

BAB 4

HASIL PENELITIAN

4.1 RINGKASAN HASIL PENELITIAN

Hasil penelitian analisis sentimen pengguna Twitter terhadap Perusahaan Daerah Air Minum (PDAM) menggunakan metode naive bayes *classifier* dengan kata kunci “pdam”, “pdam sleman”, “pelayanan pdam”, dan “kualitas pdam”. Data Twitter yang diambil sebanyak 2.530 data mentah dengan menggunakan Google Colab. Data mentah tersebut dilakukan *preprocessing* sehingga mendapatkan 1.077 data bersih. Sebanyak 754 data digunakan untuk proses *training* berdasarkan kelas puas 377 data dan kelas kurang puas 377 data. Pada proses *testing* data yang digunakan sebanyak 323 data. Analisis sentimen ini dibangun menggunakan bahasa pemrograman Python dengan metode Naive Bayes *classifier* untuk menghasilkan hasil perhitungan sentimen menggunakan *framework* Flask.

4.2 PENGAMBILAN DATA TWITTER

Pengambilan data *tweet* yang berkaitan dengan kata kunci “pdam”, “pdam sleman”, “pelayanan pdam”, dan “kualitas pdam”. Data *tweet* yang diambil mulai dari tanggal Maret - Juni 2023 dengan 2.530 data *tweet*. Proses pengambilan data menggunakan Tweet Harvest sehingga data yang diperoleh terbatas. Data *tweet* yang diambil dari Google Colab berformat file CSV sehingga dapat dibuka melalui Microsoft Excel.

4.2.1 PENGUMPULAN DATA TWITTER

Pengumpulan data Twitter pada penelitian ini dilakukan dengan bahasa pemrograman Python. Penelitian ini mendapatkan data *tweet* berdasarkan beberapa kata kunci yaitu “pdam”, “pdam sleman”, “pelayanan pdam”, dan “kualitas pdam”. Tampilan pengambilan data *tweet* dapat ditunjukkan pada Gambar 4.1.

```

| pip install pandas
| curl -sL https://deb.nodesource.com/setup_18.x | sudo -E bash -
| sudo apt-get install -y nodejs

Tampilkan output tersembunyi

[ ] # Crawl Data

filename = '2.csv'
search_keyword = 'pdam until:2023-04-30 since:2023-04-01'
limit = 2000

lnpx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token ""

```

Gambar 4.1 Pengambilan data *tweet*

Dalam pengambilan data *tweet* menggunakan *tweet harvest* dengan format file CSV. Contoh data *tweet* yang sudah diambil dapat ditunjukkan pada Tabel 4.1.

Tabel 4.1 Contoh data *tweet* mentah

No	Data Tweet
1	@PDAMSleman Selamat pagi pak. Saya masih pelanggan PDAM, bukan pelanggan Air Coklat lho. Mohon perhatiannya. Sudah DM. https://t.co/ney2T5V3X3
2	Lupa bayar pdam alhasil kena dendaðŸ¥² Ni kalo ketauan bakal dapat ceramah betolðŸ¥²
3	@herjunxt @Biasabangetaja air PDAM mati? sumur kering?
4	@PDAMSleman Sampai jam berapa ya matinya? Gak bisa buat aktifitas lho kalo airnya mati
5	@Setyawan_Utomo @PDAMSleman Beda ih air sama ipal :/ðŸ©”

Tabel 4.1 menunjukkan hasil pengambilan data *tweet* yang belum terstruktur sehingga diperlukan proses *preprocessing* untuk menghilangkan variasi yang tidak digunakan dalam proses analisis sentimen.

4.2.2 PREPROCESSING

Preprocessing dilakukan untuk proses pengolahan data teks yang belum terstruktur untuk diperbaiki dengan tahapan-tahapan agar menjadi data yang terstruktur. Contoh data *tweet* yang sudah dilakukan proses *preprocessing* ditunjukkan pada Tabel 4.2.

Tabel 4.2 Contoh data *tweet* hasil *preprocessing*

No	Data Tweet
1	macet air pdam xixixi
2	pdam mati kemarin
3	pasok air pdam kurang
4	pdam hasil tuntas bocor pipa beringin cepat target
5	mantap sih pdam mati pln mati gitu mandi pagi

Tabel 4.2 menunjukkan hasil data *tweet* yang sudah dilakukan *preprocessing*. Data *tweet* lebih terstruktur dan dapat dilakukan perhitungan pada tahap selanjutnya. Berikut tahapan proses *preprocessing* data:

4.2.2.1 Number Removal

Number removal merupakan proses membersihkan karakter angka yang terdapat dalam data teks. Contoh data hasil proses *number removal* ditunjukkan pada Gambar 4.2.

	full_text	tweet
0	[Sertu, Sukarno, Babinsa, Kalurahan, Purwoharj...	['Sertu', 'Sukarno', 'Babinsa', 'Kalurahan', '...
1	[Beli, Rumah, Gede, Tanah, Luas, Harga, Murah,...	['Beli', 'Rumah', 'Gede', 'Tanah', 'Luas', 'Ha...
2	[06:32], Situasi, lalu, lintas, pagi, ini, si,...	[':', 'Situasi', 'lalu', 'lintas', 'pagi', 'i...
3	[@, KompasTV, Langsung, ke, toko, bahan, bangu...	['@', 'KompasTV', 'Langsung', 'ke', 'toko', 'b...
4	[AERNYA, BLOM, NYALAAAA, ASTAGAA, NYAWA, GUE, ...	['AERNYA', 'BLOM', 'NYALAAAA', 'ASTAGAA', 'NYA...

Gambar 4.2 Contoh data hasil *number removal*

4.2.2.2 Punctuation Removal

Punctuation removal merupakan proses penghapusan karakter khusus yang tidak memiliki pengaruh signifikan dalam analisis. Contoh data hasil proses *punctuation removal* dapat ditunjukkan pada Gambar 4.3.

	full_text	tweet
0	Sertu Sukarno Babinsa Kalurahan Purwoharjo Kor...	Sertu Sukarno Babinsa Kalurahan Purwoharjo Kor...
1	Beli Rumah Gede Tanah Luas Harga Murah Dengan ...	Beli Rumah Gede Tanah Luas Harga Murah Dengan ...
2	06:32 Situasi lalu lintas pagi ini si Simpang...	06:32 Situasi lalu lintas pagi ini si Simpang P...
3	@KompasTV Langsung ke toko bahan bangunan aja ...	KompasTV Langsung ke toko bahan bangunan aja l...
4	AERNYA BLOM NYALAAAA ASTAGAA NYAWA GUE SEHARI ...	AERNYA BLOM NYALAAAA ASTAGAA NYAWA GUE SEHARI ...

Gambar 4.3 Contoh data hasil *punctuation removal*

4.2.2.3 Whitespaces Removal

Whitespaces removal merupakan proses penghapusan spasi di awal dan akhir kalimat. Contoh data hasil proses *whitespaces removal* dapat ditunjukkan pada Gambar 4.4.

	full_text	tweet
0	Sertu Sukarno Babinsa Kalurahan Purwoharjo Kor...	SertuSukarnoBabinsaKalurahanPurwoharjoKoramil0...
1	Beli Rumah Gede Tanah Luas Harga Murah Dengan ...	BeliRumahGedeTanahLuasHargaMurahDenganCaraMuda...
2	06:32 Situasi lalu lintas pagi ini si Simpang...	06:32 SituasialulintaspagiiniSimpangPDAMdar...
3	@KompasTV Langsung ke toko bahan bangunan aja ...	@KompasTVLangsungketokobahanbangunanajalaluutk...
4	AERNYA BLOM NYALAAAA ASTAGAA NYAWA GUE SEHARI ...	AERNYABLOMNYALAAAAASTAGAANYAWAGUESEHARILEBIHPA...

Gambar 4.4 Contoh data hasil *whitespaces removal*

4.2.2.4 Case Folding

Case folding merupakan perubahan bentuk huruf dari bentuk awal (huruf kecil atau *lowercase*). Contoh data hasil proses *case folding* dapat ditunjukkan pada Gambar 4.5.

	full_text	tweet
0	Sertu Sukarno Babinsa Kalurahan Purwoharjo Kor...	sertu sukarno babinsa kalurahan purwoharjo kor...
1	Beli Rumah Gede Tanah Luas Harga Murah Dengan ...	beli rumah gede tanah luas harga murah dengan ...
2	06:32 Situasi lalu lintas pagi ini si Simpang...	06:32 situasi lalu lintas pagi ini si simpang...
3	@KompasTV Langsung ke toko bahan bangunan aja ...	@kompastv langsung ke toko bahan bangunan aja ...
4	AERNYA BLOM NYALAAAA ASTAGAA NYAWA GUE SEHARI ...	aernya blom nyalaaaa astagaa nyawa gue sehari ...

Gambar 4.5 Contoh data hasil *case folding*

4.2.2.5 Tokenization

Tokenization merupakan proses pemecahan kalimat, penghapusan simbol dan tanda baca dalam data teks. Contoh data hasil proses *tokenization* dapat ditunjukkan pada Gambar 4.6.

	full_text	tweet
0	Sertu Sukarno Babinsa Kalurahan Purwoharjo Kor...	[sertu, sukarno, babinsa, kalurahan, purwoharj...
1	Beli Rumah Gede Tanah Luas Harga Murah Dengan ...	[beli, rumah, gede, tanah, luas, harga, murah,...
2	06:32 Situasi lalu lintas pagi ini si Simpang...	[situasi, lalu, lintas, pagi, ini, si, simpang...
3	@KompasTV Langsung ke toko bahan bangunan aja ...	[kompastv, langsung, ke, toko, bahan, bangunan...
4	AERNYA BLOM NYALAAAAASTAGAA NYAWA GUE SEHARI ...	[aernya, blom, nyalaaaa, astagaa, nyawa, gue, ...

Gambar 4.6 Contoh data hasil *tokenization*

4.2.2.6 Stopword Removal

Stopword removal merupakan proses penghapusan kata yang memiliki informasi rendah dalam data teks. Contoh data hasil proses *stopword removal* dapat ditunjukkan pada Gambar 4.7.

0	[sertu, sukarno, babinsa, kalurahan, purwoharj...
1	[beli, rumah, gede, tanah, luas, harga, murah,...
2	[situasi, lalu, lintas, pagi, ini, si, simpang...
3	[kompastv, langsung, ke, toko, bahan, bangunan...
4	[aernya, blom, nyalaaaa, astagaa, nyawa, gue, ...
	...

Gambar 4.7 Contoh data hasil *stopword removal*

4.2.2.7 Stemming

Stemming merupakan proses penghapusan infleksi kata menjadi bentuk dasarnya. Contoh data hasil proses *stemming* dapat ditunjukkan pada Gambar 4.8.

```

sertu = sertu
sukarno = sukarno
babinsa = babinsa
kalurahan = kalurahan
purwoharjo = purwoharjo
koramil = koramil
samigaluh = samigaluh
komunikasi = komunikasi
sosial = sosial
dengan = dengan
kontraktor = kontraktor
bapak = bapak
didif = didif
dan = dan
mandor = mandor
lapangan = lapang
nurhadi = nurhadi
bertempat = tempat
di = di

```

Gambar 4.8 Contoh data hasil *stemming*

4.2.2.8 Normalization

Proses standarisasi *term* yang mengalami kesalahan penulisan atau penggunaan bahasa yang tidak lazim. Contoh normalisasi dapat ditunjukkan pada Tabel 4.3.

Tabel 4.3 Contoh normalisasi

Before	After
naek	naik
jg	juga
ttp	tetap
pake	pakai
bgt	banget
sy	saya

4.2.3 PELABELAN MANUAL

Pelabelan manual penelitian ini memberikan label puas dan kurang puas terhadap data *tweet* yang sudah dilakukan proses *preprocessing*. Pelabelan dilakukan di Microsoft Excel secara manual agar data dapat dianalisis. Data *tweet* sebanyak 1.077 diberi label yang akan digunakan untuk *training* sebanyak 754 data sehingga 377 data *tweet* berlabel puas dan 377 data *tweet* berlabel kurang puas. Contoh data yang sudah dilakukan pelabelan manual dapat ditunjukkan pada Tabel 4.4.

Tabel 4.4 Contoh pelabelan manual

No	Tweet	Kelas	Label
1	pdam tulungagung lambat asik susah air langgan kecewa	Kurang Puas	0
2	nunggu air pdam kaya nunggu album blink	Kurang Puas	0
3	pdam komitmen jaga kualitas air	Puas	1
4	pdam nya lancar mas siram terus	Puas	1
5	barusan ngalir terimakasih	Puas	1

Tabel 4.4 menunjukkan hasil pelabelan manual bahwa data *tweet* dengan kelas puas diberi label 1 dan kelas kurang puas diberi label 0. Data pelabelan ini digunakan untuk menghitung nilai akurasi sentimen.

4.2.4 TRAINING DATA

Training data dilakukan untuk menghasilkan model klasifikasi secara otomatis yang digunakan pada proses klasifikasi. Proses ekstraksi TF-IDF pada penelitian ini secara otomatis melalui sistem yang dibangun. Berikut tampilan kode program untuk menghitung nilai TF-IDF secara otomatis di sistem dan model klasifikasinya.

```
X = text['full_text']
y = text['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(text['full_text'])
X_test = vectorizer.transform(X_test)
return render_template('training.html',y=y, X=X, X_train=X_train,
                      X_test=X_test,
                      y_train=y_train,
                      y_test=y_test)
```

Berikut tampilan kode program untuk menghitung *pipeline* model klasifikasi yang ada pada proses *training* data.

```
bow_transformer = CountVectorizer().fit(text['full_text'])
messages_bow = bow_transformer.transform(text['full_text'])
tfidf_transformer = TfidfTransformer().fit(messages_bow)
messages_tfidf = tfidf_transformer.transform(messages_bow)

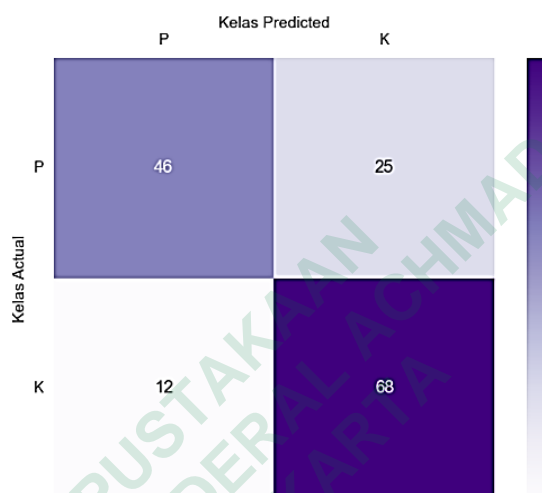
pipeline = Pipeline([
    ('bow', CountVectorizer()),
    ('tfidf', TfidfTransformer()),
    ('classifier', MultinomialNB())
])
```

4.2.5 TESTING DATA

Proses yang digunakan untuk memprediksi kelas dan label berdasarkan data *training* yang sudah dibangun untuk mengetahui tingkat keakuratan pemodelan.

4.2.6 HASIL MODEL KLASIFIKASI DAN EVALUASI KLASIFIKASI

Evaluasi klasifikasi data *training* dengan menggunakan perhitungan *confusion matrix* untuk menemukan nilai aktual dan nilai prediksi berdasarkan akurasi yang diberikan oleh sistem. Data *training* yang digunakan sebanyak 754 data *tweet* yang diberi label puas dan label kurang puas secara manual. Hasil perhitungan *confusion matrix* pada data *training* dapat ditunjukkan pada Gambar 4.9.



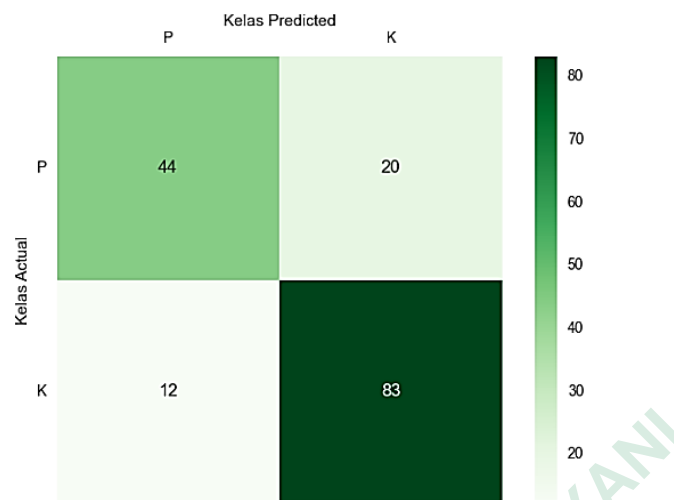
Gambar 4.9 *Confusion matrix* data *training*

Gambar 4.9 menjelaskan bahwa hasil perhitungan *confusion matrix* data *training* didapatkan berupa TPos = 46, FPostNeg = 25, FNegPost = 12, TNeg = 68. Kemudian dilanjutkan dengan perhitungan klasifikasi data *training* yang dapat ditunjukkan pada Tabel 4.5.

Tabel 4.5 Hasil perhitungan klasifikasi data *training*

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
75.50	75.85	75.50	75.40

Kemudian terdapat hasil *confusion matrix* dari data *testing* yang dapat ditunjukkan pada Gambar 4.10.



Gambar 4.10 *Confusion matrix* data *testing*

Gambar 4.10 menjelaskan bahwa hasil perhitungan *confusion matrix* data *testing* didapatkan berupa $TPos = 44$, $FPostNeg = 20$, $FNegPost = 12$, $TNeg = 83$. Kemudian dilanjutkan dengan perhitungan klasifikasi data *testing* yang dapat dilihat pada Tabel 4.6.

Tabel 4.6 Hasil perhitungan klasifikasi data *testing*

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
79.88	79.77	79.87	79.61

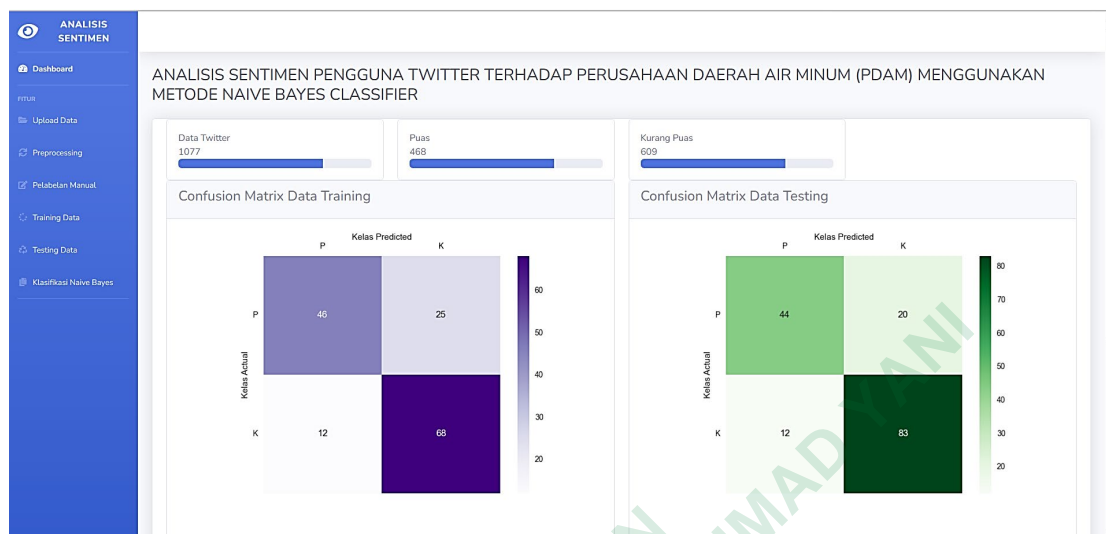
4.3 IMPLEMENTASI DESAIN ANTARMUKA

Implementasi desain halaman sistem tentang “Analisis Sentimen Pengguna Twitter Terhadap Perusahaan Daerah Air Minum (PDAM) Menggunakan Metode Naive Bayes *classifier*” dengan menggunakan bahasa pemrograman Python *framework* Flask. Tampilan halaman sistem dan kode program dapat dilihat dibawah ini.

4.3.1 TAMPILAN HALAMAN DASHBOARD ANALISIS SENTIMEN

Halaman dashboard memuat informasi hasil perhitungan *confusion matrix* data *training*, *testing*, hasil klasifikasi naive bayes dan jumlah data Twitter yang diperoleh berdasarkan komentar puas dan kurang puas. Implementasi halaman

dashboard dapat ditunjukkan pada Gambar 4.11 dan implementasi halaman klasifikasi NBC ditunjukkan pada Gambar 4.12.



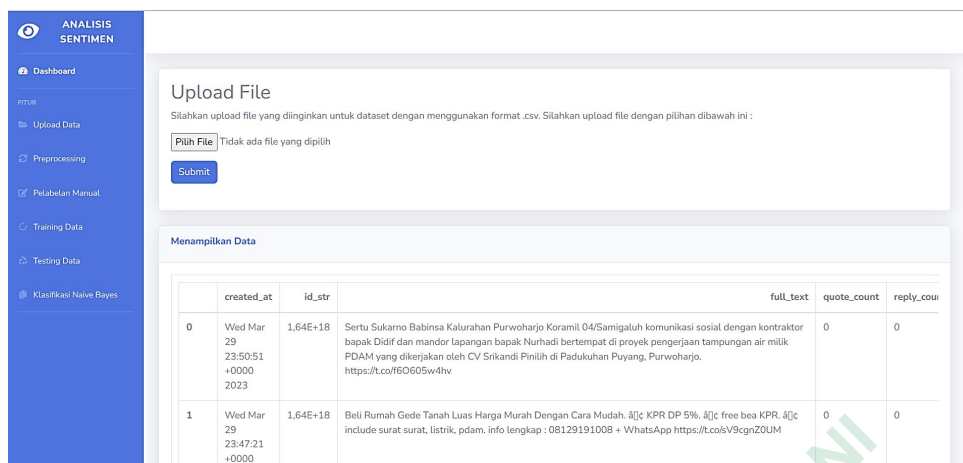
Gambar 4.11 Implementasi halaman dashboard *confusion matrix*



Gambar 4.12 Implementasi halaman dashboard klasifikasi

4.3.2 TAMPILAN HALAMAN UPLOAD DATA

Halaman data Twitter yang digunakan untuk melakukan upload file hasil *scraping* yang selanjutnya akan dilakukan proses pengolahan data seperti proses *preprocessing*, *pelabelan manual*, *training testing* dan *klasifikasi*. Implementasi halaman dapat ditunjukkan pada Gambar 4.13.



Gambar 4.13 Implementasi halaman upload data

4.3.3 TAMPILAN HALAMAN PREPROCESSING DATA

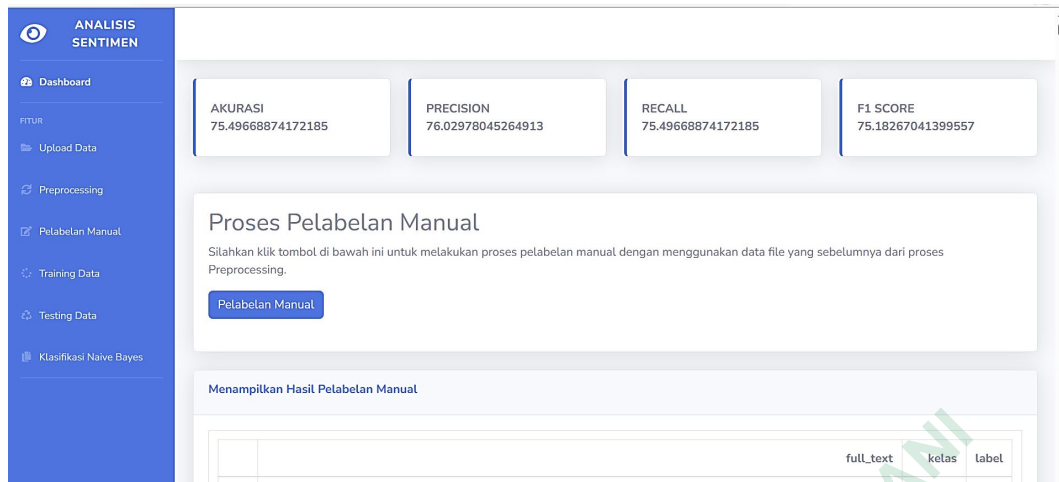
Halaman ini digunakan untuk melakukan dan menampilkan hasil proses pengolahan data teks yang diperoleh dari *scraping*. Proses ini berfungsi untuk memperbaiki data *tweet* yang belum terstruktur dan dapat ditunjukkan pada Gambar 4.14.



Gambar 4.14 Implementasi halaman *preprocessing*

4.3.4 TAMPILAN HALAMAN PELABELAN MANUAL

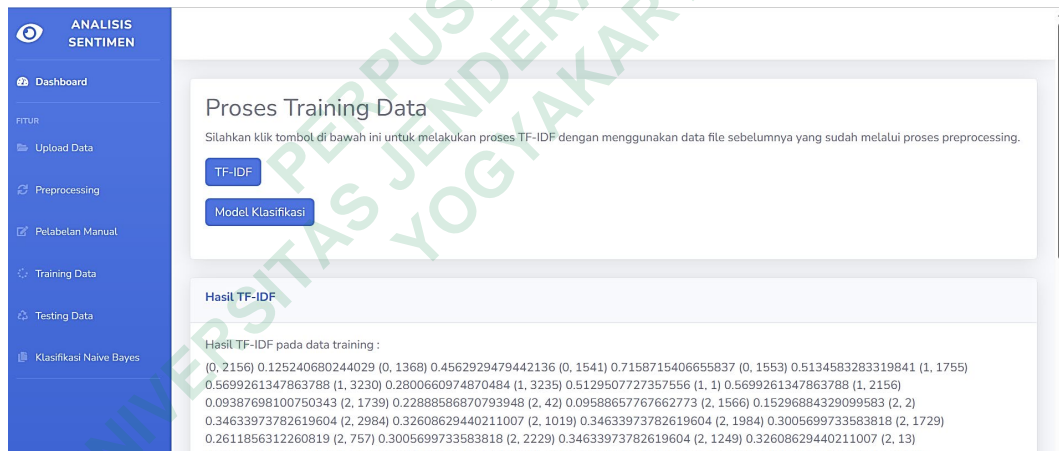
Halaman ini berfungsi untuk menampilkan hasil *Accuracy*, *Precision*, *Recall* dan *F1-score* berdasarkan data *tweet* yang sudah diberi kelas dan label secara manual yang dapat ditunjukkan pada Gambar 4.15.



Gambar 4.15 Implementasi halaman pelabelan manual

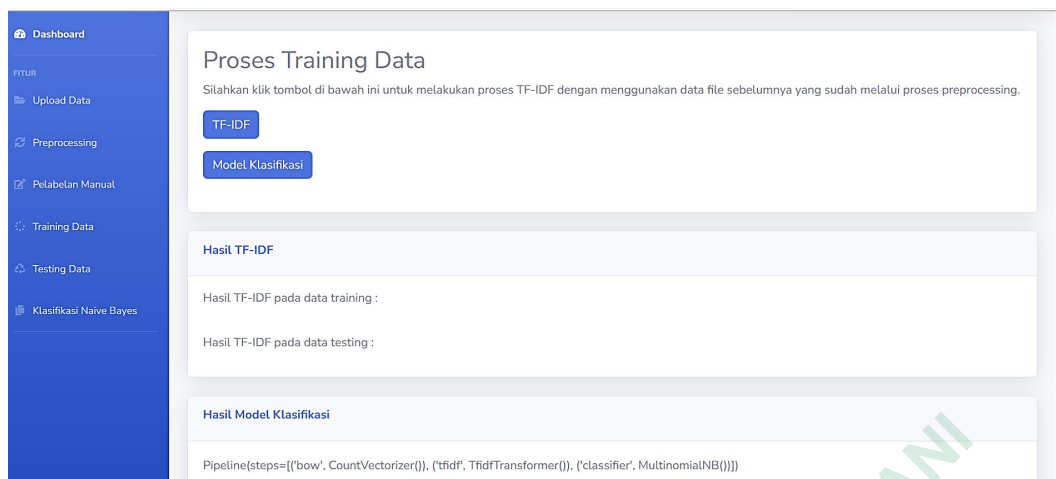
4.3.5 TAMPILAN HALAMAN TRAINING DATA

Halaman ini digunakan untuk membuat model klasifikasi dan menghitung TF-IDF secara otomatis pada data teks. Implementasi halaman *training* sistem ini dapat ditunjukkan pada Gambar 4.16.



Gambar 4.16 Implementasi halaman *training* TF-IDF

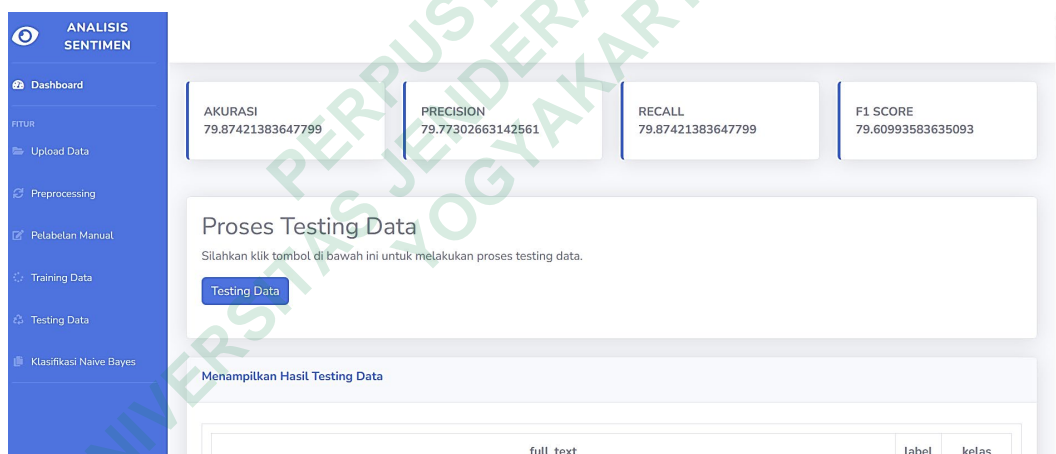
Setelah melakukan perhitungan TF-IDF secara otomatis pada sistem dilakukan permodelan klasifikasi dengan tombol button Model Klasifikasi. Implementasi halaman pembuatan model klasifikasi dapat ditunjukkan pada Gambar 4.17.



Gambar 4.17 Implementasi halaman *training* model klasifikasi

4.3.6 TAMPILAN HALAMAN TESTING DATA

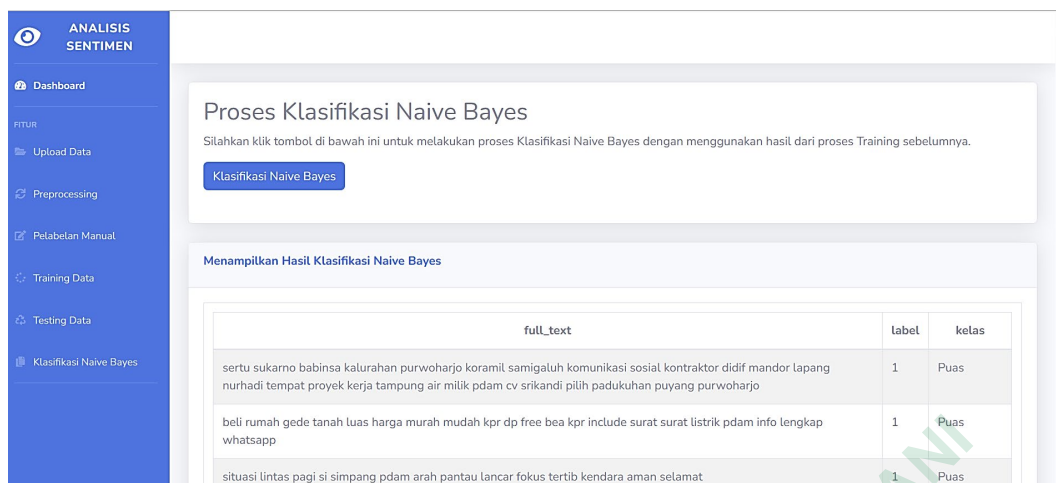
Halaman ini digunakan untuk menghitung nilai *Accuracy*, *Precision*, *Recall* dan *F1-score* data tweet yang sudah diprediksi kelas dan labelnya secara otomatis. Implementasi halaman *testing* dapat ditunjukkan pada Gambar 4.18.



Gambar 4.18 Implementasi halaman *testing* data

4.3.7 TAMPILAN HALAMAN KLASIFIKASI NAIVE BAYES

Halaman ini digunakan untuk menampilkan data *tweet* yang sudah di prediksi secara otomatis kelas dan labelnya berdasarkan model yang dibangun pada tahap *training*. Implementasi halaman ini dapat ditunjukkan pada Gambar 4.19.



Gambar 4.19 Implementasi halaman klasifikasi NBC

4.4 PEMBAHASAN

Analisis sentimen pada penelitian ini berdasarkan kata kunci “pdam”, “pdam sleman”, “pelayanan pdam”, dan “kualitas pdam” yang mendapat 2.530 data. Sebanyak 1.077 data preprocessing digunakan untuk analisis. Data tweet sebanyak 754 data yang diberi kelas dan label secara manual dengan perbandingan sama yaitu 377 data kelas puas dan 377 kelas kurang puas dengan rata-rata akurasi 75.50% dan data *tweet*.

Sistem analisis sentimen pengguna PDAM di Twitter dibangun dengan *framework* Flask. Sistem yang dibangun memiliki beberapa menu, yaitu menu upload data, *preprocessing*, pelabelan manual, *training* data, *testing* data dan klasifikasi NBC. Pada menu dashboard digunakan untuk memuat informasi hasil perhitungan *confusion matrix* data *training*, *testing*, hasil *klasifikasi* naive bayes dan jumlah data Twitter yang diperoleh berdasarkan komentar puas dan kurang puas. menu upload data digunakan untuk mengupload data file yang sudah diambil dan tersimpan di *directory*. Menu *preprocessing* digunakan untuk melakukan dan menampilkan hasil proses pengolahan data teks yang diperoleh dari hasil *scraping*. Menu pelabelan manual digunakan untuk menampilkan hasil akurasi data *tweet* yang sudah diberi kelas dan label secara manual. Menu *training* data digunakan untuk membangun model klasifikasi sentimen secara otomatis dengan menggunakan TF-IDF pada data teks. Menu *testing* data digunakan untuk menampilkan nilai akurasi berdasarkan data tweet yang di prediksi secara

otomatis untuk menentukan seberapa akurat model yang dibangun pada tahap *training*. Menu klasifikasi naive bayes digunakan untuk menampilkan hasil klasifikasi dari metode naive bayes *classifier* dengan data yang di prediksi secara otomatis.

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YANI
YOGYAKARTA