

## **BAB 4**

### **HASIL PENELITIAN**

#### **4.1 RINGKASAN HASIL PENELITIAN**

Hasil penelitian dari penerapan analisis metode *support vector machine* untuk melakukan analisis data *tweet* berbahasa Indonesia dari media sosial Twitter baik sentimen positif maupun negatif berdasarkan data yang sudah diperoleh dari tanggal 1 januari – 30 juni 2020 dengan menggunakan kata kunci ”#ruangguru, #zenius, #quipper, #rumah belajar, dan #cakap” yang berjumlah 24.311 data yang merupakan gabungan data *tweet* dan *re-tweet*. Pada tahap *training* data yang digunakan sejumlah 798 *tweet* dengan masing-masing 399 positif dan 399 negatif yang sudah diberi label secara manual. Sedangkan untuk data *testing* jumlah data yang digunakan adalah 23.511 data yang sudah diberi label positif dan negatif. Hasil dari akurasi data *training* sebesar 91.81% dan nantinya hasil penelitian juga diharapkan dapat memberikan dampak yang baik mengenai ulasan tentang penggunaan aplikasi bimbingan belajar *online* pada masa pandemi di media sosial Twitter. Berikut merupakan penerapan analisis sentimen Twitter dengan menggunakan bahasa pemrograman Python dengan metode *support vector machine* dengan bantuan beberapa *library* dari Python seperti *pandas*, *numpy*, *nlk*, *sklearn*, *sastrawi*, *pickle* dan *matplotlib* dalam pemrosesan analisis sentimen pada Jupyter Notebook.

#### **4.2 HASIL EVALUASI DAN MODEL KLASIFIKASI**

Evaluasi model klasifikasi menggunakan 798 data *tweet* untuk melakukan data *training* yang dimana masing-masing data 399 positif dan 399 negatif yang sudah dilabel secara manual. Kemudian perlu dilakukan pengujian *confusion matrix* untuk melihat nilai data aktual dan prediksi guna untuk memeriksa keakuratan aplikasi yang dibuat. Berikut hasil dari pengujian *confusion matrix* dapat dilihat pada Tabel 4.1

Setelah itu dapat melakukan perhitungan pada akurasi model yaitu *cross validation*, dengan penghitungan *confusion matrix* pada *cross validation* untuk menguji keakuratan dengan melihat nilai data aktual dan prediksi. Berikut hasil perhitungan *cross validation* dapat dilihat pada Tabel 4.1

Tabel 4.1 Confusion Matrix Cross Validation

Kelas Prediksi	Kelas Aktual	
	Positif	Negatif
Positif	69	9.9
Negatif	3.8	77.3

Setelah mendapatkan nilai *confusion matrix* dapat dilanjutkan untuk penghitungan akurasi model dari *cross validation*. Menguji k-fold cross validation dilakukan agar mengetahui angka yang sesuai dari perulangan yang sebanyak 10 kali. Dan perhitungan tiap fold yang terdapat pada k-fold cross validation sudah mendapatkan nilainya sendiri-sendiri dan hasil yang berbeda pada tiap-tiap foldnya. Berikut hasil perhitungan 10 fold cross validation pada *accuracy* dan *f-measure* terdapat pada Tabel 4.2

Tabel 4.2 Nilai K-fold Accuracy dan F-measure

Fold	Accuracy	F-measure
Fold 1	89,4 %	89,37 %
Fold 2	94,4 %	94,36%
Fold 3	89,4 %	89,36%
Fold 4	91,2 %	91,25%
Fold 5	91,9 %	91,87%
Fold 6	89,4 %	89,26%
Fold 7	96,6 %	96,62%

Fold 8	92,5 %	92,49%
Fold 9	90,6 %	90,62%
Fold 10	93,8 %	93,75%

Dari hasil yang sudah didapatkan dari perhitungan *cross validation* yang sudah dilakukan sebanyak 10 kali maka didapatkan nilai rata-rata akurasi yang baik yaitu 0,96 untuk *accuracy* dan 0,96 untuk *f-measure*.

Pada tahap selanjutnya yaitu data *testing* jumlah data *tweet* yang digunakan adalah 200 dan masing-masing telah diberi label 100 positif dan 100 negatif secara manual. Dan untuk mengetahui nilai akurasi pada tahap testing ini diperlukan penghitung *confusion matrix* agar mengetahui perbedaan nilai antara data *training* dan data *testing*. Berikut hasil perhitungan *confusion matrix* data *testing* dapat dilihat pada Tabel 4.3

Tabel 4.3 Confusion Matrix Data Testing

Kelas Prediksi	Kelas Aktual	
	Positif	Negatif
Positif	143	57
Negatif	52	148

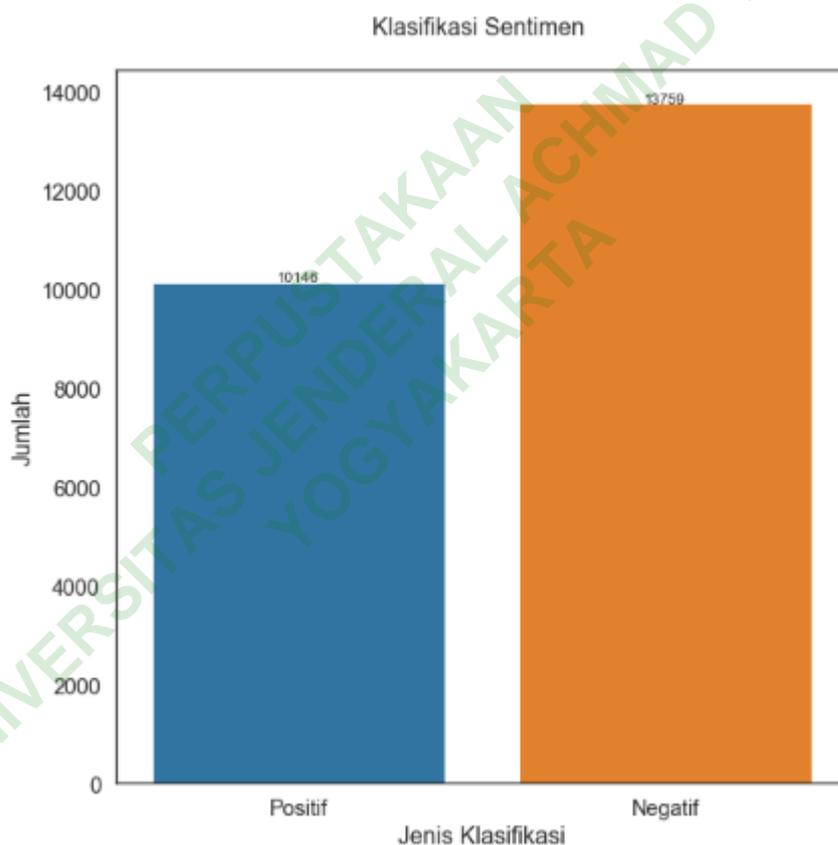
Maka dapat menghitung nilai akurasi pada data testing yang dimana sudah diketahui nilai TP = 92, TN = 90, FP = 8 dan FN = 10 dengan menggunakan rumis persamaan (1) untuk *accuracy* dan rumus persamaan (5) untuk menghitung nilai *f-measure*. Berikut untuk mengetahui nilai *accuracy* dan *f-measure* pada data *testing* pada Tabel 4.4

Tabel 4.4 Akurasi Data Testing

Akurasi Data Testing	Hasil
Accuracy	72 %
F-measure	72 %

### 4.3 HASIL ANALISIS

Langkah untuk tahap selanjutnya pada hasil analisis yaitu menghitung jumlah data keseluruhan data *tweet* yang sudah terambil dengan total 24.311 data, dengan rincian 798 data *tweet* yang telah dilabeli secara manual dengan masing-masing nilai yang sama dalam proses *training* kemudian dari data 23.913 diambil 200 data *tweet* untuk proses testing yang dimana telah didapatkannya nilai akurasi yang baik dari tahap *training* dan *testing*, hasil prediksi pada data *tweet* keseluruhan yang sudah diketahui hasil positif dan negatif dapat dilihat pada Gambar 4.1



Gambar 4.1 Histogram data positif dan negatif

Pada Gambar 4.1 telah dijelaskan bahwa didapatkan data 10.146 dengan label positif dan 13.759 dengan label negatif. Pada sentimen positif banyak warga Twitter yang membahas mengenai potongan harga pada biaya langganan aplikasi bimbingan belajar *online*, admin dengan sigap dalam menanggapi keluhan

pengguna aplikasinya, kemudian membahas paket belajar yang lengkap. Berikut Contoh beberapa data *tweet* positif terdapat pada Tabel 4.5

Tabel 4.5 Data Tweet Positif

No	Data Tweet
1	siang kak sarah mimin bantu jelasin kalo paket bebas akses ilmupedia pakai buat akses beberapa aplikasi ikut quipper zenius cakap bahasa rumah belajar
2	yuk dirumahterusbelajar bahasa inggris aplikasi cakap biar dapat akses kamu tukar telkomselpoin kamu fitur live chatnya
3	mau jago bahasa inggris lancar tetap hemat solusi aplikasi cakap sob cobain dulu nih sesi live chatnya
4	quipper mantep sih pengetahuan dan pemahaman umum nya gw
5	quipper gratis sampe malem besok malem gais
6	belajar zenius sama ruangguru ikut tryout masukkampus
7	alhamdulillah dijelasin bentar sudah aga nyambung bank soal ruangguru

Sedangkan untuk sentimen negatif pada data *tweet* banyak yang membahas beberapa permasalahan yang dialami para warga Twitter diantara lain permasalahan jaringan yang lambat pada pengguna aplikasi, soal-soal latihan yang terlalu susah, bug pada aplikasi, kemudian ada beberapa materi yang tidak tercover. Dari beberapa tanggapan warga Twitter banyak kritikan dan keluhan mengenai aplikasi bimbingan belajar. Berikut contoh beberapa data *tweet* negatif terdapat pada Tabel 4.6

Tabel 4.6 Data Tweet Negatif

No	Data Tweet
1	quipper diselesain hari hari gak sih gue mo nonton keburu males lihat durasi
2	inten mirip darimananya sih susah banget gitu
3	iya anjir soal paling susah yang aku kerjain
4	iya gua bodoh makanya download ruangan guru biar pintar
5	karena dimensi tiga tuh lihat susah tidak sih btw shaa testimoninya

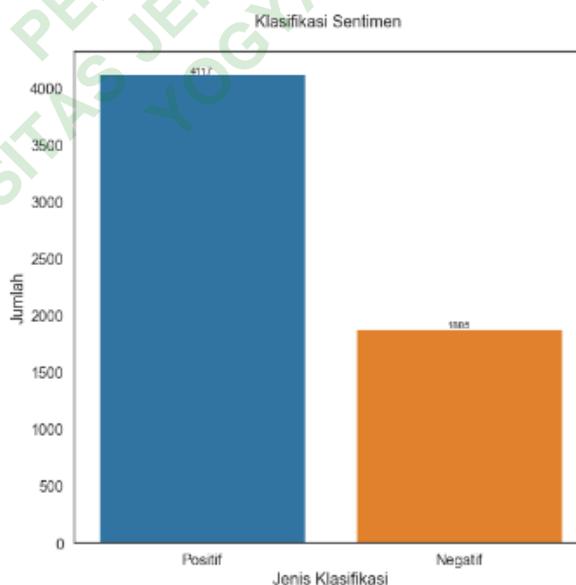
	make quipper dongg
6	latihan soal quipper gaada akhlak banget woeeeeeeee susah banget pola tp gpp moga makin jago wkwkw
7	pengetahuan dan pemahaman umum quipper tuh emang susah apa emang saya yang bodoh pengetahuan dan pemahaman umum

Berdasarkan hasil klasifikasi sentimen pada data *tweet* banyak yang diklasifikasikan ke negative dikarenakan permasalahan jaringan yang lambat pada pengguna aplikasi, soal-soal latihan yang terlalu susah, bug pada aplikasi, kemudian ada beberapa materi yang tidak tercover.

#### 4.4 HASIL REKOMENDASI KLASIFIKASI

##### 4.4.1 Quipper

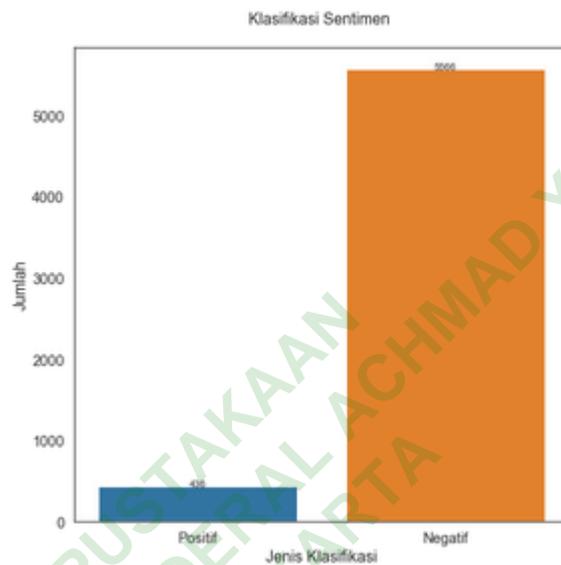
Pengambilan data Quipper sebanyak 6002 data. Dari data tersebut di proses menggunakan model SVM pickle, untuk mendapatkan hasil klasifikasi. Hasil akhir dari proses klasifikasi mendapatkan 4117 data sentimen positif dan 1885 sentimen negatif. Adapun grafik klasifikasi seperti Gambar 4.2



Gambar 4.2 Hasil klasifikasi Quipper

#### 4.4.2 Rumah Belajar

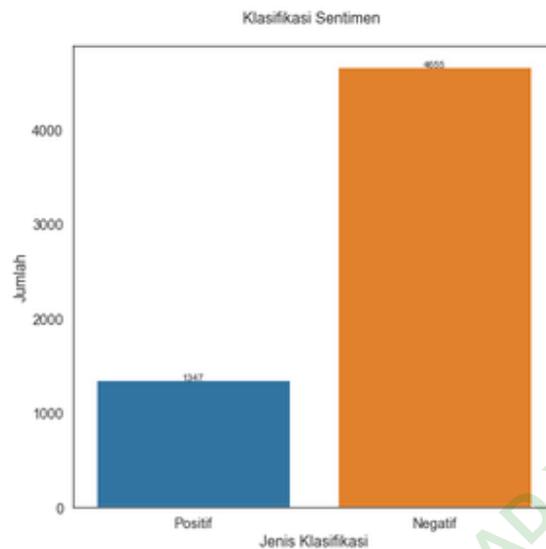
Pengambilan data Rumah Belajar sebanyak 6002 data. Dari data tersebut di proses menggunakan model SVM pickle, untuk mendapatkan hasil klasifikasi. Hasil akhir dari proses klasifikasi mendapatkan 436 data sentimen positif dan 5566 sentimen negatif. Adapun grafik klasifikasi seperti Gambar 4.3



Gambar 4.3 Hasil klasifikasi Rumah Belajar

#### 4.4.3 Zenius

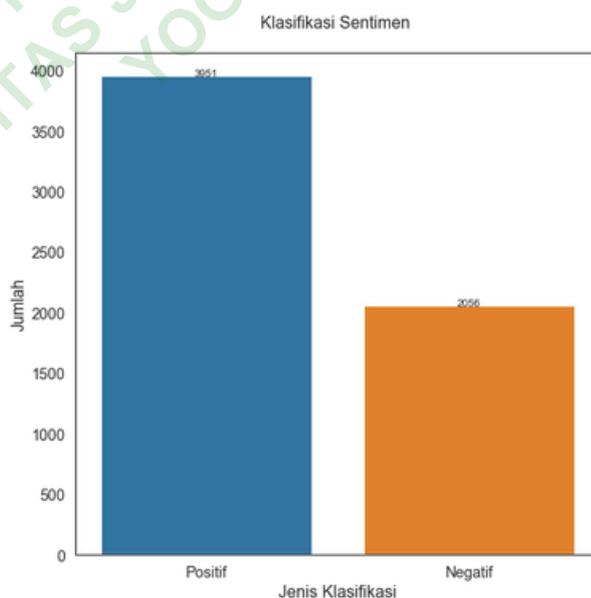
Pengambilan data Zenius sebanyak 6002 data. Dari data tersebut di proses menggunakan model SVM pickle, untuk mendapatkan hasil klasifikasi. Hasil akhir dari proses klasifikasi mendapatkan 1347 data sentimen positif dan 4655 sentimen negatif. Adapun grafik klasifikasi seperti Gambar 4.4



Gambar 4.4 Hasil klasifikasi Zenius

#### 4.4.4 Ruang Guru

Pengambilan data Ruang Guru sebanyak 6007 data. Dari data tersebut di proses menggunakan model SVM pickle, untuk mendapatkan hasil klasifikasi. Hasil akhir dari proses klasifikasi mendapatkan 3951 data sentimen positif dan 2056 sentimen negatif. Adapun grafik klasifikasi seperti Gambar 4.5



Gambar 4.5 Hasil klasifikasi Ruang guru

## 4.5 IMPLEMENTASI SISTEM

### 4.5.1 Fitur Upload File

Tampilan ini berguna untuk mengupload file dari hasil *crawling* data dalam bentuk format CSV. Berikut tampilan menu upload filenya dapat dilihat pada Gambar 4.6



Gambar 4.6 Upload File

### 4.5.2 Fitur Dataset Raw

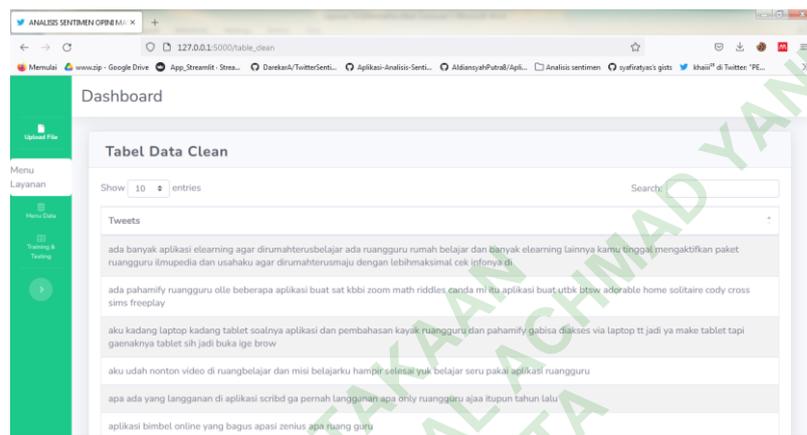
Tampilan menu dataset raw ini berfungsi untuk menampilkan data file CSV dari hasil upload. Isinya ada beberapa kolom yaitu Tanggal yang dimana data tersebut belum dilakukannya proses *pre-processing*, *username* dan *tweets*, Berikut tampilan menu *dataset raw* dapat dilihat pada Gambar 4.7

Tanggal	Username	Tweets
2020-06-12 15:20:05+00:00	Lyrelai	@IM3Ooredoo min, kuota imclassnya tidak bisa sama sekali, tidak bisa akses ruangguru via web ataupun pun via aplikasi, tolong perbaiki secepatnya
2020-06-13 09:01:19+00:00	nuangguru	@ESTev_17 Halo, @ESTev_17, Apakah ada yang bisa bantu terkait Aplikasi Ruangguru? 🙏
2020-06-13 08:05:05+00:00	ShTarkiyah	Bimbel CPNS 🙏🙏🙏 Dapatkan diskon untuk pembelian paket ruangbelajar dengan kode TARKIFLA4ZCTRHSU, Kamu juga bisa follow aku di Ruangguru, cari aku dengan kode TARKIFLA4ZCTRHSU juga yal Download aplikasi Ruangguru https://t.co/XcStaisiNen
2020-06-13 08:06:50+00:00	nuangguru	@awdult Halo, @awdult, Apakah ada yang bisa bantu terkait Aplikasi Ruangguru? 🙏
2020-06-13 10:46:16+00:00	nuangguru	@Izardhalcyon @intheadark @yopiedan_ @mandala_sertens @ejkayama @za_risaldy21

Gambar 4.7 Dataset Raw

### 4.5.3 Fitur Tabel *Cleaning*

Tampilan ini menjelaskan *dataset* raw yang sudah diproses *Cleaning* kemudian didapatkan data yang telah bersih dan dapat digunakan untuk ke tahap selanjutnya yaitu tahap *tokenizing*, Berikut hasil tabel data *cleaning* dapat dilihat pada Gambar 4.8

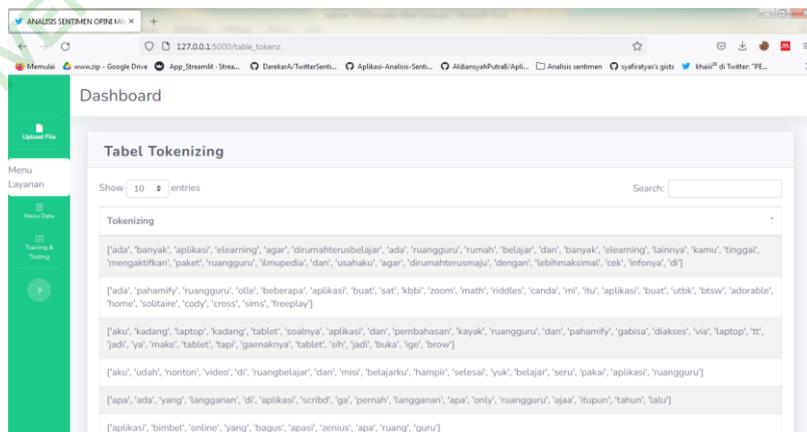


Tweets
ada banyak aplikasi elearning agar dirumahterusbelajar ada ruangguru rumah belajar dan banyak elearning lainnya kamu tinggal mengaktifkan paket ruangguru ilmupedia dan usahaku agar dirumahterus maju dengan lebih maksimal cek infonya di
ada pahamify ruangguru olle beberapa aplikasi buat sat kbbi zoom math riddles canda mi ita aplikasi buat utbk btw adorable home solitaire cody cross sims freeplay
aku kadang laptop kadang tablet soalnya aplikasi dan pembahasan kayak ruangguru dan pahamify gabisa diakses via laptop tt jadi ya make tablet tapi gaenaknya tablet sih jadi buka ige brow
aku udah nonton video di ruangbelajar dan misi belajarku hampir selesai yuk belajar seru pakai aplikasi ruangguru
apa ada yang langganan di aplikasi scribd ga pernah langganan apa only ruangguru ajaa stupun tahun lalu
aplikasi bimbel online yang bagus apasi zenius apa ruang guru

Gambar 4.8 Tabel *Cleaning*

### 4.5.4 Fitur Tabel *Tokenizing*

Pada tampilan ini menjelaskan data yang sudah dilakukan *cleaning* kemudian lanjut ke tahap *tokenizing* yang dimana teks pada data *tweet* dipecah menjadi potongan-potongan kata, Berikut hasil data proses *tokenizing* dapat dilihat pada Gambar 4.9

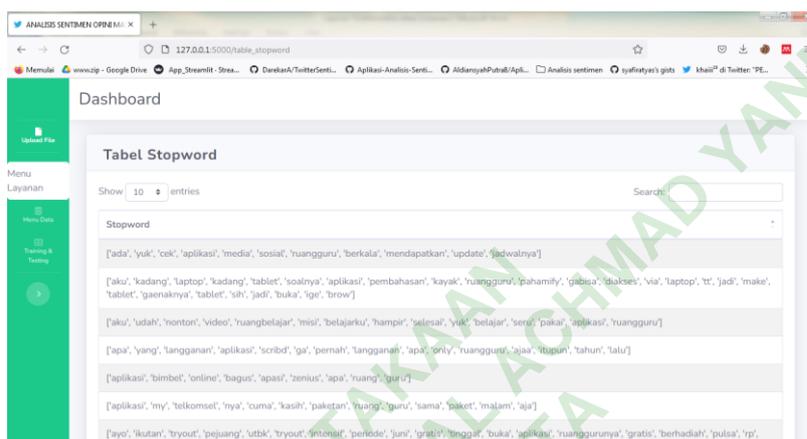


Tokenizing
[ada, 'banyak', 'aplikasi', 'elearning', 'agar', 'dirumahterusbelajar', 'ada', 'ruangguru', 'rumah', 'belajar', 'dan', 'banyak', 'elearning', 'lainnya', 'kamu', 'tinggal', 'mengaktifkan', 'paket', 'ruangguru', 'ilmupedia', 'dan', 'usahaku', 'agar', 'dirumahterusmaju', 'dengan', 'lebihmaksimal', 'cek', 'infonya', 'di']
[ada, 'pahamify', 'ruangguru', 'olle', 'beberapa', 'aplikasi', 'buat', 'sat', 'kbbi', 'zoom', 'math', 'riddles', 'canda', 'mi', 'ita', 'aplikasi', 'buat', 'utbk', 'btsw', 'adorable', 'home', 'solitaire', 'cody', 'cross', 'sims', 'freeplay']
[aku, 'kadang', 'laptop', 'kadang', 'tablet', 'soalnya', 'aplikasi', 'dan', 'pembahasan', 'kayak', 'ruangguru', 'dan', 'pahamify', 'gabisa', 'diakses', 'via', 'laptop', 'tt', 'jadi', 'ya', 'make', 'tablet', 'tapi', 'gaenaknya', 'tablet', 'sih', 'jadi', 'buka', 'ige', 'brow']
[aku, 'udah', 'nonton', 'video', 'di', 'ruangbelajar', 'dan', 'misi', 'belajarku', 'hampir', 'selesai', 'yuk', 'belajar', 'seru', 'pakai', 'aplikasi', 'ruangguru']
[apa, 'ada', 'yang', 'langganan', 'di', 'aplikasi', 'scribd', 'ga', 'pernah', 'langganan', 'apa', 'only', 'ruangguru', 'ajaa', 'stupun', 'tahun', 'lalu']
[aplikasi, 'bimbel', 'online', 'yang', 'bagus', 'apasi', 'zenius', 'apa', 'ruang', 'guru']

Gambar 4.9 Tabel *Tokenizing*

#### 4.5.5 Fitur Tabel *Stopword*

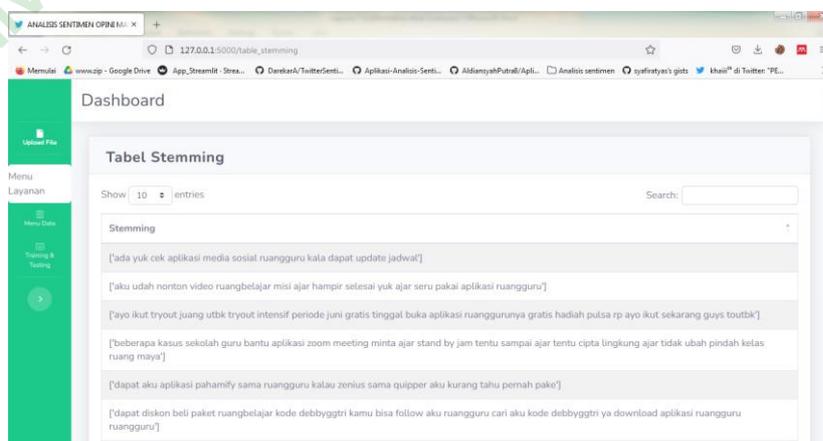
Pada tampilan ini berfungsi melakukan penghapusan kata yang tidak penting pada sebuah teks yang terdapat pada *library sastra* yang dimana fungsi *library* ini untuk bahasa Indonesia, Berikut hasil dari proses *stopword* dapat dilihat pada Gambar 4.10



Gambar 4.10 Tabel *Stopword*

#### 4.5.6 Fitur Tabel *Stemming*

Tampilan ini adalah tahap terakhir pada proses *pre-processing* yaitu *stemming* yang merupakan proses pemetaan dan penguraian bentuk kata ke bentuk kata dasarnya, yang mengacu pada data *stopword*, Berikut hasil pada data tabel *stemming* dapat dilihat pada Gambar 4.11



Gambar 4.11 Tabel *Stemming*

#### 4.5.7 Fitur Tabel *Training*

Tampilan ini menjelaskan bagaimana mengetahui hasil nilai *accuracy* dan *f-measure* secara otomatis oleh sistem dari data pelabelan manual yang dimana data pelabelan manual tersebut didapatkan dari hasil *pre-processing* dan disamping itu, fitur ini juga terdapat fungsi untuk melakukan upload file excel dengan format XLSX yang sebatas untuk memasukan file ke dalam sebuah folder. Berikut hasil dari perhitungan *accuracy* dan *f-measure* pada sistem terdapat pada Gambar 4.12

Username	Tweets	Label	Kelas
1	bab quipper latihan soal lulus negeri soal proses belajar mengajar soal pengetahuan dan pemahaman umum inggris makin dikit aja bro belajar allah dt cuma inget	P	1
2	kalimat paling teta ingat lang keluar ruang guru harap harap cemas tutor beri pak agus untuk takut kalau kalau	N	0

Gambar 4.12 Tabel *Training*

#### 4.5.8 Fitur *Predict Testing*

Pada tampilan ini menjelaskan pembuatan model *.pickle support vector machine* pada sistem dimana data yang digunakan dalam tahap ini adalah data yang berjumlah 200 masing-masing 100 positif dan 100 negatif maka akan didapatkan hasil kelas (*predicted*) dari mesin tersebut. Berikut hasil sentimen dari mesin yang telah dibuat model *pickle* terdapat pada Gambar 4.13

ANALISIS SENTIMEN OPINI MA... | Dashboard

Upload File

Menu Layanan

Menu Data

Training & Testing

Tabel Predicted Testing

Show 10 entries

Search:

Tweets	Kelas(Predicted)
aaaa aku gayakin aku ngajarin aku cuma anak suka baca aja penting kalo pengetahuan dan pemahaman umum aku paham baca terus aku pakai nalar sendiri mungkin coba quipper semangat huhu moga ngerjain pas ujian tertulis berbasis komputer	P
aamiin jelek hyung	P
adek gua minta ajarin kimia gua suka banget kimia dulu waktu sma cuma gua udah lupa cara ngerjain runtut nya alias ingetnya pakai cara cepet hitung runtut akhir merecall ingat dengan cara belajar dulu dikit ngajarin adek haha	N
akhir gw mutusin kek juni mau rehat dulu tidak mau terlalu maksain target pd nyata emg gw baru bs belajar julu karena gw belum duit buat beli akun quipper pahamify	N
akhir tidak sendiri truss punya masih kyy gt ak kalau mau tryout jadi tidak enak baca huhu	N

Gambar 4.13 Predict Testing

#### 4.5.9 Fitur Testing

Pada tampilan ini berfungsi untuk mengetahui nilai akurasi dari *testing* yang dimana data hasil pelabelan manual(actual) dan kelas(predicted) digabungkan menjadi satu file untuk menentukan hasil dari TP, TN, FP dan FN. Berikut hasil dari perhitungan *accuracy*, *precision*, *recall* dan *f-measure* terdapat pada Gambar 4.14

ANALISIS SENTIMEN OPINI MA... | Dashboard

Upload File

Menu Layanan

Menu Data

Training & Testing

ACCURACY 0.7375

PRECISION 0.725

RECALL 0.7435897435897436

F-MEASURE(F1 SCORE) 0.7341772151898733

Testing

Show 10 entries

Search:

Tweets	Kelas(Actual)	Kelas(Predicted)	Keterangan
aaaa aku gayakin aku ngajarin aku cuma anak suka baca aja penting kalo pengetahuan dan pemahaman umum aku paham baca terus aku pakai nalar sendiri mungkin coba quipper semangat huhu moga ngerjain pas ujian tertulis berbasis komputer	1	1	TP
aamiin jelek hyung	0	0	TN
adek gua minta ajarin kimia gua suka banget kimia dulu waktu sma cuma gua udah lupa cara ngerjain runtut nya alias ingetnya pakai cara cepet hitung runtut akhir merecall ingat dengan cara belajar dulu dikit ngajarin adek haha	1	0	FP
akhir gw mutusin kek juni mau rehat dulu tidak mau terlalu maksain target pd nyata emg	0	0	TN

Gambar 4.14 Fitur Testing