

BAB III

METODOLOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

Subjek pada penelitian ini adalah pengguna sosial media X atau X yang memberikan komentar terhadap rangka eSAF Honda. Objek pada penelitian ini yaitu data cuitan atau komentar X yang menggunakan kata kunci terkait rangka eSAF Honda. Data ini akan dianalisis menggunakan algoritma *Support Vector Machine* (SVM) untuk menentukan polaritas sentimennya yaitu positif dan negatif.

3.2 Alat Dan Bahan Penelitian

3.2.1 Perangkat Keras

Penelitian yang dilakukan menggunakan sebuah laptop dengan spesifikasi:

- 1) Merk : Dell Latitude E7470
- 2) Processor : intel core-i5
- 3) Memori : RAM 8GB, SSD 256GB
- 4) Sistem operasi : Windows 11

3.2.2 Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini diantaranya:

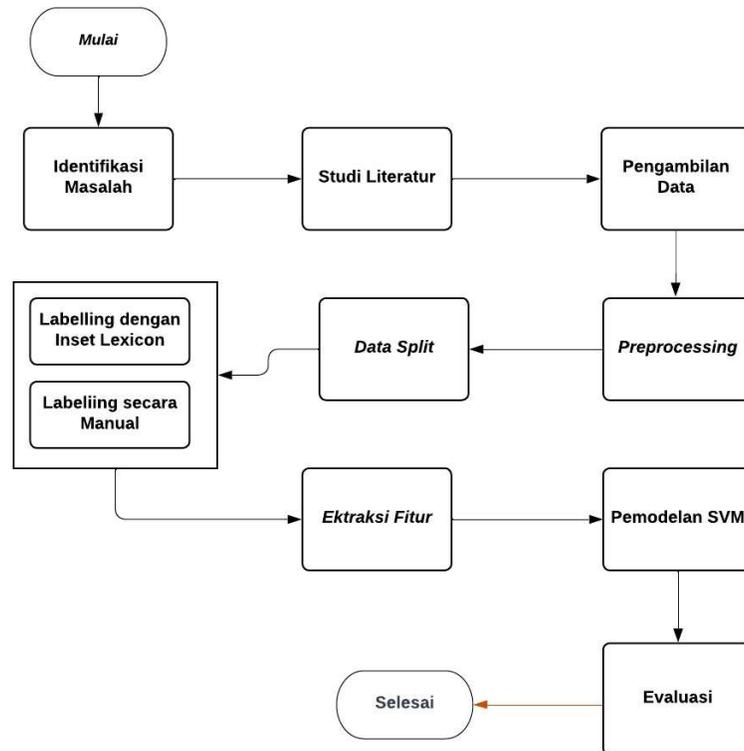
- 1) Microsoft Word 2019 sebagai media penyusunan laporan penelitian.
- 2) Google Chrome yang digunakan sebagai media untuk membuka Google Colaboratory.
- 3) Google Colaboratory digunakan sebagai alat bantu analisis.

3.2.3 Bahan Penelitian

Bahan penelitian yang digunakan pada penelitian ini yaitu data komentar atau tweet yang berisi opini tentang rangka esaf honda.

3.3 Diagram Alir Penelitian

Diagram alir penelitian merupakan representasi grafis dari langkah-langkah atau proses yang dilalui dalam suatu penelitian. Diagram ini membantu menggambarkan secara sistematis bagaimana penelitian dilakukan dari awal hingga akhir. Diagram alir penelitian terdapat pada gambar 3.1



Gambar 3.1 Diagram Alir

3.3.1 Identifikasi Masalah

Pada tahapan identifikasi masalah, dimulai dengan merumuskan permasalahan yang berkaitan dengan pro-kontra penggunaan rangka eSAF pada sepeda motor Honda pada komentar pengguna X. Penyusunan latar belakang menjadi dasar utama untuk melaksanakan penelitian ini, dan selama tahap ini, metode penelitian yang akan diterapkan sudah ditentukan dengan menggunakan *Support Vector Machine*. Selain latar belakang, juga telah ditegaskan rumusan masalah, pertanyaan penelitian, tujuan, dan manfaat yang ingin diperoleh dari hasil penelitian.

3.3.2 Studi Literatur

Tahapan studi literatur melibatkan pengkajian terhadap penelitian-penelitian sebelumnya yang berkaitan dengan analisis sentimen di X, penggunaan algoritma *Support Vector Machine*(SVM)mencari informasi yang berkaitan dengan rangka eSAF Honda. Pada tahap ini, diperoleh perbandingan antara penelitian yang telah dilakukan sebelumnya dengan penelitian yang sedang direncanakan. Hal ini membantu untuk memahami kontribusi penelitian sebelumnya, mengidentifikasi gap dalam pengetahuan, dan merinci bagaimana penelitian yang sedang direncanakan akan memberikan nilai tambah atau perspektif baru.

3.3.3 Pengambilan Data

Metode pengambilan dan pengumpulan data dalam penelitian ini melibatkan pengumpulan tweet dari pengguna X dengan kata kunci "esaf honda" dalam rentang waktu 1 Agustus 2023 hingga 31 Desember 2023. Proses dilakukan melalui *crawling* dari platform X dengan menggunakan program. Bahasa pemrograman yang digunakan adalah Python, dengan eksekusi *tweet-harvest* melalui Node.js melalui npm (*Node Package Manager*).

3.3.4 Preprocessing

Tahapan selanjutnya yaitu *preprocessing*. Tujuan dari *preprocessing* adalah untuk meningkatkan kualitas data, mengatasi masalah-masalah yang mungkin muncul, dan membuat data siap untuk diproses. Tahapan dalam *preprocessing* terdiri dari *case folding*, *cleansing*, *tokenizing*, *stopword*, dan *stemming*.

1. *Case Folding*, adalatahap mengubah semua huruf menjadi huruf kecil, sebagai contoh dapata dilihat pada tabel 3.1.

Tabel 3.1 Contoh Case Folding

Sebelum	Sesudah
Kebetulan saya menggunakan motor Honda yang menggunakan sasis Esaf, Scoopy dan Beat Deluxe tahun 2021 yang konon katanya sasis nya mudah keropos	kebetulan saya menggunakan motor honda yang menggunakan sasis esaf, scoopy dan beat deluxe tahun 2021 yang konon katanya sasis nya mudah keropos dan

dan patah. Insya Allah minggu ini akan saya bongkar untuk ngecek kondisi sasis nya. Sambil nunggu saya membuktikan sendiri, https://t.co/rICuoLawyF	patah. insya allah minggu ini akan saya bongkar untuk ngecek kondisi sasis nya. sambil nunggu saya membuktikan sendiri, https://t.co/ricuolawyf
---	---

2. *Cleansing*, pembersihan data yang berguna untuk menghilangkan elemen-elemen yang tidak perlu seperti angka, tanda baca, tautan, mention, dan emoticon, sebagai contoh dapat dilihat pada tabel 3.2

Tabel 3.2 Contoh *Cleansing*

Sebelum	Sesudah
kebetulan saya menggunakan motor honda yang menggunakan sasis esaf, scoopy dan beat deluxe tahun 2021 yang konon katanya sasis nya mudah keropos dan patah. insya allah minggu ini akan saya bongkar untuk ngecek kondisi sasis nya. sambil nunggu saya membuktikan sendiri, https://t.co/ricuolawyf	kebetulan saya menggunakan motor honda yang menggunakan sasis esaf scoopy dan beat deluxe tahun yang konon katanya sasis nya mudah keropos dan patah insya allah minggu ini akan saya bongkar untuk ngecek kondisi sasis nya sambil nunggu saya membuktikan sendiri

3. *Tokenizing*, berupa memecah dokumen kalimat menjadi unit yang lebih kecil menjadi kata atau disebut token, sebagai contoh dapat dilihat pada tabel 3.3.

Tabel 3.3 Contoh *Tokenizing*

Sebelum	Sesudah
kebetulan saya menggunakan motor honda yang menggunakan sasis esaf scoopy dan beat deluxe tahun yang konon katanya sasis nya mudah keropos dan patah insya allah minggu ini akan saya bongkar untuk ngecek kondisi sasis nya sambil nunggu saya membuktikan sendiri	['kebetulan', 'saya', 'menggunakan', 'motor', 'honda', 'yang', 'menggunakan', 'sasis', 'esaf', 'scoopy', 'dan', 'beat', 'deluxe', 'tahun', 'yang', 'konon', 'katanya', 'sasis', 'nya', 'mudah', 'keropos', 'dan', 'patah', 'insya', 'allah', 'minggu', 'ini', 'akan', 'saya', 'bongkar', 'untuk', 'ngecek', 'kondisi', 'sasis', 'nya', 'sambil', 'nunggu', 'saya', 'membuktikan', 'sendiri']

4. *Stopword*, proses ini melibatkan penghapusan kata-kata yang memiliki sedikit makna atau tidak relevan, seperti contohnya “yang”, “di”, “ke”, dan sejenisnya yang sering muncul dalam kalimat. Contoh dapat dilihat pada tabel 3.4.

Tabel 3.4 Contoh *Stopword*

Sebelum	Sesudah
['kebetulan', 'saya', 'menggunakan', 'motor', 'honda', 'yang', 'menggunakan', 'sasis', 'esaf', 'scoopy', 'dan', 'beat', 'deluxe', 'tahun', 'yang', 'konon', 'katanya', 'sasis', 'nya', 'mudah', 'keropos', 'dan', 'patah', 'insya', 'allah', 'minggu', 'ini', 'akan', 'saya', 'bongkar', 'untuk', 'ngecek', 'kondisi', 'sasis', 'nya', 'sambil', 'nunggu', 'saya', 'membuktikan', 'sendiri']	['motor', 'honda', 'sasis', 'esaf', 'scoopy', 'beat', 'deluxe', 'konon', 'sasis', 'mudah', 'keropos', 'patah', 'insya', 'allah', 'bongkar', 'cek', 'sasis', 'tunggu', 'membuktikan']

5. *Stemming*, proses pembersihan suatu kata dengan menghapus imbuhan sehingga diperoleh bentuk dasar atau akar kata, sebagai contoh dapat dilihat pada tabel 3.5.

Tabel 3.5 Contoh *Stemming*

Sebelum	Sesudah
['motor', 'honda', 'sasis', 'esaf', 'scoopy', 'beat', 'deluxe', 'konon', 'sasis', 'mudah', 'keropos', 'patah', 'insya', 'allah', 'bongkar', 'ngecek', 'sasis', 'tunggu', 'membuktikan']	motor honda sasis esaf scoopy beat deluxe konon sasis mudah keropos patah insya allah bongkar cek sasis tunggu bukti

3.3.5 Data Split

Tahap berikutnya adalah *data split* atau pemisahan data, dimana dataset akan dibagi menjadi dua bagian. Bagian pertama berfungsi sebagai data pelatihan atau *training data* yang akan digunakan untuk mengembangkan model menggunakan metode *Support Vector Machine*. Sementara bagian kedua, berperan sebagai data uji atau *testing data*, digunakan untuk mengukur performa dan kualitas dari model yang telah

dibuat. Pada penelitian [31] menggunakan metode SVM membagi data latih dan uji sebesar 80%:20%, yang mendapatkan tingkat akurasi senilai 74%. Berdasarkan pada penelitian tersebut, pada penelitian ini akan menggunakan perbandingan data latih dan data uji 80%:20%. *Data training* akan menggunakan hasil pelabelan dengan Inset lexicon, sedangkan *data training* akan menggunakan hasil pelabelan secara manual.

3.3.6 Labelling

Labelling atau pelabelan data adalah tahap menentukan kalimat termasuk ke dalam kategori positif atau negatif. Labeling teks untuk analisis sentimen dengan menggunakan *lexicon* melibatkan penggunaan kamus kata atau lexicon yang memiliki nilai sentimen tertentu untuk setiap kata. *Lexicon* ini mencakup daftar kata-kata beserta nilai sentimen positif, negatif, atau netral yang terkait dengan masing-masing kata. Pada penelitian ini menggunakan kamus kata opini (*lexicon*) dari penelitian sebelumnya yaitu InSet Lexicon[39]. Metode ini akan mengklasifikasikan *tweet* menjadi 2 kategori yaitu positif dan negatif, sehingga dalam penelitian ini juga akan menggunakan dua kelas dalam klasifikasinya. Selanjutnya, penelitian ini juga akan menggunakan pelabelan secara manual, menentukan label positif atau negatif untuk setiap *tweet* secara manual.

Berikut contoh dari pelabelan, dapat dilihat pada tabel 3.6.

Tabel 3.6 Contoh Pelabelan

Kalimat	Penghitung Kata	Score	Label
iya rangka esaf rangka motor honda patah	Patah = -4	-4	Negatif
honda beat scoopy vario produk laris manis genio laris honda eksperimen kurang biaya produksi pangkas sisi blunder rangka esaf beat scoopy vario fenomenal	Manis = 5 Produk= 3	8	Positif

3.3.7 Ekstraksi Fitur

Tahap pembobotan adalah langkah dalam proses ekstraksi fitur di mana mengubah data teks menjadi bentuk numerik yang bisa diolah oleh komputer. Metode TF-IDF digunakan untuk menghitung frekuensi kemunculan suatu kata dalam suatu dokumen dan frekuensi kemunculan kata tersebut dalam seluruh dokumen. Nilai TF-IDF menunjukkan seberapa penting suatu kata dalam suatu dokumen.

3.3.8 Pemodelan dengan SVM

Tahapan berikutnya adalah pemodelan, pada tahap pemodelan menggunakan *Support Vector Machine(SVM)* model *machine learning* dibangun untuk melakukan klasifikasi sentimen positif dan negatif pada data latih yang telah dipisahkan sebelumnya. SVM, dalam upayanya menghasilkan model yang dapat menggeneralisasi pola-pola dalam dataset, berfokus pada penemuan *hyperplane* terbaik yang memisahkan kelas sentimen tersebut.

3.3.9 Evaluasi

Tahapan selanjutnya, model ini akan dievaluasi dengan menggunakan data uji untuk mengukur kinerja dan kemampuan prediktifnya. *Confusion Matrix* digunakan sebagai alat untuk mengukur *accuracy*, *precision*, *recall*, dan *F1-Score* sebagai metrik evaluasi. Proses *preprocessing* yang efektif, menghasilkan data yang matang dan bersih, akan meningkatkan performa hasil uji coba dengan menghasilkan nilai yang lebih tinggi.