

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Subyek dan Obyek Penelitian

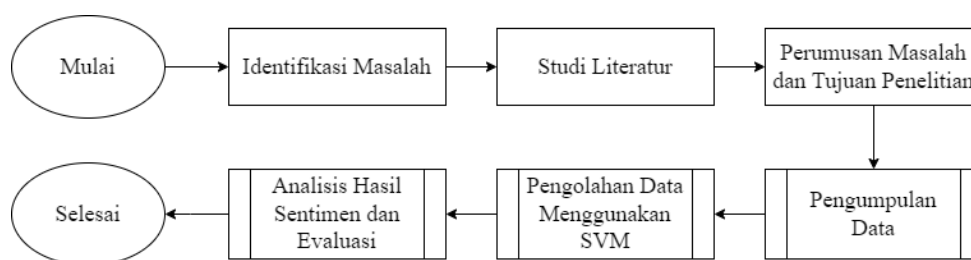
Subyek dari penelitian adalah media dari diambilnya dataset. Dataset pada penelitian ini diambil dari tagar atau *hashtag* “Percuma Lapor Polisi” di media sosial *Twitter*, jadi subyek dari penelitian ini adalah *hashtag*

“Percuma Lapor Polisi” pada *Twitter*. Obyek penelitian adalah fokus data yang akan diteliti atau dataset penelitian. Untuk penelitian ini, yang ditetapkan sebagai obyek penelitian sekaligus dataset adalah *Tweet* dengan *hashtag* “Percuma Lapor Polisi”.

#### 3.2 Alat dan Bahan Penelitian

Terdapat *software* sebagai alat pada penelitian ini berupa *Microsoft Word 2019*, *Jupyter Notebook*, dan *Python*. *Microsoft word 2019* digunakan untuk penulisan naskah tugas akhir. *Jupyter Notebook* digunakan untuk menulis *code* yang diperlukan untuk pengolahan data *Tweet*. *Python* untuk pemrosesan dan analisa data yang diperlukan. Data *Tweet* dengan *hashtag* “Percuma Lapor Polisi” pada *Twitter* digunakan sebagai bahan penelitian.

#### 3.3 Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian

Berdasarkan Gambar 3.1 maka alur penelitian dapat dijelaskan seperti di bawah ini:

##### 3.3.1 Identifikasi Masalah

Identifikasi masalah ialah proses pertama dimana peneliti mencari permasalahan yang ada di sekitar. Masalah tersebut selanjutnya diangkat

menjadi topik penelitian. Pencarian topik penelitian dengan mengangkat suatu masalah bertujuan untuk menemukan solusi yang tepat dan dapat diterapkan dalam masalah tersebut. Pada penelitian ini, masalah yang diidentifikasi ialah keresahan masyarakat dengan instansi kepolisian pada *trending topics* di *Twitter* dengan *hashtag* “Percuma Lapor Polisi”. *Tweet* yang pastinya mengandung opini dapat berupa positif ataupun negatif, maka dari itu dilakukanlah analisis sentimen.

### **3.3.2 Studi Literatur**

Studi literatur diterapkan setelah peneliti mengetahui masalah yang hendak diangkat menjadi topik penelitian untuk mencari metode yang tepat. Studi literatur yang dilakukan ialah mempelajari jurnal, artikel, buku, dan skripsi dengan rentang waktu 2016 – 2022 atau dengan kata lain rentang waktu lima tahun kebelakang.

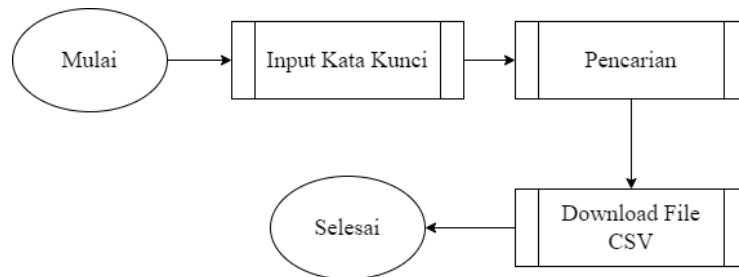
### **3.3.3 Perumusan Masalah dan Tujuan Penelitian**

Langkah selanjutnya peneliti menetapkan metode SVM sebagai metode untuk menyelesaikan analisis sentimen pada penelitian ini, peneliti merumuskan masalah dan tujuan penelitian. Rumusan masalah diterapkan untuk mempersempit batasan yang sudah akan diteliti agar penelitian lebih terfokus. Tujuan penelitian ditetapkan agar penelitian lebih terarah.

### **3.3.4 Pengumpulan Data**

Pengumpulan data diterapkan dengan cara *crawling* data pada *jupyter notebook* dengan bahasa pemrograman *python*. Data yang dikumpulkan adalah *Tweet* dengan *hashtag* “Percuma Lapor Polisi” pada tanggal 11 hingga 17 November 2021. *Tweet* yang berhasil terkumpul sebanyak 3.533 data *Tweet*. Proses pertama yang pengumpulan data yaitu menuliskan *code* dengan bantuan *library tweepy* pada bahasa pemrograman *python* di dalam *Jupyter Notebook*. Pencarian dilakukan dengan memasukkan kata kunci, token pada API *twitter*, dan rentang waktu yang ditentukan. Selanjutnya pencarian dimulai dan memunculkan *Tweet* yang relevan dengan kata kunci, lalu *download* data dengan format *Comma Separated Values* (CSV) dan data berhasil dikumpulkan [29]. Langkah

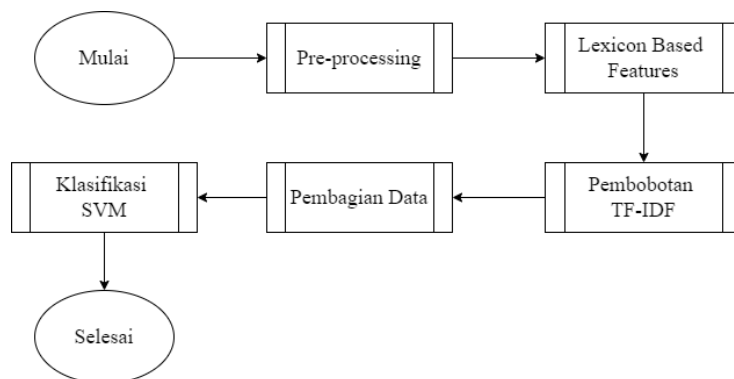
pengumpulan data yang lebih runtut dipaparkan dalam Gambar 3.2 Alur Pengumpulan Data.



Gambar 3.2 Alur Pengumpulan Data

### 3.3.5 Pengolahan Data Menggunakan SVM

Adapun proses dari klasifikasi SVM setelah berhasil melakukan pengumpulan data adalah *pre-processing*, pelabelan sentimen positif dan negatif menggunakan *lexicon based features* dua kelas data yang ditentukan yaitu, data latih dan data uji, pembobotan menggunakan TF-IDF, klasifikasi menggunakan SVM. Gambar 3.3 Alur Pengolahan Data di bawah ini menunjukkan alur metode dari pengolahan data menggunakan SVM.

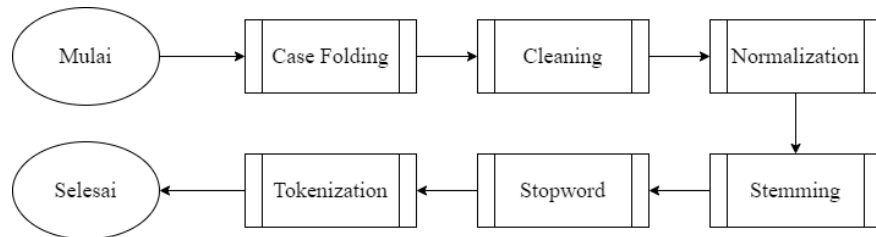


Gambar 3.3 Alur Pengolahan Data

#### 3.3.5.1 Pre-processing

*Pre-processing* ialah proses pembersihan data dari hal-hal yang tidak diperlukan. Contoh hal yang tidak diperlukan dalam kasus ini adalah simbol dan tanda baca. *Pre-processing* diperlukan untuk mengubah data teks yang mana dalam kasus ini data teks diambil dari *Tweet*, data teks yang sebelumnya adalah data tidak

terstruktur diubah menjadi data terstruktur [18]. Adapun proses dari *pre-processing* seperti pada Gambar 3.4 Alur *Pre-processing* di bawah ini [12].



Gambar 3.4 Alur *Pre-processing*

Pada Gambar 3.4 Alur *Pre-processing* dapat dijelaskan seperti berikut:

1. *Case Folding filtering*

Pada proses ini dilakukan penyamarataan pada huruf dengan cara merubah format huruf menjadi huruf kecil [12].

2. *Cleaning*

Proses *Cleaning* ialah menghapus elemen yang tidak diperlukan seperti tanda baca, hashtag, URL, *username* [12].

3. *Normalization*

*Normalization* adalah proses untuk mengubah kata agar sesuai dengan kaidah KBBI. dalam proses penelitian ini menggunakan penelitian milik Salsabila, Nikmatun Aliyah dan kawan-kawan pada tahun 2018. Penelitian mereka membentuk kamus *out-of-vocabulary* (OOV) berbahasa Indonesia yang dapat digunakan untuk sebagian besar kasus [13].

4. *Stemming*

Proses ini akan mengembalikan setiap kata dalam dokumen ke dalam kata dasar. Bahasa Indonesia sangatlah kompleks, untuk mengubah ke dalam kata

dasar dan penghapusan imbuhan pada kata. Saat ini sudah ada *library* dalam *python* untuk proses *stemming* berbahasa Indonesia bernama Sastrawi [12].

#### 5. *Stopword Removal*

Penghapusan kata yang tidak mengandung sentimen dilakukan pada proses ini. *Stopword* sendiri memerlukan *database* yang berisi daftar kata-kata tidak mengandung sentimen. Beberapa peneliti sudah melakukan pengumpulan daftar kata *stopword*, ialah Fadhillah Z. Talla, Damian Doyle dan Yudi Wibisono [33]. Pada penelitian ini menggunakan daftar *stopword* milik Talla yang dilakukan saat beliau menyelesaikan studi masternya. Tesis beliau dapat diunduh siapa saja secara mudah untuk selanjutnya dijadikan pijakan ilmiah [34].

#### 6. *Tokenization*

Proses ini memisahkan kalimat menjadi kata demi kata [12].

### 3.3.5.2 *Lexicon Based Features*

Pelabelan data dilakukan menggunakan *lexicon*, cara kerja pelabelan *lexicon* yaitu dengan penilaian bobot positif atau negatif pada setiap kata yang muncul berdasarkan kamus *lexicon*, dimana pada kamus *lexicon* terdapat kata-kata dengan sentimen positif dan negatif. Penggunaan pelabelan *lexicon* membutuhkan kamus yang dijadikan sebagai acuan untuk menghitung polaritas opini atau sentimen. Pada penelitian ini kamus yang digunakan berasal dari penelitian [13]. Kamus ini memiliki kamus kecil *lexicon* positif dan negatif, serta kamus kecil berisi kata-kata umpatan dan kamus kecil lainnya. Semua kamus kecil akan digabungkan menjadi kamus *lexicon* utama untuk proses pelabelan data.

Proses pelabelan data dimulai dengan mengecek kata sentimen atau opini pada setiap data *Tweet*, kemudian dihitung

bobotnya berdasarkan kamus *lexicon*. Nilai hasil dari bobot pada tiap kata dalam *Tweet* kemudian diakumulasikan. Apabila nilai akumulasi pada *Tweet* bernilai  $0 (> 0)$ , maka label dari *Tweet* bernilai positif. Jika hasil nilai akumulasi pada *Tweet* bernilai kurang dari sama dengan  $0 (\leq 0)$ , maka label dari *Tweet* bernilai negatif [21]. Tabel 3.2 di bawah merupakan contoh dari kamus *lexicon*.

Tabel 3.2 Kamus *Lexicon*

<i>word</i>	<i>weight</i>	<i>number_of_words</i>
hai	3	1
merekam	2	1
ekstensif	3	1
paripurna	1	1
detail	2	1
pernik	3	1
belas	2	1
welas	4	1
kabung	1	1

### 3.3.5.3 Pembobotan dengan TF-IDF

Pembobotan kata dilakukan setelah data ya.ng berupa kata melewati proses *pre-processing*. Pembobotan dilakukan dengan TF-IDF yang memiliki kepanjangan sebagai *Term Frequency-Inverse Document Frequency*. TF ialah banyaknya kemunculan (frekuensi) suatu kata (*term*) dalam suatu dokumen. Sedangkan *Document Frequency* (DF) ialah frekuensi dari jumlah dokumen yang memiliki kemunculan kata (*term*) tersebut [27]. Proses TF-IDF menggunakan *library sklearn*.

### 3.3.5.4 Pembagian Data

Pembagian data adalah proses pemecahan data menjadi dua kelas data. Rasio yang digunakan yaitu perbandingan 80:20. Pembagian data menjadi 80% data latih dan 20% data uji [35]. Data latih berguna sebagai pembelajaran untuk sistem dan akan dijadikan

sebagai acuan oleh sistem, sedangkan data uji berguna sebagai pengetahuan letak kelas suatu data [18]. Pembagian data akan menggunakan *library sklearn*.

#### **3.3.5.5 Klasifikasi dengan SVM**

Klasifikasi dilakukan setelah pembagian dua kelas data. Penelitian ini melakukan klasifikasi SVM dengan *library sklearn*.

#### **3.3.6 Analisis dan Evaluasi Hasil Sentimen**

Evaluasi adalah proses menguji keakuratan klasifikasi yang sudah diimplementasikan dengan data latih. Proses selanjutnya adalah hasil dari dua dataset yang berbeda yaitu data latih dan data uji dibandingkan untuk mendapatkan nilai *accuracy*, *precision recall*, dan *f1-score* [18]. Metode tahap evaluasi ini adalah *confusion matrix* [32].