

BAB 3

METODE PENELITIAN

Penelitian analisis sentimen di media sosial twitter dengan studi kasus vaksinasi covid-19 ini adalah penelitian analisis sentimen positif dan negatif pada data *Twitter*. Penelitian ini menggunakan metode *Naïve Bayes Classification*. Penggunaan metode *Naïve Bayes Classifier* (NBC) dalam penelitian analisis sentiment di media sosial twitter dengan studi kasus vaksinasi covid-19 dipilih dikarenakan pada algoritma NBC dapat melakukan proses pengolahan data diskrit dan data kuantitatif dengan menggunakan sampel yang relative sedikit dan juga perhitungan pada algoritma NBC lebih cepat (Ndruru, 2022).

Penelitian analisis sentimen di media sosial twitter dengan studi kasus vaksinasi covid-19 membutuhkan data *tweet* yang didapatkan dari *Twitter* yang berkaitan tentang anti vaksinasi, selanjutnya dilakukan pengolahan data berupa *preprocessing* untuk mendapatkan hasil yang diinginkan. Data tersebut nantinya digunakan untuk memetakan informasi atau sentimen dari *netizen* di *Twitter* mengenai opini Vaksinasi pada kelompok anti vaksin, sehingga didapatkan informasi yang sesuai mengenai proses program Vaksinasi terhadap kelompok Anti Vaksin. Penelitian ini berawal dari latar belakang permasalahan yang ada, mengolah data yang sudah didapatkan dan menentukan sentimen yang tepat sehingga informasi yang diperoleh sesuai dengan apa yang diharapkan.

Penelitian sentimen di media sosial twitter dengan studi kasus vaksinasi covid-19 ini bermula dari identifikasi permasalahan, kemudian pengolahan data yang sudah didapat dan menentukan sentimen analisis, sehingga hasil penelitian ini berupa informasi yang tepat dan sesuai.

3.1 BAHAN PENELITIAN

Dalam penelitian analisis sentimen di media sosial dengan studi kasus vaksinasi covid-19 ini, materi utama berupa data tweet dan retweet di media sosial *Twitter* yang berkaitan dengan topik program vaksinasi. penelitian analisis

sentimen di media sosial dengan studi kasus vaksinasi covid-19 ini menggunakan data tweet Twitter dengan kata kunci "Antivaksin Covid" atau dengan mengambil data berdasarkan akun yang terkait dengan berita tentang program vaksinasi seperti @KemenkesRI. Data tersebut dikumpulkan dari periode bulan Agustus 2021 hingga Desember 2021.

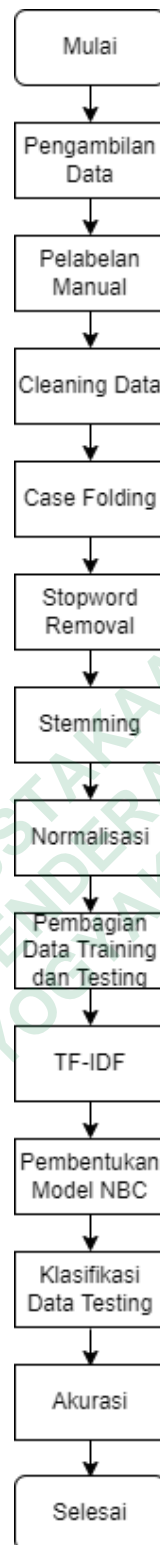
3.2 ALAT PENELITIAN

Alat yang digunakan dalam penelitian analisis sentimen di media sosial dengan studi kasus vaksin Covid-19 berupa laptop dengan spesifikasi yang cukup untuk menjalankan proses pengolahan data dan mampu terhubung ke Internet. Sistem operasi dan program aplikasi yang digunakan dalam pengembangan aplikasi ini adalah sebagai berikut:

1. OS: *Windows 10 64-bit*
2. Bahasa Pemrograman: *Python 3.10.2*
3. *Microsoft Excel 2019*
4. *Jupyter Notebook*

3.3 JALAN PENELITIAN

Dalam penelitian analisis sentimen di media sosial dengan studi kasus vaksin Covid-19 ini, tahap pertama adalah penggunaan bahasa pemrograman *Python* untuk mengambil dan memvisualisasikan data yang ditampilkan di Microsoft Office Excel menggunakan *library Python*. Gambar 1 adalah jalan penelitian :



Gambar 1 Jalan Penelitian

3.3.1 Pengambilan Data

Pada penelitian analisis sentimen di media sosial dengan studi kasus vaksin Covid-19 ini bahan utama yaitu berupa data *tweet* dan *retweet* di media sosial Twitter yang berkaitan tentang tema Program Vaksinasi. Penelitian ini menggunakan data *tweet* yang didapatkan dari Twitter dengan keyword/hastag “Antivaksin Covid” atau dengan pengambilan data berdasarkan akun-akun terkait berita mengenai program Vaksinasi seperti @KemenkesRI. Pengumpulan data dilakukan pada periode Bulan Agustus 2021- Desember 2021. Proses ini dilakukan dengan menggunakan software Anaconda dengan bantuan tools *JupyterNotebook*. Proses selanjutnya menampilkan hasil data yang sudah berhasil di ambil, data tersebut akan ditampilkan ke dalam format csv atau Microsoft Office Excel. Data *tweet* di ambil dari periode 1 Agustus 2021- Desember 2021, dengan jumlah data total 889.

3.3.2 Pelabelan Manual

Pada tahap pelabelan manual, tahap ini merupakan proses memberikan level terhadap kata pada dokumen sehingga dapat dianalisis lebih lanjut mengenai sifat yang positif atau negatif. Pada proses pelabelan manual peneliti membagi data *tweet* yang berisi tentang anti vaksin dengan data training label positif dan label negatif. Data *tweet* yang sudah dilabeli dengan jumlah data 889 *tweet* dengan masing-masing 277 *tweet* positif dan 612 *tweet* negatif dari data *training*.

3.3.3 Cleaning Data

Pada proses cleaning data, proses ini berfungsi untuk menghilangkan link url, tanda baca, angka, simbol dan username yang berada pada *tweet* dan *re-tweet*.

3.3.4 Case Folding

Proses pengambilan data *tweet* “Antivaksin Covid” masih memiliki bentuk yang tidak standar. Jika data tersebut dianalisis maka hasil tidak akan akurat, sehingga diperlukan konversi dari bentuk awal menjadi bentuk standar.

3.3.5 Stopword Removal

Stopword removal yaitu melakukan penghapusan kata-kata yang memiliki informasi rendah dari sebuah teks (“yang”, “dan”, “di”, “dari” dll). Peneliti menggunakan proses

3.3.6 Stemming

Stemming merupakan proses penghilangan infleksi kata menjadi bentuk dasar. Peneliti menggunakan *Pyhton Sastrawi* dalam proses *Stemming*, hal ini dikarenakan data *tweet* yang berbahasa Indonesia semua kata tambahan sufiks dan prefix harus dihilangkan

3.3.7 Normalisasi

Normalization merupakan penyeragaman pada term yang mengalami kesalahan penulisan atau menggunakan bahasa yang tidak baku. Hasil dari proses *Normalization* berupa data *tweet* yang lebih terstruktur serta dapat dilakukan perhitungan pada proses selanjutnya.

3.3.8 Pembagian Data Training dan Data Testing

Pada proses ini digunakan untuk membagi data training dan data testing menggunakan Data frame agar data dilakukan training dan testing pada data tweet.

3.3.9 TF-IDF

TF-IDF digunakan untuk memberikan bobot pada hubungan sebuah kata (term) dengan sebuah dokumen. Metode ini menggabungkan dua konsep untuk menghitung bobot, yaitu frekuensi kemunculan suatu kata dalam dokumen tertentu dan frekuensi kebalikan dari dokumen yang mengandung kata tersebut.

3.3.10 NBC

NBC digunakan untuk menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan.

3.3.11 Testing

Pada tahap testing merupakan tahapan untuk mengetahui tingkat keakuratan pemodelan yang telah dibangun serta untuk memprediksi label atau kelas dari data uji yang tersedia.

3.3.12 Akurasi

Menghitung nilai akurasi digunakan untuk menguji kebenaran penelitian yang diteliti.

```

result_tweet=[]
for i in range(len(predicted)):
    if(predicted[i]==1):
        sentiment_result='Positif'
    elif(predicted[i]==0):
        sentiment_result='Negatif'
# result_tweet.append({'class':prediction_linear[i], 'result_nbc':sentiment_result})
result_tweet.append({'Cleaned_Text':data_tweet[i], 'class':predicted[i] })

```

Gambar 2. Kode Pemanggilan Prediksi

3.3.13 Hasil Klasifikasi

```

import pickle

with open('model_klasifikasi_tweet.pickle', 'wb') as files:
    pickle.dump(NB_classifier, files)

```

Setelah mendapatkan nilai akurasi yang baik dari proses *training* dan *testing* dari pemodelan klasifikasi maka dari itu dilakukannya tahap klasifikasi untuk data keseluruhan agar mendapatkan hasil sentimen positif dan negatif yang telah diuji dari tahapan *training* dan *testing*. Pada gambar 3 merupakan hasil klasifikasi seluruh tweet.

	normalized_tweet	label	predicted_label
0	adonii antivaksin nie kan duk kesah side effect...	0	0
1	pelik geng antivaksin ni kecil dulu kau tak cu...	1	0
2	facebook padam ratus akaun kempen antivaksin c...	1	0
3	aku ja antivaksin ni share pasai amal depa ceg...	0	0
4	mati covid tu kata salah hospital covid ni asa...	0	0

Gambar 3. Hasil Klasifikasi