

## BAB III METODE PENELITIAN

### 3.1 Objek Penelitian

Objek pada penelitian tugas akhir ini adalah produk motor *matic* Honda Beat dan *Scoopy* adapun sumber data yang diambil adalah dari media sosial *Twitter* dan perancangannya dilaksanakan di Universitas Buana Perjuangan Karawang. Lokasi ini diambil karena memiliki berbagai aspek yang mendukung.

Tabel 3.1 Jadwal Penelitian

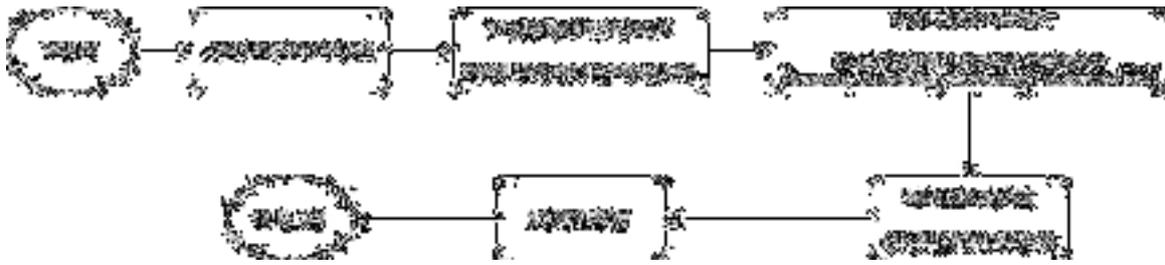
No	Kegiatan	Bulan					
		1	2	3	4	5	6
1	Analisis Masalah	■					
2	Pengumpulan Data		■	■	■		
3	<i>Text Processing</i>			■	■	■	
4	Implementasi					■	
5	Pengujian						■

### 3.2 Peralatan Penelitian

Penelitian ini tentunya membutuhkan peralatan yang dapat membantu proses penelitian, adapun alat yang digunakan adalah perangkat keras dengan spesifikasi Processor Intel Core i5-7200, 1 TB HDD, 8192 RAM. Kemudian perangkat lunak yang digunakan adalah RapidMiner Studio sebagai alat yang digunakan untuk membantu mengambil data dari *Twitter*, *Twitter* sebagai media utama pencarian data, Google Chrome yang digunakan untuk browsing serta mencari referensi jurnal maupun artikel. Selanjutnya Microsoft Office 2013 Sebagai software yang membantu membuat laporan dan proposal.

### 3.3 Prosedur Penelitian

Adapun prosedur penelitian yang digambarkan dengan menggunakan bentuk *Flowchart* dalam Gambar 3.1. Penelitian ini dimulai dengan analisis masalah lalu dilanjutkan ke tahap pengumpulan data atau *crawling data* di *Twitter* menggunakan RapidMiner Studio. Data diambil dari bulan November 2021 sampai Maret 2022. Kemudian data yang sudah didapatkan akan diproses pada bagian pemrosesan data. Setelah itu data akan diklasifikasikan menjadi tiga kategori yaitu positif, netral dan negatif, data yang sudah diklasifikasi akan implementasikan kepada algoritma kemudian dilanjutkan ke tahap pengujian untuk mendapatkan hasil akurasi terbaik.



Gambar 3.1. Prosedur Penelitian

### 3.3.1 Analisis Masalah

Pada tahap ini, proses pencarian permasalahan data dilakukan dengan tujuan untuk mengetahui fakta data yang sebenarnya. Cara menentukan fakta data sentimen yang bernilai positif, netral dan negatif dapat digambarkan sebagai berikut :

1. Sentimen positif dapat diartikan sebagai suatu tindakan atau tanggapan yang dilakukan dengan tidak adanya penolakan atau bersifat mendukung.
2. Sentimen netral dapat diartikan sebagai suatu tindakan atau tanggapan yang tidak memihak kepada positif ataupun negatif, bisa disimpulkan bahwa sentimen netral yaitu tidak adanya sifat mendukung atau tidak mendukung.
3. Sentimen negatif dapat diartikan sebagai suatu tindakan atau tanggapan yang dilakukan dengan adanya penolakan, atau bisa disimpulkan bahwa sentimen negatif yaitu bersifat kurang baik atau tidak mendukung.

Tahapan dalam proses analisis masalah diantaranya mengumpulkan data, identifikasi sentimen, menentukan metode yang digunakan dan menafsirkan hasil dari analisis. Selain itu analisis masalah digunakan untuk mendapatkan kesimpulan dari data yang telah di analisis berdasarkan hasil pengujian.

### 3.3.2 Pengumpulan Data

Pengumpulan data dari penelitian ini dibagi menjadi dua bagian yaitu :

1. Studi Literatur

Studi literatur dilakukan untuk pencarian landasan teori dan data dari berbagai jurnal, buku, dan juga internet seperti definisi menggunakan aplikasi web browser, contoh penelitian terkait, metode, data mining, analisis sentimen dan data yang berkaitan dengan penelitian ini.

2. *Crawling Data*

*Crawling Data* juga dilakukan dengan cara melakukan pengambilan data yang ada pada *Twitter* menggunakan aplikasi RapidMiner Studio, dimana data tersebut diambil dari tanggapan masyarakat di *Twitter* terhadap produk motor

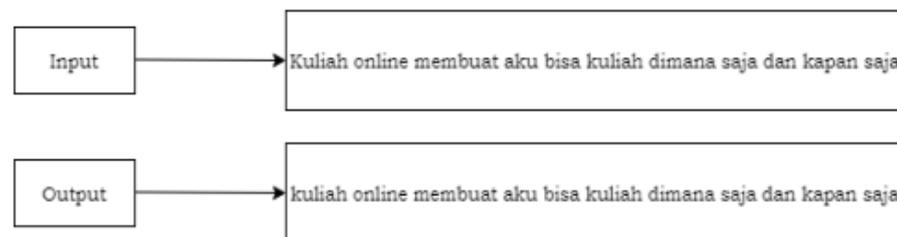
*matic* Honda Beat dan *Scoopy* sejak bulan November 2021 sampai maret 2022. Data yang diambil dari *Twitter* sebanyak 205 data yang dibagi menjadi dua, 141 data dari motor Honda Beat dan 64 data dari motor Honda *Scoopy*. Hasil yang didapat nantinya berupa informasi-informasi yang berkaitan dengan penelitian seperti tanggapan terhadap Motor Beat dan *Scoopy*, kelebihan, kekurangan dan saran yang nantinya masuk kedalam kategori positif, netral dan negatif.

### 3.3.3 *Text Processing*

Sebelum dilakukannya proses klasifikasi, teks dokumen atau data harus disiapkan terlebih dahulu, proses tersebut biasanya dinamakan dengan *Text Processing*. Tahapan *Text Processing* ini berguna agar data teks yang masih terdapat banyak *noise* atau tidak terstruktur menjadi terstruktur. *Text Processing* memiliki beberapa tahapan yaitu *Case Folding*, *Tokenizing*, menghapus angka, *Filtering* dan TF IDF.

#### 1. *Case Folding*

*Case Folding* adalah sebuah proses untuk merubah huruf kapital menjadi huruf biasa atau standar. Proses ini dilakukan dengan tujuan mempermudah pencarian, dikarenakan tidak semua dokumen teks konsisten dengan huruf kapital. Berikut adalah contoh *Case Folding* :



Gambar 3.2 Contoh *Case Folding*

#### 2. *Cleaning*

*Cleaning* adalah proses untuk membersihkan data yang tidak diperlukan dari hasil *crawling* data.

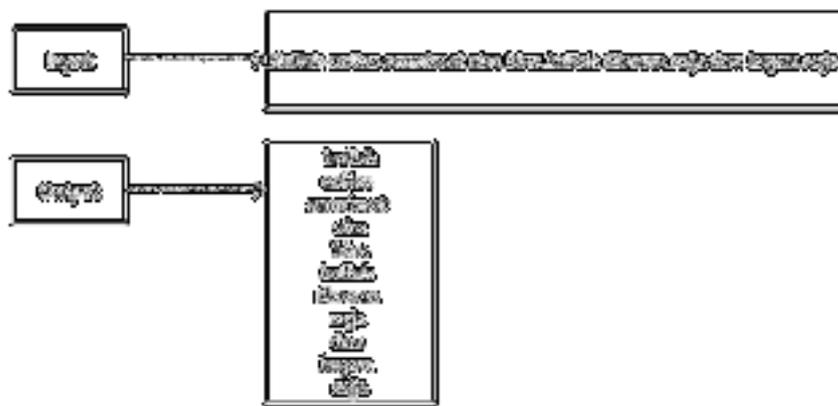
#### 3. *Stopwords Removal*

*Stopwords Removal* adalah proses membuang kata-kata kurang penting atau menyimpang dari kosa kata. *Stopwords* adalah proses pembuangan *term* yang tidak memiliki arti atau tidak relevan. *Term* yang diperoleh akan dicek dalam suatu daftar *stopword*, apabila sebuah kata masuk di dalam daftar *stopword* maka kata tersebut tidak akan diproses lebih lanjut. Maksud

penggunaan *stopwords* adalah dengan menghilangkan kata-kata informasi rendah dari teks, kita dapat fokus pada kata-kata penting.

#### 4. *Tokenizing*

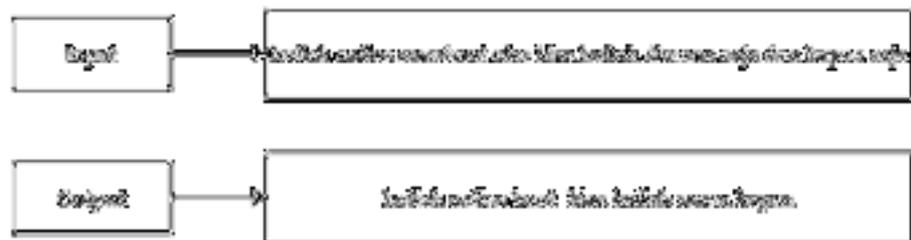
*Tokenizing* adalah sebuah proses untuk memecah kalimat menjadi kata – kata yang akan menjadikan kalimat lebih bermakna. Ada beberapa tahapan pada proses ini, yang pertama adalah melakukan normalisasi kata dengan mengubah semua karakter huruf menjadi huruf kecil atau *Case Folding*. Tahap kedua adalah dengan melakukan penguraian teks yang semula berupa kalimat yang berisi kata – kata. Berikut adalah contoh *Tokenizing* :



Gambar 3.3 Contoh *Tokenizing*

#### 5. *Stemming*

*Stemming* adalah tahapan untuk memperkecil jumlah indeks yang berbeda dari satu data sehingga kata yang memiliki suffix maupun prefix akan Kembali ke bentuk dasarnya. Berikut adalah contoh dari *Stemming* :



Gambar 3.4 Contoh *Stemming*

#### 6. Pelabelan Data

Pelabelan setiap data dilakukan secara manual, label yang diberikan ada 3 kelas, yaitu kelas positif, kelas negatif, dan kelas netral. (Khaira et al., 2020). Pelabelan data yang digunakan dalam penelitian ini sebanyak 205 data, data tersebut dibagi menjadi 2 kategori yaitu 141 data untuk Motor *Matic* Honda

Beat dan 64 data untuk Motor Matic Honda Scoopy. Kemudian data tersebut diberikan label secara manual untuk menentukan sentiment positif, netral atau negatif.

### 3.3.4 Implementasi

Setelah melakukan tahapan *text processing* dan pelabelan data, kemudian data akan diimplementasikan dengan tahapan sebagai berikut :

#### 1. TF-IDF

*Term Frequency - Inverse Document Frequency* atau TF-IDF merupakan metode algoritma yang umum digunakan untuk menghitung bobot setiap kata. Metode ini juga dikenal efisien, sederhana dan akurat. Metode ini akan menghitung *term frequency (TF)* dan *inverse document frequency (IDF)*. Selain itu TF IDF juga dapat mengetahui kata apa yang sering muncul pada sebuah dokumen. Berikut adalah contoh TF IDF :

anies	0.047619047619047616
penting	0.047619047619047616
masyarakat	0.047619047619047616

Gambar 3.5 Contoh TF IDF

#### 2. Klasifikasi *Naive Bayes*

Setelah dilakukan proses TF-IDF atau pembobotan kata, kemudian data akan di klasifikasi dengan menggunakan algoritma *naive bayes*. Klasifikasi merupakan metode untuk mengelompokkan dan menemukan *itemset* sesuai dengan nilai bobot yang ada dari setiap *itemset*. Dalam penelitian ini data yang dikelompokkan dibagi menjadi tiga bagian yaitu positif, netral dan negatif.

### 3.3.5 Pengujian

Metode yang digunakan dalam pengujian adalah *Confusion Matrix*. *Confusion Matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pengujian pada penelitian ini digunakan untuk mengevaluasi performa algoritma yang digunakan. Proses evaluasi dimulai dengan matrik konfusi untuk menilai akurasi algoritma. Kemudian dilakukan pengujian algoritma menggunakan metode *Confusion Matrix*. Metode ini cukup membantu untuk melakukan proses analisis kualitas *Classifier*.

Manual	Sistem		
	Positif	Netral	Negatif
Positif	True Positive (TP)	False Netral (FNt)	False Negative (FN)
Netral	False Positive (FP)	True Netral (TNt)	False Negative (FN)
Negatif	False Positive (FP)	False Netral (FNt)	True Negative (TN)

Pengujian dilakukan dengan menghitung nilai *accuracy*, *precision* dan *recall* dan ditampilkan dalam bentuk persentase.

1. *Accuracy*

Akurasi merupakan persentase dari total sentimen yang benar dikenali. Untuk menghitung nilai akurasi digunakan persamaan dibawah ini :

$$\text{Akurasi} = \frac{TP+TNt+TN}{TP+TN+TNt+FP+FNt+FN} \times 100\% \dots\dots\dots(3.1)$$

2. *Precision*

*Precision* merupakan perbandingan data relevan yang ditemukan terhadap jumlah data ditemukan. Untuk menghitung nilai *precision* digunakan persamaan dibawah ini :

$$\text{Precision} = \frac{TP}{FP+TP} * 100\% \dots\dots\dots(3.2)$$

3. *Recall*

*Recall* merupakan perbandingan jumlah materi relevan yang ditemukan terhadap jumlah materi relevan. Untuk menghitung nilai *recall* digunakan persamaan dibawah ini :

$$\text{Recall} = \frac{TP}{FN+FNt+TP} * 100\% \dots\dots\dots(3.3)$$