

BAB III

METODE PENELITIAN

3.1 Pengumpulan Data

Penelitian ini menggunakan data berupa *tweet* dan *retweet* dari pengguna *Twitter* (Nugroho, 2018). Data terbatas pada 3 aplikasi pinjaman online di Indonesia yaitu AdaKami, Kredit Pintar, dan Tunaiku. Pengumpulan data dilakukan mulai 2 Desember 2023 – 5 Juni 2024 secara terpisah antara aplikasi pinjaman online satu dengan yang lainnya, tujuan dari penelitian ini adalah untuk mengidentifikasi aplikasi pinjaman online yang paling populer berdasarkan komentar positif yang terdapat pada masing-masing aplikasi. Pengambilan data secara keseluruhan dianggap tidak adil karena hasil *crawling* data mungkin tidak memberikan jumlah data yang sama untuk setiap aplikasi pinjaman online. Oleh karena itu, penelitian ini menggunakan 500 data *tweet* untuk setiap aplikasi pinjaman online, sehingga total data yang digunakan adalah 1.500 *tweet*.

Dalam penelitian ini, penulis menggunakan teknik *crawling* data melalui *Google Colab* untuk mengumpulkan data. *Crawling* data di *Twitter* adalah proses pengambilan atau pengunduhan data dari *server Twitter* menggunakan *Application Programming Interface (API) Twitter*, baik itu data pengguna maupun data *tweet* (Pintoko et al., 2018). Penetapan data ini dilakukan menggunakan cara mengumpulkan *tweet* dari media *Twitter* dengan memasukkan kata kunci atau *keyword* yang diinginkan, dalam penelitian ini penulis menggunakan kata kunci atau *keyword* “AdaKami”, “Kredit Pintar”, dan “Tunaiku”.

3.2 Teknik Analisis Data

Dalam penelitian ini, penulis menggunakan perangkat lunak *Rapid Miner* untuk menganalisis data yang dikumpulkan. *Rapid Miner* adalah perangkat lunak yang digunakan dalam lingkungan *machine learning*, *data mining*, *text mining*, dan *predictive analytics* (Rachman & Wasianty, 2019). *Rapid Miner* dikenal sebagai perangkat lunak *open source* yang dapat

digunakan untuk analisis *data mining*. Dengan prinsip dan algoritma data mining, Rapid Miner dapat menemukan pola dari kumpulan data yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan, dan basis data. Penggunaannya dipermudah melalui operator – operator yang memodifikasi data, di mana data dihubungkan dengan node – node pada operator, lalu hasilnya dapat dihubungkan ke node output. Rapid Miner juga dapat menampilkan hasil secara visual dalam bentuk grafik, menjadikannya pilihan yang tepat untuk ekstraksi data menggunakan metode data mining.

3.3 Metode Yang Diusulkan

Teorema *Bayes* dikombinasikan dengan konsep '*Naive*' yang berarti bahwa setiap atribut atau variabel dianggap independen. *Naive Bayes Classifier* dapat dilatih secara efisien dalam pembelajaran terawasi (*supervised learning*). Salah satu keuntungannya adalah bahwa metode ini hanya memerlukan sejumlah kecil data pelatihan untuk memperkirakan parameter (varian dan sarana dari variabel) yang diperlukan untuk klasifikasi. Karena variabel diasumsikan independen, hanya varian dari variabel pada setiap kelas yang harus ditentukan, bukan seluruh matriks kovarians (Hikmah & Utammimah, 2017).

Metode *Bayesian Learning* ini adalah yang paling mudah, cepat, dan sederhana. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya memerlukan sedikit data pelatihan (Training Data) untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi (Saleh, 2015). Berdasarkan teorema *Bayes* dan hipotesis independensi, metode ini menghasilkan pengklasifikasi statistik berdasarkan peluang. *Naive Bayes* merupakan metode pengklasifikasian dengan probabilitas dan statistik yang diperkenalkan oleh ilmuwan Inggris, Thomas Bayes (Idris, 2019). Algoritma ini mengasumsikan bahwa atribut objek adalah independen. *Naive Bayes Classifier* memiliki keunggulan karena tidak memerlukan banyak data latih untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi, hanya memerlukan sedikit data latih. Berikut

adalah langkah-langkah yang digunakan dalam perhitungan metode *Naïve Bayes* :

1. Menentukan kriteria – kriteria
2. Membuat tabel aturan
3. Menhitung probabilitas yang sering muncul setiap nilai untuk atribut dari setiap kriteria
4. Menghitung nilai dari setiap kriteria
5. Menghitung nilai probabilitas

3.4 Eksperimen dan Cara Pengujian Model/Metode

Dalam penelitian ini pengujian melalui beberapa langkah, yaitu sebagai berikut :

1. Pengambilan Data

Data yang diambil dalam penelitian ini merupakan data yang diperoleh dari hasil *crawling* data yang dilakukan pada media sosial *Twitter* dengan keyword “*AdaKami*”, “*KreditPintar*”, dan “*Tunaiku*” dengan rentang waktu mulai 2 Desember 2023 – 5 Juni 2024.

2. *Preprocessing* Data

Dalam melakukan *preprocessing* data ini berfungsi mengubah data yang tidak terstruktur. Tahapan dalam *preprocessing* data ini terdapat 5 tahapan diantaranya, *cleaning data*, *case folding*, *tokenizing*, *filtering*, dan *stemming*.

3. Labeling Data

Dalam proses pelabelan data, dilakukan untuk mengkategorikan data menjadi tiga kategori, yaitu positif, netral dan negatif.

4. Data Training dan Testing

Dalam tahap ini penulis menggunakan rasio 80 : 20 untuk data set *training* dan *testing*.

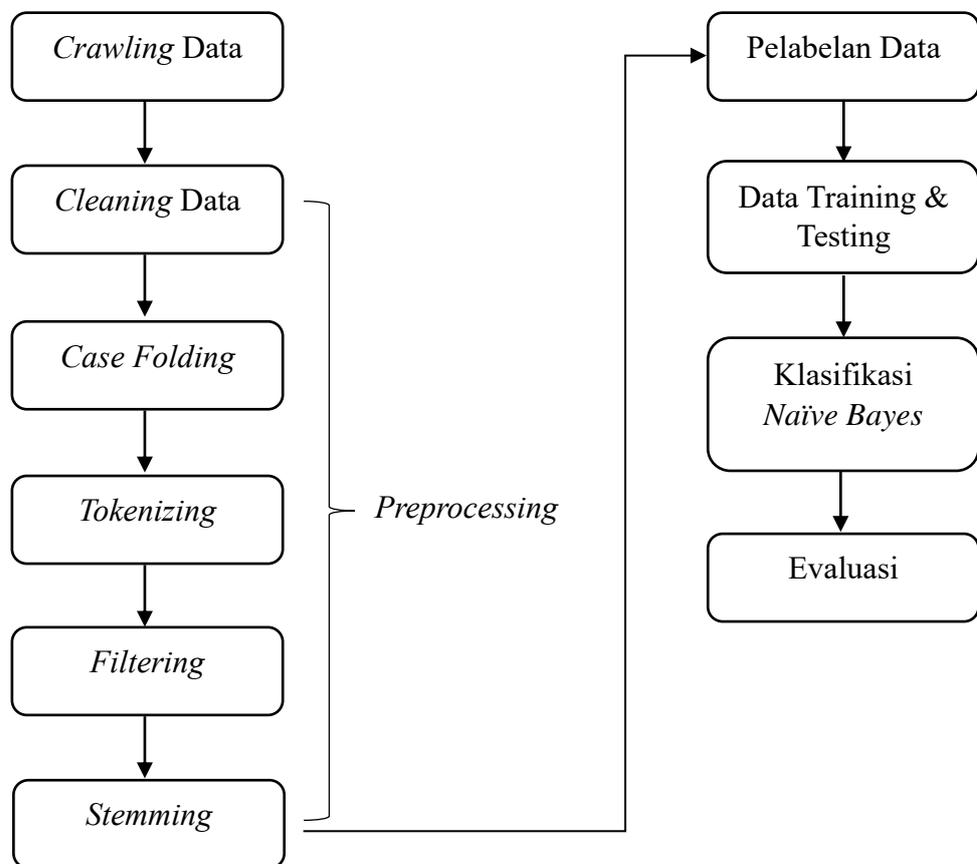
5. Klasifikasi *Naïve Bayes*

Klasifikasi menggunakan algoritma *Naïve Bayes*. Tahap ini dilakukan menggunakan *software* Rapid Miner, ketika semua data

sudah terkumpul yang diambil dari *tweet* pada *Twitter* dan penerapannya di sebuah model berdasarkan teorema *Bayes*.

6. Evaluasi

Langkah terakhir melakukan evaluasi yang dilakukan dengan perhitungan *accuracy*, *precision*, *recall* dan *f1-score*. Evaluasi ini menggunakan *library* yang dimiliki *software* Rapid Miner.



Gambar 3.1 Alur Penelitian