



### **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. Objek Penelitian



Gambar 3.1. Logo MRT Jakarta

**Sumber** (“MRT Jakarta,” 2019)

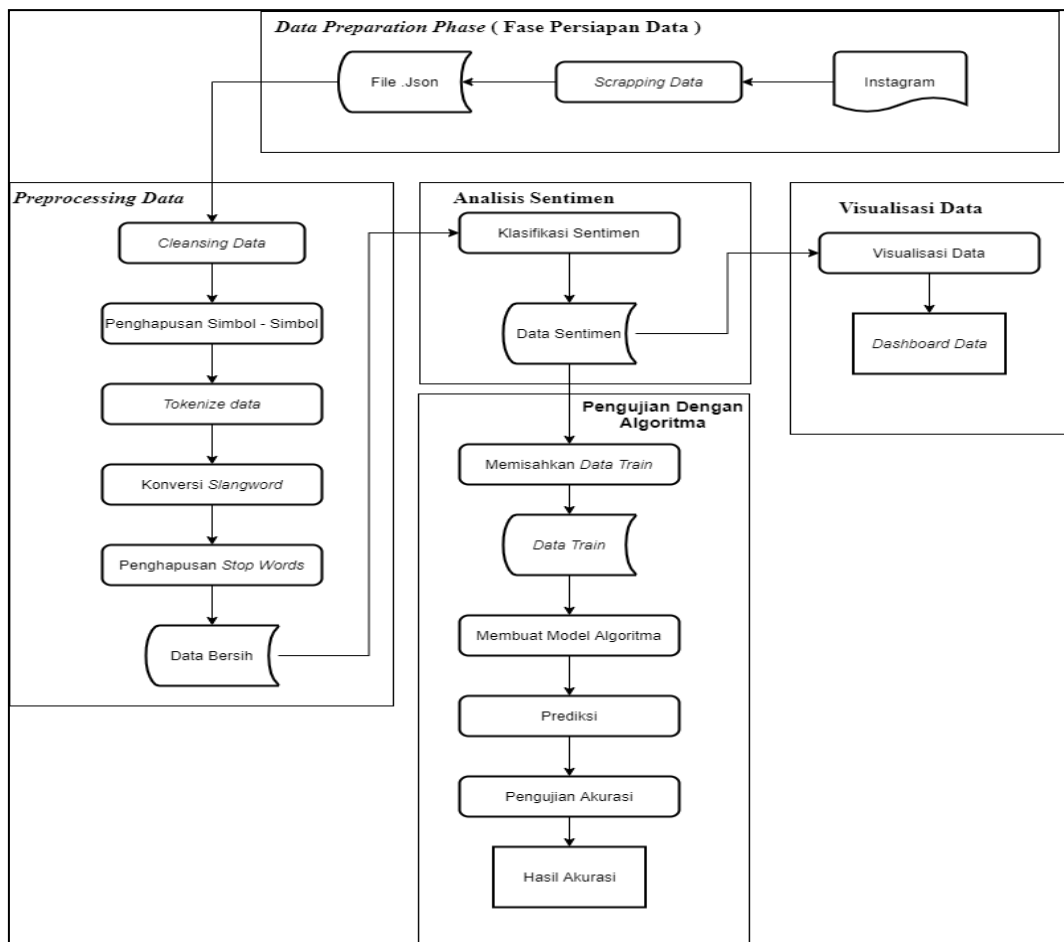
*Mass Rapid Transit* (MRT) merupakan kendaraan transportasi umum yang terbilang sangat baru dibandingkan dengan transportasi umum lainnya seperti kereta, bus, dan transportasi umum lainnya. Dikarenakan merupakan transportasi umum yang baru, pelayanan yang berada di MRT masih ada beberapa kekurangan lainnya, seperti contohnya kurang tegasnya pengawasan terhadap peraturan yang berlaku baik didalam MRT ataupun di stasiun MRT, sehingga ada beberapa orang yang memanfaatkan kelemahan tersebut untuk melanggar peraturan yang berlaku. Maka dari itu pihak PT.MRT Jakarta sendiri membuat media sosial untuk membantu melaporkan apakah ada pelayanan atau permasalahan lainnya yang harus diperbaiki dan ditingkatkan lagi. Salah satu media sosial MRT adalah *Instagram* dengan nama akun @mrtjkt, sehingga jika ada pengguna MRT melihat

masalah terhadap pelayanan dan permasalahan lainnya bisa melaporkan melalui aplikasi *Instagram* tersebut.

### 3.2. Metode Penelitian

#### 2 3.2.1 Alur Penelitian

Pada penelitian ini menggunakan alur penelitian yang digunakan pada penelitian yang dilakukan oleh Ira Zulfa dan Edi Winarko dengan judul penelitian, “Sentimen Analisis Tweet Berbahasa Indonesia dengan *Deep Belief Network*” (Zulfa & Winarko, 2017) dan alur penelitian tersebut diubah untuk memenuhi tujuan pada penelitian ini. Berikut gambar yang merupakan alur penelitian yang digunakan :



Gambar 3.2. Alur Penelitian

### 3.2.1.1 *Data Preparation Phase ( Fase Persiapan Data )*

Pada fase ini yang dilakukan adalah persiapan *data source* atau dataset yang akan digunakan untuk mengukur analisis sentimen. Pada pengambilan *data source* yang dilakukan diaplikasi *Instagram* menggunakan *instagram-scrapers* yang berguna untuk pengambilan *comments* yang ada di post pada *instagram*. Akan tetapi tidak semua data *comments* pada setiap posting diambil, data yang diambil merupakan 500 posting terakhir yang dilakukan oleh akun *@mrtjkt*, yang memiliki periode 30 Desember 2018 hingga 29 Februari 2020 dengan total komen 69492.

### 3.2.1.2 *Preprocessing Data*

Tahap *preprocessing* sangat penting dalam proses sentimen analisis, dan tahap ini juga bisa meningkatkan tingkat akurasi pada algoritma – algoritma yang digunakan (Zulfa & Winarko, 2017). Pada tahap ini yang dilakukan adalah melakukan pembersihan pada data, sehingga data tersebut tidak terdapat *error* ataupun data yang tidak konsisten. Tahap ini terdiri dari beberapa proses diantaranya sebagai berikut :

#### 1. *Cleansing Data*

Pada tahap ini yang dilakukan adalah proses membersihkan data atau *cleansing data* yang bertujuan untuk, agar data tersebut tidak terdapat *error*, *null*, atau data tersebut tidak konsisten. Pada proses membersihkan *data source* aplikasi atau *software* yang digunakan adalah menggunakan *R Studio*. Data yang akan dihapus atau data yang akan dibersihkan adalah data yang tidak berkaitan dengan MRT, seperti

contohnya adalah komen tentang penjualan produk, komen yang berisi hanya *mentions* terhadap seseorang, dan komen yang lainnya yang tidak berkaitan dengan MRT.

## 2. Penghapusan Simbol - Simbol

Pada fase ini yang dilakukan adalah menghapus simbol – simbol, karakter khusus, dan *emoticon* yang diberikan oleh *user instagram*. Penghapusan simbol – simbol karakter khusus, dan *emoticon* yang tidak penting berguna agar nantinya, saat pengolahan data tidak mengalami tingkat *error* yang tinggi. Contoh simbol yang tidak penting adalah koma (,), titik (.), tanda seru (!), dan lain sebagainya. Contoh karakter khusus yang tidak penting adalah angka numerik 0-9, dan karakter lainnya seperti (~,!,@,\$,%,^,&,\*?,\_ ,dll).

## 3. *Tokenize Data*

Pada fase ini yang dilakukan adalah mengubah kalimat – kalimat dari *data source* yang sudah ada menjadi kata – kata. Pada proses ini dilakukan dengan melihat setiap spasi yang ada dikalimat, sehingga kalimat tersebut dapat dipecah menjadi kata – kata yang nantinya akan digunakan untuk mencari sentiment, sehingga pada akhirnya dapat memudahkan dalam mengerjakan proses selanjutnya.

## 4. *Konversi Slangword*

Pada fase ini yang dilakukan adalah mengubah kata yang tidak baku menjadi kata baku dan mengubah kata yang disingkat menjadi kata yang seharusnya seperti contohnya adalah mengubah kata “yg” menjadi kata

“yang”. Pada proses ini menggunakan kamus tambahan yang didapatkan dari website <https://github.com/nasalsabila/kamus-alay>, yang merupakan kamus acuan untuk mengubah kata tidak baku dan mengubah kata yang disingkat. Pada proses ini kata yang sudah ada akan dicocokkan dengan kamus tambahan tersebut, sehingga jika ada kata yang tidak baku dan kata yang disingkat akan diganti atau diubah menjadi kata yang tepat dari kamus tersebut.

#### 5. Penghapusan *Stop Words*

Pada fase ini yang dilakukan adalah menghapus kata penghubung atau kata yang tidak penting, seperti contohnya adalah “dan”, “setelah”, “atau”, dan kata penghubung lainnya, sehingga saat memudahkan untuk tahap memproses analisis sentimen, dan memudahkan saat mencari kata – kata yang sering muncul.

#### 3.2.1.3 Analisis Sentimen

Pada fase ini yang dilakukan adalah mengelompokkan kata – kata serta tanggapan menjadi 3 kategori, yaitu: positif, netral, dan negatif. Pengelompokan kategori tersebut berdasarkan hasil terjemahan dari *Bing Liu Sentiment Lexicon* yang dapat diunduh pada website <https://github.com/masdevit/ID-OpinionWords>, sehingga pengelompokan kata tersebut dapat menggunakan Bahasa Indonesia.

#### 3.2.1.4 Pengujian dengan Algoritma

Pada tahap ini yang dilakukan adalah menguji tingkat akurasi terhadap sentimen yang sudah. Pada proses ini menggunakan algoritma

*Support Vector Machine* sebagai algoritma utamanya, dan algoritma *Naïve Bayes* sebagai algoritma pembanding.

### 3.2.1.5 Visualisasi Data

Pada fase ini yang dilakukan adalah visualisasi data berdasarkan hasil yang sudah dikerjakan pada tahap – tahap sebelumnya. Visualisasi data berupa dashboard agar dapat dengan mudah dipahami oleh orang lain, sehingga informasi yang telah didapatkan dapat dengan mudah dimengerti dan dipahami oleh orang lain.

### 3 3.2.2 Problem Solving

Untuk mendapatkan hasil sesuai klasifikasi yang sudah ada, maka penelitian ini menggunakan algoritma yang bisa mengklasifikasi sesuai kategori yang sudah ditentukan, sehingga hasil dari penelitian ini dapat digunakan untuk mendukung keputusan yang tepat. Berikut adalah perbandingan algoritma *Support Vector Machine*, *Naïve Bayes* dan *Deep Belief Network*.

Tabel 3.1. Perbandingan *Support Vector Machine* dengan *Deep Belief Network* dan *Naïve Bayes*

**Sumber** ( (Zulfa & Winarko, 2017), (Kaviani & Dhotre, 2017), (Saraswati, 2013)

Algoritma	Kelebihan	Kekurangan	Fungsi
<i>Naïve Bayes</i>	<ul style="list-style-type: none"> <li>• Lebih cepat untuk memprediksi dibanding algoritma lainnya.</li> <li>• Mudah digunakan saat menggunakan dataset yang kecil</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Zero conditional probability problem</i> yaitu algoritma ini menghapus seluruh informasi dari probabilitas lainnya.</li> <li>• Pada algoritma ini menggunakan fitur kelas yang <i>independence</i> yang</li> </ul>	<ul style="list-style-type: none"> <li>• Prediksi</li> <li>• Klasifikasi</li> </ul>

		sangat kuat, sehingga sulit mencari data untuk algoritma ini	
Deep Belief Network	<ul style="list-style-type: none"> <li>• Pengembangan dari metode deep learning lainnya.</li> <li>• Memudahkan mencari parameter yang optimal.</li> <li>• Memanfaatkan <i>resources</i> secara optimal</li> </ul>	<ul style="list-style-type: none"> <li>• Memiliki 2 proses, yaitu : tahap <i>training</i> dan tahap klasifikasi, sehingga menyebabkan lamanya proses pengolahan</li> </ul>	<ul style="list-style-type: none"> <li>• Klasifikasi</li> </ul>
Support Vector Machine	<ul style="list-style-type: none"> <li>• Banyak digunakan dalam klasifikasi teks</li> <li>• Memiliki akurasi yang cukup tinggi dalam melakukan analisis sentimen</li> </ul>	<ul style="list-style-type: none"> <li>• Jika pelatihan tidak tepat maka hasil akhir akan cenderung buruk</li> </ul>	<ul style="list-style-type: none"> <li>• Klasifikasi</li> <li>• Prediksi</li> </ul>

Dari perbandingan antara 3 algoritma tersebut, maka penelitian ini menggunakan algoritma *support vector machine* yang digunakan untuk mengklasifikasikan data – data yang telah diambil berdasarkan 3 tipe sentimen yaitu positif, netral, dan negatif, sehingga memudahkan penelitian ini mendapatkan hasil yang baik.

### 3.3. Variabel Penelitian

Variabel yang digunakan untuk penelitian berupa 2 variabel yang berbeda, yaitu variabel *independent* dan *dependent*. Variabel *independent* yang digunakan dalam penelitian ini adalah pelayanan yang berada di MRT seperti contohnya adalah keamanan saat didalam MRT ataupun di stasiun MRT, kebersihan yang berada di stasiun MRT, rute pelayanan yang ada. Sedangkan variabel *dependent* yang digunakan dalam penelitian ini adalah sentimen seluruh pengguna *Instagram* tanpa adanya kategori yang spesifik terhadap pelayanan MRT.



### **3.4. Teknik Pengumpulan Data**

Teknik yang digunakan untuk pengambilan data untuk penelitian ini adalah dengan menggunakan *instagram-scraper* yang didapatkan diwebsite [github.com/rarcega](https://github.com/rarcega). Data tersebut diperoleh berdasarkan komen yang dilakukan oleh pengguna lain terhadap 500 postingan terakhir yang dilakukan pada akun *@mrtjkt*, yaitu 30 Desember 2018 hingga 29 Februari 2020 dengan total komen 69492.

### **3.5. Teknik Pengambilan Sampel**

Teknik pengambilan sampel pada penelitian ini menggunakan teknik *random sampling*. Teknik yang digunakan ini mengambil secara acak sampel dari *data source* yang sudah ada, sehingga data sampel tersebut dapat mewakili populasi.