

BAB III METODE PENELITIAN

3.1 Metode Pengembangan

Penelitian tugas akhir ini menggunakan metodologi *Waterfall* sebagai metode pengembangan. Proses *waterfall* dimulai pada tahap pengumpulan data, analisis kebutuhan, perancangan, eksekusi dan pengujian sistem.

a. Pengambilan data

Pada langkah ini dilakukan dengan mengambil *postingan tweet* dari twitter.com. Teknik yang di ambil untuk mengumpulkan data teks menggunakan teknik *Crawling* dengan memanfaatkan Twitter *developer* yaitu API Key Twitter. Peneliti harus mendaftar terlebih dahulu pada pihak Twitter developer di app.twitter.com untuk mendapatkan *key number* dan *secret number*. Pada saat mendaftar pihak Twitter developer akan mengajukan pertanyaan tentang kegunaan untuk apa API Key digunakan sesuai dengan kebutuhan peneliti.

b. Analisis kebutuhan

Pada langkah ini, menganalisis dan menentukan apa yang dibutuhkan sistem. Hal ini yang dibutuhkan menganalisis sentimen terhadap *tweet* yang sudah berhasil di ambil. Setelah data berhasil terkumpul peneliti akan mempersiapkan *tools* seperti *anaconda* dan memakai *python* sebagai Bahasa pemrograman, serta *library* pendukung untuk melakukan analisis sentimen dengan menggunakan metode *naïve bayes classifier* dan *library* untuk melakukan visualisasi *word cloud*

c. Analisis perancangan

Pada langkah ini, peneliti membuat gambaran aliran analitis melalui diagram alir setelah melakukan analisis kebutuhan.

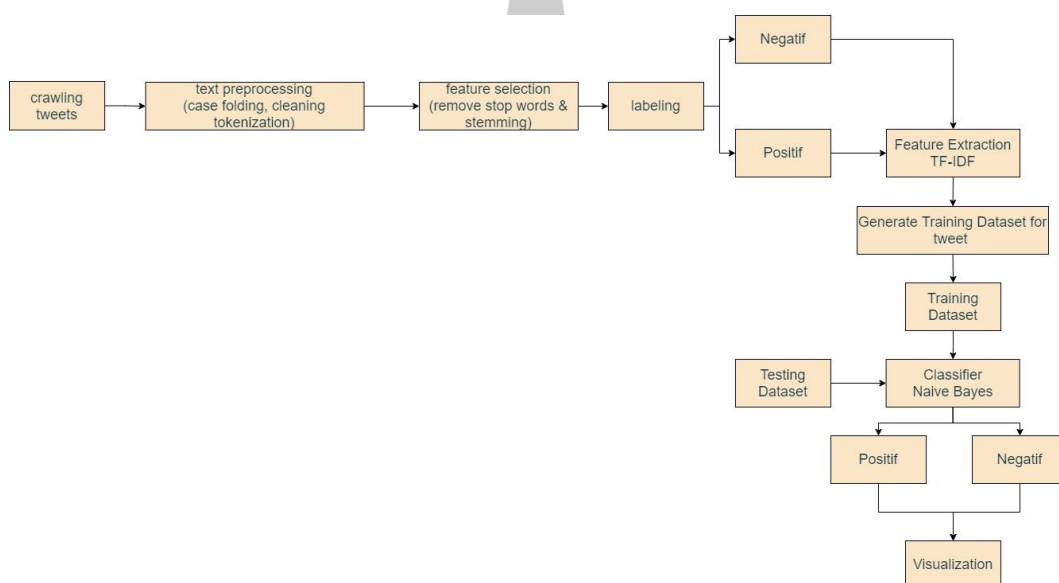
d. Implementasi

Pada langkah ini peneliti akan mengimplementasi analisis kebutuhan dan analisis perancangan yang sudah dibuat dan menganalisis sesuai dengan diagram alir

e. Pengujian sistem

Pada tahap ini, setelah mengimplementasikan peneliti akan menguji untuk meminimalisir terjadinya eror. Pengujian sistem dapat membantu mengidentifikasi kesalahan dan kekurangan dari analisis sentimen.

3.2 Flow Chart



Gambar 3. 1 Flow Chart

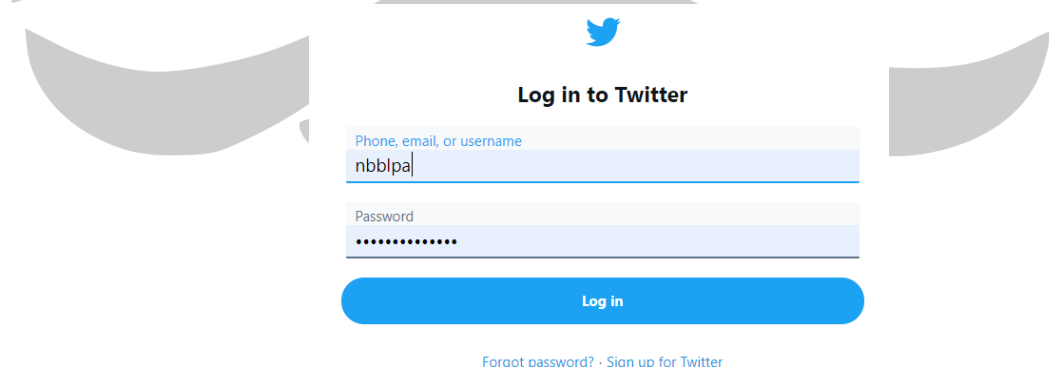
Langkah awal dalam tugas akhir ini adalah pengumpulan data *tweet* dari pengguna Twitter dengan metode *crawling* seperti informasi yang didapat pada Gambar 3.1. Lalu masuk tahap selanjutnya yaitu tahap *preprocessing*, seleksi fitur dan labeling. Pada proses *preprocessing* melalui 3 proses yaitu *case folding*, *cleansing* dan tokenisasi. Lalu masuk ke tahap seleksi fitur menggunakan proses *remove stop words* dan *stemming*. Dan masuk ke tahap labeling, pada tahap labeling di bagi menjadi 2 kelas yaitu negatif dan positif dengan pelabelan otomatis menggunakan kamus lexicon, tahap labeling ini dilakukan pada dataset setelah dilakukan proses

cleansing dikarenakan pada proses hasil dari stop words terdapat tanda baca “[]” dan “ ‘ ‘ ” yang menyulitkan saat proses labeling. Di tahap *cleaning*, *tweet* akan dibersihkan seperti menghilangkan *link*, *mention*, *hashtag*, lalu mengubah huruf menjadi *lower case* semua. Selanjutnya *tokenization*, dilakukan untuk memotong-motong teks. Lalu ada normalisasi, yang dilakukan seperti untuk mengubah kata tidak baku menjadi kata baku. Dan di tahap *feature selection* terdapat *remove stopwords*, dalam tahap *remove stopwords* akan melakukan penghapusan kata tidak bermakna seperti kata: yang, tau, dan. Lalu yang terakhir ada *stemming*, pada tahap ini mengubah setiap kata ke bentuk aslinya. Langkah selanjutnya ada *feature extraction* dimana pada langkah ini dilakukan proses pembobotan kata menggunakan TF-IDF. Setelah melakukan *feature extraction* selanjutnya melakukan *generate dataset* dengan membagi 2 dataset menjadi data *training* dan data *testing* dengan perbandingan rasio 80:20. Setelah itu masuk ke *classifier* dan didapatkan hasil akurasi serta *precision*, *recall* dan *f-measure* atau *F1- Score* untuk masing-masing kelas. Setelah melakukan klasifikasi lalu visualisasi menggunakan *word cloud*

3.3 Metode Pengumpulan Dataset

3.3.1 Pengambilan Dataset

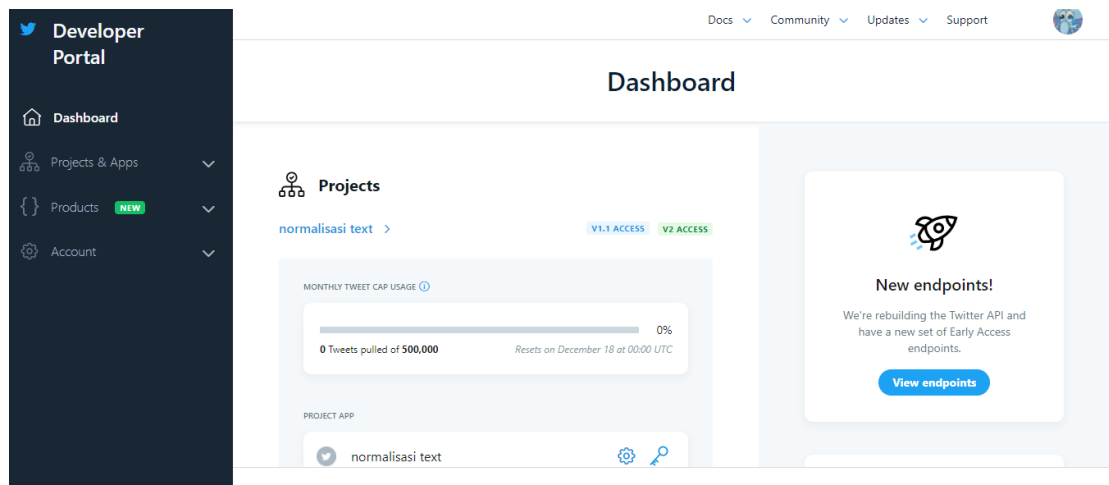
Data yang digunakan dalam penelitian ini adalah data tweet yang di *posting* oleh pengguna Twitter. informasi yang dikumpulkan menggunakan metode *crawling* dan API Twitter.



Gambar 3. 2 Halaman Log in Twitter

Selama pengambilan data, peneliti perlu mendaftar terlebih dahulu ke pengembang Twitter untuk mengakses akun dengan nomor kunci dan nomor rahasia di API Key Twitter agar akun dapat diambil.

Untuk melakukan pendaftaran peneliti harus mengakses *website* Twitter developer di <https://developer.twitter.com/apps> yang dimana pihak Twitter developer akan mengajukan beberapa pertanyaan yang berisi kegunaan API yang akan di *request* oleh peneliti.



Gambar 3. 3 Halaman *Dashboard Developer*

Jika pengembang Twitter menyetujui, API Key Twitter akan diberikan ke akun yang diinginkan. Setelah itu, peneliti dapat melakukan prosedur *tweet crawling*.

1	tanggal	tweet
2	12/1/2020 2:40	b'RT @TelkomIndonesia: Sobat, selagi menunggu pengujian vaksin yang sedang diupayakan pemerintah, masyarakat wajib mengikuti protokol kesehatan
3	12/1/2020 2:40	b'RT @RakyatMerdeka99: Top News Koran Rakyat Merdeka\n\nKasus Positif Melonjak\n\nMuhadjir: Vaksin Bukan Senjata Pamungkas Basmi Virus Corona
4	12/1/2020 2:40	b'Hadirnya vaksin Covid-19 masih menjadi perbincangan yang hangat di tengah-tengah masyarakat. #Vaksin #Covid19
5	12/1/2020 2:37	b'RT @AchuNadia: Terkini\n\nModerna umum vaksin ciptaan mereka adalah 100% berkesan terhadap jangkitan COVID-19.
6	12/1/2020 2:35	b'Selamat datang bulan Desember, doaku masih sama. Semoga vaksin Covid-19 cepat dilakukan #DecemberWish
7	12/1/2020 2:35	b'RT @bernamadotcom: Penerapan vaksin COVID-19 faktor utama penurunan mendadak harga emas
8	12/1/2020 2:30	b'30 Menit lagi di Selamat Siang Indonesia, akan membahas seputar vaksin Covid-19 yang sedang dalam uji coba. Membaha
9	12/1/2020 2:30	b'Apakah vaksin Covid-19 aman atau tetap beresiko? Bahas langsung bersama Dr. dr. Hariadi Wibisono MPH. Ketua Umum Per
10	12/1/2020 2:29	b'hmetromy: #kualalumpur Berita pengumuman penemuan vaksin Covid-19 yang memberi kelegaan kepada semua orang di dunia termasuk di Malaysi
11	12/1/2020 2:28	b'Penemuan vaksin COVID-19 faktor utama penurunan mendadak harga emas
12	12/1/2020 2:28	b'RT @AchuNadia: Terkini\n\nModerna umum vaksin ciptaan mereka adalah 100% berkesan terhadap jangkitan COVID-19.
13	12/1/2020 2:27	b'Penemuan vaksin Covid-19 punca harga emas turun mendadak
14	12/1/2020 2:24	b'RT @hmetromy: #kualalumpur Berita pengumuman penemuan vaksin Covid-19 yang memberi kelegaan kepada semua orang di dunia termasuk di Malaysi
15	12/1/2020 2:23	b'Penemuan Vaksin COVID-19 Dilapor Jadi Faktor Utama Harga Emas Turun Mendadak
16	12/1/2020 2:23	b'@Syarman59 Jelas ya @LisaAmartatara3\n\nBahwa Dirut RS sendiri yang mengakui bahwa ada "pelanggaran" dari proses test
17	12/1/2020 2:22	b'RT @pikiran_rakyat: Kim Jong-un Diam-Diam Suntik Vaksin dari Tiongkok Karena Takut Mati Terinfeksi Covid-19
18	12/1/2020 2:22	b'taun depan vaksinasi covid 19 uda mulai dilaksanakan nih? karena lagi nunggu kerjaan datang, kalo buat thread tenta
19	12/1/2020 2:21	b'hmetromy: #kualalumpur Berita pengumuman penemuan vaksin Covid-19 yang memberi kelegaan kepada semua orang di dunia termasuk di Malaysi
20	12/1/2020 2:20	b'ikuti Rakor Pengadaan Vaksin, Pjs Bupati Gowa Harap Covid-19 Segera Berakhir -
21	12/1/2020 2:20	b'Kim Jong-un Diam-Diam Suntik Vaksin dari Tiongkok Karena Takut Mati Terinfeksi Covid-19
22	12/1/2020 2:19	b'DPR RI dan Masyarakat Dukong Program Vaksin Guna Memutus mata Rantai Penularan Covid-19.\n\nVaksin #Vaccine
23	12/1/2020 2:19	b'Pemerintah Indonesia seial awal pandemi sudah melibatkan Maieis Ulama Indonesia (MUI), termasuk dalam hal rencana

Gambar 3. 4 Hasil *Crawling*

Data *crawling* yang berhasil diselesaikan dan memasuki tahap pelabelan ditunjukkan pada Gambar 3.4. Metode ini dipisahkan menjadi dua kelas dalam proses pelabelan ini. Kelas positif dan negatif khususnya.

3.3.2 Tahap Proses *Preprocessing*

Preprocessing adalah tahap awal di mana kata-kata atau tweet dibersihkan tanpa struktur. Tahap *crawling* Twitter didapatkan pada langkah ini. Adapun proses dari tahapan *preprocessing* sebagai berikut:

a. *Case folding*

Case Folding adalah prosedur persiapan teks untuk mengubah huruf besar menjadi huruf kecil (*lower case*)

b. *Cleaning*

Cleaning adalah proses penghapusan tanda baca, angka, simbol serta huruf kapital yang tidak diperlukan tujuannya untuk menghilangkan *noise* yang dapat menyebabkan proses perhitungan pada klasifikasi kurang optimal [26]. Proses *cleaning* dijalankan di *python* dengan menggunakan program yang otomatis serta hasil yang di dapatkan berbentuk file excel.

c. *Tokenization*

Tokenisasi adalah proses memecah pernyataan dan kemudian mengubahnya menjadi token. Tokenisasi juga berguna sebagai pemisah kata simbol, atau frase dari teks.

d. Normalisasi

Normalisasi adalah proses pengubahan kata tidak baku menjadi kata baku yang sesuai kaidah KBBI atau membenarkan kata yang disingkat. Proses normalisasi perlu menggunakan dataset [27] yang berisi dari kata singkatan dan kata tidak baku dan kata yang sesuai KBBI. Dataset normalisasi berguna untuk memperbaiki kata-kata yang tidak sesuai di dataset vaksin covid dengan menyamakan kata yang terdapat di dataset vaksin covid dengan dataset normalisasi sehingga kata yang di dataset vaksin covid sesuai dengan KBBI.

e. *Remove Stopword*

Remove stopwords merupakan *filtering* kata dengan cara membuat kata-kata yang tidak memiliki makna atau dianggap tidak penting pada dokumen. Contohnya seperti kata “yang”, “dari”, “dan” . kata yang dihapus oleh *stopword* dianggap memiliki informasi yang rendah

f. *Stemming*

Stemming adalah proses mengubah kata dengan menghapus awalan atau akhiran menjadi kata-kata penting atau kata dasar.

3.3.3 Analisis Sentimen

Analisis sentimen dilakukan dengan menggunakan metode *lexicon-based*. Kamus positif dan negatif pada *lexicon-based* yang digunakan pada penelitian ini yaitu kamus positif dan negatif dari beberapa dataset yang dijadikan satu data, datasetnya adalah penelitian oleh Koto & Rahmingtyas [28] yang melakukan penelitian kamus lexicon umum di Twitter dan penelitian oleh Liu, Hu, & Cheng [29] yang telah diterjemahkan ke dalam Bahasa Indonesia & telah melakukan penyesuaian Bahasa. Kamus ini sebelumnya telah di aplikasikan pada beberapa penelitian sebelumnya [30][31].

3.4 Feature Extraction

Dalam tahapan ekstraksi fitur setelah tweet berhasil di *crawling* dan telah berhasil di *preprocessing* lalu dilakukan analisis sentimen, selanjutnya adalah proses ekstraksi dimana pada proses ini data teks akan di berikan pembobotan kata menggunakan TD-IDF (*Term Frequency-Inverse Document Frequency*). Untuk perhitungan bobot, pendekatan TF-IDF menggunakan 2 konsep, yaitu berapa kali suatu kata ditemukan dalam teks, dan kebalikannya adalah berapa kali dalam dokumen tersebut. Jumlah kata dalam teks mengungkapkan betapa pentingnya dan betapa pentingnya dalam frekuensi dokumen dengan kata-kata ini. Kumpulan data dapat digunakan untuk data latih melalui metode *Naïve Bayes Classifier*.

3.5 Klasifikasi

Peneliti menggunakan metode *Naïve Bayes Classifier* sebagai metode klasifikasi. Untuk melakukan pengklasifikasian diperlukan data yang sudah di

olah sebelumnya. Data *training* yang sudah berhasil di dapatkan akan masuk ke dalam tahap pengujian untuk menguji ketetapan dari hasil klasifikasi.

3.6 Uji model

Setelah prosedur data latih selesai dilakukan, maka dilakukan proses pengujian model. Pengumpulan data dilakukan secara acak melalui pustaka *python*. Penyiapan data pelatihan dan pengujian dibagi dengan rasio 80:20. 80% untuk data *training* dan 20% untuk pengujian data

3.7 Evaluasi Model

Untuk menentukan besarnya presisi kinerja model digunakan penilaian model. Untuk mendapat nilai akurasi digunakan metode *confusion matrix*. Di *confusion matrix* selain melihat nilai akurasi terdapat juga nilai *precision*, *recall*, dan juga *f-measure*. *Confusion matrix* menghasilkan *matrix* berupa kelas *actual* dan kelas *predict*.

