

**ANALISIS SENTIMEN KEPUASAN PELANGGAN PERUSAHAAN
TELEKOMUNIKASI SELULER TELKOMSEL DI TWITTER**

TUGAS AKHIR

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
Program Studi Informatika



Disusun oleh:

MELIA HAERUNNISSA
182102022

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN TEKNOLOGI INFORMASI
UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA
2022**

HALAMAN PENGESAHAN

TUGAS AKHIR

ANALISIS SENTIMEN KEPUASAN PELANGGAN PERUSAHAAN TELEKOMUNIKASI SELULER TELKOMSEL DI TWITTER

Diajukan oleh:

MELIA HAERUNNISSA

182102022

Telah dipertahankan di depan Dewan Penguji dan dinyatakan sah
sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta

Tanggal: 29 Agustus 2022

Mengesahkan:

Pembimbing I




Agung Priyanto, S.T., M.Eng.
NIDN: 0008107301

Pembimbing II



Choerun Ashawi, S.Kom., M.Kom
NIDN: 0011077702

Penguji I



Andika Bayu Saputra, S.Kom., M.Kom
NIDN: 0521018603

Penguji II



Dayat Subekti, S.Si., M.Kom
NIDN: 0507037401

Ketua Program Studi Informatika
Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta



Puji Winar Cahyo, S.Kom., M.Cs.
NPP: 2018.13.0109

HALAMAN PERNYATAAN

Saya yang bertanda tangan di bawah ini, adalah mahasiswa Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta,

Nama : Melia Haerunnissa
NPM : 182102022
Program Studi : Informatika (S-1)
Judul Tugas Akhir : Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel Di Twitter.

Menyatakan bahwa hasil penelitian dengan judul tersebut di atas adalah asli karya saya sendiri dan bukan hasil plagiarisme. Semua referensi dan sumber terkait yang dikutip dalam karya ilmiah ini telah ditulis sesuai kaidah penulisan ilmiah yang berlaku. Dengan ini, saya menyatakan untuk menyerahkan hak cipta penelitian kepada Universitas Jenderal Achmad Yani Yogyakarta guna kepentingan ilmiah.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya tanpa ada paksaan dari pihak mana pun. Apabila terdapat kekeliruan atau ditemukan adanya pelanggaran akademik di kemudian hari, maka saya bersedia menerima konsekuensi yang berlaku sesuai ketentuan akademik.

Yogyakarta, 22 Agustus 2022



Melia Haerunnissa

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul: “Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel Di Twitter”. Penyusunan laporan ini merupakan salah satu persyaratan untuk menyelesaikan studi di Program Studi Informatika (S-1) Fakultas Teknik & Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta. Laporan ini dapat diselesaikan atas bimbingan, arahan, dan bantuan dari berbagai pihak. Pada kesempatan ini penulis dengan rendah hati mengucapkan terima kasih dengan setulus-tulusnya kepada:

1. Bapak Aris Wahyu Murdiyanto, S.Kom., M.Cs. selaku Dekan Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
2. Bapak Puji Winar Cahyo, S.Kom., M.Cs. selaku Ketua Program Studi Informatika (S-1) Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
3. Bapak Agung Priyanto, S.T., M.Eng. dan Choerun Asnawi, S.Kom., M.Kom. selaku Dosen Pembimbing Tugas Akhir;
4. Para dosen yang telah memberikan banyak bekal ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
5. Ayah, ibu, dan adik saya Rizky Dwi Waluyo yang telah memberikan dukungan semangat serta doa restu kepada saya, sehingga dapat menyelesaikan studi saya;
6. Rekan-rekan mahasiswa Informatika (S-1) di Universitas Jenderal Achmad Yani Yogyakarta yang sudah memberi dukungan dan kerja sama selama pembuatan tugas akhir.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kata sempurna. Maka dari itu dengan segala kerendahan hati penulis sangat

menghargai adanya kritik dan saran yang membangun dari semua pihak yang bersedia meluangkan waktu untuk membaca laporan tugas akhir ini.

Yogyakarta, 22 Agustus 2022

A handwritten signature in black ink, featuring a large, stylized 'M' followed by 'ha' and a horizontal line.

Melia Haerunnissa

DAFTAR ISI

Halaman Judul	i
Halaman Pengesahan.....	ii
Halaman Pernyataan	iii
Kata Pengantar	iv
Daftar Isi	vi
Daftar Tabel.....	ix
Daftar Gambar	x
Daftar Lampiran	xi
Daftar Singkatan	xii
Intisari	xiii
<i>Abstract</i>	xiv
Bab 1 Pendahuluan	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Pertanyaan Penelitian	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Hasil Penelitian	4
Bab 2 Tinjauan Pustaka dan Landasan Teori.....	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori.....	8
2.2.1 Twitter	8
2.2.2 Text Mining.....	9
2.2.3 Python.....	9
2.2.4 Flask	9
2.2.5 Analisis Sentimen.....	10
2.2.6 Model Evaluasi & Classifier	10
2.2.7 Naive Bayes Classification.....	11
Bab 3 Metode Penelitian	12

3.1	Bahan dan Alat Penelitian	12
3.2	Jalan Penelitian.....	13
3.3	Desain Interface	15
3.3.1	Dashboard Analisis Sentimen	15
3.3.2	<i>Preprocessing</i>	16
3.3.3	Pelabelan Manual	16
3.3.4	<i>Training</i>	17
3.3.5	<i>Testing</i>	17
3.3.6	Klasifikasi Naive Bayes	18
Bab 4	Hasil Penelitian.....	19
4.1	Ringkasan Hasil Penelitian	19
4.2	Pengambilan Data Twitter.....	19
4.2.1	Autentifikasi API Twitter	19
4.2.2	Pengumpulan Data Twitter.....	20
4.2.3	<i>Preprocessing</i>	21
a.	Case Folding	22
b.	Number Removal	22
c.	Punctuation Removal	23
d.	Whitespaces Removal	23
e.	Tokenize	23
f.	Stopword Removal.....	23
g.	Normalisasi	23
h.	Stemming	24
4.2.4	Pelabelan manual.....	24
4.2.5	<i>Training</i>	25
4.2.6	<i>Testing</i>	31
4.2.7	Hasil Model Klasifikasi.....	32
4.2.8	Hasil Evaluasi Klasifikasi	33
4.3	Implementasi Disain Interface	34
4.3.1	Halaman Dashboard Analisis Sentimen	34
4.3.2	Halaman Data Twitter	35

4.3.3	Halaman <i>Preprocessing</i> Data.....	36
4.3.4	Halaman Pelabelan Manual.....	38
4.3.5	Halaman <i>Training</i> Data.....	39
4.3.6	Halaman <i>Testing</i> Data	42
4.3.7	Halaman Klasifikasi Naive Bayes Classification.....	43
4.4	Pembahasan.....	45
Bab 5	Kesimpulan dan Saran	47
5.1	Kesimpulan	47
5.2	Saran.....	47
	Daftar pustaka.....	48
	Lampiran	50

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	6
Tabel 3.1 Data Tweet	21
Tabel 3.2 Data <i>Tweet</i> Hasil <i>Preprocessing</i>	22
Tabel 3.3 Data Normalisasi	23
Tabel 3.4 Tabel Pelabelan Manual	24
Tabel 3.5 Data <i>Tweet</i> TF-IDF	25
Tabel 3.6 Perhitungan <i>Term Frequency</i> (TF).....	26
Tabel 3.7 Perhitungan <i>Invers Document Frequency</i> (IDF)	28
Tabel 3.8 Perhitungan (TF-IDF).....	30
Tabel 3.9 Hasil Confusion Matrix Data <i>Training</i>	32
Tabel 3.10 Hasil Klasifikasi Data <i>Training</i>	32
Tabel 3.11 Hasil Klasifikasi Data <i>Testing</i>	33
Tabel 3.12 Hasil <i>Confusion Matrix</i> Data <i>Testing</i>	33

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	13
Gambar 3.2 Desain <i>Interface</i> Halaman Dasboard	15
Gambar 3.3 Desain <i>Interface</i> Halaman <i>Preprocessing</i> Data.....	16
Gambar 3.4 Desain <i>Interface</i> Halaman Pelabelan Manual.....	16
Gambar 3.5 Desain <i>Interface</i> Halaman <i>Training</i>	17
Gambar 3.6 Desain <i>Interface</i> Halaman <i>Testing</i>	18
Gambar 3.7 Desain <i>Interface</i> Klasifikasi Naive Bayes	18
Gambar 3.8 Tampilan Autentifikasi API Twitter	20
Gambar 3.9 Library Pengambilan Data Twitter	20
Gambar 4.1 Implementasi Halaman Dashboard Confusion Matrix	34
Gambar 4.2 Implementasi Halaman Dashboard Grafik	35
Gambar 4.3 Implementasi Halaman Data Tweet.....	36
Gambar 4.4 Implementasi Halaman <i>Preprocessing</i> Data	37
Gambar 4.5 Implementasi Halaman Pelabelan Manual	38
Gambar 4.6 Implementasi Halaman Training Data.....	40
Gambar 4.7 Pembuatan Model Klasifikasi.....	41
Gambar 4.8 Implementasi Halaman <i>Testing</i>	42
Gambar 4.9 Implementasi Halaman Klasifikasi Naive Bayes Classification	44

DAFTAR LAMPIRAN

Lampiran 1 Jadwal Penelitian	50
Lampiran 2 Lembar Bimbingan Dosen	51
Lampiran 3 Hasil Cek Plagiarisme.....	52

DAFTAR SINGKATAN

HTML	Hypertext Markup Language
CSS	Cascading Stylesheet
URL	<i>Uniform Resource Locator</i>
API	<i>Application Programming Interface</i>
CSV	<i>Comma Separated Value</i>
NLTK	<i>Natural Language Toolkit</i>
TF	<i>Term Frequency</i>
IDF	<i>Invers Document Frequency</i>
TF-IDF	<i>Term Frequency-Invers Document Frequency</i>
TPost	<i>True Positive</i>
TNeg	<i>True Negative</i>
TNet	<i>True Netral</i>
FPost	<i>False Positive</i>
FNeg	<i>False Negative</i>
FNet	<i>False Netral</i>

ANALISIS SENTIMEN KEPUASAN PELANGGAN PERUSAHAAN TELEKOMUNIKASI SELULER TELKOMSEL DI TWITTER

Melia Haerunnissa, Agung Priyanto, Choerun Asnawi

INTISARI

Latar Belakang: Telkomsel adalah operator terbesar di Indonesia yang masih menjadi perusahaan dengan pengguna terbanyak. Dengan bertambahnya jumlah pengguna begitu pula pengikut media sosial Telkomsel memuat banyak data *tweet* baik pro maupun kontra terhadap layanan internet Telkomsel di media sosial Twitter yang kurang maksimal. Data dalam bentuk *tweet*, *re-tweet*, dan komentar dari media sosial Twitter dapat digunakan sebagai bahan penelitian untuk dilakukan analisis dan diidentifikasi kebenarannya agar didapatkan informasi yang sesuai tentang layanan internet yang diberikan berdasarkan *tweet*, *re-tweet*, dan komentar dari pengguna Telkomsel di media sosial Twitter.

Tujuan: Membangun model analisis sentimen tentang Telkomsel untuk mengambil informasi yang relevan berdasarkan *tweet*, *re-tweet*, dan komentar dari pengguna Telkomsel di media sosial Twitter, nantinya dapat digunakan sebagai saran pendukung dalam mengevaluasi produk layanan menjadi lebih baik serta menjadi sarana edukasi untuk para pengguna Telkomsel.

Metode Penelitian: Analisis sentimen dengan kata kunci “Telkomsel” menggunakan data sebanyak 13.239 data *tweet* yang diambil pada periode 20 April-30 Juli. Analisis sentimen memiliki beberapa tahapan untuk proses pengolahan data teks, yaitu pengambilan data, *preprocessing*, *training*, *testing*, klasifikasi dan menampilkannya dalam bentuk grafik *bar* menggunakan bahasa pemrograman Python, dengan *framework* Flask.

Hasil: Berdasarkan hasil penelitian analisis sentimen kepuasan pelanggan perusahaan telekomunikasi seluler Telkomsel di Twitter diperoleh bahwa perhitungan akurasi didapatkan nilai 81,7% untuk data *training*, dan 84% untuk *testing* data.

Kesimpulan: Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel di Twitter dibangun menggunakan bahasa pemrograman Python, *framework* Flask dengan menggunakan metode Naive Bayes Classification, menunjukkan akurasi 81,7% pada data *training* dan 84% pada data *testing*. Berdasarkan hasil klasifikasi Telkomsel memiliki sentimen negatif.

Kata-kunci: Telkomsel, Naive Bayes Classification, WebApp, Analisis Sentimen, Python, Flask.

CUSTOMER SATISFACTION ANALYSIS OF TELKOMSEL CELLULAR TELECOMMUNICATIONS COMPANY ON TWITTER

Melia Haerunnissa, Agung Priyanto, Choerun Asnawi

ABSTRACT

Background: *Telkomsel is the largest operator in Indonesia which is still the company with the most users. With the increase in the number of users as well as social media followers, Telkomsel contains a lot of tweet data, both the pros and cons of Telkomsel's internet service on Twitter social media which is less than optimal. Data in the form of tweets, re-tweets, and comments from Twitter social media can be used as research material for analysis and identification of the truth in order to obtain appropriate information about internet services provided based on tweets, re-tweets, and comments from Telkomsel users in the media Twitter social.*

Objective: *Building a sentiment analysis model about Telkomsel to retrieve relevant information based on tweets, re-tweets, and comments from Telkomsel users on Twitter social media, which can later be used as supporting advice in evaluating service products for the better and as an educational tool for Telkomsel users.*

Method: *Sentiment analysis with the keyword "Telkomsel" uses 13,239 tweets taken from April 20 to July 30. Sentiment analysis has several stages for preprocessing text data, namely data retrieval, preprocessing, training, testing, classification and displaying it in the form of a bar graph using the Python programming language, with the Flask framework.*

Result: *Based on the results of the analysis of customer satisfaction sentiment analysis of the Telkomsel cellular telecommunications company on Twitter, it was found that the accuracy calculation obtained a value of 81,7% for training data, and 84% for testing data.*

Conclusion: *Sentiment Analysis of Customer Satisfaction Telkomsel Cellular Telecommunication Company on Twitter was built using the Python programming language, the Flask framework using the Naive Bayes Classification method, showing an accuracy 81,7% on training data, and 84% on testing data. Based on the classification results, Telkomsel has a negative sentiment.*

Keywords: *Telkomsel, Naive Bayes Classification, WebApp, Sentiment Analysis, Python, Flask.*

BAB 1

PENDAHULUAN

1.1 LATAR BELAKANG

Layanan Internet menjadi salah satu kebutuhan yang sangat penting di era revolusi industri 4.0 dengan pengguna internet di dunia sebanyak 4,95 miliar orang per Januari tahun 2022. Di Indonesia pengguna internet juga mengalami peningkatan dari tahun ke tahun, tercatat pada Maret 2021 mencapai 202,6 juta pengguna dan mengalami peningkatan hingga 2,1 juta orang (Agustini, 2021). Hingga total pengguna per Januari 2022 di Indonesia mencapai 204,7 juta orang, jumlah ini mencangkup 73,7% dari total populasi Indonesia (Pahlevi, 2022). Meningkatnya pengguna internet menjadikan jasa layanan komunikasi sebagai perusahaan yang paling dibutuhkan untuk berkomunikasi dan mengakses informasi.

PT Telekomunikasi Seluler Indonesia sering disebut sebagai Telkomsel merupakan salah satu operator telekomunikasi seluler terbesar di Indonesia dengan pengguna sebanyak 173,5 juta pelanggan pada tahun 2021 (Jatmiko, 2021). Telkomsel masih menjadi salah satu perusahaan operator yang paling banyak digunakan di Indonesia dengan jumlah pengikut media sosial Twitter sebanyak 1.6 juta *followers* dan paling banyak memberikan komentar dibandingkan layanan internet lainnya. Twitter merupakan salah satu media sosial yang saat ini menjadi pilihan masyarakat untuk memperoleh informasi, karena kemudahan dalam mengakses, memberikan opini dan bertukar informasi (Retnasary & Rohman, 2020). Seperti informasi mengenai layanan internet Telkomsel yang kurang maksimal, seperti banyak terkendala sinyal, penggunaan kuota yang tidak sesuai dan ada juga keluhan mengenai pulsa yang tiba-tiba berkurang, sehingga belum memberikan dampak yang baik bagi operator dan pengguna Telkomsel di media sosial Twitter serta dapat menguntungkan pesaing karena komentar negatif yang ada.

Analisis sentimen merupakan klasifikasi untuk mengelompokkan teks berdasarkan opini. Analisis sentimen di media sosial Twitter untuk layanan telekomunikasi seluler Telkomsel menjadi penting dilakukan analisis untuk mengetahui bagaimana respon pengguna terhadap layanan yang diberikan berdasarkan *tweet*, *retweet*, dan komentar-komentar yang ada. Data *tweet* yang ada dikelompokkan berdasarkan nilai sentimen positif, netral atau negatif, sehingga menghasilkan data teks yang dapat dianalisis dan menghasilkan informasi yang akurat (Asror, et al., 2019).

Berdasarkan permasalahan tersebut, dilakukan penelitian berdasarkan analisis sentimen Telkomsel di media sosial Twitter untuk mengetahui respon pengguna terhadap layanan internet. Informasi yang dihasilkan nantinya dapat digunakan sebagai saran pendukung dalam mengevaluasi dan mengembangkan produk layanan menjadi lebih baik serta menjadi sarana edukasi untuk para pengguna Telkomsel. Hal ini penting dilakukan karena berdasarkan penelitian sebelumnya dapat disimpulkan bahwa kualitas layanan mengenai suatu perusahaan memberikan pengaruh besar terhadap loyalitas pelanggan (Abubakar, et al., 2018).

Dalam penelitian ini menggunakan metode Naive Bayes Classification karena memiliki akurasi dan kecepatan yang maksimal dengan data latih yang sedikit untuk mengklasifikasi teks, serta memudahkan dalam menghitung probabilitas yang melibatkan set data *training*. kelebihan pengklasifikasian Naive Bayes Classification dibandingkan metode lainnya adalah sekumpulan probabilitas dapat dengan mudah dihitung dengan menambahkan nilai frekuensi dan kombinasi dari data *training* dan kategori independen tidak bergantung satu sama lain (Damaiyanti, et al., 2021). Penelitian sebelumnya membandingkan pengklasifikasian Naive Bayes Classification dan K-Nearest Neighbor untuk proses klasifikasi teks menggunakan Parameter akurasi *Area Under Curve* (AUC), *CA*, *F1 score*, *Precision* dan *Recall*. Berdasarkan hasil evaluasi metode K-Nearest Neighbor yang dipengaruhi oleh jumlah data latih yang digunakan, pengklasifikasian Naive Bayes Classification lebih unggul dengan akurasi sebesar

82,7%. Pengklasifikasian Naive Bayes Classification dapat mencapai akurasi tinggi pada sebagian besar data latih (Fitriyah & Waliyansyah, 2019).

Penelitian ini mengimplementasikan metode Naive Bayes Classification ke dalam suatu sistem untuk mengklasifikasi data ke dalam sentimen positif, netral dan negatif berdasarkan data dari Twitter tentang Telkomsel. Sistem yang akan dibangun terdiri dari proses *preprocessing*, pelabelan manual, *training*, *testing*, klasifikasi dan visualisasi. Hasil perhitungan dapat divisualisasikan dengan sistem berbasis web yang dibangun dengan bahasa pemrograman Python menggunakan *framework* Flask, dalam sistem ini nantinya akan terdapat menu yang dapat digunakan untuk melakukan proses pengolahan data *tweet* berdasarkan *preprocessing*, *training*, *testing*, klasifikasi dan visualisasi dari hasil analisis sentimen yang didapat.

1.2 PERUMUSAN MASALAH

Berdasarkan uraian dari latar belakang tersebut, maka permasalahan yang dapat diambil adalah banyaknya *tweet*, *re-tweet* dan komentar dari media sosial Twitter tentang Telkomsel mengenai layanan internet yang kurang maksimal. *Tweet* tersebut dapat menjadi bahan penelitian untuk dilakukan analisis sentimen dan diidentifikasi kebenarannya, agar menghasilkan informasi yang sesuai mengenai sentimen pengguna Telkomsel terhadap layanan internet yang diberikan.

1.3 PERTANYAAN PENELITIAN

1. Bagaimana penelitian ini dapat menghasilkan analisis sentimen positif, negatif dan netral pengguna Telkomsel di media sosial Twitter?
2. Bagaimana pengolahan data *tweet* tentang Telkomsel yang didapatkan dari media sosial Twitter?
3. Bagaimana penerapan metode Naive Bayes Classification dalam penelitian ini ?

1.4 TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah membuat suatu model analisis sentimen menggunakan metode Naive Bayes Classification yang dapat digunakan untuk melakukan analisis sentimen media sosial Twitter tentang Telkomsel. Analisis sentimen dilakukan untuk mengetahui respon pengguna mengenai layanan internet. Informasi yang dihasilkan nantinya dapat digunakan sebagai saran pendukung dalam mengevaluasi dan mengembangkan produk layanan menjadi lebih baik serta menjadi sarana edukasi untuk para pengguna Telkomsel.

1.5 MANFAAT HASIL PENELITIAN

Manfaat dari penelitian ini adalah:

1. Memberikan informasi mengenai bagaimana sentimen pengguna telekomunikasi seluler Telkomsel di media sosial Twitter.
2. Informasi yang dihasilkan dapat digunakan sebagai saran pendukung dalam mengevaluasi dan mengembangkan produk layanan menjadi lebih baik.
3. Hasil analisis yang sudah didapat sebagai sarana edukasi untuk para pengguna Telkomsel.

BAB 2

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 TINJAUAN PUSTAKA

Landasan teori yang akan digunakan sebagai dasar dalam mengerjakan penelitian ini berasal dari beberapa jurnal yang terkait dengan penelitian yang hampir sama. Berikut merupakan penjelasan mengenai penelitian sebelumnya:

Pertama, penelitian yang dilakukan oleh Robi Kurniawan dan Aulia Apriliani dengan penelitian yang berjudul "Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper". Dalam penelitian ini membahas tentang pengumpulan data dengan teknik scrapper untuk mengekstraksi data *tweet* mengenai virus corona berdasarkan opini dari Twitter yang bersifat positif, negatif dan netral, sehingga dengan adanya analisis sentimen ini mampu menghasilkan informasi tampilan dashboard tentang persentase *tweet* positif, negatif dan netral agar mengetahui dampak virus corona di Indonesia (Apriliani & Kurniawan, 2020).

Kedua, penelitian yang dilakukan oleh Billy Gunawan, Helen Sasty Pratiwi dan Enda Esyudha Pratama dengan penelitian yang berjudul "Sistem Analisis Sentimen Pada Ulasan Produk Menggunakan Metode Naive Bayes Classification". Dalam penelitian ini membahas tentang sistem analisis sentimen untuk mengelompokkan opini dan *review* secara otomatis berdasarkan sentimen positif, netral dan negatif, sehingga dapat membantu pihak manajemen dan perusahaan agar mengetahui umpan balik terhadap merk dan masyarakat dalam menilai produk berdasarkan opini dan *review* (Gunawan, et al., 2018).

Ketiga, penelitian yang dilakukan oleh Kevin Perdana, Titani Pricilia dan Zulfachmi dengan penelitian yang berjudul "Optimasi TextBlob Menggunakan Support Vector Machine Untuk Analisis Sentimen Layanan Telkomsel". Dalam penelitian ini membahas tentang mengidentifikasi opini dan mengklasifikasi kelas berdasarkan pola dari proses *training*, sehingga dapat melihat meningkatnya nilai

akurasi dari klasifikasi menggunakan TextBlob dengan mengambil data mengenai pelayanan Telkomsel (Perdana, et al., 2021).

Tabel 2.1 Tinjauan Pustaka

No	Nama Penulis	Judul	Tahun	Isi	Perbedaan
1	Robi Kurniawan, Aulia Apriliani	Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari <i>Twitter</i> Berbasis Web Scrapper	2020	Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari <i>Twitter</i> . Analisis sentimen mengambil data dengan teknik scrapper untuk mengekstraksi data <i>tweet</i> .	Analisis Sentimen Kepuasan Pelanggan Perusahaan Jasa Penyedia Telekomunikasi Seluler Di Twitter akan dianalisis menggunakan bahasa pemrograman python dengan memanfaatkan Jupyter Notebook untuk proses analisis.

No	Nama Penulis	Judul	Tahun	Isi	Perbedaan
2	Billy Gunawan, Helen Sasty Pratiwi dan Enda Esyudha Pratama	Sistem Analisis Sentimen Pada Ulasan Produk Menggunakan Metode Naive Bayes Classification	2018	Membangun Sistem Analisis Sentimen secara otomatis untuk mengelompokkan opini dan <i>review</i> ke dalam <i>class</i> positif dan negatif.	Analisis Sentimen Kepuasan Pelanggan Perusahaan Jasa Penyedia Telekomunikasi Seluler pada analisis ini akan dibangun sistem berbasis Python dengan <i>framework</i> Flask untuk menampilkan data <i>training</i> , data <i>testing</i> dan persentase hasil analisis.
3	Kevin Perdana, Titania Pricillia, Zulfachmi	Optimasi TextBlob Menggunakan Support	2021	Pada penelitian ini berfokus pada bagaimana	Metode yang digunakan oleh penelitian

No	Nama Penulis	Judul	Tahun	Isi	Perbedaan
		Vector Machine Untuk Analisis Sentimen Layanan Telkomsel		Mengklasifikasi kelas berdasarkan pola dari proses <i>training</i> agar menghasilkan nilai akurasi meningkat ketika menambah <i>supervised learning</i> dari klasifikasi TextBlob dan dibangun menggunakan metode Support Vector Machine.	Sebelumnya adalah klasifikasi TextBlob menggunakan metode <i>support vector machine</i> sedangkan pada analisis sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler menggunakan metode Naive Bayes Classification.

2.2 LANDASAN TEORI

2.2.1 Twitter

Twitter adalah layanan media sosial yang berkembang pesat karena pengguna dapat berinteraksi melalui pesan dengan pengguna yang dikenal dengan nama *tweet*. Twitter merupakan salah satu layanan media sosial yang menjadi

pilihan masyarakat untuk memperoleh dan berbagi informasi antar pengguna dengan perangkat elektronik (Retnasary & Rohman, 2020).

2.2.2 Text Mining

Text mining adalah proses yang dilakukan untuk mengekstrak sebuah informasi yang tersembunyi agar mampu menghasilkan informasi melalui pemrosesan, pengelompokan dan analisis data dalam jumlah besar tidak terstruktur, informasi yang di ekstrak secara otomatis berupa data teks dari sumber yang berbeda-beda (Abidin, et al., 2021). *Text mining* merupakan teknik yang mempelajari prosedur serta metode untuk mencari dan menemukan kembali informasi yang relevan.

2.2.3 Python

Python (bahasa pemrograman) merupakan bahasa pemrograman tinggi yang bisa melakukan eksekusi sejumlah instruksi multiguna secara langsung (interpretatif) dengan metode *object oriented programming* dan juga menggunakan semantik dinamis untuk memberikan tingkat keterbacaan *syntax* (Rahmadhika & Thantawi, 2021). Sebagai bahasa pemrograman tingkat tinggi Python memiliki instruksi yang mirip dengan bahasa manusia, instruksi atau kode yang digunakan pada Python adalah menggunakan bahasa inggris. Python dikenal sebagai bahasa yang mudah dipelajari karena struktur *syntax* yang rapi serta mudah dipahami atau mudah dibaca.

2.2.4 Flask

Flask adalah *microframework* yang dipelopori oleh Armin Ronacher. Flask jauh lebih ringan dan cepat karena Flask dibuat dengan ide menyederhanakan inti dari *framework* seminimal mungkin. Dengan *tagline* "*web development, one drop at a time*", Flask dapat membantu membuat situs dengan sangat cepat meskipun dengan *library* yang sederhana (Putera & Putra, 2019). Flask merupakan web *framework* dari bahasa pemrograman Python, *framework* ini menyediakan *libraries* dan kumpulan kode yang bisa digunakan untuk membangun sebuah *website* tanpa perlu melakukan semuanya dari awal.

2.2.5 Analisis Sentimen

Analisis sentimen adalah proses menentukan sentimen, memahami, mengelompokkan, mengekstrak dan mengolah data teks secara otomatis untuk mendapatkan informasi yang bersifat positif atau negatif. Analisis sentimen bertujuan untuk menganalisis pendapat dengan suatu topik, produk, layanan sehingga dapat ditentukan berdasarkan kalimat positif dan negatif (Ambiyar, et al., 2021).

2.2.6 Model Evaluasi & Classifier

Confusion matrix merupakan label yang menggambarkan performa dari sebuah algoritma secara spesifik, dilakukan untuk melakukan perhitungan akurasi yang memberikan perbandingan antara hasil klasifikasi oleh sistem atau pemodelan klasifikasi yang sebenarnya pada data. *Confusion matrix* memiliki istilah sebagai representasi untuk hasil proses klasifikasi diantaranya *true positive* (TP), *true negative* (TN), *false positive* (FP) dan *false negative* (FN) (Saputro & Sari, 2019).

Accuracy merupakan nilai yang menggambarkan seberapa akurat sistem dapat mengklasifikasi data secara benar antara data yang terklasifikasi benar dengan keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Precision merupakan nilai yang menggambarkan jumlah data positif yang diklasifikasikan secara benar dengan total data yang diklasifikasi positif.

$$Precision = \frac{TP}{FP + TP} * 100\%$$

Recall menunjukan berapa persen data positif yang terklasifikasi dengan benar oleh sistem.

$$Recall = \frac{TP}{FN + TP} * 100\%$$

F-measure merupakan perhitungan evaluasi untuk menemukan informasi kembali yang mengkombinasikan *precision* dan *recall*.

$$f - measure = \frac{2 * TP}{2 * TP + FP + FN} * 100\%$$

2.2.7 Naive Bayes Classification

Naive Bayes Classification adalah metode pengklasifikasian dengan probabilistik sederhana setiap kategorinya tidak memiliki ketergantungan satu sama lain. Metode Naive Bayes Classification merupakan metode *supervised document classification* yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan dan membutuhkan data *training* sebelum melakukan proses klasifikasi (Devita, et al., 2018).

Term Frequency-Invers Document Frequency (TF-IDF) merupakan metode untuk mencari bobot *term* (kata) dengan menghitung nilai *term frequency* (TF) dan *invers document frequency* (IDF) pada sebuah dokumen. Metode ini memiliki tingkat keakuratan yang tinggi dan menghitung bobot pada setiap *term* (kata) dengan rumus:

$$TF - IDF(t) = TF * \log \frac{N}{df}$$

Keterangan:

t : *term*(kata)

N : total dokumen

df : jumlah dokumen yang termuat pada t

Pada perhitungan klasifikasi menggunakan algoritma Naive Bayes Classification menggunakan data latih yang sudah ditentukan label (kelas) untuk melakukan perhitungan probabilitas untuk setiap kelas data *training* yang akan dilakukan pengujian dengan rumus:

$$P(wk) = \frac{pk + 1n}{|vocab|}$$

Dimana:

$P(wk)$: Probabilitas *term*(kata)

$pk+1$: *query* + angka satu agar hasil perhitungan tidak nol

n : semua jumlah kata yang muncul di dokumen

$|vocab|$: total kata keseluruhan

BAB 3

METODE PENELITIAN

Penelitian ini merupakan penelitian analisis sentimen dengan menggunakan data dari Twitter. Penelitian ini menggunakan metode Naive Bayes Classification. Penelitian dimulai dengan latar belakang permasalahan yang ada, memetakan proses-proses, mencari sumber permasalahan, dan akhirnya menganalisis data mengenai Telkomsel di media sosial Twitter untuk mengatasi permasalahan yang ada. Penelitian ini membutuhkan data *tweet* yang diperoleh melalui media sosial Twitter yang berkaitan dengan Telkomsel. Dibawah ini adalah bahan, alat, metode dan jalannya penelitian analisis sentimen tentang Telkomsel beserta langkah-langkah penelitian untuk menyelesaikan analisis sentimen menggunakan data *tweet* dan merancang sistem untuk menampilkan hasil analisis yang diperoleh.

3.1 BAHAN DAN ALAT PENELITIAN

Penelitian ini akan mengidentifikasi data dan informasi dari berbagai sumber. Sumber utama penelitian ini diamati dari *tweet* terkait Telkomsel di media sosial Twitter, buku tentang analisis sentimen. Selain itu, referensi lain yang terkait dengan penelitian ini diperoleh dari *literature* yang relevan seperti jurnal, penelitian sebelumnya, dan informasi internet.

Alat yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi yang cukup untuk menjalankan sistem operasi dan pengembangan perangkat lunak serta koneksi Internet.

Sistem Operasi dan program aplikasi yang digunakan untuk mengembangkan aplikasi ini adalah:

1. Processor : Intel(R) Celeron(R) CPU 1005M @ 1.90GHz (2CPUs),
~1.9GHz
2. RAM : 10 GB

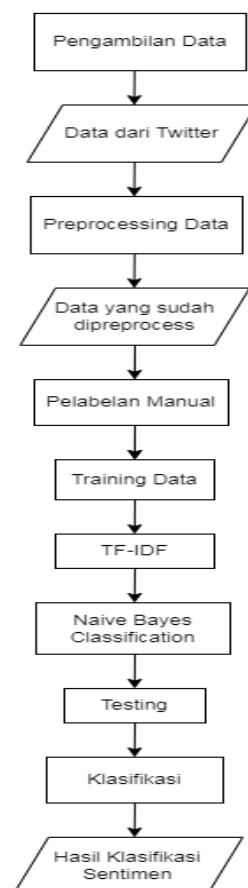
Sistem Operasi dan program-program aplikasi yang dipergunakan dalam pengembangan aplikasi ini adalah :

1. Sistem Operasi : Windows 7

- 2. Bahasa Pemrograman : Python 3.7.3
- 3. Framework : Flask, Bootstrap 4
- 4. Text editor : Sublime Text, Jupyter Notebook

3.2 JALAN PENELITIAN

Naive Bayes Classification merupakan salah satu metode *machine learning* dalam klasifikasi data teks yang memanfaatkan perhitungan probabilitas dan statistik dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset (Devita et al., 2018). Dalam metode ini, semua langkah akan dijalankan selangkah demi selangkah dan dieksekusi secara berurutan. Dibawah ini adalah alur penelitian dan langkah-langkah metode Naive Bayes Classification yang akan digunakan dalam penelitian ini seperti terlihat pada Gambar 3.1 dibawah ini.



Gambar 3.1 Alur Penelitian

1. Tahap Pengambilan data, pada tahap ini dilakukan pengambilan data di media sosial Twitter untuk mendapatkan *tweet* mengenai Telkomsel. Proses ini dilakukan menggunakan Anaconda Prompt dan dieksekusi pada Jupyter Notebook, data *tweet* yang di dapat berformat file *Comma Separated Values* (CSV) sehingga dapat dibuka di Microsoft Excel.
2. Tahap *preprocessing* data, pada tahap ini dilakukan proses pengolahan data teks untuk data yang belum terstruktur, sehingga perlu diperbaiki dengan melakukan tahapan-tahapan berikut ini:
 - a. *Tokenization* merupakan proses untuk pemecahan atau pemisahan kalimat dalam suatu teks yang disebut token.
 - b. *Case folding* merupakan proses untuk mengubah setiap huruf pada *tweet* menjadi *lowercase* atau huruf kecil.
 - c. *Stopword removal* merupakan penghapusan kata yang memiliki informasi rendah pada teks.
 - d. *Number removal* merupakan proses untuk membersihkan data teks dari karakter angka.
 - e. *Punctuation removal* merupakan proses penghapusan karakter pada data teks.
 - f. *Whitespaces removal* merupakan proses menghilangkan spasi dari awal dan di akhir kalimat.
 - g. *Normalization* merupakan standarisasi kata dengan salah eja atau penggunaan bahasa yang tidak biasa.
 - h. *Stemming* merupakan penghilangan infleksi pada kata menjadi bentuk dasarnya.
3. Tahap Pelabelan manual, pada tahap ini dilakukan proses pelabelan manual untuk memberikan sentimen positif, netral atau negatif terhadap kata yang ada pada dokumen agar dapat dianalisis.
4. Tahap *training* data, pada tahap ini dilakukan proses *training data* yang diawali dengan fitur ekstraksi dengan menggunakan *term frequency-invers document frequency* (TF-IDF) pada data teks. Proses *training data* digunakan untuk melakukan klasifikasi sentimen secara otomatis.

5. Tahap *testing*, pada tahap ini dilakukan proses *testing* untuk mengukur dan menentukan seberapa akurat model yang telah dibangun pada tahap *training*, hasil *testing* yang diprediksi kelas dan label akan dihitung dengan cara menghitung *confusion matrix* untuk mengetahui *persentase* dari setiap pengujian.
6. Tahap Klasifikasi, pada tahap ini merupakan proses dimana hasil data *tweet* yang sudah diprediksi kelas dan labelnya secara otomatis akan ditampilkan dalam bentuk dashboard menggunakan *framework* Flask bahasa pemrograman Python.

3.3 DESAIN INTERFACE

Interface merupakan gambaran yang berhubungan langsung dengan pengguna sistem. Perancangan antar muka dibuat agar sistem yang dihasilkan lebih menarik dan mudah dipahami pada saat perancangan. Berikut ini adalah deskripsi atau desain-desain *interface* pada Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel Di Twitter:

3.3.1 Dashboard Analisis Sentimen

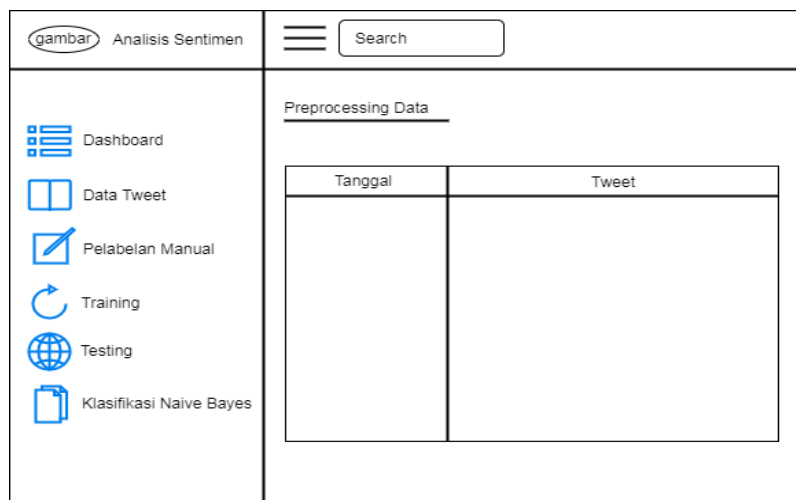
Pada sistem analisis sentimen dapat mengetahui data *tweet* yang telah didapatkan kelas dan labelnya berdasarkan prediksi yang sudah dibuat pada model *training* dan ditampilkan pada menu dashboard. Tampilan menu dashboard ditunjukkan pada Gambar 3.2.



Gambar 3.2 Desain *Interface* Halaman Dashboard

3.3.2 Preprocessing

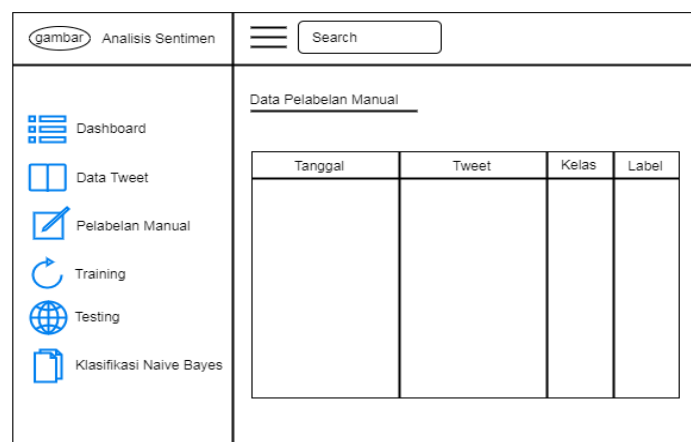
Preprocessing pada sistem analisis sentimen merupakan proses pengolahan data *tweet* yang ada dengan melakukan langkah-langkah untuk menyempurnakan data *tweet* yang masih memiliki komponen sehingga menjadi data yang diinginkan. Tampilan menu *preprocessing* data ditunjukkan pada Gambar 3.3.



Gambar 3.3 Desain *Interface* Halaman *Preprocessing* Data

3.3.3 Pelabelan Manual

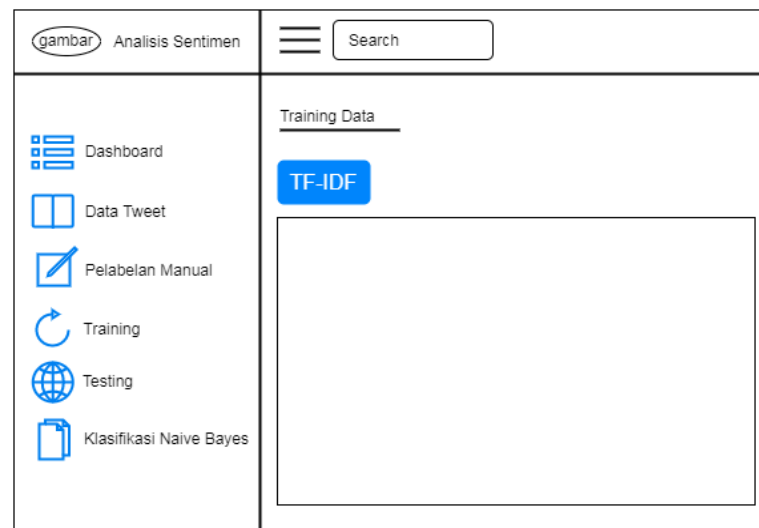
Pelabelan manual pada sistem analisis sentimen merupakan proses untuk menampilkan data *tweet* yang sudah diberi kelas dan label terhadap kalimat yang ada pada data *tweet* secara manual. Proses pelabelan manual dilakukan pada Microsoft Excel. Tampilan menu pelabelan manual ditunjukkan pada Gambar 3.4.



Gambar 3.4 Desain *Interface* Halaman Pelabelan Manual

3.3.4 Training

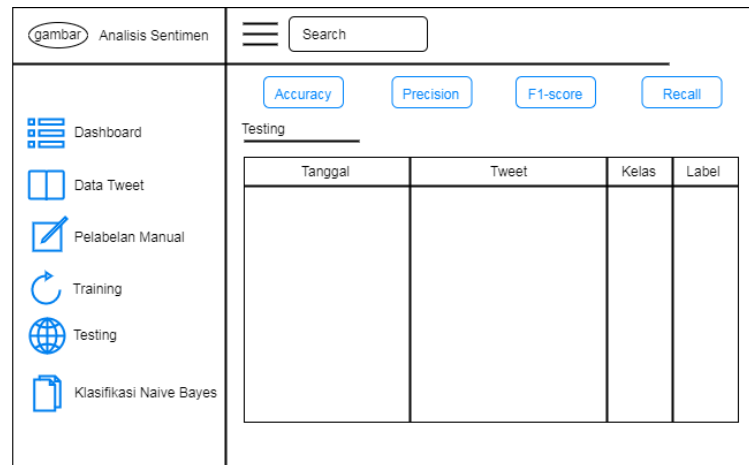
Training pada sistem analisis sentimen menggunakan ekstraksi *Term Frequency-Invers Document Frequency* (TF-IDF) pada data teks untuk menghasilkan model klasifikasi yang dapat digunakan untuk menampilkan proses klasifikasi sentimen secara otomatis. Tampilan menu *training* data ditunjukkan pada Gambar 3.5.



Gambar 3.5 Desain *Interface* Halaman *Training*

3.3.5 Testing

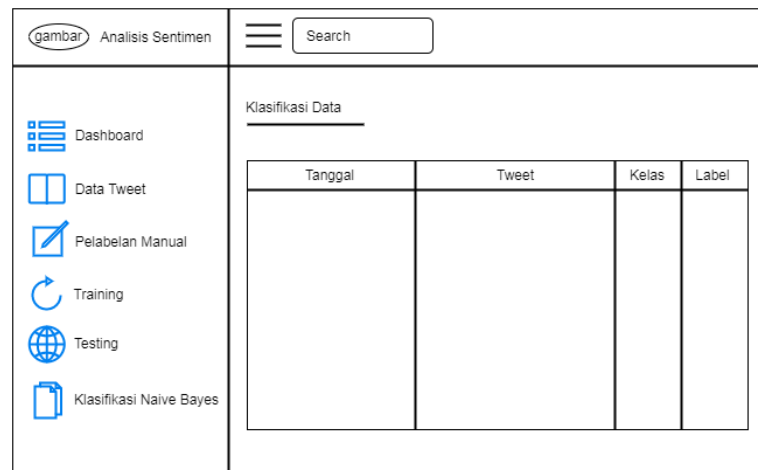
Testing pada sistem analisis sentimen dilakukan untuk mengetahui seberapa akurat model yang dibangun dalam *training* data digunakan untuk memprediksi kelas atau label dari data uji yang tersedia. Tampilan menu *training* data ditunjukkan pada Gambar 3.6.



Gambar 3.6 Desain *Interface* Halaman *Testing*

3.3.6 Klasifikasi Naive Bayes

Klasifikasi naive bayes pada sistem analisis sentimen dilakukan untuk menampilkan data *tweet* yang akan diprediksi kelas, label yang sudah dibangun pada *training* data. Tampilan menu klasifikasi Naive Bayes ditunjukkan pada Gambar 3.7.



Gambar 3.7 Desain *Interface* Klasifikasi Naive Bayes

BAB 4

HASIL PENELITIAN

4.1 RINGKASAN HASIL PENELITIAN

Hasil penelitian analisis sentimen dengan kata kunci “Telkomsel” menggunakan metode Naive Bayes Classification untuk menganalisis data dari *tweet* berbahasa Indonesia baik positif, negatif maupun netral. Data Twitter yang diambil selama periode 20 April-30 Juli 2022 dengan jumlah 13,239 data *tweet* dan *retweet*. Data yang digunakan selama *training* sebanyak 600 data dengan 200 data berlabel positif, 200 data berlabel negatif dan 200 data berlabel netral. Pada proses *testing* menggunakan sebanyak 336 data dengan 112 data berlabel positif, 112 data berlabel negatif dan 112 data berlabel netral. Analisis sentimen media sosial Twitter dibangun untuk mengolah data *tweet* menggunakan bahasa pemrograman Python dengan metode Naive Bayes Classification untuk mendapatkan hasil perhitungan sentimen menggunakan *framework* Flask. Penelitian ini akan memberikan informasi tentang analisis sentimen pengguna Telkomsel di media sosial Twitter tentang Telkomsel. .

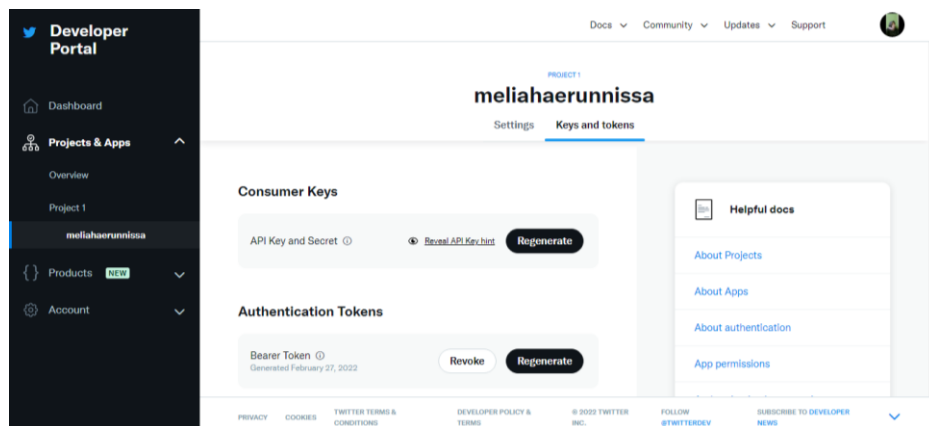
4.2 PENGAMBILAN DATA TWITTER

Pengambilan data dilakukan di media sosial Twitter untuk mendapatkan data *tweet* yang berkaitan dengan kata kunci “Telkomsel”. Data *tweet* yang diambil mulai dari tanggal 20 Mei – 30 Juli 2022 dengan 13.239 data *tweet*, *re-tweet* dan komentar. Proses pengambilan data *tweet* menggunakan API Twitter dilakukan di Anaconda Prompt dan dieksekusi di Jupyter Notebook. Data *tweet* yang dapat diambil berformat file *Comma Separated Values* (CSV) sehingga dapat dibuka di Microsoft Excel.

4.2.1 Autentifikasi API Twitter

Autentikasi API Twitter adalah kerangka kerja untuk secara aman dan ringkas mencapai akses terbatas dengan mengisi data pada *developer portal*

Twitter. *Access token* digunakan untuk mengambil data *tweet* sesuai kebutuhan yang ditetapkan oleh Twitter, setelah mendapatkan *access token*, *access token secret*, *consumer key*, *consumer secret* API berguna untuk mendapatkan data *tweet* untuk pemrosesan data. Tampilan autentifikasi API Twitter ditunjukkan pada Gambar 3.8.



Gambar 3.8 Tampilan Autentifikasi API Twitter

4.2.2 Pengumpulan Data Twitter

Pengumpulan data Twitter dilakukan dengan bahasa pemrograman Python. Penelitian ini menggunakan *library tweepy* dengan melakukan akses pada API Twitter untuk mendapatkan data *tweet* berdasarkan kata kunci “Telkomsel”. Tampilan import *library* ditunjukkan pada Gambar 3.9.

```

1 import tweepy
2 import csv
3 import sys
4 import re
5 import pandas as pd
6 import numpy as np
7 import string
8 import datetime
9 import time

```

Gambar 3.9 Library Pengambilan Data Twitter

Library Python yang digunakan untuk proses pengumpulan data *tweet* kemudian disimpan sebagai file *Comma Separated Values* (CSV). Dalam pengambilan data *tweet* menggunakan API Twitter, data *tweet* yang dihasilkan

memiliki variasi dalam berbagai bentuk, termasuk *hashtag*, tautan URL, angka, username sehingga diperlukan proses *preprocessing* untuk mendapatkan data *tweet* yang diinginkan. *Pandas* memiliki manipulasi dengan pembersihan data dan fungsi *datetime* menjalankan operasi terkait waktu. Contoh data *tweet* yang sudah diambil ditunjukkan pada Tabel 3.1.

Tabel 3.1 Data Tweet

No	Data Tweet
1	b'@scorpeeoss aku pk Telkomsel 25-29k perbulan udh sama Disney+'
2	b'TELKOMSEL MAHAL BGT ANJIR KUOTANYA NANGISSSSSA https://t.co/3FSOI9HPiu
3	b'Telkomsel knp siiii'
4	b'woi @Telkomsel kuota msh ada 10 gb knp lemot yh ini hp w yg lemot ap jaringan nya'
5	b'min @Telkomsel tolong dong ini udah hampir sebulan gini kok jaringan ku lemot terus ya :('
6	b'@Meitryparamita1 Halo Kak Mei. Terkait paket internet infoin nomor Hp nya yuk Kak melalui DM agar Sulis cek dan\xe2\x80\xa6 https://t.co/n4aHNoJE1J
7	b'@Telkomsel Nah kan bener dilempar lempar'
8	b'Semoga besok sinyal telkomsel gak bermasalah : \xe2\x80\x99)
9	b'Ada apa sih telkomsel dan indihome hr ini??? Kok jelek bgt'
10	b'@Felixie04 coba lgsg dtg ke gerai telkomsel aja kayanya bs kalau mau di aktivasi lg'

Tabel 3.1 menunjukan hasil pengambilan data *tweet* yang belum terstruktur dan memiliki variasi bentuk, sehingga diperlukan beberapa *preprocessing* karena masih banyak variasi bentuk yang perlu dihilangkan karena tidak digunakan dalam pembuatan analisis sentimen agar menjadi data yang terstruktur.

4.2.3 Preprocessing

Preprocessing dilakukan untuk proses pengolahan data yang berfungsi memperbaiki data *tweet* yang belum terstruktur dengan melakukan tahapan-

tahapan agar menjadi data yang terstruktur. Contoh data *tweet* yang sudah dilakukan proses *preprocessing* ditunjukkan pada Tabel 3.2.

Tabel 3.2 Data *Tweet* Hasil *Preprocessing*

No	Data Tweet
1	my telkomsel ganggu kah
2	kenapa tidak masuk langgan disney lagi
3	app my telkomsel error mau panjang paket gimana ini
4	siap kak putra kami akan cek dm dari kakak tunggu konfirmasi lanjut di dm ya kak makasih gundi
5	memang lagi ganggu apa gimana sih kok tiba tiba sinyal hilang
6	baik kak jika belum kakak sudah pernah kirim data yang minta belum yuk siapa kami di dm
7	aku dapat bulan khusus disney di telkomsel
8	mati lampu mati juga lah telkomsel
9	sih telkomsel mahal banget kouta nya boros pula
10	maaf banget ya kak jeje jadi tidak nyaman kena keluhan tidak bisa akses apps mytelkomsel yuk infoin nomor

Tabel 3.2 menunjukkan hasil data *tweet* yang sudah dilakukan proses *preprocessing*. Data *tweet* tersebut menjadi lebih terstruktur agar dapat digunakan untuk melakukan perhitungan ditahap selanjutnya. Berikut merupakan tahapan-tahapan dalam proses *preprocessing* data:

a. Case Folding

Case Folding merupakan tahapan yang dilakukan untuk mengubah setiap huruf pada data *tweet* menjadi *lowercase* atau huruf kecil.

b. Number Removal

Number removal merupakan tahapan yang dilakukan untuk membersihkan data dari karakter angka.

c. Punctuation Removal

Punctuation removal merupakan tahapan yang dilakukan untuk penghapusan karakter pada data teks.

d. Whitespaces Removal

Whitespaces removal merupakan langkah yang dilakukan untuk menghilangkan spasi diawal dan akhir kalimat.

e. Tokenize

Tokenizing merupakan tahapan yang dilakukan untuk pemecahan atau pemisahan kalimat dalam suatu teks yang disebut token.

f. Stopword Removal

Stopword removal merupakan tahapan yang dilakukan untuk penghapusan kata yang memiliki informasi rendah pada teks.

g. Normalisasi

Normalization merupakan tahapan yang dilakukan membakukan kata-kata dengan salah eja atau penggunaan bahasa yang tidak biasa. Contoh normalisasi dapat dilihat pada Tabel 3.3.

Tabel 3.3 Data Normalisasi

Before	After
dtg	datang
lgsg	langsung
pk	pakai
jgn	jangan
cmn	hanya
tuku	beli
kmren	kemarin
gbs	tidak bisa
dipake	dipakai

h. Stemming

Stemming merupakan tahapan yang dilakukan untuk penghilangan infleksi pada kata menjadi bentuk dasarnya.

4.2.4 Pelabelan manual

Pelabelan manual merupakan proses memberikan label positif, negatif dan netral terhadap data *tweet* yang sudah dilakukan proses *preprocessing*. Proses pelabelan manual dilakukan di Microsoft Excel secara manual agar data *tweet* dapat dianalisis. Data *tweet* yang sudah diberi label didapatkan 600 data *tweet* yang digunakan untuk *training* dengan data berlabel positif 200 data *tweet*, berlabel netral 200 data *tweet* dan berlabel negatif 200 data *tweet*. Contoh data *tweet* yang sudah dilakukan pelabelan manual ditunjukkan pada Tabel 3.4.

Tabel 3.4 Tabel Pelabelan Manual

No	Tweet	Kelas	Label
1	ngebut banget asli si paling ngebut sudah pokok	Positif	1
2	ada cashback nampol sampai dari mytelkomsel yang paling bisa bagi bagi promo ntransaksi minimal rp ribu di mytelkomsel	Positif	1
3	hai kak maaf ya untuk keluhan akses internet lambat yuk infoin nomor handphone tanggal jadi loka	Netral	0
4	tenang kak gilang kalau yang kakak maksud kendala akses aplikasi mytelkomsel yuk infoin nomor handphone tanggal	Netral	0
5	jelek banget kalau mati lampu sinyal ikut ngilang	Negatif	-1
6	telkomsel memang asa boros banget apa gimana sih pakai bentar doang sudah habis banyak	Negatif	-1
7	hai kak maaf ya jadi tidak nyaman inara akan bantu kendala kakak yuk dm nomor handphone kakak lokasi detail	Netral	0
8	harga naik disney hotstar yang murah banget kalau pakai telkomsel cuma rb bulan	Positif	1
9	telkomsel legal gift dari mytelkomsel proses cepat langgan disney bulan	Positif	1

No	Tweet	Kelas	Label
10	usaha donk buat telkomsel jangan kartu baru mati yang lama juga bisa di bantu agar lebih byk yang guna kayu telkomsel dan puas	Negatif	-1

Tabel 3.4 menunjukkan hasil pelabelan manual bahwa data *tweet* dengan kelas positif diberi label 1, kelas negatif diberi label -1, dan kelas netral diberi label 0. Data pelabelan ini digunakan untuk memberikan nilai sentimen yang dihitung akurasinya.

4.2.5 Training

Training data merupakan proses yang dilakukan untuk menghasilkan model klasifikasi yang secara otomatis digunakan untuk proses klasifikasi, proses *training* menggunakan metode Naive Bayes Classification dengan fungsi ekstraksi TF-IDF pada data teks. Contoh perhitungan TF-IDF secara manual ditunjukkan pada Tabel 3.5.

Tabel 3.5 Data *Tweet* TF-IDF

Dokumen (d)	Tweet
d1	harga naik disney hotstar yang murah banget kalau pakai telkomsel cuma rb bulan
d2	maaf banget ya kak sane jadi tidak nyaman kena keluhan tidak bisa akses apps mytelkomsel yuk infoin nomor handphone
d3	jaring sangat buruk sangat lambat
d4	coba langsung datang ke gerai telkomsel saja kaya bisa kalau mau di aktivasi lg
d5	parah banget memang sekarang enggak kalau jaring tidak dimaksimalin harga paket jangan mahal dong

Tabel 3.5 menjelaskan data *tweet* akan digunakan sebagai contoh perhitungan TF-IDF, perhitungan TF-IDF dilakukan secara manual menggunakan 5 data *tweet* yaitu d1, d2, d3, d4, d5 dan beberapa komponen seperti t yaitu kata, d yaitu kalimat atau data *tweet* untuk mengetahui berapa banyak data *tweet* dimana suatu kata muncul. Perhitungan TF-IDF ditunjukkan pada Tabel 3.6.

Tabel 3.6 Perhitungan *Term Frequency* (TF)

term/kata	d1	d2	d3	d4	d5	df
akses		1				1
apps		1				1
banget	1	1			1	3
bisa		1		1		2
bulan	1					1
buruk			1			1
coba				1		1
cuma	1					1
datang				1		1
diaktivasi				1		1
dimaksimalkan					1	1
disney	1					1
dong					1	1
gerai				1		1
handphone		1				1
harga	1				1	2
hotstar	1					1
infoin		1				1
jangan					1	1
jaring			1		1	2
kak	1					1
kalau	1			1	1	3
kaya				1		1
ke				1		1
keluh		1				1
kena		1				1
lagi				1		1
lambat			1			1
langsung				1		1

term/kata	d1	d2	d3	d4	d5	df
maaf	1					1
mahal					1	1
mau				1		1
memang					1	1
murah	1					1
mytelkomsel		1				1
naik	1					1
nomor		1				1
nyaman		1				1
pakai	1					1
paket					1	1
parah					1	1
ribu	1					1
saja				1		1
sane	1					1
sangat			2			1
sekarang					1	1
telkomsel	1			1		2
tidak		2			2	2
ya	1					1
yang	1					1
yuk		1				1

Tabel 3.6 menjelaskan perhitungan *term frequency* (TF) sehingga setiap kata dari kalimat yang terdapat dalam dokumen akan dilakukan perhitungan *invers document frequency* (IDF). Proses perhitungan tersebut menghitung sebuah kata yang terdapat pada dokumen dengan N banyaknya dokumen menggunakan komponen *term* atau kata, *document frequency* (DF) dan *invers document frequency* (IDF). Perhitungan IDF ditunjukkan pada Tabel 3.7.

Tabel 3.7 Perhitungan *Invers Document Frequency* (IDF)

term/kata	Df	Idf	idf(N=4)	idf(N=1000)
akses	1	1	0,602059991	3
apps	1	1	0,602059991	3
banget	3	0,333333333	0,124938737	2,522878745
bisa	2	0,5	0,301029996	2,698970004
bulan	1	1	0,602059991	3
buruk	1	1	0,602059991	3
coba	1	1	0,602059991	3
cuma	1	1	0,602059991	3
datang	1	1	0,602059991	3
diaktivasi	1	1	0,602059991	3
dimaksimalkan	1	1	0,602059991	3
disney	1	1	0,602059991	3
dong	1	1	0,602059991	3
gerai	1	1	0,602059991	3
handphone	1	1	0,602059991	3
harga	2	0,5	0,301029996	2,698970004
hotstar	1	1	0,602059991	3
infoin	1	1	0,602059991	3
jangan	1	1	0,602059991	3
jaring	2	0,5	0,301029996	2,698970004
kak	1	1	0,602059991	3
kalau	3	0,333333333	0,124938737	2,522878745
kaya	1	1	0,602059991	3
ke	1	1	0,602059991	3
keluh	1	1	0,602059991	3
kena	1	1	0,602059991	3
lagi	1	1	0,602059991	3
lambat	1	1	0,602059991	3
langsung	1	1	0,602059991	3

term/kata	Df	Idf	idf(N=4)	idf(N=1000)
maaf	1	1	0,602059991	3
mahal	1	1	0,602059991	3
mau	1	1	0,602059991	3
memang	1	1	0,602059991	3
murah	1	1	0,602059991	3
mytelkomsel	1	1	0,602059991	3
naik	1	1	0,602059991	3
nomor	1	1	0,602059991	3
nyaman	1	1	0,602059991	3
pakai	1	1	0,602059991	3
paket	1	1	0,602059991	3
parah	1	1	0,602059991	3
ribu	1	1	0,602059991	3
saja	1	1	0,602059991	3
sane	1	1	0,602059991	3
sangat	1	1	0,602059991	3
sekarang	1	1	0,602059991	3
telkomsel	2	0,5	0,301029996	2,698970004
tidak	2	0,5	0,301029996	2,698970004
ya	1	1	0,602059991	3
yang	1	1	0,602059991	3
yuk	1	1	0,602059991	3

Tabel 3.7 menjelaskan perhitungan *invers document frequency* (IDF) dengan rumus $idf = \log \frac{(N)}{(df)}$ untuk mengurangi bobot setiap *term* atau kata dalam kalimat yang terdapat dalam dokumen untuk mengetahui banyaknya *term* atau kata yang muncul. Perhitungan TF-IDF ditunjukkan pada Tabel 3.8.

Tabel 3.8 Perhitungan (TF-IDF)

term/kata	d1	d2	d3	d4	d5
akses		0,60206			
apps		0,60206			
Banget	0,124939	0,124939			0,124939
Bisa		0,30103		0,30103	
Bulan	0,60206				
Buruk			0,60206		
Coba				0,60206	
Cuma	0,60206				
Datang				0,60206	
Diaktivasi				0,60206	
Dimaksimalkan					0,60206
Disney	0,60206				
Dong					0,60206
Gerei				0,60206	
Handphone		0,60206			
Harga	0,30103				0,30103
Hotstar	0,60206				
Infoin		0,60206			
Jangan					0,60206
Jaring			0,30103		0,30103
Kak	0,60206				
Kalau	0,124939			0,124939	0,124939
Kaya				0,60206	
Ke				0,60206	
Keluh		0,60206			
Kena		0,60206			
Lagi				0,60206	
Lambat			0,60206		
Langsung				0,60206	

term/kata	d1	d2	d3	d4	d5
Maaf	0,60206				
Mahal					0,60206
Mau				0,60206	
Memang					0,60206
Murah	0,60206				
Mytelkomsel		0,60206			
Naik	0,60206				
Nomor		0,60206			
Nyaman		0,60206			
Pakai	0,60206				
Paket					0,60206
Parah					0,60206
Ribu	0,60206				
Saja				0,60206	
Sane	0,60206				
Sangat			0,60206		
Sekarang					0,60206
Telkomsel	0,30103			0,30103	
Tidak		0,30103			0,30103
Ya	0,60206				
Yang	0,60206				
Yuk		0,60206			

Tabel 3.8 menjelaskan perhitungan *term frequency-invers document frequency* (TF-IDF) dengan perhitungan secara manual.

4.2.6 Testing

Testing merupakan proses yang digunakan untuk memprediksi kelas, label berdasarkan data *training* yang sudah dibangun untuk mengetahui tingkat keakuratan pemodelan.

4.2.7 Hasil Model Klasifikasi

Melakukan evaluasi model klasifikasi data *training* menggunakan perhitungan *confusion matrix* untuk menemukan nilai aktual dan nilai prediksi berdasarkan akurasi yang diberikan oleh aplikasi. Penelitian ini menggunakan *multiclass confusion matrix* karena memiliki 3 kelas sentimen, yaitu positif, negatif dan netral. Data *training* yang akan dihitung adalah 600 data *tweet* yang diberi label secara manual sebagai positif, negatif dan netral. *Multiclass confusion matrix* memiliki istilah yang mewakili hasil dari proses klasifikasi diantaranya *true positive* (TPos), *true negative* (TNeg), *true netral* (TNet), *false positive* (FPos) dan *false negative* (FNeg), *false netral* (FNet). Hasil perhitungan *multiclass confusion matrix* pada data *training* ditunjukkan pada Tabel 3.9.

Tabel 3.9 Hasil Confusion Matrix Data *Training*

Kelas Aktual	Kelas Prediksi		
	Positif	Negatif	Netral
Positif	30	4	7
Negatif	1	37	5
Netral	2	3	31

Tabel 3.9 menjelaskan bahwa hasil perhitungan menggunakan *confusion matrix* didapatkan nilai berupa TPos = 30, FPosNeg = 4, FPosNet = 7, FNegPost = 1, TNeg = 37, FNegNet = 5 dan FNetPos = 2, FNegNet = 3, TNet = 31 maka dilanjutkan perhitungan klasifikasi. Hasil klasifikasi data *training* dapat dilihat pada Tabel 3.10.

Tabel 3.10 Hasil Klasifikasi Data *Training*

Accuracy	0,82
F1-Score	0,82
Precision	0,83
Recall	0,82

4.2.8 Hasil Evaluasi Klasifikasi

Hasil evaluasi klasifikasi untuk data *testing* dilakukan dengan menggunakan perhitungan *confusion matrix* untuk menghitung nilai akurasi untuk menentukan perbedaan antara data *training* dan *testing*. Data *testing* akan dihitung sebanyak 336 data *tweet* yang telah diberi kelas positif, negatif dan netral masing-masing 112 label data *tweet*. Hasil perhitungan klasifikasi pada data *testing* ditunjukkan pada Tabel 3.11.

Tabel 3.11 Hasil Klasifikasi Data *Testing*

Accuracy	0,84
F1-Score	0,84
Precision	0,84
Recall	0,83

Tabel 3.11 menunjukkan hasil perhitungan klasifikasi memiliki akurasi yang cukup baik yaitu 0,84, selama pengujian menggunakan *confusion matrix* untuk mengetahui nilai aktual dan prediksi dalam data *testing*. Perhitungan *confusion matrix* pada data *training* dapat dilihat pada Tabel 3.12.

Tabel 3.12 Hasil *Confusion Matrix* Data *Testing*

Kelas Aktual	Kelas Prediksi		
	Positif	Negatif	Netral
Positif	19	3	1
Negatif	1	22	3
Netral	3	0	16

Tabel 3.12 bahwa hasil perhitungan menggunakan *confusion matrix* didapatkan nilai berupa TPos = 19, FPostNeg = 3, FPostNet = 1, FNegPost = 1, TNeg = 22, FNegNet = 3 dan FNetPost = 3, FNetNeg = 0, TNet = 16, kemudian

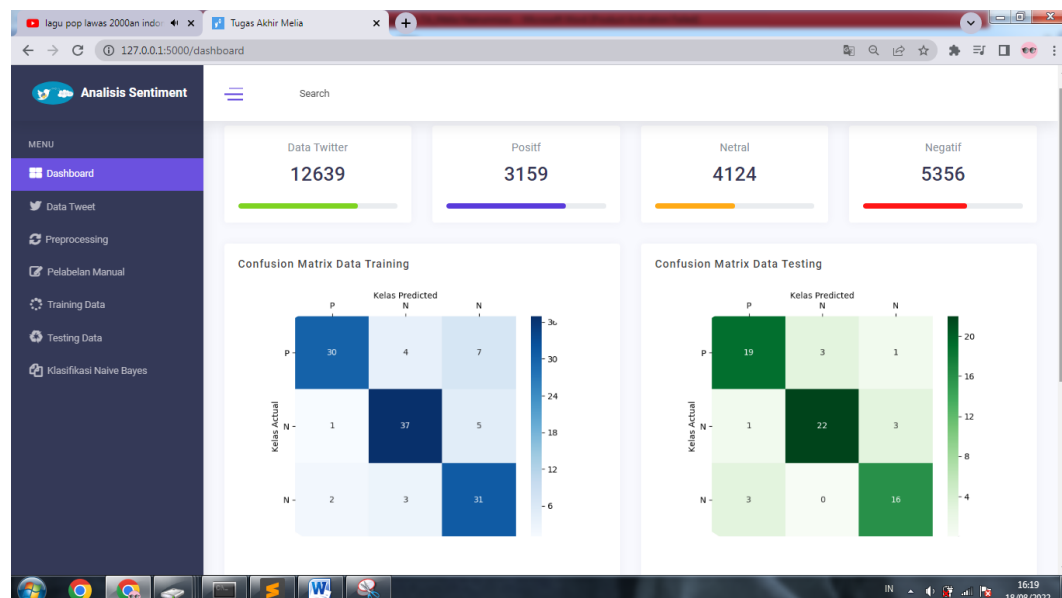
hasil perbandingan data *training* sebanyak 600 data *tweet* dengan data *testing* sebanyak 336 data *tweet* memiliki pengujian yang hampir sama.

4.3 IMPLEMENTASI DISAIN INTERFACE

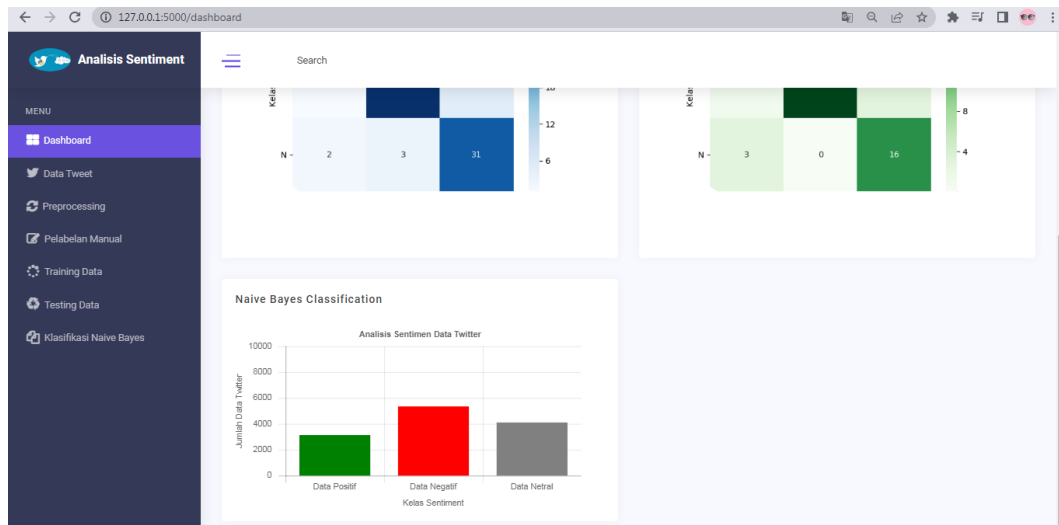
Implementasi *desain interface* merupakan implementasi desain tampilan sistem. Setiap desain yang telah dilakukan sebelumnya di realisasikan menggunakan bahasa pemrograman Python. Dalam Analisis sentimen kepuasan pelanggan perusahaan telekomunikasi seluler Telkomsel di Twitter ini yang digunakan adalah Python dengan memanfaatkan *framework* Flask. Dibawah ini adalah beberapa halaman dan contoh kode program yang termasuk dalam sistem analisis sentimen kepuasan pelanggan perusahaan telekomunikasi seluler Telkomsel di Twitter.

4.3.1 Halaman Dashboard Analisis Sentimen

Halaman dashboard memuat beberapa informasi berdasarkan hasil dari perhitungan *confusion matrix* data *training*, *testing* dan hasil data *tweet* yang didapat seperti banyaknya data Twitter berdasarkan komentar positif, netral, dan negatif. Implementasi halaman dashboard ditunjukkan pada Gambar 4.1.



Gambar 4.1 Implementasi Halaman Dashboard Confusion Matrix



Gambar 4.2 Implementasi Halaman Dashboard Grafik

Berikut ini potongan kode pada *controller* yang berfungsi untuk menampilkan informasi dalam bentuk grafik.

```
#<----- DASHBOARD ----->
@application.route('/dashboard')
def dashboard():
    data_testing = pd.read_excel('TA-
Melia/static/uploads/hasil_data_keseluruhan.xlsx')
    data_testing.dropna()

    positif = data_testing.kelas.value_counts().Positif
    netral = data_testing.kelas.value_counts().Netral
    negatif = data_testing.kelas.value_counts().Negatif
    total = positif + netral + negatif

    return render_template('dashboard.html', total=total,
    positif=positif, netral=netral, negatif=negatif)
```

Fungsi diatas digunakan untuk menampilkan hasil dari data *tweet* menggunakan `value_counts()` untuk mengetahui total dari data *tweet* dan menampilkan total sentiment positif, netral, dan negatif dengan membuat variabel.

4.3.2 Halaman Data Twitter

Halaman data Twitter digunakan untuk melakukan upload file dan menampilkan file CSV yang sudah terupload ke dalam directory agar memudahkan proses *preprocessing*, *training*, *testing* dan *klasifikasi*. Implementasi halaman data Twitter ditunjukkan pada Gambar 4.3.

	2022-04-20 15:27:28+00:00	b'@scorpeeoss aku pk Telkomsel 25-29k perbulan udh sama Disney+'''';
0	2022-04-20 15:25:44+00:00	b'@yunointherain Oke Kak. Gyan cek dulu DM nya ya. Makasih :) -Gyan'''';
1	2022-04-20 15:25:43+00:00	b'@Telkomsel udah min cek dmnya yaa'''';
2	2022-04-20 15:23:59+00:00	b'@iskameraa Oke Kak. Gyan cek dulu DM nya ya. Makasih :) -Gyan'''';
3	2022-04-20 15:23:58+00:00	b'@Telkomsel sudah ada di dm kakak :)'''';
4	2022-04-20 15:23:33+00:00	b'TELKOMSEL MAHAL BGT ANJIR KUOTANYA NANGISSSSA https://t.co/3FSOI9HPiu'''';

Gambar 4.3 Implementasi Halaman Data Tweet

Berikut ini potongan kode pada *controller* yang berfungsi untuk melakukan upload file.

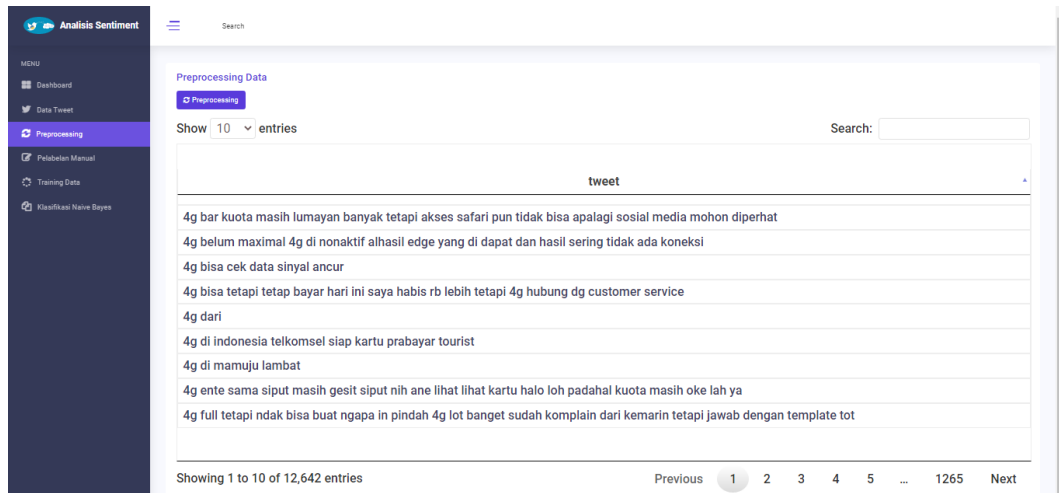
```
#<----- UPLOAD FILE ----->
@app.route('/data_tweet', methods=['GET', 'POST'])
def data_tweet():
    data = []
    data = pd.DataFrame()
    if request.method == "POST":
        uploaded_file = request.files['filename']
        file = os.path.join(application.config['UPLOAD_FOLDER'],
uploaded_file.filename)
        uploaded_file.save(file)
        try:
            data = pd.read_excel(file)
            data.drop(['Unnamed: 1', 'Unnamed: 2', 'Unnamed: 3',
'Unnamed: 4', 'Unnamed: 5', 'Unnamed: 6', 'Unnamed: 7',
'Unnamed: 8', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed:
11', 'Unnamed: 12'], axis=1, inplace=True)
            data.append(data)
        except:
            data = pd.read_csv(file, on_bad_lines='skip')
            data.dropna()
            data.append(data)

    return render_template('data_tweet.html',
data=[data.to_html(justify='center', classes=['table-striped',
'table-bordered', 'dt-responsive', 'table-style'],
table_id='example')])
```

4.3.3 Halaman *Preprocessing* Data

Halaman *preprocessing* berfungsi untuk menampilkan dan melakukan proses pengolahan data teks yang sudah tersedia. *Preprocessing* berfungsi untuk

memperbaiki data *tweet* dengan melakukan tahapan-tahapan agar menjadi data yang terstruktur. Implementasi halaman *preprocessing* ditunjukkan pada Gambar 4.4.



Gambar 4.4 Implementasi Halaman *Preprocessing* Data

Berikut ini potongan kode pada *controller* yang berfungsi untuk melakukan *preprocessing* data.

```
#<----- PREPROCESSING DATA ----->
@application.route('/preprocessing', methods=['GET', 'POST'])
def preprocessing():
    if request.method == 'POST':
        data = pd.read_csv('TA-
Melia/static/uploads/hasil_15April.csv', names=['tanggal',
'tweet'], encoding='latin-1')
        data.dropna()
        data.drop(['tanggal'], axis=1, inplace=True)

        data['tweet'] = data['tweet'].map(lambda x: lower(x))
        data['tweet'] = data['tweet'].map(lambda x:
remove_punctuation(x))
        data['tweet'] = data['tweet'].map(lambda x:
normalized_term(x))
        data['tweet'] = data['tweet'].map(lambda x:
remove_stopwords(x))
        data['tweet'] = data['tweet'].map(lambda x: stem_text(x))
        data['tweet'] = data['tweet'].map(lambda x: tokenize(x))

        data = data.drop_duplicates(subset=['tweet'], keep=False)
        data.to_excel('TA-
Melia/static/uploads/hasil_preprocessing.xlsx')

        data_preprocessing = pd.read_excel('TA-
Melia/static/uploads/hasil_preprocessing.xlsx')
        data_preprocessing.dropna()
```

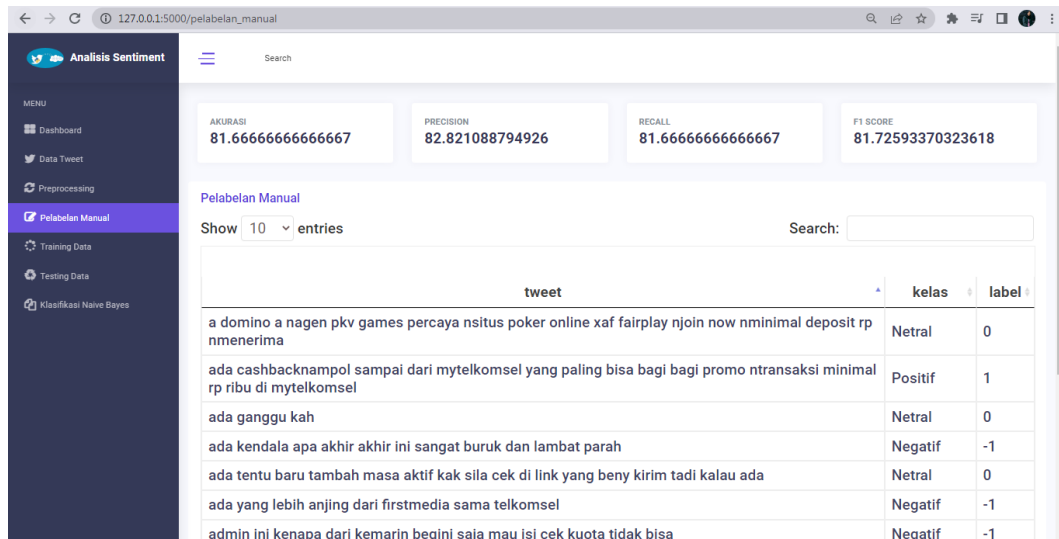
```
data_preprocessing.drop(['Unnamed: 0'], axis=1, inplace=True)

return render_template('preprocessing.html',
data_preprocessing=[data_preprocessing.to_html(index=False,
justify='center', classes=['table-striped', 'table-bordered', 'dt-
responsive', 'table-style'], table_id='example')])
```

Fungsi `.map` diatas digunakan untuk mengaplikasikan suatu fungsi pada semua anggota array yang terdapat pada variabel data *tweet*, `lambda` digunakan untuk melakukan operasi tanpa harus mendefinisikan fungsinya, memiliki beberapa argumen 1 ekspresi.

4.3.4 Halaman Pelabelan Manual

Halaman pelabelan manual digunakan untuk menampilkan hasil *Accuracy*, *Precision*, *Recall* dan *F1-score* berdasarkan data *tweet* yang sudah diberi kelas, label terhadap kalimat yang ada pada data *tweet* secara manual. Proses pelabelan manual dilakukan menggunakan Microsoft Excel. Implementasi halaman pelabelan manual ditunjukkan pada Gambar 4.5.



tweet	kelas	label
a domino a nagen pkv games percaya nsitus poker online xaf fairplay njoin now nminimal deposit rp nmenerima	Netral	0
ada cashbacknampol sampai dari mytelkomsel yang paling bisa bagi bagi promo ntransaksi minimal rp ribu di mytelkomsel	Positif	1
ada ganggu kah	Netral	0
ada kendala apa akhir akhir ini sangat buruk dan lambat parah	Negatif	-1
ada tentu baru tambah masa aktif kak sila cek di link yang beny kirim tadi kalau ada	Netral	0
ada yang lebih anjing dari firstmedia sama telkomsel	Negatif	-1
admin ini kenapa dari kemarin begini saja mau isi cek kuota tidak bisa	Negatif	-1

Gambar 4.5 Implementasi Halaman Pelabelan Manual

Berikut ini potongan kode pada *controller* yang berfungsi untuk menampilkan data *tweet* yang sudah dilakukan pelabelan.

```
#<----- PELABELAN MANUAL ----->

@application.route('/pelabelan_manual', methods=['GET', 'POST'])
def pelabelan_manual():
```

```

data_pelabelan = pd.read_excel('TA-
Melia/static/uploads/data_training.xlsx')
data_pelabelan.dropna()
data_pelabelan.drop(['Unnamed: 1', 'Unnamed: 2', 'Unnamed: 3',
'Unnamed: 4', 'Unnamed: 5', 'Unnamed: 6', 'Unnamed: 7',
'Unnamed: 8', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed:
11', 'Unnamed: 12'], axis=1, inplace=True)
data_pelabelan.drop_duplicates(['tweet'])

```

Fungsi diatas berfungsi untuk menampilkan data *tweet* yang sudah dilakukan pelabelan secara manual.

Berikut ini potongan kode pada *controller* yang berfungsi untuk melakukan perhitungan akurasi pada data *tweet* yang sudah dilakukan pelabelan secara manual.

```

#<----- AKURASI TRAINING ----->

X = data_pelabelan['tweet']
Y = data_pelabelan['label']
X_train,X_test,Y_train,Y_test =
train_test_split(X,Y,test_size=0.2,random_state=42)

vect = TfidfVectorizer(analyzer =
"word",min_df=0.0004,max_df=0.115, ngram_range=(1,3))
vect.fit(X_train)
X_train_dtm = vect.transform(X_train)
X_test_dtm = vect.transform(X_test)

nbmodel = MultinomialNB(alpha=0.1)
nbmodel = nbmodel.fit(X_train_dtm,Y_train)
Y_pred = nbmodel.predict(X_test_dtm)

akurasi = accuracy_score(Y_test, Y_pred) * 100
flscore = f1_score(Y_test, Y_pred, average='weighted') * 100
presision = precision_score(Y_test, Y_pred,
average='weighted') * 100
recall = recall_score(Y_test, Y_pred, average='weighted') *
100

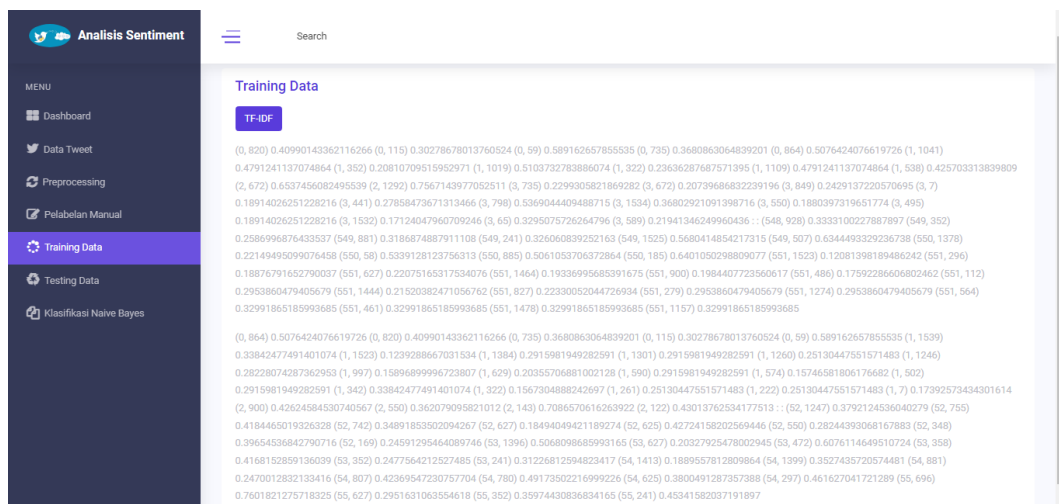
return render_template('pelabelan_manual.html',
data_pelabelan=[data_pelabelan.to_html(index=False,
justify='center', classes=['table-striped', 'table-bordered',
'table-hover', 'table-condensed', 'dt-responsive', 'table-style'],
table_id='example')], accuracy_score=akurasi, f1_score=flscore,
precision_score=presision, recall_score=recall)

```

4.3.5 Halaman *Training* Data

Halaman *training* data digunakan untuk membuat model klasifikasi yang digunakan untuk mengklasifikasi sentimen secara otomatis. Langkah-langkah

dalam *training* data menggunakan metode Naive Bayes Classification dan *library* Multinomial Naive Bayes dengan fitur ekstraksi menggunakan *term frequency-invers document frequency* (TF-IDF) pada data teks. Implementasi halaman *training* data ditunjukkan pada Gambar 4.6.



Gambar 4.6 Implementasi Halaman Training Data

Berikut ini potongan kode pada *controller* yang berfungsi untuk melakukan perhitungan secara otomatis dengan pembobotan pada kata yang ada pada data *training*.

```
#<----- TRAINING DATA ----->

@app.route('/training', methods=['GET', 'POST'])
def training():
    data_testing = pd.read_excel('TA-
Melia/static/uploads/data_training.xlsx')
    data_testing.dropna()
    data_testing.drop(['Unnamed: 3', 'Unnamed: 4', 'Unnamed: 5',
'Unnamed: 6', 'Unnamed: 7',
'Unnamed: 8', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed:
11', 'Unnamed: 12'], axis=1, inplace=True)

    X = data_testing['tweet']
    y = data_testing['label']

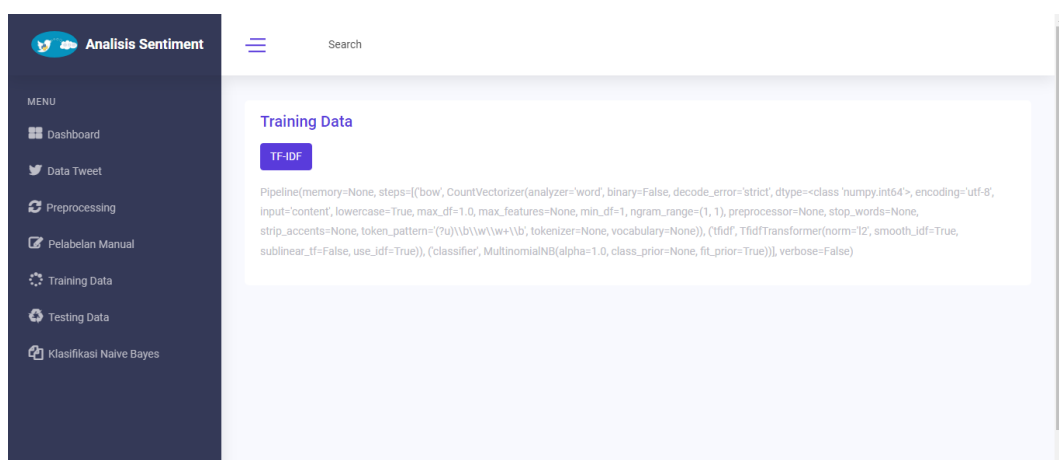
    X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.1, random_state=13)

    vectorizer = TfidfVectorizer()

    X_train = vectorizer.fit_transform(data_testing['tweet'])
    X_test = vectorizer.transform(X_test)
```

```
return render_template('testing_tfidf.html', y=y, X=X,
X_train=X_train, X_test=X_test, y_train=y_train, y_test=y_test)
```

Setelah melakukan perhitungan *term frequency-invers document frequency* (TF-IDF) selanjutnya dilakukan pembuatan model klasifikasi yang dibuat dalam sebuah fungsi untuk memudahkan dalam pemanggilan. Tombol Button TF-IDF berfungsi untuk melakukan pembuatan model yang akan disimpan dalam bentuk file *.pickle* yang diberi nama variabel pipeline. Implementasi halaman pembuatan model klasifikasi dapat dilihat pada Gambar 4.7.



Gambar 4.7 Pembuatan Model Klasifikasi

Berikut ini potongan kode pada *controller* yang berfungsi untuk pembuatan model klasifikasi yang disimpan dalam bentuk file *.pickle*.

```
#<----- FILE PICKLE ----->

@app.route('/testing_tfidf', methods=['GET', 'POST'])
def testing_tfidf():
    data_testing = pd.read_excel('TA-
Melia/static/uploads/data_training.xlsx')
    data_testing.dropna()
    data_testing.drop(['Unnamed: 3', 'Unnamed: 4', 'Unnamed: 5',
'Unnamed: 6', 'Unnamed: 7',
'Unnamed: 8', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed:
11', 'Unnamed: 12'], axis=1, inplace=True)

    bow_transformer = CountVectorizer().fit(data_testing['tweet'])
    messages_bow =
bow_transformer.transform(data_testing['tweet'])

    tfidf_transformer = TfidfTransformer().fit(messages_bow)

    messages_tfidf = tfidf_transformer.transform(messages_bow)
```

```

pipeline = Pipeline([
    ('bow', CountVectorizer()),
    ('tfidf', TfidfTransformer()),
    ('classifier', MultinomialNB())
])

X_train = np.asarray(data_testing['tweet'])
pipeline = pipeline.fit(X_train,
np.asarray(data_testing['label']))

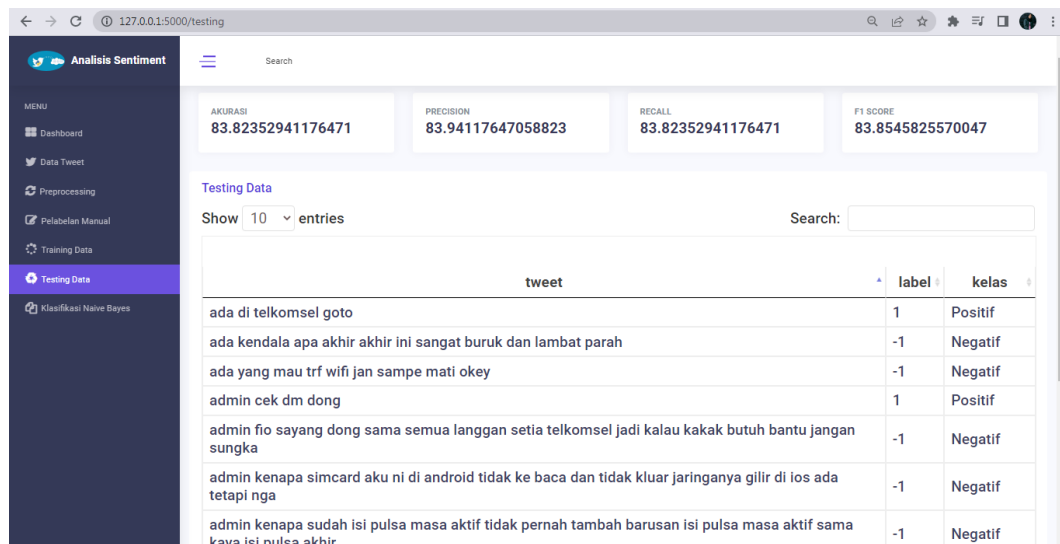
file_data = pickle.dump(pipeline, open('TA-
Melia/static/uploads/latihan_ta.pickle', 'wb'))

return render_template('testing_tfidf.html',
pipeline=pipeline)

```

4.3.6 Halaman *Testing Data*

Halaman *testing* digunakan untuk menampilkan hasil *Accuracy*, *Precision*, *Recall* dan *F1-score* berdasarkan data *tweet* yang sudah diprediksi kelas dan labelnya secara otomatis, untuk menentukan seberapa akurat model yang dibangun pada *training* data. Implementasi halaman *testing* data ditunjukkan pada Gambar 4.8.



The screenshot shows a web application interface for sentiment analysis. At the top, there are four boxes displaying performance metrics: AKURASI (83.82352941176471), PRECISION (83.94117647058823), RECALL (83.82352941176471), and F1 SCORE (83.8545825570047). Below these is a section titled 'Testing Data' with a search bar and a 'Show 10 entries' dropdown. A table displays the test data with three columns: 'tweet', 'label', and 'kelas'.

tweet	label	kelas
ada di telkomsel goto	1	Positif
ada kendala apa akhir akhir ini sangat buruk dan lambat parah	-1	Negatif
ada yang mau trf wifi jan sampe mati okey	-1	Negatif
admin cek dm dong	1	Positif
admin fio sayang dong sama semua langgan setia telkomsel jadi kalau kakak butuh bantu jangan sungka	-1	Negatif
admin kenapa simcard aku ni di android tidak ke baca dan tidak keluar jaringanya gilir di ios ada tetapi nga	-1	Negatif
admin kenapa sudah isi pulsa masa aktif tidak pernah tambah barusan isi pulsa masa aktif sama kava isi nuls akhir	-1	Negatif

Gambar 4.8 Implementasi Halaman *Testing*

Berikut ini potongan kode pada *controller* yang berfungsi untuk menampilkan data *tweet* yang sudah diprediksi untuk mengetahui tingkat keakuratan pemodelan yang sudah dibangun pada *training*.


```

#<----- TESTING ----->

@app.route('/testing', methods=['GET', 'POST'])
def testing():
    import pickle
    vect = pickle.load(open('TA-
Melia/static/uploads/latihan_ta.pickle', 'rb'))

    testing = pd.read_excel('TA-
Melia/static/uploads/testing.xlsx')
    testing.dropna()
    testing.drop(['Unnamed: 0', 'Unnamed: 3', 'Unnamed: 4',
'Unnamed: 5', 'Unnamed: 6', 'Unnamed: 7',
'Unnamed: 8', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed:
11', 'Unnamed: 12'], axis=1, inplace=True)

    testing = testing['tweet'].fillna(' ')
    prediction = vect.predict(testing)

    result = []

    for i in range(len(prediction)):
        if(prediction[i] == 1):
            sentiment = 'Positif'
        elif(prediction[i]==0):
            sentiment = 'Netral'
        else:
            sentiment = 'Negatif'

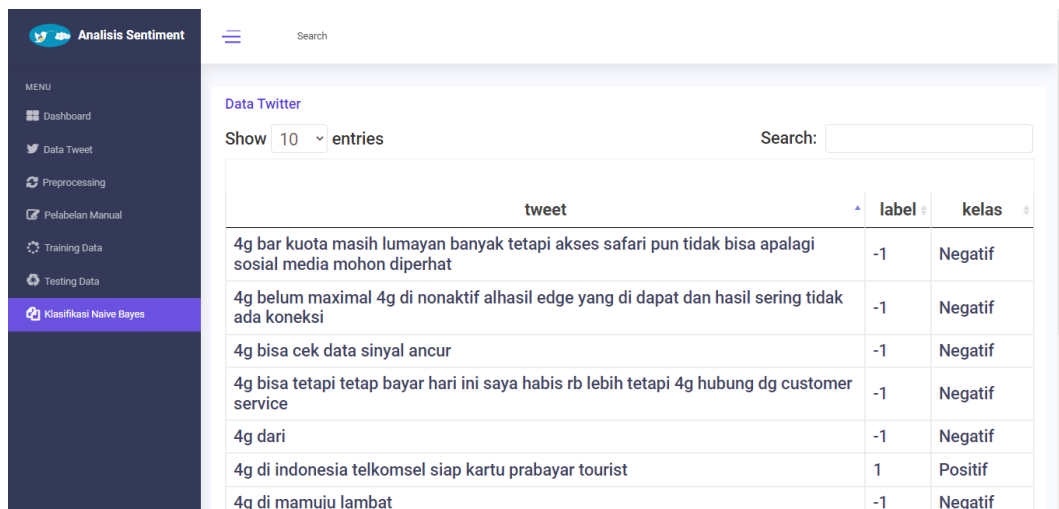
        result.append({'tweet':testing[i], 'label':prediction[i],
'kelas':sentiment })

    testing = pd.DataFrame(result)
    testing = testing.dropna()
    testing.to_excel('TA-Melia/static/uploads/hasil_testing.xlsx')

```

4.3.7 Halaman Klasifikasi Naive Bayes Classification

Halaman klasifikasi Naive Bayes Classification digunakan untuk menampilkan data *tweet* yang sudah diprediksi kelas dan labelnya secara otomatis berdasarkan model yang sudah dibangun pada tahap *training*. Implementasi halaman klasifikasi Naive Bayes Classification ditunjukkan pada Gambar 4.9.



tweet	label	kelas
4g bar kuota masih lumayan banyak tetapi akses safari pun tidak bisa apalagi sosial media mohon diperhat	-1	Negatif
4g belum maximal 4g di nonaktif alhasil edge yang di dapat dan hasil sering tidak ada koneksi	-1	Negatif
4g bisa cek data sinyal ancur	-1	Negatif
4g bisa tetapi tetap bayar hari ini saya habis rb lebih tetapi 4g hubung dg customer service	-1	Negatif
4g dari	-1	Negatif
4g di indonesia telkomsel siap kartu prabayar tourist	1	Positif
4g di mamuju lambat	-1	Negatif

Gambar 4.9 Implementasi Halaman Klasifikasi Naive Bayes Classification

Berikut ini potongan kode pada *controller* yang berfungsi untuk menampilkan data *tweet* yang sudah diprediksi.

```
#<----- KLASIFIKASI NAIVE BAYES ----->

@appapplication.route('/klasifikasi', methods=['GET', 'POST'])
def klasifikasi():
    import pickle
    vect = pickle.load(open('TA-
Melia/static/uploads/latihan_ta.pickle', 'rb'))

    klasifikasi = pd.read_excel('TA-
Melia/static/uploads/data_tweet_keseluruhan.xlsx')
    klasifikasi.dropna()
    klasifikasi.drop(['Unnamed: 0', 'Unnamed: 3', 'Unnamed: 4',
'Unnamed: 5', 'Unnamed: 6', 'Unnamed: 7',
'Unnamed: 8', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed:
11', 'Unnamed: 12', 'Unnamed: 13'], axis=1, inplace=True)

    klasifikasi = klasifikasi['tweet'].fillna(' ')
    prediction = vect.predict(klasifikasi)

    result = []

    for i in range(len(prediction)):
        if(prediction[i] == 1):
            sentiment = 'Positif'
        elif(prediction[i]==0):
            sentiment = 'Netral'
        else:
            sentiment = 'Negatif'

    result.append({'tweet':klasifikasi[i], 'label':prediction[i],
'kelas':sentiment })
```

```

    klasifikasi = pd.DataFrame(result)
    klasifikasi = klasifikasi.dropna()
    klasifikasi.to_excel('TA-
Melia/static/uploads/hasil_data_keseluruhan.xlsx')

    return render_template('klasifikasi_naive_bayes.html',
    klasifikasi=[klasifikasi.to_html(index=False, justify='center',
    classes=['table-striped', 'table-bordered', 'table-hover', 'table-
condensed', 'dt-responsive', 'table-style'], table_id='example')])

```

4.4 PEMBAHASAN

Analisis sentimen berdasarkan kata kunci “Telkomsel” menggunakan data *tweet* sebanyak 13.239 data untuk dilakukan analisis. Data *tweet* sebanyak 600 data diberi kelas dan label secara manual dengan rata-rata akurasi yang cukup bagus sebesar 81,7% dan data *tweet* sebanyak 336 data yang sudah diprediksi secara otomatis berdasarkan model yang sudah dibangun pada *training* dengan rata-rata akurasi 84%. Pada pengujian Data *tweet* untuk dilakukan prediksi yang sudah dibangun pada *training* sebanyak 12.639 data, hasil prediksi didapatkan data *tweet* berlabel positif sebanyak 3.159 data, data *tweet* berlabel netral sebanyak 4.124 data dan data *tweet* berlabel negatif sebanyak 5.356 data.

Analisis sentimen kepuasan pelanggan perusahaan telekomunikasi seluler di Twitter dibangun menggunakan bahasa pemrograman Python dengan *framework* Flask, pada sistem analisis sentimen memiliki beberapa menu, yaitu menu dashboard, menu data *tweet*, menu *preprocessing*, menu pelabelan manual, menu *training* data, menu *testing* dan menu klasifikasi Naive Bayes Classification. Pada menu dashboard berfungsi untuk memberikan informasi berdasarkan hasil dari proses klasifikasi Naive Bayes Classification, menu data *tweet* digunakan untuk melakukan upload file dan menampilkan file yang sudah terupload didalam directory, menu *preprocessing* digunakan untuk melakukan proses pengolahan data teks dengan memerlukan beberapa tahapan-tahapan, menu pelabelan manual berfungsi untuk menampilkan hasil akurasi dari data *tweet* yang sudah diberi kelas atau label berdasarkan kalimat secara manual, menu *training* digunakan untuk membangun model klasifikasi sentimen secara otomatis dengan fungsi ekstraksi menggunakan *term frequency-invers document frequency* (TF-IDF) pada data teks, menu *testing* berfungsi untuk menampilkan hasil akurasi

berdasarkan data *tweet* yang diprediksi secara otomatis untuk menentukan seberapa akurat model yang dibangun pada *training*, menu klasifikasi Naive Bayes Classification berfungsi untuk menampilkan data *tweet* yang sudah diprediksi secara otomatis.

Grafik yang digunakan dalam sistem analisis sentimen ini adalah grafik bentuk *Bar*. Grafik tersebut secara otomatis menunjukkan jumlah data *tweet* yang diprediksi berdasarkan kelas positif, negatif dan netral. Sistem analisis sentimen ini tentunya masih terdapat banyak kekurangan seperti dalam pemanggilan file belum bisa dilakukan pemanggilan secara otomatis.

BAB 5

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah :

1. Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel di Twitter telah berhasil dibangun menggunakan *framework* Flask dengan bahasa pemrograman Python.
2. Penelitian ini berhasil membuat model Analisis sentimen menggunakan metode Naive Bayes Classification dengan mengklasifikasikan dan memprediksi data *tweet* yang bersifat positif, netral dan negatif.
3. Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel di Twitter berhasil melakukan pengujian dengan akurasi 81,7% untuk data *training* dan 84% untuk klasifikasi Naive Bayes Classification.
4. Berdasarkan hasil klasifikasi dan perhitungan pada model dan data *tweet* memberikan hasil bahwa Telkomsel memiliki sentimen negatif dikarenakan layanan internet yang tidak stabil.

5.2 SARAN

Adapun saran yang dapat dilakukan untuk Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel Di Twitter adalah sebagai berikut:

1. Proses pemanggilan file sebaiknya dibuat secara otomatis untuk memudahkan dalam proses analisis sentimen.
2. Menambahkan fitur grafik untuk menampilkan informasi hasil analisis sentimen berdasarkan bulan.

DAFTAR PUSTAKA

- Abidin, Z., Darwis, D. & Siskawati, N., 2021. Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), pp. 131-145.
- Abubakar, H., Khaidar, A. M. A. & Nur, I., 2018. Pengaruh Kualitas Layanan Terhadap Kepuasan Pelanggan Pada PT Telekomunikasi Selular Indonesia Grapari Telkomsel Makassar (Studi Kasus: Aplikasi My Telkomsel). *Economics Bosowa*, 4(1), pp. 56-66.
- Agustini, P., 2021. *Warganet Meningkat, Indonesia Perlu Tingkatkan Nilai Budaya di Internet*. [Online]
Available at: <https://aptika.kominfo.go.id/2021/09/warganet-meningkat-indonesia-perlu-tingkatkan-nilai-budaya-di-internet/>
[Diakses 06 April 2022].
- Ambiyar, Edi, F., Samsir, Verawardani, U., & Watrianthos, R., 2021. Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(1), pp. 157-163.
- Apriliani, A. & Kurniawan, R., 2020. Analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter berbasis web scraper. *Jurnal INSTEK (Informatika Sains dan Teknologi)*, 5(1), pp. 67-75.
- Asror, I., Darwiyanto, E. & Nisa, A., 2019. Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi. *eProceedings of Engineering*, 6(2), pp. 8650-8659.
- Damaiyanti, R., Desinta, D., Khairi, F., Naraswati, N. P., Nooraeni, R., & Rosmilda, D. C., 2021. Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *Sistemasi: Jurnal Sistem Informasi*, 10(1), pp. 222-238.
- Devita, R. N., Hermanto, H. W. & Wibawa, A. P., 2018. Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *JTIK (Jurnal Teknologi Informasi dan Ilmu Komputer)*, Volume 5, pp. 427-434.
- Fitriyah, C. & Waliyansyah, R. R., 2019. Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN). *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 5(2), pp. 157-163.

- Gunawan, B., Pratiwi, H. S. & Pratama, E. E., 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 4(2), pp. 113-118.
- Jatmiko, L. D., 2021. *Rekor! Telkomsel Layani 173,5 Juta Pelanggan per Kuartal III/2021*. [Online]
Available at:
<https://teknologi.bisnis.com/read/20211125/101/1470356/rekor-telkomsel-layani-1735-juta-pelanggan-per-kuartal-iii2021>
[Diakses 6 April 2022].
- Pahlevi, R., 2022. *Pengguna Internet di Dunia Capai 4,95 Miliar Orang Per Januari 2022*. [Online]
Available at:
<https://databoks.katadata.co.id/datapublish/2022/02/07/pengguna-internet-di-dunia-capai-495-miliar-orang-per-januari-2022>
[Diakses 2022 April 06].
- Perdana, K., Pricillia, T. & Zulfachmi, 2021. Optimasi TextBlob Menggunakan Support Vector Machine untuk Analisis Sentimen (Studi Kasus Layanan Telkomsel). *Jurnal Bangkit Indonesia*, 10(1), pp. 13-15.
- Putera, M. I. A. & Putra, M. G. L., 2019. Analisis Perbandingan Metode SOAP dan REST yang digunakan pada Framework Flask untuk Membangun Web Service. *SCAN-Jurnal Teknologi Informasi dan Komunikasi*, 14(2), pp. 1-7.
- Rahmadhika, M. K. & Thantawi, A. M., 2021. Rancang Bangun Aplikasi Face Recognition Pada Pendekatan CRM Menggunakan Opencv Dan Algoritma Haarcascade. *ikraith-informatika*, 5(1), pp. 109-118.
- Retnasary, M. & Rohman, R. E., 2020. Penggunaan Media Sosial Twitter Masyarakat di Kota Bandung sebagai Media untuk Memperoleh Informasi mengenai New Normal dalam akun @humasjabar. *Jurnal Politikom Indonesiana*, 5(2), pp. 18-28.
- Saputro, I. W. & Sari, B. W., 2019. Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), pp. 1-11.

LAMPIRAN

LAMPIRAN 1 JADWAL PENELITIAN

No	URAIAN KEGIATAN	BULAN					
		Februari	Maret	April	Mei	Juni	Juli
TAHAPAN PENYUSUNAN							
1. Tahap Identifikasi dan Analisis							
1	Rumusan Masalah	■					
	Studi literatur dan studi lapangan	■	■				
	Pengambilan data			■	■	■	■
2. Tahap <i>Preprocessing</i>							
1	Pengolahan data			■	■	■	■
3. Tahap Pelabelan manual							
1	Pelabelan data			■	■	■	■
4. Tahap <i>Training data</i>							
1	Proses klasifikasi secara otomatis				■	■	■
5. Tahap <i>Testing</i>							
1	Uji tingkat keakuratan				■	■	■
6. Tahap Penulisan Laporan							
1	Proposal penelitian	■					
2	Laporan Hasil Penelitian	■	■	■	■	■	■

LAMPIRAN 2 LEMBAR BIMBINGAN DOSEN



YAYASAN KARTIKA EKA PAKSI
UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA
FAKULTAS TEKNIK DAN TEKNOLOGI INFORMASI (FTTI)
Jl. Silwangi Pengasih Barat, Bantul, Sleman, Yogyakarta 55283 Telp. (0274) 522485, 522501 Fax. (0274) 527228
Website : www.unjpa.ac.id E-mail : info@unjpa.ac.id



Prodi : INFORMATIKA (S1)

KARTU BIMBINGAN TUGAS AKHIR - GENAP 2021/2022

NPM : 182102022
NAMA MAHASISWA : Melia Haerunnissa
NO. HP MAHASISWA : 8994121133
PEMBIMBING TA : Agung Priyanto, S.T., M.Eng
JUDUL TA : Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler
Telkomsel Di Twitter

NO	TANGGAL	MATERI BIMBINGAN
1	Selasa, 05 Juli 2022	Bimbingan awal sistem.
2	Kamis, 07 Juli 2022	Bimbingan pelabelan manual dan perhitungan akurasi.
3	Rabu, 13 Juli 2022	Bimbingan perhitungan akurasi <i>training</i> .
4	Senin, 18 Juli 2022	Bimbingan perhitungan akurasi <i>training</i> .
5	Senin, 25 Juli 2022	Bimbingan perhitungan akurasi <i>testing</i> .
6	Senin, 01 Agustus 2022	Bimbingan Laporan Bab 3-5
7	Jumat, 05 Agustus 2022	Bimbingan Perhitungan manual TF-IDF.
8	Senin, 08 Agustus 2022	Bimbingan Laporan.
9	Jumat, 12 Agustus 2022	Bimbingan Laporan.
10	Kamis, 18 Agustus 2022	Bimbingan Sistem dan Laporan.
11		
12		
13		
14		

Persetujuan Pendadaran
Ttd Dosen Pembimbing Tugas Akhir

(Agung Priyanto, S.T., M.Eng)

Ttd Mahasiswa

(Melia Haerunnissa)

LAMPIRAN 3 HASIL CEK PLAGIARISME

Cek Plagiarisme
Skripsi_Final_Analisis Sentimen
Kepuasan Pelanggan
Perusahaan

by Melia Haerunnissa 182102022

Submission date: 31-Aug-2022 12:45AM (UTC+0700)

Submission ID: 1889518892

File name: 182102022_Melia_Haerunnissa_Informatika_Final.docx (1.53M)

Word count: 6068

Character count: 41065

Cek Plagiarisme Skripsi_Final_Analisis Sentimen Kepuasan Pelanggan Perusahaan

ORIGINALITY REPORT

14%	14%	4%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.unjaya.ac.id Internet Source	9%
2	danielyu-cua.com Internet Source	1%
3	123dok.com Internet Source	1%
4	Dspace.Uii.Ac.Id Internet Source	1%
5	hackernoon.com Internet Source	1%
6	stackoverflow.com Internet Source	1%
7	academy.hsoub.com Internet Source	1%
8	Submitted to University of Southampton Student Paper	<1%
9	repository.uin-suska.ac.id Internet Source	<1%

10

Submitted to Universitas Brawijaya

Student Paper

<1%

Exclude quotes On

Exclude bibliography On

Exclude matches

< 25 words