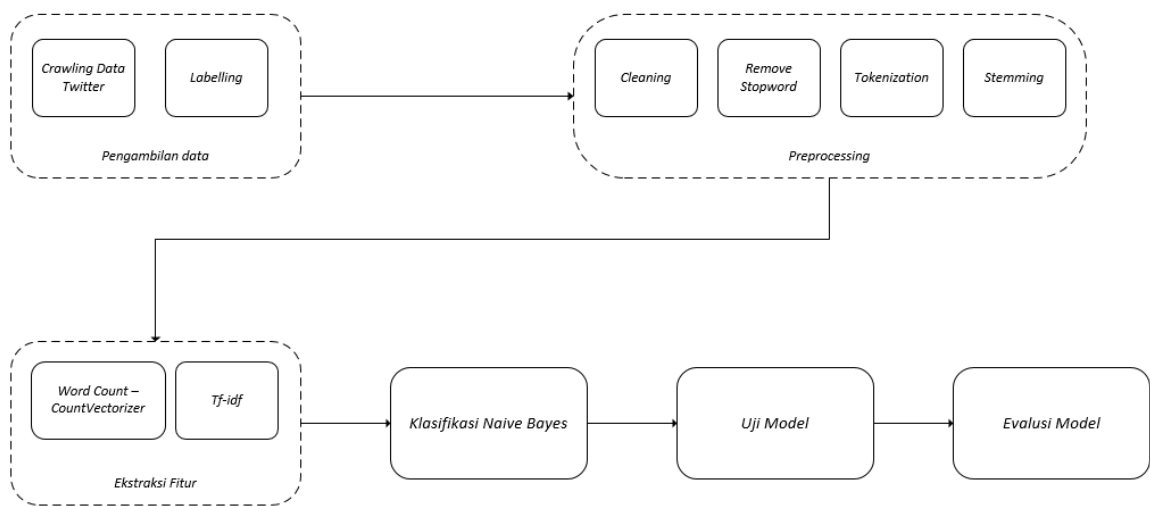


BAB III METODOLOGI PENELITIAN

3.1 Alur Pengerjaan Tugas Akhir

Perancangan alur pengerjaan tugas akhir adalah gambaran umum terkait alur dari penelitian yang akan dilakukan dalam pengerjaan tugas akhir dari awal hingga akhir. Alur kerja dari pengerjaan tugas akhir penelitian dapat dilihat pada Gambar 3.1 berikut:



Gambar 3.1 Alur Pengerjaan Tugas Akhir

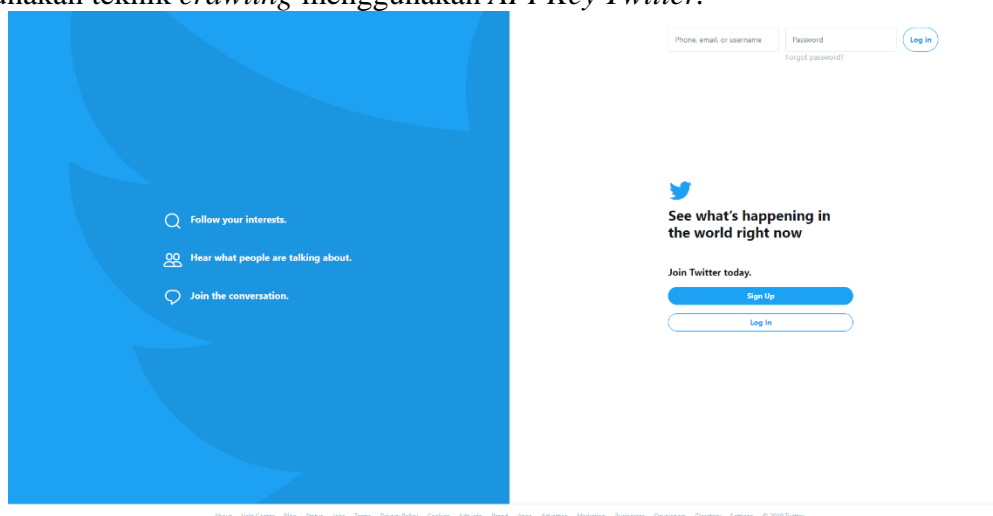
Alur pertama dalam pengerjaan tugas akhir penelitian adalah mendapatkan data dari postingan atau *tweet* pengguna Twitter dengan menggunakan teknik *crawling* data kemudian setelah semua data dikumpulkan selanjutnya tahap proses *labelling* data untuk menentukan sentimen terhadap postingan pengguna Twitter yang didapatkan. Langkah kedua, dilakukan proses *preprocessing* berguna sebagai menyeleksi data serta mengubahnya menjadi data yang lebih terstruktur. Pada proses *preprocessing* terdapat 4 tahapan yang dilakukan, yaitu *Cleaning*, *Remove Stopword*, *Tokenization* dan *Stemming*. Pada tahapan *Cleaning* berguna sebagai membersihkan kata-kata yang tidak diperlukan guna mengurangi *noise* seperti *html*, *link*, dan *script*. Selain kata-kata yang tidak perlu dihilangkan pada tahap ini juga menghilangkan tanda baca seperti titik(.), koma(,) dan juga tanda baca yang lainnya. Selain menghilangkan kata-kata dan tanda baca, pada tahap *Cleaning* juga mengubah kata menjadi

lower-case (huruf kecil) semua. Tahapan kedua adalah tahap *Remove Stopword* dalam tahap ini kata-kata yang kurang bermakna atau tidak mempunyai arti akan dilakukan penghapusan, seperti kata: saya, dan, atau. Selanjutnya masuk pada tahapan ke tiga yaitu tahap *Tokenization* digunakan sebagai indentifikasi kata-kata yang terdapat di dalam teks menjadi beberapa urutan yang terpotong oleh spasi atau juga dengan karakter spesial. Tahapan terakhir pada proses *preprocessing* adalah tahap *Stemming*, pada tahapan ini mengubah kata yang berimbuhan kembali ke kata bentuk aslinya. Langkah ketiga adalah proses ekstraksi fitur dalam proses ini dilakukan pembuatan fitur sebagai mempermudah bekerjanya proses *learning Naïve Bayes Classifier*. Langkah keempat adalah proses pengklasifikasian data menggunakan metode *Naïve Bayes Classifier* proses ini data akan diklasifikasi berdasarkan sentimen yang terdapat dalam dokumen. Setelah proses klasifikasi akan menghasilkan model yang nantinya akan dipergunakan sebagai menunjukkan ketepatan hasil pengklasifikasi. Langkah kelima adalah uji model sebagai pengukuran nilai performa pengklasifikasian yang telah dikerjakan. Langkah terakhir, setelah uji model selesai maka evaluasi model dengan cara melihat tingkat akurasi metode menggunakan *confusion matrix* dan tabel akurasi serta presisi pada setiap model.

3.2 Uraian Metodologi

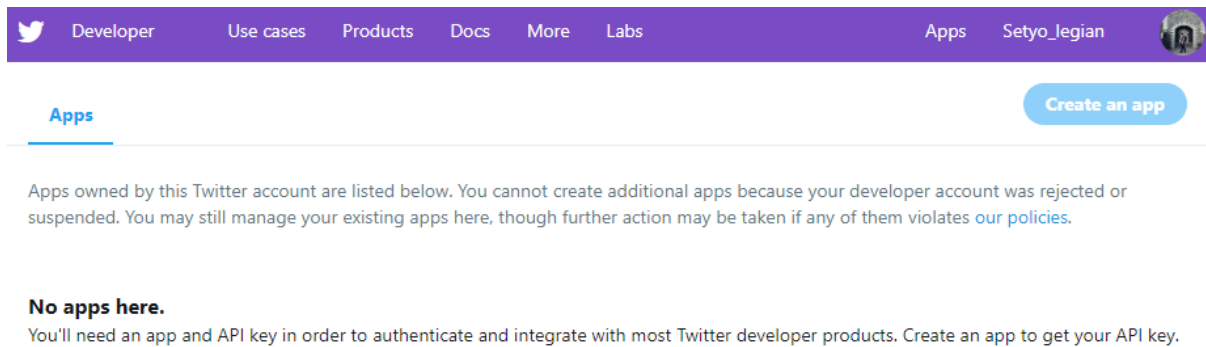
3.2.1 Pengambilan Data

Data yang digunakan dalam penelitian ini adalah data postingan pengguna Twitter yang terdapat pada situs *Twitter.com*. Data yang dikumpulkan berupa data teks yang diambil menggunakan teknik *crawling* menggunakan *API Key Twitter*.



Gambar 3.2 Halaman Awal Twitter

Tahap awal untuk melakukan proses pengambilan data dari Twitter, peneliti harus memiliki *key number* dan *secret number* dari *API Key Twitter*. Mendapatkan *API Key Twitter* peneliti harus mendaftarkan atau melakukan pengajuan terhadap pihak *developer* atau pengembang Twitter untuk mendapatkan *API Key Twitter* tersebut seperti Gambar 3.3 di bawah.



Gambar 3.3 Halaman *Developer* Twitter

Setelah mengakses website <https://developer.twitter.com/en/apps> nantinya diminta untuk mengisi form pengajuan terkait permintaan *API Key Twitter* yang berisi tentang kegunaan *API Key Twitter* yang peneliti *request*. Seperti Gambar 3.4 di bawah contoh halaman pada saat pengguna *Request API Key Twitter*.

Gambar 3.4 Halaman *Request API Key Twitter*

Setelah *API Key Twitter* didapatkan maka peneliti sudah dapat melakukan proses *crawling* data.

| Waktu | Tweet | label |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| 05/05/2019 09:56 | b'RT @lahan_poker: Andre Taulany Minta Maaf Pasca Hina Nabi, Proses Hukum Tetap Jalan! https://t.co/eq7xrexua5 #AndreTau | |
| 05/05/2019 09:50 | b'Jaman sekarang, apa aja di politikin #SaveAndreTaulany #AndreTaulanyHinaRasulullah' | |
| 05/05/2019 08:16 | b'RT @RiswandiDito: Dari sini kita bisa ambil pelajaran, Bercanda boleh tpi jangan berlebihan dan jangan bawa-bawa itu kedalar | |
| 05/05/2019 07:51 | b'Andre Taulany Minta Maaf Pasca Hina Nabi, Proses Hukum Tetap Jalan! https://t.co/eq7xrexua5 https://t.co/U3D | |
| 05/05/2019 07:49 | b'#andretaulany\n#AndreTaulanyMakinSongong\n#AndreTaulanyKufurNikmat\n#AndreTaulanyHinaRasulullah\nCoba cek ini:\n | |
| 05/05/2019 07:20 | b'@PartaiSocmed Akun abal anal\n#AndreTaulanyHinaRasulullah' | |
| 05/05/2019 07:18 | b'@hudlaha @PartaiSocmed Tangkap si pelawak cebong \n#AndreTaulanyKufurNikmat\n#AndreTaulanyHinaRasulullah' | |
| 05/05/2019 07:14 | b"@umardhan @jokowi Kelakuan'a sama...apakah nasibnya akan sama??? we'll see\n#AndreTaulanyKufurNikmat\Xe2\x80\xa6 | |
| 05/05/2019 07:10 | b'RT @RiswandiDito: Dari sini kita bisa ambil pelajaran, Bercanda boleh tpi jangan berlebihan dan jangan bawa-bawa itu kedalar | |
| 05/05/2019 07:05 | b'@aburasyid13 Proses hukum tetap harus berjalan\n#AndreTaulanyHinaRasulullah\n#BoikotNetTV' | |
| 05/05/2019 06:32 | b'RT @lahan_poker: BREAKING NEWS: Andre Taulany Minta Maaf Usai Hina Nabi, Pelapor: Proses Hukum Tetap Jalan! https://t.c | |
| 05/05/2019 06:21 | b'RT @RiswandiDito: Dari sini kita bisa ambil pelajaran, Bercanda boleh tpi jangan berlebihan dan jangan bawa-bawa itu kedalar | |
| 05/05/2019 06:19 | b'Saya setuju #PenjarakanAndreTaulany \n#AndreTaulanyHinaRasulullah https://t.co/4GjagcDvfV ' | |
| 05/05/2019 06:08 | b'@iswan214 @prabowo siap2 nnti paspampers kewalahan\Xf0\x9f\xa4\xa3\xF0\x9f\xa4\xa3 pak prabowo juga tak mau Ada jar | |

Gambar 3.5 Hasil *Crawling* Data

Hasil dari proses *crawling* data dalam file excel pada Gambar 3.5 di atas yang nantinya akan proses *labelling* data untuk menentukan klasifikasi pendapat atau pandangan dari hasil tweet yang telah di *crawling* tadi. Pada proses *labelling* ini dibedakan menjadi 2 kelas. Yaitu *class positif* dan *class negative*. Contoh dari proses *labelling* data seperti Tabel 3.1 di bawah.

Tabel 3.1 Contoh Data Hasil *Labelling*

| Tweet | Clean Text | Label |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|-------|
| b'RT @lahan_poker: Andre Taulany Minta Maaf Pasca Hina Nabi, Proses Hukum Tetap Jalan! https://t.co/eq7xrexua5 #AndreTaulanyKufurNikmat #Andr\Xe2\x80\xa6' | andre taulany minta maaf pasca hina nabi proses hukum tetap jalan andretaulanykufur nikmat andr | 0 |
| b'Kontol kau Andre ...anjing...biadab....taik babi ko memang udah bauk haram lagi #AndreTaulanyHinaRasulullah' | kontol kau andre anjing biadab taik babi ko udah bauk haram andretaulanyhinara sulullah | 1 |
| b'@detikcom Cebong bangsat\n#AndreTaulanyHinaRasulullah' | cebong bangsat andretaulanyhinara sulullah | 1 |

| | | |
|---------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|---|
| b'baperan. habib kok baperan. harusnya lbh kenal agama lbh bijaksana #AndreTaulanyHinaRasulullah https://t.co/t9GZ7T5v39' | baperan habib baperan kenal agama bijaksana andretaulanyhinara sulullah | 1 |
| b'@aburasyid13 Proses hukum tetap harus berjalan\n#AndreTaulanyHinaRasulullah\n#BoikotNetTV' | proses hukum tetap jalan andretaulanyhinara sulullah boikotnettv | 0 |

Dalam kasus ini *class positif* label 1 menyatakan bahwa tweet tersebut adalah kata-kata yang mengandung unsur *hatespeech* atau ujaran kebencian, sedangkan *class negative* berlabel kan 0 adalah kata-kata yang netral atau tidak mengandung unsur *hatespeech*.

3.2.2 Tahapan Proses *Preprocessing*

Preprocessing adalah tahapan proses untuk membersihkan data dari kata-kata atau tweet yang tidak di perlukan serta kata-kata yang tidak memiliki makna. Proses ini dilakukan sesuai dengan isi data dari proses pengambilan data atau *crawling* data Twitter. Adapun proses beberapa langkah dari proses *preprocessing* memiliki urutan sebagai berikut :

a. *Cleaning*

Cleaning adalah proses penghapusan simbol, tanda baca, huruf kapital dan bilangan angka yang sering muncul pada tweet pengguna Twitter sehingga data tersebut menjadi data yang tidak efektif dan tidak memiliki arti. Proses ini dijalankan menggunakan program, sehingga *cleaning* ini berjalan secara otomatis sebelum menyimpan hasil *decode* dalam bentuk file excel. Contoh penerapan proses *cleaning* dapat dilihat seperti Tabel 3.2 di bawah.

Tabel 3.2 Contoh Hasil *Cleaning*

| <i>Tweet Sebelum Cleaning</i> | <i>Tweet Sesudah Cleaning</i> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| b'RT @lahan_poker: Andre Taulany Minta Maaf Pasca Hina Nabi, Proses Hukum Tetap Jalan! https://t.co/eq7xrexua5 #AndreTaulanyKufurNikmat #Andr\xe2\x80\xa6' | andre taulany minta maaf pasca hina nabi proses hukum tetap jalan andretaulanykufurnikmat andra |
| b'Kontol kau Andre ...anjing...biadab....taik babi ko memangnya udah bauk haram lagi #AndreTaulanyHinaRasulullah' | kontol kau andre anjing biadab taik babi ko memangnya udah bauk |

| | |
|------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| | haram lagi andretaulanyhinarasulullah |
| b'@detikcom Cebong bangsat\n#AndreTaulanyHinaRasulullah' | cebong bangsat andretaulanyhinarasulullah |
| b'baperan. habib kok baperan. harusnya lbh kenal agama lbh bijaksana #AndreTaulanyHinaRasulullah https://t.co/t9GZ7T5v39' | baperan habib baperan kenal agama bijaksana andretaulanyhinarasulullah |
| b'@aburasyid13 Prosesnya hukum tetap harus berjalan\n#AndreTaulanyHinaRasulullah\n#Bo ikotNetTV' | prosesnya hukum tetap jalan andretaulanyhinarasulullah ah boikotnettv |

b. *Remove Stopword*

Remove Stopword adalah proses pehapusan kata-kata yang kurang bermakna atau kata yang tidak memiliki arti seperti kata dan, atau, kamu, saya. Contoh proses penerapan pada tahap *Remove Stopword* dapat dilihat pada Tabel 3.3 di bawah.

Tabel 3.3 Contoh Data Hasil *Remove Stopword*

| Tweet Sebelum <i>Remove Stopword</i> | Tweet Sesudah <i>Remove Stopword</i> |
|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| andre taulany minta maaf pasca hina nabi proses hukum tetap jalan andretaulanykufurnikmat andra | andre taulany minta maaf pasca hina nabi proses hukum tetap jalan andretaulanykufurnikmat andra |
| kontol kau andre anjing biadab taik babi ko memangnya udah baik haram lagi andretaulanyhinarasulullah | kontol kau andre anjing biadab taik babi ko udah baik haram andretaulanyhinarasulullah |
| cebong bangsat andretaulanyhinarasulullah | cebong bangsat andretaulanyhinarasulullah |

| | |
|------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| baperan habib kok baperan harus lbh kenal agama lbh bijaksana andretaulanyhinarasulullah | baperan habib baperan kenal agama bijaksana andretaulanyhinarasulullah |
| prosesnya hukum tetap harus jalan andretaulanyhinarasulullah boikotnettv | prosesnya hukum tetap jalan andretaulanyhinarasulullah boikotnettv |

c. *Tokenization*

Tokenization adalah proses untuk memecahkan kalimat untuk menjadi beberapa bagian yang dinamakan *token*. Sebuah *token* dapat dianggap menjadi satu bentuk sebuah kata, frasa, atau suatu elemen yang berarti. Contoh proses pada tahap *tokenization* dapat dilihat pada Tabel 3.4 di bawah.

Tabel 3.4 Contoh Data Hasil *Tokenization*

| Tweet Sebelum <i>Tokenization</i> | Tweet Sesudah <i>Tokenization</i> |
|-------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| andre taulany minta maaf pasca hina nabi proses hukum tetap jalan andretaulanykufurnikmat andra | ['andre','taulany', 'minta','maaf','pasca', 'hina','nabi','proses','hukum', 'tetap','jalan', 'andretaulanykufurnikmat','andra'] |
| kontol kau andre anjing biadab taik babi ko udah bauk haram andretaulanyhinarasulullah | ['kontol','kau','andre','anjing','biadab','taik', 'babi','ko','udah','bauk','haram','andretaulanyhinarasulullah'] |
| cebong bangsat andretaulanyhinarasulullah | ['cebong','bangsat','andretaulanyhinarasulullah'] |
| baperan habib baperan kenal agama bijaksana andretaulanyhinarasulullah | ['baperan','habib','baperan','kenal','agama', 'bijaksana','andretaulanyhinarasulullah'] |
| prosesnya hukum tetap jalan andretaulanyhinarasulullah boikotnettv | ['prosesnya','hukum','tetap','jalan','andretaulanyhinarasulullah','boikotnettv'] |

d. *Stemming*

Stemming adalah proses untuk merubah kata kedalam bentuk dasar dengan menghapus kata imbuhan di depan maupun di belakang kata. Contoh proses penerapan *stemming* dapat dilihat pada Tabel 3.5 di bawah.

Tabel 3.5 Contoh Data Hasil *Stemming*

| Tweet Sebelum <i>Stemming</i> | Tweet Sesudah <i>Stemming</i> |
|-----------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| ['andre', 'taulany', 'minta', 'maaf', 'pasca', 'hina', 'nabi', 'proses', 'hukum', 'tetap', 'jalan', 'andretaulanykufurnikmat', 'andra'] | andre taulany minta maaf pasca hina nabi proses hukum tetap jalan andretaulanykufurnikmat andra |
| ['kontol', 'kau', 'andre', 'anjing', 'biadab', 'taik', 'babi', 'ko', 'udah', 'bauk', 'haram', 'andretaulanyhinarasulullah'] | kontol kau andre anjing biadab taik babi ko udah bauk haram andretaulanyhinarasulullah |
| ['cebong', 'bangsat', 'andretaulanyhinarasulullah'] | cebong bangsat andretaulanyhinarasulullah |
| ['baperan', 'habib', 'baperan', 'kenal', 'agama', 'bijaksana', 'andretaulanyhinarasulullah'] | baperan habib baperan kenal agama bijaksana andretaulanyhinarasulullah |
| ['prosesnya', 'hukum', 'tetap', 'jalan', 'andretaulanyhinarasulullah', 'boikotnettv'] | proses hukum tetap jalan andretaulanyhinarasulullah boikotnettv |

3.2.3 Ekstraksi Fitur

Dalam kasus ini setelah semua tweet dikumpulkan dari hasil *crawling* dan hasil data tersebut telah melewati proses *preprocessing*, proses selanjutnya adalah membuat fitur yang berguna untuk mempermudah proses pengklasifikasian data tweet tersebut biasa proses ini dibuat dengan proses ekstraksi fitur. Pada proses ekstraksi fitur ini terdapat dua proses yang dilakukan yaitu proses pembuatan *word vector* dalam proses ini sistem akan mengubah suatu teks menjadi representasi *vector* dan proses pembobotan kata menggunakan $Tf-idf(term$

frequency-inverse document frequency). *Word vector* ini sendiri dapat diartikan dalam Bahasa Indonesia sebagai vektor kata proses pembuatan kalimat yang sudah ada menjadi sekumpulan *array* menjadi suatu matriks, setiap baris matriks tersebut mewakili dari baris dokumen, sedangkan setiap kolom matriks akan mewakili seluruh kata yang terdapat di dalam teks dari suatu data tweet. Setelah semua kata diproses dan berubah menjadi vektor kata, selanjutnya adalah proses pemberian bobot dari setiap kata pada setiap kalimat atau dokumen menggunakan metode *Uni gram* dan *Tf-idf* (*term frequency-inverse document frequency*) menggunakan rumus yang sudah dijelaskan pada bab sebelumnya. Jika proses pembobotan selesai maka *dataset* dapat digunakan dalam *training* menggunakan perhitungan *Naïve Bayes Classifier*.

3.2.4 Klasifikasi

Dalam kasus ini peneliti menggunakan Metode *Naïve Bayes Classifier* sebagai metode pengklasifikasian data Twitter untuk mendapatkan sentimen analisis. Untuk melakukan proses pengklasifikasian ini diperlukan hasil dari data yang sudah diolah dari proses sebelumnya yaitu hasil dari proses *preprocessing* dan hasil dari pembobotan kata dengan *Tf-idf*. Setelah data berhasil di- *training* maka akan dilakukan proses pengujian menggunakan *datatest* sebagai pengujian hasil ketepatan klasifikasi yang dilakukan.

Selain menggunakan metode *Naïve Bayes Classifier* dalam penelitian ini juga menggunakan metode *logistic regression model*. Metode *logistic regression model* ini digunakan sebagai pembandingan terhadap metode *Naïve Bayes Classifier*. *Logistic regression model* adalah suatu metode regresi (metode melihat pengaruh antara dua atau lebih variabel), *Logistic regression* ini menghubungkan antara satu atau beberapa variabel bebas (variabel *independen*) dengan variabel *dependen* yang kategori variabel ini 0 dan 1.

3.2.5 Uji Model

Proses uji model dapat dilakukan ketika proses dari *training* data telah selesai dikerjakan. Pengujian model ini dilakukan untuk mengetahui bagaimana kinerja model. Jumlah data yang dijadikan sebagai bahan pengujian diambil dari data *training* sebesar 33% sama dengan 0,33. Pengambilan data ini dilakukan secara *random* dengan menggunakan bantuan *library* dari *Python*. Setelah semua proses selesai maka sistem akan menampilkan besar akurasi dari model yang dikerjakan.

3.2.6 Evaluasi Model

Evaluasi model ini berguna untuk mengetahui tingkat keakurasian dari kinerja model. Untuk mendapatkan tingkat keakurasian dari kinerja model dalam kasus ini digunakan metode *confusion Matrix* dan tabel akurasi serta melihat presisi untuk setiap model. Setelah *datatest* dilakukan pengujian terhadap data *training*, maka akan menghasilkan beberapa kelas dari *datatest*, biasa disebut prediksi kelas. Kemudian prediksi kelas tadi yang sebenarnya berasal dari *datatest* sebelumnya tadi disembunyikan, sehingga dapat ditampilkan dan dihitung nilai dari *accuracy*, *precision*, *recall*, dan *f1-score*.