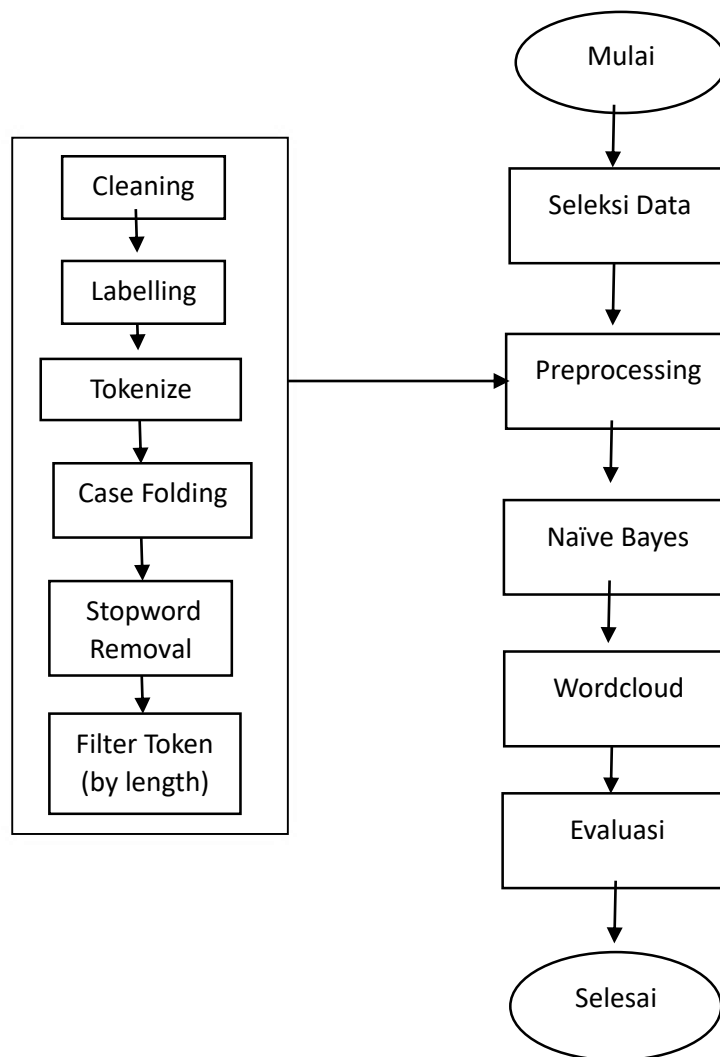


BAB III METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metodologi yang digunakan pada penelitian ini yaitu *Knowledge Discovery in Database (KDD)* sendiri merupakan suatu teknik untuk mendapatkan informasi atau pengetahuan dari suatu data yang telah dipilih [Siroj, Arwani, & Ratnawati, 2021]. Terdapat 5 tahapan sebagai berikut, Seleksi Data, Preprocessing, Transformasi, Data Mining, Evaluasi.

Berikut adalah Alur Penelitian yang akan dilakukan :



Gambar 3.1 Alur Penelitian

No	User	Ulasan
7	Imam	barang bagus, mantap. langsung buat Nanak nasi. tidak ada kendala. Aman. jos brott
8	Trijayanto	Full set. Sementara ini oke, semoga awet. Fast response
9	c***a	barang ok, packingnyaaa aman bangettt, siiiiipppp..
10	Teguh	brg sangat baik dan pengiriman cpt semoga awet dan terima kasih semoga tambah rame tokonya kak
11	Trifena	pengiriman cepat, packing bagus..puas beli di store ini
12	Yuanita	sudah 3000 tahun saya pake rice cooker. kalo rusak selalu beli lagi yg merek ini
13	N***a	pengiriman cepat, hari ini pesan besok sudah terima, terimakasih banyak. barang bagus.
14	Yosi	packing rapi, barang sesuai order
15	Buchari	bagus dan berkualitas

2. Preprocessing

Tahapan yang selanjutnya adalah *preprocessing*, dimana diharapkan dapat menghasilkan informasi yang dapat dianalisis, Informasi tersebut harus diproses terlebih dahulu untuk memastikan kualitas yang baik sebelum digunakan untuk analisis. Tujuan dari Prapemrosesan ini adalah untuk membersihkan dan menyiapkan data teks yang siap untuk di ekstraksi.

Tahapan *Preprocessing* :

1) Cleaning

Data hasil ulasan sebanyak 1009 data kemudian di lakukan proses pembersihan data agar tanda baca dan emoji pada komentar di

bersihkan dan menyisakan data saja. Data yang diperoleh dari proses pembersihan data tersebut berubah menjadi 966 data ulasan.

2) Labeling Data

Langkah selanjutnya setelah semua data dibersihkan melalui *text processing* tahap selanjutnya adalah pelabelan data secara manual. adalah pelabelan pada kelas sentimen menggunakan kategori positif dan kategori negatif.

3) Tokenize

Proses memisahkan kata dari data ulasan yang berjumlah 965 dan di pisah menjadi sebuah kata dengan jumlah 1854 kata.

4) Case Folding

Data yang digunakan ialah dari ulasan produk Cosmos di Tokopedia, *Case Folding* adalah proses merubah huruf besar menjadi huruf kecil atau huruf standar. Proses ini mempermudah pencarian karena tidak semua dokumen teks memiliki kapitalisasi yang konsisten.

5) Stopwords Removal

Pada proses menghilangkan kata-kata yang kurang penting atau menyesatkan dari kosakata yang tidak mempunyai arti untuk mendapatkan kamus datanya berasal dari Kaggle.com Versi Indonesia.

6) Filter Token (by length)

Filter Token (by length) adalah penghapusan sejumlah kata singkatan yang minimal karakternya digunakan peneliti menggunakan panjang minimal 4 karakter dan maksimal 25 karakter.

3. Naïve Bayes

Pada tahapan ini melaksanakan Analisa informasi bersumber pada tata cara yang sudah ditetapkan ialah Naïve Bayes. Tahapan ini dicoba untuk cara pengelompokkan bersumber pada berat nilai yang sudah dicoba lebih dulu. Hasil yang dicoba merupakan memilah label positif serta negative pada keterangan komentar Ulasan produk Cosmos di

Tokopedia menggunakan algoritma Naïve Bayes untuk memperoleh nilai ketepatan terbaik.

Pengelompokan Naïve Bayes salah satu tata cara yang dipakai dalam informasi mining yang bertabiat supervised learning. Naïve Bayes didasarkan teorema Bayes yang mempunyai keahlian pengelompokan seragam dengan decision tree serta neural jaringan. Tidak hanya itu, tata cara Naïve bayes mempunyai durasi pengelompokan yang pendek memusatkan cara system Analisa afeksi, dengan Teorema sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

P(H) : Probabilitas hipotesis H (prior probability)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

4. Wordcloud

Pada Tahap ini, tampilan bobot kata yang dominan muncul, divisualisasikan dalam bentuk wordcloud.

5. Evaluasi

Pengujian pada penelitian ini untuk mengevaluasi hasil dari Algoritma yang digunakan. Metode evaluasi yang digunakan Confusion Matrix, metode ini cukup membantu untuk proses analisis kualitas Classifier. Pengujian ini akan menghitung nilai accuary, recall, precision, dan fl-score yang akan ditampilkan dalam bentuk persentase. Perhitungan confusion matrix untuk menghitung akurasi pengujian, sebagai berikut :

$$\text{Accuary} = \frac{TP+TN}{TP+FN+FP+TN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 - Score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Keterangan :

TP : nilai *True Positive*

FP : nilai *False Positive*

FN : nilai *False Negative*

TN : nilai *True Negative*.