

## **BAB 3**

### **METODE PENELITIAN**

Penelitian dimulai dengan latar belakang masalah yang ada, memetakan proses, mencari akar penyebab masalah, dan akhirnya merancang dan mengembangkan sistem yang dapat memitigasi atau menghilangkan masalah yang ada. Di bawah ini adalah bahan, alat, metode pengembangan sistem, dan tahap penelitian yang diperlukan untuk penelitian ini.

#### **3.1 BAHAN DAN ALAT PENELITIAN**

Bahan yang akan dibutuhkan dalam penelitian ini adalah data *tweet*, data *retweet* maupun komentar di Twitter yang berkaitan dengan saham sejak Januari 2022 – Mei 2022.

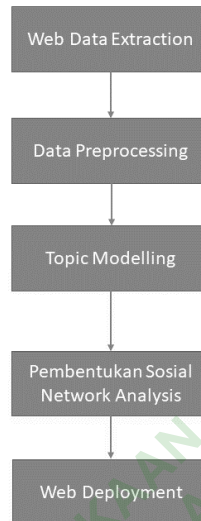
Alat yang digunakan dalam penelitian ini adalah laptop dengan spesifikasi cukup untuk menjalankan proses pengolahan data dan mampu untuk koneksi jaringan internet. Sistem Operasi dan program-program aplikasi yang dipergunakan dalam pengembangan aplikasi ini adalah:

1. Sistem Operasi Windows 10 64 bit.
2. Bahasa Pemrograman Python 3.8.8
3. Microsoft Office Excel 2016
4. Anaconda 3 64-bit
5. Visual Studio Code
6. MongoDB
7. Web framework : Flask
8. Twitter Scraping : snsrapper

#### **3.2 JALAN PENELITIAN**

Penelitian ini menggunakan Bahasa pemrograman Python, Anaconda 3 dan Jupyter Notebook untuk melakukan pengambilan data yang akan disimpan pada Microsoft Office Excel dengan format CSV dan akan dilakukan *preprocessing data*, pemodelan topik, pembentukan *Sosial Network Analysis* pertopik dan kemudian

hasil dari *Sosial Network Analysis* ditampilkan melalui *website dashboard* yang akan dibuat menggunakan bahasa pemrograman Python. Adapun tahapan pada jalan penelitian ini yaitu :



Gambar 3.1 Diagram Jalan Penelitian

### 3.2.1 Web Data Extraction

Pada proses dilakukannya pengambilan data *tweet* dari website Twitter yang berhubungan dengan saham menggunakan library *snscraper* kemudian akan disimpan kedalam file berbentuk CSV.

### 3.2.2 Data Preprocessing

Hasil data *tweet* dari *scrapping* menggunakan library *snscraper* yang sudah terkumpul akan dilakukannya tahapan *text cleaning*, *casefolding*, *tokenizing*, menghilangkan *stopwords* dan penyaringan *Bigram* dan *Trigram*.

1. *Text Cleaning*

*Text cleaning* bertujuan untuk menghilangkan karakter – karakter yang tidak dibutuhkan pada proses pembuatan model. Contohnya karakter – karakter seperti tanda baca.

2. *Casefolding*

*Casefolding* bertujuan untuk menyamakan ratakan besar kecil huruf yang terdapat pada data yang ada.

### 3. *Tokenizing*

*Tokenizing* adalah proses pemotongan kalimat menjadi beberapa bagian kata yang terpisah dan pada saat yang sama menghilangkan karakter tertentu seperti tanda baca, angka, dan karakter non-abjad. Ini karena karakter ini dianggap sebagai pembatas kata dan tidak memengaruhi pemrosesan kata (Rahutomo et al., 2019).

### 4. menghilangkan *stopwords*

*Stopword* adalah beberapa kata yang umum digunakan dan tidak banyak memengaruhi pemrosesan kata. Sebagai contoh *stopword* pada bahasa Indonesia diantaranya : “yang”, “ini”, “dari”, “ke” (Rahutomo et al., 2019)

### 5. Penyaringan *Bigram* dan *Trigram*

Pada penelitian ini dilakukan identifikasi frasa yang mengandung 2 kata misalnya “media sosial” dan frasa yang mengandung 3 kata misalnya “Sekolah menengah atas”, oleh karena itu diperlukan penyaringan *bigram* dan *trigram*

Tabel 3.1 Data Mentah Hasil scrapping

@cjakesim Buat apa emangnya Jake main saham? Mending main monopoli aja gak sih? Bisa keliling dunia gratis tinggal ngocok dadu.
@Orenji1708 @starfess Kata gua si mending gitu, pacaran sama lee soo man, rebut saham SM nyaðŸ~”
@moggynow @shahrol_shiro Ckp kt dia, ada saham dia sbb menfitnah. Lain kalo jgn buat lagi.

Pada **tabel 3.1** merupakan data mentah hasil dari dilakukannya scrapping dengan library *snscraping*, terdapat karakter - karakter tanda baca, link URL, dan juga hastag yang tidak digunakan dalam pembuatan model.

Setelah dilakukannya *preprocessing* data akan terlihat seperti pada **gambar 3.2** adalah data yang sudah siap untuk dilakukan pembuatan model.

```

[['jake',
 'main',
 'saham',
 'mending',
 'main',
 'monopoli',
 'keliling',
 'dunia',
 'gratis',
 'tinggal',
 'kocok',
 'dadu'],
 ['gua', 'mending', 'pacar', 'lee', 'soo', 'man', 'rebut', 'saham'],
 ['saham', 'menfitnah']]

```

Gambar 3.2 Hasil dari *preprocessing*

### 3.2.3 Topic Modelling

Setelah data selesai dilakukan *preprocessing*, akan dibentuk pemodelan topik menggunakan metode *Latent Dirichlet Allocation*. Dua masukan utama untuk pemodelan topik menggunakan metode LDA adalah *dictionary* dan *corpus*. Dari data yang telah dilakukan *preprocessing* akan dibentuk *dictionary* berupa pemetaan kata-kata yang ada kedalam id. Kemudian akan dibentuk *corpus* yang berisi frekuensi kemunculan suatu kata. Sebelum dibentuk model LDA, *corpus* akan diolah lagi dengan menghitung frekuensi suatu kata dan memberi bobot relevansi kata tersebut di dalam dokumen tertentu.

Setelah semua masukan yang diperlukan telah ada, akan dibentuk model LDA dan dilakukan evaluasi dengan metode *coherence measurement* untuk memperoleh model yang optimal.

Hasil pemodelan topik dapat dilihat pada Tabel 3.2 di mana tiap topik akan tersusun atas beberapa kumpulan kata dengan frekuensi kemunculan kata tertentu. Kemudian hasil pemodelan topik akan disimpan pada database. Dari susunan kata yang ada dalam topic ini nanti akan digunakan untuk membuat SNA relasi antar topik.

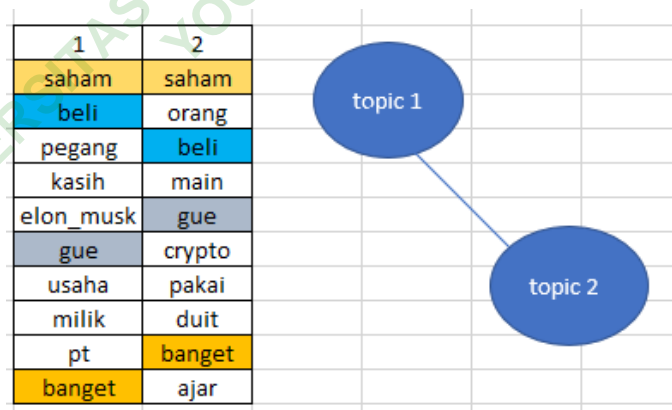
Tabel 3.2 Hasil Dari *Topic Modelling*

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4
ihsg	saham	saham	saham	saham
saham	beli	orang	goto	investasi

kuat	pegang	beli	rp	pasar
indeks	kasih	main	mei	turun
tutup	elon_musk	gue	dividen	harga
buka	gue	crypto	triliun	uang
persen	usaha	pakai	pegang	reksadana
dagang	milik	duit	bank	suku_bunga
bursa	pt	banget	selenggara	wall_street
merah	banget	ajar	gojek_tokopedia	pilih

### 3.2.4 Pembentukan *Social Network Analysis*

Topik yang telah berhasil diekstraksi pada tahap *topic modeling* akan dijadikan sebagai komponen utama pembentuk relasi pada model *social network analysis*. Relasi pada setiap node di bentuk dengan kedekatan antar topik yang dengan menggunakan kata pembentuk topic akan saling tumpang tindih dengan kata yang terdapat pada topik yang lain seperti contoh pada gambar pada topik 1 dan topik 2 terdapat kata yang sama yaitu : saham , beli, gue, dan banget. Maka akan adanya relasi antara kedua topik.



Gambar 3.3 Pembentukan Relasi

### 3.2.5 Web Deployment

Hasil model dari *topic modeling* dan *social network analysis* akan ditampilkan pada *platform website* berupa grafik *social network*. Dari model tersebut akan disajakannya dalam bentuk *website*.