

BAB III

METODOLOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

Subjek penelitian adalah pendapat masyarakat terhadap PPKM di DKI Jakarta pada media sosial YouTube. Sedangkan, pada penelitian ini yang sebagai objek penelitian merupakan konten berita yang membahas tentang naiknya level PPKM di DKI Jakarta pada *social media* YouTube.

3.2 Alat dan bahan penelitian

Alat:

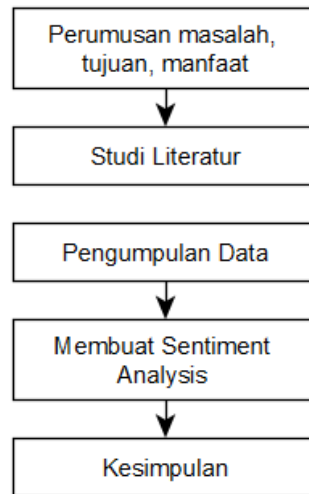
1. *Laptop* (Prosesor i5 generasi 11, Ram 16GB, GPU RTX 3050).
2. Jupyter Notebook.

Bahan:

1. Data komentar dari YouTube.

3.3 Diagram alir penelitian

Pada penelitian ini, pertama peneliti menentukan permasalahan yang akan di selesaikan, tujuan dan manfaat. Lalu peneliti mencari studi literatur untuk membantu menyelesaikan permasalahan yang sudah di tentukan. Setelah mendapatkan studi literatur, peneliti merumuskan hipotesis. Setelah itu peneliti mencari data untuk di analisis dengan metode Naïve Bayes. Setelah mendapatkan data, peneliti melanjutkan untuk membuat sentimen analisis mengenai Covid-19 dan akurasi dari metode *naïve bayes* untuk membuat analisis. Lalu, berdasarkan hasil sentimen analisis tersebut, peneliti membuat kesimpulan. Gambar 3.1 menunjukkan diagram alir penelitian



Gambar 3. 1 Diagram alir penelitian

3.3.1 Perumusan Masalah, Tujuan dan Manfaat

Tahap pertama, peneliti membuat perumusan masalah dengan melihat berita mengenai PPKM di DKI Jakarta. Berita tersebut di ambil dari sumber yang terpercaya seperti detik dan cnn Indonesia. Selanjutnya peneliti menentukan tujuan dan manfaat dari penelitian ini.

3.3.2 *Studi Literatur*

Tahap kedua, peneliti mencari jurnal yang relevan dengan penelitian ini. Peneliti menggunakan jurnal atau artikel ilmiah yang terdapat dari sumber yang terpercaya. Sumber yang digunakan yaitu IEEE, Scopus dan Google Scholar dengan rentang waktu 5 tahun terakhir.

3.3.3 Pengumpulan Data

Tahap ketiga, peneliti mengunpulkan komentar YouTube dengan menggunakan API YouTube dengan bantuan *apps script*. Untuk konten yang diambil oleh peneliti merupakan berita yang membahas naiknya PPKM ke level 2 di DKI Jakarta. Karena PPKM tersbut hanya berlangsung selama 2 hari, peneliti akan mengambil paada rentang waktu 5-6 juli 2022.

3.3.4 Membuat Sentiment Analysis

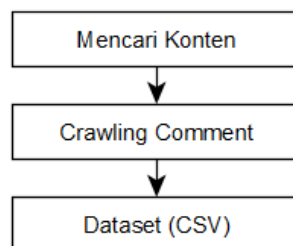
Tahap keempat, peneliti membuat sebuah sistem untuk membuat menganalisa dari data yang telah dikumpulkan. Pada penelitian ini, peneliti menggunakan sentiwordnet untuk melakukan label dan menggunakan Naïve Bayes untuk melakukan klasifikasi.

3.3.5 Kesimpulan

Tahap kelima, peneliti melihat hasil dari diagram data yang telah dilabeli dan melihat kualitas dari Naïve Bayes untuk melakukan klasifikasi. Dari hasil tersebut, peneliti akan membuat sebuah kesimpulan yang bisa diambil.

3.4 Pengumpulan Data

Penelitian ini mendapatkan data komentar dari YouTube. Dalam penelitian ini tahapan sentimen analisis terdiri dari mencari konten pada YouTube, melakukan *crawling data* dengan API v3, menggubah menjadi .csv. Gambar 3.2 menunjukkan diagram alir *Crawling data*:



Gambar 3. 2 Diagram alir Crawling Data

3.4.1 Mencari Konten

Dalam mencari konten berita, peneliti menggunakan kata kunci “PPKM di DKI Jakarta pada 5-6 juli 2022”. Dengan kata kunci tersebut peneliti akan mengambil setiap alamat pada setiap video YouTube yang ditemukan. Nantinya setiap alamat yang disimpan akan dipanggil untuk mendapatkan komentarnya dengan API.

3.4.2 *Crawling Comment*

Crawling Comment adalah tahapan untuk mendapatkan data komentar dari YouTube. Untuk mendapatkan komentar, peneliti menggunakan API YouTube dengan Apps Script yang merupan

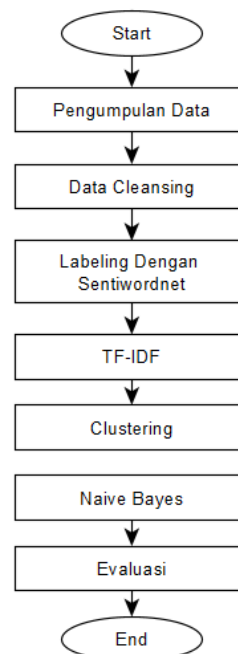
layanan dari Google. Untuk mendapatkannya hanya perlu memasukkan alamat pada google sheet dan menjalankan Apps Script.

3.4.3 Dataset

Tahap ini merupakan menggabungkan file-file yang telah didapatkan menjadi satu file. Selain itu peneliti juga menindahkan setiap *replay* ke dalam kolom komentar. Hal ini ditujukan agar semua koementar dapat lebih mudah untuk dianalisa dalam penelitian.

3.5 Analisis Data

Variabel penelitian terdiri dari pendapat masyarakat mengenai Covid-19 dan akurasi dari metode *naïve bayes* untuk membuat analisis. Untuk mendapatkan akurasi yang baik tentu memerlukan tahapan yang tepat. Dalam penelitian ini tahapan sentimen analisis terdiri dari pengumpulan data, data cleansing, Labeling, TF-IDF, Split data, Pemodelan Naïve bayes, dan evaluasi. Gambar 3.3 menunjukkan diagram alur tahapan sentimen analisis:



Gambar 3. 3 Diagram Tahapan Analisis Sentimen

3.5.1 Data Cleansing

Pada tahapan ini peneliti melakukan data *cleansing* yang meliputi penghapusan *missing value*, pembersihan element yang tidak digunakan, *stemming* menggunakan sastrawi, melakukan *stop words* dan *tokenisasi*. Pada tahapan ini akan menghasilkan data sebagai berikut :

	Komentar	POS tagged	Lemma
1042		[]	
956	aaamiin ya allah	[(aaamiin, n), (ya, n), (allah, n)]	aaamiin ya allah
436	aamiin	[(aamiin, n)]	aamiin
572	aamiin	[(aamiin, n)]	aamiin
723	aamiin moga doa kabul aamiin	[(aamiin, n), (moga, n), (doa, n), (kabul, n), ...]	aamiin moga doa kabul aamiin

Gambar 3. 4 Hasil Data Cleansing

3.5.2 Labeling

Pada bagian *labeling*, peneliti menggunakan *sentimentnet* untuk mengidentifikasi data menjadi 3 kategori yang meliputi *Netral*, *Positive* dan *Negative*. Dengan proses ini menemukan 896 komentar netral, 113 komentar negative dan 55 komentar positive. Pada tahapan ini menghasilkan output sebagai berikut:

	Komentar	POS tagged	Lemma	SIM analysis
1042		[]		Neutral
956	aaamiin ya allah	[(aaamiin, n), (ya, n), (allah, n)]	aaamiin ya allah	Neutral
436	aamiin	[(aamiin, n)]	aamiin	Neutral
572	aamiin	[(aamiin, n)]	aamiin	Neutral
723	aamiin moga doa kabul aamiin	[(aamiin, n), (moga, n), (doa, n), (kabul, n), ...]	aamiin moga doa kabul aamiin	Neutral

Gambar 3. 5 Hasil Labeling

3.5.3 TF-IDF

Setelah proses labeling, selanjutnya data akan diekstraksi feature dengan menggunakan TF-IDF. Data yang di lakukan TF-IDF yaitu data yang di bersihkan dari nilai netral.

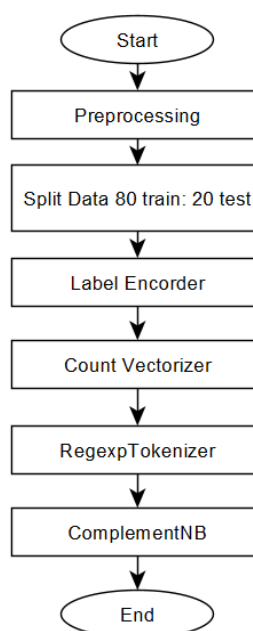
3.5.4 Clustering

Pada tahapan ini, data yang telah di TF-IDF akan di clustering dengan menggunakan K-Means. Clustering ini di

gunakan untuk sebagai memberi label dan meningkatkan akurasi saat di modelkan dengan Naïve Bayes.

3.5.5 Pemodelan Naïve Bayes

Pada tahap ini peneliti melakukan pemodelan klasifikasi dengan Algoritma *Complementary Naive Bayes*. Pada tahapan ini berisi:



Gambar 3. 6 Alur Naïve Bayes

3.5.5.1. Preprocessing

Pada tahap ini, peneliti mempersiapkan data yang akan diolah. Data yang digunakan yaitu data yang telah diberi label.

3.5.5.2. Split Data

Pada tahapan ini, peneliti membagi 2 data dengan perbandingan 80:20 (80% data latih, 20% data uji). Data yang split yaitu data yang telah di hapus yang *Netral*, lalu di split menjadi Train_X, Test_X, Train_Y, Test_Y dan kemudian Train_X, Test_X di training menggunakan TF-IDF menjadi Train_X_tfidf dan Test_X_tfidf.

3.5.5.3. Label Encoder

Pada tahap ini, peneliti melakukan mengubah isi kolom menjadi angka dengan *library* `sci-kit` `LabelEncoder`.

3.5.5.4. Count Vectorizer

Pada tahap ini, peneliti melakukan `Count Vectorizer` agar data bisa di hitung dengan `Naïve Bayes`.

3.5.5.5. RegexpTokenizer

Pada tahap ini, peneliti melakukan tokenisasi dengan menggunakan `library` `RegexpTokenizer`

3.5.5.6. ComplementNB

Pada tahap ini, peneliti melakukan klasifikasi pada data yang telah di olah sebelumnya.

3.5.6 Evaluasi

Tahap terakhir ini adalah melakukan evaluasi terhadap system yang telah di buat. Untuk evaluasi ini, peneliti melihat akurasi dari system yang telah di buat sebelumnya. Pada tahapan ini peneliti mendapatkan output sebagai berikut:

	precision	recall	f1-score	support
0	0.96	0.85	0.90	26
1	0.64	0.88	0.74	8
accuracy			0.85	34
macro avg	0.80	0.86	0.82	34
weighted avg	0.88	0.85	0.86	34

Gambar 3. 7 Hasil Evaluasi