

## PENGUNAAN ALGORITMA NAÏVE BAYES DALAM MENGANALISIS SENTIMEN ULASAN APLIKASI ADAKAMI DI GOOGLE PLAY STORE

Evianti, Bambang Irawan, Agus Bahtiar

Program Studi Teknik Informatika, STMIK IKMI Cirebon

Jalan Perjuangan No.10 B, Kota Cirebon

nengvhiant@gmail.com

### ABSTRAK

Aplikasi Adakami adalah *platform* pinjaman *online* yang menyediakan layanan pinjaman tunai kepada penggunanya. Aplikasi ini sudah diunduh oleh lebih dari 10 juta pengguna di *Google Play Store*. Permasalahan penelitian ini berfokus pada sejauh mana Algoritma *Naïve Bayes* dapat berhasil mengenali dan membedakan sentimen positif dan negatif dalam ulasan pengguna. Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma *Naïve Bayes* dalam mengevaluasi sentimen pengguna yang terungkap dalam ulasan untuk aplikasi Adakami di *Google Play Store*. Dalam konteks penelitian ini, analisis sentimen terhadap ulasan aplikasi Adakami di *Google Play Store* akan dilakukan dengan menggunakan Algoritma *Naïve Bayes*. Data ulasan aplikasi Adakami di *Google Play Store* akan dikumpulkan melalui teknik *web scraping*. Data yang terkumpul terdiri dari ulasan yang ditulis dalam bahasa Indonesia. Data tersebut akan dianalisis menggunakan *Rapidminer*, termasuk pembersihan data, *tokenisasi*, penghapusan kata-kata umum. Model *Naïve Bayes* yang telah disusun akan diuji menggunakan data uji. Hasil penelitian ini menunjukkan bahwa aplikasi Adakami mendapatkan respons positif dari penggunanya, yang tercermin dari dominasi sentimen positif dalam ulasan. Hasil evaluasi menunjukkan bahwa algoritma *Naïve Bayes* berhasil mencapai tingkat akurasi sebesar 84,55%, dengan nilai presisi mencapai 83%, dan nilai *recall* mencapai 86%. Algoritma *Naïve Bayes* pada aplikasi Adakami telah mencapai tingkat kinerja yang sangat memuaskan.

**Kata Kunci:** Analisis sentimen, Aplikasi Adakami, Algoritma *Naïve Bayes*

### 1. PENDAHULUAN

Dalam beberapa tahun terakhir, kemajuan pesat di bidang Informatika telah secara mendasar mengubah cara kita berinteraksi dengan dunia sekitar. Perkembangan teknologi informasi telah menjadi pendorong utama dalam memajukan berbagai aspek kehidupan, memengaruhi sektor-sektor kunci seperti teknologi, bisnis, dan pendidikan. Secara khusus, penetrasi teknologi informasi ke dalam kehidupan sehari-hari telah membawa perubahan signifikan dalam cara kita mengakses informasi, berkomunikasi, dan menjalankan kegiatan sehari-hari. Di era kontemporer ini, ungkapan opini dan sentimen masyarakat dapat lebih mudah disebarkan dan diungkapkan secara bebas, termasuk melalui berbagai platform media seperti *Google Play Store*. Fenomena ini memberikan peluang penting bagi perusahaan yang ingin mengumpulkan umpan balik terkait aplikasi yang mereka tawarkan. Pemanfaatan aplikasi semakin mencerminkan pola perilaku masyarakat dalam merangkul kemajuan teknologi informasi di berbagai dimensi kehidupan, termasuk dalam bidang hiburan, dunia kerja, dan ekspresi kreatif[1].

Tujuan utama dari penelitian ini adalah menerapkan pendekatan analisis sentimen menggunakan metode *Naive Bayes* pada ulasan aplikasi AdaKami, sebuah layanan pinjaman uang digital yang tersedia di *Google Play*. Fokus penelitian ini adalah mengklasifikasikan ulasan pengguna menjadi kategori positif dan negatif, dengan penekanan khusus pada evaluasi kinerja algoritma dalam konteks aplikasi keuangan tertentu. Penelitian ini tidak hanya bertujuan untuk memahami persepsi

pengguna terhadap aplikasi AdaKami, tetapi juga bertujuan untuk berkontribusi pada pengembangan analisis sentimen dalam bidang ilmu komputer, memfasilitasi pengguna dalam meminjam uang secara online. Pentingnya penelitian ini terletak pada kontribusinya untuk memahami respons pengguna terhadap layanan keuangan digital, terutama dalam konteks aplikasi peminjaman uang. Hasil penelitian ini dapat memberikan manfaat praktis dengan memberikan pedoman bagi pengembang aplikasi untuk meningkatkan kualitas layanan mereka berdasarkan masukan pengguna. Selain itu, penelitian ini memberikan dukungan informasi kepada calon pengguna dan membantu mereka membuat keputusan yang terinformasi mengenai penggunaan aplikasi AdaKami.

Hasil evaluasi pemodelan algoritma *Naive Bayes* mencapai akurasi sebesar 84,55%, presisi sebesar 83%, dan *recall* sebesar 86%. Akurasi kelas yang diprediksi sebagai negatif mencapai 86,13%, dengan 276 *true negatives (TN)* dan 1714 *true positives (TP)*. Namun, akurasi kelas yang diprediksi sebagai positif mencapai 82,96% dengan *true negatives (TN)* sebanyak 1651 dan *true positives (TP)* sebanyak 339. Dari hasil ini, dapat disimpulkan bahwa klasifikasi *Naive Bayes* merupakan pilihan yang efektif untuk analisis data yang baik.

### 2. TINJAUAN PUSTAKA

#### 2.1. Analisis Sentimen

Analisis sentimen adalah langkah otomatis yang memahami, mengekstrak, dan memproses data informasi teks dengan tujuan untuk memperoleh

informasi tentang emosi yang terkandung dalam kalimat[2]. Tujuan utama dari analisis sentimen adalah untuk menilai apakah teks atau data mengandung sentimen positif, negatif, atau netral. Proses analisis sentimen melibatkan penerapan teknik pemrosesan bahasa alami, pembelajaran mesin, dan algoritma klasifikasi untuk mendeteksi serta mengklasifikasikan ekspresi emosional dan penilaian dalam teks seperti ulasan pelanggan, tweet, artikel berita, dan pos blog. Analisis sentimen telah memperluas cakupan penelitian *Pemrosesan Bahasa Alami (NLP)*, karena hal ini menimbulkan banyak pertanyaan penelitian yang sulit yang sebelumnya belum diteliti.[3].

## 2.2. Scraping Data

*Data scraping*, juga dikenal sebagai *web scraping*, adalah metode otomatis untuk mengambil informasi dari situs web. Teknik data *scraping* umumnya digunakan untuk mengumpulkan sejumlah besar data dari berbagai sumber *online* seperti situs *e-commerce*, situs berita, dan platform media sosial. Dalam proses data *scraping* ini, para peneliti mengekstrak berbagai informasi relevan seperti ulasan dan peringkat, serta menyimpan data dalam format file CSV[2]. Secara umum, data komentar yang dihasilkan dari proses *scraping* terdiri dari data asli yang ada pada elemen *XML* atau *HTML* di halaman-halaman website[4].

## 2.3. RapidMiner

*RapidMiner* adalah platform perangkat lunak sumber terbuka yang digunakan untuk analisis data, pemrosesan data, dan pembangunan model prediktif. Keuntamaan utama dari *RapidMiner* adalah kemampuannya untuk membangun model prediktif menggunakan berbagai algoritma pembelajaran mesin.

## 2.4. Aplikasi Shopee

*Shopee* adalah sebuah aplikasi yang sering digunakan untuk melakukan proses transaksi secara online yang dapat digunakan melalui *smartphone*[5]. Sebagai aplikasi *e-commerce*, *Shopee* memungkinkan pengguna untuk mencari, membeli, dan menjual berbagai produk secara digital. Pengguna dapat menemukan berbagai jenis produk di platform ini, termasuk pakaian, elektronik, barang rumah tangga, dan lain sebagainya, yang dijual oleh penjual yang terdaftar.

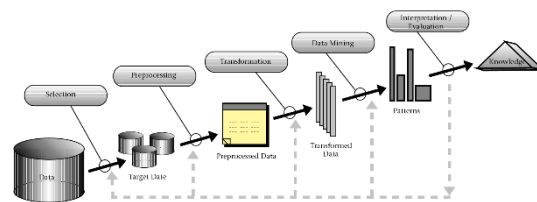
## 2.5. Naïve Bayes

Algoritma *Naïve Bayes* adalah algoritma klasifikasi *probabilitas* yang didasarkan pada teorema Bayes. Klasifikasi *Naïve Bayes* dapat diinterpretasikan dengan cara bahwa nilai atribut kelas tidak dipengaruhi atau tidak mempengaruhi nilai atribut lainnya[6]. Ciri utama dari klasifikasi *Naïve Bayes* adalah adanya asumsi kemandirian yang sangat kuat[7]. Dalam penelitian ini, menggunakan algoritma *Naïve Bayes* untuk mengelompokkan ulasan aplikasi Adagami di *Google Play Store* menjadi dua kelas:

positif dan negatif. Kelas positif menunjukkan bahwa ulasan mencerminkan sentimen positif, sedangkan kelas negatif menunjukkan bahwa ulasan mengandung sentimen negatif.

## 3. METODE PENELITIAN

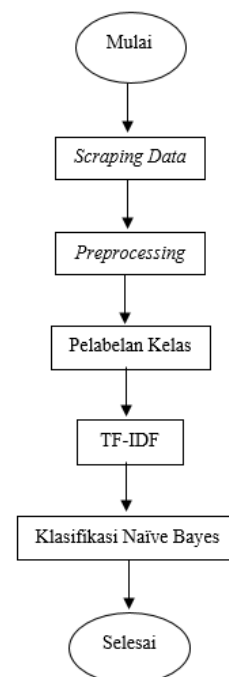
Proses penemuan pengetahuan dalam basis data (*Knowledge Discovery in Databases/KDD*) diterapkan secara bertahap untuk menganalisis data ulasan aplikasi AdaKami-Digital Money Loans di *Google Play Store*. Tahapan ini mencakup serangkaian langkah seperti pemilihan data, *pra-pemrosesan*, *transformasi*, penambangan data, *evaluasi*, dan *Knowledge*:



Gambar 1. Metode Penelitian

KDD adalah singkatan dari Knowledge Discovery in Database. Ini adalah proses mengidentifikasi pola yang berguna, pengetahuan baru, dan informasi berguna dari data yang besar dan kompleks. Metode KDD melibatkan serangkaian langkah atau proses untuk mencapai tujuan tersebut.

Berikut adalah langkah untuk melakukan klasifikasi menggunakan Algoritma *Naïve Bayes* melalui serangkaian *Scraping Data*, *Preprocessing*, Pelabelan, *TF-IDF*, dan melakukan Klasifikasi menggunakan *Naïve Bayes*.



Gambar 2 Perancangan

### 3.1. Selection Data

Informasi mengenai ulasan aplikasi AdaKami di *Google Play Store* dikumpulkan pada tanggal 30 November 2023, mencakup dataset sebanyak 3000 ulasan. Setelah itu, dilakukan proses seleksi data dengan penekanan pada atribut target. Proses ini melibatkan pengenalan pola atau wawasan dalam atribut data, yang kemudian disimpan dalam file terpisah dari basis data.

### 3.2. Preprocessing

*Preprocessing* teks adalah proses mengubah data teks yang tidak terstruktur menjadi format terstruktur sesuai kebutuhan untuk mengekstrak informasi selama proses penambangan. Teks yang telah diproses disebut sebagai istilah, dan setiap istilah dalam dokumen dapat memiliki berbagai makna, termasuk beberapa yang tidak relevan. Menghilangkan istilah yang tidak relevan dari dokumen merupakan bagian dari proses tersebut, sementara istilah yang bermakna tetap ada dan berfungsi sebagai representasi dokumen[8]. Menghapus kata bertujuan untuk menghilangkan istilah-istilah yang tidak perlu, sehingga mengurangi dimensi dataset[9].

Tahapan dari *pra-pemrosesan* mencakup *Tokenize*, *Transform Cases*, *Filter Tokens (by Length)*, *Filter stopwords*, dan *SMOTE Upsampling*. Penjelasan untuk masing-masing tahapannya adalah sebagai berikut:

#### 3.2.1. Tokenize

Dalam *RapidMiner*, proses tokenisasi, yang dikenal sebagai operator "*Tokenize*," adalah tahap dalam *pra-pemrosesan* teks di mana teks dipisahkan menjadi token-token individual, biasanya kata-kata atau frasa-frasa. Operator *Tokenize* memungkinkan pemrosesan teks lebih lanjut dengan memecah teks menjadi elemen-elemen yang dapat diolah lebih lanjut, seperti mendeteksi frekuensi kata, menganalisis pola kata, atau membangun model teks lebih lanjut. Dalam konteks *RapidMiner*, pengguna dapat mengonfigurasi operator *Tokenize* untuk menyesuaikan cara pemisahan teks, misalnya, dengan mempertimbangkan spasi, tanda baca, atau aturan khusus lainnya. Setelah proses tokenisasi, data teks dapat diolah lebih lanjut dalam alur kerja analisis data menggunakan berbagai operator lainnya untuk memahami dan mendapatkan wawasan dari informasi teks yang dihasilkan.

#### 3.2.2. Transform Cases

*Transform Cases* adalah langkah dalam pemrosesan teks di mana huruf teks diubah menjadi format tertentu. Ini dapat melibatkan perubahan teks menjadi huruf besar, huruf kecil, atau mengapit huruf pertama setiap kata dengan huruf besar (*title case*). Dalam konteks *RapidMiner* atau *platform* analisis data serupa, *Transform Cases* digunakan untuk menciptakan konsistensi dalam representasi huruf dalam teks. Ini berkontribusi pada analisis data teks dengan menangani variasi huruf (seperti huruf besar dan huruf kecil) secara seragam.

#### 3.2.3. Filter Tokens (by Length)

*Filter Tokens (by Length)* adalah suatu langkah dalam pemrosesan teks yang melibatkan penghapusan token (kata atau frasa) berdasarkan panjangnya. Proses ini memungkinkan pengguna untuk menyaring token berdasarkan jumlah karakter atau panjang kata tertentu. Dalam konteks *RapidMiner* atau platform analisis data serupa, *Filter Tokens (by Length)* dapat digunakan untuk membersihkan data teks dengan mengeliminasi token yang mungkin kurang relevan atau tidak signifikan dalam analisis. Pengguna dapat menyesuaikan kriteria panjang token yang diinginkan untuk menjaga akurasi dan relevansi dalam representasi teks yang dihasilkan. Secara umum, *Filter Tokens (by Length)* membantu menyaring token yang dapat mempengaruhi hasil analisis, sehingga meningkatkan kualitas pemrosesan teks dan wawasan yang dihasilkan.

#### 3.2.4. Filter stopwords

*Filter Stopwords* adalah teknik pemrosesan teks yang melibatkan penghapusan kata-kata isapan atau "*stopwords*" dari suatu teks. *Stopwords* adalah kata-kata umum yang sering muncul dalam suatu bahasa dan umumnya dianggap tidak memiliki makna yang signifikan dalam analisis teks. Contoh *stopwords* dalam bahasa Inggris mencakup "*the*," "*and*," "*is*," dan sebagainya. Tujuan utama dari penerapan *Filter Stopwords* adalah membersihkan teks dengan menghilangkan kata-kata yang cenderung memberikan kontribusi yang minim terhadap pemahaman atau analisis teks. Dengan menghapus *stopwords*, fokus analisis dapat lebih terarah pada kata-kata yang memiliki makna lebih mendalam dan memberikan wawasan yang lebih signifikan.

#### 3.2.5. SMOTE Upsampling

*SMOTE (Synthetic Minority Over-sampling Technique) Upsampling* adalah metode dalam pemrosesan data yang digunakan untuk mengatasi ketidakseimbangan kelas dalam sebuah dataset. Teknik ini terutama berguna dalam kasus di mana kelas minoritas dalam dataset memiliki representasi yang jauh lebih rendah dibandingkan dengan kelas mayoritas. Dengan menerapkan *SMOTE Upsampling*, model pembelajaran mesin cenderung menjadi lebih mampu mengenali pola dalam kelas minoritas, menghasilkan kinerja yang lebih baik dalam tugas-tugas seperti klasifikasi di mana ketidakseimbangan kelas menjadi masalah.

### 3.3. Transformation

*Transformasi* dalam konteks *KDD (Knowledge Discovery in Databases)* mengacu pada proses mengubah atau mentransformasi data yang telah dikumpulkan menjadi bentuk atau format yang lebih sesuai untuk analisis lebih lanjut. Proses transformasi ini dapat melibatkan berbagai tindakan, seperti penggabungan, normalisasi, penghapusan *outlier*, atau pembuatan fitur baru. Transformasi melibatkan

konversi atau modifikasi data awal menjadi data yang lebih terstruktur untuk mempermudah proses analisis[5].

### 3.4. Data Mining

Data mining melibatkan ekstraksi informasi yang bermanfaat atau pengetahuan berharga dari dataset yang luas atau kompleks. Tujuannya utama adalah untuk mengungkap pola, hubungan, atau *tren* yang mungkin tidak langsung terlihat melalui metode analisis *konvensional*. Proses ini melibatkan penggunaan teknik dan algoritma analisis data untuk mengidentifikasi pola tersembunyi dalam data, memberikan wawasan berharga, dan mendukung pengambilan keputusan.

### 3.5. Evaluation

*Evaluasi* adalah proses penilaian atau pengukuran kinerja, *efektivitas*, atau nilai suatu objek, program, atau sistem. Tujuan utama dari evaluasi adalah untuk memahami secara menyeluruh sejauh mana suatu entitas mencapai tujuannya, keberhasilan implementasinya, atau dampak yang diinginkan.

## 4. HASIL DAN PEMBAHASAN

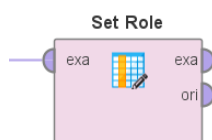
### 4.1. Data Selection

Dalam penelitian ini, data yang digunakan berasal dari data sekunder. Data tersebut bersifat publik yang diperoleh dengan cara scraping menggunakan *Google Colab*. Dataset yang digunakan berjumlah 2000 data sampel ulasan sentimen Aplikasi AdaKami di *Google Play Store*.

Tabel 1. Data

No	Atribut	Type
1.	Ulasan	<i>Polynomial</i>
2.	Label	<i>Bionominal</i>

Dalam penelitian ini, terdapat dua atribut utama yang menjadi fokus, yaitu ulasan dan sentimen. Kedua atribut ini tidak memiliki pengaruh terhadap tahapan data mining yang digunakan untuk mengukur tingkat akurasi. Untuk memberikan atribut *ID* pada dataset, digunakan operator *Set Role*. Operator *Set Role* akan mengubah peran atribut menjadi *ID* dan sejenisnya. Tampilan dari operator *Set Role* seperti yang terlihat pada gambar di bawah ini:



Gambar 3. Set Role

Pada operator *Set Role*, terdapat beberapa parameter yang perlu disesuaikan untuk mengubah peran atribut. Untuk atribut yang akan diubah menjadi peran target, yaitu Sentimen (*Label*), diubah menjadi *Label*. Hasil dari operator *Set Role* dapat dijelaskan sebagai berikut:

Gambar 4. Hasil Set Role

Dengan melihat hasil dari *Set Role* pada gambar tersebut, dapat disimpulkan bahwa atribut label akan berfungsi sebagai target dalam penelitian ini. Untuk mengatur peran data di RapidMiner, dapat memanfaatkan operator "*Set Role*".

### 4.2. Preprocessing

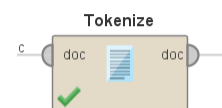
*Preprocessing* teks adalah proses mengubah data teks yang tidak terstruktur menjadi format terstruktur sesuai kebutuhan untuk mengekstrak informasi selama proses penambangan.

Gambar 5 Data Sebelum di Processing

Pada gambar 5 menunjukkan bahwa data tersebut sebelum dilakukannya *Preprocessing* masih terdapat karakter seperti `[-!"#$%&'()*+,-./:;<=>?@[\|_`{}~]`. Untuk melakukan *Claning* data seperti penjelasan di bawah ini:

#### 4.2.1. Tokenize

Untuk menghilangkan atau menghapus karakter yang terdapat pada data yang telah di infutkan, dengan menggunakan *Tokenize*. Terdapat beberapa parameter yang harus di ubah pada Mode sebagai *regular expression* dan *expression* di isi sebagai berikut: `[-!"#$%&'()*+,-./:;<=>?@[\|_`{}~]`. Tampilan *Tokenize* seperti yang terlihat pada gambar di bawah ini:



Gambar 6 Tokenize

Berikut hasil dari operator *Tokenize* seperti yang terlihat pada gambar di bawah ini:

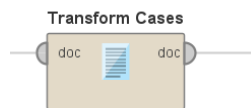
Row No.	Index	Text	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	prosedur	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	2	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	3	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	4	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	5	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	6	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	7	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	8	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	9	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	10	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	11	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	12	prosedur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Gambar 7. Hasil Tokenize

Hasil dari proses *Tokenize* adalah membersihkan sentimen dengan tujuan mengidentifikasi bagian data yang mungkin tidak lengkap, salah, tidak akurat, atau hilang. Proses pembersihan pada token melibatkan penghapusan semua karakter atau simbol yang ada pada dataset.

#### 4.2.2. Transform Cases

*Transform Cases* berfungsi untuk mengubah bentuk huruf pada data. Fungsi ini untuk melakukan transformasi pada kasus teks, seperti mengubah teks menjadi huruf besar (*uppercase*), huruf kecil (*lowercase*), atau huruf awal setiap kata menjadi huruf besar (*title case*). Tampilan *Transform Cases* seperti yang terlihat pada gambar di bawah ini:



Gambar 8 Transform Cases

Berikut hasil dari operator *Transform Cases* seperti yang terlihat pada gambar di bawah ini:

Row No.	Index	Text	1	2
1	1	semangat untuk bisa menggapai impian yang lebih mulia para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
2	2	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
3	3	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
4	4	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
5	5	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
6	6	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
7	7	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
8	8	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
9	9	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
10	10	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
11	11	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1
12	12	semangat yang harus dimiliki oleh para pemimpin saat ini sangat penting yang harus dimiliki oleh para pemimpin yang akan memimpin bangsa ini	1	1

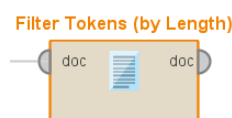
Excel 2010 2.000 words

Gambar 9. Hasil Transform Cases

Hasil dari *Transform Cases*, seperti yang terlihat seperti yang terlihat pada gambar di atas menunjukkan bahwa semua kalimat dalam data tersebut telah diubah menjadi huruf kecil (*lowercase*).

#### 4.2.3. Filter Tokens (by Length)

Operator ini dapat digunakan untuk berbagai keperluan, seperti menghapus token yang terlalu pendek atau terlalu panjang dari data teks dan memfokuskan analisis pada token dengan panjang tertentu.

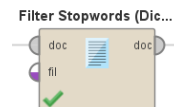


Gambar 10. Filter Tokens (by Length)

Hasil dari *Filter Tokens (by Length)* adalah mengeliminasi kata-kata dalam teks yang memiliki panjang kurang dari 4 karakter dan juga menghilangkan kata-kata yang memiliki panjang lebih dari 25 karakter.

#### 4.2.4. Filter Stopwords

*filter stopwords* ini Bertujuan untuk menghapus kata-kata yang tidak memberikan informasi penting dalam analisis teks, dengan demikian meningkatkan kualitas hasil analisis. Tampilan *Filter stopwords* seperti yang terlihat pada gambar di bawah ini:

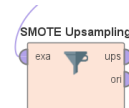


Gambar 11. Filter stopwords

*Filter stopwords* dapat digunakan dengan menggunakan kamus (*dictionary*) yang berisi daftar kata-kata yang dianggap *stopwords*. Fungsi utama dari *filter stopwords* ini adalah untuk membersihkan teks dari kata-kata yang sering muncul tetapi tidak memberikan makna yang kaya dalam konteks analisis.

#### 4.2.5. SMOTE Upsampling

Untuk melakukan *upsampling* atau peningkatan jumlah sampel pada kelas minoritas dalam suatu dataset. Fungsi ini dirancang khusus untuk menangani ketidakseimbangan kelas dalam konteks klasifikasi. Tampilan *SMOTE Upsampling* dapat dilihat pada gambar tersebut:

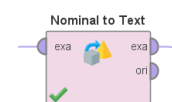


Gambar 12. SMOTE Upsampling

Hasil dari *SMOTE Upsampling* adalah dataset yang seimbang, yaitu dengan jumlah data yang sama pada setiap kelas. Hal ini dapat meningkatkan kinerja model pembelajaran mesin yang digunakan untuk klasifikasi, seperti akurasi, *precision*, dan *recall*.

#### 4.3. Transformation

Fungsi *Nominal to Text* adalah Untuk mengubah jenis atribut nominal yang dipilih menjadi teks, operator ini juga memetakan semua nilai dari atribut ini ke nilai string yang sesuai. Tampilan *Nominal to Teks* dapat dilihat pada gambar di bawah ini:



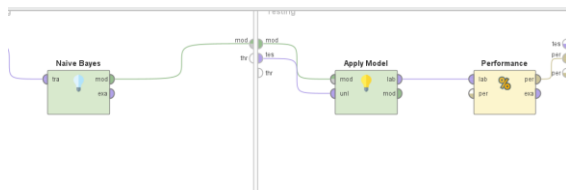
Gambar 13. Nominal to Teks



Hasil dari *Nominal to Teks* pada *RapidMiner* adalah dataset yang memiliki kolom baru dengan nilai teks. Kolom baru ini berisi nilai teks dari data *nominal* sebelumnya. Untuk mengubah data *nominal* menjadi data teks pada *RapidMiner*, dapat dilakukan dengan menggunakan operator *Nominal to Text*.

#### 4.4. Data Mining

Maksud dari *data mining* adalah untuk mengeksplorasi wawasan yang bernilai atau mendapatkan pengetahuan baru dari data yang besar, rumit, dan mungkin terstruktur secara tidak teratur. Sementara itu, *Naïve Bayes* merupakan salah satu metode klasifikasi dalam *machine learning* yang bergantung pada teorema *Bayes*.



Gambar 14. Model Naïve Bayes

Hasil *Performance* yang diperoleh dari metode *Naïve Bayes* dapat dilihat pada gambar di bawah ini:

	True Negatif	True Positif	Class Precision
pred. negatif	1714	339	83.48%
pred. positif	276	1651	85.98%
class recall	86.13%	82.96%	

Gambar 15. Hasil Model Naïve Bayes

Berdasarkan ilustrasi di atas, performa analisis sentimen ulasan Aplikasi AdaKami di *Google Play Store* dengan menggunakan metode *Naïve Bayes* menunjukkan hasil akurasi sebesar 84.55%. Sementara untuk kelas yang diprediksi sebagai negatif, akurasinya mencapai 86.13%, dengan *True Negative (TN)* sebanyak 276 dan *True Positive (TP)* sebanyak 1714. Sedangkan untuk kelas yang diprediksi sebagai positif, akurasinya mencapai 82.96%, dengan *True Negative (TN)* sebanyak 1651 dan *True Positive (TP)* sebanyak 339.

#### 4.5. Evaluation

Pengevaluasian pola adalah tahap di mana pola yang secara konsisten muncul berdasarkan metrik yang ditetapkan diidentifikasi. Penilaian ini dilakukan setelah data melalui proses *Naïve Bayes*, yang memberikan jumlah sentimen positif dan negatif. *Confusion matrix* digunakan sebagai alat evaluasi, memungkinkan estimasi objek yang diklasifikasikan dengan benar dan yang salah oleh model. Terdiri dari *True Positives (TP)*, yang merupakan jumlah data kelas positif yang diklasifikasikan sebagai kelas positif, dan *True Negatives (TN)*, yang merupakan

jumlah data kelas negatif yang diklasifikasikan sebagai kelas negatif. Sementara itu, *False Positives (FP)* mewakili jumlah data kelas negatif yang salah diklasifikasikan sebagai kelas positif, dan *False Negatives (FN)* mewakili jumlah data kelas positif yang salah diklasifikasikan sebagai kelas negatif [10]. Tabel berikut menunjukkan gambaran *Confusion Matrix*:

Tabel 2. *Confusion Matrix*

Prediksi	Actual	
	TP	FP
	FN	TN

Rumus untuk melakukan perhitungan *Accuracy*:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$Accuracy = \frac{1714+1651}{1714+1651+339+276} = 0.84\% \quad (2)$$

Dengan:

*TP* = *True Positive*

*TN* = *True Negative*

*FP* = *False Positive*

*FN* = *False Negative*

Rumus untuk melakukan perhitungan *Recall*:

$$Recall = \frac{(TP)}{(TP+FP)} = \frac{1714}{1714+276} = 0.86\% \quad (3)$$

Dengan:

*TP* = *True Positive*

*FP* = *False Positive*

Rumus untuk melakukan perhitungan *Precision*:

$$Precision = \frac{(TP)}{(TP+FN)} = \frac{1714}{1714+339} = 0.83\% \quad (4)$$

Dengan:

*TP* = *True Positive*

*FN* = *False Negative*

## 5. KESIMPULAN DAN SARAN

Hasil evaluasi menunjukkan bahwa algoritma *Naïve Bayes* mencapai tingkat akurasi sebesar 84.55%, dengan tingkat *Precision* sebesar 83%, dan tingkat *Recall* sebesar 86%. Untuk kelas yang diprediksi sebagai negatif, akurasi mencapai 86.13%, dengan *True Negative (TN)* sebanyak 276 dan *True Positive (TP)* sebanyak 1714. Sementara untuk kelas yang diprediksi sebagai positif, akurasi mencapai 82.96%, dengan *True Negative (TN)* sebanyak 1651 dan *True Positive (TP)* sebanyak 339. Dari hasil ini, dapat disimpulkan bahwa metode klasifikasi *Naïve Bayes* merupakan pilihan yang efektif untuk menganalisis data dengan baik. Penelitian ini memiliki beberapa kekurangan dan kelemahan yang perlu diperbaiki. Oleh karena itu, peneliti memberikan beberapa rekomendasi perbaikan. Pertama, penelitian mendatang sebaiknya mempertimbangkan pemanfaatan dua atau lebih metode atau algoritma untuk meningkatkan tingkat akurasi. Kedua, penelitian berikutnya dapat mempertimbangkan penggunaan

algoritma *alternatif* guna mencapai tingkat akurasi yang lebih baik. Terakhir, penting bagi pengguna pinjaman uang online untuk berhati-hati dalam menjaga keamanan data nasabah pada platform AdaKami. *Implementasi* perbaikan-perbaikan ini diharapkan dapat meningkatkan kualitas dan validitas hasil penelitian secara keseluruhan.

#### DAFTAR PUSTAKA

- [1] F. S. Mufidah, S. Winarno, F. Alzami, E. D. Udayanti, and R. R. Sani, "Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier," *JOINS (Journal Inf. Syst.*, vol. 7, no. 1, pp. 14–25, 2022, doi: 10.33633/joins.v7i1.5883.
- [2] Gilbert, Syariful Alam, and M. Imam Sulisty, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 100–108, 2023, doi: 10.55123/storage.v2i3.2333.
- [3] Ernianti Hasibuan and Elmo Allistair Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
- [4] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.
- [5] 2022 Kurnia & Nurhaliza Agustina. C.A., "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [6] I. Verawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1411, 2022, doi: 10.30865/mib.v6i3.4132.
- [7] T. A. Sari, E. Sinduningrum, and F. N. Hasan, "Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 773–779, 2023, doi: 10.30865/klik.v3i6.884.
- [8] F. Prasetya and F. Ferdiansyah, "Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 132, 2022, doi: 10.30865/json.v4i1.4852.
- [9] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Sattriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 234–241, 2022, doi: 10.37859/coscitech.v3i3.4343.
- [10] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.