

BAB III

METODE PENELITIAN

3.1 Subyek dan Obyek Penelitian

Subyek dalam penelitian ini adalah pengguna Twitter di Indonesia. Obyek dalam penelitian ini adalah data komentar pengguna Twitter terhadap isu kemahalan minyak goreng di Indonesia. Data komentar yang digunakan merupakan data berupa teks berbahasa Indonesia dengan data awal sebanyak 10.000 data.

3.2 Alat Penelitian

Untuk mendukung keberhasilan penelitian yang dilakukan, maka diperlukan analisis terhadap kebutuhan sistem. Adapun komponen yang dibutuhkan yaitu:

3.2.1. Perangkat Keras

Penelitian yang dilakukan menggunakan sebuah laptop dengan spesifikasi:

- 1) Merk : Lenovo Ideapad 330
- 2) Processor : AMD A4
- 3) Memori : RAM 4GB, SSD 128GB
- 4) Sistem operasi : Windows 10

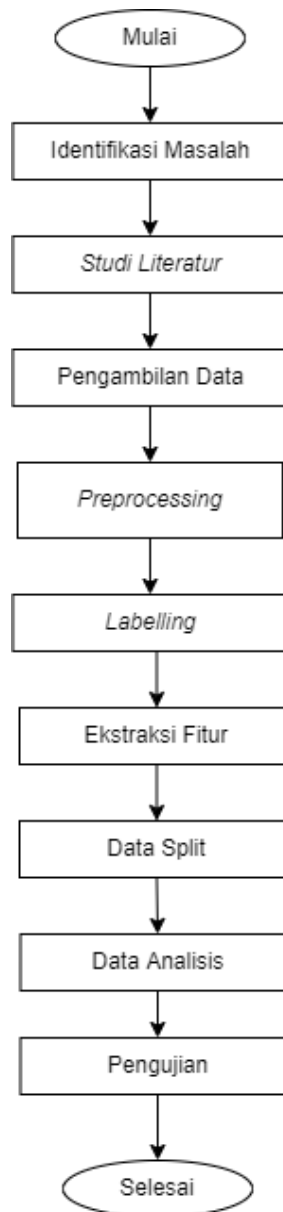
3.2.2. Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini diantaranya:

- 1) Microsoft Word 2013 sebagai media penyusunan laporan penelitian
- 2) Google Chrome yang digunakan sebagai media untuk membuka teks editor Google Colaboratory
- 3) Anaconda Navigator digunakan sebagai alat bantu analisis.

3.3 Diagram Alur Penelitian

Diagram alir penelitian berisi mengenai Langkah Langkah yang akan dilakukan peneliti dalam menganalisa sentimen komentar masyarakat di media sosial Twitter pada gambar 3.1.



Gambar 3. 1 Gambar Alur Penelitian

3.3.1. Identifikasi Masalah

Tahap identifikasi masalah adalah tahap penulis merumuskan masalah yang akan diangkat, tahap ini terdiri dari beberapa bagian. Perumusan latar belakang dibuat untuk dijadikan sebagai dasar penelitian yang dilakukan, metode yang akan digunakan juga sudah ditentukan. Selain latar belakang juga ditentukan rumusan masalah, pertanyaan, tujuan dan manfaat penelitian.

3.3.2. Studi Literature

Tahap studi literature adalah tahap penulis melakukan pengkajian terhadap penelitian lain yang sudah dilakukan peneliti lain. Pada tahap ini akan didapatkan perbandingan antara penelitian sebelumnya dengan penelitian yang akan dilakukan.

3.3.3. Pengumpulan Data

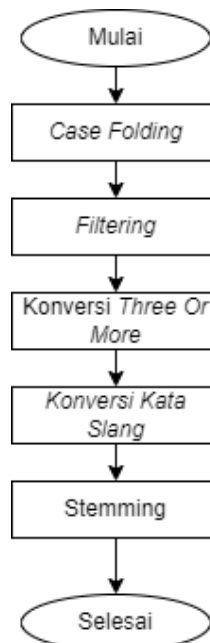
Teknik mengumpulkan data yang dilakukan adalah dengan mengumpulkan *tweet* pengguna Twitter dengan pencarian “minyak goreng mahal” dengan periode waktu 25 Maret 2022 hingga 13 Agustus 2022. Metode yang digunakan adalah dengan cara *crawling* dari Twitter menggunakan program. Bahasa pemrograman yang digunakan adalah bahasa pemrograman python menggunakan *library socrab*. Data yang diambil adalah data *tweet* atau teks sejumlah 10.000 baris data. Sebagai contoh dapat diamati potongan data pada gambar 3.2.

tweet
@ourprettystuffs Memang sekarang yang sangat dibutuhkan tuh minyak goreng karna sulit dan harganya pun mahal
@ourprettystuffs Memang sekarang yang sangat dibutuhkan tuh minyak goreng karna sulit dan harganya pun mahal
@era_gunawan @khairn_nisa Minyak mahal. Jgn goreng dulu🙄
@Mukidioon2 Lengkap sudah penderitaan rakyat ini,,sembako mahal,bbm naik, gas naik,,minyak goreng mahal,, hutang mem
@ayamsuwirr minyak goreng? Mahal dik
@TarunaAdjie1 @msaid_didu Kalau di negeri Antah Berantah semua lini dipenuhi kebobrokan jadi susah mencari yg baik, raky
@Askrifess Sok keras wkwk. Jamanku pkkmb disuru baris aku kabur balik ke kos. Hari ke 3 disuru bawa minyak goreng, susu, yg
@flippedtable @kemkominfo @ditjen_migrasi @PNS_Ababil harga cabe mahal, ya tanam sendiri di pekarangan,minyak mah
@papa_loren Minyak goreng mahal pakai dikukus, digodok aja...solusi pejabat sekarang
Biayanya murah krn risetnya abal2, spt:1) sth harga minyak goreng naik - keinginan menggoreng menurun2) terigu bisa diganti
@kegblgnunfaedh Olinya ganti sama minyak goreng aja biar ga mahal
@msaid_didu @prastow perang ga ngaruh... komitmen ke tionskok ngaruh bgt lho.... tionskok butuh jaminan milyaran rakyatn
@RamiRizal Coba dicek apa bener semua lahan kelapa sawit yg diindonesia adalah milik orang pribumi??...ya wajar saja harga
@TRANSTV_CORP Min..goreng katsu mah jangan pake minyak tanah dong. Emang efek migor mahal gitu?
@DanielS19494891 @brin_watch @yasmin_amalia @ConversationIDN Oo setelah ada BRIN jadi maju ya. BRIN buat survey mer
Halo guis aku mau nanya kalo minyak goreng 1L 19k itu mahal gak ya? Harga pasarannya sekarang berapa? Maaf aku bener-ben
@karmatiranda @alextham878 Buktinya Setuju dgn TKA China, Bahagia dgn Minyak goreng mahal, Happy dgn BBM mahal, itu!
@tanyakanrl telor rebus karena minyak goreng mahal
harga minyak goreng emang lagi mahal, tapi kalo untuk kebutuhan menggoreng isu isu nak ormek...aku beli 20 karton deh
@_ekokuntadhi Minyak goreng naik, bbm naik, sembarang mahal, preaidennya ada gk bong
Mendag Akui Minyak Goreng di Kawasan Timur Masih Mahal https://t.co/7rEj7MgOb
Ini negara mau di bawa kemana si arahnya???🙄,mulai dari mahal minyak goreng,pas ktauan dalangnya ada cukong dibalikny

Gambar 3. 2 Gambar Potongan *Dataset*

3.3.4. Preprocessing

Salah satu hal yang harus dilakukan dalam pengolahan data teks adalah tahap pengolahan data menjadi lebih bersih atau dikenal dengan *preprocessing*. Tahap ini diperlukan karena data yang didapatkan tidak seluruhnya dapat digunakan dalam proses analisis, terdapat beberapa bagian yang dihilangkan atau dirubah. Dengan tahap ini maka data yang digunakan akan sesuai dengan kebutuhan pada proses analisis sehingga didapatkan hasil yang baik [11]. Secara sistematis tahap *preprocessing* dilakukan dengan tahapan seperti pada Gambar 3.3.



Gambar 3. 3 Gambar Alur Penelitian

- 1) *Case folding*, adalah tahap konversi huruf besar untuk dijadikan huruf kecil, sebagai contoh dapat diamati pada tabel 3.1.

Tabel 3. 1 Tabel *Case Folding*

Sebelum	Sesudah
Heraaaan!!! harga minyak goreng sekarang mahalnya kelewatan, padahal Indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari @beritaviral https://pemkomedan.go.id/arti	heraaaaan!!! harga minyak goreng sekarang mahalnya kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari @beritaviral https://pemkomedan.go.id/arti

Sebelum	Sesudah
kel-22322-tidak-langka-minyak-goreng-hanya-harganya-mahal.html #minyagoreng #minyakmahal	kel-22322-tidak-langka-minyak-goreng-hanya-harganya-mahal.html #minyagoreng #minyakmahal

2) Filtering

- a. *Remove link web*, adalah menghapus link yang mengacu kesuatu web, ditandai dengan 'http' diawalnya seperti tertera pada tabel 3.2.

Tabel 3. 2 Tabel *Remove Link Web*

Sebelum	Sesudah
heraaaaan!!! harga minyak goreng sekarang mahalnya kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari @beritaviral https://pemkomedan.go.id/artikel-22322-tidak-langka-minyak-goreng-hanya-harganya-mahal.html #minyagoreng #minyakmahal	heraaaaan!!! harga minyak goreng sekarang mahalnya kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari @beritaviral #minyagoreng #minyakmahal

- b. *Remove username*, adalah menghapus *username* atau mention yang terdapat pada data *tweet* seperti tertera pada tabel 3.3.

Tabel 3. 3 Tabel *Remove Username*

Sebelum	Sesudah
heraaaaan!!! harga minyak goreng sekarang mahal nya kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari @beritaviral #minyakgoreng #minyakmahal	heraaaaan!!! harga minyak goreng sekarang mahal nya kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari #minyakgoreng #minyakmahal

- c. *Remove* tagar adalah tahap menghapus hastag atau tagar seperti tertera pada tabel 3.4.

Tabel 3. 4 Tabel *Remove Tagar*

Sebelum	Sesudah
heraaaaan!!! harga minyak goreng sekarang mahal nya kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari #minyakgoreng #minyakmahal	heraaaaan!!! harga minyak goreng sekarang mahal nya kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari

- d. *Remove punctuation*, adalah menghapus karakter selain alphabet seperti tertera pada tabel 3.5.

Tabel 3. 5 Tabel *Remove Punctuation*

Sebelum	Sesudah
heraaaaan!!! harga minyak goreng sekarang mahal nya	heraaaaan harga minyak goreng sekarang mahal nya

Sebelum	Sesudah
kelewatan, padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari	kelewatan padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari

- 3) Konversi *Three or More* adalah mengubah kata yang terdapat 3 atau lebih huruf yang sama dalam satu kata menjadi bentuk aslinya, sebagai contoh kata 'heraaan' akan diubah menjadi 'heran' seperti tertera pada tabel 3.6.

Tabel 3. 6 Tabel Konversi *Three or More*

Sebelum	Sesudah
heraaaaan harga minyak goreng sekarang mahalnnya kelewatan padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari	heran harga minyak goreng sekarang mahalnnya kelewatan padahal indonesia punya banyak sawit buat produksi migor sendiri. berita terbaru dari

- 4) Konversi kata slang adalah tahap mengubah kata slang atau kata gaul menjadi bentuk kata yang baku berdasarkan kamus yang dibuat seperti tertera pada tabel 3.7.

Tabel 3. 7 Tabel Konversi Kata Slang

Sebelum	Sesudah
heran harga minyak goreng sekarang mahalnnya kelewatan padahal indonesia punya banyak sawit buat produksi	heran harga minyak goreng sekarang mahalnnya kelewatan padahal indonesia punya banyak sawit buat produksi minyak goreng sendiri. berita terbaru dari

Sebelum	Sesudah
migor sendiri berita terbaru dari	

- 5) *Stemming* adalah membersihkan sebuah kata dengan menghapus imbuhan pada sebuah kata sehingga akan didapatkan kata dasarnya seperti tertera pada tabel 3.8.

Tabel 3. 8 Tabel *Stemming*

Sebelum	Sesudah
heran harga minyak goreng sekarang mahalnya kelewatan padahal indonesia punya banyak sawit buat produksi minyak goreng sendiri berita terbaru dari	heran harga minyak goreng sekarang mahal lewat padahal indonesia punya banyak sawit buat produksi minyak goreng sendiri berita baru dari

3.3.6 Pelabelan Data

Proses analisis sentimen penelitian ini menggunakan metode supervised learning. Metode supervised learning membutuhkan label pada masing-masing baris datanya untuk pembangunan model. Pelabelan adalah tahap menentukan kalimat termasuk ke dalam kategori positif atau negatif. Tahap ini akan menggunakan metode *lexicon based* dengan menggunakan kamus lexicon berbahasa Inggris dari *library python TextBlob*. Konversi data dari bahasa Indonesia ke bahasa Inggris harus dilakukan karena perbedaan bahasa antara kamus lexicon yang digunakan dengan data yang akan dicari polaritasnya. *Library TextBlob* memiliki 2919 kata berbahasa Inggris dimana masing-masing kata sudah memiliki polaritas masing-masing. Dalam menentukan polaritas *TextBlob* akan melihat apakah kalimat tersebut memiliki skor positif atau negatif [9].

Penelitian analisis teks sudah dilakukan sebelumnya oleh E.Indrayuni pada tahun 2019 dan T.Krisdiyanto tahun 2021 menggunakan metode Naïve Bayes serta R.Tineges pada tahun 2020 dan S.Azza Amira pada tahun 2020 menggunakan metode SVM. Keempat penelitian tersebut membagi klasifikasinya menjadi dua kelas yaitu positif dan negatif. Hasil pengujian yang dilakukan didapatkan hasil akurasi yang cukup baik yaitu diatas 85%, sehingga dalam penelitian ini juga akan digunakan dua kelas dalam klasifikasinya.

3.3.7 Ekstraksi Fitur

Tahap pembobotan adalah tahap yang termasuk ke dalam tahap ekstraksi fitur. Tahap ini akan mengubah data teks menjadi data numerik yang dapat diolah oleh komputer. Pembobotan dalam penelitian ini dilakukan dengan metode Term Frequency Inverse Document Frequency (TF-IDF). TF-IDF akan memberikan nilai pada masing-masing kata pada setiap record data. Langkah-langkah dalam menentukan bobot tiap kata yaitu:

- 1) Menghitung frekuensi kemunculan kata pada setiap record data atau disebut sebagai tf .
- 2) Menghitung nilai IDF dengan persamaan:

$$idf = \log\left(\frac{n}{df}\right) \quad (3.1)$$

Keterangan:

n = jumlah record data

df = jumlah kemunculan kata pada n

- 3) Menghitung nilai $tf-idf$ pada masing-masing record data menggunakan persamaan:

$$tfidf = tf * idf \quad (3.2)$$

Keterangan:

tf = kemunculan kata pada setiap record data

idf = nilai *invers* dari kemunculan data

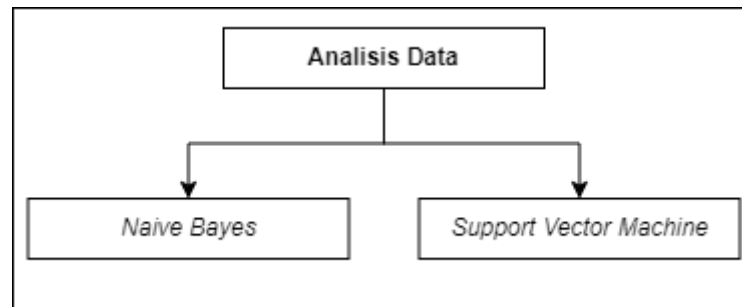
- 4) Lakukan langkah tiga untuk setiap kata pada masing-masing baris data.

3.3.8 Data Split

Tahap selanjutnya adalah *data split* dimana data akan dibagi menjadi dua bagian. Bagian pertama adalah data latih sebagai data untuk membuat model pada metode Support Vector Machine dan Naïve Bayes. Bagian kedua adalah data tes untuk mengukur kualitas model yang ada. Penelitian yang dilakukan oleh Muhammad Akbar S dengan metode SVM membagi *dataset* dengan perbandingan data latih dan data tes 75%:25%, 80%:20% dan 90%:10% menghasilkan tingkat akurasi tertinggi pada perbandingan 80%:20% dibandingkan dua perbandingan lainnya [19]. Penelitian yang dilakukan oleh Winda Christina Widyaningtyas dengan metode Naïve Bayes membagi data latih dan data uji dengan perbandingan 20%:80%, 50%:50% dan 80%:20% menghasilkan tingkat akurasi tertinggi pada perbandingan 80%:20% dibandingkan dua perbandingan lainnya [20]. Berdasarkan dua penelitian tersebut pada penelitian ini akan menggunakan perbandingan data latih dan data uji 80%:20%.

3.3.9 Data Analisis

Proses analisis menggunakan data *training* yang didapatkan pada langkah data *split*. Secara umum tahap dalam data analisis dapat diamati dalam gambar 3.4.



Gambar 3. 4 Gambar Alur Analisis Data

Hasil dari data analisis adalah dibuatnya model pada masing-masing metode yang digunakan. Model yang terbentuk kemudian akan dilakukan proses pengujian model.

3.3.10 Pengujian

Setelah model dari masing-masing metode terbentuk langkah selanjutnya adalah menguji model tersebut. Confusion Matrix digunakan untuk mencari *accuracy*, *precision*, *recall* dan *F1-Score* sebagai bahan pengujian. Tahap *preprocessing* yang baik dan menghasilkan data yang matang dan bersih akan membuat hasil pengujian performa semakin tinggi.