

BAB III

METODE PENELITIAN

3.1. Subjek dan Objek Penelitian

Subjek penelitian ini menganalisis komentar masyarakat dan membandingkan akurasi algoritma SVM dan *Naïve Bayes* dari komentar yang didapatkan pada platform *Twitter* diambil dari tanggal 23 mei sampai 17 juni 2020 dikarenakan penelitian dimulai pada tahun 2020. Objek pada penelitian ini adalah komentar pengguna *Twitter* terkait dengan pelayanan kinerja PT. Telkom Indonesia di *social media Twitter studi kasus Indihome*.

3.2. Alat Dan Bahan

Alat dan Bahan yang digunakan dalam penelitian antara lain :

3.2.1. Perangkat Keras

Perangkat Keras yang digunakan dalam penelitian dengan sebuah laptop yang memiliki spesifikasi :

1. Prosesor Intel Core i5-8250U 1.6Ghz
2. Graphics Nvidia GeForce MX 150 2 GB VRAM
3. RAM 12 GB
4. Storage 1TB HDD
5. Display 14 inch @1366 x 768 pixels, 60 Hz Full HD

3.2.2. Perangkat Lunak

Perangkat Lunak yang digunakan dalam penelitian :

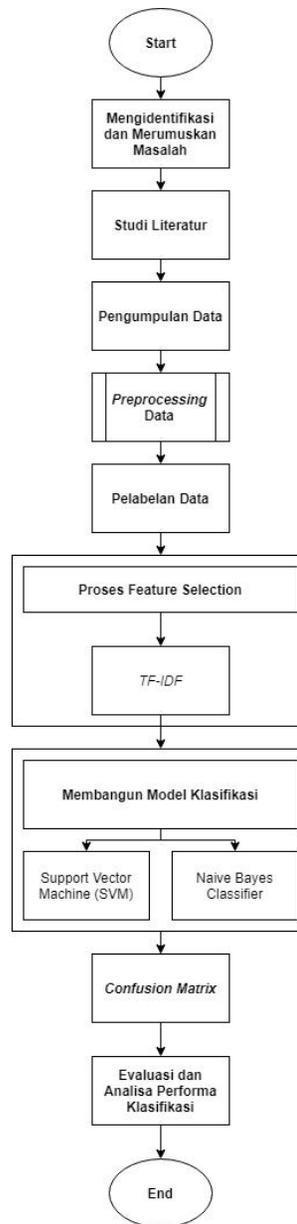
1. Sistem Operasi Windows 10
2. Jupyter Notebook
3. Anaconda
4. Google Colab

5. Python 3.8.7
6. Command Prompt
7. Browser (Google Chrome)
8. Tweepy

3.3. Bahan

Bahan penelitian yang digunakan untuk penelitian ini merupakan kumpulan data tweet pada media sosial Twitter dengan username @Indihome yang dapat dikategorikan atau berlabel menjadi *sentiment* berbentuk *negative* atau *sentiment* berbentuk *positive*.

3.4. Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian

Gambar 3.1 menjelaskan diagram alir penelitian yang akan dikembangkan oleh peneliti yang akan dijelaskan melalui poin-poin.

3.4.1. Mengidentifikasi dan Merumuskan Masalah

Tahap awal yang dilakukan peneliti adalah melakukan identifikasi dan merumuskan masalah yang akan diangkat dalam proses pengembangan dalam penelitian. Menentukan bidang, topik, masalah penelitian serta mengusulkan metode yang akan digunakan pada penelitian. Tahap ini mempelajari masalah yang masih terjadi di kehidupan sehari-hari, lalu tujuan, ruang lingkup serta metodologi penelitian. Pada penelitian ini ide atau masalah sudah ada penelitian terdahulunya, namun terdapat perbedaan pada algoritma atau metode yang diusulkan.

3.4.2. Studi Literatur

Peneliti melakukan studi literatur atau pengumpulan data pustaka, membaca, dan mencatat seluruh penelitian terdahulu yang memiliki kesamaan atau kemiripan terkait dengan penelitian yang akan dikembangkan oleh peneliti. Tahap ini, peneliti melakukan pengumpulan data – data yang berkaitan dengan topik permasalahan yang akan diangkat yakni tentang data mining, klasifikasi, analisis sentimen, tentang tweet di *social media twitter* serta algoritma yang akan diusulkan yaitu support vector machine dan naive bayes classifier. Studi literatur untuk penelitian ini diperoleh dari jurnal, buku, buku elektronik, situs internet, media elektronik. Studi literatur memiliki tujuan untuk memperkuat permasalahan yang dibahas pada penelitian serta menjadi dasar untuk melakukan pengembangan dan langkah-langkah selanjutnya.

3.4.3. Pengumpulan Data

Langkah pengumpulan data merupakan langkah penting dalam penelitian dimana akan dilakukan semua bentuk pengolahan atau pembuatan penelitian sangat bergantung kepada data, pengambilan data dari tanggal 23 mei sampai 17 juni 2020. Supaya mendapatkan hasil data yang maksimal yang kemudian akan dilakukan tahap pengembangan atau pengolahan terhadap data

tersebut, penelitian ini menggunakan bantuan Google Collab dalam pengumpulan data dimana menggunakan library *tweepy*, dan *API Twitter* untuk pengumpulan data terhadap cuitan *twitter* oleh pengguna dengan username *@Indihome* dalam penelitian ini pengumpulan data menggunakan Bahasa Indonesia.

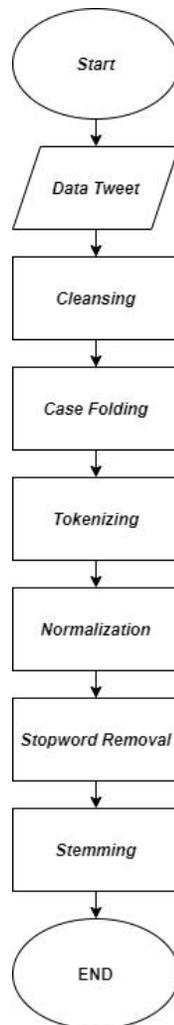
Tabel 3.1. Contoh Sentimen Positif dan Sentimen Negatif

Sentimen Positif	Sentimen Negatif
alhamdulillah internet kenceng murah rumah layan nya bagus	maklum bro indihome emang gitu layan kualitas aplikasi aja gk yg beres
sebenarnya gue bingung ko orang teriak indihome ancur dll gue ga rasa layan bagus orang teknis gue lapor trs langsung dateng speed ping stabil gue tinggal kebon jeruk jakarta barat	tunggu teknisi kalo gak dateng putus langgan udah muak layan bumn kualitas bintang maaf provider aja

3.4.4. *Preprocessing Data*

Langkah *preprocessing* data merupakan tahap dimana diharuskan peneliti melakukan pemrosesan terhadap data yang telah dikumpulkan dikarenakan data *tweet* yang telah dikumpulkan tadi beragam jenis penulisan oleh pengguna *twitter* dan belum memiliki kesiapan data untuk dilakukan proses langkah selanjutnya dalam tahap pengolahan data.

Diagram alir *preprocessing* data dapat dilihat pada gambar 3.2.



Gambar 3.2 Diagram Alir Preprocessing Data

1. *Cleansing*

Data cleansing merupakan tahapan pembersihan data dari hal-hal yang tidak diperlukan seperti koma (,), titik (.), URL, *emoticon*, dan tanda baca lainnya, pembersihan terdapat *null values* atau data yang kosong.

Tabel 3.2 Tahap Data Cleansing

Dokumen	Sebelum Data <i>Cleansing</i>	Setelah melakukan Data <i>Cleansing</i>
1	@TelkomCare Selamat siang. saya ingin berhenti layanan indihome dan mohon bantuannya	Salamat siang. saya ingin berhenti layanan indihome dan mohon bantuannya

Tabel 3.2 menjelaskan bahwa contoh data sebelum dilakukan data *cleansing* masih terdapat *username twitter* dalam kalimat cuitan pengguna twitter dan setelah dilakukannya data *cleansing* sudah tidak muncul lagi *username* dengan @TelkomCare

2. *Case Folding*

Case folding merupakan tahap awal *pre-processing* untuk mengubah setiap bentuk kata menjadi sama. Merubah semua kata menjadi *lowercase*.

Tabel 3.3 Tahap *Case Folding*

Dokumen	Sebelum Case Folding	Setelah Case Folding
1	Salamat siang. saya ingin berhenti layanan indihome dan mohon bantuannya	slamat siang. saya ingin berhenti layanan indihome dan mohon bantuannya

Tabel 3.3 menjelaskan *case folding* berfungsi untuk membuat seluruh kalimat menjadi berbentuk *lowercase*.

3. *Tokenizing*

Tokenizing yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan *delimiter-delimiter* seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata. Contoh proses tokenizing dapat dilihat pada tabel berikut ini.

Tabel 3.4 Tahap *Tokenizing*

Dokumen	Sebelum Tokenizing	Setelah Tokenizing
1	slamat siang. sya ingin berhenti layanan indihome dan mohon bantuannya	slamat
		siang
		saya
		ingin
		berhenti
		layanan
		indihome

Tabel 3.4 menjelaskan bahwa fungsi kegunaan dari *tokenizing* yakni adalah merubah kalimat menjadi berbentuk penggalan kata.

4. *Normalization*

Normalization merupakan perbaikan dan substansi kata yang salah eja ataupun disingkat dengan bentuk tertentu. Substitusi kata dilakukan untuk menghindari perhitungan dimensi kata yang melebar. Perhitungan dimensi kata akan melebar jika kata disingkat atau salah eja tidak diubah karena kata tersebut memiliki kontribusi dalam mempresentasikan dokumen tetapi akan dianggap sebagai entitas yang berbeda proses penyusunan matriks.

Tabel 3.5 Tahap *Normalization*

No	Sebelum Normalization	Setelah Normalization
1	slamat	selamat
2	siang	siang
3	saya	saya

4	ingin	ingin
5	berhenti	berhenti
6	layanan	layanan
7	indihome	indihome
8	dan	dan
9	mohon	mohon
10	bantuannya	bantuannya

Tabel 3.5 menjelaskan bahwa fungsi kegunaan dari normalization yakni adalah merubah kata yang sebelumnya memiliki *typo* atau singkatan kata tertentu dan selanjutnya diubah menjadi kata yang utuh sesuai dengan Kamus Besar Bahasa Indonesia (KBBI), contoh diatas yakni sebelum dilakukannya normalization berbentuk kata slamat dan setelah dilakukannya normalization berubah menjadi selamat.

5. *Stopword Removal*

Stopword Removal yaitu proses penghapusan kata-kata yang terdapat pada stoplist. Stoplist itu sendiri berisi kosakata-kosakata yang bukan merupakan ciri dari suatu dokumen dan bertujuan agar menghilangkan kata yang tidak berguna. Contoh stopwords removal bahasa Indonesia adalah kata hubung “dan”, “yang”, dan lain-lain. Contoh proses stopwords removal dapat dilihat pada tabel berikut ini.

Tabel 3.6 Tahap *Stopword Removal*

No	Sebelum Stopword Removal	Setelah Stopword Removal
1	selamat	selamat
2	siang	siang
3	saya	saya
4	ingin	ingin
5	berhenti	berhenti
6	layanan	layanan
7	indihome	indihome
8	dan	mohon
9	mohon	bantuannya
10	bantuannya	

Tabel 3.6 menjelaskan bahwa fungsi kegunaan dari *Stopword Removal* yakni adalah menghapus kata yang tidak berguna berdasarkan stoplist yang akan digunakan oleh peneliti, contoh diatas yakni adalah penggunaan “dan” memiliki dampak kedepannya dan termasuk kedalam daftar stoplist sehingga kata tersebut dihapus.

6. *Stemming*

Stemming adalah suatu proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*). Stemming bertujuan untuk menghilangkan berbagai imbuhan seperti prefiks,

sufiks, maupun konfiks yang ada pada setiap kata. Contoh proses stemming dapat dilihat pada tabel berikut ini.

Tabel 3.7 Tahap *Stemming*

No	Sebelum Stemming	Setelah Stemming
1	selamat	selamat
2	siang	siang
3	saya	saya
4	ingin	ingin
5	berhenti	henti
6	layanan	layanan
7	indihome	indihome
8	mohon	mohon
9	bantuannya	bantu

Tabel 3.7 menjelaskan bahwa fungsi kegunaan dari *Stemming* yakni menghilangkan berbagai imbuhan baik prefiks, sufiks, dan konfiks yang ada pada setiap kata, dapat dilihat dari tabel 3.6 yakni terdapat kata berhenti, bantuannya setelah dilakukan *stemming* kata tersebut berubah menjadi henti dan bantu

3.4.5. Pelabelan *Data*

Langkah Pelabelan Data merupakan langkah lanjutan dari pengumpulan data dimana data yang sudah dikumpulkan oleh peneliti selanjutnya dilakukan langkah pelabelan data. Penelitian yang akan dikembangkan oleh peneliti akan menggunakan *Lexicon Based Approach* dengan bantuan InSet (Indonesia Sentiment Lexicon) dan Elang dimana merupakan *Education Toolkit* untuk Bahasa Indonesia (NLP) dalam tahap

pelabelan data, tahap pelabelan data setiap kalimat akan digolongkan menjadi 2 kategori diantaranya tergolong sentimen *positive* atau sentimen *negative*.

3.4.6. Proses Feature Selection

Selanjutnya data akan melakukan proses feature selection atau seleksi fitur menggunakan pembobotan TF-IDF. Adanya pembobotan fitur ini sebelum klasifikasi dilakukan agar membantu meningkatkan akurasi klasifikasi.

Tabel 3.8 Contoh Dokumen

No	Username	Tweet/Dokumen
1	FinnoNino	layanan indihome emang bagus tapi mungkin sedikit ada respon lambat dari admin
2	danadyaksa	menurut aku bagus aja layanan indihome cuman kadang ada masalah jaringan jadi lambat wajar aja

- Menghitung *Term Frequency* (TF), yakni frekuensi kemunculan *term* (t) pada dokumen (D) dimana setiap *term* akan membaca seluruh dokumen yang tersedia, contoh kata layanan terdapat pada dokumen 1 dan dokumen 2 sedangkan kata emang hanya terdapat pada dokumen 1 saja.

Tabel 3.9 Menghitung TF

No	Term(t)	D1	D2
1	layanan	1	1
2	indihome	1	1
3	emang	1	0
4	bagus	1	1
5	tapi	1	0
6	mungkin	1	0
7	sedikit	1	0
8	ada	1	1
9	respon	1	0
10	lambat	1	1
11	dari	1	0
12	admin	1	0
13	menurut	0	1
14	aku	0	1
15	aja	0	1
16	cuman	0	1
17	kadang	0	1
18	masalah	0	1
19	jaringan	0	1
20	jadi	0	1
21	wajar	0	1
22	aja	0	1

- Menghitung *Document Frequency* (DF), yaitu banyaknya dokumen dimana suatu *term* (t) muncul. Kemudian menghitung DF yakni total yang telah dijumlahkan dari seluruh kata yang muncul dalam dokumen contoh kata layanan memiliki 2 kemunculan kata, dan kata emang hanya terdapat 1 total kemunculan kata terhadap dokumen.

Tabel 3.10 Menghitung DF

No	Term(t)	DF
1	layanan	2
2	indihome	2
3	emang	1
4	bagus	2
5	tapi	1
6	mungkin	1
7	sedikit	1
8	ada	2
9	respon	1
10	lambat	2
11	dari	1
12	admin	1
13	menurut	1
14	aku	1
15	aja	1
16	cuman	1
17	kadang	1
18	masalah	1
19	jaringan	1
20	jadi	1
21	wajar	1
22	aja	1

- Menghitung IDF. Menghitung IDF merupakan perhitungan dari log total dokumen dibagi dengan total kemunculan kata pada dokumen, contoh pada penelitian ini terhadap 2 dokumen dan pada kata layanan terdapat 2 total kemunculan kata setiap dokumen maka yang akan dihitung adalah

log 2 dari total dokumen dibagi dengan 2 total kemunculan kata mendapatkan hasil 0.1505 *ratio* dari kemungkinan kata yang muncul atau biasa disebut IDF.

Tabel 3.11 Menghitung IDF

No	Term(t)	DF	IDF
1	layanan	2	$\log 2/2 = 0.1505$
2	indihome	2	$\log 2/2 = 0.1505$
3	emang	1	$\log 2/1 = 0.3010$
4	bagus	2	$\log 2/2 = 0.1505$
5	tapi	1	$\log 2/1 = 0.3010$
6	mungkin	1	$\log 2/1 = 0.3010$
7	sedikit	1	$\log 2/1 = 0.3010$
8	ada	2	$\log 2/2 = 0.1505$
9	respon	1	$\log 2/1 = 0.3010$
10	lambat	2	$\log 2/2 = 0.1505$
11	dari	1	$\log 2/1 = 0.3010$
12	admin	1	$\log 2/1 = 0.3010$
13	menurut	1	$\log 2/1 = 0.3010$
14	aku	1	$\log 2/1 = 0.3010$
15	aja	1	$\log 2/1 = 0.3010$
16	cuman	1	$\log 2/1 = 0.3010$
17	kadang	1	$\log 2/1 = 0.3010$
18	masalah	1	$\log 2/1 = 0.3010$
19	jaringan	1	$\log 2/1 = 0.3010$
20	jadi	1	$\log 2/1 = 0.3010$
21	wajar	1	$\log 2/1 = 0.3010$
22	aja	1	$\log 2/1 = 0.3010$

- Menghitung TF-IDF

Tabel 3.12 Menghitung TF-IDF

Term(t)	TF		IDF	TF-IDF	
	D1	D2		D1	D2
layanan	1	1	0,1505	0.07525	0.07525
indihome	1	1	0,1505	0.07525	0.07525
emang	1	0	0,301	0,301	0
bagus	1	1	0,1505	0.07525	0.07525
tapi	1	0	0,301	0,301	0
mungkin	1	0	0,301	0,301	0
sedikit	1	0	0,301	0,301	0
ada	1	1	0,1505	0.07525	0.07525
respon	1	0	0,301	0.301	0
lambat	1	1	0,1505	0.07525	0.07525
dari	1	0	0,301	0.301	0
admin	1	0	0,301	0.301	0
menurut	0	1	0,301	0	0,301
aku	0	1	0,301	0	0,301
aja	0	1	0.301	0	0,301
cuman	0	1	0.301	0	0,301
kadang	0	1	0.301	0	0.301
masalah	0	1	0.301	0	0.301
jaringan	0	1	0.301	0	0.301
jadi	0	1	0.301	0	0.301
wajar	0	1	0.301	0	0.301

3.4.7. Membangun Model Klasifikasi

Selanjutnya membangun model klasifikasi menggunakan metode Naïve Bayes Classifier dan Support Vector Machine yang merupakan bagian data mining, proses ini merupakan inti dari proses ekstraksi pengetahuan dari data. Pengujian dataset pada penelitian ini dilakukan menggunakan dua algoritma klasifikasi yaitu Naïve Bayes Classifier dan Support Vector Machine yang akan diimplementasikan menggunakan bahasa pemrograman Python. Pada model klasifikasi terdapat proses training dan testing, dalam analisis ini data dibagi menjadi 80% dan 20%. Proses training menggunakan dataset adalah proses melatih mesin tentang pengetahuan dataset yang sudah diberi label sebelumnya. Kemudian proses testing, mesin akan diuji menggunakan 20% dari total dataset untuk mengetahui tingkat akurasi.

3.4.8. Evaluasi dan Analisa Performa Klasifikasi

Langkah evaluasi ini parameter yang digunakan adalah confusion matrix. Tujuan dari adanya evaluasi ini agar dapat melihat nilai *accuracy* dimana berdasarkan True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Nilai True Negative (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan False Positive (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, True Positive (TP) merupakan data positif yang terdeteksi benar. False Negative (FN) merupakan kebalikan dari True Positive, sehingga data positif, namun terdeteksi sebagai data negatif, *precision dan recall* dengan melihat performa model klasifikasi Naïve Bayes Classifier dan SVM untuk perbandingan algoritma NBC dan SVM terhadap pelayanan kinerja PT Telkom Indonesia pada studi kasus Indihome pada *social media Twitter*. Setelah mendapatkan hasil dari evaluasi maka dapat ditarik kesimpulan dari penelitian ini.