kesederhanaannya, algoritma ini juga memiliki performa yang baik untuk tugas klasifikasi teks.

Tabel 9. Pembagian Data 80:20

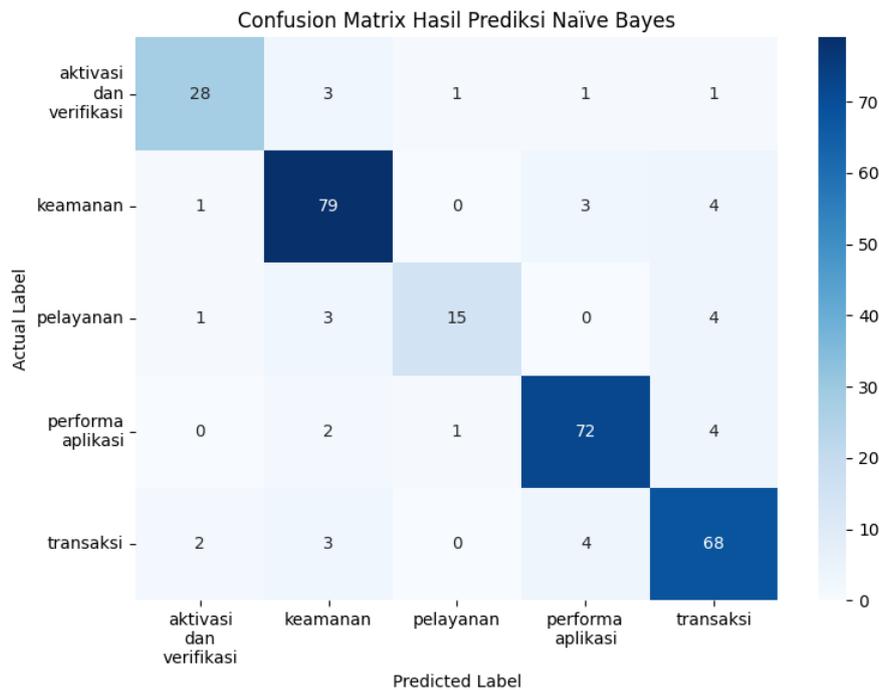
Data Latih	Data Uji
1.199	300

Evaluasi

Setelah model dilatih menggunakan algoritma *Multinomial Naïve Bayes*, dilakukan evaluasi dan visualisasi untuk melihat performa model secara menyeluruh. Dari hasil evaluasi, diketahui bahwa model berhasil mencapai akurasi sebesar 87%, yang mengindikasikan bahwa model mampu mengklasifikasikan seluruh kategori.

Tabel 10. *Classification Report*

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
aktivasi dan verifikasi	0.88	0.82	0.85	34
keamanan	0.88	0.91	0.89	87
pelayanan	0.88	0.65	0.75	23
Performa aplikasi	0.90	0.91	0.91	79
transaksi	0.84	0.88	0.86	77
accuracy			0.87	300
macro avg	0.87	0.84	0.85	300
weighted avg	0.87	0.87	0.87	300



Gambar 3. Confusion Matrix Hasil Prediksi Naïve Bayes

Selain metrik evaluasi seperti akurasi, presisi, *recall*, dan *f1-score*, visualisasi dalam bentuk *confusion matrix* juga dimanfaatkan untuk memperlihatkan detail klasifikasi dari setiap kategori.

Berdasarkan *confusion matrix*, diketahui bahwa:

1. Model berhasil mengklasifikasikan 28 dari 34 data kategori aktivasi dan verifikasi dengan benar, sedangkan sisanya keliru diprediksi ke kategori lain.
2. Untuk kategori keamanan, model mengklasifikasikan 79 data benar dari 87 data.
3. Pada kategori pelayanan, hanya 15 dari 23 data yang diklasifikasikan dengan benar. Sisanya keliru diklasifikasikan sebagai aktivasi dan verifikasi, keamanan, dan transaksi.
4. Kategori performa aplikasi menunjukkan performa yang sangat baik dengan 71 dari 79 data diklasifikasikan dengan benar.
5. Pada kategori transaksi, 69 dari 77 data berhasil dikenali dengan benar.

Berdasarkan hasil klasifikasi dan prediksi terhadap dataset uji yang terdiri dari 300 ulasan, ditemukan bahwa sebanyak 262 ulasan diprediksi akurat, sementara 38 ulasan diprediksi tidak tepat. Adapun dibawah ini contoh hasil klasifikasi.

Tabel 11. Contoh Hasil Klasifikasi Label

No.	Teks Ulasan	Label Manual	Label Prediksi
1	Tolong bantuannya, saya tidak bisa cek riwayat transaksi. Menu Aktivitas tidak berfungsi sama sekali, muter2 loading di tunggu berjam jam tetep tidak terbuka. Upaya yang sudah saya lakukan bersihkan chace, data sudah.	Performa Aplikasi	Performa Aplikasi
2	Yang mau instal aplikasi ini mending mikir" lagi, apalagi ini aplikasi buat transaksi. Saya pakai ini berjalan 1 tahun normal" saja. Baru hari ini tadi tiba" saldo saya hilang. Ada notifikasi pembayaran tanpa konfirmasi pin sama sekali.	Keamanan	Keamanan
3	HAI min aku mau kasih tahu ada masalah di bagian verifikasi wajah sudah berapa kali coba di pencahayaan yang terang sampai di bawah lampu tapi tetep saja hasil enggak bisa dengan alasan pencahayaan kurang tolong di perbaiki.	Aktivasi dan Verifikasi	Aktivasi dan Verifikasi
4	semudah itu aplikasi kalian down hanya karena banyak yang sedang transaksi, cuma aplikasi keuangan biasa bisa-bisanya down tolong tingkatkan lagi fungsi aplikasi kalian	Performa Aplikasi	Keamanan
5	Kenapa aplikasi begini ya.. Jelek banget, saya yang dari kemarin punya keluhan tidak di respon dengan baik malah di persulit, padahal nomor handphone selalu aktif...	Pelayanan	Pelayanan
6	Terulang lagi, tahun kemarin kirim saldo tidak terkirim ndak ada pengembalian atau pemberitahuan lebih lanjut. Ehh sekarangg transfer dua kali ke no rek bank, yang pertama tulisannya sukses tapi belum masuk ke rek tujuan	Transaksi	Transaksi



KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan, klasifikasi dan prediksi terhadap ulasan aplikasi DANA pada *Google Play Store* yang berjumlah 1.500 data ulasan, dengan pembagian data 80:20 diperoleh bahwa 262 ulasan diprediksi benar, sedangkan 38 ulasan diprediksi salah. Untuk kategori “aktivasi dan verifikasi” dari 34 terdapat 28 yang benar, kategori “keamanan” dari 87 terdapat 79 yang benar, kategori “pelayanan” dari 23 terdapat 15 yang benar, kategori “performa aplikasi” dari 79 terdapat 71 yang benar, dan pada kategori “transaksi” dari 77 terdapat 69 yang benar. Untuk hasil evaluasi dari model menggunakan algoritma *Multinomial Naïve Bayes* menunjukkan tingkat akurasi 87%, presisi 87%, *recall* 84%, dan *f1-score* 85%. Dapat disimpulkan bahwa model mampu dalam mengklasifikasikan seluruh kategori. Pada hasil tersebut diketahui bahwa kategori Keamanan merupakan kategori dengan jumlah ulasan terbanyak, yaitu sebanyak 87 data ulasan. Hal ini menunjukkan bahwa isu terkait perlindungan data, keamanan transaksi, dan keamanan akun menjadi perhatian utama bagi pengguna aplikasi DANA. Adapun saran untuk penelitian berikutnya adalah pada tahap *preprocessing* gunakan teks berbahasa Inggris agar hasil dari pembersihan datanya lebih akurat, dan juga disarankan untuk melakukan komparasi dengan berbagai algoritma lainnya, seperti *K-Nearest Neighbour*, *Support Vector Machines*, dan lainnya, agar memiliki hasil komparasi yang beragam.

DAFTAR PUSTAKA

- [1] F. Puspitaningrum, S. C. Kusumastuti, and A. Rimbawati, “Penggunaan QRIS Dalam Transaksi Jual Beli di Tengah Masyarakat UMKM Ketintang Surabaya,” *Prosiding Seminar Nasional*, pp. 603–614, 2023.
- [2] S. S. Hilabi, “Transformasi Digitalisasi Marketing pada UMKM (Study Kasus Desa Gintung Kerta),” *Konf. Nas. Penelit. dan Pengabd.*, vol. 2, pp. 1855–1863, 2022.
- [3] D. Mertiwi, B. Santoso, and B. Intan, “Analisis Sentimen Aplikasi Dana Pada Google Play Store Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor,” 2024.
- [4] D. Anjani, H. Awali, D. N. Misidawati, F. Ekonomi, B. Islam, and I. Pekalongan, “Analisis Faktor-Faktor Yang Mempengaruhi Minat Generasi Z Dalam Menggunakan Sistem Pembayaran E-Wallet,” 2022. [Online]. Available: www.dana.id.
- [5] R. Ardila and T. Sutabri, “IJM: Indonesian Journal of Multidisciplinary Analisis Sudut Pandang Mahasiswa Dalam Performa dan Efisiensi Aplikasi Dana Dalam Kehidupan Sehari Hari Menggunakan Metode Deskriptif,” *IJM: Indonesian Journal of Multidisciplinary*, vol. 2, 2024, [Online].
- [6] A. Hissi Heryanti, S. Karawang, J. H. Ronggo Waluyo, T. Timur, and J. Barat, “Pengaruh Kualitas Layanan Aplikasi Dana terhadap Kepuasan Pelanggan dalam Melakukan Transaksi Secara Online sebagai Alat Pembayaran Elektronik (E-Payment),” *Journal on Education*, vol. 05, no. 03, pp. 8080–8096, 2023.
- [7] G. Riansyah Ramadhan and C. Agus Sugianto, “Analisis Sentimen Ulasan Aplikasi Dana Di Google Play Store Menggunakan Algoritma Naïve Bayes,” 2024.
- [8] I. N. KUSUMA ET AL., “Analisis Sentimen Pada Pengguna Aplikasi Dana Menggunakan Algoritma Naive Bayes,” 2024. [Online]. Available: <https://play.google.com/store/apps/details?id=id.dana>.
- [9] S. A. Helmayanti, F. Hamami, and R. Y. Fa’rifah, “Penerapan Algoritma Tf-Idf Dan Naïve Bayes Untuk Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Flip Pada Google Play Store,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 1822–1834, Sep. 2023, doi: 10.35870/jimik.v4i3.415.
- [10] Tukino, “Analisis Prediksi Penjualan Sayuran Hidroponik Menggunakan Algoritma C4.5 pada PT Batam Indo Agri Perkasa,” *Jurnal Desain Dan Analisis Teknologi (JDDAT)*, 2023.
- [11] M. A. Mukti, W. Wiyanto, and I. Nawangsih, “Implementasi Algoritma Naive Bayes Untuk Prediksi Calon Penerima Bt (Studi Kasus Desa Tridayasakti, Kabupaten Bekasi),” 2023.
- [12] A. Fahreza, “Penerapan Data Mining dengan Metode Single Moving Average dalam Pengolahan Data Penerimaan Siswa Baru,” 2022.
- [13] Rian Pratama, B. Huda, E. Novalia, and H. Kabir, “Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Menentukan Persediaan Stok,” *Metik Jurnal*, vol. 6, no. 2, pp. 115–122, Dec. 2022, doi: 10.47002/metik.v6i2.379.
- [14] Mursyidah, M. Davi, and N. S. Dheyai, “Klasifikasi Sentimen Review Pengguna terhadap Aplikasi Instagram menggunakan Algoritma Random Forest,” 2024.
- [15] I. Muhammad Falih, N. Hafifah Matondang, and N. Chamidah, “Seleksi Fitur Information Gain pada Analisis Sentimen Terhadap Ulasan Aplikasi Flip dengan Algoritma Support Vector Machine,” 2022.
- [16] G. Riansyah Ramadhan and C. Agus Sugianto, “Analisis Sentimen Ulasan Aplikasi Dana Di Google Play Store Menggunakan Algoritma Naïve Bayes,” 2024.
- [17] Tukino, A. Hananto, R. A. Nanda, E. Novalia, E. Sedyono, and J. Sanjaya, “LSTM and Word Embedding: Classification and Prediction of Puskesmas Reviews Via Twitter,” in *E3S Web of Conferences*, EDP Sciences, Mar. 2024. doi: 10.1051/e3sconf/202450001018.
- [18] M. Irfan and D. Widiyanto, “Analisis Sentimen Pada Ulasan Aplikasi Samsat Mobile Jawa Barat (Sambara) Menggunakan algoritma Naive bayes Classifier dengan seleksi fitur Chi Square,” 2022.
- [19] R. Zulfiqri, B. N. Sari, and T. N. Padilah, “Analisis Sentimen Ulasan Pengguna Aplikasi Media Sosial Instagram Pada Situs Google Play Store Menggunakan Naive Bayes Classifier,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4995.
- [20] L. Kusneti, A. Ratu, and A. Wijaya, “Analisis Sentimen Ulasan Aplikasi LinkedIn Dalam Google Play Store Dengan Model Naïve Bayes,” *Jurnal Teknologi Informasi*, vol. 4, no. 2, 2023, doi: 10.46576/djtechno.



- [21] A. Rohanah *et al.*, “Klasifikasi Ulasan Pengguna Zoom Cloud Meetings Menggunakan Metode Information Gain dan Naïve Bayes Classifier,” vol. 6, no. 2, pp. 348–357, 2021, doi: 10.32493/informatika.v6i2.10728.
- [22] Cindy Caterine Yolanda, Syafriandi Syafriandi, Yenni Kurniawati, and Dina Fitria, “Sentiment Analysis of DANA Application Reviews on Google Play Store Using Naïve Bayes Classifier Algorithm Based on Information Gain,” *UNP Journal of Statistics and Data Science*, vol. 2, no. 1, pp. 48–55, Feb. 2024, doi: 10.24036/ujsds/vol2-iss1/147.
- [23] J. Sanjaya, B. Priyatna, S. Shofia Hilabi, S. Informasi, F. Ilmu Komputer, and U. Buana Perjuangan Karawang, “Analisis Sentimen Terhadap Opini Proyek Kereta Cepat Menggunakan Metode Naïve Bayes Classifier,” *Jurnal Fasilkom*, 2024.
- [24] F. M. Delta Maharani, A. Lia Hananto, S. Shofia Hilabi, F. Nur Apriani, A. Hananto, and B. Huda, “Perbandingan Metode Klasifikasi Sentimen Analisis Penggunaan E-Wallet Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor,” *Metik Jurnal*, vol. 6, no. 2, pp. 97–103, Dec. 2022, doi: 10.47002/metik.v6i2.372.