

BAB 4

HASIL PENELITIAN

4.1 RINGKASAN HASIL PENELITIAN

Hasil dari kombinasi metode *Lexicon based* dan *Naïve Bayes Classifier* untuk menganalisis data *tweet* dari Twitter untuk mendapatkan kalimat negatif dan positif dengan 3 kata kunci yaitu “BPJSTK, BPJS dan JHT” yang diambil mulai tanggal 14 Februari 2022 – 12 Juli 2022 dengan jumlah data 4.154 data *tweet* dan *retweet*. Data yang digunakan untuk *testing Lexicon Based* yaitu 600 data, untuk *training Naïve Bayes Classifier* 1.000 data terlabeli otomatis, dan *testing Naïve Bayes Classifier* 500 data terlabeli otomatis, dan untuk data klasifikasi yaitu 4.154 data terlabeli otomatis.

Penerapan analisis sentimen Twitter menggunakan metode *Lexicon Based* dan *Naïve Bayes Classifier* dengan bahasa pemrograman Python dimana menggunakan *library* yang sudah ada yaitu *VaderSentiment, pandas, numpy, nltk, sastrawi, sklearn, pickle, seaborn* dan *matplotlib* yang ditampilkan menggunakan *tools* Jupyter Notebook.

4.2 HASIL EVALUASI LEXICON BASED

Evaluasi metode *Lexicon Based* menggunakan data *testing* 600 data dimana data tersebut sudah terlabeli manual dan nantinya di bandingkan dengan pelabelan otomatis, tujuannya adalah untuk menentukan akurasi dari pelabelan otomatis menggunakan *confusion matrix* . Menggunakan rumus pada persamaan (3). Hasil perhitungan *confusion matrix* dapat dilihat pada tabel 4.1

Tabel 4. 1 Hasil Confusion Matrix

Actual	Predicted	
	Positif	Negatif
Positif	296	36
Negatif	12	256

Dari tabel 4.1 mendapatkan nilai $TP = 296$, $FP = 36$, $TN = 256$ dan $FN = 12$. Lalu lakukan perhitungan menggunakan hasil *confusion matrix* tersebut untuk mendapatkan nilai *accuracy*, *precision*, *recall* dan *F1 score*. Menggunakan rumus pada persamaan (6). Hasil perhitungan *confusion matrix* dapat dilihat pada tabel 4.2.

Tabel 4. 2 Hasil Perhitungan Confusion Matrix

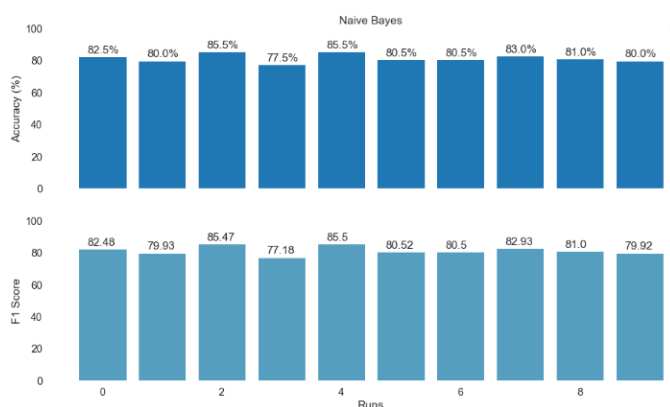
Nilai	Hasil
Accuracy	92%
Precision	89%
Recall	96%
F1-Score	92%

4.3 HASIL EVALUASI NAÏVE BAYES CLASSIFIER

4.3.1 Hasil Evaluasi Training Naïve Bayes Classifier

Evaluasi model klasifikasi menggunakan data *training* 1.000 data yang sudah terlabeli oleh mesin, dimana masing-masing terlabeli 5.00 negatif dan positif.

Kemudian dilakukan pengujian menggunakan *k-fold cross validation* dengan percobaan sebanyak 10 kali untuk mendapatkan nilai akurasi nya. Perhitungan *k-fold cross validation* dilakukan secara berulang dan memiliki akurasi yang berbeda beda. Kode untuk menghitung *k-fold cross validation* dapat dilihat seperti pada persamaan gambar 3.49 dan Tampilan presentase akurasi *k-fold cross validation* dapat dilihat pada Gambar 4.1..



Gambar 4. 1 Presentase Akurasi K-Fold Cross Validation

Setelah mendapatkan nilai presentase dari percobaan *k-fold cross validation*, tahap selanjutnya adalah menghitung nilai rata-rata *accuracy* dan *f1 score* dari presentase yang didapat. Hasil perhitungan dari nilai *accuracy* dan *f1 score* dapat dilihat di bawah ini.

Average accuracy across folds: 81.60%

Average F1 score across folds: 81.54%

Average Confusion Matrix across folds:
[[77.8 20.7]
[16.1 85.4]]

4.3.2 Hasil Evaluasi Testing Naïve Bayes Classifier

Evaluasi *testing naïve bayes classifier* menggunakan data *testing* 5.00 data yang sudah terlabeli oleh mesin, dimana masing-masing terlabeli 4.02 negatif dan 98 positif. Menggunakan rumus pada persamaan (3) Hasil *confusion matrix* dapat dilihat pada tabel 4.3.

Tabel 4. 3 Hasil Confusion Matrix

Actual	Predicted	
	Positif	Negatif
Positif	82	5
Negatif	16	397

Dari tabel 4.3 mendapatkan nilai TP = 82, FP = 5, TN = 397 dan FN = 16. Kemudian dilakukan perhitungan menggunakan *confusion matrix* untuk mendapatkan nilai *accuracy*, *precision*, *recal* dan *F1 score*. Menggunakan rumus pada persamaan (6). Hasil perhitungan *confusion matrix* dapat dilihat pada tabel 4.4

Tabel 4. 4 Hasil Perhitungan Confusion Matrix

Nilai	Hasil
Accuracy	95%
Precision	84%
Recall	83%
F1-Score	88%

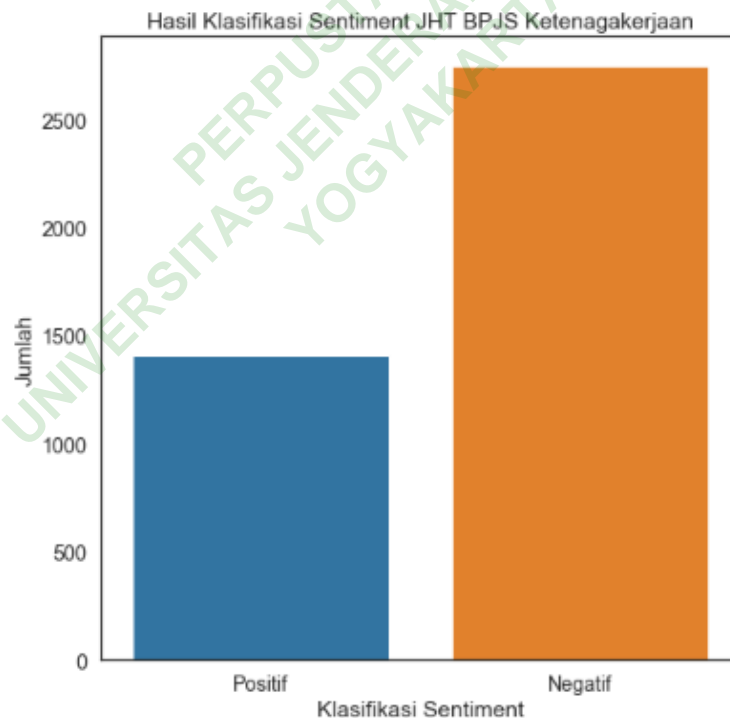
4.3.3 Hasil Evaluasi Klasifikasi

Evaluasi klasifikasi *naïve bayes classifier* menggunakan data keseluruhan yaitu 4.154 data yang sudah terlabeli oleh mesin, alasan pelabelan menggunakan mesin yaitu dikarenakan data untuk *testing* memiliki akurasi yang cukup tinggi yaitu 95%. Dari data keseluruhan dimana masing-masing terlabeli 2.746 negatif dan 1.408 positif. Jumlah tersebut dapat dilihat pada Tabel 4.5.

Tabel 4. 5 Jumlah Negatif dan Positif

Tweet	
Negatif	2.746
Positif	1.408

Lalu tampilkan jumlah dari hasil negatif dan positif dalam bentuk *histogram*. Tampilan *histogram* dapat dilihat pada Gambar 4.2.



Gambar 4. 2 Histogram

Pada hasil klasifikasi 4.154 data terlabeli 1.408 positif dan 2.746 negatif. Dimana sentimen positif pada data *tweet* membahas mengenai beberapa hal yaitu mengenai gaji dan bonus insentif dan asuransi kesehatan jaminan hari tua. Contoh data *tweet* negatif dapat dilihat pada Tabel 4.6.

Tabel 4. 6 Data Tweet Positif

NO	Data Tweet
1	pns jamin tua mana sangon gaji pensiun tetep jalan hidup sederhana bumn gaji golong atas umr lipat contoh umr semarang gaji bumn sampe bonus sangon juga
2	gapapa tidak dapat bonus tidak dapat bpjs ketenagakerjaan
3	bedain gaji sampai jamin tua tidak saya kerja gaji saya terima alhamdulillah usaha saya profit saya bonus lihat tidak paham utuh oh saya muslim baik bawa agama
4	dapat gaji sangon bonus insentif asuransi sehat thr jamin tua gaji dan lain-lain
5	nanya potong bpjstk ambil gji pokok saja bonus

Sedangkan sentimen negatif pada data *tweet* membahas mengenai beberapa hal yaitu penolakan terhadap rencana undang-undang terbaru dari kementerian ketenagakerjaan dimana dana Jaminan Hari Tua (JHT) yang hanya bisa dicairkan ketika peserta BPJS Ketenagakerjaan menginjak usia 56 tahun. Contoh data *tweet* sentimen negatif dapat dilihat pada tabel 4.7.

Tabel 4. 7 Data Tweet Negatif

NO	Data Tweet
1	ngomong fakta nama jamin tua undang undang tahun bilang khianat buruh klaim buruh tidak pilih jokowi pake senjata buruh pilih jokowi set your fact straight
2	ribut urus jaminan hari tua tidak sesuai jamin tua pada nahan hak buruh biar jadi manfaat tua atas mending ganti saja jaminan hari tua jadi jamin hak bayar apa lebih cocok
3	tolak peraturan menteri ketenagakerjaan permenaker nomor tahun cair jaminan hari tua jamin tua silah like retweet

4	tolak permenaker tahun jamin tua anggota fpks rampas hak rakyat cabut permen jht tahun
5	tolak permenaker nomor tahun dana jamin tua jaminan hari tua milik kerja perintah kerja hak ambil hak cabutpermenjhttahun

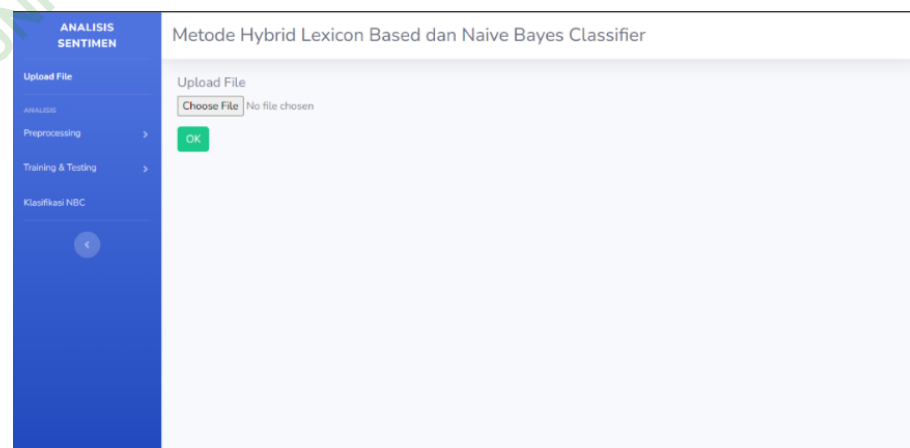
Dari hasil klasifikasi data keseluruhan dengan total data 4.154, sentimen yang didapatkan lebih banyak sentimen bersifat negatif daripada sentimen bersifat positif dimana lebih banyak masyarakat tidak menerima atau menolak mengenai rencana rancangan undang-undang Kementerian Kesehatan mengenai dana jaminan hari tua (JHT) yang hanya bisa dicairkan ketika peserta BPJS Ketenagakerjaan menginjak usia 56 tahun.

Oleh karena itu dari nilai sentimen negatif dapat digunakan untuk mengetahui mengenai opini dari masyarakat di Twitter tentang jaminan hari tua (JHT) BPJS Ketenagakerjaan yang merugikan masyarakat.

4.4 HASIL WEB DASHBOARD

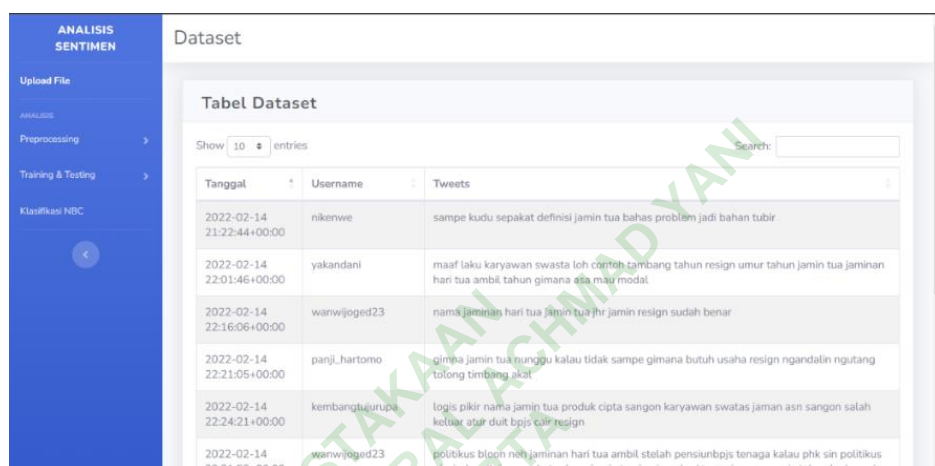
Dari hasil penelitian menggunakan metode *Lexicon Based* dan *Naive Bayes Classifier* maka tercipta sebuah sistem analisis sentimen berupa *web dashboard*, berikut adalah tampilan dari dashboard web sistem analisis sentimen.

1. Tampilan awal *dashboard*, pada tahap ini kita melakukan upload file data hasil dari scrapping data pada *jupyter notebook*. Tampilan awal *dashboard* dapat dilihat pada Gambar 4.3.



Gambar 4. 3 Tampilan Awal Dashboard

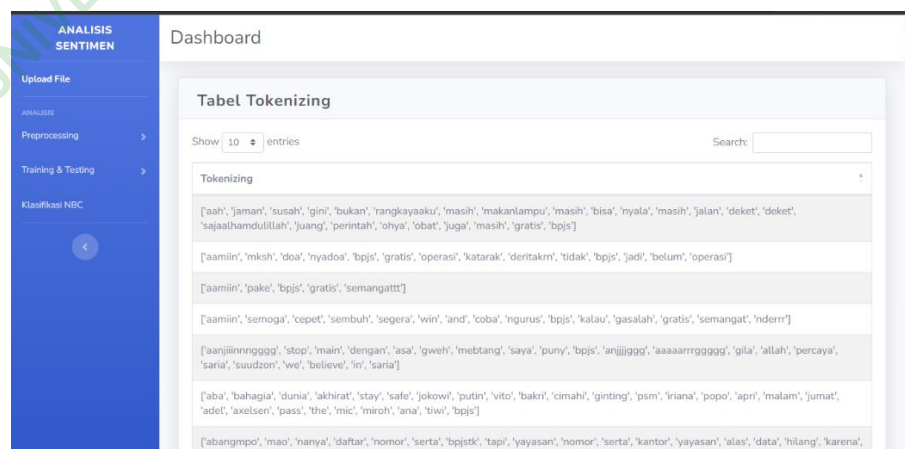
2. Berikutnya merupakan tampilan dari menu data *preprocessing*, pada menu *preprocessing* terdapat beberapa menu lagi antara lain dataset, tabel *tokenizing*, tabel *stopword*, dan tabel *stemming*, tabel *cleaning*, Tabel dataset, pada menu ini akan menampilkan file yang telah kita *upload*. Tampilan dataset dapat dilihat pada Gambar 4.4.



Tanggal	Username	Tweets
2022-02-14 21:22:44+00:00	nikenwe	sampe kudu sepatok definisi jamin tua bahas problem jadi bahan tubir
2022-02-14 22:01:46+00:00	yakandani	maaf laku karyawan swasta loh canteh tambang tahun resign umur tahun jamin tua jaminan hari tua ambil tahun gimana asa mau modal
2022-02-14 22:16:06+00:00	wanwijoged23	nama jaminan hari tua jamin tua ikr jamin resign sudah benar
2022-02-14 22:21:05+00:00	panji_hartomo	gimana jamin tua nunggu kalau tidak sampe gimana butuh usaha resign ngandalin ngutang tolong timbang akal
2022-02-14 22:24:21+00:00	kembangjurupa	logis pikir nama jamin tua produk cipta sangon karyawan swatas jaman asn sangon salah ketuar atur duit bpjs car resign
2022-02-14 22:31:52+00:00	wanwijoged23	politik bloon neh jaminan hari tua ambil stelah pensiunbpjs tenaga kalau phk sin politikus pik instruksi memusd atur baru jamin tua jaminan hari tua cair aneun usia tahun kerja nusi

Gambar 4. 4 Tabel Dataset

3. Tabel *tokenizing*, *tokenizing* merupakan proses untuk melakukan pemecahan atau pemisahan karakter dalam suatu teks yang didefinisikan sebagai pemisah kata atau bukan. Tampilan tabel *tokenizing* dapat dilihat pada Gambar 4.5.



Tokenizing
['aah', 'jaman', 'susah', 'gini', 'bukan', 'rangkayaaku', 'masih', 'makanlampu', 'masih', 'bisa', 'nyala', 'masih', 'jalan', 'deket', 'deket', 'sajaalhamdulillah', 'juang', 'perintah', 'ohya', 'obat', 'juga', 'masih', 'gratis', 'bpjs']
['aamiin', 'mksh', 'doa', 'nyadoa', 'bpjs', 'gratis', 'operasi', 'katarak', 'deritakrn', 'tidak', 'bpjs', 'jadi', 'betum', 'operasi']
['aamiin', 'pake', 'bpjs', 'gratis', 'semangatt!']
['aamiin', 'semoga', 'cepat', 'sembuh', 'segera', 'win', 'and', 'coba', 'ngurus', 'bpjs', 'kalau', 'gasalah', 'gratis', 'semangat', 'nderrn']
['aanjinngggg', 'stop', 'main', 'dengan', 'asa', 'gweh', 'mebtang', 'saya', 'puny', 'bpjs', 'anjijggg', 'aaaaarrggggg', 'gila', 'allah', 'percaya', 'saria', 'suudzon', 'we', 'believe', 'in', 'saria']
['aba', 'bahagia', 'dunia', 'akhirat', 'stay', 'safe', 'jokowi', 'putih', 'vito', 'bakri', 'cimahi', 'ginting', 'psm', 'iriana', 'popo', 'apri', 'malam', 'jumat', 'adet', 'axelsen', 'pass', 'the', 'mic', 'miroh', 'ana', 'tiwi', 'bpjs']
['abangmpo', 'mao', 'nanya', 'daftar', 'nomor', 'serta', 'bpjst', 'tapi', 'yayanan', 'nomor', 'serta', 'kantor', 'yayanan', 'alas', 'data', 'hilang', 'karena', 'kita']

Gambar 4. 5 Tabel Tokenizing

4. Berikutnya merupakan tampilan *stopword*, tahap ini Melakukan penghapusan kata-kata yang memiliki informasi rendah dari sebuah teks (“yang”, “dan”, “di”, “dari” dll). Tampilan *stopword* dapat dilihat pada Gambar 4.6.

Stopword
[sah, 'aman', 'susah', 'gini', 'rangkayaku', 'makanlampu', 'nyala', 'jalan', 'deket', 'deket', 'sajaahmdullillah', 'juang', 'perintah', 'ohya', 'obat', 'gratis', 'bpjs]
[aamin, 'mksh', 'doa', 'nyadoa', 'bpjs', 'gratis', 'operasi', 'katarak', 'deritakm', 'bpjs', 'operasi]
[aamin, 'pake', 'bpjs', 'gratis', 'semangatt]
[aamin, 'semoga', 'cepat', 'sembuh', 'win', 'and', 'coba', 'ngurus', 'bpjs', 'gasalah', 'gratis', 'semangat', 'hderrr]
[aanginnngggg, 'stop', 'main', 'asa', 'gweh', 'mebtang', 'puny', 'bpjs', 'anjingg', 'aaaaarrggggg', 'gila', 'allah', 'percaya', 'saria', 'suudzon', 'we', 'believe', 'in', 'saria]
[aba, 'bahagia', 'dunia', 'akhirat', 'stay', 'safe', 'jokowi', 'putin', 'vito', 'bakri', 'cuma', 'ginting', 'psm', 'riana', 'popo', 'apri', 'malam', 'Jumat', 'adel', 'axelsen', 'pass', 'the', 'mic', 'miroh', 'ana', 'tww', 'bpjs]
[abangppo, 'mao', 'nanya', 'daftar', 'nomor', 'bpjs', 'yayasan', 'nomor', 'kantor', 'yayasan', 'alas', 'data', 'hilang', 'instalasi', 'data', 'hilang', 'hderrr', 'hderrr]

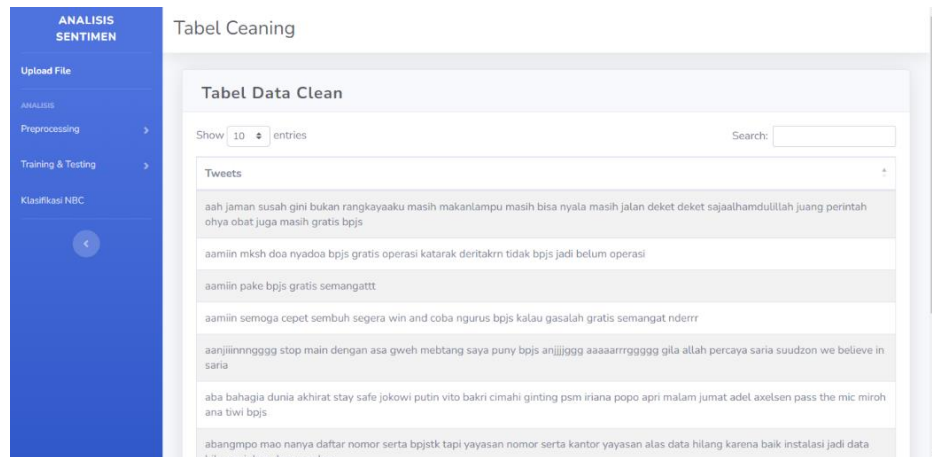
Gambar 4. 6 Tabel Stopword

5. Berikutnya tampilan *stemming*, Proses untuk merubah semua kata menjadi kata dasar. Tampilan *stemming* dapat dilihat pada Gambar 4.7.

Stemming
[admin tolong balas mana urgent terimakasih]
[agreed jalan big ngerasin rada oot dikit mental issue notabene gabakal sembuh sampe mati psikiater its more more better gaada duit psikiater make bpjs]
[aktif tingkat persen banding tahun bpjs ketenagakerjaan service level agreement sla bayar klaim]
[booby sehat obat habis buru obat sanggup bayar bpjs beli bbm ojo gratis]
[bpjs iur ha ha ha ha]
[bpjs kesehatanbayar gratis lohhtapi layan rs layan rs utama bayar tunaiiii]
[bpjs nanggung sakit kasi sakit nakal makan hak]
[bpjs nanggung obat generik harga seribuankalau obat mahal beli sendrikalau biaya inap operasi gratis]

Gambar 4. 7 Tabel Stemming

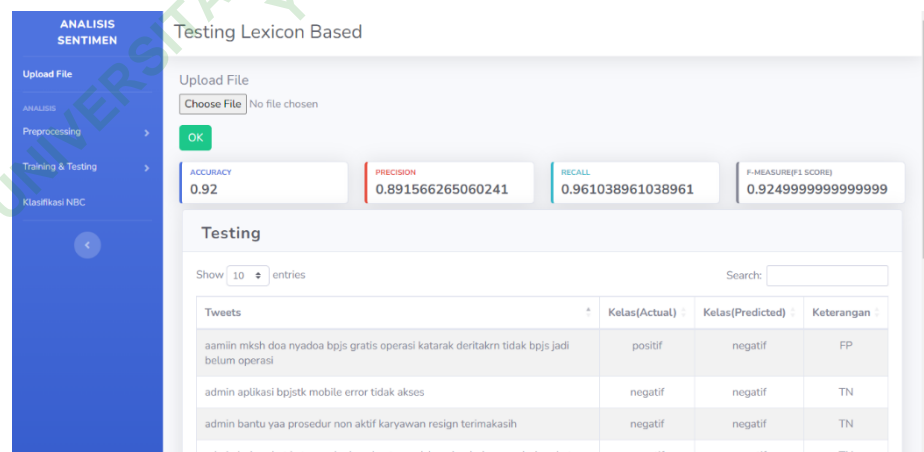
6. Tabel *cleaning*, itu sendiri merupakan proses pembersihan atribut-atribut yang tidak penting yang terdapat dalam dokumen. Tampilan data *cleaning* dapat dilihat pada Gambar 4.8.



Gambar 4. 8 Tabel Cleaning

Berikutnya merupakan menu *training* dan *testing*, pada menu ini terdapat beberapa menu lagi yaitu, *testing Lexicon Based*, *training Naïve Bayes Classifier*, *predicted*, dan *testing Naïve Bayes Classifier*.

1. Tampilan *testing Lexicon based*, pada tampilan ini terdapat menu *upload file* dimana berfungsi untuk *mengupload* file yang akan digunakan untuk *testing* pada *Lexicon Based* dan langsung menampilkan data yang telah di upload dan menampilkan nilai dari *accuracy*, *precision*, *recall* dan *f1-score*. Tampilan *testing Lexicon Based* dapat dilihat pada Gambar 4.9.



Gambar 4. 9 Testing Lexicon Based

2. Tampilan *training naïve bayes classifier*, pada tampilan ini terdapat menu *upload file* dimana berfungsi untuk *mengupload* file yang akan digunakan untuk *training* pada *naïve bayes classifier* dan langsung

menampilkan data yang telah di upload dan menampilkan jumlah nilai negatif dan positif, nilai *confusion matrix*, nilai *accuracy* dan nilai *f-measure* dari data *training*. Tampilan *training naïve bayes classifier* dapat dilihat pada Gambar 4.10.

Tweets	Label	Kelas
aamin mksh doa nyadoo bpjs gratis operasi katarak deritakrn tidak bpjs jadi belum operasi	negatif	-1
aamin pake bpjs gratis semangatt	positif	1
aanjilnngggg stop main dengan asa gw eh mbuang saya punya bpjs anjilngggg aaaaaarrggggg gila allah percaya saria suudzon we believe in saria	negatif	-1

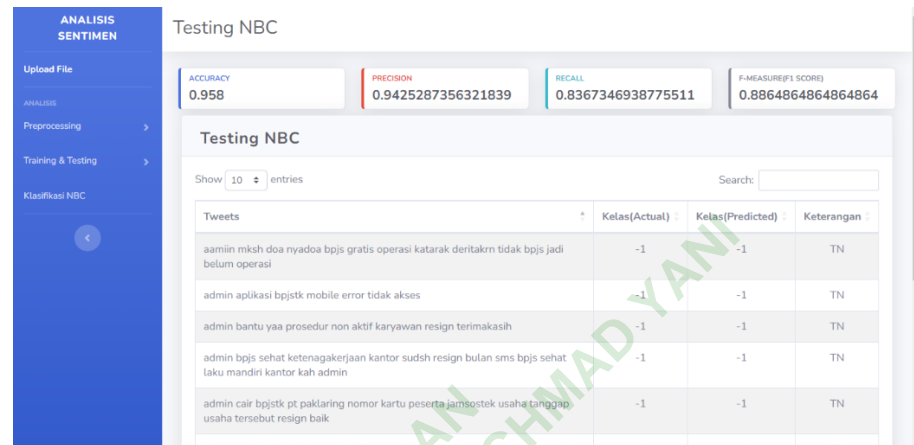
Gambar 4. 10 Training Naive Bayes Classifier

3. Tampilan *predicted naïve bayes classifier*, pada tampilan ini menampilkan data hasil dari prediksi mesin atau pelabelan otomatis dari mesin. Tampilan *predicted naïve bayes classifier* dapat dilihat pada Gambar 4.11.

Tweets	Kelas(Predicted)
aamin mksh doa nyadoo bpjs gratis operasi katarak deritakrn tidak bpjs jadi belum operasi	-1
admin aplikasi bpjstk mobile error tidak akses	-1
admin bantu yaa prosedur non aktif karyawan resign terimakasih	-1
admin bpjs sehat ketenagakerjaan kantor sudsh resign bulan sms bpjs sehat laku mandiri kantor kah admin	-1
admin cair bpjstk pt paklaring nomor kartu peserta jamsstek usaha tanggap usaha tersebut resign baik	-1
admin cairin bpjstk tpi status aktif hari usaha kasih rekomendasi ngurus bpjstk resign	-1
admin cairin bpjstk usaha alas aktif cek aktif admin	-1
admin daftar bpjs mobile jaminan kesehatan nasional pas proses autodebet error aplikasi pas tidak gimana nerusin daftar nomor bpjs juga belum terima	-1

Gambar 4. 11 Tabel Predicted Naive Bayes Classifier

4. Tampilan *testing naïve bayes classifier*, pada tampilan ini menampilkan data *testing naïve bayes classifier*, dan menampilkan nilai dari *accuracy*, *precision*, *recall* dan *f1-score*. Tampilan *testing naïve bayes classifier* dapat dilihat pada Gambar 4.12.

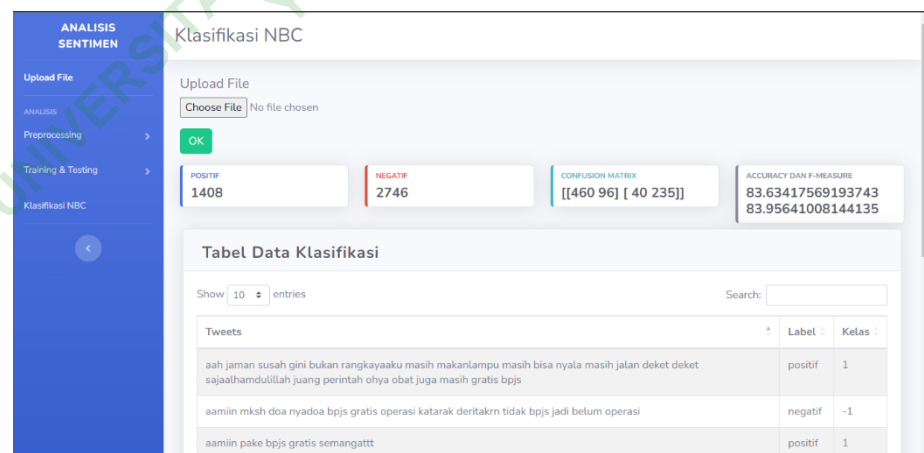


The screenshot shows the 'Testing NBC' interface. At the top, there are four performance metrics: ACCURACY (0.958), PRECISION (0.9425287356321839), RECALL (0.8367346938775511), and F-MEASURE(F1 SCORE) (0.8864864864864864). Below these is a table titled 'Testing NBC' with columns for Tweets, Kelas(Aktual), Kelas(Predicted), and Keterangan. The table contains five rows of tweet text and their corresponding classifications.

Tweets	Kelas(Aktual)	Kelas(Predicted)	Keterangan
aamin mksh doa nyadoo bpjs gratis operasi katarak deritakrn tidak bpjs jadi belum operasi	-1	-1	TN
admin aplikasi bpjstc mobile error tidak akses	-1	-1	TN
admin bantu yaa prosedur non aktif karyawan resign terimakasih	-1	-1	TN
admin bpjs sehat ketenagakerjaan kantor sudsh resign bulan sms bpjs sehat laku mandiri kantor kah admin	-1	-1	TN
admin cair bpjstc pt paklaring nomor kartu peserta jamkostek usaha tanggap usaha tersebut resign baik	-1	-1	TN

Gambar 4. 12 Testing Naïve Bayes Classifier

Berikutnya merupakan tampilan klasifikasi, pada tampilan ini terdapat fungsi untuk *upload* file data klasifikasi atau data keseluruhan dan nantinya akan menghitung nilai dari jumlah negatif dan positif, *confusion matrix* dan *f measure*. Tampilan kladifikasi dapat dilihat pada Gambar 4.13.



The screenshot shows the 'Klasifikasi NBC' interface. It features an 'Upload File' section with a 'Choose File' button and an 'OK' button. Below this, there are four performance metrics: POSITIF (1408), NEGATIF (2746), CONFUSION MATRIX ([460 96] [40 235]), and ACCURACY DAN F-MEASURE (83.63417569193743, 83.95641008144135). Below these is a table titled 'Tabel Data Klasifikasi' with columns for Tweets, Label, and Kelas. The table contains three rows of tweet text and their corresponding labels and classes.

Tweets	Label	Kelas
aah jaman susah gini bukan rangkayaaku masih makanlampu masih bisa nyala masih jalan dekat dekat sajaahamdulillah juang perintah ohya obat juga masih gratis bpjs	positif	1
aamin mksh doa nyadoo bpjs gratis operasi katarak deritakrn tidak bpjs jadi belum operasi	negatif	-1
aamin pake bpjs gratis semangattt	positif	1

Gambar 4. 13 Tabel Klasifikasi