

BAB 3

METODE PENELITIAN

Penelitian ini adalah penelitian analisis sentimen berdasarkan opini pengguna PDAM *platform* Twitter yang menghasilkan sentimen puas dan tidak puas menggunakan metode Naive Bayes *classification*. Penelitian ini dilatar belakangi permasalahan yang ada kemudian menganalisis data PDAM di *platform* Twitter untuk mengatasi permasalahan tersebut. Berikut ini bahan, alat, metode dan jalannya penelitian analisis sentimen tentang PDAM beserta langkah-langkah penelitian untuk menyelesaikan analisis menggunakan data *tweet* dan merancang sistem berbasis web untuk menampilkan hasil analisis yang diperoleh.

3.1 BAHAN DAN ALAT PENELITIAN

Penelitian ini membutuhkan data *tweet* Twitter dengan menggunakan kata kunci "pdam", "pelayanan pdam", "kualitas pdam", dan "pdam sleman". Data *tweet* diambil dari bulan Maret – Juni 2023 dengan jumlah data yang ada.

Alat yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi cukup untuk menjalankan sistem operasi dan perangkat lunak pengembangan serta koneksitas Internet.

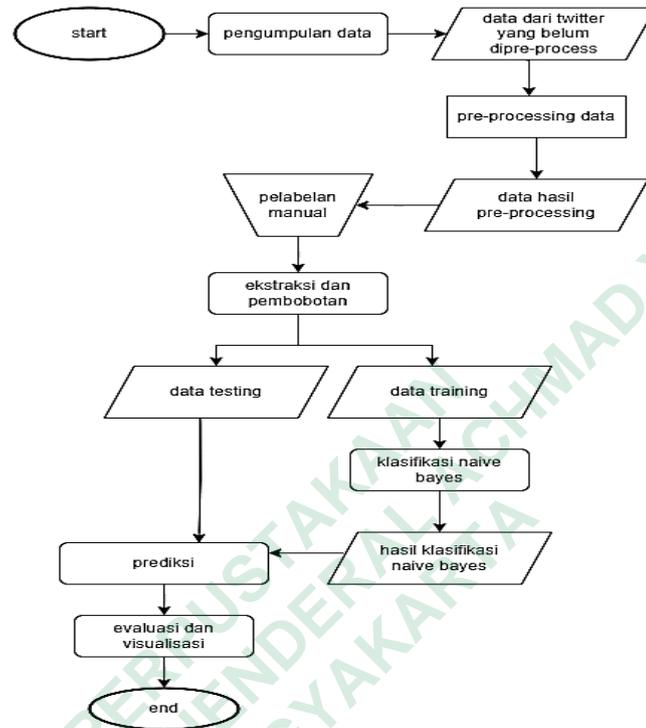
Sistem Operasi dan program aplikasi yang digunakan dalam pengembangan aplikasi ini adalah:

1. Sistem Operasi: Windows 11
2. Bahasa Pemrograman Python 3.10.6
3. *Framework* Flask 2.2.2
4. Microsoft Office Excel 2016
5. Sublime Text Build 4107
6. Google Collab

3.2 JALAN PENELITIAN

Naive Bayes *classification* adalah metode pengklasifikasian probabilitas sederhana yang menghitung jumlah probabilitas dengan meningkatkan frekuensi dan menggabungkan nilai dari dataset yang disediakan dan setiap kategori tidak

memiliki ketergantungan satu sama lain. NBC adalah metode *supervised document classification* yang memerlukan data *training* sebelum memulai proses klasifikasi (Devita et al., 2018). Berikut ini alur penelitian dan tahapan dari metode Naive Bayes *classification* pada Gambar 3.1.



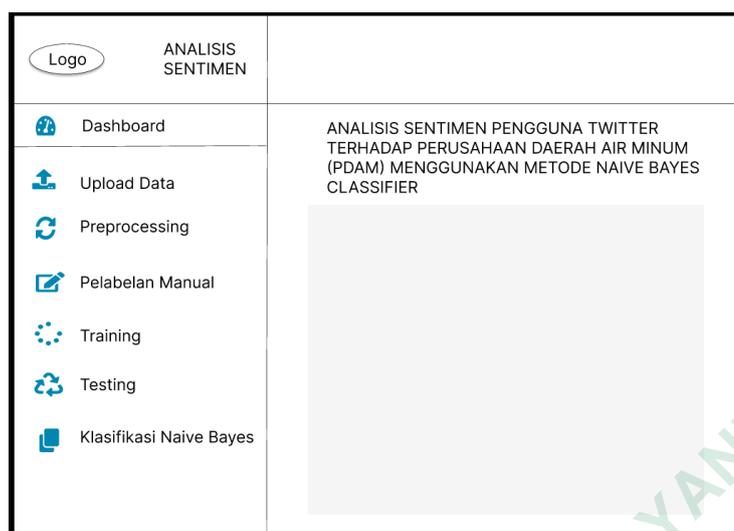
Gambar 3.1 Alur penelitian

1. Tahap pengumpulan data, dilakukan pengambilan data *tweet* dari Twitter tentang PDAM dengan menggunakan *tools* Google Collab, data yang diperoleh berformat *comma separated values* (CSV) yang dapat dibuka di Microsoft Excel.
2. Tahap *preprocessing data*, tahap ini dilakukan untuk proses pengolahan data teks yang belum terstruktur untuk diperbaiki dengan tahapan-tahapan berikut ini:
 - a. *Tokenization* adalah proses pemecahan kalimat dan penghapusan simbol dan tanda baca dalam data teks.
 - b. *Case folding* adalah perubahan bentuk huruf dari bentuk awal (huruf kecil atau *lowercase*).

- c. *Stopword removal* adalah proses penghapusan kata yang memiliki informasi rendah dalam data teks.
 - d. *Number removal* adalah proses membersihkan karakter angka yang terdapat dalam data teks.
 - e. *Punctuation Removal* adalah proses penghapusan karakter khusus yang tidak memiliki pengaruh signifikan dalam analisis.
 - f. *Whitespaces Removal* adalah proses menghilangkan spasi di awal dan akhir kalimat.
 - g. *Normalization* adalah proses standarisasi *term* yang mengalami kesalahan penulisan atau penggunaan bahasa yang tidak lazim.
 - h. *Stemming* adalah proses penghapusan infleksi kata menjadi bentuk dasarnya.
3. Tahap pelabelan *manuals*, tahap pelabelan *manuals* ini untuk memberikan sentimen puas atau tidak puas terhadap kata atau kalimat yang ada pada dokumen agar dapat dianalisis.
 4. Tahap *training data*, tahap ini diawali dengan fitur ekstraksi data teks menggunakan TF-IDF. Proses *training data* digunakan untuk membuat model klasifikasi sentimen secara otomatis.
 5. Tahap *testing*, tahap ini untuk mengetahui tingkat keakuratan pemodelan yang dibangun, hasil testing dihitung dengan rumus *confusion matrix* untuk mengetahui *persentase* dari setiap prediksi pengujian kelas dan label.
 6. Tahap klasifikasi, pada tahap ini menampilkan hasil analisis ke dalam bentuk dashboard menggunakan bahasa pemrograman Python dan *framework* Flask.

3.3 DESAIN INTERFACE

Perancangan antarmuka dashboard analisis sentimen dibuat untuk menampilkan representasi data agar lebih mudah dipahami. Desain rancangan antarmuka dashboard analisis sentimen ditunjukkan pada Gambar 3.2.



Gambar 3.2 Desain *interface*

Pada antar muka dashboard, terdapat beberapa tombol menu, yaitu menu upload data, *preprocessing*, pelabelan manual, *training*, *testing* dan klasifikasi naive bayes. Menu upload data digunakan untuk melakukan upload file data mentah ke *directory* untuk mempermudah proses pengolahan data selanjutnya, menu *preprocessing* digunakan untuk melakukan proses pengolahan data teks dan menampilkan hasil proses pengolahan, menu pelabelan manual digunakan untuk menghitung nilai *accuracy*, *precision*, *recall* dan *f1-score* yang sudah diberi label dan kelas secara manual, menu traing digunakan untuk membuat model klasifikasi sentiment secara otomatis dan menghitung TF-IDF pada data teks, menu *testing* digunakan untuk menghitung nilai *accuracy*, *precision*, *recall* dan *f1-score* yang sudah di prediksi secara otomatis kelas dan labelnya, dan menu klasifikasi naive bayes digunakan untuk menampilkan data *tweet* yang sudah di prediksi label dan kelasnya pada tahap *training*.