BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Objek dalam penelitian ini merupakan data *review e commerce* yang diambil dari kaggle. Kaggle adalah komunitas yang bertujuan untuk belajar permasalahan data *science* dari dataset, berikut link kaggle yang dipakai https://www.kaggle.com/christofel04/review-lapak-sentiment . data yang digunakan yaitu sebanyak 500 dataset. Pelaksanaan penelitian dilakukan kurang lebih 6 bulan, berikut adalah tabel perinciaan pelaksanaan kegiatan penelitian :

Tabel 3. 1 Tabel Penelitian

Kegiatan	Bulan						
	Oktober	November	Desember	Januari	Februari	Maret	April
Pengumpulan data							
Analisis data							
Pengolahan data							
Pengujian							
Evaluasi							
Tugas akhir							

3.2 Peralatan Penelitian

Penelitian ini dilakukan dengan menggunakan peralatan berupa *Hardware* dan *Software*.

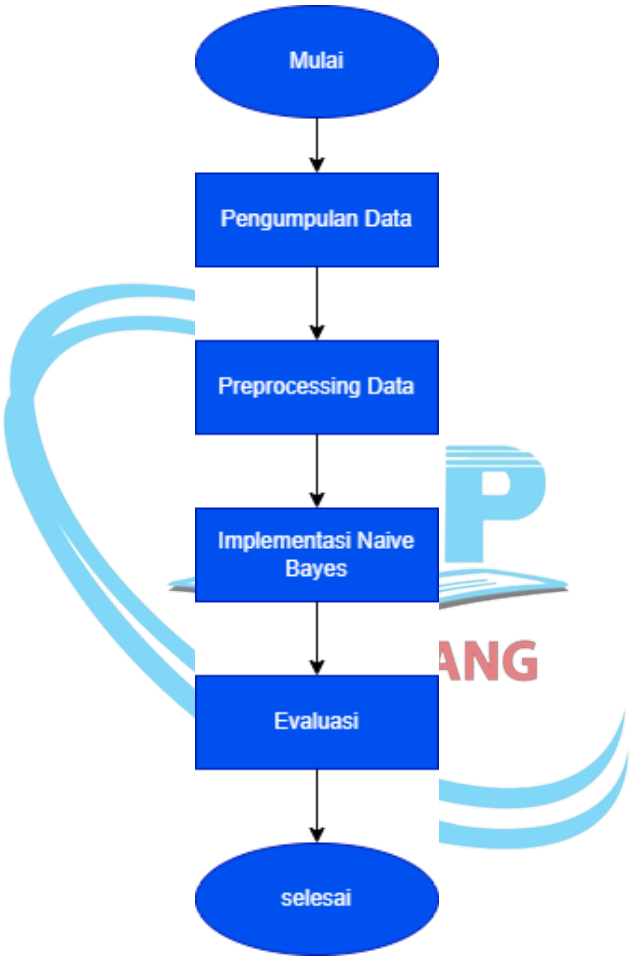
1. *Hardware* yang digunakan dalam penelitian ini adalah
   a. Laptop Asus A442U
   b. Intel CORE i5 8th Gen
   c. *Memory size* 12 GB
2. *Software* yang digunakan dalam penelitian ini adalah
   a. *Microsoft word* 2013

b. *Microsoft exel* 2013

c. Google Colaboratory

3.3 Prosedur Penelitian

Tahapan pada penelitian Tugas Akhir yang akan dilakukan tertera pada gambar berikut :



Gambar 3. 1 Prosedur Penelitian

3.3.1 Pengumpulan Data

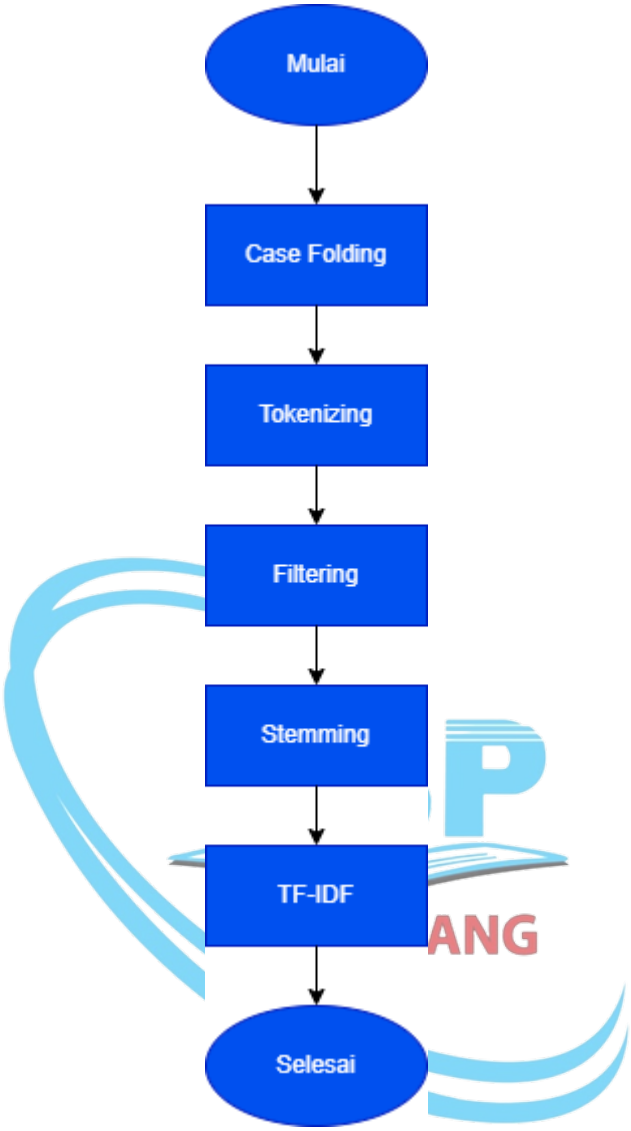
Penelitian diawali dengan pengumpulan data, tahapan ini proses yang dilakukan adalah dengan mencari data pada *website* kaggle. Contoh data yang didapat dan digunakan dapat dilihat pada tabel berikut :

Tabel 3. 2 Contoh dataset

Review	Label
Keren bgt sepatu nya... Nyaman dikaki, ringan pokok nya mantap..	1
kecil ga muat di pakai dan ga bisa melar	0
Murah banget dan sesuai orderan	1
barang sobek main kirim aja,payah lapak ni nyari rizki tp kain lap di kirim	0
Barang sampai dengan sempurna tanpa kerusakan sedikitpun	1
Barang sesuai foto. Tidak mengcewakan. Pengiriman tepat waktu. Oke semua	1
Barang sesuai pesanan, packing rapih, pengiriman cuma ke tunda libur lebaran aja. Makasih gan barangnya bagus nyaman di pakai	1
barang rusak tidak bisa dipakai nelpon & sms	0
Barang sesuai dengan pesanan dan pelayanannya memuaskan	1
Baru sampe saya buka kabelnya udah putus. Udah komplain saya tunggu 24 jam gak ada respon. Tipe penjual yang lepas tanggung jawab, cuma tau untung.	0

3.3.2 Preprocessing Data

Pada tahapan ini teks pada kalimat akan diseragamkan agar data bisa dilanjutkan ke tahap selanjutnya. *Text processing* meliputi *case folding*, *tokenizing*, *filtering*, *stemming* dan TF-IDF.



Gambar 3. 2 Preprocessing Data

1. *Case folding,* tujuan dari *case folding* yaitu mengubah semua huruf menjadi huruf kecil dan menghilangkan tanda baca seperti koma, titik dan tanda baca lainya.

   Input ⟶ packing lama !! Chat gak di bales bales !! tapi lumayan bagus hasilnya gan :^_ ^

   Output ⟶ packing lama chat gak di bales bales tapi lumayan bagus hasilnya gan

2. *Tokeniizing*, pada tahapan ini bertujuan untuk menyusun setiap kata dan menjadikanya sebagai sebuah token agar bisa diproses pada tahap selanjutnya.

   Input ⟶ packing lama chat gak di bales bales tapi lumayan bagus hasilnya gan

   Output ⟶ ‘packing’, ‘lama’, ‘chat’, ‘gak’, ‘di’, ‘bales’, ‘bales’, ‘tapi’, ‘lumayan’, ‘bagus’, ‘hasilnya’, ‘gan’

3. *Filtering,* kata-kata yang kurang relevan atau tidak bermakna akan dihilangkan .

   Input ⟶ packing lama chat gak di bales bales tapi lumayan bagus hasilnya gan

   Output ⟶ packing lama chat gak di bales bales tapi lumayan bagus hasilnya

4. *Stemming*, pada tahapan ini bertujuan untuk mencari kata dasar, algoritma yang dipakai yaitu algoritma Nazief dan Adriani.

   Input ⟶ packing lama chat gak di bales bales tapi lumayan bagus hasilnya

   Output ⟶ lama bales lumayan bagus hasil

5. Langkah selanjutnya yaitu mencari frekuensi kata yang banyak digunakan. Setelah frekuensi kata didapatkan tahap selanjutnya adalah memberikan bobot pada setiap kata yang ada menggunakan rumus TF-IDF. Rumus TF-IDF sebagai berikut :

$$TF\ (t) = f_{t,d} / \Sigma\ t,d$$

Dimana $f_{t,d}$ merupakan frekuensi sebuah kata muncul didalam dokumen, sedangkan Σ t,d merupakan total keseluruhan kata yang terdapat didalam dokumen.

Kemudian menghitung IDF (*Inverse Document Frequency*) dari

sebuah kata di dalam sekumpulan dokumen menggunakan persamaan :

$$IDF(t) = \log(|D|/ f_{t,D})$$

|D| merupakan jumlah dokumen yang ada didalam koleksi, sedangkan $f_{t,D}$ merupakan merupakan jumlah dokumen dimana t muncul didalam D.

Setelah TF dan IDF nilainya didapatkan selanjutnya yaitu menghitung TF-IDF dengan persamaan sebagai berikut :

$$TF\text{-}IDF = TF(t) * IDF(t)$$

3.3.3 Evaluasi

Evaluasi dilakukan untuk mengetahui kinerja algortitma yang dipakai, metode yang dipakai untuk evaluasi adalah *confusion matrix*. Setelah *confusion matrix* dilakukan maka nilai bisa dilihat dan dihitung nilai *accuration*, *precision, recall*, dan *f1-score*. Tabel *confusion matrix* dapat dilihat sebagai berikut.

Tabel 3. 3 Confusion Matrix

		Class Predicted	
		1	0
Actual Class	1	True Positive	False Positive
	0	False Negative	True Negatif

Keterangan dari tabel diatas dinyatakan sebagai berikut :

1. *True Positive* (TP), yaitu jumlah dari kelas Positif yang benar dan diklasifikasikan sebagai kelas Positif.
2. *True Negative* (TN), yaitu jumlah dari kelas Negatif yang benar dan diklasifikasikan sebagai kelas Negatif.
3. *False Positive* (FP), yaitu jumlah dari kelas Negatif yang salah dan diklasifikasikan sebagai kelas Positif.
4. *False Negative* (FN), yaitu jumlah dari kelas Positif yang salah dan

diklasifikasikan sebagai kelas Negatif.

Untuk melihat peforma dari algoritma adalah menghitung akurasi dengan persamaan sebagai berikut :

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} * 100\%$$

Selanjutnya menghitung nilai *precision* yang bertujuan untuk melakukan perbandingan antara *true positive* dengan banyaknya data yang diprediksi *positive*. Persamaanya dapat dilihat sebagai berikut :

$$\text{Precission} = \frac{TP}{TP+FP} * 100\%$$

Selanjutnya menghitung nilai *recall* yang bertujuan untuk melakukan perbandingan antara *true positive* dengan banyaknya data yang sebenarnya p*ositive.* Persamaanya dapat dilihat sebagai berikut :

$$\text{Recall} = \frac{TP}{TP+FN} * 100\%$$

Selanjutnya menghitung nilai *F1-score* yang bertujuan untuk melihat parameter tunggal keberhasilan *retrieval*. Persamaanya sebagai berikut :

$$\text{F1-Score} = 2 * \frac{\text{Precission} * \text{Recall}}{\text{Precission} + \text{Recall}}$$

UPB KARAWANG