JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 13 No. 2, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i2.6423

IMPLEMENTASI METODE NAÏVE BAYES PADA ANALISIS SENTIMEN PENGGUNA APLIKASI MOBILE KITA BISA

I Gede Surya Dharma Putra[1*], I Nyoman Tri Anindia Putra[2]
[1,2]Sistem Informasi, Universitas Pendidikan Ganesha; Jalan Udayana No 11 Singaraja, Banjar Tegal, Singaraja, Kabupaten Buleleng, Bali 81116; (0362) 22570

Received: 8 Maret 2025
Accepted: 27 Maret 2025
Published: 14 April 2025

Keywords:
Kita Bisa;
Naïve Bayes;
Sentiment Analysis;
Data Mining;
Mobile Application;
KDD Methodology

Corespondent Email:
surya.dharma@student.undiksha.ac.id

Abstrak. Kita Bisa merupakan aplikasi filantropi digital yang diperuntukkan mengumpulkan dana dari berbagai aktivis guna membantu mereka yang sedang membutuhkan. Meskipun memiliki tujuan yang baik dalam pengembangan inovasi aplikasi ini, terdapat pengguna yang kurang puas terhadap fitur dan alur kerja yang ditawarkan aplikasi tersebut. Dalam memahami bagaimana sentimen pengguna, penelitian ini melakukan analisis dengan metode Naïve Bayes dan Knowledge Discovery in Database (KDD) sebagai kerangka kerja dalam analisis. Data yang digunakan diambil dari ulasan pengguna dari platform Google Play Store dengan jumlah sebanyak 1000 ulasan dan dari data tersebut yang akan digunakan hanya data dengan rentang waktu dari januari 2024 sampai dengan maret 2025. Hasil penelitian dengan implementasi metode naïve bayes menunjukkan nilai akurasi 81% yang mengindikasikan bahwa metode telah berhasil dalam menganalisis sentimen pengguna. Temuan dalam penelitian menunjukkan bahwa sentimen positif lebih dominan dan meskipun sentimen negatif berjumlah sedikit masih diperlukan adanya evaluasi dan perbaikan berdasarkan keluhan pengguna untuk pengembangan aplikasi.

Abstract. *Kita Bisa is a digital philanthropy application aimed at collecting funds from various activists to help those in need. Although it has a good objective in the development of this application innovation, there are users who are dissatisfied with the features and workflows offered by the application. To understand how user sentiment works, this study conducted an analysis using the Naïve Bayes method and Knowledge Discovery in Database (KDD) as the framework for the analysis. The data used is taken from user reviews from the Google Play Store platform with a total of 1000 reviews, and from this data only reviews within the time range from January 2024 to March 2025 will be used. The results of the study implementing the Naïve Bayes method show an accuracy of 81%, indicating that the method has been effective in analyzing user sentiment. The findings in the study show that positive sentiment is more dominant and although negative sentiment is small, evaluation and improvement based on user complaints is still needed for application development.*

1. PENDAHULUAN

Dalam beberapa tahun terakhir, banyak dari masyarakat yang terdampak oleh musibah yang disebabkan oleh bencana alam, krisis kesehatan, dan kecelakaan. Untuk mengatasi musibah yang dialami, banyak masyarakat yang peduli dengan melakukan penggalangan dana. Tetapi penggalangan dana sangat terbatas oleh jarak dan waktu, biasanya hanya dilakukan melalui informasi mulut ke mulut, turun ke jalan dan melalui brosur. Dan seiring berkembang pesatnya era teknologi dari tahun ke tahun, aplikasi mobile turut mengalami perkembangan yang signifikan dan telah

menjadi kebutuhan masyarakat karena mudah diakses, nyaman untuk digunakan dan tidak terbatas jarak sosial. Dengan teknologi seperti AI (*Artificial Intelegence*), *Blockchain*, *Machine learning*, IoT (*Internet of Things*), AR (*Augmented Reality*) aplikasi mobile telah meningkat dari segi ketersediaannya pada platform Google Play Store maupun App Store, dan menjangkau seluruh aspek yang ada pada masyarakat seperti pendidikan, kesehatan, ekonomi dan sosial[1], [2], [3], [4], [5]. Khususnya pada aspek sosial saat ini telah muncul inovasi aplikasi filantropi digital yaitu Kita Bisa, aplikasi yang tersedia pada platform Google Play Store dan App Store ini telah berdiri sejak tahun 2013 sebagai aplikasi yang diperuntukkan mengumpulkan dana dari berbagai aktivis guna membantu mereka yang sedang membutuhkan. Dengan adanya aplikasi ini juga memungkinkan individu atau organisasi penggalang dana untuk menjangkau lebih banyak audiens, terbukti dengan 10 Juta orang yang sudah berpartisipasi dalam kegiatan penggalangan dana ini. Informasi yang didapat dari aplikasi ini setiap bulannya sudah memfasilitasi 2,5 Juta transaksi, 28.000 galang dana sosial, ratusan yayasan dan lembaga yang tersebar di 34 Provinsi dan termasuk ribuan pasien yang membutuhkan di 100 lebih rumah sakit di Indonesia. Meskipun dengan tujuan yang baik dalam pengembangan inovasi aplikasi ini, masih banyak pengguna yang kurang puas terhadap fitur dan alur kerja yang ditawarkan aplikasi tersebut. Dengan demikian, analisis terhadap faktor-faktor yang mempengaruhi pendapat, kepuasan maupun ketidakpuasan dari pengguna dalam menggunakan aplikasi Kita Bisa sangat penting dikarenakan dapat menjadi pedoman dalam pengembangan atau perbaikan selanjutnya dari aplikasi. Dalam menganalisis sentimen dari pengguna aplikasi, data mining akan digunakan sebagai solusi untuk mengatasi permasalahan tersebut. Data mining merupakan metode yang digunakan untuk memprediksi bagaimana fenomena dari suatu objek data besar yang digunakan untuk diteliti lebih lanjut, baik dengan mengetahui atau tidak mengetahui tentang kelas dari data tersebut[6].

Pada data mining, salah satu hal yang menjadi komponen penting adalah teknik klasifikasi, yaitu metode untuk mempelajari kumpulan data dengan tujuan menciptakan aturan yang dapat mengklasifikasikan atau mengenali data baru yang belum pernah dianalisis sebelumnya[7]. Dan klasifikasi adalah proses menetapkan suatu objek data ke dalam salah satu kategori yang telah ditentukan sebelumnya. Ada beberapa metode teknik klasifikasi yang dapat digunakan yaitu *Decision Tree*, *Rainforest*, *Naïve Bayesian*, *Support Vector Machine*, *Neural Network*, *Genetic Algorithm*, *Fuzzy*, *Case-Based Reasoning*, dan *K-Nearest Neighbor*[8], [9]. Pada penelitian ini mengimplementasikan metode dari *Naïve Bayes* untuk menganalisis bagaimana sentimen pengguna dari data set yang diambil dari ulasan aplikasi di Google Play Store, Metode ini dipilih karena sangat efektif untuk digunakan dalam menganalisis sentimen publik. Selain itu, pada *Naïve Bayes* juga didasarkan oleh teorema bayes yang memiliki kemampuan klasifikasi yang mirip dengan decision tree dan neural network, kemudian naïve bayes juga sering digunakan karena kesederhanaanya saat mengklasifikasikan sebuah teks pada data[10], [11]. Pada penelitian sebelumnya, terlihat metode *naïve bayes* dapat bekerja dengan baik dalam menganalisis sentimen publik terhadap pariwisata bali saat masa pandemi dengan nilai akurasi sebesar 81,70% dan nilai F-Measure sebesar 82,02%, meskipun nilai ini masih dipengaruhi oleh kualitas data yang masih memiliki banyak komentar singkatan, kata non-standar dan penggunaan emoji[11]. Selanjutnya, penelitian pada [12] menunjukkan peningkatan akurasi dalam penggunaan metode *Multinomial Naïve Bayes* untuk menganalisis aplikasi MyBlueBird. Performa model naïve bayes yang didapat yaitu tingkat akurasi sebesar 92%, dengan evaluasi per sentimen yaitu presisi 91%, recall 100%, dan F1-Score 95% untuk sentimen Positif, serta presisi 100%, recall 81%, dan F1-Score 90% untuk sentimen Negatif, sedangkan netral memiliki presisi 100%, netral 10% dan F1-Score hanya 18%. Sementara itu, penelitian lain tentang analisis sentimen terhadap aplikasi “Ojol The Game” menunjukkan hasil yang positif dengan akurasi sebesar 94.12% dari 1000 ulasan data yang digunakan[13].

Tujuan dari penelitian ini adalah menerapkan metodologi KDD (*Knowledge Discovery in Database*) sebagai kerangka kerja dalam proses analisis data dan model naïve bayes dalam mengklasifikasikan teks ulasan

dan komentar pengguna secara efesien menjadi beberapa kategori yaitu positif, negatif, dan neutral pada studi kasus aplikasi Kita Bisa. Dan penelitian ini juga akan mengukur kinerja dari model naïve bayes yang akan terlihat pada tingkat akurasi, precision, recall, dan F1-score. Selain itu, data yang dihasilkan diharapkan dapat memberikan landasan yang kuat untuk pengambilan keputusan yang lebih akurat dan strategis mengenai peningkatan fitur aplikasi dan perbaikan kualitas layanan publik yang diberikan. Dan analisis ini juga dapat memberikan wawasan yang lebih mendalam tentang faktor-faktor utama yang secara signifikan memengaruhi tingkat kepuasan masyarakat atau ketidakpuasan masyarakat terhadap suatu layanan tertentu.

2. TINJAUAN PUSTAKA

2.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Databases (KDD) merupakan proses sistematis yang digunakan dalam Data Mining untuk menemukan pola yang bermakna, tren terkini, dan hubungan tersembunyi dalam data besar sehingga sangat efektif saat digunakan untuk menganalisis data. Proses KDD mencakup beberapa tahapan utama, yaitu pemilihan data, pembersihan data, transformasi, proses Data Mining, dan interpretasi hasil. KDD sangat penting dalam mendukung pengambilan keputusan berbasis data di berbagai bidang, termasuk desain berkelanjutan dalam arsitektur dan konstruksi[14], [15].

2.2. Naïve Bayes

Naïve Bayes merupakan salah satu algoritma dalam data mining dan machine learning yang berdasarkan dari *Teorema Bayes*. Digunakan untuk klasifikasi data dengan pendekatan probabilistik, di mana setiap atribut tidak saling ketergantungan yang berarti atribut bernilai independen[16], [17]. Metode ini sering memberikan hasil yang akurat dalam berbagai pengaplikasian, seperti klasifikasi teks, deteksi spam, dan analisis sentimen. *Naïve Bayes* memiliki berbagai jenis yaitu *Gaussian*, *Multinomial*, *Bernoulli Bayes*. Meskipun dapat memberikan hasil yang akurat, algoritma *naive bayes* memiliki kelemahan yang cukup signifikan saat melakukan pemilihan atribut, karena naive bayes merupakan metode klasifikasi data mining yang menggunakan konsep statistik yang berdasarkan pada Teorema Bayes sehingga hanya dapat digunakan untuk memprediksi peluang keanggotaan pada suatu grup atau kelas[18].

2.3. Data Mining

Data mining merupakan bidang yang mencakup banyak disiplin ilmu, termasuk pembelajaran mesin, metode matematika, statistik, administrasi basis data, dan kecerdasan buatan. Banyak metode yang dapat digunakan untuk pemecahan masalah, analisis, perencanaan, diagnosis, deteksi, integrasi, pencegahan, pembelajaran, dan inovasi dalam teknologi ini. Kebutuhan akan metode canggih untuk menganalisis, memahami, dan menampilkan sejumlah besar data yang dikumpulkan dari aplikasi ilmiah dan bisnis adalah motivasi utama untuk pengembangan data mining. Oleh karena itu, data mining sangat penting dalam banyak bidang[19]. Pada data mining ada enam teknik yang dapat digunakan dalam menemukan sebuah hubungan, pola, kepingan informasi tersembunyi dalam data, teknik tersebut yaitu *prediction*, *classification*, *estimation*, *description*, *clustering* dan *association*[20]. Tahap awal yang sangat krusial dalam Data Mining adalah *preprocessing data*, dikarenakan data yang diperoleh dari berbagai sumber sering kali tidak terstruktur, mengandung nilai yang hilang, atau tidak konsisten, sehingga dapat memengaruhi akurasi model analisis[21].

2.3 Analisis Sentimen

Analisis sentimen merupakan suatu proses yang termasuk ke dalam *Natural Language Processing* (NLP) digunakan untuk memperoleh pendapat, perasaan atau emosi pengguna yang dilakukan secara sadar maupun tidak sadar oleh individu tersebut, hal tersebut biasanya disebabkan oleh berbagai alasan seperti ketidaksesuaian pada suatu topik, produk, layanan aspek tertentu atau keadaan diluar ekspektasi individu. Perkembangan teknologi saat ini sangat memudahkan dalam menganalisis sentimen publik dari data tidak terstruktur dengan menggunakan pendekatan *Machine Learning* (ML) dan *Deep Learning* (DL) dengan tingkat akurasi dan klasifikasi sentimen yang lebih baik. Dalam analisis sentimen dari perspektif granularitas dibagi menjadi 3 tingkatan yaitu *Document Sentiment Analysis* yang berfokus untuk menganalisis

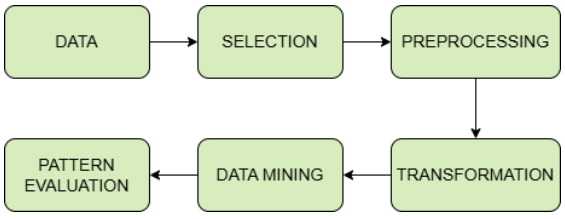
sentimen positif, netral dan negatif dari isi dokumen, *Sentence Sentiment Analysis* yaitu analisis sentimen berdasarkan per kalimat dan *Aspect Sentiment Analysis* menganalisis berdasarkan aspek-aspek tertentu[22], [23], [24], [25].

3. METODE PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian ini disusun secara sistematis dan terstruktur dari proses pengumpulan data yang relevan sampai dengan pengolahan data untuk mendukung tercapainya tujuan dari penelitian. Langkah awal dalam penelitian ini adalah melakukan scraping yaitu mengekstrak atau mengumpulkan data ulasan pengguna aplikasi Kita Bisa dari platform Google Play Store. Kemudian, dilakukan penyeleksian data yang akan digunakan dan proses preprocessing yaitu pembersihan data yang mencakup case folding, tokenization, stopword removal, stemming, cleaning. Setelah itu melakukan pelabelan dan transformasi data yang sudah bersih menjadi format numerik menggunakan TF-IDF. Jika data sudah menjadi format numerik dilakukan split data 30% testing dan 70% training kemudian pengklasifikasian data menggunakan metode Naïve Bayes. Terakhir dilakukan analisis berdasarkan hasil klasifikasi-an data, dengan tahapan yang terstruktur ini diharapkan menghasilkan data yang berkualitas dan analisis sentimen yang optimal.

3.2 Tahapan Pengolahan Data



Gambar 1. Knowledge Discovery in Database (*KDD*)

Penelitian ini menggunakan konsep *Knowledge Discovery in Database* (KDD) sebagai pedoman dalam pengolahan data. KDD terdiri dari beberapa tahapan yaitu *Selection*, *Preprocessing*, *Transformation*, *Data Mining* dan *Pattern Evaluation*. Sebelum melakukan tahapan KDD, terlebih dahulu peneliti akan menyiapkan data yang akan dilakukan pengolahan.

3.2.1 Data

Penelitian analisis sentimen ini dimulai dengan tahap pengumpulan data yang akan digunakan untuk analisis. Data yang akan digunakan dalam penelitian ini berasal dari ulasan atau komentar pengguna sebanyak 1000 data pada aplikasi Kita Bisa di Google Play Store. Data yang dikumpulkan yaitu berupa teks dari ulasan penggguna, rating pengguna yang mencerminkan opini masyarakat terhadap aplikasi Kita Bisa.

3.2.2 Selection

Setelah pengumpulan data, selanjutnya yaitu pemilihan atau seleksi data yang sesuai dengan kriteria penelitian. Tidak semua data yang dikumpulkan akan digunakan karena kemungkinan adanya data yang tidak relevan, tidak lengkap, atau duplikat. Data yang digunakan pada penelitian adalah data real-time dengan rentang waktu bulan januari 2024 sampai dengan bulan maret 2025, dengan tujuan agar mendapatkan pola yang terbaru dan aktual dalam melakukan analisis sentimen publik, karena data statis atau historis tidak relevan dan data cepat berubah berdasarkan waktu, kondisi tren sosial dan update pada aplikasi.

3.2.3 Preprocessing

Tahap preprocessing yaitu proses untuk meningkatkan kualitas data dengan melakukan pembersihan dan standarisasi teks agar siap dianalisis menggunakan teknik pemrosesan bahasa alami (*Natural Language Processing*). Langkah-langkah preprocessing dalam penelitian ini mencakup:

1. *Case Folding* yaitu melakukan perubahan pada semua kata menjadi huruf kecil untuk menyamakan format dari teks ulasan agar mudah dalam pemprosesan data.
2. *Cleaning Data* yaitu menghapus karakter khusus, angka, tanda baca, serta URL atau tautan yang tidak memiliki relevansi dalam analisis sentimen.
3. *Stopword Removal* yaitu menghilangkan kata umum atau penghubung yang tidak memiliki makna signifikan dalam analisis sentimen
4. *Stemming* yaitu mengubah setiap kata pada teks ulasan menjadi kata baku untuk menyederhanakan kata dengan makna yang sama

5. *Tokenization* yaitu memecah setiap kata yang ada pada teks ulasan menjadi kata-kata individual atau token untuk mempermudah analisis data.

Tahap processing ini sangat penting dalam pengolahan data dikarenanakan dapat meningkatkan akurasi model dalam menginterpretasikan sentimen pengguna dengan lebih baik.

3.2.4 Transformation

Setelah teks telah dibersihkan dan diproses, tahap berikutnya adalah melakukan palabelan pada data untuk mengetahui sentimen dari pengguna apakah termasuk positif, negatif atau netral berdasarkan dari rating pengguna. Kemudian, melakukan perubahan data teks menjadi representasi numerik agar dapat digunakan oleh algoritma machine learning. Pada penelitian ini menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) yang dimana metode ini mengubah teks menjadi representasi berbasis bobot berdasarkan frekuensi kemunculan kata dalam dokumen. Kata yang sering muncul dalam satu ulasan tetapi jarang muncul di seluruh dataset akan memiliki bobot lebih tinggi.

3.2.5 Data Mining

Tahap ini merupakan inti dari penelitian, di mana metode data mining atau machine learning diterapkan untuk mengklasifikasikan sentimen dari data ulasan pengguna yang telah diproses sebelumnya. Model yang digunakan dalam penelitian ini yaitu *Naïve Bayes*, sebelum eksekusi model yang telah dipilih, dilakukan pelatihan dan pengujian dengan membagi dataset menjadi data training dan testing untuk mengukur performa model.

3.2.6 Pattern Evaluation

Tahap akhir dalam penelitian ini adalah evaluasi pola atau hasil yang diperoleh dari analisis sentimen. Evaluasi ini bertujuan untuk mengukur akurasi dan efektivitas model yang digunakan. Beberapa metrik yang umum digunakan dalam evaluasi adalah:

1. *Acuracy*, untuk mengukur persentase prediksi yang benar dari keseluruhan data.
2. *Precision* dan *Recall*, presisi untuk mengukur seberapa banyak prediksi sentimen positif benar-benar positif. Dan recall mengukur seberapa banyak data positif yang berhasil diklasifikasikan dengan benar.
3. *F1-Score*, kombinasi antara presisi dan recall yang memberikan gambaran lebih seimbang terkait performa model.

Selain evaluasi kuantitatif, dilakukan juga analisis pola sentimen yang ditemukan. Misalnya, apakah mayoritas pengguna memberikan ulasan positif terhadap fitur tertentu dari aplikasi Kita Bisa, atau apakah ada keluhan yang sering muncul dalam ulasan negatif.

4. HASIL DAN PEMBAHASAN

4.1 Data

Dalam penelitian ini tahap pertama yang dilakukan adalah pengambilan atau mengekstrak data ulasan pengguna pada aplikasi Kita Bisa sebanyak 1000 data.

```
Jumlah data:  1000
```

Gambar 2. Jumlah data scraping

Pada gambar 3 merupakan hasil dari scraping data yang ditampilkan secara acak, data tersebut memiliki beberapa atribut penting seperti tanggal ulasan, nama pengguna, ulasan, dan rating dari pengguna. Selanjutnya data yang akan digunakan diseleksi terlebih dahulu berdasarkan rentang waktu.

57	2025-02-16 22:44:11	joko susilo	Mau buka aplikasi ya aja harus nunggu 2 jam ba...	1
547	2024-01-25 21:43:17	Bahrul Zaman	MasyaAllah terima kasih telah membantu barbagi...	5
689	2023-11-23 05:39:38	Putri Munalisa	Bagus aplikasi. Sayangnya ini nggak bisa balas...	5
108	2024-12-25 22:28:49	Suryadi Eko	Agak kecewa isi saldo di kita bisa padahal isi...	1
836	2023-09-12 10:40:34	Danang Pas	Tolong hapus data saya ,saya tidak pernah pinj...	1
838	2023-09-12 02:58:05	Yashy Grow	Keren dan manfaat buat semua	5
570	2024-01-13 03:28:20	Dini Marca	Untuk kitabisa, kami pasien berharap pencairan...	1
372	2024-05-16 06:48:42	rika walas	Luar biasa,, sangat memudahkan kita untuk berd...	5
423	2024-04-07 05:00:22	Rosna Bahar	Sangat bermanfaat & memudahkan kita dalam beri...	5
582	2024-01-08 13:54:20	indra kurniawan	Ngambil untung banyak banget,,, kemanusian dib...	1
814	2023-09-28 12:17:41	Nur Syafarudin	Ga bisa isi saldo ,lewat BSI . Sudah saya lapo...	2
971	2023-07-15 07:34:08	Anak Lanang	Oky	5
54	2025-02-18 23:04:17	Dimas Poetrapratama	tolong tambah audio di bacaan Al Qur'an	5
157	2024-11-10 06:43:44	Fajerry sangat ganteng sekali duwa	Jangan dikit² update	5
807	2023-10-02 03:20:12	Isti 6062	Kayaknya aplikasi pinjol ya	3
810	2023-09-30 22:56:38	Setia Rahmayanti	Alhamdulillah ada cara mudah berzakat dan bers...	5

Gambar 3. Hasil *scraping* data aplikasi Kita Bisa

4.2 Selection

Tahap selanjutnya seleksi untuk memilih dan hanya menggunakan data yang relevan dengan topik penelitian yaitu analisis sentimen. Seleksi yang dilakukan adalah rentang waktu pada data yang hanya menggunakan data dengan rentang waktu dari januari 2024 sampai dengan maret 2025, setelah melakukan seleksi data didapatkan jumlah data yaitu 591 dari 1000 data.

```
Jumlah data:  591
```

Gambar 3. Jumlah data seleksi

Pada gambar 4 terlihat bahwa terdapat data yang hilang dari hasil *data scraping*, dikarenakan rentang waktu yang tidak cocok berdasarkan seleksi. Dengan demikian data dapat dikatakan relevan untuk melakukan analisis sentimen.

57	2025-02-16 22:44:11	joko susilo	Mau buka aplikasi ya aja harus nunggu 2 jam ba...	1
547	2024-01-25 21:43:17	Bahrul Zaman	MasyaAllah terima kasih telah membantu barbagi...	5
108	2024-12-25 22:28:49	Suryadi Eko	Agak kecewa isi saldo di kita bisa padahal isi...	1
570	2024-01-13 03:28:20	Dini Marca	Untuk kitabisa, kami pasien berharap pencairan...	1
372	2024-05-16 06:48:42	rika walas	Luar biasa,, sangat memudahkan kita untuk berd...	5
423	2024-04-07 05:00:22	Rosna Bahar	Sangat bermanfaat & memudahkan kita dalam beri...	5
582	2024-01-08 13:54:20	indra kurniawan	Ngambil untung banyak banget,,, kemanusian dib...	1
54	2025-02-18 23:04:17	Dimas Poetrapratama	tolong tambah audio di bacaan Al Qur'an	5
157	2024-11-10 06:43:44	Fajerry sangat ganteng sekali duwa	Jangan dikit² update	5
256	2024-08-03 13:00:56	Reyes Giska	Aplikasi tabungan akherat	5
109	2024-12-25 00:06:14	Ranjo R	Bagus, semoga kedepannya bisa lebih baik lagi	5
47	2025-02-24 03:07:26	Dersiana Rambu Podu	Aplikasinya bagus dan sangat bermanfaat,slalu ...	5
244	2024-08-14 20:45:15	Tito Sidarta	Assalamualaikum kakak kakak yg di sana semoga ...	5
198	2024-09-28 10:02:25	M Rido	Nice Aplikasi	5
232	2024-08-26 09:10:19	aning pratikasari	aplikasi rekomen, bgs bgt. pelayananya juga ok...	5
278	2024-07-18 20:34:41	Ners Ashari	Terbaik	5

Gambar 4. Hasil seleksi data

4.3 Preprocessing

Selanjutnya merupakan tahap preprocessing atau pembersihan data yang dilakukan secara bertahap mulai dari *case folding* yaitu menyamakan format kalimat menjadi huruf kecil sampai dengan *stemming* yaitu merubah setiap kata menjadi kata baku. *Preprocessing* dilakukan dengan tujuan untuk meningkatkan kualitas data sebelum dianalisis.

4.3.1 Case Folding

Tahap pertama dalam preprocessing adalah case folding yaitu menyamakan semua kata dalam kalimat menjadi huruf kecil atau *lowercase*. Seperti yang terlihat pada gambar 5, contohnya yaitu "Mau" menjadi "mau" dan "Luar biasa" menjadi "luar biasa". Hal ini dilakukan supaya kata dengan penulisan berbeda yaitu besar atau kecil tapi berarti sama dapat dihitung menjadi satu kata.

57	Mau buka aplikasi ya aja harus nunggu 2 jam ba...	mau buka aplikasi ya aja harus nunggu 2 jam ba...
547	MasyaAllah terima kasih telah membantu barbagi...	masyaallah terima kasih telah membantu barbagi...
108	Agak kecewa isi saldo di kita bisa padahal isi...	agak kecewa isi saldo di kita bisa padahal isi...
570	Untuk kitabisa, kami pasien berharap pencairan...	untuk kitabisa, kami pasien berharap pencairan...
372	Luar biasa,, sangat memudahkan kita untuk berd...	luar biasa,, sangat memudahkan kita untuk berd...
423	Sangat bermanfaat & memudahkan kita dalam beri...	sangat bermanfaat & memudahkan kita dalam beri...
582	Ngambil untung banyak banget,,, kemanusian dib...	ngambil untung banyak banget,,, kemanusian dib...
54	tolong tambah audio di bacaan Al Qur'an	tolong tambah audio di bacaan al qur'an
157	Jangan dikit² update	jangan dikit² update
256	Aplikasi tabungan akherat	aplikasi tabungan akherat
109	Bagus, semoga kedepannya bisa lebih baik lagi	bagus, semoga kedepannya bisa lebih baik lagi
47	Aplikasinya bagus dan sangat bermanfaat,slalu ...	aplikasinya bagus dan sangat bermanfaat,slalu ...
244	Assalamualaikum kakak kakak yg di sana semoga ...	assalamualaikum kakak kakak yg di sana semoga ...
198	Nice Aplikasi	nice aplikasi
232	aplikasi rekomen, bgs bgt. pelayananya juga ok...	aplikasi rekomen, bgs bgt. pelayananya juga ok...
278	Terbaik	terbaik

Gambar 5. Hasil tahap Case Folding

4.3.2 Cleaning Data

Selanjutnya, dilakukan pembersihan data dengan menghilangkan karakter yang tidak diperlukan dalam proses tahapan selanjutnya, yaitu tanda baca, url, angka, emoji dan simbol tertentu.

57	mau buka aplikasi ya aja harus nunggu 2 jam ba...	mau buka aplikasi ya aja harus nunggu jam bar...
547	masyaallah terima kasih telah membantu barbagi...	masyaallah terima kasih telah membantu barbagi...
108	agak kecewa isi saldo di kita bisa padahal isi...	agak kecewa isi saldo di kita bisa padahal isi...
570	untuk kitabisa, kami pasien berharap pencairan...	untuk kitabisa kami pasien berharap pencairan ...
372	luar biasa,, sangat memudahkan kita untuk berd...	luar biasa sangat memudahkan kita untuk berdon...
423	sangat bermanfaat & memudahkan kita dalam beri...	sangat bermanfaat memudahkan kita dalam berinfaq
582	ngambil untung banyak banget,,, kemanusian dib...	ngambil untung banyak banget kemanusian dibiki...
54	tolong tambah audio di bacaan al qur'an	tolong tambah audio di bacaan al quran
157	jangan dikit² update	jangan dikit update
256	aplikasi tabungan akherat	aplikasi tabungan akherat
109	bagus, semoga kedepannya bisa lebih baik lagi	bagus semoga kedepannya bisa lebih baik lagi
47	aplikasinya bagus dan sangat bermanfaat,slalu ...	aplikasinya bagus dan sangat bermanfaatslalu j...
244	assalamualaikum kakak kakak yg di sana semoga ...	assalamualaikum kakak kakak yg di sana semoga ...
198	nice aplikasi	nice aplikasi
232	aplikasi rekomen, bgs bgt. pelayananya juga ok...	aplikasi rekomen bgs bgt pelayananya juga oke ...
278	terbaik	terbaik

Gambar 6. Hasil tahap Cleaning Data

4.3.3 Stopword Removal

Setelah mendapatkan data yang bersih selanjutnya, *stopword removal* akan diproses untuk penghilangan kata-kata umum yang tidak memiliki makna dalam analisis sentimen seperti 'ya', 'yg', 'agak', 'untuk','tolong','di', termasuk kata 'dan'. Hal tersebut dilakukan agar model lebih berfokus pada kata-kata yang memiliki makna penting.

57	mau buka aplikasi ya aja harus nunggu jam bar...	mau buka aplikasi aja nunggu jam baru kebuka
547	masyaallah terima kasih telah membantu barbagi...	masyaallah terima kasih membantu barbagi yg wa...
108	agak kecewa isi saldo di kita bisa padahal isi...	kecewa isi saldo kita padahal isi k anehnya ha...
570	untuk kitabisa kami pasien berharap pencairan ...	kitabisa pasien berharap pencairan dana tepat ...
372	luar biasa sangat memudahkan kita untuk berdon...	luar biasa sangat memudahkan berdonasi
423	sangat bermanfaat memudahkan kita dalam berinfaq	sangat bermanfaat memudahkan dalam berinfaq
582	ngambil untung banyak banget kemanusian dibiki...	ngambil untung banyak banget kemanusian dibiki...
54	tolong tambah audio di bacaan al quran	tambah audio bacaan al quran
157	jangan dikit update	jangan dikit update
256	aplikasi tabungan akherat	aplikasi tabungan akherat
109	bagus semoga kedepannya bisa lebih baik lagi	bagus semoga kedepannya lebih baik
47	aplikasinya bagus dan sangat bermanfaatslalu j...	aplikasinya bagus sangat bermanfaatslalu jadi ...
244	assalamualaikum kakak kakak yg di sana semoga ...	assalamualaikum kakak kakak yg sana semoga sem...
198	nice aplikasi	nice aplikasi
232	aplikasi rekomen bgs bgt pelayananya juga oke ...	aplikasi rekomen bgs bgt pelayananya oke bgt t...
278	terbaik	terbaik

Gambar 7. Hasil tahap Stopword Removal

4.3.4 Stemming

Tahap selanjutnya yaitu melakukan penyederhanaan pada kata yang memiliki imbuhan pada awalan, akhiran, sisipan maupun gabungan menjadi kata dasarnya. Tujuan melakukan tahap *stemming* adalah untuk menjadikan beberapa kata dengan makna yang sama ke dalam satu kata yaitu bentuk dasar dari kata tersebut. Contohnya terlihat pada gambar 9 yaitu 'membantu' menjadi 'bantu', 'berharap' menjadi 'harap', 'memudahkan' menjadi 'mudah'.

57	mau buka aplikasi aja nunggu jam baru kebuka	mau buka aplikasi aja nunggu jam baru buka
547	masyaallah terima kasih membantu barbagi yg wa...	masyaallah terima kasih bantu barbagi yg wajib...
108	kecewa isi saldo kita padahal isi k anehnya ha...	kecewa isi saldo kita padahal isi k aneh hari ...
570	kitabisa pasien berharap pencairan dana tepat ...	kitabisa pasien harap cair dana tepat waktu ka...
372	luar biasa sangat memudahkan berdonasi	luar biasa sangat mudah donasi
423	sangat bermanfaat memudahkan dalam berinfaq	sangat manfaat mudah dalam berinfaq
582	ngambil untung banyak banget kemanusian dibiki...	ngambil untung banyak banget kemanusian bikin ...
54	tambah audio bacaan al quran	tambah audio baca al quran
157	jangan dikit update	jangan dikit update
256	aplikasi tabungan akherat	aplikasi tabung akherat
109	bagus semoga kedepannya lebih baik	bagus moga depan lebih baik
47	aplikasinya bagus sangat bermanfaatslalu jadi ...	aplikasi bagus sangat bermanfaatslalu jadi jem...
244	assalamualaikum kakak kakak yg sana semoga sem...	assalamualaikum kakak kakak yg sana moga semua...
198	nice aplikasi	nice aplikasi
232	aplikasi rekomen bgs bgt pelayananya oke bgt t...	aplikasi rekomen bgs bgt pelayananya oke bgt t...
278	terbaik	baik

Gambar 8. Hasil tahap Stemming

4.3.5 Tokenization

Tokenization merupakan tahapan dalam memotong atau memecah kata yang ada pada kalimat menjadi token dalam bentuk potongan dari kata per kata. Gambar 8 merupakan hasil yang memperlihatkan bahwa setiap kata pada kalimat dibuat menjadi potongan-potongan kata. Pada tahap ini sering dilakukan untuk membantu dalam proses selanjutnya yaitu menghitung frekuensi data.

57	mau buka aplikasi aja nunggu jam baru buka	[mau, buka, aplikasi, aja, nunggu, jam, baru, ...
547	masyaallah terima kasih bantu barbagi yg wajib...	[masyaallah, terima, kasih, bantu, barbagi, yg...
108	kecewa isi saldo kita padahal isi k aneh hari ...	[kecewa, isi, saldo, kita, padahal, isi, k, an...
570	kitabisa pasien harap cair dana tepat waktu ka...	[kitabisa, pasien, harap, cair, dana, tepat, w...
372	luar biasa sangat mudah donasi	[luar, biasa, sangat, mudah, donasi]
423	sangat manfaat mudah dalam berinfaq	[sangat, manfaat, mudah, dalam, berinfaq]
582	ngambil untung banyak banget kemanusian bikin ...	[ngambil, untung, banyak, banget, kemanusian, ...
54	tambah audio baca al quran	[tambah, audio, baca, al, quran]
157	jangan dikit update	[jangan, dikit, update]
256	aplikasi tabung akherat	[aplikasi, tabung, akherat]
109	bagus moga depan lebih baik	[bagus, moga, depan, lebih, baik]
47	aplikasi bagus sangat bermanfaatslalu jadi jem...	[aplikasi, bagus, sangat, bermanfaatslalu, jad...
244	assalamualaikum kakak kakak yg sana moga semua...	[assalamualaikum, kakak, kakak, yg, sana, moga...
198	nice aplikasi	[nice, aplikasi]
232	aplikasi rekomen bgs bgt pelayananya oke bgt t...	[aplikasi, rekomen, bgs, bgt, pelayananya, oke...
278	baik	[baik]

Gambar 9. Hasil tahap Tokenization

4.4 Transformation

4.4.1 Pelabelan

Setelah mendapatkan data yang bersih selanjutnya dilakukan pengklasifikasian ulasan pengguna ke dalam 3 kategori yaitu positif, netral dan negatif berdasarkan rating dari pengguna, ulasan dengan rating diatas 4 akan dimasukkan ke dalam kategori positif, rating 3 menjadi netral dan sisa data dengan rating dibawah 3 masuk ke kategori negatif. Berdasarkan gambar 10 menunjukkan bahwa kategori positif bernilai lebih banyak yaitu 456, dibandingkan negatif dengan jumlah 111.

```
label
Positif    456
Negatif    111
Netral      24
Name: count, dtype: int64
```

Gambar 10. Jumlah ulasan berdasarkan label

4.4.2 TF-IDF

Setelah melalui proses pelabelan, proses selanjutnya yaitu TF-IDF. Pada proses ini data pada teks nantinya akan diubah menjadi bentuk numerik digunakan untuk mengukur bobot dari beberapa kata dalam ulasan berdasarkan frekuensi kemunculan kata tersebut. Semakin sering kata tersebut muncul maka semakin tinggi nilai dari TF-IDF, tetapi pada penelitian ini dilakukan pembatasan terhadap nilai dari TF-IDF hanya sebanyak 500 kata yang memiliki frekuensi terbanyak dengan tujuan membatasi jumlah kata yang akan dihitung pada tahap selanjutnya supaya model menjadi lebih efektif. Selain itu, dengan membatasi dapat mengurangi redundan kata-kata umum yang tidak informatif dan relevan dengan penelitian seperti kata yang typo, nama seseorang, atau nama tempat.

```
[45] from sklearn.feature_extraction.text import TfidfVectorizer

     tfidf = TfidfVectorizer(max_features=500)
     tfidf_matrix = tfidf.fit_transform(df_filtered['tokens']).toarray()

     print("TF-IDF Matrix Shape:", tfidf_matrix.shape)

     TF-IDF Matrix Shape: (591, 500)
```

Gambar 11. Code TF-IDF pada Data

4.5 Data Mining

Sebelum melanjutkan ke tahap data mining dengan metode *Naïve Bayes*, data akan dibagi terlebih dahulu menjadi data *training* sebesar 70% dan data *testing* sebesar 30%. Pada gambar 12 data training didapatkan dengan jumlah 413, dan data testing sebanyak 178.

```
from sklearn.model_selection import train_test_split

X = tfidf_matrix
y = df_filtered['label']

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, random_state=42, stratify=y
)

print("Jumlah data training:", len(X_train))
print("Jumlah data testing:", len(X_test))
```

```
Jumlah data training: 413
Jumlah data testing: 178
```

Gambar 12. Code Split Data

Gambar 13 merupakan potongan code dengan menerapkan algoritma *Multinomial Naïve Bayes* untuk mengklasifikasi data yang sudah melakukan pembagian menjadi training dan testing menjadi 3 kategori yaitu positif, netral dan negatif.

```
[17] from sklearn.naive_bayes import MultinomialNB
     from sklearn.metrics import classification_report, accuracy_score, precision_score, recall_score, f1_score, c

     model = MultinomialNB()
     model.fit(X_train, y_train)
     y_pred = model.predict(X_test)
```

Gambar 13. Code Naïve Bayes

4.6 Pattern Evaluation

Penggunaan metode naïve bayes dalam menganalisis sentimen dari segi performa menunjukkan hasil yang baik dengan tingkat akurasi sebesar 81% dari data yang diuji, diikuti dengan nilai precision atau prediksi benar yaitu 79%, recall atau total data yang seharusnya benar bernilai 81% dan f1 score bernilai 76%. Pada confusion matrix kategori negatif terlihat 8 benar dan 26 salah , netral sebanyak 7 benar, sedangkan positif 136 benar, dan 1 dianggap salah, makna salah pada *Confusion Matrix* berarti terdapat ulasan yang labelnya berbeda saat sebelum dan sesudah diproses, contoh negatif 8 benar tetapi 26 salah karena diprediksi sebagai positif.

```
MultinomialNB Accuracy: 80.90%
MultinomialNB Precision: 0.79
MultinomialNB Recall: 0.81
MultinomialNB f1_score: 0.76

Confusion Matrix:
 [[  8   0  26]
 [  0   0   7]
 [  1   0 136]]
========================================================================
              precision    recall  f1-score   support

     Negatif       0.89      0.24      0.37        34
      Netral       0.00      0.00      0.00         7
     Positif       0.80      0.99      0.89       137

    accuracy                           0.81       178
   macro avg       0.56      0.41      0.42       178
weighted avg       0.79      0.81      0.76       178
```

Gambar 14. Hasil Inteprestasi Data setelah Tahap Data Mining

Berdasarkan gambar 14 menunjukkan bahwa aplikasi Kita Bisa setelah dilakukan analisis sentimen memperlihatkan hasil yang didominasi mengarah ke kategori positif dengan hasil *Precision* sebesar 80% didukung oleh *Recall* sebesar 99% dan *F1-Score* sebesar 89% dengan data sebanyak 137, sedangkan negatif memiliki hasil *Precision* yang tinggi yaitu 89% tetapi hasil *Recall* hanya 24% dikarenakan banyak data negatif yang dianggap positif. Dan netral tidak menunjukkan hasil apapun, kemungkinan karena didukung oleh data hanya sebanyak 7. Dan untuk macro avg didapat hasil sebesar 42% dikarenakan kelas netral yang buruk dan weighted avg bernilai 76% disebabkan oleh kelas positif yang terlalu dominan.

5. KESIMPULAN

Berdasarkan hasil analisis dari 1000 data ulasan aplikasi Kita Bisa yang diambil dari platform Google Play Store dan diseleksi hanya menggunakan 591 data ulasan, menunjukkan hasil yang positif masih sesuai dengan rating dari goole play store yaitu 4.5. Dan model *Naïve Bayes* dari 178 data testing yang digunakan menunjukkan hasil yang cukup baik dengan akurasi sebesar 81% dengan tingkat pengenalan ulasan kelas positif sebesar 99%.

Namun demikian, tetap ditemukan ulasan negatif dari pengguna meskipun jumlahnya tidak terlalu banyak tetapi dengan adanya ulasan tersebut menunjukkan bahwa terdapat masalah dan ketidakpuasan pengguna terhadap fitur atau alur kerja dari aplikasi tersebut. Hasil yang ditunjukkan oleh *Confusion Matrix* pada kelas negatif dari 34 data, hanya 8 yang diprediksi sebagai negatif sedangkan sisanya tidak diprediksi negatif oleh model.

Sentimen negatif ini disebabkan oleh fitur yang kurang optimal pada aplikasi, dan terjadinya bug seperti setelah top up tidak ada penambahan pada kantong donasimu berdasarkan ulasan pengguna. Dengan demikian, pengembang harus selalu melakukan evaluasi dan perbaikan berdasarkan aspek-aspek yang sering dikeluhkan oleh pengguna, walaupun saat ini hasil analisis menunjukkan sentimen positif yang terbanyak. Dan untuk penelitian selanjutnya disarankan untuk menambah jumlah data dan memperpanjang rentang waktu untuk mendapatkan hasil yang

lebih menggambarkan ulasan dari pengguna. Selain itu penelitian selanjutnya juga dapat melakukan analisis dengan metode lain dari *Deep Learning* atau *Machine Learning* dan dapat menganalisis faktor-faktor penyebab banyaknya ulasan negatif untuk memberikan saran perbaikan pada aplikasi.

UCAPAN TERIMA KASIH

Terimakasih dari penulis kepada pihak-pihak yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

[1] W. Cao, A. A. Kadir, W. Tang, J. Wang, J. Yuan, and I. I. Hassan, "Effectiveness of mobile application interventions for stroke survivors: systematic review and meta-analysis," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, pp. 1–22, 2024, doi: 10.1186/s12911-023-02391-1.

[2] F. Xin, H. Yu, X. Dong, and C. Chen, "When the customers comes to you: mobile apps and corporate investment efficiency," *Humanit. Soc. Sci. Commun.*, vol. 12, no. 1, pp. 1–18, 2025, doi: 10.1057/s41599-025-04625-1.

[3] M. Y. T. Martha and I. N. Tri Anindia Putra, "Penerapan Pembelajaran Secara Daring Aplikasi Google Classroom Di Smptk Galang Kasih," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 1, pp. 97–105, 2022, doi: 10.29100/jipi.v7i1.2414.

[4] I. Larasati, A. N. Yusril, and P. Al Zukri, "Systematic Literature Review Analisis Metode Agile Dalam Pengembangan Aplikasi Mobile," *Sistemasi*, vol. 10, no. 2, p. 369, 2021, doi: 10.32520/stmsi.v10i2.1237.

[5] S. Ziha Fidela, M. Putri Azizah, and S. Rizka Hidayah, "Tren Pengembangan Aplikasi Mobile: Sebuah Tinjauan Literatur," *J. Tek. Mesin, Ind. Elektro Dan Inform.*, vol. 2, no. 4, pp. 30–48, 2023, [Online]. Available: https://doi.org/10.55606/jtmei.v2i4.2848

[6] I Gusti Agung Putu Bagus Satria Wicaksana, Agus Aan Jiwa Permana, and Ni Putu Novita Puspa Dewi, "Pengumpulan Data Tweet Berdasarkan Kata Kunci Depresi Dan Kisah Hidup Di Kalangan Mahasiswa Berbasis Phq-9," *J. Pendidik. Teknol. dan Kejuru.*, vol. 21, no. 1, pp. 24–33, 2024, doi: 10.23887/jptkundiksha.v21i1.66460.

[7] N. W. Wardani, P. G. S. C. Nugraha, and G. S. Mahendra, "Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 3, pp. 656–668, 2024, doi: 10.23887/jstundiksha.v12i3.38605.

[8] D. D. Hariyanti, G. A. Pradnyana, and I. G. M. Darmawiguna, "Kombinasi Metode Naive Bayes dan K-Medoid dalam Memprediksi Penjurusan Siswa di Sekolah Menengah Atas," *J. Ilmu Komput.*, vol. 14, no. 2, p. 88, 2021, doi: 10.24843/jik.2021.v14.i02.p03.

[9] B. I. Nugroho, Z. Ma'arif, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Metode Klasifikasi Untuk Menganalisa Penyalahgunaan Sosial Media," *J. Sist. Inf. dan …*, vol. 3, no. 2, pp. 46–51, 2022, [Online]. Available: http://journal.peradaban.ac.id/index.php/jsitp/article/download/1265/860

[10] I. B. G. Sarasvananda, D. Selivan, M. L. Radhitya, and I. N. T. A. Putra, "Analisis Sentimen Pada Pembelajaran Daring Di Indonesia Melalui Twitter Menggunakan Naïve Bayes Classifier," *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 2, pp. 227–233, 2022, doi: 10.31598/sintechjournal.v5i2.1241.

[11] G. A. Pradnyana and I. G. M. Darmawiguna, "Web-Based System for Bali Tourism Sentiment Analysis during The Covid-19 Pandemic using Django Web Framework and Naive Bayes Method," *Proc. 4th Int. Conf. Innov. Res. Across Discip. (ICIRAD 2021)*, vol. 613, no. Icirad, pp. 316–320, 2022, doi: 10.2991/assehr.k.211222.050.

[12] D. Prasetia, N. Rahaningsih, R. D. Dana, and C. L. Rohmat, "ANALISIS SENTIMEN PENGGUNA APLIKASI MYBLUEBIRD DENGAN ALGORITMA NAÏVE BAYES DI PLAYSTORE," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, pp. 602–608, 2025.

[13] R. Rahmadani, A. Rahim, and Rudiman, "Analisis Sentimen Ulasan 'Ojol the Game' Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2928–2936, 2024, doi: 10.23960/jitet.v12i3.4988.

[14] C. Llatas, B. Soust-Verdaguer, L. C. Torres, and D. Cagigas, "Application of Knowledge Discovery in Databases (KDD) to environmental, economic, and social indicators used in BIM workflow to support sustainable design," *J. Build. Eng.*, vol. 91, no. May, p. 109546, 2024, doi: 10.1016/j.jobe.2024.109546.

[15] A. S. Llige and C. J. T. Escuadra, "Mapping the Terrain: A Systematic Review of Drama Research Using Data Mining," *SAGE Open*, vol. 14, no. 3, pp. 1–18, 2024, doi: 10.1177/21582440241285341.

[16] Komang Aditya Pratama, Gede Aditra Pradnyana, and I Ketut Resika Arthana, "Pengembangan Sistem Cerdas Untuk Prediksi Daftar Kembali Mahasiswa Baru Dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha)," *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 1, pp. 22–34, 2020, doi: 10.31598/sintechjournal.v3i1.523.

[17] Y. J. Lee, S. H. O, and J. E. Eck, "Improving Recidivism Forecasting With a Relaxed Naïve Bayes Classifier," *Crime Delinq.*, vol. 71, no. 1, pp. 89–117, 2023, doi: 10.1177/00111287231186093.

[18] I. M. A. A. D. Putra, I. M. G. Sunarya, and I. G. A. Gunadi, "Perbandingan Algoritma Naive Bayes Berbasis Feature Selection Gain Ratio dengan Naive Bayes Kovensional dalam Prediksi Komplikasi Hipertensi," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 6, no. 1, pp. 37–49, 2024, doi: 10.35746/jtim.v6i1.488.

[19] C. Sicakyuz, S. A. Edalatpanah, and D. Pamucar, "Data mining applications in risk research : A systematic literature review," *Int. J. Knowledge-Based Intell. Eng. Syst.*, pp. 1–40, 2024, doi: 10.1177/13272314241296866.

[20] L. P. Wiwien Widhyastuti, I. N. Sukajaya, and K. Y. Ernanda Aryanto, "Customer Profiling berdasarkan Model RFM dengan Metode K-Means pada Institusi Pendidikan untuk menunjang Strategi," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 2, pp. 94–108, 2022.

[21] R. Shetty, M. Geetha, U. Dinesh Acharya, and G. Shyamala, "Enhancing Ovarian Tumor Dataset Analysis through Data Mining Preprocessing Techniques," *IEEE Access*, vol. 12, no. August, pp. 122300–122312, 2024, doi: 10.1109/ACCESS.2024.3450520.

[22] W. A. Awadh, R. B. Sulaiman, and M. A. Mahmoud, *Aspect-based sentiment analysis in MOOCs: a systematic literature review introducing the MASC-MEF framework*, vol. 37, no. 1. 2025. doi: 10.1007/s44443-025-00018-1.

[23] L. Davoodi, J. Mezei, and M. Heikkilä, *Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms*, no. 0123456789. Springer US, 2025. doi: 10.1007/s10660-025-09948-4.

[24] M. Maghsoudi, M. Bakhtiari, and H. Bakhtiari, "Global Perspectives on Laptop Features: A Sentiment Analysis of User Preferences in Developed and Developing Countries," *IEEE Access*, vol. 13, no. November 2024, pp. 9102–9119, 2025, doi: 10.1109/ACCESS.2025.3528324.

[25] A. Albladi, M. Islam, and C. Seals, "Sentiment Analysis of Twitter data using NLP Models: A Comprehensive Review," *IEEE Access*, vol. 13, no. February, pp. 30444–30468, 2025, doi: 10.1109/ACCESS.2025.3541494.