

BAB III

METODOLOGI PENELITIAN

3.1. Alat dan Bahan Penelitian

3.1.1. Alat Penelitian

Spesifikasi kebutuhan yang digunakan dalam pelaksanaan penelitian ini *hardware* dan *software*. Masing-masing spesifikasi yang dibutuhkan dijelaskan sebagai berikut.

- a. Spesifikasi Hardware terdiri dari PC / *Notebook* dengan spesifikasi *Processor* Intel(R) Celeron(R) CPU N3060 @ 1.60GHz (2 CPUs), 1.6GHz, RAM 4 Gb, System type 64-bit operating system, x64-based processor, Harddisk 500 GB; Mouse; dan Keyboard.
- b. Spesifikasi Software, terdiri dari:
 - 1) Operating System: Windows 10
 - 2) Web Browser: Google Chrome, digunakan untuk pengunduhan data.
 - 3) Python 3, digunakan sebagai bahasa pemrograman.
 - 4) Google Colab untuk membantu merunning program untuk kcepatan data yang diolah.

3.1.2. Bahan Penelitian

Bahan penelitian ini salah satunya adalah *dataset*. *Dataset* analisis sentimen berasal dari media sosial Youtube. Penulis memilih Youtube karena kebijakan Twitter yang memberikan akses ke data mereka. Twitter menyediakan application programming interface (API) untuk mengakses data-data yang diperlukan dengan terlebih dahulu mengajukan permintaan kepada Twitter untuk mendapatkan API key. API Twitter merupakan bagian dari pengembang Twitter yang berupa program yang memudahkan pengembang lain untuk mengakses informasi yang terdapat pada Twitter.

Selain itu, terdapat tiga alasan utama pemilihan media sosial Twitter untuk data mining [9], yaitu a) API Twitter dirancang dengan baik dan mudah diakses, b) data Twitter tersedia dalam format yang mudah untuk dianalisis, dan c) kebijakan data Twitter yang relatif liberal dibandingkan dengan API lainnya. Penulis juga menggunakan data Twitter karena beberapa tren analisis data saat ini menggunakan Twitter untuk melakukan analisis sentimen untuk mengevaluasi komentar pengguna yang positif atau negatif.

Untuk mendapatkan data, penulis menggunakan bahasa pemrograman Python 3 yang dikelola oleh google colabs untuk data mining, text mining, dan analisis prediktif. Python memiliki beberapa teknik deskriptif dan prediktif yang memberikan informasi kepada pengguna dan membantu mereka membuat keputusan terbaik. Pustaka Python mencakup berbagai operator data mining, termasuk operator input, output, pra-pemrosesan data, dan visualisasi yang dapat digunakan untuk melakukan operasi data mining dengan mudah dan cepat.

Seluruh proses penelitian ini dari proses awal data mining hingga visualisasi menggunakan Python melalui Google Colabs.

Data dikumpulkan dengan menggunakan metode data mining Twitter API dan disimpan dalam format csv. Dalam proses data mining, diperlukan kunci identifikasi berupa token dan kunci akses ke Twitter API yang diperoleh melalui langkah-langkah yang ditentukan oleh Twitter.

3.2. Tahapan Penelitian

Penelitian ini melalui beberapa tahapan yang harus dilakukan, yaitu dari awal mengumpulkan data atau penarikan data (*crawling*) dari Twitter hingga visualisasi. Adapun tahapan penelitian ini sebagaimana dalam gambar berikut ini.



Gambar 3. 1. Tahapan Penelitian

Berdasarkan gambar alur tahapan penelitian tersebut, dijabarkan lebih lanjut sebagai berikut.

a. Tahap pengumpulan data

Data yang digunakan ialah komentar pengguna media sosial Youtube yang ada pada channel “Pendidikan Vox Point” dengan judul video unggahan "Memaksakan Kurikulum Merdeka, Masih Waraskah Bangsa Ini???". Data dikumpulkan (*crawling*) secara langsung terhubung dari media sosial Youtube

dengan menggunakan google colabs, yang kemudian disimpan dalam bentuk file csv.

b. Tahap Pengolahan Data

Tujuan dari tahap ini adalah untuk mengekstrak dan membersihkan data yang diperoleh selama eksplorasi sebelumnya. Langkah-langkah dalam tahap ini: 1) penghapusan data berulang (*remove duplicate*), 2) pemilihan *attribute* 3) penghapusan tanda baca (*replace*) 4) *filterisasi stopword*, 5) *tokenisasi*, dan 6) mencari frekuensi kata-kata yang sering muncul.

c. Tahap Analisis Sentimen

Pada tahap ini, data diproses sehingga pelabelan dan penambahan atribut dapat dilakukan untuk menentukan nilai sentimen dari data tersebut (negatif, positif, atau netral). Langkah-langkah yang dilakukan dalam analisis sentimen adalah sebagai berikut[9].

1. Tahap 1: mencari sentimen negatif, positif, dan netral di setiap baris.
2. Tahap 2: menganalisis sentimen seluruh dokumen sebagai negatif, positif, serta netral.
3. Tahap 3: menerapkan pengelompokan untuk mengumpulkan semua atribut yang ada dengan skor sentimen yang sama.
4. Tahap 4: menggunakan visualisasi data dari analisis sentimen untuk interaksi antar pengguna.

d. Tahap Implementasi Algoritma

Setelah analisis sentimen, langkah selanjutnya adalah memproses data yang telah dilabeli untuk klasifikasi lebih lanjut. Metode yang digunakan untuk klasifikasi yaitu menggunakan metode KNN dan *Naive Bayes* dengan bantuan *tools google colabs*.

e. Tahap Visualisasi

Langkah terakhir adalah memeriksa hasil analisis sentimen menggunakan wordcloud di google colabs untuk mendapatkan gambaran umum tentang frekuensi kata-kata yang sering muncul.

3.3. Metode Yang Digunakan

Metode atau model yang diusulkan dalam penelitian ini yaitu algoritma KNN dan *Naive Bayes* untuk membandingkan algoritma mana yang memiliki akurasi lebih tinggi. KKN adalah salah satu algoritma klasifikasi yang konsepnya memeriksa jarak antara tetangga terdekat dengan data baru. Algoritma ini menemukan jarak terdekat antara set pelatihan dan k tetangga terdekat dari set pengujian. Algoritma KNN mengklasifikasikan dokumen uji X sebagai berikut [9].

- a. Tentukan parameter K, yang merupakan jumlah tetangga terdekat.
- b. Pada data training dilakukan, hitung kuadrat dari jarak Euclidean.
- c. Urutkan data tersebut dalam urutan menaik.
- d. Gabungkan kategori Y sesuai dengan parameter K..

Keunggulan metode *Naive Bayes* dalam proses klasifikasi ada sejumlah kecil data pelatihan yang diperlukan untuk penentuan estimasi parameter. Perbedaan yang lebih mendasar antara *teorema Bayes* untuk metode lain, parameter Bayesian dianggap sebagai variabel acak, sedangkan dalam statistik

klasik parameternya tidak dapat diketahui. Hubungan antara probabilitas bersyarat kejadian H dan X disebut teorema Bayes. Nilai akurasi yang diperoleh lebih tinggi dibandingkan metode lainnya[39].

Data yang dikumpulkan akan diproses terlebih dahulu agar tidak menjadi tidak akurat atau tidak relevan, sehingga hasil prediksi dapat diperhitungkan.

Pada tahap ini dilakukan pengklasifikasian teks atau sentimen melalui tahapan *preprocessing* agar teks mempunyai isi yang tidak sempurna seperti data yang hilang, tidak valid, salah ketik, atau juga atribut-atribut data yang tidak relevan. Data tersebut dibuang karena keberadaannya bisa mengurangi mutu atau akurasi. Teks yang belum diolah biasanya memiliki karakteristik dimensi yang tinggi, terdapat *noise* pada data dan terdapat struktur teks yang tidak baik.

3.4. Evaluasi Metode

Skenario yang diterapkan untuk melakukan evaluasi dan validasi agar memastikan validitas metode dan hasil penelitian, diusulkan model dengan menerapkan algoritma KNN dan *Naïve Bayes*. Algoritma KNN dan *Naïve Bayes* diyakini dapat menghasilkan akurasi yang baik.

Proses dilakukan dengan melakukan pengujian dengan mengukur *performance* dari algoritma klasifikasi *text mining* yang dipakai. Dalam penelitian ini *performance* diukur menggunakan *Accuracy* dan AUC serta akan ditampilkan dalam bentuk kurva ROC.

No	Kegiatan	Bulan																
		Juli				Agustus				September					Oktober			
		1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4
5	Seminar proposal																	
6	Perbaikan proposal																	
7	Penyusunan Tesis																	
8	Sidang Tesis																	
9	Perbaikan dan finalisasi Tesis																	
10	Penggandaan dan pengumpulan Tesis																	