

**ANALISIS SENTIMEN PADA MEDIA SOSIAL *TWITTER*
TERHADAP PENGGUNAAN KENDARAAN LISTRIK
DENGAN MENGGUNAKAN METODE *NAÏVE BAYES*
CLASSIFIER DAN *INFORMATION GAIN***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-I Pada Jurusan Teknik Informatika



Oleh:

Ivando Sibarani

NIM: 09021282025046

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024**

LEMBAR PENGESAHAN SKRIPSI

**ANALISIS SENTIMEN PADA MEDIA SOSIAL *TWITTER*
TERHADAP PENGGUNAAN KENDARAAN LISTRIK
DENGAN MENGGUNAKAN METODE *NAÏVE BAYES*
CLASSIFIER DAN *INFORMATION GAIN***

Oleh :

Ivando Sibarani

NIM: 09021282025046

Inderalaya, 8 Oktober 2024

Pembimbing I

Pembimbing II

Rizki Kurniati, S.Kom, M.T.

NIP 199107122019032016

Desty Rodiah, S.Kom, M.T.

NIP 198912212020122011

Mengetahui

Dekan Fakultas Ilmu Komputer



Hadipurnawan Satria, Ph.D.

NIP 198004182020121001

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Jum'at tanggal 13 September 2024 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Ivando Sibarani
NIM : 09021282025046
Judul : Analisis Sentimen pada Media Sosial Twitter terhadap Penggunaan Kendaraan Listrik dengan menggunakan Metode *Naïve Bayes Classifier* dan *Information gain*
dan dinyatakan LULUS

1. Ketua Penguji

Yunita, S.Si., M.Cs.
NIP 198306062015042002

2. Pembimbing 1

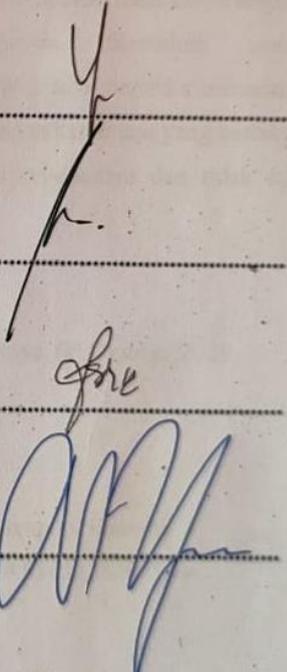
Rizki Kurniati, S.Kom., M.T.
NIP 199107122019032016

3. Pembimbing 2

Desty Rodiah, S.Kom., M.T.
NIP 198912212020122011

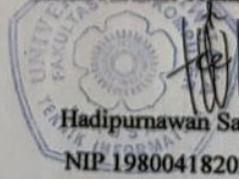
4. Penguji 1

Novi Yusliani, S.Kom., M.T.
NIP 198211082012122001



Mengetahui,

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.

NIP 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Ivando Sbarani
NIM : 09021282025046
Program Studi : Teknik Informatika
Judul : Analisis Sentimen pada Media Sosial Twitter terhadap Penggunaan
Kendaraan Listrik dengan menggunakan Metode *Naïve Bayes Classifier* dan *Information gain*

Hasil Pengecekan Turnitin : 9%

Menyatakan bahwa Laporan Penelitian saya merupakan hasil karya sendiri dan bukan hasil penjiplakkan/plagiat. Apabila ditemukan unsur penjiplakkan/plagiat dalam laporan penelitian ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Indralaya, 08 Oktober 2024



Ivando Sibarani

NIM 09021282025046

Motto dan Persembahan

“Hidup Memang Tidak Adil, Jadi Biasakan Dirimu”

Kupersembahkan Karya Tulis ini Kepada:

- Orang Tua
- Keluarga Besar
- Teman-teman TI REG 2020
- Para Dosen dan staff Fakultas Ilmu
Komputer Universitas Sriwijaya

ABSTRACT

The use of electric vehicles is increasing worldwide as a solution to reduce carbon emissions and dependence on fossil fuels. However, the adoption of this technology faces various challenges, including public acceptance. Given the large number of opinions on Twitter, it is important to classify sentiment to facilitate the identification of opinions on electric vehicles. This study aims to analyze public sentiment towards the use of electric vehicles through data from Twitter. This study uses 4174 tweet data with 1903 positive sentiments, 1603 negative sentiments, and 668 neutral sentiments. The method used in this study is the Naïve Bayes Classifier and Information gain feature selection. The results of the study with the Naïve Bayes Classifier model applied with Information gain obtained the highest accuracy, precision, recall, and f1-score at a ratio of 80:20 of 61.77%, 40.1%, 47.8%, and 43.13%, with a training time of 0.0054 seconds. Meanwhile, by only using the naive bayes method, the values obtained were 60.2%, 40.75%, 46.7% and 42.55%, with a training time of 0.0061 seconds.

Keywords: sentiment analysis, electric vehicles, Twitter, Naïve Bayes Classifier, Information gain

ABSTRAK

Penggunaan kendaraan listrik semakin meningkat di seluruh dunia sebagai solusi mengurangi emisi karbon dan ketergantungan pada bahan bakar fosil. Namun, adopsi teknologi ini menghadapi berbagai tantangan, termasuk penerimaan dari masyarakat. Melihat banyaknya opini yang ada di Twitter, penting untuk melakukan klasifikasi sentimen untuk memudahkan identifikasi opini terhadap kendaraan listrik. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap penggunaan kendaraan listrik melalui data dari Twitter. Penelitian ini menggunakan data *tweet* sebanyak 4174 dengan 1903 sentimen positif, 1603 sentimen negatif, dan 668 sentimen netral. Metode yang digunakan dalam penelitian ini adalah *Naïve Bayes Classifier* dan seleksi fitur *Information gain*. Hasil penelitian dengan model *Naïve Bayes Classifier* yang diterapkan dengan *Information gain* mendapatkan *accuracy*, *precision*, *recall*, dan *f1-score* tertinggi pada rasio 80:20 sebesar 61.77%, 40.1%, 47.8%, dan 43.13%, dengan waktu latih selama 0.0054 detik. Sedangkan dengan hanya menggunakan metode *naive bayes* menfapatkan nilai sebesar 60.2%, 40.75%, 46.7% dan 42.55%, dengan waktu latih selama 0.0061 detik.

Kata kunci : analisis sentimen, kendaraan listrik, Twitter, *Naïve Bayes Classifier*, *Information gain*

KATA PENGANTAR

Puji dan Syukur kehadirat Allah SWT atas rahmat, nikmat, dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul **“Analisis Sentimen Pada Media Sosial Twitter Terhadap Penggunaan Kendaraan Listrik Dengan Menggunakan Metode *Naïve Bayes Classifier* Dan *Information gain*”**. Tugas Akhir ini disusun dalam rangka memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika Fakultas Ilmu Komputer, Universitas Sriwijaya.

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan dan bantuan baik dalam segi materil maupun moril selama proses pembuatan tugas akhir ini. Adapun penulis merujuk secara khusus pihak yang telah membantu sebagai berikut :

1. Kedua orang tua saya dan kedua kakak saya, yang selalu memberikan dukungan, bantuan serta kritik dan saran.
2. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika.
4. Ibu Rizki Kurniati S.Kom., M.T. selaku pembimbing akademik yang telah memberikan bimbingan serta arahan dalam urusan akademik.
5. Ibu Rizki Kurniati S.Kom., M.T. dan ibu Desty Rodiah, S.Kom., M.T. selaku dosen pembimbing I dan dosen pembimbing II yang telah memberikan arahan, bimbingan, kritik, dan saran kepada penulis sehingga dapat menyelesaikan tugas akhir.
6. Seluruh Bapak dan Ibu Dosen Program Studi Teknik Informatika yang telah memberikan ilmu kepada penulis selama perkuliahan.
7. Staff Jurusan Teknik Infomatika beserta staff Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi dan akademik selama masa perkuliahan.
8. Teman - temanku “ERSIZ” yang yang selalu memberikan semangat, masukan, serta hiburan selama perkuliahan.

9. Teman - teman kelas Reguler angkatan 2020 yang selalu membagikan informasi dan menciptakan banyak momen kebersamaan selama perkuliahan yang tentu saja tidak akan pernah dilupakan.
10. Semua pihak lainnya yang tidak dapat disebutkan satu persatu yang telah berperan bagi penulis dalam menyelesaikan tugas akhir, penulis ucapkan Terima kasih banyak atas semuanya.

Semoga tugas akhir ini dapat menjadi manfaat bagi penulis mau pun pembaca pada umumnya, serta dapat menjadi referensi dan rujukan untuk peneliti-peneliti selanjutnya. Penulis menyadari bahwa dalam proses penyelesaian tugas akhir ini, masih terdapat beberapa kekurangan, Oleh karena itu, penulis berharap akan kritik dan saran yang membangun agar tugas akhir ini menjadi karya tulis yang lebih baik agar dapat menjadi bekal pengetahuan yang bermanfaat bagi peneliti di masa depan.

Indralaya, 08 Oktober 2024

Ivando Sibarani

DAFTAR ISI

Motto dan Persembahan	i
<i>ABSTRACT</i>	ii
ABSTRAK	iii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-4
1.5 Manfaat penelitian	I-4
1.6 Batasan Masalah	I-5
1.7 Sistematika penulisan	I-5
1.8 Kesimpulan	I-6
BAB II LANDASAN TEORI	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 <i>Text Mining</i>	II-1
2.2.2 <i>Text Preprocessing</i>	II-1
2.2.3 Analisis Sentimen	II-3
2.2.4 TF-IDF	II-3
2.2.5 Seleksi fitur	II-4
2.2.6 <i>Information gain</i>	II-4
2.2.7 <i>Naive bayes Classifier</i>	II-6
2.2.8 Evaluasi	II-8
2.2.9 Metode Waterfall	II-12
2.3 Penelitian yang relevan	II-13
2.4 Kesimpulan	II-15

BAB III METODOLOGI PENELITIAN	III-1
3.1 Pendahuluan	III-1
3.2 Deskripsi Umum	III-1
3.3 Dataset	III-2
3.4 Pengumpulan data	III-3
3.5 Preprocessing Data	III-3
3.6 <i>TF-IDF</i>	III-7
3.7 <i>Information gain</i>	III-10
3.8 <i>Naïve Bayes Classifier</i>	III-16
3.9 Pengembangan Perangkat Lunak	III-21
3.10 Menejemen Proyek Penelitian	III-23
3.11 Kesimpulan	III-27
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 Metode Pengembangan Perangkat Lunak	IV-1
4.2.1 <i>Requirement</i>	IV-1
4.2.2 <i>System Design</i>	IV-3
4.2.3 <i>Coding and Testing</i>	IV-14
4.2.4 Implementasi	IV-21
4.2.5 <i>Maintenance</i>	IV-26
4.3 Kesimpulan	IV-28
BAB V HASIL DAN PEMBAHASAN	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Pengujian	V-1
5.3 Analisis Data Hasil Pengujian	V-3
5.4 Kesimpulan	V-7
BAB VI KESIMPULAN DAN SARAN	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran	VI-2
DAFTAR PUSTAKA	iv

DAFTAR TABEL

Tabel II- 1 Tabel Confusion Matrix	II-8
Tabel III- 1 Dataset	III-2
Tabel III- 2 Case Folding	III-4
Tabel III- 3 Cleaning	III-4
Tabel III- 4 Tokenizing	III-5
Tabel III- 5 Normalisasi	III-5
Tabel III- 6 Stopword Removal	III-6
Tabel III- 7 Stemming	III-6
Tabel III- 8 Tabel TF.....	III-7
Tabel III- 9 Tabel DF	III-8
Tabel III- 10 TF-IDF.....	III-9
Tabel III- 11 Data Training	III-11
Tabel III- 12 Perhitungan Entropy data	III-12
Tabel III- 13 Perhitungan Entropy Kata.....	III-12
Tabel III- 14 Perhitungan Entropy negasi Kata	III-13
Tabel III- 15 Nilai Gain.....	III-15
Tabel III- 16 Nilai n	III-17
Tabel III- 17 Nilai n1	III-17
Tabel III- 18 Nilai $P(a_i v_j)$	III-18
Tabel III- 19 Data Testing Naïve Bayes	III-18
Tabel III- 20 Fitur setelah Information gain	III-19
Tabel III- 21 Hasil Klasifikasi Naïve Bayes	III-21
Tabel III- 22 Jadwal Penelitian dalam <i>Work Breakdown Structure</i> (WBS).....	III-23
Tabel IV- 1 Rincian Dataset.....	IV-2
Tabel IV- 2 Skenario preprocess data	IV-3
Tabel IV- 3 Skenario Analisis sentiment teks.....	IV-6
Tabel IV- 4 function dari preprocess_ignb.py	IV-15
Tabel IV- 5 function dari classify.py	IV-17
Tabel IV- 6 function dari app.py.....	IV-19
Tabel IV- 7 Rencana Pengujian Preprocess Data dan Analisis Sentimen	IV-20
Tabel IV- 8 Rencana Pengujian Klasifikasi teks.....	IV-20
Tabel IV- 9 Pengujian Preprocess Data dan analisis sentimen	IV-24
Tabel IV- 10 Pengujian Klasifikasi teks	IV-25
Tabel V- 1 Rincian data Pengujian	V-1
Tabel V- 2 Hasil pengujian	V-2

DAFTAR GAMBAR

Gambar II- 1 Metode Waterfall.....	II-12
Gambar III- 1 Alur Program	III-1
Gambar III- 2 Information gain.....	III-11
Gambar III- 3 Naïve Bayes	III-16
Gambar IV- 1 Use case Diagram	IV-3
Gambar IV- 2 Activity diagram Preprocess Data dan Sentiment Analisis	IV-9
Gambar IV- 3 Activity Diagram Analisis sentimen teks	IV-10
Gambar IV- 4 Analisis Sentimen Naïve Bayes.....	IV-13
Gambar IV- 6 Sequence Diagram sentimen analisis teks	IV-13
Gambar IV- 7 Class Diagram.....	IV-14
Gambar IV- 32 Interface halaman awal	IV-21
Gambar IV- 33 Interface setelah input file	IV-22
Gambar IV- 34 Interface setelah tombol 'Preprocess Data' ditekan	IV-22
Gambar IV- 35 Hasil Sentimen Analisis Naive bayes	IV-22
Gambar IV- 36 Sentimen Analisis Information gain dan Naive bayes.....	IV-23
Gambar IV- 37 Interface awal klasifikasi teks.....	IV-23
Gambar IV- 38 Interface setelah input teks dan menekan 'analisis'	IV-23
Gambar V- 1 Diagram Perbandingan Hasil Evaluasi	V-3

BAB I PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan ini membahas latar belakang masalah, rumusan masalah, tujuan, manfaat, batasan masalah, sistematika penulisan serta kesimpulan dari skripsi. Bab ini akan memberikan penjelasan umum mengenai keseluruhan isi laporan. Pendahuluan dimulai dengan penjelasan mengenai masalah yang ada dan bagaimana penyelesaian suatu masalah.

1.2 Latar Belakang

Penggunaan kendaraan listrik semakin mendapatkan perhatian di seluruh dunia sebagai solusi untuk mengurangi emisi karbon dan ketergantungan pada bahan bakar fosil (Alfarizi dan Fitriani. 2023). Namun, tantangan yang dihadapi dalam mengadopsi teknologi kendaraan listrik tidak hanya sebatas pada aspek tersebut, tetapi juga mencakup pada penerimaan dari masyarakat (Ernawati et al. 2023).

Twitter sebagai platform media sosial yang populer di Indonesia, dengan 14.75 juta pengguna aktif pada April 2023, menjadi sumber data yang potensial untuk memahami opini publik¹. Oleh karena itu, pemahaman mendalam tentang sentimen, dan sikap masyarakat terhadap kendaraan listrik sangatlah penting untuk melihat respon masyarakat, pengembangan kebijakan publik, merancang strategi yang efektif dalam pemasaran dan mendorong adopsi teknologi ini.

¹ Databoks “Jumlah Pengguna Twitter di Indonesia Capai 14,75 Juta per April 2023”(<https://databoks.katadata.co.id/datapublish/2023/05/31/jumlah-pengguna-twitter-di-indonesia-capai-1475-juta-per-april-2023-peringkat-keenam-dunia>, diakses pada 15 Maret 2024)

Analisis sentimen menjadi hal yang efektif dalam menggali data opini publik. Dengan mengumpulkan dan mengolah opini yang tersebar di *Twitter*, analisis sentimen dapat membantu mengklasifikasikan respons masyarakat ke dalam kategori positif, negatif, atau netral (Alfarizi dan Fitriani, 2023).

Dalam analisis sentimen, salah satu metode klasifikasi yang umum digunakan adalah *Naïve Bayes Classifier*. Algoritma ini didasarkan pada teorema Bayes, yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Metode ini menggunakan konsep probabilitas dan statistik, dimana *Naïve Bayes Classifier* mencari nilai probabilitas tertinggi untuk mengelompokkan data uji ke dalam kategori yang tepat (Ramadhani dan Suryono. 2024). Namun, ada masalah umum dalam klasifikasi teks yang disebabkan oleh banyaknya dimensi ruang fitur. Pemilihan fitur dapat menjadi solusi untuk masalah ini dengan *Information gain*. Dengan menggunakan *Information gain*, fitur yang kurang atau tidak relevan akan dihilangkan, sehingga dimensi data berkurang dan dapat meningkatkan kinerja (Anggita dan Abdulloh. 2023). Beberapa penelitian telah menunjukkan bahwa penerapan *Information gain* pada algoritma klasifikasi mampu meningkatkan akurasi yang dihasilkan.

Pada penelitian yang berjudul “Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma *naive bayes* dengan Seleksi Fitur *Information gain* dan *Particle Swarm Optimization* ” oleh Salman Alfarizi dan Eka Fitriani pada 2023. Hasil yang diperoleh dengan metode *Naïve Bayes* sebesar 79.43%. Hasil yang didapatkan setelah menggunakan algoritma *Naïve Bayes*, *Information gain* dan *PSO* meningkat menjadi 84.54%.

Penelitian selanjutnya dilakukan oleh Amelia Isnanda, Yuyun Umidah, dan Jajam Haerul Jaman pada 2021, dengan judul “Implementasi *Naïve Bayes Classifier* Dan *Information gain* Pada Analisis Sentimen Penggunaan *E-Wallet* Saat Pandemi”. Hasil akurasi yang didapat dengan metode *naïve bayes* sebesar 84%. Sedangkan hasil akurasi dengan menggunakan *Naïve Bayes* dan *Information gain* sebesar 92%.

Penelitian lainnya dengan judul “Perbandingan Metode Klasifikasi *Support Vector Machine* Dan *Naïve Bayes* Pada Analisis Sentimen Kendaraan Listrik” oleh Ni Wayan Ernawati, I Nyoman Satya Kumara, dan Widyadi Setiawan pada 2023. Pada penelitian ini hasil *accuracy* yang didapatkan dengan metode *Naïve Bayes* sebesar 82% dan hasil dari SVM sebesar 81%. Dengan waktu latih yang dibutuhkan oleh metode SVM yaitu 37.42 detik sedangkan metode *naive bayes* hanya membutuhkan 0.10 detik. Hal ini membuktikan *Naïve Bayes* menjadi metode penelitian yang baik karena memiliki waktu latih yang cepat dan akurasi yang tinggi.

Tujuan dari penelitian ini adalah untuk mengetahui sentimen yang diberikan oleh masyarakat terhadap kendaraan listrik dengan menggunakan metode *naive bayes* dan seleksi fitur *information gain*.

1.3 Rumusan Masalah

1. Bagaimana membangun sebuah sistem analisis sentimen masyarakat terkait penggunaan kendaraan listrik dapat dianalisis melalui data yang terdapat pada media sosial Twitter dengan menggunakan metode *Naïve Bayes* dan *Informaiton Gain*?

2. Bagaimana akurasi model *Naïve Bayes* dalam menganalisis sentiment terhadap penggunaan kendaraan Listrik di twitter setelah diterapkan metode *Information gain*?

1.4 Tujuan Penelitian

1. Tujuan dari penelitian ini adalah untuk membangun sebuah sistem yang mampu menganalisis sentimen masyarakat terkait penggunaan kendaraan listrik berdasarkan data yang diambil dari media sosial Twitter. Sistem ini akan dibangun dengan menggunakan metode *Naïve Bayes Classifier* dan teknik seleksi fitur *Information gain*, sehingga dapat meningkatkan akurasi analisis sentimen terhadap penggunaan kendaraan listrik.
2. Mengetahui model *Naïve Bayes Classifier* dalam analisis sentimen setelah diterapkan metode *Information gain*.

1.5 Manfaat penelitian

1. Dapat mengetahui dan mempelajari hasil kinerja dari metode *naive bayes classifier* dalam melakukan analisis sentimen setelah ditambahkan seleksi fitur *information gain*.
2. Hasil penelitian dapat digunakan oleh industri kendaraan listrik untuk mengetahui persepsi publik mengenai kendaraan listrik.
3. Hasil penelitian dapat digunakan sebagai dasar untuk penelitian lanjutan dalam bidang yang sama dan berhubungan.

1.6 Batasan Masalah

1. Data yang dianalisis adalah data dari media sosial *Twitter*.
2. Kategori sentimen analisis dibagi menjadi 3, yaitu positif, negatif, dan netral.

1.7 Sistematika penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan

BAB II KAJIAN LITERATUR

Pada bab ini membahas mengenai dasar-dasar teori yang digunakan dalam penelitian, seperti definisi dari *Text Mining*, *Text Preprocessing*, analisis sentimen, *Information gain*, *naive bayes*, evaluasi, dan metode *waterfall*.

BAB III METODOLOGI PENELITIAN

Pada bab ini membahas mengenai tahapan yang dilakukan pada penelitian. Setiap rencana tahapan penelitian dideskripsikan dengan rinci dengan berfokus pada satu kerangka kerja. Di akhir bab berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Bab ini menjelaskan proses pengembangan perangkat lunak yang digunakan dalam penelitian ini. Pembahasannya meliputi metode pengembangan perangkat lunak, tahap-tahap yang dilalui mulai dari analisis kebutuhan, perancangan sistem, implementasi, pengujian, hingga pemeliharaan. Setiap tahap dijelaskan secara

terperinci untuk memberikan gambaran lengkap tentang bagaimana perangkat lunak dikembangkan untuk keperluan analisis sentimen pada media sosial Twitter terhadap kendaraan listrik.

BAB V HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil dari pengujian perangkat lunak yang telah dikembangkan. Analisis terhadap hasil pengujian dijelaskan secara mendalam, termasuk data yang diperoleh dari pengujian *naïve bayes classifier* dengan dan tanpa seleksi fitur menggunakan *information gain*. Bab ini juga membahas perbandingan hasil akurasi, presisi, recall, dan F1-score, serta interpretasi dari hasil tersebut dalam konteks sentimen masyarakat terhadap kendaraan listrik.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan, yaitu mengenai efektivitas penggunaan *naïve bayes classifier* dan *information gain* dalam analisis sentimen terhadap kendaraan listrik. Selain itu, bab ini juga menyajikan saran yang dapat dijadikan acuan untuk penelitian lebih lanjut, termasuk pengembangan metode lain yang mungkin dapat meningkatkan performa klasifikasi atau penerapan dataset yang lebih besar untuk hasil yang lebih representatif.

1.8 Kesimpulan

Dalam penelitian ini, analisis sentimen pada media sosial *Twitter* terhadap penggunaan kendaraan listrik dilakukan menggunakan metode *naïve bayes classifier* dan *information gain*. Tujuan utama penelitian adalah untuk mengidentifikasi sentimen pengguna *Twitter* terhadap kendaraan listrik,

mengimplementasikan metode *naïve bayes classifier*, dan menganalisis efektivitas *information gain* dalam meningkatkan akurasi klasifikasi sentimen.

Dengan demikian, penelitian ini tidak hanya memberikan pemahaman yang lebih baik mengenai sentimen pengguna terhadap kendaraan listrik, tetapi juga menyediakan wawasan yang berguna bagi industri kendaraan listrik dalam merancang strategi pemasaran dan pengembangan produk. Selain itu, penelitian ini juga memberikan kontribusi terhadap pengembangan teknologi analisis sentimen, yang dapat bermanfaat bagi berbagai aplikasi lainnya dalam pemrosesan bahasa alami.

BAB II

LANDASAN TEORI

2.1 Pendahuluan

Disini akan menjelaskan secara terperinci tentang landasan teori dari konsep-konsep yang digunakan dalam penelitian ini, serta memberikan penelitian-penelitian lain yang relevan.

2.2 Landasan Teori

2.2.1 *Text Mining*

Text mining mempunyai pengertian yang sama dengan data *mining*, namun tidak untuk beberapa metode dan data yang dikelolanya seperti data teks yang tidak terstruktur, terstruktur sebagian, maupun terstruktur seperti teks email, teks HTML, maupun teks komentar dari berbagai sumber. *Text Mining* merupakan pengetahuan di *database* dalam bentuk tekstual atau *knowledge discovery in textual database* (KDT), merupakan pencarian data-data yang berbentuk teks yang merupakan ketertarikan terhadap pengetahuan yang baru dibuat, diartikan sebagai bagian dari proses pencarian data teks yang sebelumnya tidak diketahui (Firdaus, & Istalma, 2021).

2.2.2 *Text Preprocessing*

Text Preprocessing merupakan salah satu bagian dalam proses *Text Mining*, proses mengolah teks yang bertujuan untuk mengubah bentuk dokumen menjadi data yang terstruktur sesuai dengan kepentingannya agar bisa diproses lebih lanjut pada proses *text mining*. Langkah *preprocessing* dalam klasifikasi bertujuan untuk memaksimalkan akurasi dalam klasifikasi data (Ridwansyah, 2022). Ada beberapa proses dalam *Text Preprocessing* sebagai berikut (Rahman Isnain et al., 2021).

2.2.2.1 Case Folding

Proses *Case Folding* melibatkan pemisahan setiap kata dari teks, dengan mengubah semua huruf menjadi huruf kecil. Hasil dari proses *Case Folding* adalah kata yang tidak mengandung karakter selain huruf kecil.

2.2.2.2 Cleaning

Cleaning merupakan proses yang bertujuan untuk menghapus berbagai informasi yang tidak diperlukan dalam analisis sentimen, seperti symbol, angka, tanda baca, link (http, https, pic.twitter), *hashtag*, *username*, dan *retweet*.

2.2.2.3 Tokenizing

Tokenizing adalah proses pembagian suatu teks besar menjadi beberapa bagian, yang dapat berupa kalimat. Kalimat yang dihasilkan kemudian dibagi menjadi kata-kata.

2.2.2.4 Normalisasi

Normalisasi adalah proses yang bertujuan untuk memperbaiki kata-kata yang mungkin memiliki kesalahan penulisan atau pengejaan, serta kata-kata yang dituliskan dengan singkatan.

2.2.2.5 Stemming

Stemming adalah proses untuk mengubah kata-kata hasil *filtering* menjadi kata dasar dengan cara menghilangkan semua imbuhan, baik itu awalan, akhiran, atau sisipan. Kombinasi awalan dan akhiran pada kata turunan juga harus dihilangkan. *Stemming* digunakan untuk menyederhanakan kata-kata dari bentuk asli yang masih mengandung imbuhan menjadi kata dasar (Ulgasesa et al., 2021).

2.2.2.6 Stopword Removal

Stopword merupakan proses untuk menghilangkan kata-kata yang dianggap tidak penting dari hasil Tokenisasi. *Stopword* digunakan dengan memeriksa setiap kata dari hasil *Case Folding*, dan jika kata tersebut terdapat dalam daftar *stopword*, maka kata tersebut dihapus (Nofiyani dan Wulandari. 2021).

2.2.3 Analisis Sentimen

Analisis sentimen merupakan proses yang dilakukan untuk memahami, menganalisis dan mengklasifikasikan informasi yang terkandung dalam suatu teks dengan analisis teks. Analisis sentimen digunakan untuk mengevaluasi dan memahami opini yang terdapat dalamn teks dari berbagai sumber (Ernwati et al., 2023).

Analisis senitmen adalah merupakan sebuah proses pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistic dan *text mining* yang memiliki tujuan menganalisa pendapat, sentimen, dan sikap dari seseorang (Novina dan Rasal, 2023). Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seorang, apakah beropini positif, negatif atau netral.

2.2.4 TF-IDF

TF-IDF adalah salah satu teknik dalam pengolahan teks untuk memberikan bobot pada kata-kata dalam sebuah data. TF-IDF bertujuan untuk mengidentifikasi fitur yang paling penting dalam suatu data. *Term Frequency* (TF) adalah nilai frekuensi kemunculan suatu kata dalam sebuah dokumen. Fitur yang sering muncul dalam suatu data, seperti kata hubung atau kata umum cenderung memiliki nilai TF yang tinggi, tetapi tidak memiliki makna yang penting dalam data tersebut. Oleh

karena itu, dibutuhkan teknik lain, yaitu *Inverse Document Frequency* (IDF) yang memberikan bobot pada kata-kata yang jarang muncul di dokumen (Wati et al., 2023). Nilai TF dan IDF dapat dihitung dengan menggunakan rumus:

$$TF = \frac{\text{jumlah kemunculan fitur dalam teks}}{\text{jumlah kata dalam teks}} \quad (\text{II-1})$$

$$IDF = \log \frac{N}{m} \quad (\text{II-2})$$

Keterangan

N adalah jumlah dokumen dalam kumpulan dokumen

m adalah jumlah dokumen yang mengandung kata tersebut.

Setelah nilai TF dan IDF diperoleh, nilai TF-IDF dapat dihitung dengan.

$$TF - IDF = TF \times IDF \quad (\text{II-3})$$

Fitur yang memiliki nilai TF-IDF tinggi dianggap penting dan memberikan kontribusi yang lebih besar dalam menentukan topik dokumen atau kumpulan dokumen tersebut.

2.2.5 Seleksi fitur

Seleksi fitur adalah suatu metode yang digunakan untuk mengurangi dimensi atribut dengan memilih sejumlah atribut yang dianggap relevan untuk proses klasifikasi. Terlalu banyak atribut, terutama jika tidak relevan, dapat mempengaruhi efektivitas pengenalan pola selama tahap pemrosesan menggunakan metode yang telah ditentukan (Nur et al., 2022). Tujuan utama dari seleksi fitur adalah untuk memilih fitur terbaik dari suatu kumpulan fitur.

2.2.6 *Information gain*

Information gain biasanya digunakan untuk menilai dan menentukan atribut

mana yang dianggap berpengaruh terhadap kelas. *Information gain* digunakan untuk menghasilkan atribut yang berkaitan terhadap kelas target, karena setiap atribut memiliki nilai dan dipilih dengan nilai yang terbaik (Nur et al., 2022).

Information gain dihitung berdasarkan pengaruh fitur terhadap keseragaman entropy pada data yang dibagi jadi subdata dengan nilai fitur tertentu. Keseragaman entropy dihitung pada data sebelum dipecah dengan persamaan II-1 dan pada data setelah dipecah dengan persamaan II-2 berikut ini.

$$\text{entropy}(S) = \sum_{i=1}^k (P_i) \log_2(P_i) \quad (\text{II-4})$$

Keterangan:

- Entropy (S) : nilai entropy data S
 P_i : probabilitas kelas i dalam Data.
 K : jumlah kelas.

$$\text{Entropy}(S, A) = \sum_{v=1}^v \left(\frac{S_v}{S} * \text{Entropy}(S_v) \right) \quad (\text{II-5})$$

Keterangan

- S : seluruh nilai yang mungkin dari atribut A.
 S_v : subset dari S dimana atribut A bernilai v.
 v : jumlah nilai unik beratribut A.
 Entropy (S, A) : nilai entropy setelah dipisah berdasarkan A.

Nilai dari *information gain* dapat dihitung dengan persamaan berikut

$$Gain(S, A) = Entropy(S) - Entropy(S, A) \quad (II-6)$$

Keterangan:

Gain (S,A) : nilai *information gain*.

2.2.7 Naive bayes Classifier

Fungsi lain dari *Data Mining* adalah untuk mengklasifikasikan data, yaitu mengelompokkan data ke dalam satu atau beberapa kelas yang telah ditentukan. Salah satu metode yang digunakan dalam klasifikasi data adalah *Naive Bayes*, sebuah metode yang menggunakan perhitungan probabilitas yang dikembangkan oleh ilmuwan Inggris Thomas Bayes. *Naive Bayes* bekerja dengan memprediksi probabilitas kejadian di masa depan berdasarkan data dari kejadian-kejadian sebelumnya (Darwis et al).

Teorema Bayes merupakan torema yang mengacu pada konsep probabilitas bersyarat. Teorema Bayes dapat dituliskan dengan:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (II-7)$$

Keterangan

$P(A|B)$: Probabilitas dari kelas A yang diberikan Fitur B

$P(B|A)$: Probabilitas mengamati fitur B dari kelas A

$P(A)$: Probabilitas dari kelas A

$P(B)$: Probabilitas fitur B

Pada *naive bayes*, setiap *tweet* direpresentasikan dalam atribut

$(a_1, a_2, a_3, \dots, a_n)$ dimana a_1 merupakan kata pertama, a_2 merupakan kata kedua, a_3 merupakan kata ketiga dan seterusnya, sedangkan V merupakan himpunan kelas. Saat proses klasifikasi, *Naive Bayes* akan menghasilkan kategori / kelas yang paling tinggi probabilitasnya (V_{MAP}) dengan memasukkan atribut $(a_1, a_2, a_3, \dots, a_n)$. Untuk persamaan V_{MAP} dapat ditulis dengan:

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2, a_3, \dots, a_n) \quad (\text{II-8})$$

Dengan menggunakan teorema Bayes, maka dapat ditulis menjadi

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} \frac{P(v_j | a_1, a_2, a_3, \dots, a_n) P(v_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (\text{II-9})$$

$P(a_1, a_2, a_3, \dots, a_n)$ memiliki nilai yang konstan pada semua v_j sehingga dapat ditulis dengan

$$V_{MAP} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1, a_2, a_3, \dots, a_n) P(v_j) \quad (\text{II-10})$$

Keterangan :

V_{MAP} : Semua Kategori yang diujikan

$P(v_j | a_1, a_2, a_3, \dots, a_n)$: probabilitas kelas v_j yang diberi fitur

$a_1, a_2, a_3, \dots, a_n$

$P(v_j)$: Probabilitas dari v_j

$P(a_1, a_2, a_3, \dots, a_n)$: probabilitas dari fitur

Naive bayes menyederhanakan hal ini dengan mengasumsikan bahwa setiap kategori, setiap atribut bebas bersyarat satu dengan yang lain. Maka:

$$P(a_1, a_2, a_3, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \quad (\text{II-11})$$

Jika kedua persamaan diatas digabungkan maka akan menghasilkan persamaan :

$$V_{MAP} = \underset{v_j \in V}{\text{argmax}} P(v_j) \times \prod_i P(a_i | v_j) \quad (\text{II-12})$$

$P(v_j)$ dan probabilitas kata a_i untuk setiap kategori yang dihitung pada saat *training* dapat dirumuskan dengan

$$P(v_j) = \frac{\text{docs}_j}{\text{training}} \quad (\text{II-13})$$

$$P(a_i | v_j) = \frac{n_i + 1}{n + \text{jumlah kata unik}} \quad (\text{II-14})$$

Keterangan:

docs_j : banyaknya dokumen pada kategori j

training : banyaknya fitur yang digunakan dalam proses *training*

$P(a_i | v_j)$: probabilitas a_i pada kategori v_j

n : banyaknya kata yang muncul pada kategori v_j

n_i : n_i adalah jumlah kemunculan kata a_i pada kategori v_j

jumlah kata unik : jumlah kata unik pada data *training*

2.2.8 Evaluasi

Tahap evaluasi bertujuan untuk menilai kinerja algoritma klasifikasi yang digunakan dalam penelitian (Tanggreani, & Sitokdana, 2022). Kemampuan prediksi diukur berdasarkan nilai dari Confusion Matrix, akurasi, *precision*, *recall*, dan *F1-Score* (Ramadhani & Suryono, 2024).

Tabel II- 1 *Tabel Confusion Matrix*

		Prediction Class		
		Positif	Negatif	Netral
Actual Class	Positif	A	B	C
	Negatif	D	E	F
	Netral	G	H	I

Matriks tersebut memiliki nilai yang dijadikan acuan dalam perhitungan, yaitu:

- a. 'A' adalah jumlah contoh yang sebenarnya positif dan diprediksi positif.
- b. 'B' adalah jumlah contoh yang sebenarnya positif tetapi diprediksi negatif.
- c. 'C' adalah jumlah contoh yang sebenarnya positif tetapi diprediksi netral.
- d. 'D' adalah jumlah contoh yang sebenarnya negatif tetapi diprediksi positif.
- e. 'E' adalah jumlah contoh yang sebenarnya negatif dan diprediksi negatif.
- f. 'F' adalah jumlah contoh yang sebenarnya negatif tetapi diprediksi netral.
- g. 'G' adalah jumlah contoh yang sebenarnya netral tetapi diprediksi positif.

- h. 'H' adalah jumlah contoh yang sebenarnya netral tetapi diprediksi negatif.
- i. 'I' adalah jumlah contoh yang sebenarnya netral dan diprediksi netral.

Untuk menentukan model terbaik, penelitian ini menganalisis nilai *Classification report*. Dalam *Classification report*, terdapat nilai *accuracy*, *precision*, *recall*, dan *F1-Score* yang digunakan sebagai acuan untuk memilih model terbaik.

Untuk akurasi digunakan untuk mengetahui seberapa akurat sistem dapat melakukan klasifikasi. Akurasi dapat dihitung dengan persamaan (Husada & Paramita, 2021):

$$Accuracy = \frac{A + E + I}{A + B + C + D + E + F + G + H + I} \quad (II-15)$$

Precision berfungsi untuk menghitung berapa banyak prediksi pada kelas yang benar dari semua prediksi pada kelas. Untuk menghitung *precision* dapat menggunakan.

$$Precision\ positif = \frac{A}{A + D + G} \quad (II-16)$$

$$Precision\ negatif = \frac{E}{B + E + H} \quad (II-17)$$

$$Precision\ netral = \frac{I}{C + F + I} \quad (II-18)$$

Precision – avg

$$= \frac{\textit{Precision Positif} + \textit{precision negatif} + \textit{precision netral}}{3} \quad (\text{II-19})$$

Recall berfungsi untuk mengukur berapa banyak kelas yang berhasil diidentifikasi dari semua kelas yang sebenarnya. Untuk menghitung *recall* dapat menggunakan.

$$\textit{Recall positif} = \frac{A}{A + B + C} \quad (\text{II-20})$$

$$\textit{Recall negatif} = \frac{E}{D + E + F} \quad (\text{II-21})$$

$$\textit{Recall netral} = \frac{I}{G + H + I} \quad (\text{II-22})$$

Recall – avg

$$= \frac{\textit{Recall Positif} + \textit{Recall negatif} + \textit{Recall netral}}{3} \quad (\text{II-23})$$

F1-Score berguna untuk memberikan satu nilai untuk menyeimbangkan *recall* dan *precision*. untuk menghitung f1-score dapat menggunakan.

$$\text{F1 – Score positif} = 2 \times \frac{\textit{Precision positif} \times \textit{recall positif}}{\textit{Precision positif} + \textit{recall positif}} \quad (\text{II-24})$$

F1 – Score negatif

$$= 2 \times \frac{\textit{Precision negatif} \times \textit{recall negatif}}{\textit{Precision negatif} + \textit{recall negatif}} \quad (\text{II-25})$$

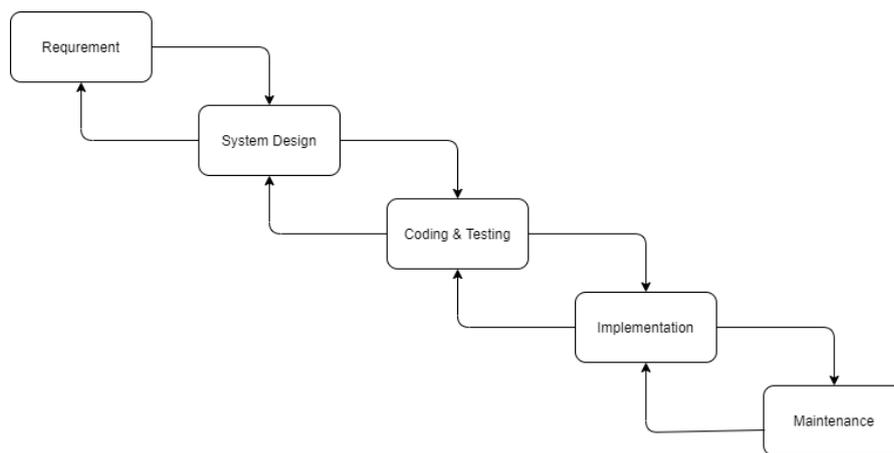
(II-26)

$$F1 - Score\ netral = 2 \times \frac{Precision\ netral \times recall\ netral}{Precision\ netral + recall\ netral}$$

$$F1 - Score - avg = \frac{F1 - Score\ Positif + F1 - Score\ negatif + F1 - Score\ netral}{3} \quad (II-27)$$

2.2.9 Metode Waterfall

Metode penelitian yang digunakan dalam penelitian ini merupakan metode *waterfall*. Karena dalam pengembangannya bersifat sistematis dan sekuensial. Selain itu metode *waterfall* dilakukan secara berurutan dan berkelanjutan. Berikut beberapa tahapan-tahapan yang dilakukan dalam penelitian ini: *requirement, system design, Coding & Testing, implementation, maintenance* (Mahardika et al., 2023).



Gambar II- 1 Metode Waterfall

Langkah-langkah dalam metode *waterfall* digambarkan sebagai berikut.

1. Requirement

Melakukan observasi data, menetapkan fitur, dan tujuan dari sistem dibuat. Semua Langkah ditetapkan secara detail dan digunakan sebagai

kualifikasi sistem.

2. *System Design*

Membuat sebuah desain sistem berdasarkan syarat sistem yang sudah di tetapkan. Pada tahap ini bertujuan untuk memberikan perancangan sistem yang harus dikerjakan.

3. *Coding & Testing*

Perancangan yang sudah di buat diterjemahkan dalam bahasa pemrograman menjadi serangkaian unit program. Kemudian dilanjutkan pengujian sistem pada setiap unit program.

4. *Implementation*

Pengintegrasian setiap unit sistem menjadi sebuah sistem utuh. Kemudian dilakukan pengujian program untuk memastikan kesesuaian syarat sistem.

5. *Maintenance*

Yang terakhir yaitu melakukan penerapan sistem. Dalam tahap ini kesalahan yang belum ditemukan pada tahap – tahap sebelumnya maka akan dilakukan perbaikan secara berkala.

2.3 Penelitian yang relevan

2.3.1 Analisis Sentimen Terhadap Kendaraan Listrik Pada Platform Twitter Menggunakan Metode *Naive bayes* (Alin et al., 2023)

Tujuan dari penelitian ini adalah untuk menganalisis sentimen masyarakat Indonesia terhadap kendaraan listrik melalui platform Twitter. Penelitian ini bertujuan untuk mengklasifikasikan opini masyarakat menjadi sentimen positif,

negatif, dan netral menggunakan metode Naïve Bayes. Dengan pendekatan ini, peneliti berharap dapat memahami kecenderungan pandangan masyarakat mengenai kendaraan listrik, yang diharapkan dapat memberikan wawasan penting bagi perkembangan teknologi kendaraan listrik di Indonesia.

2.3.2 Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma *Naive Bayes* dengan Seleksi Fitur *Information gain* dan *Particle Swarm Optimization* (Alfarizi & Fitriani, 2023)

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kendaraan listrik di media sosial Twitter. Dengan menggunakan algoritma *Naive Bayes* serta metode seleksi fitur *Information gain* dan *Particle Swarm Optimization* (PSO), penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi sentimen positif, negatif, dan netral terkait opini publik tentang kendaraan listrik. Hasil penelitian diharapkan memberikan informasi yang berguna bagi produsen kendaraan listrik dan pemerintah dalam pengembangan produk serta kebijakan yang mendukung penggunaan kendaraan listrik.

2.3.3 Perbandingan Metode Klasifikasi Support Vector Machine Dan Naïve Bayes Pada Analisis Sentimen Kendaraan Listrik (Ernawati et al., 2023)

Tujuan utama dari penelitian ini adalah untuk menganalisis sentimen masyarakat Indonesia terhadap kendaraan listrik menggunakan data dari Twitter. Penelitian ini juga bertujuan membandingkan dua metode klasifikasi, yaitu Support Vector Machine (SVM) dan Naïve Bayes, dari segi akurasi dan efisiensi waktu. Melalui analisis ini, diharapkan dapat diketahui metode yang paling efektif dalam

mengklasifikasikan sentimen positif, negatif, dan netral terhadap kendaraan listrik di Indonesia.

2.4 Kesimpulan

Secara keseluruhan, kajian literatur yang dibahas memberikan pemahaman yang komprehensif mengenai konsep-konsep utama dan metodologi yang digunakan dalam penelitian ini. Teori sentimen, analisis sentimen, dan metode *Naïve Bayes Classifier* yang diperkuat dengan seleksi fitur menggunakan *Information gain* menjadi kerangka dasar untuk analisis sentimen terhadap kendaraan listrik di Indonesia. Dengan pemahaman ini, penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam mengidentifikasi opini publik dan mendukung pengembangan kebijakan serta strategi pemasaran yang efektif untuk kendaraan listrik.

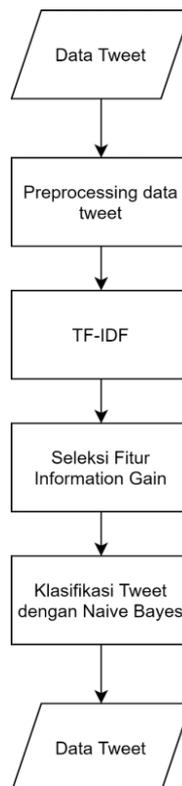
BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan

Pada bab ini membahas mengenai tahap-tahapan dan metode yang digunakan dalam menjalankan penelitian ini. Tahapan ini dimulai dengan deskripsi umum, dataset yang digunakan, sumber data, pengujian penelitian, dan metode pengembangan dari penelitian.

3.2 Deskripsi Umum

Sistem klasifikasi yang digunakan pada penelitian memiliki suatu rancangan tentang bagaimana alur ini akan berjalan. Gambaran tentang jalannya program akan dibentuk seperti gambar III-1



Gambar III- 1 Alur Program

Pada tahap *preprocessing data tweet*, data akan diproses terlebih dahulu sebelum dilakukan analisis lebih lanjut. Setelah data *tweet* diproses, tahap berikutnya adalah TF-IDF. Dimana, fitur dalam data diberikan bobot untuk mengidentifikasi fitur yang penting. Setelah itu, tahap seleksi fitur menggunakan teknik *Information gain*. *Information gain* bertujuan untuk mengurangi jumlah fitur yang digunakan dalam model klasifikasi sehingga model lebih efisien dan efektif. Pada tahap berikutnya, *tweet-tweet* yang telah diproses dan dipilih fitur-fiturnya akan diklasifikasikan menggunakan algoritma *naive bayes*. Hasil dari proses klasifikasi adalah data *tweet* yang telah diberi label kelas sesuai dengan model *naive bayes*. Data ini akan menunjukkan sentimen dari *tweet* berdasarkan model yang telah dibangun.

3.3 Dataset

Dataset yang digunakan merupakan dokumen berupa *Tweet* dari akun twitter yang merupakan opini. Dokumen yang dimasukkan adalah dokumen *tweet* berbahasa Indonesia seperti pada tabel III-1 berikut.

Tabel III- 1 Dataset

	<i>Tweet</i>	Klasifikasi
1	Pertanyaan saya, apakah mungkin produk mobil listrik digunakan untuk perusahaan sejenis ekspedisi? Berapa biaya dikeluarkan?	Netral
2	Mantap BAYAR LISTRIK PASTI SEMAKIN MAHAL .. TOKEN LISTRIK PASTI BARU 3 HARI AUTO BUNYI	Negatif
3	Semoga nanti kedepannya mobil dan motor listrik harganya akan lebih terjangkau alias murah	Positif
4	@dimsdalle @mouldie_sep Nunggu baterai murah. Bakalan terjadi kalo makin byk org pake kendaraan listrik juga	Negatif
5	Mungkin butuh waktu agak lama cari yg murah dan sporty	Netral

	untuk yg listrik	
6	Kalau pake kendaraan listrik gaperlu pajak mahal #PakaiMolis #ZonaUang #ZonaJajan https://t.co/FoWdDr4FGH	Positif
7	Klo produsennya banyak bgt Ya perang harga Awal awal saja mahal karna blom ada persaingan yg ketat	Netral
8	Untung nya apa utk rakyat?biaya hidup Bs turun ga?	Positif
9	Subsidi kalangan tengah ke atas.. Mending gratisin biaya berobat lebih manfaat	Negatif
10	Knpa ga bikin mobil listrik produk dalam negri saja.	Netral
...		

3.4 Pengumpulan data

Data yang digunakan dalam penelitian ini adalah data sekunder yang didapat dari kaggle. Dapat diakses melalui link: <https://www.kaggle.com/datasets/jovangtg13/kendaraan-listrik>. Data ini berupa dokumen *tweet* dengan format CSV. Data ini diklasifikasikan secara manual dengan label sentimen positif, negatif, dan netral.

3.5 Preprocessing Data

Tahapan penelitian adalah hal yang penting untuk tahap selanjutnya, yaitu mengurangi atribut yang kurang tidak berpengaruh terhadap proses pengklasifikasian. Data yang dimasukkan pada tahap ini masih berupa data *raw*, sehingga hasil dari proses ini adalah dokumen yang yang dapat mempermudah dalam proses pengklasifikasian. Proses ini terdiri dari beberapa tahapan.

3.5.1 Case Folding

Pada proses ini semua huruf kapital pada dokumen *tweet* diubah menjadi huruf kecil. Dengan tujuan untuk menghilangkan duplikasi data yang hanya

berbeda pada hurufnya saja. Sebagai gambaran dari *case folding*, berikut hasil dari proses.

Tabel III- 2 *Case Folding*

Input <i>Tweet</i>	Output <i>Tweet</i>
Mantap BAYAR LISTRIK	mantap bayar listrik pasti
PASTI SEMAKIN MAHAL	semakin mahal .. token
.. TOKEN LISTRIK	listrik pasti baru 3 hari auto
PASTI BARU 3 HARI AUTO	bunyi
BUNYI	

3.5.2 *Cleaning*

Atribut-atribut yang tidak berpengaruh tersebut akan dihapuskan dari dokumen kemudian akan digantikan dengan *space*. Sebagai gambaran dari *cleaning*, berikut hasil dari proses

Tabel III- 3 *Cleaning*

Input <i>Tweet</i>	Output <i>Tweet</i>
mantap bayar listrik pasti	mantap bayar listrik pasti
semakin mahal .. token	semakin mahal br br token
listrik pasti baru 3 hari auto	listrik pasti baru hari auto bunyi
bunyi	

3.5.3 Tokenizing

Tahap ini merupakan pemotongan kalimat berdasarkan kata yang menyusunnya menjadi kata tunggal. Kata yang dimaksud adalah kata yang dipisah oleh *space*. Hasil dari proses ini merupakan kata tunggal yang dimasukkan ke dalam *database* untuk pengklasifikasian. Sebagai gambaran dari *Tokenizing*, berikut hasil dari proses

Tabel III- 4 *Tokenizing*

Input <i>Tweet</i>	Output <i>Tweet</i>
mantap bayar listrik pasti	['mantap', 'bayar', 'listrik', 'pasti',
semakin mahal br br token	'semakin', 'mahal', 'br', 'br',
listrik pasti baru hari auto bunyi	'token', 'listrik', 'pasti', 'baru',
	'hari', 'auto', 'bunyi']

3.5.4 Normalisasi

Pada tahapan ini dilakukan pengubahan kata yang tidak sesuai dengan aturan EYD yang dapat mengurangi hasil sentimen dokumen. Pada tahapan ini dibagi menjadi tiga, yaitu konversi kata singkatan, konversi kata baku, dan konversi kata inggris. Sebagai gambaran dari Normalisasi, berikut hasil dari proses

Tabel III- 5 Normalisasi

Input <i>Tweet</i>	Output <i>Tweet</i>
['mantap', 'bayar', 'listrik', 'pasti',	['mantap', 'bayar', 'listrik', 'pasti',
'semakin', 'mahal', 'br', 'br',	'semakin', 'mahal', 'br', 'br',

'token', 'listrik', 'pasti', 'baru', 'hari', 'auto', 'bunyi']	'token', 'listrik', 'pasti', 'baru', 'hari', otomatis, 'bunyi']
--	--

3.5.5 *Stopword Removal*

Ditahap ini, merupakan tahap untuk menghilangkan kata yang tidak sesuai dengan topik. Kata yang akan dihilangkan dikumpulkan kedalam *data stopwords*. Jika dalam suatu dokumen *tweet* terdapat kata yang sesuai dengan kata yang berada dalam *stopword* maka kata tersebut akan dihilangkan dan diubah dengan karakter *space*. Sebagai gambaran dari *Stopword Removal*, berikut hasil dari proses

Tabel III- 6 *Stopword Removal*

Input <i>Tweet</i>	Output <i>Tweet</i>
['kalau', 'pakai', 'kendaraan', 'listrik', 'gaperlu', 'pajak', 'mahal']	['pakai', 'kendaraan', 'listrik', 'gaperlu', 'pajak', 'mahal']

3.5.6 *Stemming*

Pada tahap ini merupakan suatu proses untuk mengubah kata yang terdapat dalam suatu dokumen ke dalam kata akarnya dengan menggunakan aturan tertentu. Proses *stemming* bahasa Indonesia dilakukan dengan cara menghilangkan imbuhan pada dokumen. Sebagai gambaran dari *Stemming*, berikut hasil dari proses

Tabel III- 7 *Stemming*

Input <i>Tweet</i>	Output <i>Tweet</i>

['pakai', 'kendaraan', 'listrik', 'gaperlu', 'pajak', 'mahal']	['pakai', 'kendara', 'listrik', 'gaperlu', 'pajak', 'mahal']
--	--

3.6 TF-IDF

Pada proses ini, tiap kata dalam data akan diberikan bobot yang dianggap penting untuk klasifikasi teks. Untuk perhitungan TF-IDF dapat dilihat pada tabel-tabel berikut

Tabel III- 8 Tabel TF

kata	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
produk	1/7	0	0	0	0	0	0	0	0	1/4
mobil	1/7	0	1/8	0	0	0	0	0	0	1/4
listrik	1/7	1/6	1/8	1/6	1/5	1/6	0	0	0	1/4
usaha	1/7	0	0	0	0	0	0	0	0	0
jenis	1/7	0	0	0	0	0	0	0	0	0
biaya	1/7	0	0	0	0	0	0	1/5	1/6	0
keluar	1/7	0	0	0	0	0	0	0	0	0
mantap	0	1/6	0	0	0	0	0	0	0	0
bayar	0	1/6	0	0	0	0	0	0	0	0
mahal	0	1/6	0	0	0	1/6	1/7	0	0	0
token	0	1/6	0	0	0	0	0	0	0	0
bunyi	0	1/6	0	0	0	0	0	0	0	0
mpga	0	0	1/8	0	0	0	0	0	0	0
motor	0	0	1/8	0	0	0	0	0	0	0
Harga	0	0	1/8	0	0	0	1/7	0	0	0
jangkau	0	0	1/8	0	0	0	0	0	0	0
alias	0	0	1/8	0	0	0	0	0	0	0
murah	0	0	1/8	1/6	1/5	0	0	0	0	0
nunggu	0	0	0	1/6	0	0	0	0	0	0
baterai	0	0	0	1/6	0	0	0	0	0	0
pakai	0	0	0	1/6	0	1/6	0	0	0	0
kendara	0	0	0	1/6	0	1/6	0	0	0	0
butuh	0	0	0	0	1/5	0	0	0	0	0
cari	0	0	0	0	1/5	0	0	0	0	0
sporty	0	0	0	0	1/5	0	0	0	0	0
gaperlu	0	0	0	0	0	1/6	0	0	0	0
pajak	0	0	0	0	0	1/6	0	0	0	0
produsen	0	0	0	0	0	0	1/7	0	0	0

perang	0	0	0	0	0	0	1/7	0	0	0
karna	0	0	0	0	0	0	1/7	0	0	0
saing	0	0	0	0	0	0	1/7	0	0	0
ketat	0	0	0	0	0	0	1/7	0	0	0
untung	0	0	0	0	0	0	0	1/5	0	0
rakyat	0	0	0	0	0	0	0	1/5	0	0
hidup	0	0	0	0	0	0	0	1/5	0	0
turun	0	0	0	0	0	0	0	1/5	0	0
subsidi	0	0	0	0	0	0	0	0	1/6	0
mending	0	0	0	0	0	0	0	0	1/6	0
gratis	0	0	0	0	0	0	0	0	1/6	0
obat	0	0	0	0	0	0	0	0	1/6	0
manfaat	0	0	0	0	0	0	0	0	1/6	0
negeri	0	0	0	0	0	0	0	0	0	1/4

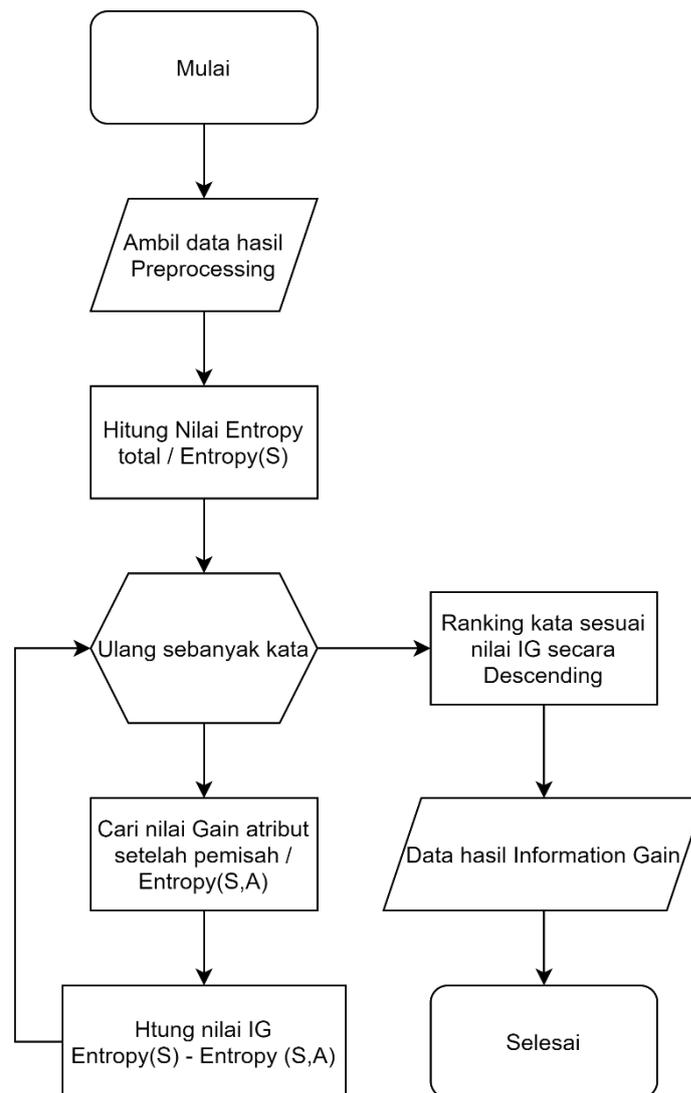
Tabel III- 9 Tabel DF

DF	IDF
2	2.321928095
3	1.736965594
7	0.514573173
1	3.321928095
1	3.321928095
3	1.736965594
1	3.321928095
1	3.321928095
1	3.321928095
3	1.736965594
1	3.321928095
1	3.321928095
1	3.321928095
1	3.321928095
2	2.321928095
1	3.321928095
1	3.321928095
3	1.736965594
1	3.321928095
1	3.321928095
2	2.321928095
2	2.321928095

murah	0	0	2/9	2/7	1/3	0	0	0	0	0
nunggu	0	0	0	5/9	0	0	0	0	0	0
baterai	0	0	0	5/9	0	0	0	0	0	0
pakai	0	0	0	2/5	0	2/5	0	0	0	0
kendara	0	0	0	2/5	0	2/5	0	0	0	0
butuh	0	0	0	0	2/3	0	0	0	0	0
cari	0	0	0	0	2/3	0	0	0	0	0
sporty	0	0	0	0	2/3	0	0	0	0	0
gaperlu	0	0	0	0	0	5/9	0	0	0	0
pajak	0	0	0	0	0	5/9	0	0	0	0
produsen	0	0	0	0	0	0	1/2	0	0	0
perang	0	0	0	0	0	0	1/2	0	0	0
karna	0	0	0	0	0	0	1/2	0	0	0
saing	0	0	0	0	0	0	1/2	0	0	0
ketat	0	0	0	0	0	0	1/2	0	0	0
untung	0	0	0	0	0	0	0	2/3	0	0
rakyat	0	0	0	0	0	0	0	2/3	0	0
hidup	0	0	0	0	0	0	0	2/3	0	0
turun	0	0	0	0	0	0	0	2/3	0	0
subsidi	0	0	0	0	0	0	0	0	5/9	0
mending	0	0	0	0	0	0	0	0	5/9	0
gratis	0	0	0	0	0	0	0	0	5/9	0
obat	0	0	0	0	0	0	0	0	5/9	0
manfaat	0	0	0	0	0	0	0	0	5/9	0
negeri	0	0	0	0	0	0	0	0	0	5/6

3.7 *Information gain*

Proses ini merupakan proses untuk menghitung skor tiap fitur dengan menggunakan *information gain*. Berikut alur dari proses *Information gain* pada gambar III-1



Gambar III- 2 *Information gain*

Pembahasan berikut merupakan proses penilaian manual dari proses *Information gain*. Dengan menghitung sesuai rumus pada persamaan II-1.

Tabel III- 11 *Data Training*

<i>Tweet</i>	Isi	Klasifikasi
1	produk mobil listrik usaha jenis biaya keluar	Netral
2	mantap bayar mahal token Listrik bunyi	Negatif
3	moga mobil motor listrik harga jangkau alias murah	Positif
4	nunggu baterai murah pakai kendara listrik	Negatif
5	butuh cari murah sporty listrik	Netral
6	pakai kendara listrik gaperlu pajak mahal	Positif
7	produsen perang harga mahal karna saing ketat	Netral
8	untung rakyat biaya hidup turun	Positif
9	subsidi mending gratisin biaya obat manfaat	Negatif
10	mobil listrik produk negeri	Netral

Tabel III- 12 Perhitungan *Entropy* data

Data	Jumlah	Probabilitas
Positif	3	0.3
Negatif	3	0.3
Netral	4	0.4
Total	10	
Entropy(S)		1.57095059

Fitur yang terdapat didalam tabel III-8 adalah kata yang akan dihitung bobotnya dengan menggunakan rumus persamaan II-2.

Tabel III- 13 Perhitungan *Entropy* Kata

Kata	Positif	Negatif	Netral	Jumlah	Entropy(kata)
Produk	2	0	0	2	0
Mobil	2	0	1	3	0.918295834
Listrik	2	2	2	6	1.584962501
Usaha	1	0	0	1	0
Jenis	1	0	0	1	0
Biaya	1	1	1	3	1.584962501
Keluar	1	0	0	1	0
Mantap	0	1	0	1	0
Bayar	0	1	0	1	0
Mahal	1	1	1	3	1.584962501
Token	0	1	0	1	0

Bunyi	0	1	0	1	0
Moga	0	0	1	1	0
Motor	0	0	1	1	0
Harga	1	0	1	2	1
Jangkau	0	0	1	1	0
Alias	0	0	1	1	0
Murah	1	1	1	3	1.584962501
Nunggu	0	1	0	1	0
Baterai	0	1	0	1	0
Pakai	0	1	1	2	1
Kendara	0	1	1	2	1
Butuh	1	0	0	1	0
Cari	1	0	0	1	0
Sporty	1	0	0	1	0
Gaperlu	0	0	1	1	0
Pajak	0	0	1	1	0
Produsen	1	0	0	1	0
Perang	1	0	0	1	0
Karna	1	0	0	1	0
Saing	1	0	0	1	0
Ketat	1	0	0	1	0
Untung	0	0	1	1	0
Rakyat	0	0	1	1	0
Hidup	0	0	1	1	0
Turun	0	0	1	1	0
Subsidi	0	1	0	1	0
Mending	0	1	0	1	0
Gratis	0	1	0	1	0
Obat	0	1	0	1	0
Manfaat	0	1	0	1	0
Negeri	1	0	0	1	0

Tabel III- 14 Perhitungan *Entropy* negasi Kata

\neg Kata	Positif	Negatif	Netral	Jumlah \neg	Entropy(\neg kata)
Produk	1	3	4	8	1.405639062
Mobil	1	3	3	7	1.448815636
Listrik	1	1	2	4	1.5
Usaha	2	3	4	9	1.530493057
Jenis	2	3	4	9	1.530493057

Biaya	2	2	3	7	1.556656707
Keluar	2	3	4	9	1.530493057
Mantap	3	2	4	9	1.530493057
Bayar	3	2	4	9	1.530493057
Mahal	2	2	3	7	1.556656707
Token	3	2	4	9	1.530493057
Bunyi	3	2	4	9	1.530493057
Mpga	3	3	3	9	1.584962501
Motor	3	3	3	9	1.584962501
Harga	2	3	3	8	1.561278124
Jangkau	3	3	3	9	1.584962501
Alias	3	3	3	9	1.584962501
Murah	2	2	3	7	1.556656707
Nunggu	3	2	4	9	1.530493057
Baterai	3	2	4	9	1.530493057
Pakai	3	2	3	8	1.561278124
Kendara	3	2	3	8	1.561278124
Butuh	2	3	4	9	1.530493057
Cari	2	3	4	9	1.530493057
Sporty	2	3	4	9	1.530493057
Gaperlu	3	3	3	9	1.584962501
Pajak	3	3	3	9	1.584962501
Produsen	2	3	4	9	1.530493057
Perang	2	3	4	9	1.530493057
Karna	2	3	4	9	1.530493057
Saing	2	3	4	9	1.530493057
Ketat	2	3	4	9	1.530493057
Untung	3	3	3	9	1.584962501
Rakyat	3	3	3	9	1.584962501
Hidup	3	3	3	9	1.584962501
Turun	3	3	3	9	1.584962501
Subsidi	3	2	4	9	1.530493057
Mending	3	2	4	9	1.530493057
Gratis	3	2	4	9	1.530493057
Obat	3	2	4	9	1.530493057
Manfaat	3	2	4	9	1.530493057
Negeri	2	3	4	9	1.530493057

Terakhir, hitung nilai Gain dengan persamaan II-3.

Tabel III- 15 Nilai Gain

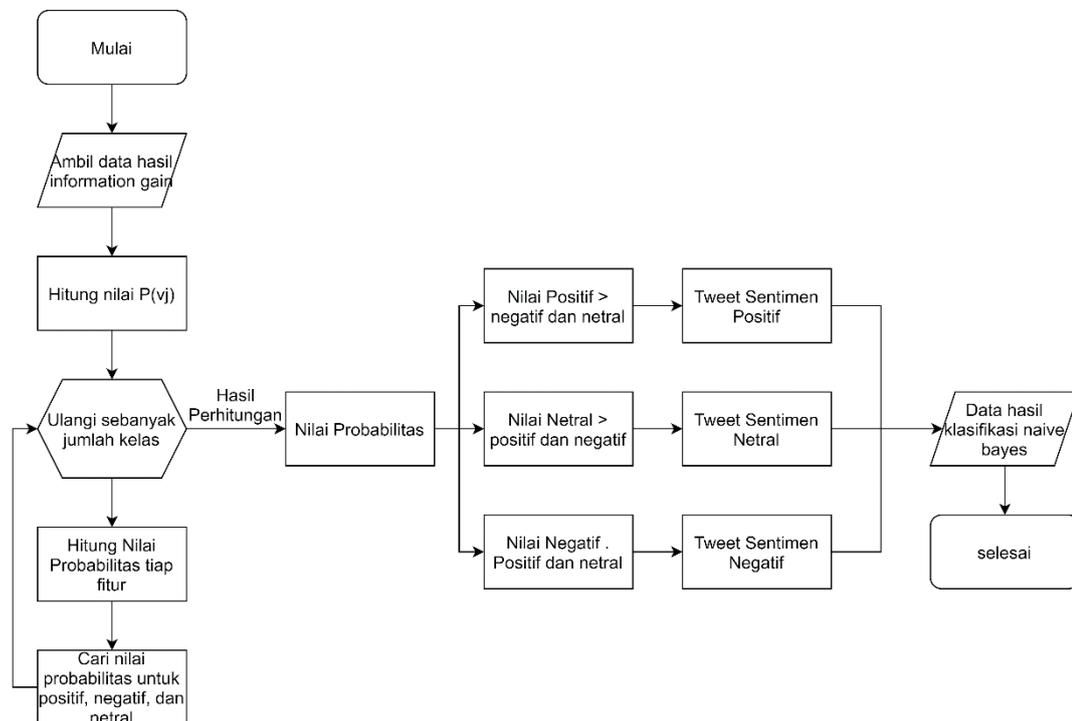
Gain
0.446439345
0.281290899
0.019973094
0.193506843
0.193506843
0.005802149
0.193506843
0.193506843
0.193506843
0.005802149
0.193506843
0.193506843
0.144484344
0.144484344
0.121928095
0.144484344
0.144484344
0.005802149
0.193506843
0.193506843
0.121928095
0.121928095
0.193506843
0.193506843
0.193506843
0.144484344
0.144484344
0.144484344
0.144484344
0.193506843
0.193506843
0.193506843
0.193506843
0.193506843
0.144484344
0.144484344
0.144484344
0.193506843
0.193506843

0.193506843
0.193506843
0.193506843
0.193506843

Dari bobot yang didapatkan, maka setiap fitur akan diranking dan akan didapatkan fitur yang memiliki bobot terbaik. Pada data *testing*, akan diambil fitur yang memiliki akurasi nilai tertinggi.

3.8 Naïve Bayes Classifier

Pada tahap ini, metode *Naïve Bayes Classifier* dibagi menjadi dua tahap, yaitu proses *training* dan proses *testing*.



Gambar III- 3 Naïve Bayes

Pada tahap ini, akan dibuat contoh perhitungan metode *Naïve Bayes* secara sederhana dengan menggunakan 10 *training* dan 2 *testing*.

3.6.1 Training

Dataset *training* yang sudah diklasifikasikan dan sudah dilakukan preprocessing seperti pada tabel III-8. Dari tabel diatas, dibuat model peluang dengan mengacu pada persamaan II-10 dan II-11. Dari hasil pembobotan *information gain* diambil $k = 20$. Kemudian hasil dari perhitungan disimpan kedalam tabel berikut

Tabel III- 16 Nilai n

	N	P(vj)
Positif	19	3/10
Negatif	18	3/10
Netral	23	4/10

Tabel III- 17 Nilai n1

N1	Positif	Negatif	Netral
Produk	0.912186037	0	0
Mobil	0.682379341	0	0.217120699
Usaha	0.474561156	0	0
Jenis	0.474561156	0	0
Keluar	0.474561156	0	0
Mantap	0	0.553654682	0
Bayar	0	0.553654682	0
Token	0	0.553654682	0
Bunyi	0	0.553654682	0
Nunggu	0	0.553654682	0
Baterai	0	0.553654682	0
Butuh	0.664385619	0	0
Cari	0.664385619	0	0
Sporty	0.664385619	0	0
Produsen	0.474561156	0	0
Perang	0.474561156	0	0
Karna	0.474561156	0	0
Saing	0.474561156	0	0
Ketat	0.474561156	0	0
Subsidi	0	0.553654682	0

Tabel III- 18 Nilai $P(a_i|v_j)$

$P(a_i v_j)$	Positif	Negatif	Netral
Produk	39/877	1/38	1/39
Mobil	34/869	1/38	23/737
Usaha	31/904	1/38	1/39
Jenis	31/904	1/38	1/39
Keluar	31/904	1/38	1/39
Mantap	1/43	24/587	1/39
Bayar	1/43	24/587	1/39
Token	1/43	24/587	1/39
Bunyi	1/43	24/587	1/39
Nunggu	1/43	24/587	1/39
Baterai	1/43	24/587	1/39
Butuh	6/155	1/38	1/39
Cari	6/155	1/38	1/39
Sporty	6/155	1/38	1/39
Produsen	31/904	1/38	1/39
Perang	31/904	1/38	1/39
Karna	31/904	1/38	1/39
Saing	31/904	1/38	1/39
Ketat	31/904	1/38	1/39
Subsidi	1/43	24/587	1/39

Hasil dari perhitungan digunakan untuk model probabilistic yang akan dipakai untuk data acuan menentukan data *testing*.

3.6.2 *Testing*

Selanjutnya merupakan perhitungan untuk data *testing*.

Tabel III- 19 Data *Testing Naïve Bayes*

<i>Tweet</i>	Isi	Klasifikasi
1	perintah produksi mobil motor listrik infrastruktur siap bangun bangkit tenaga listrik indonesia gelap gulita mobil listrik pln bayar bayar madam gilir tambah beban daya mobil listrik mikir	?
2	atur genap ganjil subsidi beli mobil listrik adalagi bijak untung produsen kendara semangkin mantap usaha kendara listrik milik tumbuh jalan non bayar banding tumbuh kendara	?

Proses *testing* dilakukan dengan menghitung nilai kemungkinan tertinggi.

Tabel III- 20 Fitur setelah *Information gain*

<i>Tweet</i>	Fitur	Klasifikasi		
		Positif	Negatif	Netral
1	produk, mobil, bayar, token	?	?	?
2	subsidi, mobil, produsen, mantap, usaha, bayar	?	?	?

$$\begin{aligned}
 P(\textit{tweet1}|V_{\textit{positif}}) &= P(\textit{Positif}) \times P(a_{\textit{produk}}|V_{\textit{positif}}) \\
 &\times P(a_{\textit{mobil}}|V_{\textit{positif}}) \times P(a_{\textit{bayar}}|V_{\textit{positif}}) \\
 &\times P(a_{\textit{token}}|V_{\textit{positif}}) \\
 &= 1.2\text{e-}06
 \end{aligned}$$

$$\begin{aligned}
 P(\textit{tweet1}|V_{\textit{negatif}}) &= P(\textit{negatif}) \times P(a_{\textit{produk}}|V_{\textit{negatif}}) \\
 &\times P(a_{\textit{mobil}}|V_{\textit{negatif}}) \times P(a_{\textit{negatif}}|V_{\textit{negatif}}) \\
 &\times P(a_{\textit{token}}|V_{\textit{negatif}}) \\
 &= 5.8\text{e-}07
 \end{aligned}$$

$$\begin{aligned}
P(\text{tweet1}|V_{\text{netral}}) &= P(\text{netral}) \times P(a_{\text{produk}}|V_{\text{netral}}) \\
&\times P(a_{\text{mobil}}|V_{\text{netral}}) \times P(a_{\text{bayar}}|V_{\text{netral}}) \\
&\times P(a_{\text{token}}|V_{\text{netral}}) \\
&= 2.3\text{e-}07
\end{aligned}$$

$$\begin{aligned}
P(\text{tweet2}|V_{\text{positif}}) &= P(\text{Positif}) \times P(a_{\text{subsidi}}|V_{\text{positif}}) \\
&\times P(a_{\text{mobil}}|V_{\text{positif}}) \times P(a_{\text{produsen}}|V_{\text{positif}}) \\
&\times P(a_{\text{mantap}}|V_{\text{positif}}) \times P(a_{\text{usaha}}|V_{\text{positif}}) \\
&\times P(a_{\text{bayar}}|V_{\text{positif}}) \\
&= 1.0\text{e-}09
\end{aligned}$$

$$\begin{aligned}
P(\text{tweet2}|V_{\text{negatif}}) &= P(\text{negatif}) \times P(a_{\text{subsidi}}|V_{\text{negatif}}) \\
&\times P(a_{\text{mobil}}|V_{\text{negatif}}) \times P(a_{\text{produsen}}|V_{\text{negatif}}) \\
&\times P(a_{\text{mantap}}|V_{\text{negatif}}) \times P(a_{\text{usaha}}|V_{\text{negatif}}) \\
&\times P(a_{\text{bayar}}|V_{\text{negatif}}) \\
&= 8.0\text{e-}10
\end{aligned}$$

$$\begin{aligned}
P(\text{tweet2}|V_{\text{netral}}) &= P(\text{netral}) \times P(a_{\text{subsidi}}|V_{\text{netral}}) \\
&\times P(a_{\text{mobil}}|V_{\text{netral}}) \times P(a_{\text{produsen}}|V_{\text{netral}}) \\
&\times P(a_{\text{mantap}}|V_{\text{netral}}) \times P(a_{\text{usaha}}|V_{\text{netral}}) \\
&\times P(a_{\text{bayar}}|V_{\text{netral}}) \\
&= 1.3\text{e-}10
\end{aligned}$$

Tabel III- 21 Hasil Klasifikasi *Naïve Bayes*

Tweet	Fitur	Klasifikasi		
		Positif	Negatif	Netral
1	Produk, mobil, bayar, token	1.2e-06	5.8e-07	2.3e-07
2	Subsidi, mobil, produsen, mantap, usaha, bayar	1.0e-09	8.0e-10	1.3e-10

Dari tabel III-13 dapat disimpulkan bahwa hasil dari data testing yaitu, *Tweet 1* mendapatkan nilai klasifikasi positif yang lebih besar dibandingkan negatif dan netral, ini menunjukkan bahwa *tweet 1* termasuk kedalam klasifikasi sentiment positif. Sedangkan *tweet 2* mendapatkan nilai klasifikasi negatif lebih besar dibandingkan dengan positif dan netral, ini menunjukkan bahwa 2 termasuk kedalam sentiment negatif.

3.9 Pengembangan Perangkat Lunak

Langkah-langkah dalam metode *waterfall* digambarkan sebagai berikut.

1. *Requirement*

Pada tahap ini melakukan observasi data, menetapkan fitur, dan tujuan dari sistem dibuat.

2. *System Design*

Pada tahap ini membuat sebuah desain sistem berdasarkan syarat sistem yang sudah ditetapkan.

3. *Coding & Testing*

Pada tahap ini perancangan yang sudah dibuat diterjemahkan dalam bahasa pemrograman menjadi serangkaian unit program.

4. *Implementation*

Tahap selanjutnya yaitu pengintegrasian setiap unit *system* menjadi sebuah *system* utuh.

5. *Maintenance*

Dalam tahap ini kesalahan yang belum ditemukan pada tahap – tahap sebelumnya maka akan dilakukan perbaikan secara berkala.

3.10 Menejemen Proyek Penelitian

Tabel III- 22 Jadwal Penelitian dalam Bentuk *Work Breakdown Structure* (WBS)

Nama Tugas	Durasi	Mulai	Akhir
Analisis Sentimen Pada Media Sosial Twitter Terhadap Penggunaan Kendaraan Listrik Dengan Menggunakan Metode <i>Naïve Bayes Classifier</i> Dan <i>Information gain</i>	156 hari	05/02/24	10/07/ 24
Menentukan Ruang Lingkup Penelitian	14 hari	05/02/24	19/02/24
Menentukan masalah penelitian	2 hari	05/02/24	06/02/24
Membuat latar belakang dan rumusan masalah	4 hari	07/02/24	11/02/24
Menentukan tujuan dan manfaat penelitian	2 hari	12/02/24	13/02/24
Menentukan batasan masalah	2 hari	14/02/24	15/02/24
Menentukan unit penelitian	2 hari	16/02/24	17/02/24
Tersedia dokumen bab pendahuluan	2 hari	18/02/24	19/02/24
Menentukan Landasan Teori Penelitian	14 hari	20/02/24	05/03/24
Mengumpulkan jurnal dan literatur ilmiah yang berkaitan dengan penelitian	5 hari	20/02/24	25/02/24
Mempelajari metode <i>naive bayes</i> dan <i>Information gain</i>	7 hari	26/02/24	02/03/24
Tersedia dokumen bab kajian literatur	2 hari	0403/24	05/03/24

Meyusun Tahapan Penelitian	22 hari	06/03/24	27/03/24
Menetapkan kerangka kerja	5 hari	06/03/24	10/03/24
Menetapkan kriteria dan format data pengujian	5 hari	11/03/24	15/03/24
Menentukan alat penelitian, pengujian penelitian, dan analisis hasil pengujian, serta metode pengembangan <i>software</i>	8 hari	16/03/24	24/03/24
Tersedia dokumen bab metodologi penelitian	4 hari	25/03/24	28/03/24
Rekayasa Perangkat Lunak dengan Waterfall	72 hari	28/03/04	08/06/24
Requirement	15 hari	28/03/24	12/04/24
Menganalisis Kebutuhan	2 hari	29/03/24	30/03/24
Spesifikasi Kebutuhan	2 hari	31/03/24	1/04/24
Mengumpulkan Kebutuhan	2 hari	2/04/24	3/04/24
Tinjauan Kebutuhan	2 hari	4/04/24	5/04/24
Membuat <i>Diagram use case</i> dan skenario <i>use case</i>	7 hari	6/04/24	12/04/24
System Design	16 hari	13/04/24	29/04/24
Mendesain arsitektur sistem	2 hari	13/04/24	14/04/24
Mendesain algoritma dan alur kerja sistem	2 hari	15/04/24	16/04/24

Membuat kelas analisis, <i>sequence Diagram</i> , dan <i>activity Diagram</i>	14 hari	17/04/24	29/04/24
Coding & Testing	28 hari	30/04/24	28/05/24
Membuat <i>Class Diagram</i>	5 hari	30/04/24	04/05/24
Membuat kode program	20 hari	06/05/24	25/05/24
Melakukan <i>testing</i>	3 hari	26/05/24	28/05/24
Implementasi	12 hari	29/05/24	10/06/24
Melakukan uji coba implementasi	7 hari	29/05/24	05/06/24
Deploy sistem pada server	2 hari	06/06/24	07/06/24
Pelatihan pengguna	1 hari	08/06/24	08/06/24
Tersedia dokumen bab pengembangan perangkat lunak	2 hari	09/06/24	10/06/24
Maintenance	-	11/06/24	-
Melaksanakan pengujian penelitian	14 hari	11/06/24	25/06/24
Membuat rancangan hasil pengujian	7 hari	11/06/24	17/06/24
Melakukan pengujian menggunakan perangkat lunak yang telah dibuat	4 hari	18/06/24	21/06/24

Tersedia dokumen bab hasil dan analisis penelitian	3 hari	22/06/24	25/06/24
Analisis hasil pengujian dan membuat kesimpulan	15 hari	26/06/24	10/07/24

3.11 Kesimpulan

Secara keseluruhan, bab ini memberikan gambaran tentang metodologi penelitian yang digunakan dalam penelitian. Rancangan penelitian, teknik pengumpulan data, dan metode analisis data telah dirancang dengan seksama untuk memastikan validitas dan reliabilitas hasil yang diperoleh. Penggunaan *Naïve Bayes Classifier* yang diperkuat dengan seleksi fitur menggunakan *Information gain* diharapkan mampu menghasilkan klasifikasi sentimen yang akurat terhadap kendaraan listrik.

Metodologi yang telah dijelaskan dalam bab ini memberikan fondasi yang untuk melanjutkan analisis dan pembahasan hasil penelitian. Dengan pendekatan yang sistematis, penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam memahami sentimen publik terhadap kendaraan listrik dan mendukung pengembangan strategi yang efektif untuk mendorong adopsi kendaraan listrik di Indonesia.

BAB IV

PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan

Pada bab ini membahas mengenai proses pengembangan perangkat lunak yang menjadi hal utama dalam penelitian ini. Proses pengembangan perangkat lunak dengan menggunakan metode *waterfall* terdapat lima fase yaitu *requirement, system design, coding and testing, implementasi, dan maintenance*.

4.2 Metode Pengembangan Perangkat Lunak

Pada penelitian ini, metode penelitian perangkat lunak yang digunakan adalah metode *waterfall* yang memiliki 5 tahapan, yaitu *requirement, system design, coding and testing, implementasi dan maintenance*.

4.2.1 Requirement

Pada tahap ini, hal yang perlu dilakukan adalah untuk menganalisis hal apa saja yang diperlukan untuk mengatasi masalah-masalah yang berkaitan dengan kebutuhan sistem. Tahap ini dilakukan dengan melakukan analisis terhadap perangkat keras, perangkat lunak, dan data.

4.2.1.1 Analisis Perangkat Keras

Dalam melakukan pengembangan perangkat lunak, diperlukan sebuah perangkat keras yang digunakan sebagai alat dalam membuat sistem. Perangkat keras yang digunakan dalam mengembangkan perangkat lunak adalah sebagai berikut:

Processor	: AMD Ryzen 5 4500U with Radeon Graphics
RAM	: 20 GB
SSD	: 512 GB

4.2.1.2 Analisis Perangkat Lunak

Dalam pengembangan perangkat lunak ini juga memerlukan perangkat lunak lain. Dalam penelitian ini perangkat lunak yang digunakan adalah sebagai berikut:

Sistem Operasi : Windows 11 64-bit

Bahasa Pemrograman : *Python*

Code Editor : *Visual Studio Code*

Library : *Pandas, sklearn, streamlit, re, nltk, sastrawi, string, os, pickle, io*

4.2.1.3 Analisis Data

Pada penelitian ini, data yang digunakan merupakan data sekunder. Data yang digunakan dalam ini diambil dari <https://www.kaggle.com/datasets/jovangt13/kendaraan-listrik>. Untuk rincian data, dapat dilihat dalam tabel V-1 berikut.

Tabel IV- 1 Rincian Dataset

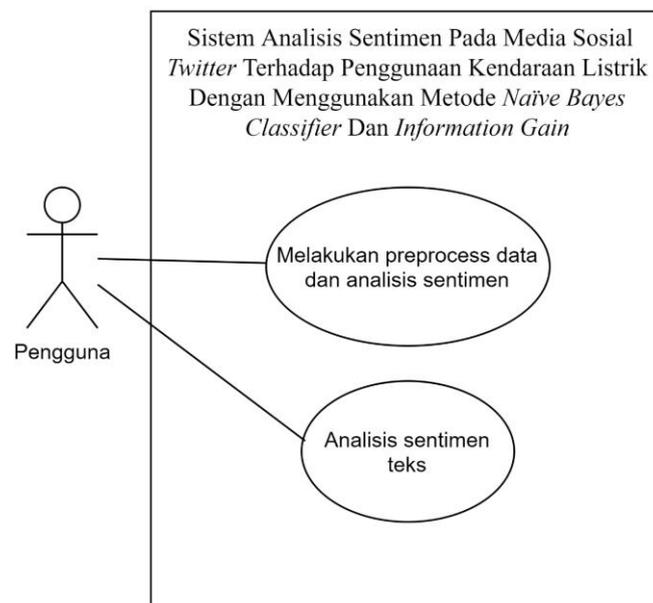
Klasifikasi	Jumlah
Positif	1903
Negatif	1603
Netral	668
Total	4174

4.2.2 System Design

Pada tahap ini hal yang dilakukan yaitu membuat *usecase*, *scenario usecase*, *Activity Diagram*, *sequence Diagram*, dan *Class Diagram*.

4.2.2.1 Use case Diagram

Use case Diagram menggambarkan *user* sebagai aktor terhadap sistem sehingga tercipta sebuah interaksi. *Diagram* usecase untuk penelitian ini digambarkan seperti dibawah ini.



Gambar IV- 1 *Use case Diagram*

4.2.2.2 Use case Skenario

Use case skenario merupakan penggambaran tindakan spesifik terhadap actor dan sistem yang dibuat pada Gambar IV-1 Berikut ini skenario penerapan dalam format tabel.

Tabel IV- 2 Skenario *preprocess data*

Identifikasi	
Nomor	1
Nama	Melakukan <i>Preprocess</i> Data dan analisis sentimen
Actor	Pengguna
Tujuan	Pengguna memasukkan data yang akan dilakukan <i>preprocessing</i> . Setelah itu, jika data sudah selesai diproses akan muncul <i>radio button</i> untuk memilih analisis dengan <i>Naive bayes</i> atau <i>Information gain dan Naive bayes</i>
Deskripsi	pada <i>use case</i> ini pengguna akan memasukkan data yang akan digunakan dalam preprocess
Kondisi awal	Belum ada data yang diinput
Skenario utama	
Pengguna	Sistem
	1. Menampilkan <i>interface</i> awal untuk menginput file csv dan tombol ' <i>browse file</i> '
2. Menekan tombol ' <i>browse files</i> '	
	3. Menampilkan <i>dialog box</i> untuk mencari <i>file</i>
4. Memilih <i>file</i> berekstensi CSV	

5. Menekan tombol ' <i>open</i> ' pada <i>dialog box</i>	
	6. Membaca <i>file</i>
	7. Menampilkan isi <i>file</i> CSV
8. Menekan tombol ' <i>Preprocess Data</i> '	
	9. Menampilkan data hasil <i>preprocessing text</i>
	10. Menampilkan <i>radio button</i> untuk memilih metode analisis sentimen.
11. Memilih metode analisis	
12. Menekan tombol ' <i>Analisis</i> '	
	13. Menampilkan hasil evaluasi berdasarkan metode yang dipilih
Kondisi Akhir	Menampilkan hasil evaluasi berdasarkan metode yang dipilih
Skenario alternatif	
Pengguna	Sistem
	1. Menampilkan interface awal dan dan button ' <i>browse file</i> '

2. Menekan tombol ' <i>browse files</i> '	
	3. Menampilkan <i>dialog box</i> untuk mencari <i>file</i>
4. Memilih <i>file</i> yang tidak berekstensi CSV atau file CSV yang dimasukkan tidak memiliki kolom yang dibutuhkan	
	5. Menampilkan pesan error
Kondisi Akhir	Sistem menampilkan pesan kesalahan input

Tabel IV- 3 Skenario Analisis sentiment teks

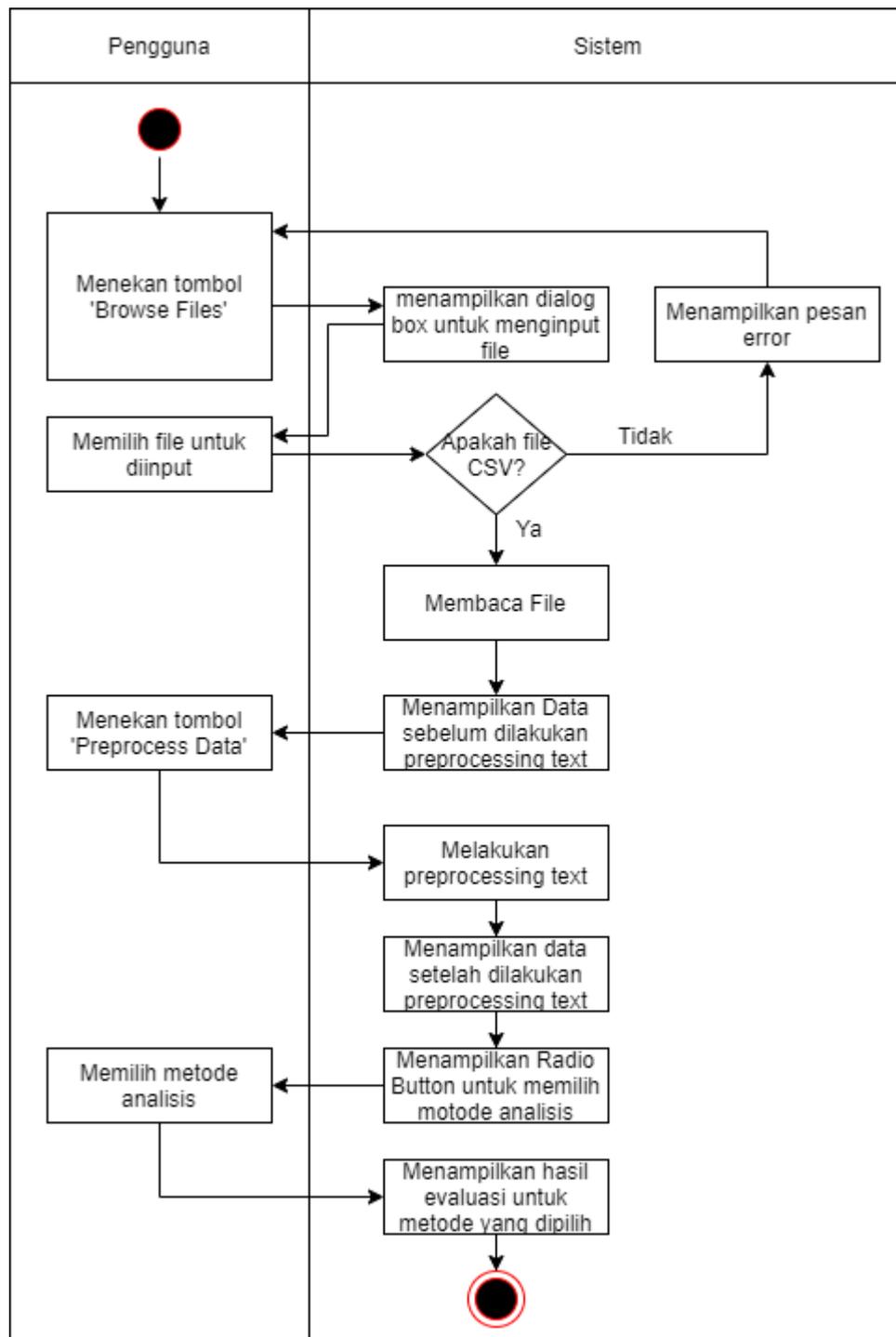
Identifikasi	
Nomor	2
Nama	Analisis sentiment teks
Actor	Pengguna
Tujuan	Pengguna memasukkan teks yang akan diklasifikasi
Deskripsi	pada use case ini pengguna akan memasukkan teks yang akan digunakan dalam proses perhitungan
Kondisi awal	Belum ada text yang diinput

Skenario utama	
Pengguna	Sistem
	1. Menampilkan interface awal dan kolom untuk input text
2. Memasukkan text	
3. Menekan tombol 'Analisis'	
	4. Melakukan preprocessing text
	5. Melakukan klasifikasi terhadap text berdasarkan model yang telah dilatih
	6. Menampilkan hasil preprocessing text dan klasifikasi text
Kondisi Akhir	interface menampilkan hasil preprocessing text dan klasifikasi text
Skenario alternatif	
Pengguna	Sistem
	1. Menampilkan interface awal dan kolom untuk input text

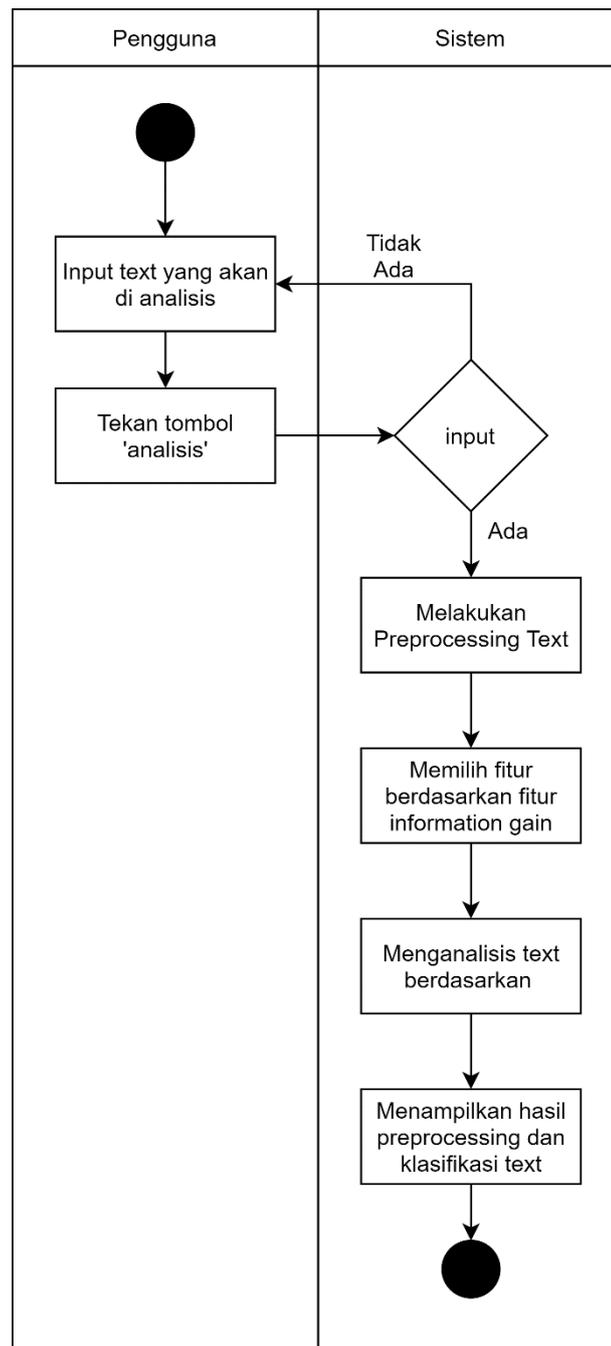
2. Menekan tombol 'analisis' tanpa memasukkan teks	
	3. Menampilkan pesan error 'silahkan masukkan text'
Kondisi akhir	Interface menampilkan pesan error 'silahkan masukkan text'

4.2.2.3 Activity Diagram

Activity *Diagram* adalah *Diagram* yang menggambarkan alur program yang perankan oleh actor dan dengan sistem. Alur program digambarkan dengan *Diagram* dibawah.



Gambar IV- 2 Activity diagram Preprocess Data dan Sentiment Analisis

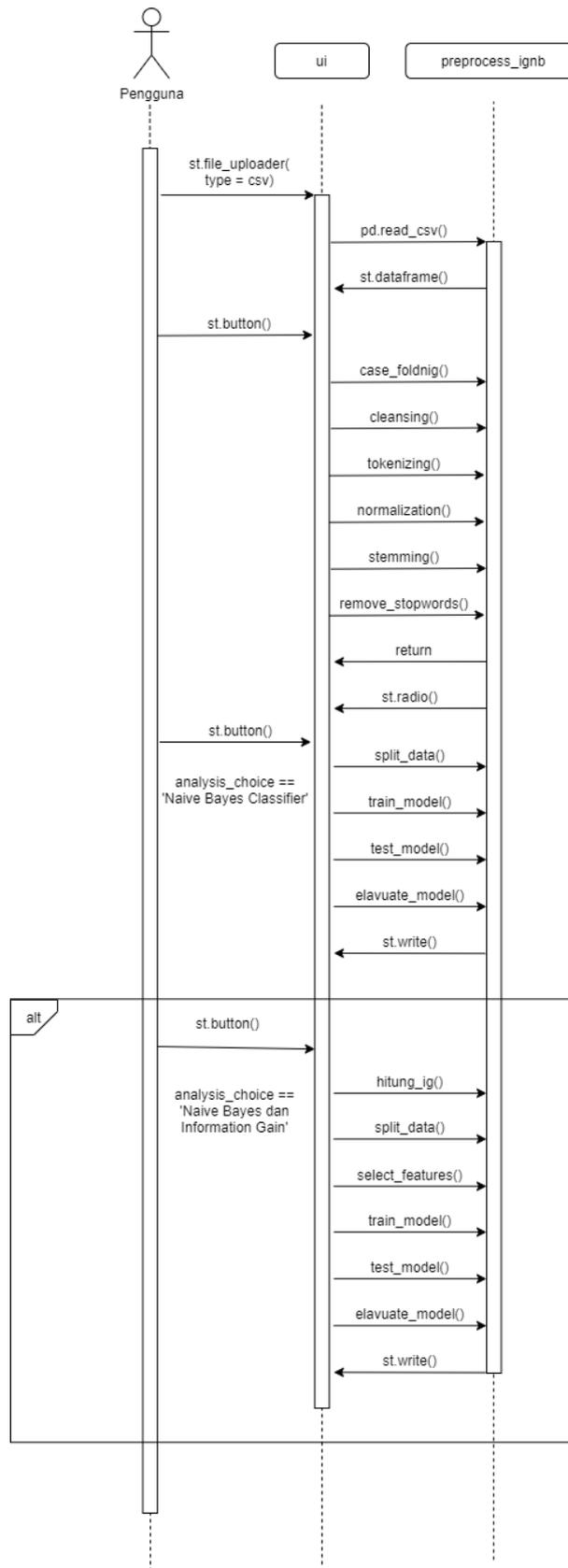


Gambar IV- 3 Activity Diagram Analisis sentimen teks

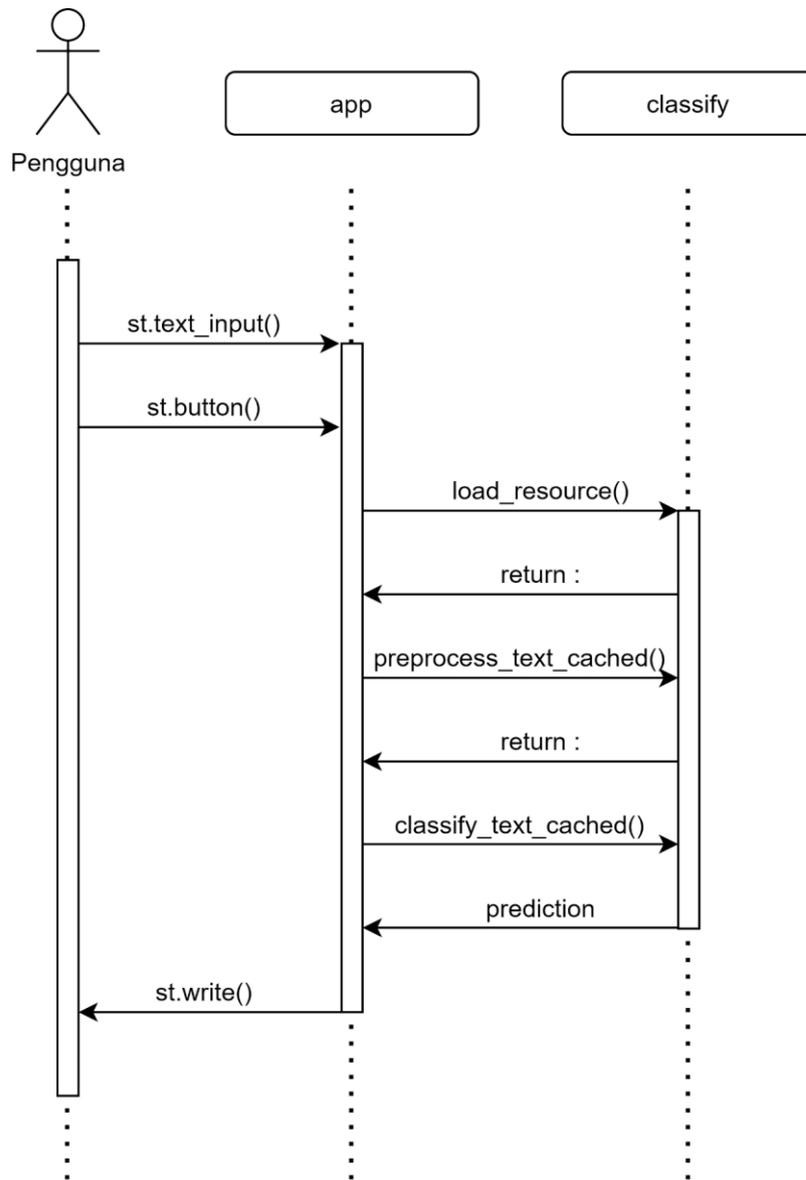
4.2.2.4 Sequence Diagram

Sequence Diagram bertujuan untuk mengetahui bagaimana suatu sistem berinteraksi satu dengan yang lain dengan urutan tertentu. Berikut adalah gambaran

dari *sequence Diagram* yang sedang dikembangkan peneliti.



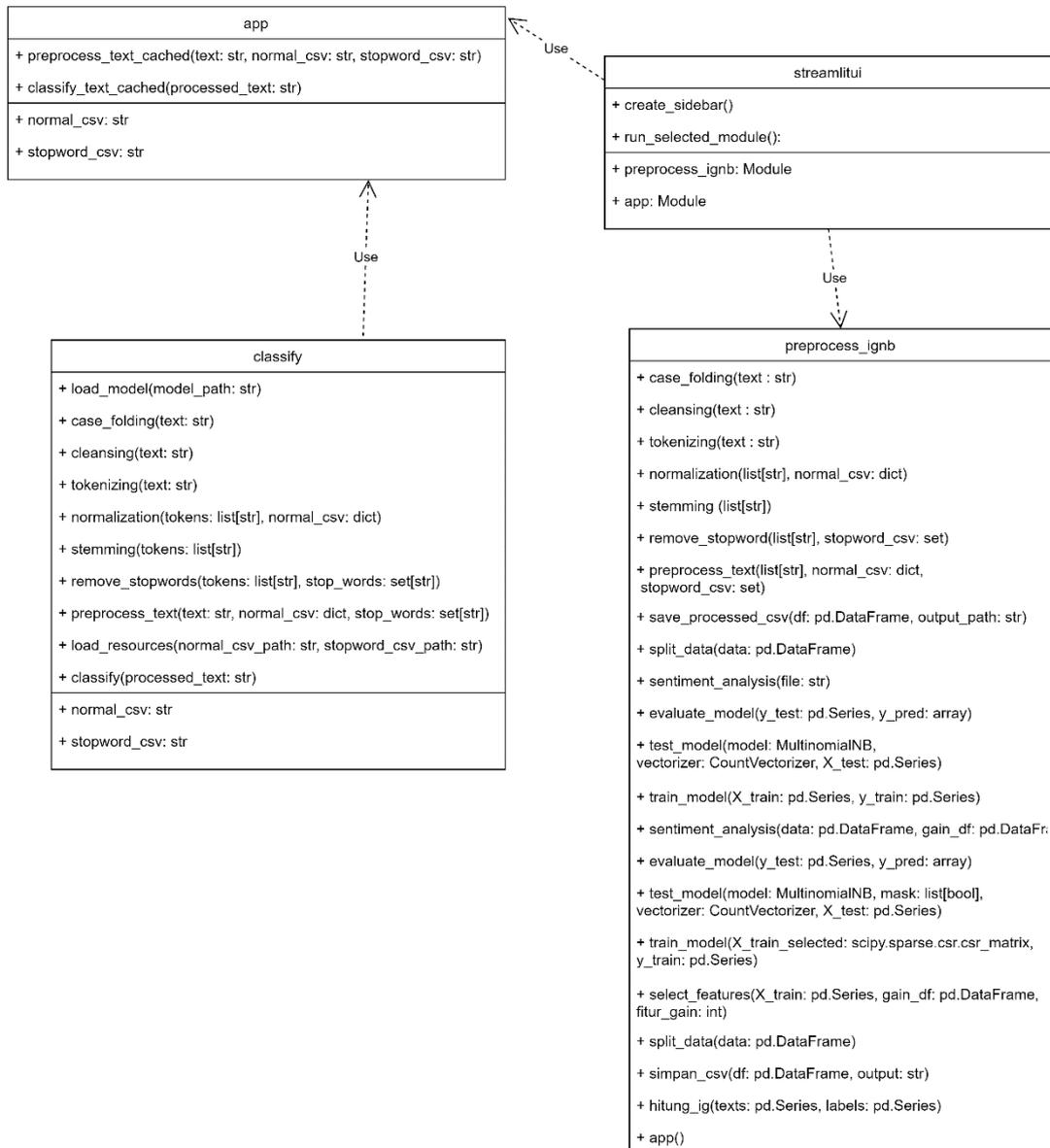
Gambar IV- 4 Analisis Sentimen Naïve Bayes



Gambar IV- 5 Sequence Diagram sentimen analisis teks

4.2.2.5 Class Diagram

Class diagram merupakan suatu *diagram* dalam UML yang diperlukan untuk menggambarkan struktur dari suatu sistem. *Diagram* ini menggambarkan kelas yang dimiliki sistem serta hubungan di antaranya. Berikut ini adalah *diagram* kelas untuk perangkat lunak yang sedang dikembangkan.



Gambar IV- 6 Class Diagram

4.2.3 Coding and Testing

Pada tahap ini dibagi menjadi dua tahap yaitu tahap *coding* dan tahap *testing*..

4.2.3.1 Coding

Pada tahap *coding*, hasil analisis dan desain sistem yang sudah dibuat

dalam tahap sebelumnya diterjemahkan kedalam bahasa pemrograman.

4.2.3.1.1 Preprcess_ignb.py

Tabel IV- 4 *function* dari *preprocess_ignb.py*

No.	Function	Keterangan
1	case_folding(text)	Mengubah semua karakter dalam teks menjadi huruf kecil.
2	cleaning(text)	Menghapus URL, symbol, tanda baca, spasi berlebih, mention, dan angka.
3	tokenizing(text)	Memecahkan teks menjadi kata Tunggal
4	normalisasi(tokens, normal_csv)	Mengubah kata slang atau informal dalam token, dengan padanan formal menggunakan file CSV yang dimasukkan
5	stopword_remove(tokens, stopword_csv)	Menghapus kata yang tidak berhubungan dengan topik dari daftar token berdasarkan CSV ynag dimasukkan
6	stemming(tokens)	Mengubah kata-kata menjadi kata dasarnya dengan stemmer Sastrawi
7	preprocess_text(text)	Menjalankan semua function prapemrosesan pada teks
8	split_data(data)	Membagi data menjadi data latih dan data uji

9	<code>evaluate_model(y_test, y_pred)</code>	Menghitung dan mengembalikan metrik evaluasi seperti akurasi, <i>confusion matrix</i> , presisi, recall, dan skor F1 untuk prediksi model.
10	<code>train_model_nb(x_train, y_train)</code>	Melatih classifier <i>Naive bayes</i> pada data pelatihan dan mengembalikan model yang dilatih serta vektorizer yang digunakan.
11	<code>test_model_nb(model, vectorizer, x_text)</code>	Menggunakan model <i>Naive bayes</i> yang dilatih dan vektorizer untuk memprediksi sentimen pada data pengujian.
12	<code>sentiment_analyze_nb(data)</code>	Melakukan analisis sentimen menggunakan classifier <i>Naive bayes</i> pada dataset, mengembalikan metrik evaluasi.
13	<code>hitung_ig(texts, labels)</code>	Menghitung <i>Information gain</i> untuk setiap fitur (kata) dalam teks dan mengembalikan DataFrame dengan nama fitur dan skor <i>Information gain</i> yang sesuai.
14	<code>select_features(x_train, gain_df,</code>	Memilih fitur dari data pelatihan berdasarkan nilai ambang untuk

	threshold)	<i>Information gain</i> , mengembalikan fitur yang dipilih, mask, vektorizer, dan jumlah fitur yang dipilih.
15	train_model_ignb(x_train_selected, y_train)	Melatih classifier <i>Naive bayes</i> pada fitur yang dipilih dari data pelatihan.
16	test_model_ignb(model, mask, vectorizer, x_text)	Menggunakan model <i>Naive bayes</i> berbasis <i>Information gain</i> yang dilatih dan vektorizer untuk memprediksi sentimen pada data pengujian.
17	sentimentt_analyze_ignb(data, gain_df, threshold)	Melakukan analisis sentimen menggunakan classifier <i>Naive bayes</i> berbasis <i>Information gain</i> pada dataset, menerapkan pemilihan fitur, dan mengembalikan metrik evaluasi.
18	app()	Fungsi utama untuk aplikasi Streamlit, menangani unggahan file, prapemrosesan, dan memanggil fungsi analisis sentimen berdasarkan pilihan pengguna.

4.2.3.1.2 Classify.py

Tabel IV- 5 *function* dari *classify.py*

No.	Function	Keterangan
1	case_folding(text)	Mengubah semua karakter dalam teks menjadi huruf kecil.
2	cleaning(text)	Menghapus URL, symbol, tanda baca, spasi berlebih, mention, dan angka.
3	tokenizing(text)	Memecahkan teks menjadi kata Tunggal
4	Normalisasi (tokens, normal_csv)	Mengubah kata slang atau informal dalam token, dengan padanan formal menggunakan file CSV yang dimasukkan
5	stopword_remove (tokens, stopwords_csv)	Menghapus kata yang tidak berhubungan dengan topik dari daftar token berdasarkan CSV yang dimasukkan
6	Stemming (tokens)	Mengubah kata-kata menjadi kata dasarnya dengan stemmer Sastrawi
7	preprocess_text(text)	Menjalankan semua function prapemrosesan pada teks
8	load_recource (normal_csv, stopwords_csv)	Memuat sumber daya yang diperlukan untuk normalisasi dan penghapusan stopwords dari file CSV

		yang diberikan. Mengembalikan kamus untuk normalisasi kata slang ke kata formal dan himpunan kata stopwords yang akan dihapus.
9	<code>classify(processed_text)</code>	Mengklasifikasikan teks yang telah diproses menggunakan model yang telah dilatih. Mengubah teks menjadi vektor fitur, memilih fitur yang relevan, dan mengembalikan hasil prediksi dari model.

4.2.3.1.3 App.py

Tabel IV- 6 *function* dari *app.py*

No.	Function	Keterangan
1	<code>preprocess_text_cached</code> (text, normal_csv, stopword_csv)	Memuat sumber daya untuk normalisasi dan penghapusan stopwords menggunakan fungsi <code>load_resources</code> dari modul <code>classify</code> , kemudian menjalankan prapemrosesan teks dengan fungsi <code>preprocess_text</code> .
2	<code>classify_text_cached</code> (processed_text)	mengklasifikasikan teks yang sudah diproses menggunakan fungsi <code>classify</code> dari modul <code>classify</code>

3	main()	Menampilkan antarmuka pengguna untuk analisis sentimen, menerima input teks dari pengguna
---	--------	---

4.2.3.2 *Testing*

Pada tahap *testing*, hasil dari tahap *coding* dilakukan *testing* dengan metode *blackbox*. *Testing* ini dilakukan untuk memastikan bahwa sistem berjalan dengan baik dan sesuai dengan sebagaimana seharusnya.

4.2.4.2.1. **Preprocess data dan Analisis sentimen**

Tabel IV- 7 Rencana Pengujian *Preprocess Data dan Analisis Sentimen*

Kode	Skenario Test
A1	Menginput <i>file</i> CSV dengan menekan tombol ' <i>browse file</i> ' untuk data dan menekan "Preprocess Data"
A2	Menginput <i>file</i> yang bukan CSV
A3	Memilih salah satu metode analisis dan menekan tombol "Analyze"

4.2.4.2.2. **Klasifikasi teks**

Tabel IV- 8 Rencana Pengujian Klasifikasi teks

Kode	Skenario Test
D1	Menginput teks yang akan dianalisis, dan menekan tombol 'analisis'

D2	Tidak ada teks yang diinput
----	-----------------------------

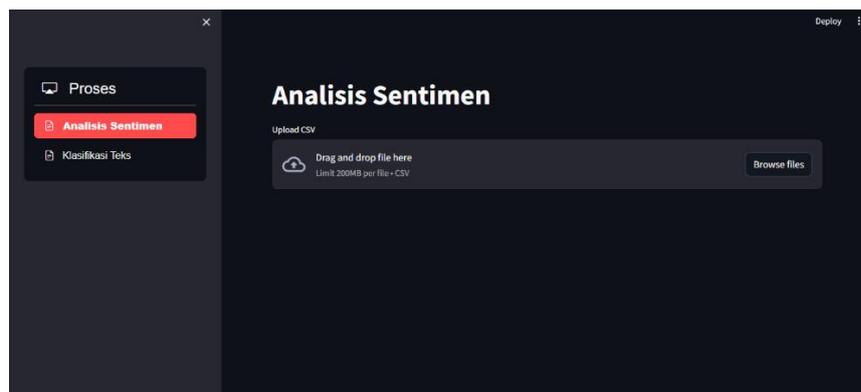
4.2.4 Implementasi

Tahap implementasi di sini mencakup serangkaian aktivitas yang dilakukan untuk menerapkan dan menguji sistem. Implementasi ini terdiri dari dua langkah, yaitu uji coba implementasi, dan pelatihan pengguna.

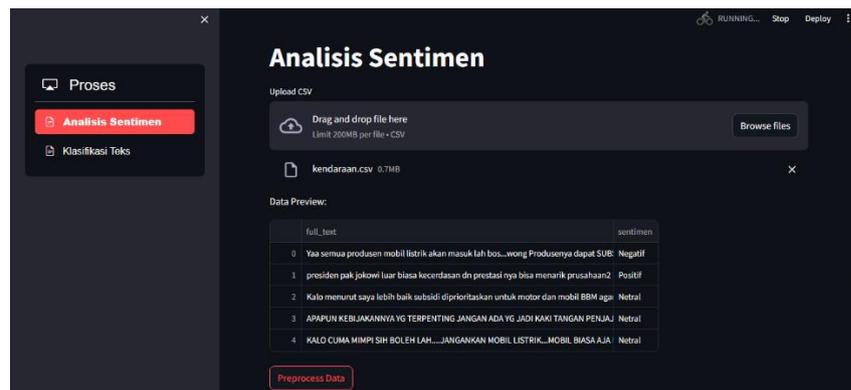
4.2.4.1. Implementasi Interface

Untuk interface sistem dapat dilihat pada gambar-gambar berikut.

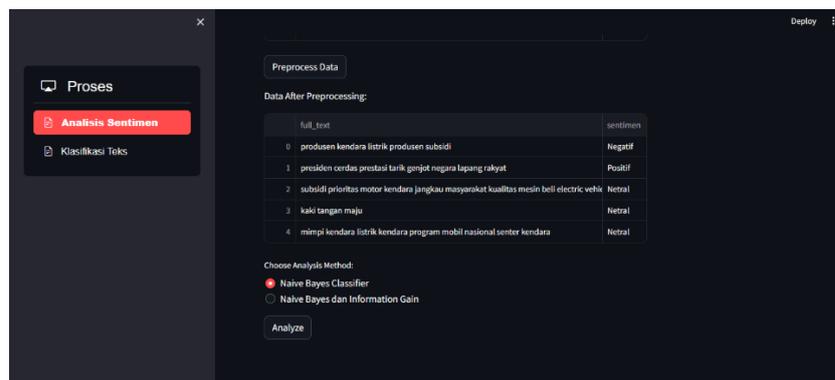
4.1.4.1.1. *Preprocess Data dan analisis sentimen*



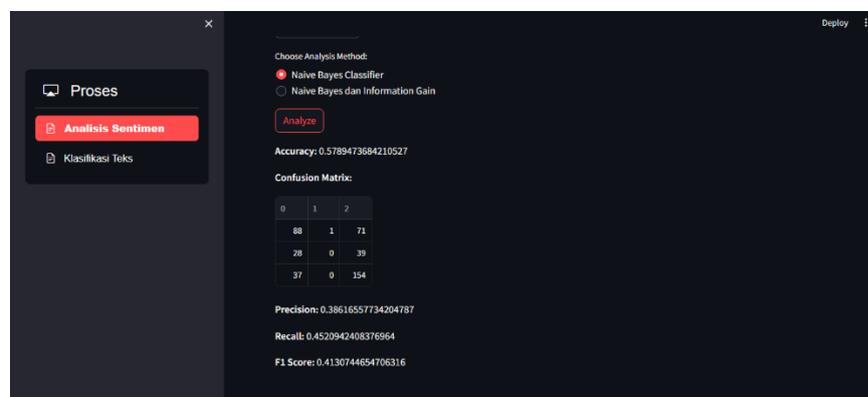
Gambar IV- 7 *Interface* halaman awal



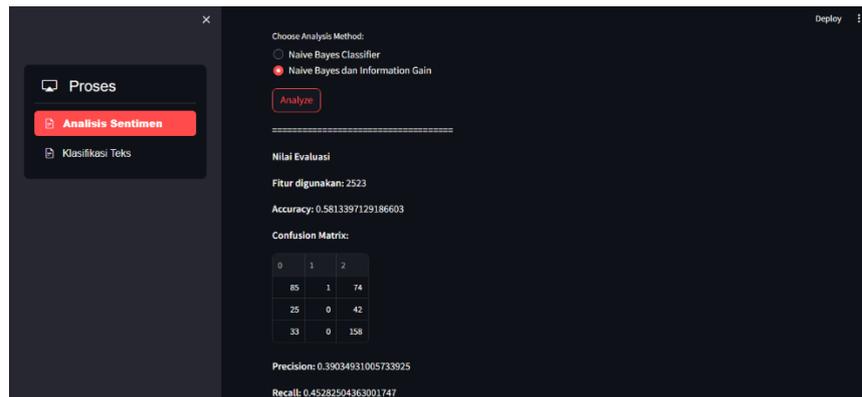
Gambar IV- 8 Interface setelah input file



Gambar IV- 9 Interface setelah tombol 'Preprocess Data' ditekan

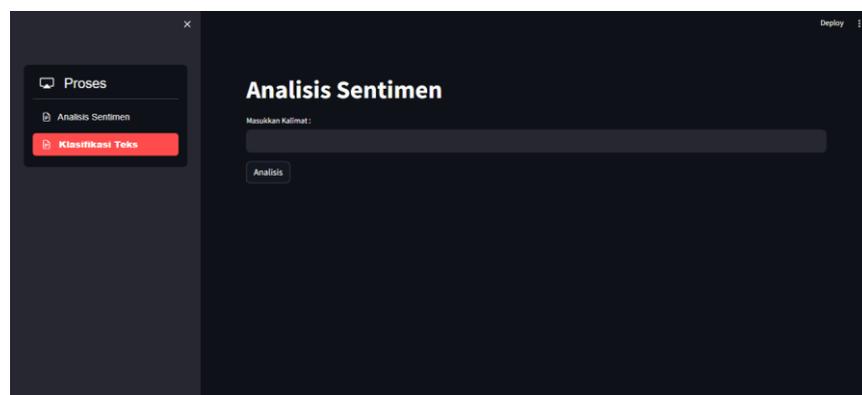


Gambar IV- 10 Hasil Sentimen Analisis Naive bayes

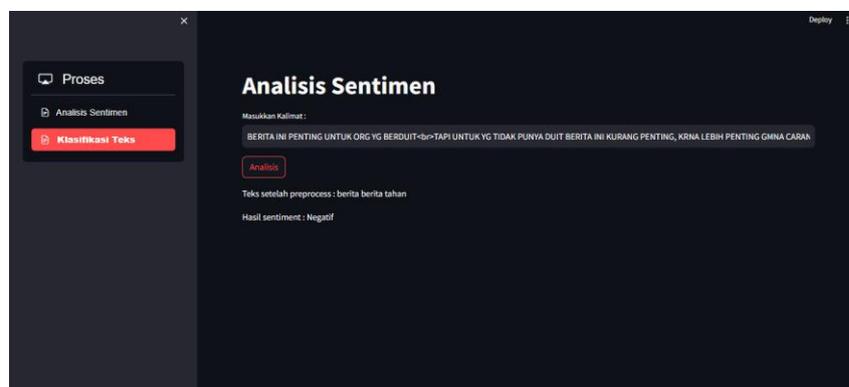


Gambar IV- 11 *Sentimen Analisis Information gain dan Naive bayes*

4.1.4.1.2. Klasifikasi Teks



Gambar IV- 12 *Interface awal klasifikasi teks*



Gambar IV- 13 *Interface setelah input teks dan menekan 'analisis'*

4.2.4.2. Implementasi Pengujian

Langkah pertama dalam tahap ini adalah melakukan uji coba terhadap sistem yang telah dikembangkan. Pengujian dilakukan dengan metode *blackbox* dengan teknik *Equivalence Partitioning*. Dalam pengujiannya menggunakan metode ini meliputi, yaitu skenario test, hasil yang diharapkan, hasil pengujian dan Kesimpulan. Uji coba ini bertujuan untuk memastikan bahwa sistem dapat berjalan sesuai dengan yang diharapkan.

4.2.4.2.1. Preprocess data dan analisis sentimen

Tabel IV- 9 Pengujian *Preprocess Data dan analisis sentimen*

Kode	Skenario Test	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
A1	Menginput <i>file</i> CSV dengan menekan tombol ' <i>browse file</i> ' dan menekan "Preprocess Data"	Menampilkan data awal yang belum dilakukan <i>preprocessing text</i> , lalu menekana tombol " <i>Preprocess Data</i> " untuk melakukan preproses teks, setelah selesai,	Sistem menampilkan data awal yang belum dilakukan <i>preprocessing text</i> , lalu menekana tombol " <i>Preprocess Data</i> " untuk melakukan preproses teks,	Berhasil

		muncul 2 radio <i>button</i> untuk memilih metode anlasis	setelah selesai, muncul 2 radio <i>button</i> untuk memilih metode anlasis	
A2	Menginput <i>file</i> yang bukan CSV	Menampilkan pesan error	Sistem menampilkan pesan error	Berhasil
A3	Memilih salah satu metode analisis dan menekan tombol “Analyze”	Menampilkan hasil evaluasi berupa <i>accuracy</i> , <i>onfusion</i> <i>matrix</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-</i> <i>score</i> untuk metode yang dipilih	Sistem menampilkan hasil evaluasi berupa <i>accuracy</i> , <i>onfusion</i> <i>matrix</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-</i> <i>score</i> untuk metode yang dipilih	Berhasil

4.2.4.2.2. Klasifikasi teks

Tabel IV- 10 Pengujian Klasifikasi teks

Kode	Skenario Test	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
D1	Menginput teks, lalu menekan tombol 'analisis'	Melakukan preprocess text dan menampilkan teks hasil preprocessing dan sentimen hasil analisis	Sistem melakukan preprocess text dan menampilkan teks hasil preprocessing dan sentimen hasil analisis	Berhasil
D2	Tidak ada teks yang diinput	Menampilkan pesan error	Sistem menampilkan pesan error	Berhasil

4.2.4.3. Pelatihan pengguna

Tahap ini adalah memberikan pelatihan kepada pengguna. Pelatihan ini bertujuan untuk memastikan bahwa pengguna dapat menggunakan sistem dengan efektif. pelatihan mencakup pengenalan antarmuka sistem, serta cara penggunaan fitur-fitur utama.

4.2.5 *Maintenance*

Tahap ini adalah langkah penting dalam memastikan keberlanjutan sistem analisis sentimen yang telah dikembangkan. Pada tahap ini, beberapa hal yang

dilakukan untuk menjaga agar sistem tetap berfungsi dengan baik terhadap perubahan atau masalah yang mungkin timbul selama penggunaan. Hal yang dapat dilakukan adalah sebagai berikut

4.2.5.1. *Monitoring*

Monitoring adalah langkah awal dalam proses maintenance, di mana sistem diawasi secara terus-menerus untuk mendeteksi setiap anomali atau masalah yang mungkin terjadi. Aktivitas monitoring meliputi pemantauan kinerja sistem dan pemantauan penggunaan sistem.

4.2.5.2. *Evaluasi sistem*

Evaluasi sistem dilakukan secara berkala untuk menilai apakah sistem masih memenuhi kebutuhan pengguna dan standar performa yang diharapkan. Evaluasi ini mencakup pengujian terhadap komponen perangkat keras dan perangkat lunak, serta analisis umpan balik dari pengguna.

4.2.5.3. *Pembaharuan sistem*

Seiring dengan perkembangan teknologi dan perubahan kebutuhan pengguna, sistem mungkin memerlukan pembaruan atau peningkatan. Proses pembaruan sistem meliputi pembaharuan perangkat lunak, dan pembaharuan data.

4.2.5.4. *Perbaikan dan pemulihan*

Jika terdeteksi masalah atau kerusakan pada sistem, langkah-langkah perbaikan dan pemulihan harus segera dilakukan untuk meminimalkan dampak negatif pada pengguna. Aktivitas perbaikan dan pemulihan meliputi identifikasi

masalah, dan *fixing bug*.

4.2.5.5. Dokumentasi

Semua aktivitas *maintenance* harus didokumentasikan dengan baik untuk tujuan audit dan referensi di masa yang akan datang. Dokumentasi ini mencakup laporan monitoring, catatan pembaruan sistem, dan log perbaikan.

4.3 Kesimpulan

Dalam penelitian ini, metode *waterfall* digunakan untuk merancang dan membangun perangkat lunak. Metode ini terdiri dari 5 tahap, yaitu *requirement*, *system design*, *coding and testing*, implementasi, dan *maintenance*. Penelitian ini menunjukkan metode *waterfall* dapat dipakai untuk mengembangkan perangkat lunak untuk melakukan analisis sentimen.

BAB V HASIL DAN PEMBAHASAN

5.1 Pendahuluan

Pada bab ini akan dibahas hasil dari penelitian mengenai analisis sentimen pada media sosial twitter terhadap penggunaan kendaraan listrik dengan menggunakan metode *naive bayes classifier* dan *information gain* dan mengukur kinerja dari hasil klasifikasi.

5.2 Data Hasil Pengujian

Disini akan dijelaskan mengenai data hasil pengujian sistem yang berfungsi untuk mengetahui kinerja program dalam proses klasifikasi. Pengujian dilakukan untuk menguji *accuracy*, *precision*, *recall*, dan *f1-score* dengan metode *naive bayes* dan *naive bayes* dengan *information gain*.

Dalam melakukan pengujian ini, akan dilakukan skenario dimana skenario ini akan dibagi berdasarkan data latih dan data uji. Dari total data sebanyak 4174 akan dibagi untuk beberapa skenario. Untuk rincian skenario dapat dilihat dalam tabel V-1

Tabel V- 1 Rincian data Pengujian

Pembagian	Data Latih	Data Uji
50 : 50	2087	2087
60 : 40	2504	1670
70 : 30	2922	1252
80 : 20	3339	835
90 : 10	3757	417

Pengujian dilakukan 2 kali, dimana pengujian pertama hanya menggunakan *naive bayes* dan pada penelitian yang kedua menggunakan seleksi fitur *information gain* dan *naive bayes*. Pengujian ini untuk melihat perbandingan

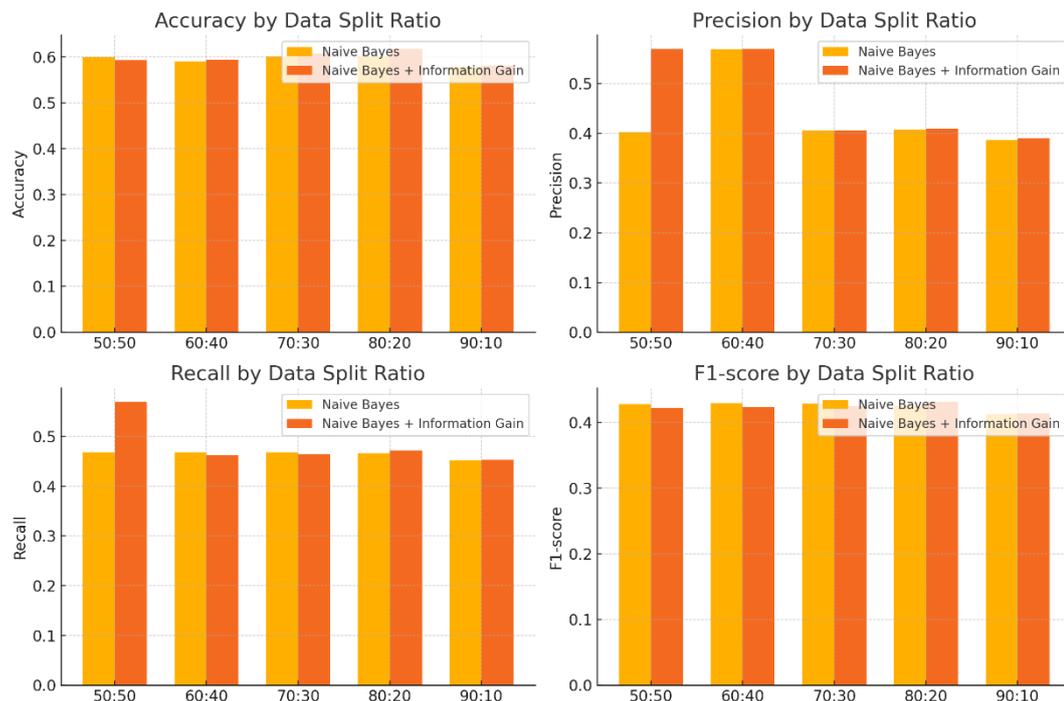
accuracy, precision, recall, dan f1-score.

Pengujian pertama, metode yang digunakan adalah *naive bayes*. Pengujian berikutnya, metode yang digunakan adalah *information gain* dan *naive bayes classifier*. Dalam pengujian ini untuk melakukan seleksi fitur, diberikan sebuah *threshold* sebesar 0.0002 untuk membatasi jumlah dari fitur berdasarkan skor dari *information gain*. Hasil dari penelitian ini akan menampilkan hasil dan menampilkan jumlah fitur yang digunakan jumlah fitur yang digunakan. Dari pengujian didapatkan data seperti berikut.

Tabel V- 2 Hasil pengujian

<i>Naive bayes</i>						
Rasio	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	Waktu (s)	Fitur
50 : 50	0.6003	0.4028	0.4686	0.4280	0.0042	2493
60 : 40	0.5904	0.5693	0.4685	0.4297	0.0054	2633
70 : 30	0.6009	0.4058	0.4689	0.4286	0.0059	2685
80 : 20	0.6019	0.4075	0.4668	0.4255	0.0061	2827
90 : 10	0.5789	0.3861	0.4520	0.4130	0.0072	3022
<i>Naive bayes dan Information gain</i>						
Rasio	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	Waktu (s)	fitur
50 : 50	0.5936	0.5697	0.5697	0.4220	0.0036	2262
60 : 40	0.5940	0.5697	0.4622	0.4231	0.0047	2379
70 : 30	0.6077	0.4065	0.4649	0.4243	0.005	2523
80 : 20	0.6177	0.4092	0.4718	0.4313	0.0054	2523
90 : 10	0.5813	0.3903	0.4528	0.4135	0.0066	2523

Perbandingan hasil elavusi digambarkan pada diagram batang berikut :



Gambar V- 1 Diagram Perbandingan Hasil Evaluasi

5.3 Analisis Data Hasil Pengujian

Berdasarkan dari data penelitian pada tabel V-2 dan gambar V-1. Pada model *Naive bayes*, dengan pembagian data 50:50, akurasi tercatat sebesar 0.6003, dengan presisi 0.4028, recall 0.4686, dan F1-score 0.4280. Meskipun akurasi cukup moderat, presisi dan F1-score relatif rendah, sementara recall lebih tinggi dari presisi, menunjukkan model lebih baik dalam mengidentifikasi klasifikasi positif. Pada pembagian 60:40, presisi meningkat signifikan menjadi 0.5693, namun akurasi dan F1-score tetap hampir tidak berubah, yang menunjukkan bahwa model menjadi lebih baik dalam hal presisi tetapi masih kesulitan untuk menyeimbangkan presisi dan recall. Pada pembagian 70:30, metrik serupa dengan pembagian 50:50,

menunjukkan kinerja yang konsisten. Pada pembagian 80:20, akurasi sedikit meningkat menjadi 0.6019, sementara presisi dan recall tetap serupa dengan pembagian sebelumnya. Namun, pada pembagian 90:10, kinerja sedikit menurun dengan akurasi 0.5789 dan presisi 0.3861.

Untuk model *Naive bayes* dengan *Information gain*, pada pembagian 50:50, akurasi tercatat sebesar 0,5936 dengan presisi dan recall yang sama yaitu 0,5697, namun F1-score lebih rendah yaitu 0,4220. Penambahan *Information gain* meningkatkan presisi dan recall secara signifikan. Pada pembagian 60:40, presisi tetap tinggi tetapi recall sedikit menurun menjadi 0,4622, dengan F1-score hampir sama. Pada pembagian 70:30, presisi menurun signifikan menjadi 0.4065, namun recall tetap serupa, dan F1-score sedikit meningkat. Pada pembagian 80:20, akurasi mencapai nilai tertinggi 0.6177, dengan presisi 0.4092 dan recall 0.4718, serta F1-score 0.4313. Pada pembagian 90:10, kinerja menurun dengan akurasi 0.5813, tetapi presisi, recall, dan F1-score lebih tinggi dibandingkan model *Naive bayes* pada pembagian yang sama.

Secara keseluruhan, *Naive bayes* ataupun *Naive bayes* dengan *Information gain* menunjukkan kinerja yang konsisten pada berbagai pembagian data latih-uji, dengan sedikit perubahan dalam akurasi, presisi, recall, dan F1-score. Penambahan *Information gain* umumnya meningkatkan presisi dan recall dibandingkan dengan *Naive bayes* saja, yang menunjukkan bahwa *Information gain* membantu dalam meningkatkan kemampuan model untuk mengidentifikasi klasifikasi positif. Pembagian data 80:20 memberikan kinerja terbaik untuk kedua model, dengan akurasi tertinggi dan keseimbangan yang relatif baik antara presisi, recall, dan F1-

score. Namun, seiring dengan penurunan ukuran data latih dari 90:10 ke 50:50, terdapat penurunan dalam kinerja untuk kedua model, yang menunjukkan bahwa kedua model mendapatkan manfaat dari lebih banyak data latih.

Penerapan *Information gain* sebagai metode seleksi fitur dalam model Naïve Bayes terbukti memberikan dampak positif terhadap kinerja model berdasarkan hasil evaluasi. Peningkatan akurasi terlihat jelas pada sebagian besar rasio data, menunjukkan bahwa *Information gain* membantu model lebih fokus pada fitur-fitur yang relevan, sehingga meningkatkan kualitas klasifikasi. Selain itu, *Information gain* secara signifikan meningkatkan precision, terutama pada rasio data 50:50 dan 60:40, yang berarti prediksi terhadap kelas positif menjadi lebih akurat dengan lebih sedikit kesalahan. Meskipun terdapat fluktuasi pada nilai recall, penurunan yang terjadi relatif kecil, sehingga *Information gain* tidak secara drastis mengurangi kemampuan model dalam mendeteksi kelas positif. Dari sisi F1-score, meskipun peningkatannya tidak selalu signifikan, *Information gain* tetap membantu menjaga keseimbangan antara precision dan recall. Secara keseluruhan, penggunaan *Information gain* memberikan peningkatan kinerja yang signifikan pada model *Naïve Bayes*, terutama dalam aspek akurasi dan precision, serta membuat model lebih efisien dengan mengurangi fitur yang tidak relevan.

Perubahan pada hasil evaluasi dari tiap percobaan disebabkan oleh beberapa faktor utama yang mempengaruhi kinerja model *Naive bayes* dan *Naive bayes* dengan *Information gain*. Pembagian data latih dan uji mempengaruhi hasil evaluasi. Jumlah data latih yang lebih besar biasanya memberikan model lebih banyak informasi untuk mempelajari pola dalam data, yang dapat meningkatkan

akurasi dan kinerja secara keseluruhan, sementara data uji yang lebih besar memberikan evaluasi yang lebih komprehensif. Variabilitas data seperti keberagaman data dan kehadiran outliers atau noise dalam data latih atau uji dapat mempengaruhi presisi, recall, dan F1-score model. Model *Naive bayes* dengan *Information gain* memilih fitur berdasarkan informasi yang paling relevan, yang dapat meningkatkan kemampuan model dalam mengidentifikasi pola penting dalam data dengan mengurangi kebisingan. Masalah *overfitting* dan *underfitting* juga dapat mempengaruhi hasil. Model yang terlalu fit pada data latih mungkin menunjukkan kinerja yang buruk pada data uji, sementara model yang tidak cukup fit pada data latih mungkin gagal mengenali pola yang relevan. Terakhir, sensitivitas model terhadap variasi dalam distribusi data dapat menyebabkan fluktuasi dalam metrik kinerja seperti presisi, recall, dan F1-score. Secara keseluruhan, model *Naive bayes* dengan *Information gain* cenderung menunjukkan kinerja yang lebih stabil dan konsisten dengan peningkatan presisi dan recall karena pemilihan fitur yang lebih relevan membantu dalam mengurangi kebisingan dan meningkatkan kemampuan model dalam mengidentifikasi pola penting dalam data.

Dari hasil pengujian menunjukkan bahwa jumlah fitur berpengaruh terhadap kinerja model *Naive bayes*, baik dengan maupun tanpa penerapan metode *Information gain*. Pada *Naive bayes* tanpa *Information gain*, jumlah fitur lebih tinggi, berkisar antara 2493 hingga 3022. Sebaliknya, ketika *Information gain* diterapkan, jumlah fitur yang dimiliki berkurang, menjadi 2262 hingga 2523. *Information gain* bekerja dengan menyaring fitur yang kurang relevan, sehingga selain mengurangi kompleksitas model, hal ini juga mempercepat waktu eksekusi,

di mana waktu pemrosesan tanpa *Information gain* lebih lama, yaitu 0.0072 detik, sedangkan dengan *Information gain* waktu dapat ditekan hingga 0.0066 detik.

Penerapan *Information gain* menghasilkan peningkatan yang lebih konsisten, terutama pada rasio data 80:20 di mana akurasi mencapai 0.6177 dengan fitur 2523, dibandingkan dengan akurasi 0.6019 tanpa *Information gain* meski menggunakan lebih banyak fitur 2827. Selain itu, precision, recall, dan F1-score juga cenderung lebih stabil saat *Information gain* diterapkan. Precision dan recall tanpa *Information gain* cenderung lebih fluktuatif, sementara dengan *Information gain*, keduanya menunjukkan peningkatan konsistensi pada setiap rasio data.

Berdasarkan waktu latih, penerapan *Information gain* mempercepat waktu komputasi dalam semua rasio data. Sebagai contoh, pada rasio 50:50, waktu eksekusi dengan *Information gain* adalah 0.0036 detik, sedangkan tanpa *Information gain* membutuhkan 0.0042 detik. Demikian pula untuk rasio lainnya, waktu dengan *Information gain* selalu lebih singkat dibandingkan tanpa *Information gain*.

5.4 Kesimpulan

Penelitian analisis sentimen pada media sosial *twitter* terhadap penggunaan kendaraan listrik dengan menggunakan metode *naïve bayes classifier* dan *information gain* dapat berhasil melakukan analisis sentimen. Sebanyak 4174 data *tweet* bahasa indonesia yang sudah diberikan label klasifikasi dengan 3 kelas, yaitu positif, negatif, dan netral. Hasil perhitungan evaluasi menunjukkan bahwa analisis sentimen dengan menggunakan seleksi fitur memberikan hasil kinerja yang lebih baik dibandingkan dengan metode *naïve bayes* tanpa menggunakan seleksi fitur.

BAB VI

KESIMPULAN DAN SARAN

6.1 Pendahuluan

.Bab ini memberikan kesimpulan dari penelitian yang telah dilakukan dan saran yang diharapkan dapat menjadi acuan bagi peneliti lain untuk melakukan penelitian di bidang yang berhubungan.

6.2 Kesimpulan

Penelitian ini membahas tentang penggunaan seleksi fitur *information gain* dan *naive bayes* dalam melakukan analisis sentimen. Dengan melakukan analisis berdasarkan hasil dari penelitian, peneliti dapat menyimpulkan hal sebagai berikut:

1. Sistem analisis sentimen berhasil dibangun dengan menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan sentimen dari *tweet* tentang kendaraan listrik menjadi positif, negatif, atau netral. Selain itu, *Information gain* digunakan sebagai teknik seleksi fitur untuk meningkatkan akurasi dan efisiensi klasifikasi. Hasil penelitian menunjukkan bahwa penerapan *Information gain* membantu mengurangi dimensi data dan meningkatkan kinerja model, seperti terlihat pada peningkatan akurasi dan waktu pelatihan yang lebih singkat.
2. Hasil penelitian menunjukkan bahwa penggunaan *Information gain* dalam seleksi fitur mampu meningkatkan akurasi model *Naïve Bayes* dalam menganalisis sentimen terhadap kendaraan listrik. Model *Naive bayes* dengan *Information gain* mendapatkan peningkatan akurasi, terutama pada rasio 80:20, yang merupakan nilai akurasi tertinggi sebesar 61.77%, lebih tinggi dibandingkan dengan model yang hanya menggunakan *Naïve Bayes* yang

mendapatkan nilai akurasi sebesar 60.2%. Dalam penelitian ini, akurasi model *Naïve Bayes* setelah diterapkan *Information gain* mengalami peningkatan, memperlihatkan peningkatan dalam kemampuan klasifikasi sentimen.

6.3 Saran

Berdasarkan penelitian yang sudah dilakukan, terdapat beberapa saran yang dapat diberikan untuk penelitian selanjutnya dan pengembangan yang lebih lanjut:

1. Penelitian ini dilakukan dengan menggunakan dataset dari satu sumber, yaitu Twitter. Disarankan agar penelitian selanjutnya mempertimbangkan penggunaan dataset yang lebih beragam dari berbagai platform media sosial seperti *Facebook*, *Youtube*, *Instagram*, atau forum-forum diskusi *online*. Hal ini dapat memberikan perspektif yang lebih luas dan mendalam mengenai sentimen masyarakat terhadap kendaraan listrik.
2. Meskipun metode *text preprocessing* yang digunakan dalam penelitian ini sudah cukup, namun untuk penelitian selanjutnya dapat mempertimbangkan penggunaan teknik pra-pemrosesan tambahan seperti *lemmatization*, *Named Entity Recognition* (NER), dan penggunaan model bahasa yang lebih canggih untuk meningkatkan kualitas data yang diolah.
3. Penelitian ini menggunakan metode *Naïve Bayes Classifier*. Disarankan agar penelitian selanjutnya mencoba menggunakan metode machine learning lain seperti *Support Vector Machine* (SVM), *Random Forest*, atau *deep learning* seperti *Long Short-Term Memory* (LSTM) dan *Transformer* untuk melihat

perbandingan performa dan akurasi dalam analisis sentimen.

4. Untuk mendapatkan hasil yang lebih relevan dan terkini, disarankan agar penelitian selanjutnya mencoba melakukan analisis sentimen dengan data real-time. Ini akan membantu dalam memahami tren sentimen yang sedang berlangsung di masyarakat.

Dengan mengikuti saran-saran tersebut, diharapkan penelitian di masa mendatang dapat memberikan kontribusi yang lebih besar dan bermanfaat dalam bidang analisis sentimen serta mendukung perkembangan industri kendaraan listrik di Indonesia.

DAFTAR PUSTAKA

- Alfarizi. S., & Fitriani. E. (2023). Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma *naive bayes* dengan Seleksi Fitur *Information gain* dan Particle Swarm Optimization. Indonesian Journal on Software Engineering (IJSE).
- Alin F. N., Totohendarto M. H., Muttaqin M. R. (2023). Analisis Sentimen Terhadap Kendaraan Listrik Pada Platform Twitter Menggunakan Metode *Naive bayes*. Informatics for Educators And Professionals : Journal of Informatics
- Anggita, S.D., & Abdulloh, F.F. (2023). Optimasi Algoritma Support Vector Machine Berbasis PSO Dan Seleksi Fitur *Information gain* Pada Analisis Sentimen. *Journal Of Applied Computer Science And Technology*.
- Darwis, D., Siskawati, N., & Abidin, Z. Penerapan Algoritma *naive bayes* untuk Analisis Sentimen Review Data Twitter BMKG Nasional. Jurnal TEKNO KOMPAK
- Ernawati. N. W., Kumara. I. N. S., & Setiawan. S. (2023). Perbandingan Metode Klasifikasi Support Vector Machine Dan Naïve Bayes Pada Analisis Sentimen Kendaraan Listrik. Jurnal SPEKTRUM.
- Firdaus, A., Istalama Firdaus, W. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). Jurnal JUPITER.

- Husada, H.C., & Paramita, A.S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). TEKNIKA
- Isnanda, A., Umaidah, Y., & Jaman, J.H. (2021). Implementasi Naïve Bayes Classifier Dan *Information gain* Pada Analisis Sentimen Penggunaan E-Wallet Saat Pandemi. Jurnal Teknologi Informatika dan Komputer MH. Thamrin
- Mahardika F., Mustofa K., Suseno A.T.(2023). Implementasi Metode Waterfall pada Sistem Informasi Penjualan Unit Motor Berbasis Web. JURNAL ILMU KOMPUTER
- Nofiyani, & Wulandari. (2021). Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining. JURNAL MEDIA INFORMATIKA BUDIDARMA
- Noviana R. & Rasal I. (2023). PENERAPAN ALGORITMA *NAIVE BAYES* DAN SVM UNTUK ANALISIS SENTIMEN BOYBAND BTS PADA MEDIA SOSIAL TWITTER. Jurnal Teknik dan science
- Nur. F., Ahsan. M., Harianto. W. (2022). Komparasi Tingkat Akurasi *Information gain* Dan Gain Ratio Pada Metode K-Nearest Neighbor. Jurnal Mahasiswa Teknik Informatika
- Rahman Isnain, A., Indra Skti, A., Satya Marga, A., & Alita. (2021). Sentimen Analisis Publik Terhadap kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM. JDMSI
- Ramadhani, B., & Suryono, R.R. (2024). Komparasi Algoritma Naïve Bayes dan

Logistic Regression Untuk Analisis Sentimen Metaverse. JURNAL
MEDIA INFORMATIKA BUDIDARMA

Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan *Naïve Bayes Classifier*. *Kajian Ilmiah Informatika dan Komputer*.

Tanggreani, A., I., & Sitokdana, M.N.N. (2022). Analisis Sentimen Aplikasi *E-Government* Pada *Google Play* Menggunakan Algoritma *Naïve Bayes*. *Jurnal Teknik Informatika dan Sistem Informasi*

Ulgasesa, R., Putra Negra, A.B., & Tursina. (2021). Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal. *Jurnal Sistem dan Teknologi Informasi*.

Wati, R., Ernawati, S., & Rachmi, H. (2023). Pembobotan TF-IDF Menggunakan *Naïve Bayes* pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH. *Jurnal Manajemen Informatika (JAMIKA)*