

## IMPLEMENTASI METODE *MULTINOMIAL NAIVE BAYES* UNTUK ANALISIS SENTIMEN TERHADAP DATA ULASAN APLIKASI MERDEKA MENGAJAR PADA *GOOGLE PLAY STORE*

Viky Fahreza Santosa<sup>1</sup>, Wahyu Pramusinto<sup>2</sup>, Hadidtyo Wisnu Wardani<sup>3</sup>

<sup>1,2,3</sup> Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1</sup>2011500234@student.budiluhur.ac.id, <sup>2</sup>wahyu.pramusinto@budiluhur.ac.id, <sup>3</sup>hadidtyo.wardani@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak-** Aplikasi Merdeka Mengajar merupakan salah satu aplikasi yang dirancang untuk mendukung proses belajar mengajar di Indonesia. Para pengguna khususnya guru, dapat dengan mudah *download* aplikasi Merdeka Mengajar secara gratis pada *Google Play Store*. Sama seperti aplikasi lainnya, pengguna yang telah menggunakan aplikasi Merdeka Mengajar dapat membagikan atau memberikan peringkat mulai dari bintang 1-5 diikuti dengan ulasan atau tanggapan mengenai pengalaman penggunaan melalui fitur *ratings* dan *reviews*. Hingga bulan Mei 2024 tercatat aplikasi Merdeka Mengajar memiliki nilai *rating* 4.8. *Rating* dan ulasan sangat berguna untuk pengambilan keputusan dan perbaikan aplikasi di masa yang akan datang. Akan tetapi terkadang dalam sebuah ulasan terdapat ketidaksesuaian antara isi ulasan dengan *rating* pada ulasan. Maka dari itu fokus penelitian ini adalah untuk membahas tentang analisis sentimen dari ulasan pengguna aplikasi Merdeka Mengajar. Tujuan penelitian ini menerapkan metode *Multinomial Naive Bayes* untuk analisis sentimen ulasan pengguna aplikasi Merdeka Mengajar melalui ulasan pada *Google Play Store*. Pengumpulan data dilakukan pada tanggal 20 Mei – 1 Juni 2024, mendapatkan 1127 dataset. Hasil penelitian ini menunjukkan bahwa sistem ini mampu mengoperasikan analisis sentimen terhadap aplikasi Merdeka Mengajar yang berada di *Google Play Store* dengan metode *Multinomial Naive Bayes*, menghasilkan akurasi 87% menggunakan 1127 dataset terdiri dari label positif berjumlah 843 dan label negatif berjumlah 284. Data ini dibagi menjadi 90% data latih dan 10% data uji. Kontribusi penelitian ini adalah mengimplementasikan metode *Multinomial Naive Bayes* untuk menganalisis sentimen terhadap data ulasan aplikasi Merdeka Mengajar pada *Google Play Store* yang dapat membantu dalam memahami persepsi pengguna terhadap aplikasi tersebut, serta memberikan wawasan untuk kedepannya supaya lebih baik dari sebelumnya.

**Kata Kunci:** *Multinomial naive bayes. google play store, analisis sentimen, aplikasi merdeka mengajar.*

## IMPLEMENTATION OF *MULTINOMIAL NAIVE BAYES* METHOD FOR SENTIMENT ANALYSIS OF INDEPENDENT TEACHING APPLICATION REVIEW DATA ON *GOOGLE PLAY STORE*

**Abstract-** The Merdeka Mengajar application is one of the applications designed to support the teaching and learning process in Indonesia. Users, especially teachers, can easily download the Merdeka Mengajar application for free on the *Google Play Store*. Just like other applications, users who have used the Merdeka Mengajar application can share or provide ratings ranging from 1-5 stars followed by reviews or feedback on the usage experience through the ratings and reviews feature. As of May 2024, the Merdeka Mengajar application has a rating value of 4.8. Ratings and reviews are very useful for decision making and future application improvements. However, sometimes in a review there is a discrepancy between the content of the review and the rating on the review. Therefore, the focus of this research is to discuss sentiment analysis of user reviews of the Merdeka Mengajar application. The purpose of this research is to apply the *Multinomial Naive Bayes* method for sentiment analysis of user reviews of the Merdeka Mengajar application through reviews on the *Google Play Store*. Data collection was carried out on 20 May - 1 June 2024, obtaining 1127 datasets. The results of this study indicate that this system is able to operate sentiment analysis of the Merdeka Mengajar application on the *Google Play Store* with the *Multinomial Naive Bayes* method, resulting in 87% accuracy using 1127 datasets consisting of 843 positive labels and 284 negative labels. This data is divided into 90% training data and 10% test data. The contribution of this research is to implement the *Multinomial Naive Bayes* method to analyse sentiment on Merdeka Mengajar application review data on the *Google Play Store* which can help in understanding user perceptions of the application, and provide insight for the future so that it is better than before.

**Keywords:** *Multinomial naive bayes. google play store, sentiment analysis, application merdeka mengajar.*

## 1. PENDAHULUAN

Perkembangan aplikasi *mobile* pada era digital saat ini telah menjadi bagian penting dari kehidupan atau aktivitas sehari-hari. Salah satu sektor yang memanfaatkan perkembangan ini adalah pendidikan. Aplikasi Merdeka Mengajar yang tersedia di *Google Play Store* merupakan salah satu aplikasi yang dirancang untuk mendukung sektor pendidikan maupun proses belajar mengajar di Indonesia.

Aplikasi Merdeka Mengajar ini membantu guru untuk mengajar, belajar, dan berkarya. Dengan fitur-fitur yang ada saat ini akan membantu guru dalam mengajar lebih baik, meningkatkan kompetensi, berkarya lebih baik untuk menunjang karir kedepannya. Aplikasi Merdeka Mengajar dapat dengan mudah diakses melalui *web browser* dan aplikasi seluler berbasis *android*. Para pengguna khususnya guru, dapat dengan mudah *mendownload* aplikasi Merdeka Mengajar secara gratis pada *Google Play Store*. Sama seperti aplikasi lainnya, pengguna yang telah menggunakan aplikasi Merdeka Mengajar dapat membagikan atau memberikan peringkat mulai dari bintang 1-5 diikuti dengan tanggapan melalui fitur *ratings* dan *reviews*.

Pada bulan Mei 2024 aplikasi Merdeka Mengajar memiliki nilai *rating* 4.8. *Rating* dan ulasan sangat berguna untuk pengambilan keputusan dan perbaikan aplikasi di masa yang akan datang. Akan tetapi terkadang dalam sebuah ulasan terdapat ketidaksesuaian antara isi ulasan dengan *rating* pada ulasan. Maka dari itu fokus penelitian ini adalah untuk membahas tentang analisis sentimen dari ulasan pengguna aplikasi Merdeka Mengajar. Analisis sentimen merupakan suatu kegiatan untuk memproses data basis teks secara otomatis dengan cara memahami dan mengekstrak data yang diperoleh dari suatu opini dan informasi berupa sentimen yang tersirat. Kategori sentimen yang dihasilkan ialah sentimen positif dan sentimen negatif [1]. Setiap ulasan yang diberikan pada aplikasi Merdeka Mengajar mengandung sentimen positif dan sentimen negatif.

Penelitian sebelumnya telah melakukan analisis sentimen dengan judul Analisis Sentimen Penerapan Kurikulum Merdeka di Twitter menggunakan metode *Naïve Bayes*, yang berhasil mencapai akurasi sebesar 96% [2]. Penelitian lain menggunakan metode *Multinomial Naïve Bayes* untuk analisis sentimen ulasan produk *Colearn* di *Google Play Store*, dengan 536 dataset yang terdiri dari 439 ulasan positif dan 97 ulasan negatif, menghasilkan akurasi 88,89% [3]. Penelitian ini bertujuan untuk menerapkan metode *Multinomial Naïve Bayes* dalam menganalisis sentimen ulasan pengguna aplikasi Merdeka Mengajar di *Google Play Store*, dengan tujuan mengkategorikan ulasan menjadi positif dan negatif serta mencapai tingkat akurasi yang optimal.

## 2. METODE PENELITIAN

### 2.1 Text Mining

*Text Mining* merupakan proses menambang data berbasis teks yang datanya bersumber dari dokumen dan proses ini bertujuan untuk mengamati kata-kata yang mewakili isi dari dokumen sehingga dapat dianalisa keterikatan antar dokumen. Tujuan adanya *text mining* ialah mengekstrak data yang bersumber dari informasi agar menjadi bermanfaat. Oleh karena itu, di dalam *text mining*, sumber data yang digunakan ialah berupa dokumen dengan format tidak sistematis yang diidentifikasi dan dieksplorasi melalui pola agar bisa dipahami [4].

### 2.2 Tahapan Text Mining

Pada tahapan *text mining* terdapat proses yang melibatkan untuk menjalankan program ini. Dapat diamati di bawah ini:

#### 2.2.1 Scraping

Teknik *web scraping* adalah teknik penarikan sebuah dokumen semi-terstruktur dari internet, dapat berbentuk *website* yang dirancang menggunakan bahasa pemrograman seperti *HTML* atau *XHTML* dengan tujuan untuk memperoleh informasi di halaman tersebut baik itu secara menyeluruh maupun sebagian akan digunakan untuk kebutuhan lain [5].

#### 2.2.2 Pelabelan Data

Pada pelabelan data adalah kegiatan untuk mengkategorikan atau melabeli untuk semua data yang berisi positif, negatif, dan netral, sehingga memudahkan proses analisis maupun penggunaannya. Proses ini dapat dilakukan dengan otomatis menggunakan algoritma maupun secara manual oleh manusia. Dalam kegiatan ini akurasi label dan ketetapan label sangatlah penting agar memastikan siap digunakan untuk maksud yang akan dicapai dan data tertata dengan baik [6].

### 2.2.3 Preprocessing

*Preprocessing* dilakukan untuk mengelola data yang ada sehingga peneliti dapat menghindari gangguan pada data-data yang tidak konsisten. Tujuannya agar hasil *output* dari klasifikasi memiliki tingkat keakuratan yang tinggi [4]. Dalam *preprocessing*, data yang tersedia harus diubah ke format agar mudah dipahami. Oleh karena itu, penelitian ini akan melakukan beberapa tahapan *preprocessing* seperti *case folding*, *cleaning*, mengganti *slang word*, menghapus *slang word*, dan *stemming*.

a. *Case Folding*

*Case folding* yaitu proses untuk mengubah semua huruf yang terdapat didalam data ulasan, dalam bentuk teks dan mengubahnya menjadi huruf kecil (*lowercase*) [7].

b. *Cleaning*

*Cleaning* yaitu teknik untuk menghilangkan beberapa atribut yang tidak digunakan, seperti membersihkan atau menghapuskan beberapa kata maupun simbol atau emotikon yang tidak terpakai [1].

c. *Mengganti Slang Word*

*Slang word* merupakan kata yang tidak mengikuti Ejaan Yang Disempurnakan (EYD), pada masa sekarang dikenal sebagai Pedoman Umum Ejaan Bahasa Indonesia (PUEBI), baik dalam bentuk singkatan, bahasa modern atau bahasa gaul, maupun kesalahan penulisan [8].

d. *Menghapus Stop Word*

Menghilangkan atau menghapus *stop word* adalah proses dihilangkannya kata-kata yang tidak memiliki makna signifikan namun sering muncul dalam dokumen atau kalimat. Kata yang ditunjukkan berupa “pada”, “dan”, “di”, dan “yang” [9]. Pada penelitian ini juga menggunakan daftar *stopword* yang bersumber dari Fadillah Z. Tala. Karena daftar ini bisa dipertanggung jawabkan secara ilmiah menurut rahutomo [10].

e. *Stemming*

Kemudian tahap terakhir dalam *preprocessing* ialah *stemming*, tujuan dari adanya proses ini adalah untuk mengubah kata berimbuhan menjadi kata dasar, yang sangat membantu mesin dalam membuat prediksi dan mengolah data [2].

### 2.3 Data Penelitian

Dataset atau sumber data yang dipakai pada penelitian ini bersumber dari ulasan aplikasi Merdeka Mengajar berada di *Google Play Store*. Sumber data dari ulasan aplikasi Merdeka Mengajar diperoleh menggunakan *platform Google Colab* yang dipadukan dengan bahasa *python*, merupakan *package* alat untuk membantu pengguna untuk pengolahan data dan proses pemrograman. Pada penelitian ini dataset yang diperoleh berjumlah 1127 data ulasan dengan label positif berjumlah 843 dan label negatif berjumlah 284 diambil pada tanggal 20 Mei - 1 Juni tahun 2024. Hasil *scraping* atau pengumpulan data dapat diamati pada tabel 1:

**Tabel 1.** Sampel Data Ulasan Dari Aplikasi Merdeka Mengajar

No.	CREATED_AT	USERNAME	REVIEWS	LABEL
1.	2024-01-05 22:17:46	A Razak Pengawas SD	Belajar adalah hal yang sangat mulia mempelajari adalah kewajiban setiap guru ataupun pengawas sekolah	Positif
2.	2024-04-19 11:57:11	aafaisal kasep	baik tuk meningkatkan kualitas guru terutama	Positif
3.	2024-05-06 22:38:42	Abasir 13	aplikasi ini sangat membantu sekali dalam pembelajaran	Positif

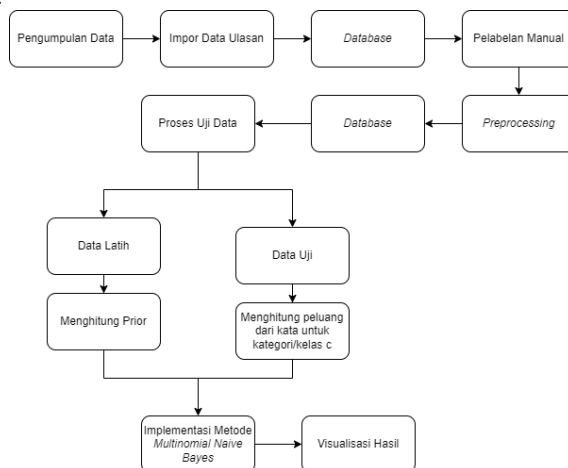
Setelah data terkumpul, teknik *stratified sampling* akan diterapkan dalam proses klasifikasi data. Teknik ini melibatkan pengambilan sampel data berdasarkan strata-strata yang ada dalam aplikasi Merdeka Mengajar. Penelitian ini berfokus pada pembagian data yang dibagi menjadi data latih dan data uji, di mana persentase rasio pembagiannya digunakan untuk menentukan akurasi tertinggi. Detail rasio pembagian data terdokumentasikan dalam tabel 2:

**Tabel 2.** Rasio Pembagian Data

Data Latih	Data Uji
90%	10%
80%	20%
70%	30%
60%	40%
50%	50%

## 2.4 Penerapan Metode

Penerapan metode pada penelitian analisis sentimen ini mempunyai beberapa tahapan. Tahapan ini berisi rincian dari rancangan dan setiap proses dalam penelitian ini dari pertama sampai terakhir. Rincian tersebut dapat diamati pada gambar 1:

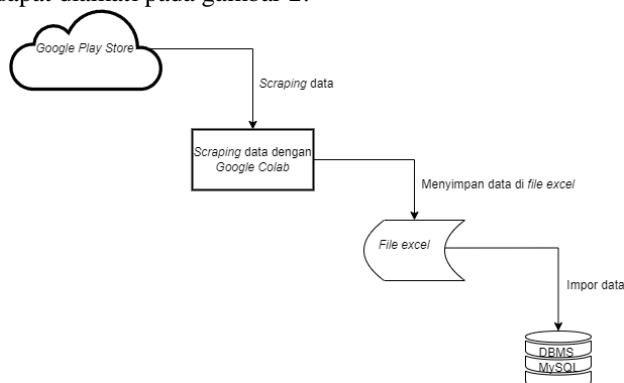


**Gambar 1.** Penerapan Metode

Pada Gambar 1, proses pertama adalah pengumpulan data dari *Google Play Store* melalui teknik *scraping*. Setelah data dikumpulkan, langkah selanjutnya adalah mengimpor dan menyimpan dataset ke dalam *database*. Data ulasan yang telah diimpor kemudian dilabeli secara manual. Setelah pelabelan, dataset yang sudah dilabeli menjalani proses *preprocessing* sebelum masuk ke tahap pengujian. Pada tahap pengujian, data dibagi menjadi dua set yaitu data latih dan data uji dengan rasio yang bervariasi seperti 9:1, 8:2, 7:3, 6:4, dan 5:5. Selanjutnya, perhitungan dilakukan untuk menentukan prior pada data latih dan peluang kata dari setiap dokumen pada data uji. Proses terakhir melibatkan klasifikasi, pengujian, perhitungan keakuratan hasil, termasuk akurasi, presisi, dan *recall*.

## 2.5 Pengumpulan Data

Pada tahap ini, proses pengumpulan data menggunakan teknik *scraping*. Proses pengumpulan data ulasan aplikasi Merdeka Mengajar yang berada di *Google Play Store* menggunakan platform bernama *Google Colab*. Platform *Google Colab* ini membantu untuk penarikan data ulasan aplikasi Merdeka Mengajar dengan menggunakan bahasa *python*. Pengumpulan data dilakukan secara *online* pada tanggal 20 Mei - 1 Juni 2024. Selanjutnya untuk proses ini dapat diamati pada gambar 2:



**Gambar 2.** Pengumpulan Data

## 2.6 Impor Data Ulasan

Dalam proses ini data ulasan diimpor ke dalam *database* sebagai dataset kotor yang akan digunakan untuk proses pelabelan terhadap data ulasan. Setelah masuk ke dalam *database*, maka akan masuk dan diproses oleh sistem. Data ulasan berasal dari ulasan aplikasi Merdeka Mengajar yang berada di *Google Play Store*. Proses impor data ulasan dapat diamati pada gambar 3:



Gambar 3. Impor Data Ulasan

## 2.7 Pelabelan Manual

Untuk proses pelabelan ini dilakukan secara manual pada dataset yang sudah masuk ke *database*, proses pelabelan terjadi di dalam sistem dimana sudah ada dua pilihan yaitu label positif dan label negatif.

Label positif mengacu pada penamaan atau penggolongan yang digunakan untuk menggambarkan atau mengidentifikasi sesuatu dengan atribut-atribut yang dianggap baik, diinginkan, atau menguntungkan. Dalam konteks yang lebih umum. Label positif bisa mengacu pada merek atau identifikasi produk yang dianggap memenuhi standar kualitas tinggi, aman, atau ramah lingkungan.

Label negatif mengacu pada penamaan atau penggolongan yang digunakan untuk menggambarkan atau mengidentifikasi sesuatu dengan atribut-atribut yang dianggap buruk, tidak diinginkan, atau merugikan. Contoh label negatif bisa berupa penandaan produk yang dianggap berisiko bagi kesehatan, tidak ramah lingkungan, atau tidak memenuhi standar keamanan tertentu. Pada pelabelan ini hanya dilakukan oleh penulis secara manual.

## 2.8 Preprocessing

Data yang sudah melalui proses pelabelan, selanjutnya akan masuk ke dalam *preprocessing*, ini melibatkan beberapa tahapan untuk memaksimalkan dari tahapan sebelumnya ke selanjutnya. Tahapan *preprocessing* terdiri dari *case folding*, *cleaning*, mengganti *slang word*, menghapus *stop word*, dan *stemming* dapat dilihat dibawah ini:

### a. Case Folding

*Case Folding* merupakan tahapan untuk melakukan proses persamaan untuk sebuah kata yang berisi huruf kapital diubah menjadi huruf kecil, contohnya dapat diamati: “Kamu” menjadi “kamu”, “Mengaji” menjadi “mengaji”, dan seterusnya.

### b. Cleaning

*Cleaning* merupakan proses untuk menghilangkan atribut yang tidak diperlukan dalam hal ini berada pada kata dan simbol serta menghilangkan emotikon yang tidak diperlukan.

### c. Mengganti Slang Word

Mengganti *slang word* adalah proses mengganti kata-kata yang tidak baku agar sesuai dengan Ejaan Yang Disempurnakan. Kata tidak baku tersebut dapat ditemui pada bahasa gaul atau Bahasa yang kekinian dan berupa singkatan. Adapun contoh dari *slang word* yaitu “pedes” dan “antri”.

### d. Menghapus Stop Word

Menghilangkan atau menghapus *stop word* adalah proses menghapus kata-kata yang tidak memiliki pengaruh besar atau kurang bermakna namun sering muncul dalam dokumen atau kalimat.

### e. Stemming

Kemudian tahap terakhir dalam *preprocessing* ialah *stemming*, tujuan dari adanya proses ini adalah untuk mengubah kata berimbuhan menjadi kata dasar, yang sangat membantu mesin dalam membuat prediksi dan mengolah data.

## 2.9 Klasifikasi Multinomial Naïve Bayes

*Multinomial Naïve Bayes* untuk analisis sentimen merupakan metode klasifikasi untuk menentukan kategori dokumen dengan cara menghitung probabilitas pada kategori tertentu dan menghitung probabilitas setiap kata dalam dokumen terhadap kategori tersebut.

### a. Menghitung sebuah peluang untuk kelas tertentu (*prior*)

Untuk tahapan ini dilanjutkan dengan penghitungan pada data latih yang berfokus priornya, lalu kemudian untuk memperoleh nilai peluang dilakukan dengan cara membagi jumlah dokumen dengan kategori tertentu tentang jumlah seluruh dokumen.

### b. Menghitung Peluang Kata

Untuk tahapan selanjutnya diterapkan perhitungan peluang untuk setiap kata yang berada di dokumen uji dan selanjutnya akan masuk ke dokumen latih.

## 2.10 Confusion Matrix

*Confusion matrix* adalah teknik yang digunakan untuk merangkum kinerja algoritma klasifikasi. Matriks ini berukuran  $n \times n$  dan menampilkan klasifikasi yang diprediksi dan *actual*, di mana  $n$  merupakan jumlah kelas yang berbeda [11]. Untuk mendapatkan akurasi maka diperlukan *confusion matrix* dari data yang dihasilkan oleh rincian klasifikasi yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Penjelasan dari TP, TN, FP, dan FN akan dijelaskan di bawah ini:

<i>True Positive</i> (TP)	:	Total data positif yang diklasifikasikan benar melalui sistem
<i>False Positive</i> (FP)	:	Total data positif yang diklasifikasikan salah melalui sistem
<i>False Negative</i> (FN)	:	Total data negatif yang diklasifikasikan salah melalui sistem
<i>True Negative</i> (TN)	:	Total data negatif yang diklasifikasikan benar melalui sistem

Di bawah ini dapat dilihat persamaan 1, 2, dan 3 berisikan akurasi, presisi, dan *recall*.

### a. Akurasi

Akurasi adalah total keseluruhan prediksi yang benar dari seluruh data, baik negatif maupun positif [12].

### b. Presisi

Presisi adalah total keseluruhan prediksi yang benar dari seluruh prediksi yang dilakukan. Presisi dibagi menjadi dua perhitungan yaitu presisi positif dan presisi negatif [12].

### c. Recall

*Recall* adalah total keseluruhan prediksi yang benar dari seluruh data *actual*. *Recall* ini terbagi menjadi dua yaitu *recall* positif dan *recall* negatif [12].

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

$$\text{Presisi Positif} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Presisi Negatif} = \frac{TN}{TN + FP}$$

$$\text{Recall Positif} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Recall Negatif} = \frac{TN}{TN + FN}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Lingkungan Percobaan

Untuk menjalankan sistem analisis pada penelitian ini maka dibutuhkan spesifikasi tertentu pada perangkat keras dan perangkat lunak agar berjalan dengan maksimal. Spesifikasi ini untuk mendukung kinerja sistem yang telah dirancang dapat berjalan dengan baik dan menjalankan fungsinya sesuai dengan tahapan-tahapan yang sudah dirancang.

#### 3.1.1 Spesifikasi Perangkat Keras

Perangkat keras atau *hardware* yang dipakai untuk membuat sistem ini dapat diamati pada tabel 3 di bawah ini:

**Tabel 3.** Spesifikasi Perangkat Keras

Nama Perangkat Keras	Detail Spesifikasi
Laptop	Legion 5
Processor	AMD Ryzen 7 5800H
Random Access Memory	16GB DDR4
Graphics Processing Unit	NVIDIA GeForce 3050 Ti

Solid State Drive	500GB
Display	15.6 FHD 165Hz

### 3.1.2 Spesifikasi Perangkat Lunak

Perangkat lunak atau *software* yang dipakai untuk membuat sistem ini dapat diamati pada tabel 4 di bawah ini:

**Tabel 4.** Spesifikasi Perangkat Lunak

Nama Perangkat Lunak	Detail Spesifikasi
Microsoft Windows	Microsoft Windows 11 Home SL
MariaDB	Versi 10.4.32
Microsoft Visual Studio Code	Versi 1.89.0
Microsoft Office	Microsoft Office 365
Python	Versi 3.10.2
Flask	Versi 2.2.2
Google Chrome	Versi 126.0.6478.114

### 3.2 Implementasi Metode

Pada sistem analisis sentimen yang dibuat oleh peneliti memiliki lima struktur. Adapun struktur tersebut dikaji secara berurutan, struktur tersebut terdiri dari tahap pengumpulan data, impor data ulasan, pelabelan manual, *preprocessing*, dan tahap *multinomial naïve bayes*.

#### 3.2.1 Pengumpulan Data

Pada tahap ini merupakan tahapan pertama untuk memastikan dataset yang berasal dari aplikasi Merdeka Mengajar. Untuk dataset ini diperoleh dari tanggal 20 Mei – 1 Juni 2024 dengan memanggil id aplikasi Merdeka Mengajar yaitu *id.belajar.app*. Selanjutnya hasil *scraping* akan disimpan dengan format *excel*. Untuk dataset ini akan diimpor ke sistem *database* untuk dilakukan proses selanjutnya. Di bawah ini merupakan sampel data dari aplikasi Merdeka Mengajar yang dapat diamati pada tabel 5:

**Tabel 5.** Sampel Data Ulasan Aplikasi Merdeka Mengajar

No.	CREATED_AT	USERNAME	REVIEWS
1.	2024-05-08 11:38:09	Abdul Aditya	Mantap
2.	2024-04-19 11:57:11	aafaisal kasep	baik tuk meningkatkan kualitas guru terutama
3.	2024-05-06 22:38:42	Abasir 13	aplikasi ini sangat membantu sekali dalam pembelajaran
4.	2024-05-08 07:21:20	abd hannan	Sangat membantu sekali utk pengembangan kompetensi guru
5.	2023-12-20 19:38:51	Adik Sa	Intinya begini, buka konten, gaperlu dilihat videonya gpp.yg penting centang, asal isi gpp yg penting terisi, aksi nyata copas ganti nama aja gpp, sertifikat keluar. Siapa juga yg mau koreksi ribuan aksi nyata. tender proyek terooooosh. Sesimpel itu.
6.	2024-01-20 12:44:50	Abdul Qodir	Kurikulum merdeka belajar merupakan kurikulum yg sangat baik dan bnyk inspiratif
7.	2024-01-30 11:46:41	Abdul Rohim	pasilitas buat para guru yg paling lengkap
8.	2024-05-04 11:13:31	Abdulumtolip Tolip	oke
9.	2024-05-03 14:09:20	Abi Habiburrahman	Sangat membantu dan mudah diakses
10.	2024-01-05 12:06:46	Abubakar sidik	selama tahun 2023 banyak hal yang saya dapat dan banyak juga cobaan yang melanda diri saya oleh karena butuh waktu dalam proses untuk mencapai suatu keberhasilan. pelajarilah hal yang menjadi suatu ujian atau cobaan karena ada hikmah yang diambil dibalik itu semua. Raihlah apayang menjadi cita-cita sebagai pedoman hidup siapa tahu 2024 akan membawa ke jalan kesuksesan.aamiin.

#### 3.2.2 Impor Data Ulasan

Untuk Langkah selanjutnya adalah tahap impor data ulasan, tahap ini dilakukan dengan cara mengklik tombol “Impor Data” memasukan *file* dengan *format excel* ke dalam sistem analisis sentimen dengan cara

mengklik tombol “Impor Data” yang berada di dalam sistem analisis sentimen. Ketika data sudah diimpor maka akan muncul tampilan data ulasan yang berisi tanggal dibuatnya ulasan, nama pembuat ulasan, dan isi ulasan.

### 3.2.3 Pelabelan Manual

Untuk langkah selanjutnya adalah tahap pelabelan ini dilakukan sesudah pengumpulan data yang diperoleh dari proses *scraping*. Tahapan ini akan dilakukan di dalam sistem yang dimana mekanismenya memilih di tombol *dropdown* untuk ulasan yang akan diberikan label positif maupun label negatif. Tahapan pelabelan tersebut dapat diamati pada tabel 6:

**Tabel 6.** Pelabelan Untuk Ulasan Aplikasi Merdeka Mengajar

No.	CREATED_AT	USERNAME	RAW REVIEWS	LABEL
1.	2024-04-19 11:57:11	aafaisal kasep	baik tuk meningkatkan kualitas guru terutama Mantap	Positif
2.	2024-05-08 11:38:09	Abdul Aditya		
3.	2024-05-06 22:38:42	Abasir 13	aplikasi ini sangat membantu sekali dalam pembelajaran	Positif
4.	2024-05-08 07:21:20	abd hannan	Sangat membantu sekali utk pengembangan kompetensi guru	Positif
5.	2023-12-20 19:38:51	Adik Sa	Intinya begini, buka konten, gaperlu dilihat videonya gpp..yg penting centang, asal isi gpp yg penting terisi, aksi nyata copas ganti nama aja gpp, sertifikat keluar. Siapa juga yg mau koreksi ribuan aksi nyata.. tender proyek terooooosh.. Sesimpel itu.	Positif
6.	2024-01-20 12:44:50	Abdul Qodir	Kurikulum merdeka belajar merupakan kurikulum yg sangat baik dan bnyk inspiratif	Positif
7.	2024-01-30 11:46:41	Abdul Rohim	pasilitas buat para guru yg paling lengkap	Positif
8.	2024-05-04 11:13:31	Abdulmuntolip Tolip	oke	Positif
9.	2024-05-03 14:09:20	Abi Habiburrahman	Sangat membantu dan mudah diakses	Positif
10.	2024-05-07 01:18:38	Ade Fiaza	Pmm Lebih Leluasa untuk berinovasi dan berkreasi	Positif

### 3.2.4 Preprocessing

Sesudah melalui tahapan pelabelan manual, maka tahap selanjutnya adalah *preprocessing*. Tahapan *preprocessing* dapat diamati pada tabel 7 di bawah ini:

**Tabel 7.** Tabel *Preprocessing*

Tahap	Sebelum	Sesudah
<i>Case Folding</i>	PMM sangat membantu pendidik dalam meningkatkan kompetensinya, terimakasih PMM.	pmm sangat membantu pendidik dalam meningkatkan kompetensinya, terimakasih pmm.
<i>Cleaning</i>	di web sama di aplikasi itemnya kok nggak sama ya? apa harus update?	di web sama di aplikasi itemnya kok nggak sama ya apa harus update
<i>Mengganti Slang Word</i>	baik tuk meningkatkan kualitas guru terutama	baik untuk meningkatkan kualitas guru terutama
<i>Menghapus Stop Word</i>	perlu update lagi biar tampilan seperi web	perlu memperbarui lagi agar tampilan seperi web
<i>Stemming</i>	aplikasi ini sangat membantu sekali dalam pembelajaran	aplikasi ini sangat bantu sekali dalam ajar

### 3.2.5 Pengujian Klasifikasi *Multinomial Naïve Bayes*

Pada pengujian ini akan menggunakan rasio 9:1 yang akan dijelaskan dan dapat diamati tabel 8 di bawah ini:

**Tabel 8.** Sampel Pengujian

No.	Reviews	Label(Actual)	Label(Predict)
1.	Dengan adanya PMM ini. sangat memudah kan guru untuk menambah ilmunya baik itu untuk belajar secara mandiri.. dan pada pmm ini banyak isi yang membuat guru memudahkan membuat perangkat ajar misalnya cara cara membuat modul ajar...	Positif	Positif
2.	Keren, karena dapat meningkatkan kompetensi guru dalam proses	Positif	Positif



- pembelajaran serta meningkatkan pembelajaran sehingga pembelajaran menjadi lebih variatif, efektif, dan efisien dan menyenangkan.
3. saya lebih puas karna adanya perangkat ajar ini membantu semua guru agar dapat bertindak dalam hal cara memecahkan masalah dalam hal mengajar murid dan membuat laporan Positif Positif

Selanjutnya pada data ulasan tersebut mendapatkan hasil *confusion matrix* yang dapat diamati pada tabel 9 di bawah ini:

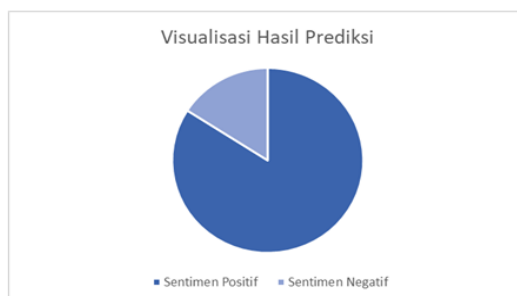
**Tabel 9.** *Confusion Matrix*

TP( <i>True Positive</i> )	84
TN ( <i>True Negative</i> )	14
FP( <i>False Positive</i> )	4
FN( <i>False Negative</i> )	11
Total	113

Selanjutnya akan dilakukan perhitungan akurasi, presisi, dan *recall*. Dapat diamati di bawah ini:

$$\begin{aligned}
 \text{Akurasi} &= \frac{84 + 14}{84 + 14 + 4 + 11} \times 100 \\
 &= \frac{98}{113} \times 100 \\
 &= 0,8672566372 \times 100 \\
 &= 86,72566372 \\
 &= 87\% \\
 \text{Presisi} &= \left( \frac{84}{84 + 4} \right) + \left( \frac{14}{14 + 11} \right) / 2 \times 100 \\
 &= \frac{84}{88} + \frac{14}{25} / 2 \times 100 \\
 &= 0,95 + 0,56 / 2 \times 100 \\
 &= 0,755 \times 100 \\
 &= 76\% \\
 \text{Recall} &= \left( \frac{84}{84 + 11} \right) + \left( \frac{14}{14 + 4} \right) / 2 \times 100 \\
 &= \left( \frac{84}{95} \right) + \left( \frac{14}{18} \right) / 2 \times 100 \\
 &= 0,88 + 0,78 / 2 \times 100 \\
 &= 0,83 \times 100 \\
 &= 83\%
 \end{aligned}$$

Dari yang sudah tertera pada tabel 9 *confusion matrix* maka akan dibuat ilustrasi visualisasi hasil prediksi dengan menggunakan diagram lingkaran (*pie chart*). Diagram lingkaran dapat diamati pada gambar 4 di bawah ini:



**Gambar 4.** Diagram Lingkaran Visualisasi Hasil Prediksi

#### 4. KESIMPULAN

Berdasarkan hasil kajian spesifik dari analisis sentimen aplikasi Merdeka Mengajar dapat ditarik kesimpulan yang dapat diamati di bawah ini:

Sistem ini mampu mengoperasikan analisis sentimen terhadap aplikasi Merdeka Mengajar yang berada di *Google Play Store*. Dengan menggunakan metode *Multinomial Naïve Bayes* untuk analisis sentimen mendapatkan hasil yang optimal dan menghasilkan akurasi dengan persentase 87% menggunakan 1127 dataset, dibagi menjadi dua yaitu data latih dan data uji dengan pembagian rasio 90%:10%. Berdasarkan hasil prediksi dengan menggunakan rasio 90%:10%, sentimen positif lebih dominan daripada sentimen negatif terkait aplikasi Merdeka Mengajar. Ini menunjukkan bahwa pengguna aplikasi Merdeka Mengajar mendapatkan pengalaman baik pada periode 20 Mei – 1 Juni 2024. Hasil akhir yang didapat dari penelitian ini adalah persentase akurasi 87%, presisi 76%, dan *recall* 83% dengan menggunakan rasio 90%:10%.

Mengenai saran yang dapat peneliti sampaikan sebagai pengembangan lebih lanjut pada sistem ini agar dapat berjalan dengan maksimal dan sempurna dengan fungsi yang lebih baik. Saran tersebut dapat diamati di bawah ini:

Pada halaman *login* diberikan fitur lupa *password* agar pengguna yang lupa bisa diberikan pilihan akses agar lebih mudah untuk membuat *password* baru atau mengubah *password* baru. Menambahkan menu untuk *scraping* data ulasan pada sistem analisis sentimen. Melakukan pelabelan secara manual dengan dukungan pakar atau ahli di bidang tata Bahasa Indonesia.

## DAFTAR PUSTAKA

- [1] E. Haerani and F. Syafrina, "Analisis Sentimen Masyarakat Terhadap Kenaikan Biaya Haji Tahun 2023 Menggunakan Metode Naïve Bayes Classifier," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1578–1584, 2023, doi: 10.30865/klik.v4i3.1457.
- [2] A. Adi Susanto, M. Syafrullah, and R. Pradana, "Analisis Sentimen Penerapan Kurikulum Merdeka Pada Twitter Dengan Metode Naive Bayes," *Prosiding Seminar Nasional Fakultas Teknologi Informasi (SENAFTI)*, vol. 2, no. 1, pp. 227-234, 2023.
- [3] F. Hadaina and U. Budiyanto, "Implementasi Metode Multinomial Naïve Bayes Untuk Sentiment Analysis Terhadap Data Ulasan Produk Colearn Pada Google Play Store Implementation Of Multinomial Naive Bayes Method For Sentiment Analysis of Colearn Product Review Data On Google Play Store," *Prosiding Seminar Nasional Fakultas Teknologi Informasi (SENAFTI)*, vol. 1, no. 1, pp. 660-666, 2022. [Online]. Available: <https://senafti.budiluhur.ac.id/index.php>
- [4] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 4, pp. 148–155, 2022, [Online]. Available: <https://djournals.com/klik>
- [5] R. Wahyudi *et al.*, "Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *Jurnal Informatika*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [6] D. Wijaya, R. A. Saputra, and F. Irwiensyah, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 4, pp. 2369-2380, 2024, doi: 10.30865/klik.v4i4.1738.
- [7] A. Yoga Pratama *et al.*, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J-SAKTI: Jurnal Sains Komputer dan Informatika*, vol. 5, no. 2, pp. 897-910, 2021.
- [8] M. Priandi, P. Painem, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor," *Prosiding SENAMIKA: Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, vol. 2, no. 2, pp. 311-319, 2021.
- [9] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA Jurnal*, vol. 10, no. 2, pp. 71-76, 2020.
- [10] F. Rahutomo, A. Retno, and T. H. Ririd, "Evaluasi Daftar Stopword Bahasa Indonesia," *Jurnal RESTI*, vol. 6, no. 1, pp. 41–48, 2019, doi: 10.25126/jtiik.201861226.
- [11] E. Hasibuan and E. A. Heriyanto, "Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier," *JTS: Jurnal Teknik dan Science*, vol. 1, no. 3, pp. 13-24, 2022.
- [12] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix," *Jurnal Informatika UPGRIS*, vol. 6, no. 2, pp. 66-75, 2020.