



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Objek Penelitian

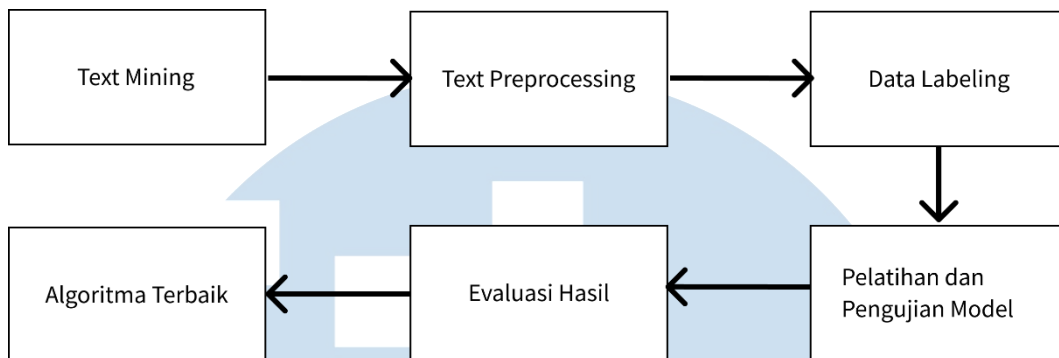
Objek yang diteliti pada penelitian ini adalah cuitan di *Twitter* berbahasa Indonesia mengenai ulasan palsu terhadap toko online. Hasil cuitan tersebut akan diproses sehingga menghasilkan cuitan tersebut termasuk ke kelas asli atau palsu dan positif atau negatif serta mendapatkan akurasi dari klasifikasi algoritma untuk mengetahui algoritma apa yang menghasilkan akurasi yang lebih tinggi,

#### 3.2 Metode Penelitian

Metode yang digunakan pada penelitian ini berdasarkan perbandingan metode-metode lain yang serupa, yang bertujuan untuk menemukan metode yang paling tepat untuk penelitian ini.

Pada penelitian ini dilakukan perbandingan antara dua buah metode penyelesaian masalah yaitu algoritma naïve bayes dan SVM (support vector machine) untuk menemukan akurasi terbaik dari kedua metode penelitian yang digunakan. Kedua metode tersebut paling sering digunakan dalam penelitian analisis sentimen dan menghasilkan tingkat akurasi yang besar. Kerangka pikir pada penelitian ini divisualisasikan pada gambar berikut:

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



Gambar 3. 1 Kerangka Pikir

Gambar kerangka pikir ini menggambarkan alur dari kerangka pikir yang telah digunakan pada penelitian terdahulu yang sudah disesuaikan untuk penelitian ini. Berikut adalah penjabaran langkah-langkah penelitian yang ada didalam kerangka pikir penelitian:

### 3.2.1 *Text Mining*

Tahap *text mining* ini data yang diambil adalah data *tweet* ulasan terhadap toko online di media sosial *Twitter*. Pengambilan data ini menggunakan aplikasi *Rapidminer Studio* menggunakan kata kunci “Tokopedia”, “Shopee”, dan “Bukalapak”. Sebelum data disimpan, akan dilakukan proses penghapusan data yang terduplikat terlebih dahulu, lalu data akan disimpan ke file excel (.xls)

### 3.2.2 *Text Preprocessing*

Tahapan *text preprocessing* merupakan lanjutan dari tahap penambangan data. Hasil penambangan data tentunya belum rapi dan tidak terstruktur. Maka dari itu dibutuhkan tahap *text preprocessing* dimana tahap ini akan mengolah informasi yang kotor tadi, akan melalui serangkaian

proses seperti berikut untuk menghasilkan data yang dapat digunakan untuk penelitian. Berikut adalah penjabaran dari tahap *text preprocessing*:

### **3.2.3.1 Data Cleansing**

Tahap *data cleansing* adalah tahap pembersihan data *tweets*.

Berikut adalah tahap membersihkan data *tweet*:

1. Penghapusan URL yang terdapat dalam *tweets*
2. Menghapus isi *tweets* yang bukan alfabet
3. Menghapus kata “RT” (*re-tweet*) yang terdapat di dalam *tweets*
4. Menghapus *tweets* yang berulang

### **3.2.3.2 Case Folding**

Tahap *Case Folding* adalah tahap mengubah semua huruf besar pada *tweets* menjadi huruf kecil.

### **3.2.3.3 Tokenizing**

Membuat tokenisasi kata dari *tweets* dan pada tahap tokenizing ini juga akan dilakukan penghapusan tanda baca, karena tanda baca akan mengganggu proses perhitungan dalam algoritma yang akan diterapkan

### **3.2.3.4 Filter Stopword**

Proses *filter stopwords* adalah proses penghilangan kata kata umum yang sering muncul. Pada penelitian ini *filter stopwords* yang digunakan adalah *filter stopwords* bahasa Indonesia. Proses ini

dilakukan untuk mengurangi kata-kata yang tidak diperlukan dalam proses labeling data dan uji akurasi data.

#### **3.2.3.5 *N-Gram Character***

Proses *N-Gram Character* adalah pembelajaran pola dari pemecahan kalimat *tweet* ini, dari pola yang ada dapat menaikkan tingkat akurasi pada klasifikasi. Pada data *tweet* ini, parameter N-Gram menggunakan pola setiap tiga kata dipecah dari kalimat.

#### **3.2.3.6 *TF-IDF***

Memberikan pembobotan terhadap kata-kata *tweets* menggunakan metode TF-IDF. Contohnya “Saya suka shopee”. Yang paling sering keluar. Misalnya saya suka. Ini lebih penting karena kata tersebut sering muncul yang dapat membantu pertimbangan dalam modeling.

### **3.2.3 *Data Labelling***

Melakukan penyeleksian *tweets* secara manual pada *tweets* yang tidak memiliki arti agar proses klasifikasi lebih baik, kemudian memberikan label pada setiap *tweets* dengan memberikan kategori sentimen

### **3.2.4 *Pelatihan dan Pengujian Model***

Data hasil dari *text preprocessing* dibagi menjadi dua untuk *data training* dan *data testing* menggunakan metode *K-fold Cross Validation*. Melatih model algoritma dengan data pelatihan untuk membentuk model, lalu

menerapkan hasil model pelatihan dengan data pengujian. Hasil dari pengujian setiap algoritma disini berupa hasil *confusion matrix*.

### **3.2.5 Evaluasi Hasil**

Data hasil pengujian dari setiap algoritma akan dilakukan seleksi hasil pengujian terbaik pada setiap algoritma untuk dijadikan perbandingan akurasi antara dua algoritma yang digunakan yaitu *Naive Bayes* dan *Support Vector Machine*.

## **3.3 Variabel penelitian**

### **3.3.1 Variabel Independen**

Variabel Independen adalah variabel yang mempengaruhi hubungan antara variabel independen dengan dependen tetapi tidak dapat diukur. Variabel Independen pada penelitian ini adalah *tweet* mengenai ulasan palsu toko online.

### **3.3.2 Variabel Dependen**

Variabel dependen adalah variabel yang menjadi akibat atau dipengaruhi. Variabel dependen pada penelitian ini adalah hasil analisis sentimen dari ulasan palsu pada toko online.

## **3.4 Teknik Pengumpulan Data**

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah data berbentuk teks diambil dari media sosial *Twitter* dengan menggunakan program *Rapidminer Studio*. Data yang dikumpulkan adalah data *tweet* dengan kata kunci

“Tokopedia”, “Shopee”, dan “Bukalapak” pada bulan November 2021. Data yang dikumpulkan akan disimpan ke dalam format excel (.xlsx).

### **3.5 Teknik Pengambilan Sampel**

Pengambilan sampel akan dilakukan seleksi terhadap data *tweets* mengenai ulasan terhadap toko online dan memiliki nilai sentimen berupa *True Positive*, *True Negative*, *Fake positive*, dan *Fake Negative*. Jumlah sampel data yang digunakan dalam analisis sentimen ini adalah sebanyak 1490 *tweets*.

UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA