

# Analisis Sentimen Ulasan Pengguna Aplikasi Atma Menggunakan Algoritma *Naïve Bayes*

Risna Kurnia Permata Ramadani<sup>1\*</sup>, Arya Soma Wicaksana<sup>2</sup>, Nunung Hidayatun<sup>3</sup>

<sup>1,2,3</sup>Teknik dan Informatika, Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Jl. Kramat Raya No.98

E-mail: <sup>1</sup>19210498@bsi.ac.id, <sup>2</sup>19210946@bsi.ac.id, <sup>3</sup>nunung.ntn@bsi.ac.id

(\*: corresponding author)

**Abstrak**— Pada masa ini kemajuan teknologi digital telah mendorong munculnya berbagai aplikasi pencari kerja berbasis mobile, salah satunya adalah Atma. Aplikasi ini dirancang untuk memudahkan pengguna dalam mencari lowongan pekerjaan, khususnya di sektor *food and beverage* (F&B). Namun, beragamnya ulasan pengguna dan volume yang terus meningkat di Google Play Store membuat analisis manual menjadi lambat dan rawan subjektivitas, sehingga menyulitkan pengembang memahami kebutuhan pengguna secara akurat. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Atma dengan menggunakan algoritma *Naïve Bayes* serta mengevaluasi performanya dalam mengklasifikasikan ulasan ke dalam kategori positif dan negatif. Pendekatan yang digunakan mengacu pada metode SEMMA (*Sample, Explore, Modify, Model, Assess*) dengan tahapan preprocessing teks, pembobotan menggunakan TF-IDF, dan penyeimbangan data menggunakan teknik SMOTE. Data yang dianalisis berjumlah 2000 ulasan, dan dilakukan tiga skenario pengujian berdasarkan rasio pembagian data latih dan data uji. Hasil pengujian menunjukkan bahwa akurasi algoritma meningkat secara signifikan setelah penerapan SMOTE. Akurasi tertinggi yang dicapai adalah sebesar 85,56% dengan proporsi 90% data latih dan 10% data uji. Temuan ini menunjukkan bahwa kombinasi *Naïve Bayes*, TF-IDF, dan SMOTE cukup efektif dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi. Penelitian ini diharapkan dapat memberikan masukan konstruktif bagi pengembang aplikasi dalam meningkatkan kualitas layanan sesuai kebutuhan pengguna.

**Kata Kunci**— Analisis Sentimen, Aplikasi Atma, *Naïve Bayes*, SEMMA.

**Abstract**— In the current digital era, rapid advancements in technology have encouraged the development of various mobile-based job search applications, one of which is Atma, focusing on the food and beverage (F&B) sector. However, the wide variety of user reviews and the continuously increasing volume of feedback on the Google Play Store make manual analysis slow and prone to subjectivity, making it difficult for developers to accurately understand user needs. This study aims to analyze user sentiment towards the Atma application using the *Naïve Bayes* algorithm and to evaluate its performance in classifying reviews into positive and negative categories. The research adopts the SEMMA methodology (*Sample, Explore, Modify, Model, Assess*), including text preprocessing, feature weighting with the TF-IDF method, and data balancing using the SMOTE technique. A total of 2,000 user reviews were analyzed under three different testing scenarios with varying proportions of training and testing data. The results show that the *Naïve Bayes* algorithm's accuracy improves significantly after applying SMOTE, with the highest accuracy of 85.56% achieved using a 90% training and 10% testing data ratio. These findings

indicate that combining *Naïve Bayes*, TF-IDF, and SMOTE is effective for sentiment analysis of user reviews and can provide valuable input for developers to enhance the Atma platform's service quality and overall user experience.

**Keyword**— Sentiment Analysis, Atma Application, *Naïve Bayes*, SEMMA

## I. PENDAHULUAN

Seiring berkembangnya teknologi digital, platform pencarian kerja online kini berperan sebagai sumber utama bagi individu dalam memperoleh peluang kerja yang sesuai dengan kualifikasi mereka [1]. Perkembangan ini mendorong pergeseran media informasi dari cetak ke digital, sehingga lowongan pekerjaan semakin mudah diakses melalui aplikasi, media sosial, maupun situs web berkat ketersediaan internet [2].

Di Indonesia, aplikasi pencari kerja semakin berkembang dengan hadirnya berbagai platform populer seperti Glints, JobStreet, Kitalulus, dan Atma. Atma menjadi salah satu platform yang cukup diminati, terlihat dari jumlah unduhan yang telah melampaui satu juta kali di Google Play Store. Sebagai aplikasi berbasis mobile, Atma berfokus pada penyediaan lowongan kerja terutama di sektor *food and beverage* (F&B), sekaligus menawarkan fitur pengembangan karier bagi para pengguna.

Ulasan atau feedback dari pengguna merupakan sumber informasi yang berharga bagi perusahaan. Melalui feedback tersebut, perusahaan dapat mengetahui berbagai permasalahan, kekurangan, serta ketidakpuasan yang dirasakan konsumen, sehingga dapat dijadikan dasar untuk melakukan perbaikan [3]. Meskipun demikian, tantangan yang muncul juga semakin besar, terutama dalam hal mengolah dan menganalisis data ulasan secara efisien [4].

Menurut [5] salah satu tantangan yang dihadapi adalah beragamnya ulasan yang diberikan, mulai dari opini positif hingga negatif, yang membuat proses pemahaman terhadap preferensi pengguna menjadi sulit. Selain itu, meningkatnya jumlah pengguna, volume ulasan yang terus bertambah setiap hari menjadikan proses analisis secara manual menjadi kurang efisien. Analisis data secara manual tidak hanya memakan waktu yang lama, tetapi juga berpotensi dipengaruhi oleh subjektivitas [6]. Oleh karena itu, diperlukan pendekatan yang lebih otomatis dan objektif, salah satunya adalah analisis sentimen. Analisis sentimen sendiri didefinisikan sebagai proses sistematis untuk mengenali serta mengelompokkan opini atau perasaan seseorang terhadap layanan, aplikasi,

maupun produk ke dalam tiga kategori, yaitu positif, negatif, dan netral [7].

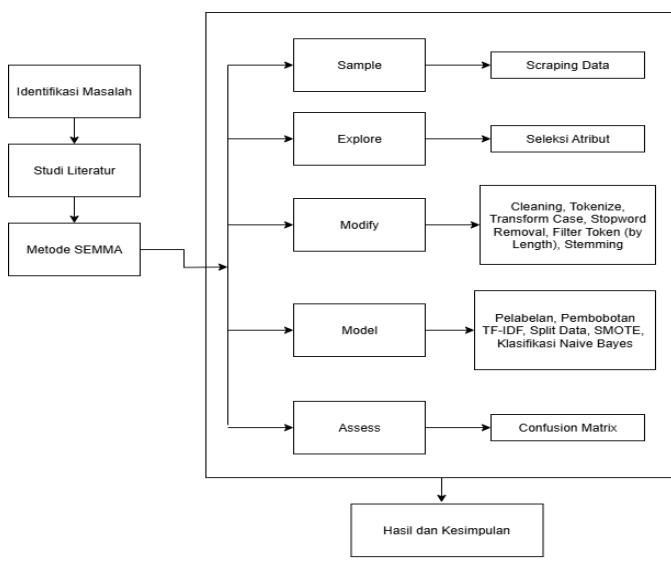
*Naive Bayes* merupakan salah satu algoritma yang banyak digunakan dalam analisis sentimen berbasis teks karena kemampuannya mengolah data dalam jumlah besar secara efisien dengan beban komputasi yang relatif ringan. Oleh karena itu, penelitian ini menerapkan algoritma *Naive Bayes* sebagai pendekatan untuk meningkatkan efektivitas dan ketepatan analisis sentimen.

Namun, dalam beberapa kasus dataset dapat mengalami masalah khusus, seperti ketidakseimbangan kelas data. Ketidakseimbangan kelas terjadi ketika jumlah data pada setiap kategori proporsinya tidak sama [8]. Untuk itu, SMOTE (*Synthetic Minority Oversampling Technique*) diterapkan dengan menambahkan data sintetis pada kelas minoritas, sehingga proporsi antar kelas lebih seimbang dan permasalahan oversampling dapat diminimalkan.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan ulasan pengguna aplikasi Atma menggunakan algoritma *Naive Bayes* guna memahami persepsi mereka terhadap aplikasi tersebut. Selain itu, penelitian ini juga berfokus pada evaluasi efektivitas algoritma *Naive Bayes* dalam melakukan klasifikasi sentimen, serta menilai apakah penerapan teknik SMOTE (*Synthetic Minority Over-Sampling Technique*) dapat meningkatkan performa model dalam analisis sentimen. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan aplikasi pencari kerja yang lebih adaptif serta mampu menjawab kebutuhan pengguna secara lebih baik. Selain itu, penelitian ini juga diharapkan dapat menjadi landasan untuk penelitian selanjutnya dalam upaya meningkatkan akurasi dan efisiensi algoritma-algoritma sejenis.

## II. METODE PENELITIAN

Penelitian ini menggunakan metode SEMMA, yaitu metodologi *data mining* yang terdiri dari 5 tahap, yaitu *Sample*, *Explore*, *Modify*, *Model*, dan *Assess*. Alur penelitian dengan metode SEMMA dapat dilihat pada Gambar 1 berikut.



Gambar 1. Alur Penelitian

### A. Sample

Sampel yang digunakan dalam penelitian ini berjumlah 2.000 ulasan yang diperoleh dari Google Play Store. Proses pengumpulan data dilakukan melalui *web scraping* menggunakan Python dengan library *google-play-scraper* pada platform Google Colab. Data hasil scraping disimpan dalam format CSV dan kemudian diseleksi agar sesuai dengan kebutuhan penelitian.

### B. Explore

Pada bagian ini membahas mengenai deskripsi dan visualisasi data. Deskripsi data bertujuan memberikan gambaran umum mengenai informasi yang terkandung dalam data, sedangkan visualisasi data menyajikan informasi tersebut dalam bentuk grafis [9].

### C. Modify

Tahap *Modify* dalam penelitian ini merupakan proses persiapan data atau *pre-processing* [10]. Teknik ini bertujuan membantu sistem komputer dalam memahami beragam jenis *dataset*. Selain itu, fase ini berfungsi untuk mengubah data dari yang tidak terstruktur menjadi data yang terstruktur [11]. Proses ini meliputi beberapa langkah, yaitu *cleaning*, *tokenize*, *transform case*, *stopword removal*, *filter token*, dan *stemming*.

#### 1) Cleaning

Tahap *cleaning* dilakukan untuk menghapus elemen-elemen yang tidak diperlukan dalam analisis, seperti URL, retweet, hashtag, mention, tanda baca, simbol, dan karakter tertentu.

#### 2) Tokenize

Pada tahap ini, setiap kalimat diuraikan menjadi potongan-potongan kata atau suku kata untuk mempermudah proses pengolahan teks selanjutnya.

#### 3) Transform Case

Transform case atau *case folding* merupakan proses penyamaan bentuk huruf dalam suatu teks, dengan cara mengubah huruf besar menjadi huruf kecil atau sebaliknya.

#### 4) Stopword Removal

Pada tahap ini, kata-kata yang tidak relevan dihapus agar dapat diperoleh inti dari teks yang akan dianalisis.

#### 5) Filter Token

Pada tahap ini, data disaring berdasarkan jumlah karakter dalam setiap kata, di mana kata-kata yang memiliki kurang dari 4 huruf atau lebih dari 25 huruf akan dihapus.

#### 6) Stemming

Stemming merupakan proses yang dilakukan untuk mengubah suatu kata menjadi bentuk dasarnya.

### D. Model

Pada tahap ini, dilakukan proses pelabelan sentimen, pemberian bobot menggunakan metode TF-IDF, pembagian data menjadi data latih dan data uji, penerapan teknik SMOTE, serta pengujian model menggunakan algoritma *Naive Bayes*.

#### 1) Pelabelan

Pelabelan merupakan proses pemberian nilai sentimen pada teks, yang dapat berupa positif, negatif, maupun netral. Proses pelabelan dilakukan secara manual oleh peneliti dengan

menggunakan Microsoft Excel. Setiap ulasan dibaca secara seksama satu per satu, kemudian diberi label sesuai dengan isi sentimennya.

### 2) TF-IDF

*Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan metode yang digunakan dalam analisis teks untuk menentukan bobot suatu kata dalam dokumen. Teknik ini bertujuan untuk mengenali kata-kata yang memiliki tingkat kepentingan tinggi dalam satu dokumen maupun kumpulan dokumen. *Term Frequency* (TF) merupakan jumlah kemunculan sebuah kata dalam satu dokumen. Namun, nilai TF saja tidak cukup untuk menunjukkan seberapa penting kata tersebut. Kata umum atau kata hubung bisa saja memiliki nilai TF yang tinggi karena sering muncul, padahal sebenarnya tidak terlalu berpengaruh terhadap makna isi dokumen. Karena itu, diperlukan konsep *Inverse Document Frequency* (IDF), yang memberikan nilai lebih tinggi pada kata-kata yang jarang muncul di seluruh dokumen. Nilai TF-IDF yang tinggi menunjukkan bahwa kata tersebut memiliki tingkat kepentingan yang besar dalam menentukan konteks atau topik dari dokumen tersebut [12].

### 3) Pembagian data

Pada tahap ini, data akan dipisahkan menjadi dua kelompok, yaitu data pelatihan dan data pengujian. Penelitian ini menggunakan tiga skenario pembagian data, yakni 90:10, 80:20, dan 70:30, di mana masing-masing menunjukkan proporsi antara data yang akan digunakan dalam proses pelatihan dan pengujian model. Data latih dimanfaatkan untuk membangun model serta melatihnya agar mampu mengenali karakteristik sentimen, sedangkan data uji digunakan sebagai

evaluasi untuk mengetahui seberapa akurat model dalam mengklasifikasikan sentimen pada setiap ulasan [13].

### 4) Synthetic Minority Oversampling Technique (SMOTE)

SMOTE merupakan pendekatan yang digunakan untuk mengatasi ketidakseimbangan distribusi kelas data. Cara kerja SMOTE yaitu dengan menambahkan data sintetis pada kelas minoritas agar jumlahnya seimbang dengan kelas mayoritas [14].

### 5) Algoritma *Naïve Bayes*

Selanjutnya, data latih dan data uji yang telah dibagi diproses menggunakan algoritma *Naïve Bayes Classifier*. Algoritma *Naïve Bayes* merupakan metode berbasis probabilitas yang menghitung peluang terbesar berdasarkan frekuensi kemunculan setiap kategori dalam data pelatihan [15]. Metode ini dipilih karena sesuai untuk analisis sentimen berbasis teks, bersifat sederhana, efisien, serta mampu memberikan kinerja yang baik pada data teks berskala besar. Selain itu, menurut penelitian yang dilakukan oleh Devita et al. menunjukkan bahwa algoritma *Naïve Bayes* dapat menghasilkan akurasi yang cukup tinggi meskipun hanya menggunakan jumlah data pelatihan yang terbatas [16].

### E. Assess

Pada tahap ini, akan dilakukan evaluasi terhadap model yang telah dikembangkan berupa nilai *confusion matrix*, yang mencakup *accuracy*, *precision*, dan *recall*. *Confusion Matrix* merupakan sebuah matriks yang digunakan untuk mengevaluasi performa algoritma klasifikasi dengan cara menyajikan jumlah prediksi benar dan keliru berdasarkan hasil klasifikasi [17].

## III. HASIL DAN PEMBAHASAN

### A. Sample

Penelitian ini menggunakan 2000 ulasan aplikasi Atma yang diperoleh dari Google Play Store melalui proses *web scraping* menggunakan Python dengan library *google-play-scraping* pada platform Google Colab. Data hasil scraping disimpan dalam format CSV dan kemudian diseleksi agar sesuai dengan kebutuhan penelitian.

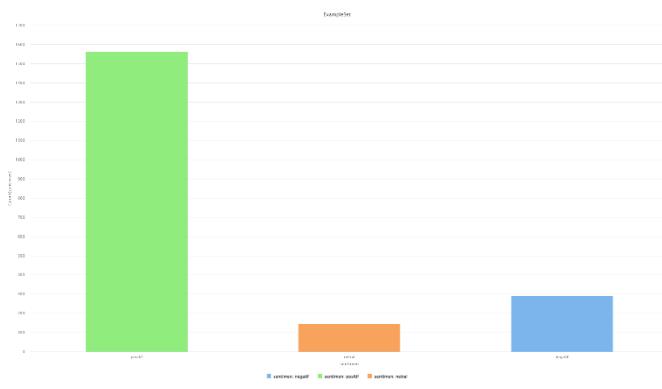
reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
2	c842e667-365f-Mahdi Yani		https://play-ll.loker nya kurang update	3	0		15/05/2025 05:05	Terima kasih atas mas	15/05/2025 06:30	
3	ab659af4-70b7-Porinus Giawa		https://play-ll.sangat memudahkan dalam men	5	0	473	13/05/2025 04:13	Terima kasih atas feec	13/05/2025 06:19	473
4	63c66333-bdd7-lilik wagiyem		https://play-ll.saya puas ,dengan informas	5	0		09/05/2025 08:48	Terima kasih banyak, I	13/05/2025 06:17	
5	ca111388-8cd3-Muhammad Nizar		https://play-ll.good jobb	5	0	473	05/05/2025 05:52	Makasih banyak ya, Ke	05/05/2025 09:02	473
6	aadd00cb-674b-Sanjes 123		https://play-ll.aplikasi caca, mau chat perusah	1	0	473	01/05/2025 08:27	Halo Kak, mohon maaf	01/05/2025 11:49	473
7	06b3aa0-0e2f-Naufal Indra		https://play-ll.gajis apknya ud dijlsin alamat mar	1	0		28/04/2025 09:22	Hai Kak, mohon maaf	30/04/2025 02:35	
8	c00e7441-dc80-Mariana		https://play-ll.kga bisa ngirim dokumen, data p	1	0	473	24/04/2025 06:38	Hai Kak Mariana, moh	24/04/2025 08:51	473
9	4f5b1bed-6337-willdan syahrial		https://play-ll.tidak rekomendasi catatan buat :	3	0		23/04/2025 02:03	Halo Kak, terima kasih	23/04/2025 09:35	
10	a30d65ad-b89b-Singaraja Bali		https://play-ll.aplikasi ini sangat membantu sel	5	0	473	22/04/2025 23:44	Wah senangnya dengi	23/04/2025 09:36	473
11	bec4f473-644d-Andri Yanto		https://play-ll.loker belum lengkap, loker bagus	2	0		21/04/2025 05:49	Halo Kak, terima kasih	21/04/2025 09:55	
12	40996e52-7d50-Rahmat Hidayat		https://play-ll.bagus apk nya	5	0	473	15/04/2025 05:04	Terima kasih banyak K	15/04/2025 09:15	473
13	0b942b55-94ea-Doni Namang		https://play-ll.de best	5	0	473	12/04/2025 10:08	Wah, terima kasih bar	14/04/2025 09:07	473
14	18370147-e6d4-Percetakan AzharPr		https://play-ll.mantap	5	0		08/04/2025 13:41	Terima kasih banyak a	09/04/2025 08:26	
15	be0a6205-9377-Abi Manyu		https://play-ll.bagus	5	0		06/04/2025 10:49	Terima kasih banyak a	07/04/2025 09:51	
16	f3137002-828b-Delviany A Manihur		https://play-ll.bagus	5	0		01/04/2025 15:36	Terima kasih banyak, I	02/04/2025 09:58	
17	d31fe541-cb3b-Nibras Hilwati		https://play-ll.Apk nya mudah dimengerti. Jobi	4	2	473	10/03/2025 10:10	Terima kasih atas mas	10/03/2025 08:27	473
18	8ab0a4d9-cbe9-Supri Yatna		https://play-ll.Sangat membantu	5	0	473	10/03/2025 08:54	Terima kasih banyak K	08/04/2025 09:45	473
19	07a775d0-feac-Juli Eka Wandana		https://play-ll.Saya ksh bintang 4 dlh, karena b	4	0	473	09/03/2025 04:53	Terima kasih banyak K	10/03/2025 08:28	473
20	1ec8626e-ed5b-Nur fika Yuniarti		https://play-ll.Susah mau di jelaskan.intinya ga n	1	0		25/02/2025 10:40	Halo kak Nur fika Yuni	26/02/2025 08:45	

Gambar 2. Hasil Scraping Data

### B. Explore

Data ulasan yang diperoleh memiliki berbagai atribut, yaitu *reviewId*, *userName*, *content*, *score*, *thumbsUpCount*, *reviewCreatedVersion*, *at*, *replyContent*, *repliedAt*, dan *appVersion*. Namun, dalam analisis ini hanya atribut *content*

yang digunakan karena atribut lainnya dianggap tidak berpengaruh secara langsung terhadap analisis sentimen. Dari total 2.000 ulasan tersebut, terdapat 1.562 ulasan positif, 292 ulasan negatif, dan 146 ulasan netral.



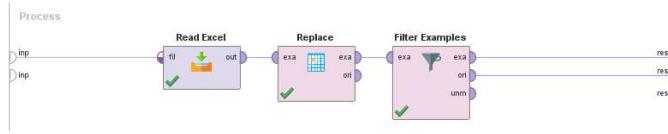
Gambar 3. Visualisasi Distribusi Sentimen

### C. Modify

Tahapan *modify* dilakukan dengan *preprocessing* teks menggunakan RapidMiner. Proses ini meliputi beberapa langkah berikut:

#### 1) Cleaning

Proses menghapus elemen yang tidak relevan, seperti URL, tanda baca, simbol, angka, dan karakter khusus. Setelah melalui tahap ini, diperoleh 1.993 data yang siap untuk diproses lebih lanjut.



Gambar 4. Proses Cleaning

#### 2) Tokenize

Proses memecah kalimat menjadi kata-kata (token) agar dapat diproses lebih lanjut.

#### 3) Transform Case

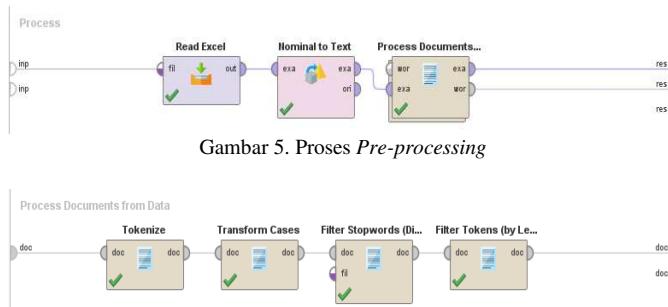
Proses mengubah seluruh huruf menjadi huruf kecil (*case folding*).

#### 4) Stopword Removal

Proses menghapus kata-kata umum yang tidak memiliki kontribusi signifikan, misalnya 'dan', 'atau', 'yang'. Kata-kata tersebut mengacu pada daftar stopwords bahasa Indonesia yang diunduh dari situs Kaggle.com.

#### 5) Filter Token (by Length)

Proses menghapus kata dengan panjang kurang dari 4 karakter atau lebih dari 25 karakter.



Gambar 5. Proses Pre-processing

### 6) Stemming

Proses mengubah kata menjadi bentuk dasarnya, misalnya kata 'mencari' menjadi 'cari'. Proses stemming dilakukan menggunakan bahasa pemrograman Python dengan menginstal sejumlah dependensi yang diperlukan seperti pandas, sastrawi, dan openpyxl.

TABEL I  
 HASIL PRE-PROCESSING

Tahap	Sebelum	Sesudah
<i>Cleaning</i>	Aplikasi ini sangat bagus sekali & sangat membantu dalam melamar pekerjaan	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan
<i>Tokenize</i>	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan
<i>Transform Case</i>	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan
<i>Stopword Removal</i>	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan	Aplikasi bagus membantu melamar pekerjaan
<i>Filter Token (by Length)</i>	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan	Aplikasi sangat bagus sekali sangat membantu dalam melamar pekerjaan
<i>Stemming</i>	Aplikasi ini sangat bagus sekali sangat membantu dalam melamar pekerjaan	Aplikasi ini sangat bagus sekali sangat bantu dalam lamar kerja

### C. Model

Pada tahap pemodelan, dilakukan proses pelabelan sentimen, pemberian bobot menggunakan metode TF-IDF, pembagian data menjadi data latih dan data uji, penerapan teknik SMOTE, serta pengujian model menggunakan algoritma *Naïve Bayes*.

#### 1) Pelabelan

Pelabelan dilakukan secara manual menggunakan Microsoft Excel dengan kategori sentimen positif dan negatif. Dari total 2.000 data ulasan, diperoleh 1.513 data berlabel positif, 291 data berlabel negatif, 146 data berlabel netral, serta 50 data tanpa label (kosong akibat proses *preprocessing*). Namun, dalam penelitian ini hanya data dengan sentimen positif dan negatif yang digunakan, sedangkan data netral dan tanpa label dikeluarkan dari proses analisis.

#### 2) Pembobotan TF-IDF

Metode ini digunakan untuk menghitung tingkat kepentingan setiap kata dalam dokumen. Kata yang jarang muncul pada keseluruhan dokumen tetapi sering muncul pada ulasan tertentu akan memperoleh bobot yang lebih tinggi. Proses pembobotan TF-IDF dalam penelitian ini dilakukan dengan menggunakan aplikasi RapidMiner.

aplikasi	apknya	aplikasi	aplikasinya	aplikas	aplikasi	aplikasibaru	aplikasidan	aplikasih
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	0	0	0	0	0	0	0
0	0	0	0	0	0.042	0	0	0
0	0.453	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0.366	0	0	0
0	0	0	0	0	0.096	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	0
0	0	0	0	0	0	0	0	0

ExampleSet (1,804 examples, 2 special attributes, 2,107 regular attributes)

Gambar 7. Hasil TF-IDF

### 3. Pembagian Data Latih dan Data Uji

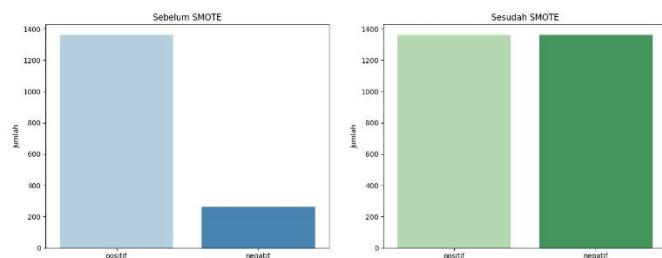
Selanjutnya data dibagi menjadi dua bagian, yaitu data latih dan data uji. Pembagian dilakukan dengan menggunakan tiga skenario, yakni perbandingan 90:10, 80:20, dan 70:30.

TABEL II  
 SKENARIO PEMBAGIAN DATA

Skenario	Data Latih	Data Uji
90:10	1624	180
80:20	1443	361
70:30	1263	541

### 4. SMOTE

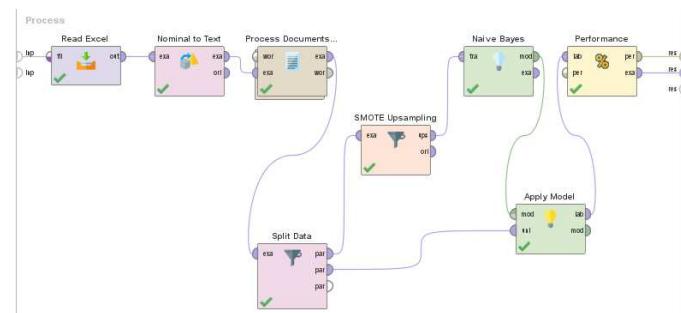
SMOTE digunakan untuk menyeimbangkan distribusi data ulasan positif dan negatif dengan cara menambahkan data sintetis pada kelas minoritas. Dalam penelitian ini terdapat ketidakseimbangan kelas, di mana jumlah sampel pada kelas positif jauh lebih banyak dibandingkan dengan kelas negatif. Untuk mengatasi hal ini, metode SMOTE diterapkan agar distribusi data menjadi lebih seimbang dan performa klasifikasi dapat ditingkatkan.



Gambar 8. Hasil SMOTE

### 5. Naïve Bayes Classifier

Model klasifikasi dibangun menggunakan algoritma *Naïve Bayes*. Algoritma ini dipilih karena sederhana, efisien, dan terbukti memberikan hasil yang baik pada dataset berbasis teks. Pada tahap ini, pengujian algoritma dilakukan menggunakan aplikasi RapidMiner.



Gambar 9. Pengujian Algoritma *Naïve Bayes*

### E. Assess

Evaluasi dilakukan menggunakan *Confusion Matrix* dengan tiga metrik utama, yaitu *accuracy*, *precision*, dan *recall*.

accuracy: 75.00%

	true negatif	true positif	class precision
pred. negatif	25	41	37.88%
pred. positif	4	110	96.49%
class recall	86.21%	72.85%	

Gambar 10. Confusion Matrix 90:10

accuracy: 77.56%

	true negatif	true positif	class precision
pred. negatif	50	73	40.65%
pred. positif	8	230	96.64%
class recall	86.21%	75.91%	

Gambar 11. Confusion Matrix 80:20

accuracy: 76.52%

	true negatif	true positif	class precision
pred. negatif	65	105	38.24%
pred. positif	22	349	94.07%
class recall	74.71%	76.87%	

Gambar 12. Confusion Matrix 70:30

Sebelum penerapan SMOTE, model cenderung bias terhadap kelas mayoritas, sehingga menghasilkan akurasi lebih rendah. Namun, setelah penerapan SMOTE terjadi peningkatan performa model secara signifikan. Hasil terbaik diperoleh pada skenario 90:10 dengan *accuracy* 85,56%, *precision* 84,12%, dan *recall* 86,40%.

TABEL III  
 AKURASI SEBELUM PENERAPAN SMOTE

Skenario	Accuracy	Precision	Recall
90:10	75.00%	67.19%	79.53%
80:20	77.56%	68.65%	81.06%
70:30	76.52%	66.16%	75.79%

TABEL IV  
 AKURASI SETELAH PENERAPAN SMOTE

Skenario	Accuracy	Precision	Recall
90:10	85.56%	75.09%	85.82%
80:20	84.49%	73.66%	83.79%
70:30	82.62%	70.87%	78.96%

#### IV. PENUTUP

Berdasarkan hasil penelitian analisis sentimen pengguna terhadap aplikasi Atma menggunakan algoritma *Naïve Bayes* dengan metode SEMMA, dapat ditarik kesimpulan bahwa (1) Analisis sentimen pengguna aplikasi Atma berhasil dilakukan dengan memanfaatkan ulasan berbahasa Indonesia dari Google Play Store. Dari 2000 data awal, setelah melalui tahap *preprocessing*, diperoleh 1804 data yang terdiri dari 1513 ulasan positif dan 291 ulasan negatif, (2) Penerapan algoritma *Naïve Bayes* dengan pembobotan TF-IDF menunjukkan bahwa model mampu mengklasifikasikan sentimen pengguna secara cukup baik. Namun, sebelum penerapan SMOTE, distribusi data yang tidak seimbang menyebabkan akurasi model kurang optimal. (3) Teknik SMOTE (*Synthetic Minority Oversampling Technique*) terbukti efektif dalam menyeimbangkan distribusi kelas data. Hasil pengujian menunjukkan adanya peningkatan performa algoritma setelah dilakukan penyeimbangan data. 4. Hasil terbaik diperoleh pada skenario 90:10, dengan akurasi 85,56%, precision 84,12%, dan recall 86,40%. Hal ini menunjukkan bahwa kombinasi *Naïve Bayes*, TF-IDF, dan SMOTE cukup efektif dalam melakukan analisis sentimen pada ulasan pengguna aplikasi Atma.

Penelitian selanjutnya dapat memperluas cakupan dengan menambahkan jumlah data lebih besar agar hasil analisis lebih representatif. Disarankan untuk membandingkan algoritma *Naïve Bayes* dengan algoritma lain seperti *Logistic Regression*, *Random Forest*, atau *Support Vector Machine* untuk mengetahui perbedaan kinerja pada dataset yang sama. Penggunaan teknik optimasi lain, seperti *ensemble learning*, dapat dipertimbangkan untuk meningkatkan akurasi model.

#### REFERENSI

- [1] D. S. Nurrochmah, N. Raharningsih, R. D. Dana, and C. L. Rohmat, "Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Ulasan Aplikasi Kitalulus Di Google Play Store," *J. Inform. Terpadu*, vol. 11, no. 1, pp. 1–11, 2025, doi: 10.54914/jit.v11i1.1544.
- [2] G. S. Al-Husna, D. Asmarajati, I. A. Ihsannuddin, and R. Mahmudati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi LinkedIn," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 3, no. 2, pp. 139–144, 2024, doi: 10.55123/storage.v3i2.3602.
- [3] S. N. Adhan, G. N. A. Wibawa, D. C. Arisona, I. Yahya, Agusrawati, and Ruslan, "Analisis Sentimen Ulasan Aplikasi Wattpad Di Google Play Store Dengan Metode Random Forest," *AnoatIK J. Teknol. Inf. dan Komput.*, vol. 2, no. 1, pp. 6–15, 2024, doi: 10.33772/anoatik.v2i1.32.
- [4] S. A. Bhat and N. F. Huang, "Big Data and AI Revolution in Precision Agriculture: Survey and Challenges," *IEEE Access*, vol. 9, pp. 110209–110222, 2021, doi: 10.1109/ACCESS.2021.3102227.
- [5] M. R. Firdaus, N. Raharningsih, and R. D. Dana, "Analisis Sentimen Aplikasi Shopee di Goole Play Store Menggunakan Klasifikasi Algoritma Naïve Bayes," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, pp. 228–237, 2024.
- [6] R. P. Kurniawan, Istiadi, and S. W. Iriananda, "Analisis Sentimen Ulasan Aplikasi LinkedIn Berbasis Lexicon dan Long Short-Term Memory (LSTM)," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 2315–2324, 2025, doi: 10.36040/jati.v9i2.13042.
- [7] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 601–609, 2020.
- [8] M. Syukron, R. Santoso, and T. Widiharih, "Perbandingan Metode Smote Random Forest dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C pada Imbalance Class Data," *J. Gaussian*, vol. 9, no. 3, pp. 227–236, 2020, doi: 10.14710/j.gauss.9.3.227-236.
- [9] M. D. Alizah, A. Nugroho, U. Radiyah, and W. Gata, "Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter," *Indones. J. Softw. Eng.*, vol. 6, no. 2, pp. 223–229, 2020, doi: 10.31294/ijse.v6i2.8991.
- [10] A. A. A. Miftachuddin and M. H. R. Fandany, "Analisis Sentimen Pengguna Aplikasi Cakap Belajar Online Menggunakan Algoritma Naïve Bayes," *J. Prosisko*, vol. 12, no. 1, 2025, doi: 10.30656/prosisko.v12i1.9117.
- [11] L. Nursinggah, Ruuhwan, and T. Mufizar, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis Dengan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 1615–1622, 2024, doi: 10.23960/jitet.v12i3.4336.
- [12] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," *J. Manaj. Inform.*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.
- [13] M. A. S. Nugroho, D. Susilo, and D. Retnoningsih, "Analisis sentimen ulasan aplikasi "access by kai" menggunakan algoritma machine learning," *J. Tekinkom*, vol. 7, no. 2, pp. 820–827, 2024, doi: 10.37600/tekinkom.v7i2.1854.
- [14] A. F. Anjani, D. Anggraeni, and I. M. Tirta, "Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 163–172, 2023, doi: 10.25077/teknosi.v9i2.2023.163–172.
- [15] M. K. Insan, U. Hayati, and O. Nurdianaw, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 478–483, 2023.
- [16] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [17] S. Muntazah, R. Astuti, and F. M. Basysyar, "Analisis Sentimen Implementasi Kurikulum Merdeka Tingkat SMP Di Kabupaten Ciamis Menggunakan Algoritma Naïve Bayes," *J. Elektron. dan Tek. Inform. Terap. (JENTIK)*, vol. 2, no. 1, pp. 06–20, 2024, doi: 10.59061/jentik.v2i1.625.