

**ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP MOBIL LISTRIK  
DI MEDIA SOSIAL TWITER MENGGUNAKAN METODE *NAÏVE*  
*BAYES CLASSIFIER* DAN *LATENT DIRICHLET ALLOCATION***

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

**Disusun Oleh :**

**PUTRA UMAMUL MUSTHOFA**

**NIM : H06219013**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA  
2023**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : PUTRA UMAMUL MUSTHOFA  
NIM : H06219013  
Program Studi : SISTEM INFORMASI  
Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP MOBIL LISTRIK DI MEDIA SOSIAL TWITTER MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER* DAN *LATENT DIRICHLET ALLOCATION*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 12 Juli 2023  
Yang menyatakan,

A handwritten signature in black ink is written over a rectangular stamp. The stamp features the Garuda Pancasila logo and the text 'METRA TELEPEK' and 'FDDC6AJX302787859'. The signature is written in a cursive style.

**Putra Umamul Musthofa**  
NIM. H06219013

## LEMBAR PERSETUJUAN PEMBIMBING

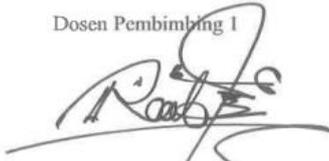
Skripsi oleh :

NAMA : PUTRA UMAMUL MUSTHOFA  
NIM : H06219013  
JUDUL : ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP  
MOBIL LISTRIK DI MEDIA SOSIAL TWITER  
MENGUNAKAN METODE *NAÏVE BAYES*  
*CLASSIFIER* DAN *LATENT DIRICHLET ALLOCATION*

Mahasiswa tersebut telah melakukan bimbingan dan dinyatakan layak untuk mengikuti Sidang Skripsi.

Surabaya, 12 Juli 2023

Dosen Pembimbing 1



Mujib Ridwan, S. Kom., M.T  
NIP.198604272014031004

Dosen Pembimbing 2



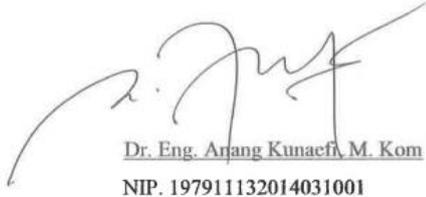
Khalid, M. Kom  
NIP.197906092014031002

## PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Putra Umamul Musthofa ini telah dipertahankan  
di depan tim penguji skripsi  
di Surabaya, 14 Juli 2023

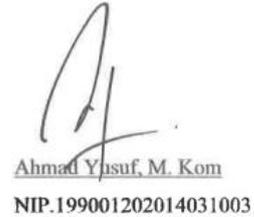
Mengesahkan,  
Dewan Penguji

Penguji I



Dr. Eng. Arjang Kunaefi, M. Kom  
NIP. 197911132014031001

Penguji II



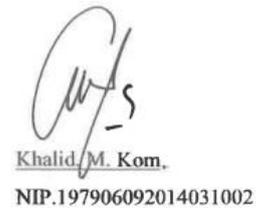
Ahmad Yusuf, M. Kom  
NIP.199001202014031003

Penguji III



Mujib Ridwan, S. Kom., M. T  
NIP.198604272014031004

Penguji IV



Khalid, M. Kom.  
NIP.197906092014031002

Mengetahui,

Dekan Fakultas Sains dan Teknologi  
Surabaya Ampel Surabaya



Agus Saepul Hamdani, M. Pd  
NIP 196507312000031002

# LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH



**KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA  
PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300  
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Putra Umamul Musthofa  
NIM : H06219013  
Fakultas/Jurusan : Sistem Informasi  
E-mail address : H06219013@student.uinsby.ac.id

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi  Tesis  Desertasi  Lain-lain (.....)

yang berjudul :

Analisis Sentimen Berbasis Aspek Terhadap Mobil Listrik Di Media Sosial

Twitter Menggunakan Metode Naïve Bayes Classifier Dan Latent Dirichlet Allocation

berserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 02 November 2023

Penulis

  
( Putra Umamul Musthofa )  
*nama terang dan tanda tangan*

## ABSTRAK

Perkembangan penjualan mobil listrik di Indonesia semakin tahun mengalami peningkatan yang signifikan. Akan tetapi kehadirannya di Indonesia menimbulkan pro dan kontra di kalangan masyarakat. Sehingga perlu diketahui pro dan kontra masyarakat terhadap mobil listrik dari aspek-aspek yang mempengaruhinya serta mengklasifikasikannya. Sebagai solusi untuk mencari pro dan kontra dari masyarakat terhadap suatu aspek yang dinilai kurang atau lebih dari mobil listrik. Metode yang digunakan untuk mencari aspek menggunakan pemodelan topik *latent Dirichlet allocation* (LDA) dan *naïve bayes classifier* untuk mengklasifikasi sentimen. Pemodelan topik LDA menghasilkan *coherence score* terbaik sebesar 0.453 dengan 6 topik. Topik yang dihasilkan kemudian diinterpretasikan menjadi nama aspek yaitu aspek pengisian baterai, produksi, perkembangan, subsidi, wilayah, dan merk. Setiap aspek kemudian diklasifikasi sentimen positif, negatif, dan netral serta menghasilkan performa dari klasifikasi *naive bayes* dengan nilai *accuracy* sebesar 53%, *precision* 52%, *recall* 53%, dan *f-1 score* 53%.

Hasil dari klasifikasi pada setiap aspek terbagi menjadi tweet positif, negatif, dan netral. Sebanyak 112 tweet *aspek pengisian baterai* terklasifikasi menjadi 80 tweet bersentimen positif, 24 tweet berentimen negatif, dan 8 tweet berentimen netral; 136 tweet *aspek produksi* terklasifikasi menjadi 46 positif, 55 negatif, dan 35 netral; 72 tweet *aspek perkembangan* terklasifikasi menjadi; 332 tweet *aspek subsidi* terklasifikasi menjadi 30 tweet positif, 29 negatif, dan 13 netral; 276 tweet *aspek wilayah* terklasifikasi menjadi 150 tweet positif, 70 negatif, dan 56 netral, dan 284 tweet *aspek merk* terklasifikasi menjadi 112 tweet bersentimen positif, 94 tweet negatif, dan 78 tweet sentimen netral.

**Kata Kunci:** *naïve bayes*, *latent Dirichlet allocation*, analisis sentimen berbasis aspek, pemodelan topik, klasifikasi, mobil listrik.

UIN SUNAN AMPEL  
S U R A B A Y A

## ABSTRACT

*The development of electric car sales in Indonesia has increased significantly over the years. However, its presence in Indonesia raises pros and cons among the public. So it is necessary to know the public's pros and cons of electric cars from the aspects that influence them and classify them. As a solution to find pros and cons from the public regarding an aspect that is considered less or more than electric cars. The method used to search for aspects uses latent Dirichlet allocation (LDA) topic modeling and naïve Bayes classifier to classify sentiment. LDA topic modeling produces the best coherence score of 0.453 with 6 topics. The resulting topics are then interpreted into aspect names, namely aspects of battery charging, production, development, subsidies, regions and brands. Each aspect was then classified as positive, negative and neutral sentiment and resulted in the performance of the naïve Bayes classification with an accuracy value of 53%, precision 52%, recall 53% and f-1 score 53%.*

*The results of the classification for each aspect are divided into positive, negative and neutral tweets. A total of 112 tweets regarding battery charging aspects were classified into 80 tweets with positive sentiment, 24 tweets with negative sentiment, and 8 tweets with neutral sentiment; 136 production aspect tweets were classified into 46 positive, 55 negative, and 35 neutral; 72 tweets on development aspects were classified into; 332 tweets regarding subsidy aspects were classified into 30 positive, 29 negative and 13 neutral tweets; 276 regional aspect tweets were classified into 150 positive, 70 negative, and 56 neutral tweets, and 284 brand aspect tweets were classified into 112 tweets with positive sentiment, 94 negative tweets, and 78 tweets with neutral sentiment.*

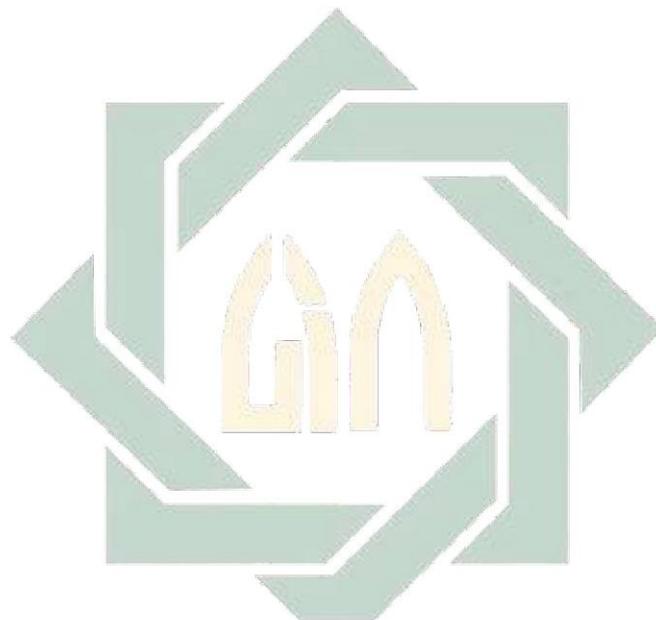
**Keywords:** *naïve Bayes, latent Dirichlet allocation, aspect-based sentiment analysis, topic modeling, classification, electric cars.*

UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR ISI

PERNYATAAN KEASLIAN .....	i
LEMBAR PERSETUJUAN PEMBIMBING.....	ii
PENGESAHAN TIM PENGUJI SKRIPSI .....	iii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	ii
BAB I.....	1
PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II .....	6
TINJAUAN PUSTAKA .....	6
2.1 Tinjauan Penelitian Terdahulu.....	6
2.2 Dasar Teori .....	9
2.2.1 <i>Text Mining</i> .....	9
2.2.2 <i>Aspect Based Sentiment Analysis (ABSA)</i> .....	9
2.2.3 Mobil Listrik.....	10
2.2.4 Twitter .....	11
2.2.5 <i>Web Scrapping</i> .....	11
2.2.6 <i>Preprocessing Text</i> .....	11
2.2.7 <i>TF - IDF</i> .....	13
2.2.8 <i>Topic Modeling</i> .....	13
2.2.9 <i>Latent Dirichlet Allocation (LDA)</i> .....	14
2.2.10 Uji Koherensi.....	17
2.2.11 <i>Cross Validation</i> .....	17
2.2.12 <i>Naïve Bayes Classifier</i> .....	18
2.2.13 <i>Confusion Matrix</i> .....	19
2.3 Integrasi Keilmuan .....	20
BAB III.....	23

METODOLOGI PENELITIAN .....	23
3.1    Desain Penelitian .....	23
BAB IV .....	30
PEMBAHASAN.....	30
4.1    Pengumpulan Data.....	30
4.2    Pemodelan Topik.....	41
4.3    Klasifikasi.....	51
4.4    Analisis Sentimen Berbasis Aspek .....	53
DAFTAR PUSTAKA.....	61



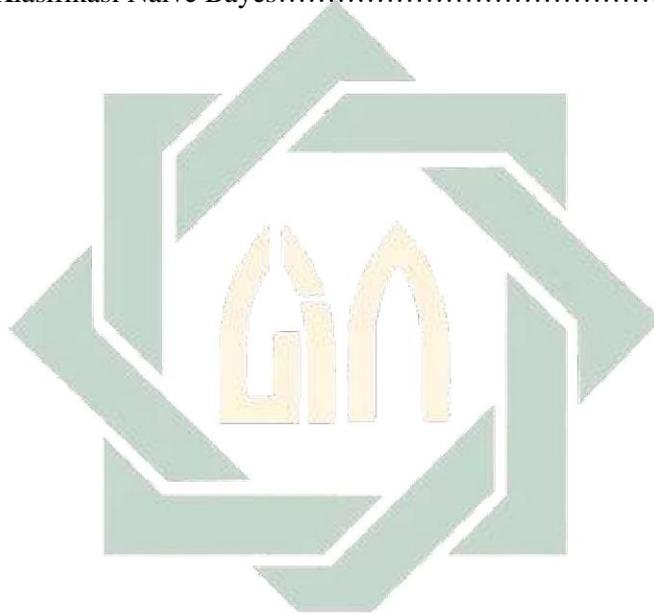
UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR TABEL

Tabel 2. 1 Penelitian-penelitian terdahulu	6
Tabel 2. 2 Contoh kata yang menginterpretasikan suatu aspek (Giffari, 2022)	16
Tabel 2. 3 Confusion Matrix (Caelen, 2017)	18
Tabel 4. 1 Data Tweet	30
Tabel 4. 2 Penghapusan Duplikasi Data	31
Tabel 4. 3 Penghapusan Missing Value	31
Tabel 4. 4 Hasil Cleaning Data	32
Tabel 4. 5 Hasil Case Folding	33
Tabel 4. 6 Data Label	34
Tabel 4. 7 Hasil Remove Number	35
Tabel 4. 8 Hasil Tokenizing	36
Tabel 4. 9 Hasil Remove Stopword	37
Tabel 4. 10 Hasil Normalization	37
Tabel 4. 11 Hasil Stemming	38
Tabel 4. 12 Hasil Filtering	39
Tabel 4. 13 Parameter LDA	40
Tabel 4. 14 Hasil parameter uji koherensi LDA	40
Tabel 4. 15 Clustering ke-1	41
Tabel 4. 16 Tweet Clustering ke-1	42
Tabel 4. 17 Clustering ke-2	42
Tabel 4. 18 Tweet Clustering ke-2	43
Tabel 4. 19 Clustering ke-3	43
Tabel 4. 20 Tweet Clustering ke-3	44
Tabel 4. 21 Clustering ke-4	44
Tabel 4. 22 Tweet Clustering ke-4	45
Tabel 4. 23 Clustering ke-5	46
Tabel 4. 24 Tweet Clustering ke-5	46
Tabel 4. 25 Clustering ke-6	47
Tabel 4. 26 Tweet Clustering ke-6	47
Tabel 4. 27 Confusion Matrix	49
Tabel 4. 28 Pengujian Performa	49
Tabel 4. 29 Klasifikasi Tweet Setiap Aspek	50

## DAFTAR GAMBAR

Gambar 2. 1 Topic Modeling.....	13
Gambar 2. 2 LDA Generative Process.....	14
Gambar 2. 3 Graphical Model LDA.....	14
Gambar 2. 4 Penjabaran Graphical Model LDA.....	15
Gambar 3. 1 Desain Penelitian.....	21
Gambar 3. 2 Alur Preprocessing.....	24
Gambar 3. 3 Pemodelan Topik.....	25
Gambar 3. 4 Proses Klasifikasi Naive Bayes.....	26



UIN SUNAN AMPEL  
S U R A B A Y A

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Teknologi dunia otomotif kini mengalami banyak perubahan dan perkembangan. Inovasi kendaraan bermotor mulai beralih dari mesin berbahan bakar diesel, bensin, dan gas kepada teknologi kendaraan listrik atau *electric vehicle (EV)*. Mobil listrik dibuat sebagai alternatif kendaraan yang ramah lingkungan dan dapat mengurangi pemakaian energi fosil (Aziz et al., 2020). Hadirnya mobil listrik juga mengurangi efek dari emisi gas rumah kaca serta mampu menghemat biaya dan lebih efisien. Bersumber dari (Dataindonesia.id, 2022) pada tahun 2019 mobil listrik di Indonesia terjual sebanyak 812, tahun 2020 sebanyak 1324, tahun 2021 sebanyak 3193, dan pada tahun 2022 mobil listrik di Indonesia sudah dipasarkan sebanyak 15437. Dari data tersebut dapat disimpulkan daya tarik masyarakat terhadap mobil listrik meningkat secara signifikan.

Meningkatnya kehadiran mobil listrik di Indonesia tidak lepas dari dukungan pemerintah. Berdasarkan peraturan presiden (Perpres) No 55 tahun 2019, pemerintah Indonesia mendukung penggunaan kendaraan listrik sebagai langkah awal untuk mengurangi pencemaran udara. Masyarakat Indonesia mulai menyadari manfaat penggunaan kendaraan listrik khususnya mobil listrik. Walaupun demikian, mobil listrik dinilai memiliki kelemahan dari segi biaya operasional, kelengkapan infrastruktur, dan performansi (Aziz et al., 2020). Fasilitas pendukung mobil listrik seperti Stasiun Pengisian Kendaraan Listrik Umum (SPKLU) atau *charger station* hanya dijumpai di beberapa kota besar dan belum merata di seluruh wilayah di Indonesia. Selain itu, teknisi ahli dalam menangani kerusakan dan perawatan mobil listrik tergolong sedikit. Muncul pro dan kontra di masyarakat dan sebagian berfikir dua kali untuk menggunakan mobil listrik. Sehingga perlu dilakukan analisis sentimen masyarakat terhadap kehadiran mobil listrik di Indonesia.

Analisis sentimen adalah studi terkait opini, penilaian, emosi, atau sikap masyarakat terhadap suatu produk, jasa, isu atau peristiwa yang terjadi (Fairuz et al., 2021). Analisis sentimen merupakan pendekatan *meachine learning* untuk menemukan pola dari data train untuk mengekstrak sebuah teks atau dokumen sehingga dapat ditentukan apakah sebuah teks bersifat positif, negatif, atau netral. Pada pendekatan terbaru, memungkinkan analisis sentimen untuk mencari subtopik dari sebuah teks. Pendekatan ini disebut sebagai “*Aspect Based Sentiment Analysis (ABSA)*” (García-Díaz et al., 2020). Sentimen analisis berbasis aspek digunakan untuk mencari subtopik dari setiap hasil sentiment positif, negatif, dan netral. Sehingga sentimen masyarakat dapat dilihat secara spesifik berdasarkan kata yang paling dominan pada setiap hasil sentimen.

Beberapa penelitian tentang analisis sentimen pernah dilakukan, salah satu diantaranya adalah penelitian yang dilakukan oleh (Yunanto & Yulianto, 2022) yang membahas tentang analisis sentimen aplikasi peduli lingkungan, analisis dilakukan terhadap 4636 data *tweet* dengan hasil klasifikasi 473 data memiliki sentimen negatif dan 4163 mempunyai sentimen positif. Analisis sentimen tersebut menghasilkan nilai akurasi sebesar 90%. Salah satu metode yang cukup populer dan memiliki tingkat akurasi yang cukup tinggi untuk melakukan analisis sentimen adalah *NBC (Naïve Bayes Classifier)* (Azhari et al., 2021). Penelitian mengenai *sentiment analysis* menggunakan metode *NBC* pernah dilakukan oleh (Gunawan et al., 2019) dengan hasil akurasi sebesar 77,78%. Dari penelitian-penelitian sebelumnya metode *naïve bayes classifier* terbukti memiliki nilai akurasi yang tinggi dan cocok digunakan untuk melakukan *sentiment analysis*.

Namun, metode *naïve bayes classifier* tidak bisa untuk mencari aspek-aspek yang terdapat di dalam dokumen, sehingga diperlukan metode *LDA (Latent Dirichlet Allocation)* untuk mencari topik yang akan digunakan sebagai aspek. Penelitian menggunakan metode *latent dirichlet allocation* pernah dilakukan oleh (Ali et al., 2022) untuk

menganalisis sentimen berbasis topik atau aspek dari web TripAdvisor. Penelitian tersebut menghasilkan empat topik yaitu *Jamaâ-el-Fna atmosphere*, *Shopping experience*, *Citizen's behaviour*, dan *Overall touristic experience*. Hasil analisis dari penggabungan klasifikasi *naïve bayes* dan pemodelan topik *latent dirichlet allocation* disebut analisis sentimen berbasis aspek. (Putu & Amrullah, 2021) pada penelitiannya tentang analisis sentimen berbasis aspek dengan metode *NBC* dan *LDA* menghasilkan nilai akurasi metode *NBC* sebesar 92%, sedangkan uji koherensi yang dilakukan pada *LDA* menghasilkan nilai koheren tertinggi sebesar 0,613 pada kelas positif dan 0,528 pada kelas negatif.

Penelitian analisis sentimen berbasis aspek menggunakan metode *naïve bayes classifier* dan juga *latent dirichlet allocation* diharapkan mampu menjawab permasalahan pro dan kontra masyarakat terhadap hadirnya mobil listrik di Indonesia. Selain itu, hasil dari penelitian ini juga dapat digunakan sebagai rujukan pemerintah dalam mengambil kebijakan yang berkaitan dengan mobil listrik di Indonesia dari segi aspek maupun ketertarikan masyarakat.

## **1.2 Perumusan Masalah**

1. Bagaimana performa *NBC* dan *LDA* dalam analisis sentimen berbasis aspek kehadiran mobil listrik di Indonesia?
2. Bagaimana analisis sentimen berbasis aspek terhadap hadirnya mobil listrik di Indonesia menggunakan *NBC* dan *LDA*?

## **1.3 Batasan Masalah**

1. Dataset bersumber dari Twitter berbahasa Indonesia.
2. Kata kunci yang digunakan dalam pengambilan data adalah *mobil listrik*.
3. Data diambil dari tanggal 01-02-2023 hingga 28-02-2023

## **1.4 Tujuan Penelitian**

1. Melakukan analisis sentimen berbasis aspek terhadap hadirnya mobil listrik di Indonesia menggunakan *naïve bayes classifier* dan *latent dirichlet allocation*.

2. Mengetahui performa *naïve bayes classifier* dan *latent dirichlet allocation* dalam mengklasifikasi sentimen terhadap hadirnya mobil listrik di Indonesia.

## **1.5 Manfaat Penelitian**

### **1.5.1 Manfaat Akademis**

1. Memberikan informasi kepada pembaca terhadap metode klasifikasi *naïve bayes classifier* dan *latent dirichlet allocation* dalam penerapannya untuk menganalisis sentimen berbasis aspek.
2. Menjadi referensi penelitian bagi mahasiswa sistem informasi.

### **1.5.2 Manfaat Praktis**

1. Membantu menjawab persoalan yang terjadi di masyarakat terkait penerimaan mobil listrik di Indonesia. Serta memberikan gambaran untuk beralih kepada mobil listrik, tetapi menggunakan mobil konvensional, atau bahkan tidak keduanya.
2. Menjadi referensi bagi pemerintah dalam menentukan kebijakan terkait kehadiran mobil listrik di Indonesia.

## **1.6 Sistematika Penulisan**

Sistematika penulisan yang digunakan pada penelitian ini sebagai berikut.

### **1. BAB I PENDAHULUAN**

Pada bagian pendahuluan, menjelaskan kronologi serta merumuskan permasalahan yang akan ditelusuri, pada bab ini juga dijelaskan tujuan serta manfaat dan sistematika kepenulisan terkait *aspect based sentiment analysis*.

### **2. BAB II TINJAUAN PUSTAKA**

Menguraikan terkait penelitian-penelitian sebelumnya, dasar teori, dan integrasi keilmuan. Penelitian terdahulu diambil dari jurnal yang berkaitan dengan metode dan topik yang relevan. Dasar – dasar teori menguraikan formula dan landasan dasar yang berkaitan dengan metode dan topik. Integrasi keilmuan dari pakar menjelaskan bagaimana dalil *Al Qur'an* dan *Hadist* yang sinkron terkait sentimen masyarakat maupun inovasi teknologi.

### 3. BAB III METODOLOGI PENELITIAN

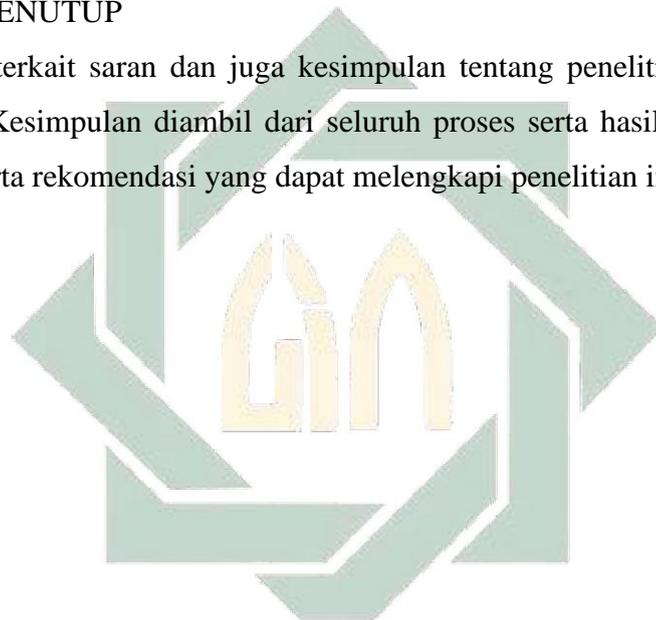
Menguraikan metode-metode yang dipakai serta alur dari penelitian yang dilakukan. Penelitian ini menggunakan jenis dan sumber data dari twitter. Alur penelitian yang dilakukan terbagi menjadi dua garis besar yaitu pengujian performa klasifikasi dan analisis sentimen berbasis aspek.

### 4. BAB IV HASIL DAN PEMBAHASAN

Menguraikan data, *output* dari klasifikasi, pemodelan topik, serta analisis hasil. Seluruh proses yang dilakukan membahas tentang jawaban dari rumusan masalah yang dibuat.

### 5. BAB V PENUTUP

Membahas terkait saran dan juga kesimpulan tentang penelitian yang dilakukan. Kesimpulan diambil dari seluruh proses serta hasil analisis sentimen serta rekomendasi yang dapat melengkapi penelitian ini.



UIN SUNAN AMPEL  
S U R A B A Y A

## BAB II TINJAUAN PUSTAKA

### 2.1 Tinjauan Penelitian Terdahulu

Tinjauan pustaka bertujuan memberikan pemahaman dan interpretasi yang sesuai dengan metode – metode yang digunakan oleh peneliti terdahulu dalam memecahkan masalah yang relevan. Relevansi penelitian yang telah dilaksanakan sebelumnya dapat ditinjau lebih detail pada tabel 2.1 sebagaimana berikut ini.

**Tabel 2. 1 Penelitian-penelitian terdahulu**

No	Judul	Metode	Hasil	Korelasi
1.	“Penerapan Metode <i>Adaboost</i> untuk Mengoptimasi Prediksi Penyakit Stroke dengan Algoritma <i>Naïve Bayes</i> ” (Byna & Basit, 2020)	<i>NBC</i> dan <i>Adaboost</i>	Metode <i>naïve bayes classifier</i> mampu menghasilkan nilai akurasi sebesar 97% dan <i>naïve bayes classifier</i> dengan <i>adaboost</i> mampu menghasilkan akurasi sebesar 98%	Penggunaan metode <i>naïve bayes classifier</i> cocok untuk digunakan pada penelitian klasifikasi karena memiliki akurasi yang tinggi tanpa harus dioptimalisasi.
2.	“Klasifikasi <i>Text Mining Review</i> Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma <i>Naïve Bayes</i> ”(Indrayuni, 2019)	<i>NBC</i> dan <i>10 Fold Cross Validation</i>	Metode <i>naïve bayes classifier</i> menghasilkan nilai akurasi terbaik sebesar 90,50%	Penggunaan metode klasifikasi <i>naïve bayes</i> dengan model evaluasi <i>10 fold cross validation</i> .
3.	“Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan <i>Naïve Bayes Classification</i> ” (Naraswati et al., 2021)	<i>NBC</i>	Berdasarkan hasil evaluasi didapatkan nilai akurasi sebesar 87%, sensitivitas 93%, dan spesifisitas 71%. Metode ini sudah cukup baik mengklasifikasikan sentimen negatif dan	Metode <i>naïve bayes classifier</i> cocok untuk diimplementasikan pada penelitian analisis sentimen

			positif.	
4.	“Perbandingan Algoritma <i>Naïve Bayes</i> , <i>K-NN</i> , dan <i>SVM</i> dalam Pengklasifikasian Sentimen Media Sosial” (Asnawi et al., 2021)	<i>NBC</i> , <i>KNN</i> , <i>SVM</i>	Dataset sebanyak 2000 data berupa ulasan di google play store, diperoleh akurasi untuk <i>NBC</i> 79%, <i>KNN</i> 50%, dan <i>SVM</i> 75%. <i>NBC</i> cocok untuk	Perbandingan metode <i>NBC</i> , <i>KNN</i> , dan <i>SVM</i> dengan bukti <i>NBC</i> pada data ulasan memiliki nilai akurasi terbaik
<b>No</b>	<b>Judul</b>	<b>Metode</b>	<b>Hasil</b>	<b>Korelasi</b>
			mengklasifikasi ulasan dari aplikasi media sosial	
5.	“ <i>Sentiment Analysis and Topic Detection of Mobile Banking Application Review</i> ”(Permana et al., 2020)	<i>NBC</i> , <i>LDA</i> , <i>K-Fold Cross Validation</i>	Data sebanyak 6194 dari google play store di klasifikasi menggunakan <i>naïve bayes</i> dan dievaluasi menggunakan <i>k-fold cross validation</i> dengan nilai akurasi sebesar 86% dan dilakukan <i>topic detection</i> algoritma <i>LDA</i>	Penggunaan <i>k-fold cross validation</i> sebagai evaluasi metode <i>naïve bayes</i> dan penggabungan metode <i>naïve bayes</i> dan <i>LDA</i>
6.	“Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode <i>Naive Bayes</i> ” (Gunawan et al., 2019)	<i>NBC</i>	Pembuatan sistem analisis sentimen ulasan dengan memanfaatkan algoritma <i>naïve bayes</i> dengan nilai evaluasi akurasi sebesar 77%	Penerapan algoritma <i>naïve bayes classifier</i> sebagai metode analisis sentimen.
7.	“Analisis Diskriminan Dengan <i>K-Fold Cross Validation</i> Untuk Klasifikasi Kualitas Air Di Kota Pontianak” (Mardiana et al.,	<i>K-Fold Cross Validation</i>	Data primer sebanyak 42 sampel air di kota Pontianak dilakukan klasifikasi menggunakan <i>K-Fold Cross Validation</i> untuk mengestimasi kesalahan prediksi.	Penggunaan <i>k-fold cross validation</i> yang digunakan untuk mencari model terbaik dengan mengestimasi kesalahan prediksi.

	2022)			
8.	<p>“Analisis Ulasan Indie Video Game Lokal pada Steam</p> <p>Menggunakan Analisis Sentimen dan Pemodelan Topik</p> <p>Berbasis <i>Latent Dirichlet Allocation</i>” (Febrianta et al., 2021)</p>	<i>NBC, LDA</i>	<p>Algoritma <i>naïve bayes classifier</i> digunakan untuk mengklasifikasi ulasan produk video game menggunakan aplikasi Rapidminer dan RStudio. Hasil dari klasifikasi sentimen menunjukkan nilai akurasi sebesar 75%. Selain itu, dihasilkan istilah yang paling sering muncul seperti “<i>story</i>”, “<i>music</i>”, “<i>character</i>”, dan “<i>art</i>” dengan metode LDA.</p>	<p>Penggabungan metode klasifikasi <i>naïve bayes</i> dengan pemodelan topik LDA untuk menganalisis sentimen berbasis aspek.</p>
<b>No</b>	<b>Judul</b>	<b>Metode</b>	<b>Hasil</b>	<b>Korelasi</b>
9.	<p>Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok</p> <p>Menggunakan Algoritma <i>Naive Bayes</i> dan <i>Latent Dirichlet Allocation</i> (Putu &amp; Amrullah, 2021)</p>	<i>NBC, LDA</i>	<p>Metode <i>naïve bayes</i> menghasilkan nilai akurasi sebesar 92% sedangkan untuk uji koherensi pada LDA menghasilkan nilai koherensi untuk kelas positif sebesar 0,613 dan kelas negatif sebesar 0,528.</p>	<p>Penggabungan metode klasifikasi <i>naïve bayes</i> dengan pemodelan topik LDA untuk menganalisis sentimen berbasis aspek.</p>
10.	<p>“Analisis Sentimen dan Pemodelan Topik Pandemi Covid-19 pada Media Sosial Twitter</p> <p>Menggunakan <i>Naïve Bayes Classifier</i> dan <i>Latent</i></p>		<p>Peneliti melakukan analisis sentimen pada 5000 data tweet menggunakan metode <i>naïve bayes</i> dan diperoleh nilai akurasi terbaik pada iterasi ke-7 sebesar 70% kemudian dilakukan pemodelan topik</p>	

	<i>Dirichlet Allocation</i> ”(Prakosa & Nasiroh, 2021)		menggunakan LDA dengan tingkat akurasi sebesar 78%.	
--	--	--	---	--

## 2.2 Dasar Teori

### 2.2.1 Text Mining

Penggalian teks (*text mining*) adalah sebuah penyelesaian masalah dengan memanfaatkan teknik *data mining*, *machine learning*, dan *natural language processing* dalam menggali informasi. Metode *text mining* digunakan untuk mengekstraksi informasi yang bersumber pada sebuah dokumen atau teks melalui proses identifikasi serta eksplorasi pola tertentu (Feldman & Sanger, 2007). Menurut (Latif, 2018) proses penggalian teks memiliki beberapa *step* yaitu *data preprocessing*, *data transformation*, serta *pattern discovery*. Jadi dapat dikatakan bahwa *text mining* merupakan sebuah langkah untuk menemukan pola yang tidak terstruktur pada teks dengan melalui beberapa tahap seperti *preprocessing*, *machine learning*, dan juga *natural language processing*.

### 2.2.2 Aspect Based Sentiment Analysis (ABSA)

Analisis sentimen menurut (Hartono et al., 2016) ialah bidang studi yang mempelajari pengolahan bahasa natural atau NLP untuk menganalisis perilaku, pendapat, serta menilai pandangan seseorang dalam sebuah teks maupun lisan yang disampaikan. Analisis sentimen juga diartikan sebagai penilaian, perilaku, emosi, dan evaluasi dari seseorang terhadap sebuah produk, jasa, organisasi, aplikasi, kejadian maupun sebuah topik (Liu, 2015). Sentimen maupun komentar yang disampaikan melalui teks maupun lisan menunjukkan ekspresi berupa positif (menguntungkan) dan negatif (tidak menguntungkan) terhadap suatu subjek (Nasukawa & Yi, 2003).

Analisis sentimen memiliki fungsi untuk mengelompokkan suatu data berupa teks maupun dokumen. Oleh karena itu, analisis sentimen juga disebut sebagai teknik *text mining*. Jadi dapat disimpulkan bahwa analisis sentimen merupakan sebuah langkah atau teknik yang digunakan untuk mencari arah pandang atau sentimen masyarakat pada sebuah topik, objek, kejadian, maupun sebuah produk. Sentimen tersebut dapat dikategorikan ke dalam sisi sentimen

positif, negatif, maupun netral. Sehingga dapat dipastikan sentimen positif adalah bentuk dukungan, negatif berupa penolakan atau tidak menguntungkan, serta netral sebagai bentuk keteguhan tidak berkomentar positif dan negatif, yang dapat disebabkan karena tidak adanya ketertarikan maupun keterkaitan pada seseorang.

Analisis sentimen berbasis aspek pada dasarnya sama dengan analisis sentimen biasa. Hanya saja terdapat penambahan *aspect-based* sebagai pengerucutan sentimen pada setiap aspeknya. Sebagai contoh dalam menganalisis mobil listrik tidak hanya menghasilkan sebuah sentimen, positif, negatif, maupun netral, namun juga aspek pada mobil listrik seperti performa, daya tahan baterai, perawatan, ketahanan pada cuaca dan lingkungan, polusi, dan lain sebagainya. Tujuan daripada *Aspect Based Sentiment Analysis (ABSA)* sejatinya ialah mengidentifikasi aspek dan sentimen dari hasil ekstraksi sebuah teks (Alghunaim, 2015).

### **2.2.3 Mobil Listrik**

Mobil listrik atau bisa disingkat molis merupakan mobil yang digerakkan menggunakan tenaga listrik (Wikipedia, 2023). Molis untuk kali pertamanya diperkenalkan oleh Robert Anderson yang berasal dari Scotlandia dalam rentang tahun 1832 hingga 1839. Pada awalnya mobil listrik ditenagai dengan *non-rechargeable primary power cells* atau baterai yang tidak dapat diisi ulang. Pada perkembangannya, mobil listrik mengalami kemunduran pada tahun 1908 dikarenakan bermunculannya mobil dengan bahan bakar fosil (BBM) yang memiliki harga beli dan bahan bakar yang murah pada tahun tersebut. Pada tahun 2000-an mobil listrik kembali bangkit dengan teknologi *hybrid* atau gabungan dari tenaga listrik dan BBM. Hingga akhirnya memasuki tahun 2020 banyak bermunculan mobil listrik yang murni dan memiliki fitur yang modern serta *rechargeable* (Maybank Finance, 2023). Pemerintah Indonesia juga mendukung adanya molis, sebagaimana dijelaskan di dalam Perpres No 55 tahun 2019 pemerintah Indonesia mendukung penggunaan kendaraan listrik sebagai langkah awal untuk mengurangi pencemaran udara, sehingga dapat dipastikan kehadiran mobil listrik di Indonesia akan meningkat pesat.

#### 2.2.4 Twitter

Salah satu platform digital yang saat ini ramai digunakan oleh masyarakat adalah Twitter. Platform ini dapat digunakan untuk mengirimkan *tweet* atau cuitan yang diciptakan oleh Jack Dorsey di tahun 2007 (Yudha Pratomo, 2021). Pengguna twitter dapat mengirimkan *tweet* berupa teks, foto, video, maupun tautan. Data hasil cuitan pengguna twitter dapat diolah dan dianalisis menjadi sebuah informasi. Informasi tersebut dapat digunakan sebagai informasi yang valid dan berharga karena mewakili pendapat maupun pemikiran setiap individu yang dituangkan pada media sosial twitter.

#### 2.2.5 Web Scrapping

*Web scrapping* merupakan salah satu *tools* yang dipakai untuk pengumpulan data yang bersumber dari internet untuk diolah dan dianalisis menjadi sebuah informasi yang diinginkan (Mitchell, 2018). Proses *scrapping* data twitter terdiri dari beberapa tahapan seperti mengunduh *tweet* dari halaman web twitter, mengekstrak data, menyimpan data berformat '.csv', lalu menganalisis data. Langkah tersebut dilakukan menggunakan bahasa pemrograman *python* karena memiliki *library* yang lengkap untuk memproses data.

#### 2.2.6 Preprocessing Text

Tahap *preprocessing text* adalah kumpulan dari beberapa tahapan pembersihan data teks dari hal-hal yang dapat mengganggu proses klasifikasi nantinya. Kumpulan dataset yang diambil dari twitter perlu dibersihkan dan diubah ke dalam format yang beraturan (Ganesan, 2019). Tahap *preprocessing text* berguna untuk meningkatkan kualitas data yang akan digunakan, sebab jika data yang digunakan memiliki kualitas yang kurang baik, maka hasil dari proses setelahnya memiliki kualitas yang kurang baik juga (Han & Kamber, 2006). Proses-proses pada tahap *preprocessing* ini diantaranya adalah (1) penyamaan format (*case-folding*); (2) pembersihan dari tanda baca, emotikon, serta *link* (*cleaning*); (3) proses pengubahan kalimat menjadi token (*tokenizing*); (4) *normalization*; (5) proses pengembalian kata kedalam bentuk dasarnya atau *stemming* (Herdhianto, 2020).

##### a. Case Folding

Menurut (Manning, 2008) *case folding* ialah proses untuk menyamaratakan teks dengan cara mengubah seluruh karakter menjadi huruf kecil atau *lowercase*. Hal ini bertujuan untuk menghindari adanya pembacaan duplikasi data. Sebagai contoh kata “Mobil” dan “mobil” akan dibaca sebagai dua kata yang berbeda padahal memiliki arti dan pelafalan yang sama. Sehingga perlu dilakukan *case folding* supaya *system* dapat membaca data dengan efektif.

b. *Cleaning*

*Cleaning* merupakan tahapan untuk membersihkan dataset dari karakter yang dapat mempengaruhi hasil sentimen. Karakter yang dibersihkan dapat berupa tanda baca, *url*, *retweet*, angka, simbol, *hashtag*, *link*, dan *mention* (Herdhianto, 2020).

c. *Tokenizing*

Menurut (Herdhianto, 2020) *tokenizing* atau tokenisasi merupakan tahapan untuk memecah *dataset* menjadi potongan – potongan kata atau *token*. Sebagai contoh kalimat “mobil listrik tidak bising dan berpolusi” akan dipecah menjadi [“mobil”, “listrik”, “tidak”, “bising”, “dan”, “berpolusi”]. Hal tersebut dapat memudahkan pada proses di tahapan selanjutnya.

d. *Normalization*

*Normalization* menurut (Manning, 2008) dapat dikatakan sebagai proses standarisasi bahasa yang tidak baku menjadi bahasa yang baku. Pada penelitian ini hanya diambil data teks berbahasa Indonesia, maka rujukan yang dipakai adalah Kamus Besar Bahasa Indonesia (KBBI). Pada tahap ini digunakan *Colloquial Indonesian Lexicon* yang memproses kata-kata gaul bahasa Indonesia menjadi kata yang baku (Salsabila et al., 2018).

e. *Stopword removal*

Langkah selanjutnya pada tahap *preprocessing* adalah *stopword removal*. Proses *stopword removal* diterapkan guna menghapus *conjunction* pada teks sehingga hanya menyisakan bagian yang diperlukan. Selain *conjunction* proses *stopword removal* juga diterapkan untuk menghapus kata yang memiliki frekuensi kemunculan paling banyak, sehingga mampu mengesampingkan kata yang lainnya dikarenakan kemunculannya terlalu

sering. Proses ini menggunakan *library stopwords\_indonesia* yang berisikan kata hubung berbahasa Indonesia seperti *dengan, dan, tidak, bukan,* dan lain lain (Albrecht et al., 2020).

*f. Stemming*

Menurut (Manning, 2008) *stemming* adalah proses menghilangkan kata yang memiliki imbuhan. Proses ini menggunakan *library sastrawi* untuk mengubah kata yang berimbuhan ke bentuk kata dasarnya, sehingga memudahkan proses pembobotan kata.

### 2.2.7 TF - IDF

Pembobotan kata atau *TF-IDF* dicetuskan oleh Sparck Jones berdasarkan intuisi heruistik (Al-Talib & Hassan, 2013). Pembobotan kata pada suatu teks lokal tanpa melihat frekuensi di dalam teks atau dokumen lainnya disebut sebagai *term frequency (TF)*. Sedangkan pembobotan kata yang mencakup keseluruhan frekuensi pada kumpulan teks atau dokumen diistilahkan dengan *invers document frequency (IDF)*. Skema dari frekuensi *term* atau bobot tiap kata akan dihitung dan dipertimbangkan. Apabila suatu kata memiliki bobot yang kecil berarti semakin jarang kemunculannya pada dokumen. Sehingga menimbulkan perbedaan yang signifikan pada kata tersebut (Karyono & Utomo, 2012).

Formula dari *TF-IDF* adalah :

$$TF - IDF_{(ti,dj)} = tf_{(ti,dj)} \log \log \frac{N}{ni} \quad (1)$$

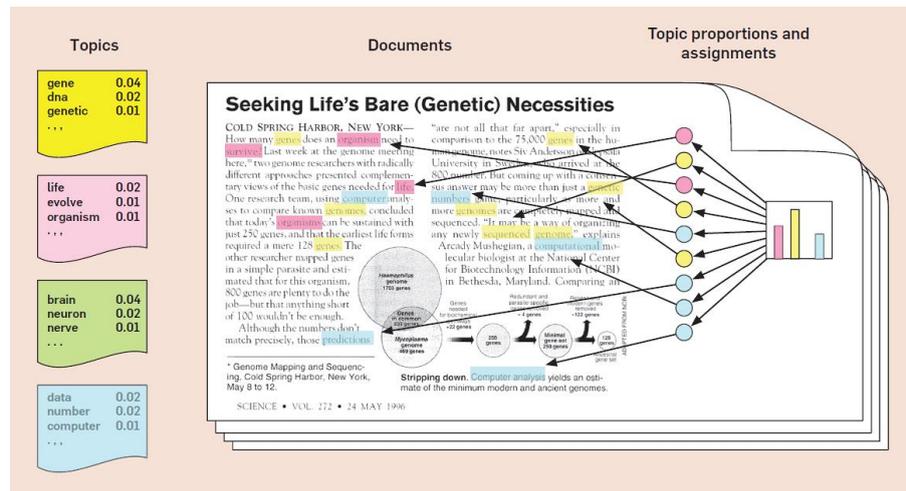
Pada formula (1) *tf* merepresentasikan frekuensi kata dari *term i* yang terdapat pada dokumen *j*. Kemudian *N* merepresentasikan total angka yang ada di dataset atau kumpulan dokumen.

### 2.2.8 Topic Modeling

Pemodelan topik merupakan sebuah konsep yang diterapkan guna mencari kata yang saling terhubung satu dengan yang lainnya di dalam entitas *word, document,* dan juga *corpus* (Blei et al., 2003). Definisi dari *word, document,* dan *corpus* menurut (Blei et al., 2003) adalah sebagai berikut.

1. *Word* atau kata ialah entitas dasar dari data tekstual di dalam dokumen.

2. *Document* ialah susunan  $N$  word dengan  $W = (w_1, \dots, w_n)$ . Pada formula tersebut  $W_n$  ialah kata ke- $n$  pada dokumen.
3. *Corpus* ialah kumpulan dari  $M$  dokumen dan sedangkan kumpulan dari *corpus* disebut *corpora*.  $D = (w_1, w_2, w_3, \dots, w_m)$  yaitu  $D_m$  merupakan dokumen ke- $m$  pada *corpus*.



Gambar 2. 1 *Topic Modeling*

Sumber : (Blei, 2012)

Secara pemahaman manusia, kumpulan dokumen adalah sebuah objek yang mampu untuk diamati secara langsung. Namun, berbeda halnya dengan topik dari sebuah dokumen dan pengelompokan setiap kata pada topik di dalam dokumen merupakan struktur yang tidak dapat diamati secara langsung (Blei, 2012). Jadi, secara sederhana *topic modeling* ialah proses penggalian topik pada suatu teks atau dokumen yang tersembunyi. Menurut (Putra & Kusumawardani, 2017), *topic modeling* memiliki tujuan meninjau kata perkata yang saling terhubung satu dengan yang lainnya sehingga mampu mewakili topik dari sebuah dokumen yang tidak terstruktur. Pada penelitian ini *topic modeling* digunakan untuk mencari nama aspek yang akan dianalisis.

### 2.2.9 *Latent Dirichlet Allocation (LDA)*

*Latent Dirichlet Allocation (LDA)* adalah salah satu teknik yang digunakan untuk pemodelan topik. Menurut (Blei et al., 2003) dalam penelitiannya menyebutkan bahwa "*LDA is a generative probabilistic model of a corpus*" atau LDA bisa disebut sebagai proses untuk menghasilkan model

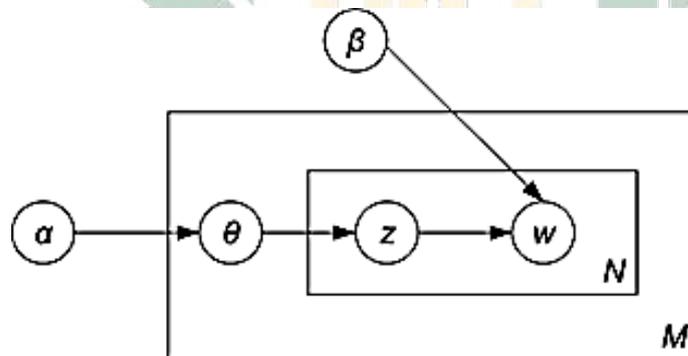
dari probabilitas yang terdapat pada corpus. LDA diasumsikan dengan proses generatif pada dokumen.

- LDA assumes the following generative process for each document  $w$  in a corpus  $D$ :
1. Choose  $N \sim \text{Poisson}(\xi)$ .
  2. Choose  $\theta \sim \text{Dir}(\alpha)$ .
  3. For each of the  $N$  words  $w_n$ :
    - (a) Choose a topic  $z_n \sim \text{Multinomial}(\theta)$ .
    - (b) Choose a word  $w_n$  from  $p(w_n | z_n, \beta)$ , a multinomial probability conditioned on the topic  $z_n$ .

Gambar 2. 2 LDA Generative Process

Sumber : (Blei et al., 2003)

Proses generatif dokumen pada gambar 2.1 menjelaskan perubahan yang terjadi secara acak pada  $N$  pada distribusi Poisson ( $\xi$ ). Setiap kata perlu diberikan topik yang sesuai dengan distribusi multinomial dari parameter  $k$ -vector  $\theta$ .  $k$ -vector  $\theta$  merupakan variabel acak yang selalu mengikuti  $\text{Dir}(\alpha)$ . Pada *graphical model* berikut ini dapat dilihat proses yang bekerja pada metode LDA.

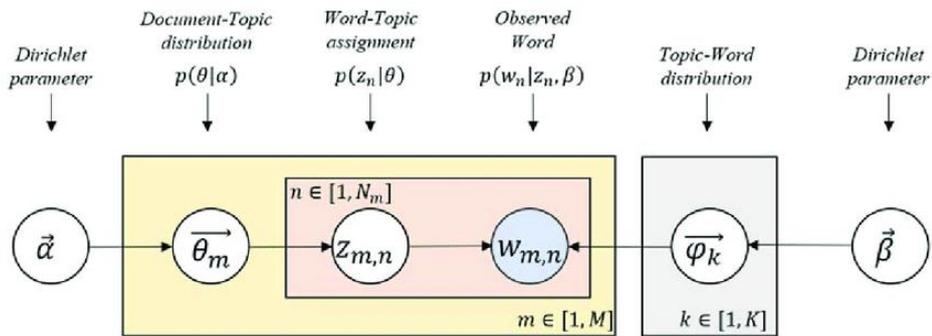


Gambar 2. 3 Graphical Model LDA

Sumber : (Blei et al., 2003)

Pada gambar 2.3 dapat dilihat parameter  $\alpha$  dan  $\beta$ . Kedua parameter menunjukkan distribusi topik. Parameter  $\alpha$  menunjukkan distribusi topik yang berasal dari dokumen, sedangkan  $\beta$  menunjukkan distribusi topik yang berasal dari kata. Semakin besar nilai pada parameter  $\alpha$  dan  $\beta$  berarti mengandung semakin banyak topik dari kata atau dokumen serta tidak banyak kata spesifik yang membedakan antara topik satu dengan lainnya. Langkah – langkah pada

LDA dimulai dari menentukan nilai  $\alpha$  dan  $\beta$  serta melakukan tokenisasi dan memberikan nilai pada setiap kata (Arslina & Liebenlito, 2019). Representasi dari model LDA pada gambar 2.3 di atas menjelaskan fungsi densitas dari distribusi marginal sesuai dengan persamaan berikut ini.



Gambar 2. 4 Penjabaran *Graphical Model LDA*

Sumber : (Lee et al., 2018)

Dari gambar di atas dapat ditarik persamaan sebagai berikut.

$$p(\alpha, \beta) = \int (\theta|\alpha) \left( \prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta \quad (2)$$

Keterangan

$p(\theta|\alpha)$  : Distribusi *dirichlect* dokumen berdasarkan topik

$p(z_n|\theta)$  : Kemungkinan kemunculan suatu topik di dalam dokumen

$p(w_n|z_n, \beta)$  : Kemungkinan kemunculan suatu kata di dalam topik

Tabel 2. 2 Contoh kata yang menginterpretasikan suatu aspek (Giffari, 2022)

Kata	Aspek
Mudah, jelas, konsisten, komunikatif, interaktif	<i>User interface</i>
Kesesuaian, efisien, bagus, membantu, visibilitas, puas	<i>User experience</i>
Proses, layanan, kinerja, mendukung, cepat, berfungsi	<i>Functionality and performance</i>
Update, kritik, saran, harapan, tanya	<i>Support and updates</i>

### 2.2.10 Uji Koherensi

Uji koherensi merupakan suatu langkah yang digunakan untuk mengukur nilai sebuah topik antara kata yang satu dengan kata lainnya. Uji koherensi perlu dilakukan untuk mengukur kesesuaian topik dari hasil *topic modeling* dengan topik dari pandangan manusia. Salah satu metode yang digunakan adalah *topic coherence*, yang mana dapat mengukur nilai suatu topik berdasarkan tingkat kesamaan semantik. (Putra & Kusumawardani, 2017) mengukur kesuksesan dari interpretasi yang dihasilkan pada *topic modeling* menggunakan metode *topic coherence*. Hasil *topic coherence* yang baik ditunjukkan dengan *coherence score* yang tinggi. Menurut Wisdom (2017) *topic coherence* mampu memberikan interpretasi topik yang lebih baik dibandingkan menggunakan *perplexity*. Terkadang, matriks *perplexity* tidak memberikan korelasi yang baik pada interpretasi model yang dilakukan oleh manusia (Listari, 2019).

### 2.2.11 Cross Validation

*Cross validation* merupakan metode yang digunakan untuk mengukur kinerja maupun performa dari sebuah *meachine learning*. Menurut (Sina et al., 2019), *cross validation* ialah metode yang dilakukan untuk membagi partisi dari data secara bergantian yang mana akan difungsikan sebagai data uji sekali. Prinsip dasarnya ialah membagi menjadi data *testing* dan juga data *training* (Davison & Hinkley, 1997). Metode *cross validation* akan mengestimasi seberapa baik model yang dihasilkan dari *meachine learning* bekerja. Proses *cross validation* melibatkan langkah-langkah sebagai berikut.

1. Pembagian dataset

Dataset akan dibagi menjadi *data training* (data latih) dan *data testing* (data uji) secara bergantian. Data training akan dicari nilai probabilitasnya sesuai dengan label yang ada. Sedangkan data testing akan dihilangkan labelnya untuk menguji model yang dihasilkan. Sehingga seluruh data akan melalui fase validasi.

2. Iterasi

Iterasi atau perulangan pada dataset akan dikelompokkan sebanyak *k* *sample* dengan ukuran yang sama. Setiap kelompok *sample* disebut

sebagai *fold*. Setiap *training set* menggunakan fungsi prediksi *k-1 fold*, sedangkan *fold* lainnya digunakan untuk *testing set*. Pada umumnya *k-fold cross validation* menggunakan  $k = 5$  dan  $k = 10$  (Sina et al., 2019).

### 3. Pengukuran kinerja

Setiap iterasi akan mengukur kinerja model klasifikasi menggunakan metrix evaluasi yang relevan. Hasil dari metrix evaluasi nantinya dijumlahkan dan dicari rata-ratanya sehingga dapat memberikan estimasi performa model secara keseluruhan.

### 4. Pemilihan model

Evaluasi dan perhitungan rata-rata hasil metrix akan memilih *best model* (model dengan performa yang terbaik).

#### 2.2.12 Naïve Bayes Classifier

Algoritma klasifikasi *naïve bayes* merupakan salah satu metode klasifikasi yang populer. Metode ini ditemukan pada tahun 1702 oleh Thomas Bayes, seorang ilmuwan asal Inggris. Algoritma ini menerapkan teorema bayes pada masa lampau yang digunakan untuk mengklasifikasi metode probabilitas dan statistik (Chai et al., 2002). *Teorema bayes* diasumsikan bahwa ada dan tidaknya ciri tertentu pada sebuah kelas yang tidak ada hubungannya dengan ciri kelas lainnya (Bustami, 2013). Berikut merupakan formula dari *teorema bayes*.

$$P(hj) = \frac{(P(hj) * P(ci))}{P(hj)} \quad (3)$$

Keterangan

$P(ci/hj)$  : kemungkinan kemunculan kejadian *ci* dengan syarat *hj*

$P(hj/ci)$  : kemungkinan kemunculan kejadian *hj* dengan syarat *ci*

$P(ci)$  : kemungkinan kemunculan kejadian *ci*

$P(hj)$  : kemungkinan kemunculan kejadian *hj*

Pada persamaan (3), formula NBC kemudian dikategorikan dan dicari probabilitas yang tertinggi ( $C_{MAP}$ ) dan dapat disederhanakan menjadi persamaan di bawah ini.

$$C_{MAP} = p(ci)p(hj|ci) \quad (4)$$

Persamaan (4) menerangkan probabilitas kemunculan hipotesis pada dokumen  $p(ci)$  dan probabilitas kemunculan hipotesis yang terdapat di dalam kelas  $p(hj|ci)$  (Ariadi, 2022).

### 2.2.13 Confusion Matrix

Salah satu model evaluasi data mining pada klasifikasi teks adalah *confusion matrix*. Metode ini melakukan kalkulasi daya uji benar dan salah hasil sentimen. Parameter yang digunakan pada model evaluasi *confusion matrix* yaitu *TP (True Positif)*, *FN (False Negative)*, *TN (True Negative)*, dan *FP (False Positif)* (Astuti, 2020). Model evaluasi *confusion matrix* dijelaskan lebih detail sebagai berikut.

Tabel 2. 3 Confusion Matrix (Caelen, 2017)

Aktual	Prediksi	
	Positif	Negatif
Positif	True Positif (TP)	False Positif (FP)
Negatif	False Negatif (FN)	True Negatif (TN)

*Confusion matrix* pada tabel 2.3 berfungsi sebagai acuan untuk mencari nilai dari *accuracy*, *recall*, *precision*, dan *f-measure*.

#### 1. Accuracy

*Accuracy* ialah perhitungan rasio total seluruh data yang sesuai antara prediksi dan aktual.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

#### 2. Precision

*Precision* ialah perhitungan rasio dari prediksi negatif maupun positif yang benar dengan total hasil prediksi.

$$Precision\ positif = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Precision negatif} = \frac{TN}{TN + FN} \quad (7)$$

### 3. Recall

*Recall* merupakan rasio dari prediksi benar baik negatif atau positif yang dibandingkan dengan total data yang aktual benar negatif atau positif. *Recall* positif disebut sebagai *sensitivity* dan *recall* negatif disebut sebagai *specificity*.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (9)$$

### 4. F-measure

*F-measure* merupakan rasio dari rata-rata *precision* dan *recall* dengan persamaan sebagai berikut.

$$F - \text{measure positif} = 2 \left( \frac{\text{precision positif} * \text{sensitivity}}{\text{precision positif} + \text{sensitivity}} \right) \quad (10)$$

$$F - \text{measure negatif} = 2 \left( \frac{\text{precision negatif} * \text{specificity}}{\text{precision negatif} + \text{specificity}} \right) \quad (11)$$

## 2.3 Integrasi Keilmuan

Integrasi keilmuan merupakan penyesuaian antara ilmu umum dengan ilmu agama. Sehingga penelitian yang dilakukan memiliki dasar keilmuan agama yang jelas. Dasar keilmuan dapat berupa ayat Al Qur'an maupun Hadits yang berhubungan dengan penelitian yang dilakukan. Integrasi keilmuan dilakukan dengan cara wawancara kepada ahli. Wawancara dilakukan kepada Bapak Tisngi Hanani, S.Ag. selaku guru agama di SMP Masyithoh Kroya, Cilacap. Hasil wawancara integrasi keilmuan tentang analisis sentimen dan inovasi teknologi mobil listrik adalah sebagai berikut.

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا يَسْخَرْ قَوْمٌ مِّنْ قَوْمٍ عَسَىٰ أَن يَكُونُوا خَيْرًا مِّنْهُمْ  
وَلَا نِسَاءٌ مِّنْ نِّسَاءٍ عَسَىٰ أَن يَكُنَّ خَيْرًا مِّنْهُنَّ وَلَا تَلْمِزُوا أَنفُسَكُمْ وَلَا

تَنَابَزُوا بِالْأَلْقَابِ بِئْسَ الْأَسْمُ الْفُسُوقُ بَعْدَ الْإِيمَانِ وَمَنْ لَّمْ يَتُوبْ فَأُولَٰئِكَ  
۝۱۱ هُمُ الظَّالِمُونَ

Terjemahan : “Wahai orang-orang yang beriman, janganlah suatu kaum mengolok-olok kaum yang lain (karena) boleh jadi mereka (yang diolok-olokkan itu) lebih baik daripada mereka (yang mengolok-olok) dan jangan pula perempuan-perempuan (mengolok-olok) perempuan lain (karena) boleh jadi perempuan (yang diolok-olok itu) lebih baik daripada perempuan (yang mengolok-olok). Janganlah kamu saling mencela dan saling memanggil dengan julukan yang buruk. Seburuk-buruk panggilan adalah (panggilan) fasik setelah beriman. Siapa yang tidak bertobat, mereka itulah orang-orang zalim.” (Q.S. Al-Hujurat : 11) (Kementerian Agama Republik Indonesia, 2023).

Dari ayat di atas dijelaskan bahwa kita sebagai umat muslim dilarang memiliki prasangka buruk atau berfikir negatif terhadap orang lain atas perbedaan pendapat maupun pilihan. Adanya sentimen negatif sebagai bentuk penolakan terhadap kehadiran mobil listrik hendaknya tidak mengolok-olok maupun merendahkan masyarakat yang memiliki sentimen atau pendapat positif terhadapnya. Sebab di dalam Q.S. Al-Hujurat ayat 11 kita dilarang untuk saling mencela dan merendahkan. Adanya perbedaan pendapat merupakan hal yang wajar, tetapi tidak dibenarkan untuk menyalahkan pendapat orang lain.

وَمَا لَهُمْ بِهِ مِنْ عِلْمٍ إِنْ يَتَّبِعُونَ إِلَّا الظَّنَّ وَإِنَّ الظَّنَّ لَا يُغْنِي مِنَ  
الْحَقِّ شَيْئًا ۝۲۸

Terjemahan : “Padahal, mereka tidak mempunyai pengetahuan tentang hal itu. Mereka tidak lain hanyalah mengikuti dugaan dan sesungguhnya dugaan itu tidak berfaedah sedikit pun terhadap kebenaran”. (Q.S. An-Najm : 28) (Kementerian Agama Republik Indonesia, 2023)

Sebagai pendukung, ayat di atas juga menjelaskan terkait etika berpendapat ataupun bersentimen. Ketika berpendapat, hendaknya kita memahami terlebih dahulu kebenaran dari suatu topik dalam hal ini adalah

kehadiran mobil listrik di Indonesia. Pada ayat di atas ditegaskan bahwasanya bersentimen tanpa memiliki dasar kebenaran maka dugaan tersebut tidaklah berfaedah sedikitpun. Artinya adalah ketika seseorang memiliki sentimen positif maupun negatif terhadap kehadiran mobil listrik, tidak berpengaruh terhadap fakta sesungguhnya. Misalnya seseorang bersentimen bahwa “mobil listrik justru lebih ribet digunakan karena sulit menemukan SPKLU” tidak berpengaruh terhadap fakta bahwasanya SPKLU sudah banyak bermunculan di kota-kota besar.



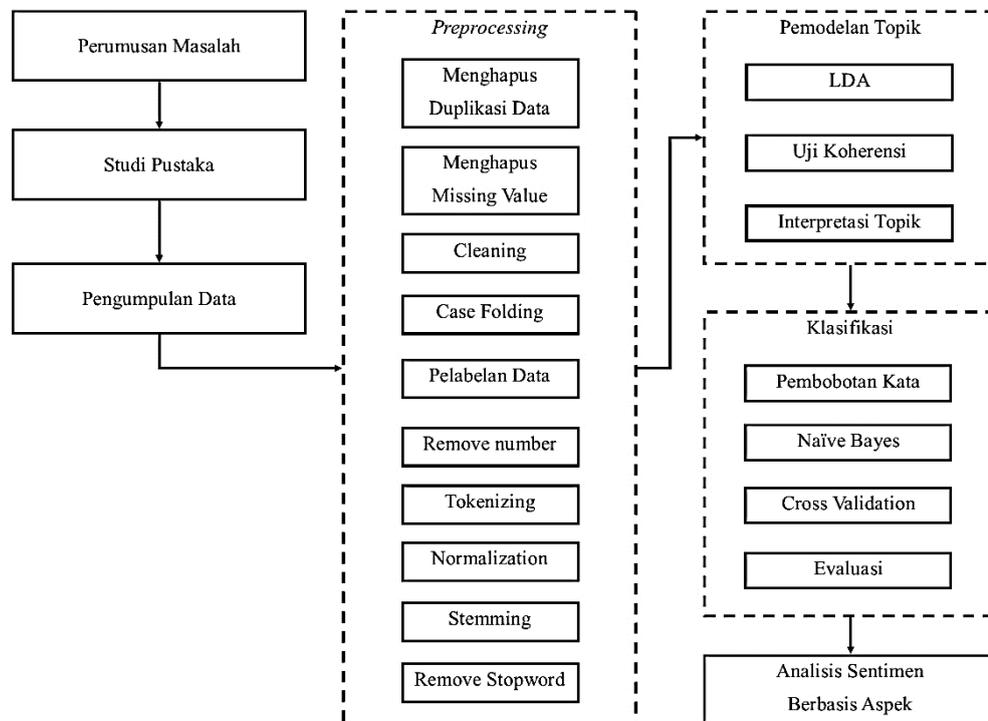
UIN SUNAN AMPEL  
S U R A B A Y A

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Desain Penelitian

Desain penelitian menjelaskan beberapa tahapan penelitian yang akan dilakukan. Pada penelitian ini terdiri dari dua *output*, *output* pertama adalah performa dari klasifikasi menggunakan *naïve bayes classifier* dan performa dari pemodelan topik menggunakan *latent dirichlet allocation*. Kedua, analisis sentimen berbasis aspek dari hasil klasifikasi dan pemodelan topik. Desain penelitian yang dilakukan dapat dilihat pada gambar 3.1 berikut.



*Gambar 3. 1* Desain Penelitian

Tahapan penelitian dimulai dari menentukan rumusan masalah. Kemudian, dari permasalahan yang didapatkan dilakukan studi pustaka untuk mencari referensi yang berkaitan dengan penyelesaian masalah. Setelah itu dilakukan pengumpulan data dan juga *preprocessing*. Setelah data bersih, dilanjutkan dengan melakukan pemodelan topik untuk mencari kemunculan setiap kemungkinan topik yang akan dijadikan nama aspek serta dilakukan uji koherensi. Berikutnya, dilakukan pelabelan data untuk menginterpretasikan

hasil dari pemodelan topik dan juga untuk pelabelan sentimen untuk klasifikasi. Selanjutnya dilakukan pembobotan kata untuk menghitung setiap bobot kata di dalam dokumen dan dilanjutkan dengan klasifikasi menggunakan *naïve bayes*. Tahapan terakhir yaitu menguji performa dari klasifikasi *naïve bayes* dan dilakukan analisis terhadap hasil klasifikasi sentimen menggunakan *naïve bayes* berdasarkan aspeknya. Penjelasan secara rinci pada setiap tahapan akan dijelaskan pada uraian berikut ini.

### **3.1.1 Perumusan masalah**

Langkah awal sebelum melakukan penelitian adalah merumuskan masalah. Pada penelitian ini permasalahan yang akan dicari solusinya adalah klasifikasi sentimen berdasarkan aspek yang terdapat pada mobil listrik serta mengukur performa dari metode klasifikasi tersebut. Metode yang digunakan adalah *naïve bayes classifier* untuk mengklasifikasi sentimen serta *latent dirichlet allocation* untuk pemodelan topik atau pencarian nama aspek.

### **3.1.2 Studi Pustaka**

Studi pustaka dilakukan untuk menggali referensi memahami solusi dari permasalahan yang didapatkan sesuai dengan metode yang akan diimplementasikan. Studi pustaka juga diperlukan untuk mencari teori-teori maupun pendapat para ahli yang mendasari topik penelitian ini. Sumber referensi yang digunakan pada penelitian ini adalah :

1. Jurnal ilmiah

Jurnal ilmiah digunakan sebagai referensi penelitian terdahulu yang memiliki korelasi dengan topik penelitian serta metode yang digunakan. Jurnal ilmiah penelitian terdahulu disajikan pada tabel 2.1 di atas.

2. E-book

E-book merupakan sumber referensi yang digunakan untuk menggali teori, konsep, serta algoritma terkait metode dan teknik-teknik yang digunakan di dalam penelitian.

3. Skripsi

Skripsi maupun laporan penelitian yang bersumber dari institusi akademis maupun pemerintah juga dapat digunakan sebagai sumber

referensi. Penelitian yang dituangkan ke dalam skripsi dijadikan acuan untuk menganalisis proses dan *output* dari metode yang digunakan.

#### 4. Artikel

Artikel yang bersumber dari media *online* dijadikan referensi pendukung untuk melihat perkembangan dan berita terbaru mengenai topik dan objek yang akan diteliti.

### 3.1.3 Pengumpulan Data

Data merupakan hal paling utama pada penelitian ini, dikarenakan ketersediaannya berpengaruh pada lamanya proses dan hasil analisis klasifikasi sentimen.

#### 1. Jenis data

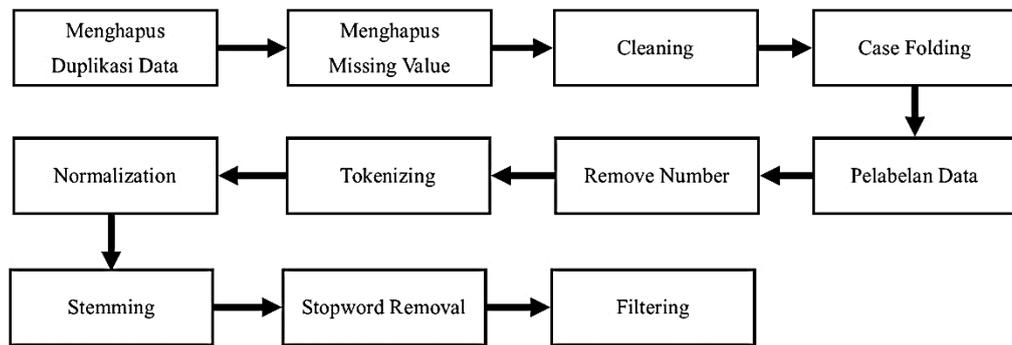
Pada penelitian ini, data yang digunakan adalah data primer. Data primer merupakan data yang diambil dan dikumpulkan secara langsung dari sumbernya oleh peneliti (Sugiyono, 2018).

#### 2. Sumber data

Pengumpulan data dilakukan menggunakan teknik *web scrapping* untuk mengambil data tweet yang berisikan opini masyarakat terkait kehadiran mobil listrik di Indonesia. *Scrapping data* yang dilakukan menggunakan pemrograman *python* di *google colab*. Sebelum dilakukan *web scrapping*, terlebih dahulu dilakukan *filtering keyword, date, dan language* melalui *tools twitter advanced*. *Keyword* yang dicari adalah “*mobil listrik*”. Data yang diambil mulai dari tanggal 1 Februari 2023 hingga 28 Februari 2023.

### 3.1.4 Preprocessing

Data yang sudah diperoleh akan dilakukan *preprocessing* untuk menghasilkan data yang siap untuk diproses lebih lanjut. Tahapan ini terdiri dari beberapa bagian sebagai berikut.



Gambar 3. 2 Alur Preprocessing

Berdasarkan gambar 3.2 di atas, terdapat beberapa tahapan *preprocessing*. Tahapan pertama yaitu menghapus duplikasi data, proses ini penting untuk dilakukan karena data yang diambil dari twitter dapat berupa *retweet*, sehingga banyak data yang sama dan perlu dihilangkan duplikasinya. Kedua menghapus *missing value*, penghapusan *missing value* dilakukan untuk mencegah kolom yang memiliki nilai *null* pada setiap baris data. Ketiga *cleaning*, merupakan tahapan untuk membersihkan dataset dari karakter yang dapat mempengaruhi hasil sentimen. Karakter yang dibersihkan berupa tanda baca, *url*, simbol, *whitespace*, *hashtag*, *link*, dan *mention*. Keempat *case folding*, proses ini digunakan untuk menyamaratakan teks dengan cara mengubah seluruh karakter menjadi huruf kecil atau *lowercase*.

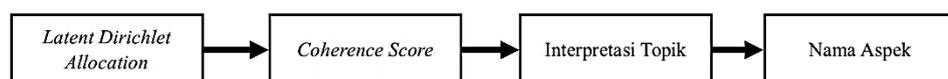
Setelah dilakukan *case folding*, data yang sudah dibersihkan dari karakter yang tidak dibutuhkan dan disamakan format teks menjadi *lowercase* dilakukan pelabelan data. Pelabelan data dibutuhkan untuk metode *supervised learning*, khususnya *naïve bayes classifier*. Pelabelan data dilakukan secara manual dengan melibatkan ahli bahasa dari Balai Bahasa Jawa Timur. Ahli bahasa merupakan orang yang mampu memberikan pelabelan data sesuai dengan teks tweet yang tersedia. Pada pelabelan yang akan dilakukan, terdapat 3 ahli bahasa yang expert dalam bidang sentimen dan bersedia untuk melabeli data yang digunakan. Label data terdiri dari label sentimen positif, negatif dan juga netral. Label positif merepresentasikan ketertarikan dan dukungan masyarakat terhadap kehadiran mobil listrik di Indonesia, label negatif merepresentasikan kebalikannya, sedangkan label netral merupakan label sentimen yang tidak tergolong positif dan negatif.

Keenam *remove number*, menghilangkan numerik yang terdapat di dalam teks. Ketujuh *tokenizing*, merupakan tahapan untuk memecah kalimat menjadi potongan-potongan kata atau *token*. Kedelapan *normalization*, proses standarisasi bahasa yang tidak baku (*slangword*) menjadi bahasa yang baku. Tahap kesembilan *stemming*, proses menghilangkan kata yang memiliki imbuhan. Proses ini menggunakan *library sastrawi* untuk mengubah kata yang berimbuhan ke dalam bentuk kata dasarnya, sehingga memudahkan proses selanjutnya. Kesepuluh *Stopword removal*, tahap ini diterapkan untuk menghapus *conjunction* pada teks, sehingga hanya menyisakan bagian yang diperlukan. Proses ini menggunakan *library stopwords\_indonesia* yang berisikan kata hubung berbahasa Indonesia seperti *dengan, dan, tidak, bukan,* dan lain lain.

Kesebelas *filtering*, proses ini diperlukan untuk memfilter data dengan panjang token atau kata lebih dari 4 kata. Proses *filtering* dibutuhkan untuk memudahkan tahap pemodelan topik dengan asumsi suatu kalimat yang dapat merepresentasikan sebuah topik setidaknya membutuhkan unsur SPOK di dalam kaidah bahasa Indonesia (Rianti, 2020) (Suprihatin, 2017).

### 3.1.5 Pemodelan Topik

Tahapan yang dilakukan untuk memodelkan sebuah topik hingga menetapkan sebuah nama aspek dapat dilihat pada gambar 3.3 dibawah ini.



Gambar 3. 3 Pemodelan Topik

Proses pemodelan topik berfungsi untuk mencari kemungkinan kemunculan topik yang terdapat pada teks. Proses ini dilakukan menggunakan metode *latent dirichlet allocation*. Metode ini akan memunculkan *topic keyword* yang akan di-*clustering* berdasarkan identifikasi topik. Selanjutnya dilakukan uji koherensi untuk mengukur sejauh mana topik-topik yang dihasilkan dapat mewakili dokumen yang terdapat dalam korpus. Uji koherensi dilakukan untuk mengukur seberapa kedekatan topik yang dihasilkan oleh

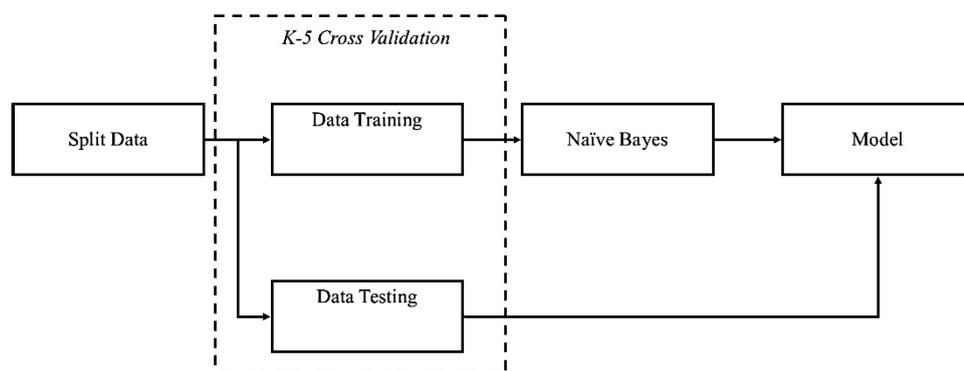
*latent dirichlet allocation*. Uji koherensi dilakukan menggunakan metode *coherence score*. Metode ini akan memberikan nilai pada suatu topik berdasarkan tingkat kesamaan semantik. Tahapan berikutnya adalah menginterpretasikan hasil dari *topic keyword* yang dimunculkan kedalam *cluster* topik. Interpretasi *topic keyword* dilakukan dengan cara wawancara kepada alumni mahasiswa teknik elektro untuk memberikan nama aspek pada setiap *cluster* topik yang muncul.

### 3.1.6 Pembobotan Kata

Sebelum dilakukan klasifikasi, terlebih dahulu data dikonversi menjadi numerik. Proses ini disebut sebagai pembobotan kata atau *TF-IDF* yang diterapkan untuk menghitung perolehan bobot pada tiap-tiap kata. Pembobotan kata dilakukan dengan menerapkan *library python TF-IDF Vectorizer*. Tahap ini sangat penting dilakukan sebelum dilakukannya klasifikasi karena mengubah struktur dokumen yang berupa kata menjadi struktur dokumen berupa numerik.

### 3.1.7 Klasifikasi

Proses klasifikasi merupakan proses utama pada penelitian ini. Proses ini akan menghasilkan model kelas sentimen menggunakan metode *naïve bayes classifier*, sehingga mampu mengklasifikasikan teks sesuai dengan label sentimennya berdasarkan aspek. Proses klasifikasi menggunakan *naïve bayes* dapat dilihat pada gambar 3.4 berikut ini.



Gambar 3. 4 Proses Klasifikasi Naive Bayes

Tahap pertama, data berlabel akan dilakukan split data dengan *ratio* 80:20, dengan *data training* sebesar 80% dan *data testing* sebesar 20%. Tahap

kedua, yaitu melakukan *k-fold cross validation* dengan jumlah  $k=5$ , sehingga setiap iterasi  $k$  akan memilih *data training* dan *data testing* untuk diimplementasikan pada *naïve bayes classifier*. Proses *cross validation* akan menghasilkan *best model* dari keseluruhan iterasi yang dilakukan. *Best model* tersebut kemudian digunakan untuk melakukan evaluasi *naïve bayes classifier* dalam mengklasifikasi teks.

### **3.1.8 Evaluasi**

Proses evaluasi atau pengujian performa dilakukan untuk mengukur akurasi yang didapatkan dari model klasifikasi *naïve bayes* pada setiap aspek yang dihasilkan pada proses pemodelan topik menggunakan *latent dirichlet allocation*. Proses evaluasi ini menggunakan metode *confusion matrix*, metode ini bertujuan untuk menilai tingkat akurasi, *recall*, presisi, serta *f-measure*. Hasil dari proses evaluasi dapat memberikan tingkat kesesuaian metode yang dilakukan dengan objek penelitian yang diambil.

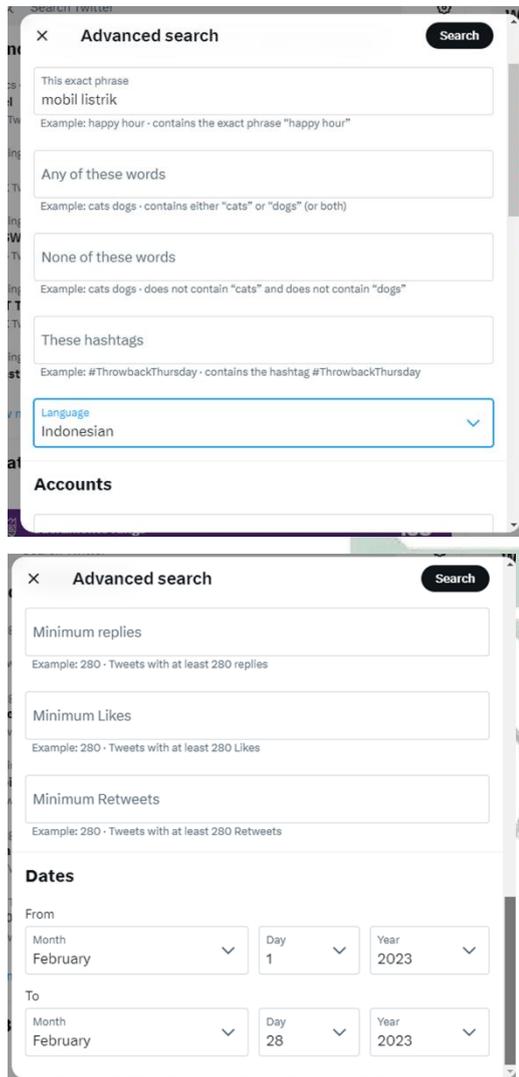
### **3.1.9 Analisis Sentimen Berbasis Aspek**

Tahap terakhir pada penelitian ini adalah memvisualisasikan dan menganalisis hasil sentimen masyarakat berdasarkan aspek yang didapatkan tentang kehadiran mobil listrik di Indonesia. Setiap aspek akan dijabarkan menjadi masing – masing aspek positif, aspek negatif, dan aspek netral. Pada masing – masing sentimen aspek inilah yang dapat dijadikan pertimbangan bagi masyarakat untuk menentukan pilihannya terhadap penggunaan mobil listrik atau tetap menggunakan mobil BBM di Indonesia.

## BAB IV PEMBAHASAN

### 4.1 Pengumpulan Data

Data yang digunakan pada penelitian ini didapatkan melalui proses *web scrapping* dari media sosial twitter. Sebelum melakukan proses *scrapping* data, terlebih dahulu dilakukan proses *filtering keyword, language, dan date* menggunakan *twitter advance*. Proses *filtering* bertujuan untuk memudahkan pencarian pada saat *scrapping* data menggunakan google colab.



The image shows two screenshots of the Twitter Advanced Search interface. The top screenshot displays the following sections:

- Advanced search** (with a 'Search' button)
- This exact phrase**: Input field containing 'mobil listrik'. Example: 'happy hour - contains the exact phrase "happy hour"'
- Any of these words**: Input field. Example: 'cats dogs - contains either "cats" or "dogs" (or both)'
- None of these words**: Input field. Example: 'cats dogs - does not contain "cats" and does not contain "dogs"'
- These hashtags**: Input field. Example: '#ThrowbackThursday - contains the hashtag #ThrowbackThursday'
- Language**: Dropdown menu set to 'Indonesian'.
- Accounts**: Input field.

The bottom screenshot displays the following sections:

- Advanced search** (with a 'Search' button)
- Minimum replies**: Input field. Example: '280 - Tweets with at least 280 replies'
- Minimum Likes**: Input field. Example: '280 - Tweets with at least 280 Likes'
- Minimum Retweets**: Input field. Example: '280 - Tweets with at least 280 Retweets'
- Dates**:
  - From**: Month (February), Day (1), Year (2023)
  - To**: Month (February), Day (28), Year (2023)

Gambar 4.1 *Filtering Twitter Advance*

Dari proses *filtering* yang dilakukan pada *twitter advance* didapatkan *twitter search* untuk memudahkan proses pengambilan data pada proses *web*

*scrapping* menggunakan google colab. Data yang dihasilkan dari proses *scrapping* berjumlah 3355 baris dan 3 kolom berupa *tweet.id*, *tweet.user.username*, dan *tweet.content*, sebagaimana pada tabel 4.1 berikut ini.

**Tabel 4. 1 Data Tweet**

Tweet Id	Tweet	Username
1630356938994360000	@abu_mahdzuroh Ngapain berdemo, ayo kita naik motor listrik dan mobil listrik, gunakan jalan raya. Hayuk hayukk 😊	si_saitama
1628610003530810000	@rahmani_ima Ya iyalah jadi n gk banyak drama, kgk dipegang anies ya diem2 aje, apalagi partai giring, pejabat pemprov DKI mau ganti mobil listrik aj didukung, mw ini event rugi jg gk peduli, motogp gk jelas untung ruginya aje diem2 bae 🤔🤔🤔	bluegogeta168
1629828635581350000	Hyundai Indonesia Pamer Ioniq 6, Apa Sudah Dijual? –Mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum. <a href="https://t.co/hqIP1fpru6">https://t.co/hqIP1fpru6</a>	otodriver
1630320000000000000	Elon Musk tamu VVIP mudah mudahan mobil listrik jadi murah di Indonesia <a href="https://t.co/bMYttoFJKX">https://t.co/bMYttoFJKX</a> #KTTG20	VionaVolandaPl s
1630220000000000000	Soal Stasiun Pengisian Mobil Listrik di SPBU Solo, Pertamina: Masih Evaluasi <a href="https://t.co/Y5TLyIqUWN">https://t.co/Y5TLyIqUWN</a>	soloposdotcom

Data yang akan diproses pada penelitian ini hanya teks cuitan pada kolom ‘Tweet’. Data ‘Tweet’ mengandung teks cuitan masyarakat yang berpendapat atau memiliki sentimen terhadap kehadiran Data tersebut masih mengandung unsur seperti *hashtag*, *mention*, *link*, *emoticon*, *number*, serta tanda baca yang memungkinkan untuk mengganggu proses pada pemodelan

topik dan klasifikasi. Perlu dilakukan *preprocessing* untuk menghasilkan data yang bersih dan siap untuk diproses.

#### 4.1. *Preprocessing*

*Preprocessing* merupakan tahap untuk mempersiapkan data yang bersih dan siap digunakan untuk pemodelan topik dan klasifikasi pada proses berikutnya. Rangkaian *preprocessing* terdiri dari menghapus duplikasi data, menghapus *missing value*, *cleaning*, *case folding*, pelabelan data, *remove number*, *tokenizing*, *stopword removal*, *normalization*, dan *stemming*. Penjelasan mendetail pada setiap rangkaian sebagai berikut.

##### 4.2.1. Menghapus Duplikasi Data

Proses penghapusan duplikasi data bertujuan untuk mengantisipasi adanya data yang sama atau duplikat. Proses ini mengaplikasikan fungsi *drop\_duplicates* pada data 'Tweet'. Hasil dari proses penghapusan duplikasi data sebagai berikut.

**Tabel 4. 2 Penghapusan Duplikasi Data**

	Sebelum	Sesudah
<b>Jumlah data</b>	3355	3355

Penghapusan data yang dilakukan tidak memberikan perubahan pada jumlah data, sehingga dapat dipastikan tidak ada data yang duplikat.

##### 4.2.2. Menghapus *Missing Value*

Sebelum menghapus *missing value* atau data yang kosong, terlebih dahulu dilakukan pengecekan data yang memiliki nilai null. Pengecekan *missing value* menggunakan fungsi *df.isna().sum()* kemudian untuk menghapus data yang *missing value* menggunakan fungsi *df.records\_dropped = df.dropna(axis=0, how='any')*. Hasil dari proses ini sebagai berikut.

**Tabel 4. 3 Penghapusan Missing Value**

	Sebelum	Sesudah

<b>Jumlah data</b>	3355	3355
--------------------	------	------

Penghapusan *missing value* yang dilakukan tidak menghasilkan perubahan pada jumlah data, sehingga data pada kolom ‘Tweet’ tidak memiliki baris yang kosong atau null.

#### 4.2.3. Cleaning

Dokumen pada kolom ‘Tweet’ masih mengandung *mention, hashtag, emoticon, special character, URL, number, multiple whitespace, single char,* dan *punctuation*. Oleh karena itu, perlu dilakukan pembersihan atau *cleaning* data. Hasil pada proses ini dapat dilihat pada tabel berikut ini.

**Tabel 4. 4 Hasil Cleaning Data**

Sebelum	Sesudah
@abu_mahdzuroh Ngapain berdemo, ayo kita naik motor listrik dan mobil listrik, gunakan jalan raya. Hayuk hayukk 😊	Ngapain berdemo ayo kita naik motor listrik dan mobil listrik gunakan jalan raya Hayuk hayukk
@rahmani_ima Ya iyalah jadi n gk banyak drama, kgk dipegang anies ya diem2 aje, apalagi partai giring, pejabat pemprov DKI mau ganti mobil listrik aj didukung, mw ini event rugi jg gk peduli, motogp gk jelas untung ruginya aje diem2 bae 😂😂😂	Ya iyalah jadi gk banyak drama kgk dipegang anies ya diem2 aje apalagi partai giring pejabat pemprov DKI mau ganti mobil listrik aj didukung mw ini event rugi jg gk peduli motogp gk jelas untung ruginya aje diem2 bae
Hyundai Indonesia Pamer Ioniq 6, Apa Sudah Dijual? Mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum. <a href="https://t.co/hqlP1fpru6">https://t.co/hqlP1fpru6</a>	Hyundai Indonesia Pamer Ioniq 6 Apa Sudah Dijual Mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum
Elon Musk tamu VVIP mudah mudahan mobil listrik jadi murah di Indonesia	Elon Musk tamu VVIP mudah mudahan mobil listrik jadi murah di Indonesia

<a href="https://t.co/bMYttoFJKX">#KTTG20</a>	
Soal Stasiun Pengisian Mobil Listrik di SPBU Solo, Pertamina: Masih Evaluasi <a href="https://t.co/Y5TLyIqUWN">https://t.co/Y5TLyIqUWN</a>	Soal Stasiun Pengisian Mobil Listrik di SPBU Solo Pertamina Masih Evaluasi

Pada tabel 4.4 dapat dilihat perbedaan pada data sebelum (sebelah kiri) dan data sesudah (sebelah kanan) dilakukan proses *cleaning*. Data yang belum *dicleaning* masih memiliki *mention*, *link*, *hashtag*, *punctuation*, *emoticon*, *number*, *special character*, *URL*, *multiple whitespace*, serta, *single char*. Setelah dilakukan *cleaning* data sudah bersih dari hal-hal tersebut. Akan tetapi teks pada dokumen masih belum memiliki standarisasi huruf besar dan kecil, sehingga dilakukan proses berikutnya yaitu *case folding*.

#### 4.2.4. Case folding

Proses *case folding* dilakukan karena data teks belum distandarisasi dan dapat menyebabkan adanya kesalahan dalam mendeteksi suatu kata yang sesungguhnya memiliki makna dan arti yang sama, tetapi terdapat perbedaan penulisan huruf besar dan kecil. Sehingga, system akan mendeteksi dan membacanya secara berbeda dan mendapatkan bobot yang berbeda. Dalam mengatasi hal tersebut perlu dilakukan penyamaan format teks yang dilakukan dengan cara *case folding*. Proses *case folding* menggunakan fungsi `data['Tweet'].str.lower()` berikut merupakan hasil dari proses *case folding*.

Tabel 4. 5 Hasil Case Folding

Sebelum	Sesudah
Ngapain berdemo ayo kita naik motor listrik dan mobil listrik gunakan jalan raya Hayuk hayukk	ngapain berdemo ayo kita naik motor listrik dan mobil listrik gunakan jalan raya hayuk hayukk
Ya iyalah jadi gk banyak drama kgk dipegang anies ya diem aje apalagi partai giring pejabat pemprov DKI mau ganti mobil listrik aj didukung mw ini event rugi	ya iyalah jadi gk banyak drama kgk dipegang anies ya diem2 aje apalagi partai giring pejabat pemprov dki mau ganti mobil listrik aj didukung mw ini event rugi

jg gk peduli motogp gk jelas untung ruginya aje diem bae	jg gk peduli motogp gk jelas untung ruginya aje diem <sup>2</sup> bae
Hyundai Indonesia Pamer Ioniq Apa Sudah Dijual Mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum	hyundai indonesia pamer ioniq 6 apa sudah dijual mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum
Elon Musk tamu VVIP mudah mudahan mobil listrik jadi murah di Indonesia	elon musk tamu vvip mudah mudahan mobil listrik jadi murah di indonesia
Soal Stasiun Pengisian Mobil Listrik di SPBU Solo Pertamina Masih Evaluasi	soal stasiun pengisian mobil listrik di spbu solo pertamina masih evaluasi

Pada tabel 4.4 dapat dilihat perbedaan pada data sebelum (sebelah kiri) dan data sesudah (sebelah kanan) dilakukan proses *cleaning*. Data yang belum *dicleaning* masih memiliki *mention, link, hashtag, punctuation, emoticon, number, special character, URL, multiple whitespace*, serta, *single char*. Setelah dilakukan *cleaning* data sudah bersih dari hal-hal tersebut. Akan tetapi teks pada dokumen masih belum memiliki standarisasi huruf besar dan kecil, sehingga dilakukan proses berikutnya yaitu *case folding*.

#### 4.2.5. Pelabelan Data

Pelabelan data dilakukan untuk melabeli sentimen yang akan digunakan untuk klasifikasi menggunakan *naïve bayes classifier*. Pelabelan data dilakukan oleh pakar dari Balai Bahasa Jawa Timur yang melibatkan 3 orang ahli bahasa. Pelabel telah melabeli data dan mengelompokkan data berdasarkan kelas sentimen positif, negatif, dan netral. Data tweet yang memiliki sentimen berupa dukungan terhadap kehadiran mobil listrik di Indonesia dilabeli dengan sentimen positif. Data yang memiliki kecenderungan berupa kritik negatif serta penolakan kehadiran mobil listrik di Indonesia dilabeli dengan sentimen negatif. Serta data tweet yang tidak memiliki sentimen positif maupun negatif termasuk kategori sentimen netral. Masing-masing dari ahli bahasa akan memberikan label sentimen sehingga didapatkan label sama pada seluruh pelabel dan disepakati sebagai label yang

digunakan. ketika label yang diberikan oleh pelabel terdapat perbedaan maka, akan dilakukan voting berdasarkan 2 label yang sama. Sedangkan, apabila dari seluruh pelabel memiliki label sentimen yang berbeda pada suatu teks maka akan dikategorikan label yang tidak disepakati dan tidak digunakan. Setelah seluruh data memiliki label, selanjutnya akan digunakan untuk membuat model klasifikasi *naïve bayes* serta menguji performanya.

**Tabel 4. 6 Data Label**

Label	Jumlah
Positif	1643
Negatif	973
Netral	696
Tidak Sepakat	43
<b>Total data</b>	3312

#### 4.2.6. Remove Number

Proses *remove number* bertujuan untuk menghilangkan number atau numerik yang masih tersisa untuk keperluan pelabelan data. Proses ini akan menghilangkan unsur angka yang terdapat pada setiap teks di dalam dokumen. Hasil dari proses *remove number* tersebut dapat dilihat pada tabel 4.6 dibawah ini.

**Tabel 4. 7 Hasil Remove Number**

Sebelum	Sesudah
ngapain berdemo ayo kita naik motor listrik dan mobil listrik gunakan jalan raya hayuk hayukk	ngapain berdemo ayo kita naik motor listrik dan mobil listrik gunakan jalan raya hayuk hayukk

ya iyalah jadi gk banyak drama kkg dipegang anies ya diem2 aje apalagi partai giring pejabat pemprov dki mau ganti mobil listrik aj didukung mw ini event rugi jg gk peduli motogp gk jelas untung ruginya aje diem2 bae	ya iyalah jadi gk banyak drama kkg dipegang anies ya diem aje apalagi partai giring pejabat pemprov dki mau ganti mobil listrik aj didukung mw ini event rugi jg gk peduli motogp gk jelas untung ruginya aje diem bae
hyundai indonesia pamer ioniq 6 apa sudah dijual mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum	hyundai indonesia pamer ioniq apa sudah dijual mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum
elon musk tamu vvip mudah mudahan mobil listrik jadi murah di indonesia	elon musk tamu vvip mudah mudahan mobil listrik jadi murah di indonesia
soal stasiun pengisian mobil listrik di spbu solo pertamina masih evaluasi	soal stasiun pengisian mobil listrik di spbu solo pertamina masih evaluasi

Hasil dari proses ini dapat dilihat pada tabel 4.6 sebelah kanan (sesudah). Teks sudah bersih dari unsur numerik dan dapat diproses lebih lanjut.

#### 4.2.7. Tokenizing

Proses *tokenizing* bertujuan untuk mengubah struktur kalimat menjadi token atau kata perkata. Pada proses ini, setiap kata yang dipisahkan dengan spasi pada satu kalimat akan dijadikan satu token. Hasil dari proses *tokenizing* tersebut dapat dilihat pada tabel 4.6 dibawah ini.

**Tabel 4. 8 Hasil Tokenizing**

Sebelum	Sesudah
hyundai indonesia pamer ioniq apa sudah dijual mobil listrik yang dilengkapi dengan berbagai macam fitur ini dapat dilihat secara langsung masyarakat umum	['hyundai', 'indonesia', 'pamer', 'ioniq', 'apa', 'sudah', 'dijual', 'mobil', 'listrik', 'yang', 'dilengkapi', 'dengan', 'berbagai', 'macam', 'fitur', 'ini', 'dapat', 'dilihat', 'secara', 'langsung', 'masyarakat',

	'umum']
solusinya semua penumpang kcl harusnya beli mobil listrik mumpung subsidi jt ye kaann	['solusinya', 'semua', 'penumpang', 'kcl', 'harusnya', 'beli', 'mobil', 'listrik', 'mumpung', 'subsidi', 'jt', 'ye', 'kaann']
mobil listrik mg ev mejeng di iims diklaim telah dipesan sebanyak lebih dari unit	['mobil', 'listrik', 'mg', 'ev', 'mejeng', 'di', 'iims', 'diklaim', 'telah', 'dipesan', 'sebanyak', 'lebih', 'dari', 'unit']
esemka memamerkan minivan esimka bima ev di iims mobil listrik ini masih diimpor secara utuh alias completely buildup unit cbu dari china	['esemka', 'memamerkan', 'minivan', 'esemka', 'bima', 'ev', 'di', 'iims', 'mobil', 'listrik', 'ini', 'masih', 'diimpor', 'secara', 'utuh', 'alias', 'completely', 'buildup', 'unit', 'cbu', 'dari', 'china']
trik manjur wuling air ev jadi mobil listrik terlaris indonesia	['trik', 'manjur', 'wuling', 'air', 'ev', 'jadi', 'mobil', 'listrik', 'terlaris', 'indonesia']

Proses *tokenizing* ini memisahkan kata atau token dengan tanda petik dan diikuti koma sebagai pemisahannya. Bentuk token seperti ini sangat berguna pada setiap proses berikutnya.

#### 4.2.8. *Remove Stopword*

Proses *remove stopword* bertujuan untuk menghilangkan kata penghubung yang tidak akan memiliki makna ketika berdiri sendiri serta kata-kata yang terlalu umum dan memiliki frekuensi kemunculan yang tinggi seperti kata 'mobil', 'listrik'. Hasil dari proses ini dapat dilihat pada tabel 4.7 sebagai berikut ini.

**Tabel 4. 9 Hasil Remove Stopword**

Sebelum	Sesudah
['hyundai', 'indonesia', 'pamer', 'ioniq', 'apa', 'sudah', 'dijual', 'mobil', 'listrik',	['hyundai', 'indonesia', 'pamer', 'ioniq', 'dijual', 'dilengkapi', 'fitur', 'langsung',

'yang', 'dilengkapi', 'dengan', 'berbagai', 'macam', 'fitur', 'ini', 'dapat', 'dilihat', 'secara', 'langsung', 'masyarakat', 'umum']	'masyarakat']
['trik', 'manjur', 'wuling', 'air', 'ev', 'jadi', 'mobil', 'listrik', 'terlaris', 'indonesia']	['trik', 'manjur', 'wuling', 'air', 'ev', 'terlaris', 'indonesia']
['mau', 'curi', 'perhatian', 'pasar', 'ngapain', 'pake', 'mobil', 'listrik', 'harga', 'setengah', 'milyar', 'mestinya', 'nyasar', 'ke', 'l'gc', 'dulu']	['curi', 'perhatian', 'pasar', 'ngapain', 'pake', 'harga', 'milyar', 'mestinya', 'nyasar', 'l'gc']

Hasil sesudah penghapusan *stopword* terlihat pada sisi sebelah kanan pada tabel 4.7 dengan hasil lebih bersih dari kata hubung daripada sebelum dilakukan penghapusan *stopword* pada sisi sebelah kiri.

#### 4.2.9. Normalization

Proses normalisasi data atau normalization bertujuan untuk mengubah kata yang tidak baku atau *slang word* ke dalam kata yang baku. Sehingga dapat memperkecil kemungkinan terjadinya kesalahan system dalam membaca kata dengan gaya bahasa yang berbeda dengan makna dan arti yang sama yang disebabkan adanya penggunaan bahasa gaul. Hasil dari proses ini dapat dilihat pada tabel 4.8 sebagai berikut ini.

**Tabel 4. 10 Hasil Normalization**

Sebelum	Sesudah
['hyundai', 'indonesia', 'pamer', 'ioniq', 'dijual', 'mobil', 'listrik', 'dilengkapi', 'fitur', 'langsung', 'masyarakat']	['hyundai', 'indonesia', 'pamer', 'ioniq', 'dilengkapi', 'fitur', 'masyarakat']
['solusi', 'mobil', 'listrik', 'udah', 'gitu', 'pemerintah', 'titik', 'ngga', 'pake', 'koma']	['solusi', 'begitu', 'pemerintah', 'titik', 'enggak', 'pakai', 'koma']

['trik', 'mobil', 'listrik', 'china', 'subsidi', 'indonesia', 'pake', 'nama', 'merk', 'lokal', 'dirakit', 'indonesia', 'komponennya', 'dr', 'china', 'dpt', 'subsidi']	['trik', 'subsidi', 'indonesia', 'pakai', 'nama', 'merk', 'lokal', 'dirakit', 'indonesia', 'komponennya', 'dari', 'dapat', 'subsidi']
--	---

Kata yang tidak baku akan dilakukan normalisasi sehingga menjadi kata yang baku. Seperti pada tabel 4.8 sebelah kiri terdapat kata ‘pake’, setelah dinormalisasi menjadi kata ‘pakai’ serta beberapa singkatan seperti ‘dpt’ diubah menjadi ‘dapat’.

#### 4.2.10. Stemming

Stemming merupakan proses yang mengubah suatu kata atau token menjadi kata dasarnya dan menghilangkan kata imbuhan menggunakan *library sastrawi*. Proses stemming menggunakan library sastrawi dikarenakan sudah umum digunakan serta berisikan dataset berbahasa Indonesia. Seperti beberapa kata yang terdapat pada dataset akan diubah menjadi kata “duga” dari asal kata “diduga”, “lengkap” dari asal kata “selengkapnya” dan “temu” dari asal kata “pertemuan”.

**Tabel 4. 11 Hasil Stemming**

Sebelum	Sesudah
['mobil', 'listrik', 'esemka', 'diduga', 'neta', 'murah', 'wuling', 'air', 'ev']	['mobil', 'listrik', 'esemka', 'duga', 'neta', 'murah', 'wuling', 'air', 'ev']
['simak', 'informasi', 'selengkapnya', 'tren', 'penggunaan', 'mobil', 'listrik', 'indonesia', 'urgensinya', 'crisis', 'bem', 'feb', 'unud', 'kreatif', 'aspiratif', 'kolaboratif']	['simak', 'informasi', 'lengkap', 'tren', 'guna', 'mobil', 'listrik', 'indonesia', 'urgensi', 'crisis', 'bem', 'feb', 'unud', 'kreatif', 'aspiratif', 'kolaboratif']
['puan', 'maharani', 'ketua', 'golkar', 'airlangga', 'hartarto', 'pertemuan',	['puan', 'maharani', 'ketua', 'golkar', 'airlangga', 'hartarto', 'temu', 'kawasan',

'kawasan', 'monas', 'mencoba', 'mobil', 'listrik']	'monas', 'coba', 'mobil', 'listrik']
--	--------------------------------------

Hasil dari proses *stemming* dapat dilihat pada sisi sebelah kanan pada tabel 4.9 yang sudah menjadi kata dasar dari kata imbuhan yang semula mengikutinya. Hasil dari proses *stemming* yang akan digunakan untuk proses pemodelan topik dan juga klasifikasi nantinya.

#### 4.2.11. Filtering

Proses *filtering* dilakukan untuk memfilter data yang tidak memiliki topik dalam suatu kalimat. Pada proses *filtering* data akan dihitung panjang token dan difilter berdasarkan jumlah token pada kalimat. Data yang memiliki panjang token kurang dari 9 akan dihapus baris datanya menggunakan fungsi `data.drop(data[data['panjang'] < 9].index)` sehingga data yang memiliki panjang token lebih dari sama dengan 9 akan dipertahankan untuk proses selanjutnya (Jha & Mahmoud, 2019).

**Tabel 4. 12 Hasil Filtering**

	Sebelum	Sesudah
<b>Jumlah data</b>	3312	1326

## 4.2 Pemodelan Topik

### 4.2.12. Uji Koherensi

Data yang sudah dilakukan *filtering* data akan digunakan untuk pencarian nama aspek menggunakan *topic modeling*. Sebelum itu, dilakukan uji koherensi untuk menemukan *coherence score* terbaik. Pengujian terdiri dari parameter  $n$  sebagai banyaknya topik, parameter  $a$  sebagai distribusi topik yang berasal dari dokumen, parameter  $b$  sebagai distribusi topik yang berasal dari kata. Kombinasi nilai pada setiap parameter dapat mempengaruhi hasil *coherence score*, sehingga diperlukan proses *looping* untuk menginputkan masing masing parameter. Parameter yang digunakan untuk pengujian koherensi dapat dilihat pada tabel berikut.

**Tabel 4. 13 Parameter LDA**

Parameter	Input
$n$	[ 2, 3, 4, 5, 6 ]
$alpha$	[ 'symmetric', 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 ]
$beta / eta$	[ 'auto', 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 ]

Pada tabel di atas, masing-masing parameter akan diinputkan dan dikombinasikan dengan parameter yang lain, sehingga muncul nilai *coherence score* yang terbaik. Semakin besar nilai  $n$  pada parameter maka semakin kompleks topik yang dihasilkan. Sedangkan pada nilai  $a$  atau *alpha*, semakin tinggi besar nilai yang diinputkan menghasilkan nilai *coherence score* yang tinggi tetapi dapat memberikan kompleksitas setiap kata yang muncul dari dokumen. Kemudian untuk parameter  $b$  atau *beta* menunjukkan seberapa banyak distribusi topik yang dihasilkan dari kata, pada parameter yang besar, distribusi topik yang dimunculkan akan semakin mengerucut dan sulit untuk dibedakan. Sehingga kombinasi dari parameter  $n$ ,  $alpha$ , dan  $beta$  juga memperhatikan dari kemudahan dalam membedakan topik yang satu dengan lainnya atau menginterpretasikan sebuah topik. Berikut merupakan hasil dari uji koherensi yang dilakukan.

**Tabel 4. 14 Hasil parameter uji koherensi LDA**

$n$	$a$	$b$	<i>coherence score</i>
3	0.6	0.3	0.407
4	0.6	0.4	0.433

6	'symmetric'	0.4	0.439
6	0.2	0.4	0.442
6	0.4	0.5	0.453

Pada tabel 4.11 di atas, dapat dilihat bahwa *coherence score* tertinggi dengan kombinasi parameter  $n = 6$ ,  $a = 0.4$ , dan  $b = 0.5$  mampu menghasilkan nilai koherensi sebesar 0.453. Sehingga terdapat 6 topik terbaik yang akan dijadikan aspek pada penelitian ini.

#### 4.2.13. Penentuan Nama Aspek

Setelah dilakukan uji koherensi, muncul 6 *cluster* topik dari hasil pemodelan topik menggunakan *latent Dirichlet allocation*. Setiap *clustering* topik diambil 10 kata yang memiliki probabilitas tertinggi dari masing-masing *cluster*. Setiap *topic keyword* atau kata kunci yang muncul pada sebuah *cluster* topik memiliki nilai probabilitas di depan setiap kata. Nilai probabilitas merepresentasikan kata yang sering muncul pada suatu topik. Semakin besar nilai probabilitas pada setiap kata, semakin sering pula kata tersebut muncul pada suatu topik. Hasil dari *clustering* topik kemudian diinterpretasikan menjadi sebuah nama aspek. Proses penentuan nama aspek dilakukan oleh pelabel yang merupakan alumni Teknik Elektronika. Hasil dari penentuan nama aspek pada setiap *cluster* dapat dilihat pada tabel – tabel *clustering* berikut ini.

**Tabel 4. 15 Clustering ke-1**

<b>Clustering ke-1</b>
$0.084 * " pln " + 0.076 * " kastalim " + 0.018 * " luhut " + 0.012 * " produksi "$ $+ 0.009 * " rumah " + 0.009 * " dahlan " + 0.008 * " pakai " + 0.008 * " baterai "$ $+ 0.008 * " segi " + 0.008 * " mobile "$

Kumpulan kata yang terdapat pada tabel 4.11 berisikan *topic keyword* pln, kastalim, luhut, produksi, rumah, dahlan, pakai, baterai, segi, mobile. Dalam menentukan nama aspek dilihat probabilitas tertinggi yang terdapat

pada setiap *topic keyword*. Kata pln memiliki probabilitas kemunculan tertinggi dengan nilai probabilitas sebesar 0.084. Kemudian dari kata tersebut dicari hubungan yang erat dengan mobil listrik sehingga dapat diinterpretasikan menjadi sebuah nama aspek yang mudah dipahami. Kata PLN pada *cluster* ke-1 merujuk kepada beberapa *statement* yang terdapat pada tweet sebagai berikut.

**Tabel 4. 16 Tweet Clustering ke-1**

Tweet
Pemilik Mobil Listrik, PLN Beri Diskon Penyambungan Baru untuk Home Charging <a href="https://t.co/blDeHFkQ1u">https://t.co/blDeHFkQ1u</a>
PLN Bakal Sediakan Charger Mobil Listrik di Rumah, Semudah Charge HP <a href="https://t.co/u8ytNUmpoY">https://t.co/u8ytNUmpoY</a>
@B1516EFG Â Baik Kak, seperti yg sudah mimin informasikan sebelumnya terkait promo diskon 30% saat ini PLN memiliki promo keringanan Biaya Penyambungan untuk Penyambungan Baru bagi Pemohon yang memiliki Mobil Listrik, dan jika telah memiliki Home Charger. 1/2
@OposisiCerdas Mobil listrik? Lalu ngisi listriknya kemana? Ke PLN yg mayoritas pakai batubara yaa sama aja dgn boong mending urus yg lain yg lbh urgent daripada kerja cuman 1 tahun APBD keluarin biaya mobil listrik 800 jt itu sih jelas bermasalah dan tidak efisien serta tidak efektif
PLN Luncurkan SPKLU Guna Support Ekosistem Mobil Listrik diÂ Kudus <a href="https://t.co/yxpYfpAnx7">https://t.co/yxpYfpAnx7</a>

Berdasarkan *statement* yang terdapat pada tweet asli, terdapat program dari PLN yang memberikan diskon untuk layanan *home charging* serta program peluncuran SPKLU sebagai support ekosistem mobil listrik. Sehingga

nama aspek yang cocok dan mudah untuk diinterpretasikan pada *clustering* ke-1 adalah **aspek Pengisian Baterai**.

**Tabel 4. 17 Clustering ke-2**

Clustering ke-2
$0.088 * " \text{produksi} " + 0.068 * " \text{mudah} " + 0.032 * " \text{riset} " + 0.008 * " \text{orang} " + 0.008 * " \text{isu} " + 0.007 * " \text{kertas} " + 0.007 * " \text{produk} " + 0.006 * " \text{jalan} " + 0.006 * " \text{elon} " + 0.005 * " \text{negara} "$

Kumpulan kata yang terdapat pada tabel 4.12 berisikan *topic keyword* produksi, mudah, riset, orang, isu, kertas, produk, jalan, elon, negara. Dalam menentukan nama aspek dilihat probabilitas tertinggi yang terdapat pada setiap *topic keyword*. Kata produksi memiliki probabilitas kemunculan tertinggi dengan nilai probabilitas sebesar 0.088. Kemudian dari kata tersebut dicari hubungan yang erat dengan mobil listrik sehingga dapat diinterpretasikan menjadi sebuah nama aspek yang mudah dipahami. Kata produksi pada *cluster* ke-2 merujuk kepada beberapa *statement* yang terdapat pada tweet sebagai berikut.

**Tabel 4. 18 Tweet Clustering ke-2**

Tweet
"Jika nanti baterai, motor, dinamo, dan controllernya diproduksi di Indonesia, dunia mobil listrik lama-lama seperti UMKM," ujar Moeldoko. <a href="https://t.co/lfKCM57Fuo">https://t.co/lfKCM57Fuo</a>
Indonesia Dorong Percepatan Produksi Mobil Listrik Berbasis Baterai Baca selengkapnya di: <a href="https://t.co/do45o292Xt">https://t.co/do45o292Xt</a> #mobillistrik #mobil #perkembanganmobillistrik #ice #internalcombustionengine #engine
Samsung akan Produksi Chip 5nm untuk Mobil Listrik Self Driving <a href="https://t.co/OkYzE2NjPl">https://t.co/OkYzE2NjPl</a>
@datanesia_id Seperti diketahui, pemerintah Indonesia sudah menetapkan target memproduksi 600 ribu mobil listrik dan 2,45 juta kendaraan roda dua listrik pada 2030.

Kebutuhan Mineral untuk Produksi Mobil Listrik Berlipat Ganda  
<https://t.co/66Ye596UM9>

Berdasarkan *statement* yang terdapat pada beberapa tweet di atas membahas tentang produksi komponen mobil listrik. Sehingga nama aspek yang cocok dan mudah untuk diinterpretasikan pada *clustering* ke-2 adalah **aspek produksi**.

**Tabel 4. 19 Clustering ke-3**

Clustering ke-3
$0.069 * "maju" + 0.064 * "nyata" + 0.021 * "unit" + 0.017 * "hasil" + 0.015 * "orang" + 0.015 * "esemka" + 0.015 * "kota" + 0.011 * "pakai" + 0.008 * "keren" + 0.007 * "banget"$

Kumpulan kata yang terdapat pada tabel 4.13 berisikan *topic keyword* maju, nyata, unit, hasil, orang, esemka, kota, pakai, keren, banget. Dalam menentukan nama aspek dilihat probabilitas tertinggi yang terdapat pada setiap *topic keyword*. Kata maju memiliki probabilitas kemunculan tertinggi dengan nilai probabilitas sebesar 0.069. Kemudian dari kata tersebut dicari hubungan yang erat dengan mobil listrik sehingga dapat diinterpretasikan menjadi sebuah nama aspek yang mudah dipahami. Kata maju pada *cluster* ke-3 merujuk kepada beberapa *statement* yang terdapat pada tweet sebagai berikut.

**Tabel 4. 20 Tweet Clustering ke-3**

Tweet
@Dandhy_Laksono wah kaget saya. cuman dijamannya jokowi astra maju, honda maju, wuling mampu memproduksi mobil listrik esemka mampu buat mobil listrik. hidup bapak otomotif dunia dan sekitarnya.
Indonesia miliki 25% dr cadangan Nickel dunia. Mobil listrik sgt perlu SDA ini. @jokowi sgt menekankan smelter dan hilirisasi industri di Indonesia. Waspada, asing sedang bermanuver krn ingin kuasai SDA, asing sedang memainkan boneka. Indonesia hrs mampu berdiri teguh agar maju <a href="https://t.co/EqPIPe1A5r">https://t.co/EqPIPe1A5r</a>

@TxtZodiak Keren banget sih zaman udah makin maju, serba digital, serba canggih, awal yang bagus karena gue yakin banget beberapa tahun ke depan pasti bakal lebih banyak pengguna mobil listrik #PLNEMobilizeYou

@TxtZodiak Wah keren ya skarang,aku baru tau lo kalo ada mobil listrik dan skrg juga udah ada tempat buat ngecas kendaraan listrik Semakin maju indonesiaku #PLNEMobilizeYou

Indonesia kaya akan Nikel, yang merupakan salah satu bahan dasar pembuatan baterai dari mobil listrik. Semoga mobil listrik nusantara maju bersama kemakmuran rakyatnya. jadi saatnya Indonesia STOP Ekspor Bahan Mentah. #Voksupdate <https://t.co/RMGrm9quP5>

Berdasarkan *statement* yang terdapat pada beberapa tweet di atas membahas tentang perkembangan dan kemajuan mobil listrik. Sehingga nama aspek yang cocok dan mudah untuk diinterpretasikan pada *clustering* ke-3 adalah **aspek perkembangan**.

**Tabel 4. 21 Clustering ke-4**

Clustering ke-4
$0.069 * " \text{subsidi} " + 0.064 * " \text{beli} " + 0.021 * " \text{motor} " + 0.017 * " \text{harga} " + 0.015 * " \text{rakyat} " + 0.015 * " \text{insentif} " + 0.015 * " \text{juta} " + 0.011 * " \text{perintah} " + 0.008 * " \text{masyarakat} " + 0.007 * " \text{pas} "$

Kumpulan kata yang terdapat pada tabel 4.13 berisikan *topic keyword* subsidi, beli, motor, harga, rakyat, insentif, juta, perintah, masyarakat, pas. Dalam menentukan nama aspek dilihat probabilitas tertinggi yang terdapat pada setiap *topic keyword*. Kata subsidi memiliki probabilitas kemunculan tertinggi dengan nilai probabilitas sebesar 0.069. Kemudian dari kata tersebut dicari hubungan yang erat dengan mobil listrik sehingga dapat diinterpretasikan menjadi sebuah nama aspek yang mudah dipahami. Kata maju pada *cluster* ke-4 merujuk kepada beberapa *statement* yang terdapat pada tweet sebagai berikut.

**Tabel 4. 22 Tweet Clustering ke-4**

Tweet
Sudah basi ngomong begitu, pupuk mahal karena subsidi dikurangi dialihkan ke motor dan mobil listrik u/mensubsidi orang kaya, ngomong sana sama presiden kebijakan itu salah berani gak?
@BramantioAb @adriansyahyasin Koplak emang solusinya, pada pengen ngatasi macet pejabatnya ehh malah ngasih subsidi mobil listrik
Pemerintah Beri Subsidi Mobil Listrik Rp80 Juta, Motor Listrik Rp8 Juta <a href="https://t.co/uLfaZ1ogwN">https://t.co/uLfaZ1ogwN</a>
Jangan sembarangan ngomong, ada datanya ngga? Distribusi subsidi diketati memang biar tepat sasaran, inilah tujuan program baik @pertamina dalam menghadirkan keadilan berenergi Dan yg pasti subsidi mobil listrik untuk mengurangi emisi sesuai SDG's #TepatGunaSasaran #BBMKonsisten
Gini koq malah pemerintah memberikan subsidi mobil listrik. Memang polusi berkurang dgn mobil listrik. Tapi macetnya bertambah.

Berdasarkan *statement* yang terdapat pada beberapa tweet di atas membahas tentang subsidi mobil listrik. Sehingga nama aspek yang cocok dan mudah untuk diinterpretasikan pada *clustering* ke-4 adalah **aspek subsidi**.

**Tabel 4. 23 Clustering ke-5**

Clustering ke-5
0.069 * " indonesia " + 0.064 * " baterai " + 0.021 * " negara " + 0.017 * " kendra " + 0.015 * " milik " + 0.015 * " ekosistem " + 0.015 * " bahan " + 0.011 * " gantung " + 0.008 * " cipta " + 0.007 * " strategi "

Kumpulan kata yang terdapat pada tabel 4.13 berisikan *topic keyword* indonesia, baterai, negara, kendra, milik, ekosistem, bahan, gantung, cipta,

strategi. Dalam menentukan nama aspek dilihat probabilitas tertinggi yang terdapat pada setiap *topic keyword*. Kata Indonesia memiliki probabilitas kemunculan tertinggi dengan nilai probabilitas sebesar 0.069. Kemudian dari kata tersebut dicari hubungan yang erat dengan mobil listrik sehingga dapat diinterpretasikan menjadi sebuah nama aspek yang mudah dipahami. Kata Indonesia pada *cluster* ke-4 merujuk kepada beberapa *statement* yang terdapat pada tweet sebagai berikut.

Tabel 4. 24 *Tweet Clustering* ke-5

<b>Tweet</b>
<p>Presiden Jokowi menyadari bhw mengintegrasikan komponen2 baterai &amp; mobil listrik tidaklah mdh mskpn Indonesia mmiliki hmpir smua bhn yg dibutuhkan. Dari segi geografis, tantangannya adalah bgamana mnyatukan brbagai bhn tambang yg lokasinya tersebar di berbagai wilayah Indonesia. <a href="https://t.co/nQngNO3FLL">https://t.co/nQngNO3FLL</a></p>
<p>@veroniicaa_06 salute sama PLN jadi udah aman gais sekarang kalo pergi pake mobil listrik krn udah banyak SPKLU di berbagai wilayah di Indonesia #PLNEMobilizeYou</p>
<p>@veroniicaa_06 Widih keren banget teknologinya makin ke sini makin berkembang ya, pengguna mobil listrik jadi nggak perlu khawatir lagi karena #PLNEMobilizeYou udah menyediakan 588 unit SPKLU dan SPBKLU yang tersebar di Indonesia. Mantapp!!!!</p>
<p>YEY skrng udah semakin banyak tempat pengisian daya mobil listrik di Indonesia karna per Januari 2023 pln sudah membangun 588 unit tempat charging yg tersebar di 257 titik di Indonesia dan unit tsb akan semakin bertambah setiap saatnya, semangat pln ðŸ’ðŸ» #PLNPimpinTransisiEnergi <a href="https://t.co/TfzWY8KCXI">https://t.co/TfzWY8KCXI</a></p>
<p>yang sebelumnya khawatir pakai mobil listrik susah untuk Charging sekarang ini sudah ada sebanyak 588 unit mesin SPKLU yang tersebar di 257 titik lokasi yang tersebar di Indonesia. mantap ga tuh #PLNEMobilizeYou <a href="https://t.co/PBLtbjW1aF">https://t.co/PBLtbjW1aF</a></p>

Berdasarkan *statement* yang terdapat pada beberapa tweet di atas

membahas tentang wilayah pengisian baterai mobil listrik. Sehingga nama aspek yang cocok dan mudah untuk diinterpretasikan pada *clustering* ke-5 adalah **aspek wilayah**.

**Tabel 4. 25 Clustering ke-6**

<b>Clustering ke-6</b>
$0.069 * " esemka " + 0.064 * " ev " + 0.021 * " pabrik " + 0.017 * " indonesia " + 0.015 * " era " + 0.015 * " bangun " + 0.015 * " china " + 0.011 * " istana " + 0.008 * " bisnis " + 0.007 * " tesla "$

Kumpulan kata yang terdapat pada tabel 4.13 berisikan *topic keyword* esemka, ev, pabrik, indonesia, era, bangun, china, istana, bisnis, tesla. Dalam menentukan nama aspek dilihat probabilitas tertinggi yang terdapat pada setiap *topic keyword*. Kata esemka memiliki probabilitas kemunculan tertinggi dengan nilai probabilitas sebesar 0.069. Kemudian dari kata tersebut dicari hubungan yang erat dengan mobil listrik sehingga dapat diinterpretasikan menjadi sebuah nama aspek yang mudah dipahami. Kata esemka pada *cluster* ke-6 merujuk kepada beberapa *statement* yang terdapat pada tweet sebagai berikut.

**Tabel 4. 26 Tweet Clustering ke-6**

<b>Tweet</b>
@ZAEffendy Diberi merk Esemka lalu ada kebijakan subsidi untuk pembelian mobil listrik? Kompensasi?
@kangdede78 @panca66 bijih nickel diangkut ke China. di sana dibikin mobil listrik lalu diekspor ke sini, diberi merk ESEMKA
@tempodotco Saking larisnya mobil esemka di Indonesia, pabrik nya membuat mobil listrik merk Esemka,, tanda tanda pemilu sudah dekat ini...

@PolJokesID Esemka sebenarnya pabrik produksinya dmn aja, perakitannya dmn, showroom dan bengkel resminya dmn, blm denger juga kemalingan atau kecelakaan mobil esemka di media manapun, kok tiba tiba kerjsama bikin mobil listrik, duitnya drmn ???? Meme ustad YM donk

Hozon Neta U kandidat mobil listrik Esemka, kenapa hanya ganti emblem dr mobil China, jaman skrng mah jamak, Proton Ertiga aja jg hanya ganti emblem dari Suzuki Ertiga, yg penting bisa alih tehnologi, komponen lokal tinggi, kenapa ga bikin dari 0, bikin dari 0 itu perlu 45 T ðŸ~Š <https://t.co/mDPedAlVfm>

Berdasarkan *statement* yang terdapat pada beberapa tweet di atas membahas tentang merk mobil listrik. Sehingga nama aspek yang cocok dan mudah untuk diinterpretasikan pada *clustering* ke-6 adalah **aspek merk**.

Hasil dari penentuan nama aspek adalah aspek Setiap aspek yang merepresentasikan masing-masing *cluster* akan dipisahkan dengan memfilter setiap aspek. Kemudian, setiap aspek ekosistem baterai, subsidi, serta promosi akan dilakukan klasifikasi menggunakan *naïve bayes classifier*.

#### **4.2.14. Pembobotan Kata**

Pembobotan kata yang telah dilakukan menggunakan TF IDF atau *Term Frequency - Invers Document Frequency* dengan menerapkan *library TF-IDF Vectorizer*. Proses ini akan menghitung bobot dari setiap kata atau token sehingga pada proses klasifikasi akan memudahkan perhitungan probabilitas setiap kemunculan kata pada dokumen.

### **4.3 Klasifikasi**

Proses klasifikasi merupakan proses utama yang telah dilakukan untuk mengklasifikasi teks menggunakan metode *naïve bayes classifier*. Setelah data sudah dilakukan pembobotan kata data dilakukan split data untuk membagi data menjadi data training dan data testing. Proses pembagian data menggunakan ratio perbandingan 80:20 dengan proporsi 80% untuk data training dan 20% untuk data testing.

#### **4.2.15. Training dan Testing**

Proses pembagian data *training* dan *testing* dilakukan menggunakan metode *K-Fold Cross Validation*. Proses *cross validation* menggunakan iterasi dengan  $K = 5$ . Sehingga terbentuk 5 iterasi yang akan memilih data training

dan data testing untuk diimplementasikan pada *naïve bayes classifier*. Setelah data dilakukan *cross validation*, data akan dilakukan proses *learning* dan pembuatan model klasifikasi *naïve bayes*.

#### 4.2.16. Evaluasi

Evaluasi atau pengujian performa pada model klasifikasi *naïve bayes* menggunakan *Confusion matrix* akan digunakan untuk menguji keberhasilan model yang dilakukan. Pada Tabel 4.14 menunjukkan hasil *confusion matrix* untuk sentimen seluruh aspek, yang artinya menentukan seberapa baik model yang telah dilakukan menentukan sentimen setiap aspek. Menggunakan rumus *accuracy* (5)

Tabel 4. 27 *Confusion Matrix*

Aktual	Prediksi		
	Negatif	Netral	Positif
Negatif	72	25	42
Netral	28	7	24
Positif	37	25	123

Dari data *confusion matrix* tersebut kemudian dicari nilai akurasi, presisi, *recall*, serta *f1-score* dengan hasil sebagai berikut.

Tabel 4. 28 Pengujian Performa

Label	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
<b>Negatif</b>	0.53	0.52	0.52
<b>Netral</b>	0.12	0.12	0.12
<b>Positif</b>	0.65	0.66	0.66
<b><i>Accuracy</i></b>			0.53

<i>Weighted avg.</i>	0.52	0.53	0.53
----------------------	------	------	------

Pengujian performa menggunakan naive bayes classifier menghasilkan nilai *accuracy* sebesar 53%, *precision* sebesar 52%, *recall* 53%, dan *F1-score* 53%.

#### 4.4 Analisis Sentimen Berbasis Aspek

Tabel 4. 29 Klasifikasi Tweet Setiap Aspek

Aspek	Tweet	Positif	Netral	Negatif
<b>Pengisian baterai</b>	112	80	8	24
<b>Produksi</b>	136	46	35	55
<b>Perkembangan</b>	72	30	13	29
<b>Subsidi</b>	332	101	57	174
<b>Wilayah</b>	276	150	56	70
<b>Merk</b>	284	112	78	94

##### 4.2.17. Analisis Sentimen Aspek Pengisian baterai

Pada aspek pengisian baterai terdapat sentimen positif yang menunjukkan dukungan, sentimen negatif sebagai penolakan, serta sentimen netral dianggap tidak menunjukkan dukungan maupun penolakan. Sebanyak 112 tweet dikategorikan ke dalam aspek pengisian baterai dengan sentimen positif sebanyak 80, sentimen negatif sebanyak 24, dan sentimen netral sebanyak 8 tweet.

##### Aspek Pengisian Baterai Positif

Adanya aspek pengisian baterai disebabkan karena banyaknya

masyarakat yang membahas terkait sumber energi yang digunakan sebagai pengisian mobil listrik berupa *home charging* atau SPKLU yang bersumber dari PLN. Sebanyak 80 tweet memiliki sentimen positif pada aspek pengisian baterai. Pada sisi positif, pengisian baterai yang digunakan oleh mobil listrik mampu memberikan banyak manfaat bagi penggunanya seperti kemudahan pengisian mobil listrik dengan adanya *home charging*, minimnya polusi yang dihasilkan oleh kendaraan, terhindar dari polusi suara dan kebisingan, serta biaya pengisian daya baterai yang *relative* lebih murah dibandingkan dengan penggunaan BBM untuk mobil konvensional. Penggunaan baterai sebagai energi untuk menggerakkan mobil listrik juga memiliki manfaat seperti; pada saat kondisi jalanan yang macet dan sering melakukan *stop and go*, mobil listrik akan melakukan *reverse* energi pada baterai. Sehingga menurut Fitra Eri, seorang pakar otomotif Indonesia pada *channel* youtubenya menyebutkan bahwasanya mobil listrik jauh lebih hemat energi ketika digunakan untuk kondisi yang macet dan jarak dekat seperti di dalam kota. Hal tersebut sangat berkesinambungan dengan pengisian baterai yang secara tidak langsung dihasilkan dari adanya *reverse* energi pengereman pada saat digunakan.

#### **Aspek Pengisian Baterai Negatif**

Sebanyak 24 tweet menolak kehadiran mobil listrik dari aspek pengisian baterai karena dianggap akan membutuhkan konsumsi listrik yang sangat besar. Sedangkan di Indonesia, sumber energi untuk menghasilkan listrik PLN bersumber dari bahan bakar fosil seperti batubara. Sehingga, mereka mengaitkan antara penggunaan listrik dari PLN tidak menimbulkan pencemaran lingkungan secara langsung di keluarkan oleh mobil listrik. Akan tetapi hakikatnya, sumber listrik yang dimiliki oleh PLN masih menggunakan bahan bakar fosil dan bukan sebuah solusi untuk mencegah dari pencemaran lingkungan secara global.

#### **4.2.18. Analisis Sentimen Aspek Produksi**

Aspek produksi mobil listrik memiliki sebanyak 136 tweet yang dikelompokkan pada kategori ini. Dengan sentimen positif sebanyak 46, sentimen negatif sebanyak 55, dan sentimen netral sebanyak 35 tweet.

### **Aspek Promosi Positif**

Aspek promosi positif yang disebabkan adanya event IIMS di Jakarta menyebabkan penjualan mobil listrik meningkat. Sehingga menjadi perbincangan dikalangan masyarakat untuk berupaya membeli mobil listrik yang telah dipamerkan pada event IIMS tersebut. Tidak hanya itu, pembeli dapat mencoba secara langsung atau *test drive* pada setiap mobil listrik yang disukainya pada event IIMS. Sehingga minat masyarakat meningkat.

### **Aspek Subsidi Negatif**

Aspek subsidi pada mobil listrik tidak sepenuhnya mendapatkan respon yang baik dimata masyarakat. Sebagian masyarakat memiliki sentimen negatif yang berupa penolakan terhadap adanya bantuan subsidi bagi mobil listrik. Penolakan tersebut dapat disebabkan karena kecemburuan sosial bagi pemilik mobil konvensional atau mobil bbm, supplier serta dealer mobil konvensional, serta pemerhati mobil konvensional yang akan berdampak pada penurunan *after sales* pada mobil konvensional.

#### **4.2.19. Analisis Sentimen Aspek Perkembangan**

Aspek promosi pada mobil listrik pada dasarnya tidak secara langsung disebutkan oleh masyarakat melalui cuitan pada twitter. Akan tetapi, berdasarkan analisis yang dilakukan menyebutkan bahwasanya aspek promosi berkaitan erat dengan adanya *event* IIMS atau Indonesia International Motor Show dan dampaknya. Sebanyak 72 tweet dikategorikan ke dalam aspek pengisian baterai dengan sentimen positif sebanyak 30, sentimen negatif sebanyak 29, dan sentimen netral sebanyak 13 tweet.

### **Aspek Promosi Positif**

Aspek promosi positif yang disebabkan adanya event IIMS di Jakarta menyebabkan penjualan mobil listrik meningkat. Sehingga menjadi perbincangan dikalangan masyarakat untuk berupaya membeli mobil listrik yang telah dipamerkan pada event IIMS tersebut. Tidak hanya itu, pembeli dapat mencoba secara langsung atau *test drive* pada setiap mobil listrik yang disukainya pada event IIMS. Sehingga minat masyarakat meningkat.

### **Aspek Promosi Negatif**

Sentimen negatif dari adanya promosi mobil listrik khususnya pada

event IIMS menyebabkan penjualan mobil listrik meningkat. Akan tetapi seiring dengan banyaknya produksi mobil listrik menyebabkan eksploitasi bahan baku pembuatan baterai mobil listrik serta bahan baku pembuatan lainnya. Hal tersebut mampu menyebabkan ketidakseimbangan lingkungan karena bahan baku seperti logam litium, aluminium, tembaga, serta nikel akan dieksploitasi secara masal. Sedangkan negara Indonesia merupakan salah satu pemasok dari bahan baku tersebut. Hal lain yang mampu mengakibatkan sentimen negatif pada aspek promosi mobil listrik adalah limbah yang dihasilkan oleh kerusakan komponen mobil listrik seperti baterai yang sukar untuk didaur ulang. Oleh karena itu, masyarakat melakukan penolakan pada aspek promosi mobil listrik.

#### **4.2.1. Analisis Sentimen Aspek subsidi**

Aspek promosi pada mobil listrik pada dasarnya tidak secara langsung disebutkan oleh masyarakat melalui cuitan pada twitter. Akan tetapi, berdasarkan analisis yang dilakukan menyebutkan bahwasanya aspek promosi berkaitan erat dengan adanya *event* IIMS atau Indonesia International Motor Show dan dampaknya. Sebanyak 332 tweet dikategorikan ke dalam aspek pengisian baterai dengan sentimen positif sebanyak 101, sentimen negatif sebanyak 174, dan sentimen netral sebanyak 57 tweet.

#### **Aspek Promosi Positif**

Aspek promosi positif yang disebabkan adanya event IIMS di Jakarta menyebabkan penjualan mobil listrik meningkat. Sehingga menjadi perbincangan dikalangan masyarakat untuk berupaya membeli mobil listrik yang telah dipamerkan pada event IIMS tersebut. Tidak hanya itu, pembeli dapat mencoba secara langsung atau *test drive* pada setiap mobil listrik yang disukainya pada event IIMS. Sehingga minat masyarakat meningkat.

#### **Aspek Promosi Negatif**

Sentimen negatif dari adanya promosi mobil listrik khususnya pada event IIMS menyebabkan penjualan mobil listrik meningkat. Akan tetapi seiring dengan banyaknya produksi mobil listrik menyebabkan eksploitasi bahan baku pembuatan baterai mobil listrik serta bahan baku pembuatan lainnya. Hal tersebut mampu menyebabkan ketidakseimbangan lingkungan

karena bahan baku seperti logam litium, aluminium, tembaga, serta nikel akan dieksploitasi secara masal. Sedangkan negara Indonesia merupakan salah satu pemasok dari bahan baku tersebut. Hal lain yang mampu mengakibatkan sentimen negatif pada aspek promosi mobil listrik adalah limbah yang dihasilkan oleh kerusakan komponen mobil listrik seperti baterai yang sukar untuk didaur ulang. Oleh karena itu, masyarakat melakukan penolakan pada aspek promosi mobil listrik.

#### **4.2.1. Analisis Sentimen Aspek wilayah**

Aspek promosi pada mobil listrik pada dasarnya tidak secara langsung disebutkan oleh masyarakat melalui cuitan pada twitter. Akan tetapi, berdasarkan analisis yang dilakukan menyebutkan bahwasanya aspek promosi berkaitan erat dengan adanya *event* IIMS atau Indonesia International Motor Show dan dampaknya. Sebanyak 276 tweet dikategorikan ke dalam aspek pengisian baterai dengan sentimen positif sebanyak 150, sentimen negatif sebanyak 70, dan sentimen netral sebanyak 56 tweet.

#### **Aspek Promosi Positif**

Aspek promosi positif yang disebabkan adanya event IIMS di Jakarta menyebabkan penjualan mobil listrik meningkat. Sehingga menjadi perbincangan dikalangan masyarakat untuk berupaya membeli mobil listrik yang telah dipamerkan pada event IIMS tersebut. Tidak hanya itu, pembeli dapat mencoba secara langsung atau *test drive* pada setiap mobil listrik yang disukainya pada event IIMS. Sehingga minat masyarakat meningkat.

#### **Aspek Promosi Negatif**

Sentimen negatif dari adanya promosi mobil listrik khususnya pada event IIMS menyebabkan penjualan mobil listrik meningkat. Akan tetapi seiring dengan banyaknya produksi mobil listrik menyebabkan eksploitasi bahan baku pembuatan baterai mobil listrik serta bahan baku pembuatan lainnya. Hal tersebut mampu menyebabkan ketidakseimbangan lingkungan karena bahan baku seperti logam litium, aluminium, tembaga, serta nikel akan dieksploitasi secara masal. Sedangkan negara Indonesia merupakan salah satu pemasok dari bahan baku tersebut. Hal lain yang mampu mengakibatkan sentimen negatif pada aspek promosi mobil listrik adalah limbah yang

dihasilkan oleh kerusakan komponen mobil listrik seperti baterai yang sukar untuk didaur ulang. Oleh karena itu, masyarakat melakukan penolakan pada aspek promosi mobil listrik.

#### **4.2.1. Analisis Sentimen Aspek merk**

Aspek promosi pada mobil listrik pada dasarnya tidak secara langsung disebutkan oleh masyarakat melalui cuitan pada twitter. Akan tetapi, berdasarkan analisis yang dilakukan menyebutkan bahwasanya aspek promosi berkaitan erat dengan adanya *event* IIMS atau Indonesia International Motor Show dan dampaknya. Sebanyak 284 tweet dikategorikan ke dalam aspek pengisian baterai dengan sentimen positif sebanyak 112, sentimen negatif sebanyak 94, dan sentimen netral sebanyak 78 tweet.

#### **Aspek Promosi Positif**

Aspek promosi positif yang disebabkan adanya event IIMS di Jakarta menyebabkan penjualan mobil listrik meningkat. Sehingga menjadi perbincangan dikalangan masyarakat untuk berupaya membeli mobil listrik yang telah dipamerkan pada event IIMS tersebut. Tidak hanya itu, pembeli dapat mencoba secara langsung atau *test drive* pada setiap mobil listrik yang disukainya pada event IIMS. Sehingga minat masyarakat meningkat.

#### **Aspek Promosi Negatif**

Sentimen negatif dari adanya promosi mobil listrik khususnya pada event IIMS menyebabkan penjualan mobil listrik meningkat. Akan tetapi seiring dengan banyaknya produksi mobil listrik menyebabkan eksploitasi bahan baku pembuatan baterai mobil listrik serta bahan baku pembuatan lainnya. Hal tersebut mampu menyebabkan ketidakseimbangan lingkungan karena bahan baku seperti logam litium, aluminium, tembaga, serta nikel akan dieksploitasi secara masal. Sedangkan negara Indonesia merupakan salah satu pemasok dari bahan baku tersebut. Hal lain yang mampu mengakibatkan sentimen negatif pada aspek promosi mobil listrik adalah limbah yang dihasilkan oleh kerusakan komponen mobil listrik seperti baterai yang sukar untuk didaur ulang. Oleh karena itu, masyarakat melakukan penolakan pada aspek promosi mobil listrik.

## BAB V

### KESIMPULAN

#### 5.1 Kesimpulan

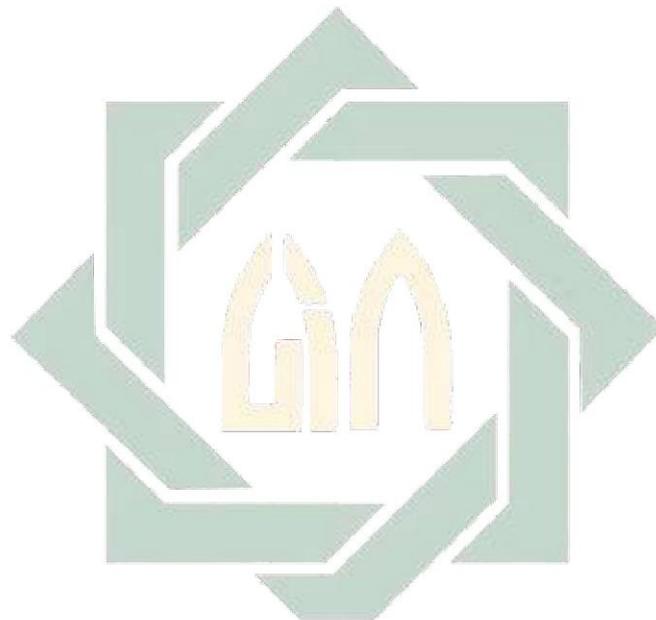
Berdasarkan hasil analisis yang telah dilakukan dapat disimpulkan sebagai berikut :

1. Pengukuran performa yang dilakukan pada pemodelan topik menggunakan metode *latent Dirichlet allocation* dalam menentukan nama aspek memiliki *coherence score* tertinggi sebesar 0.453. nilai koherensi tersebut didapatkan dengan mengkombinasikan parameter  $n = 6$ ,  $a = 0.4$ , dan  $b = 0.5$ . Pemodelan topik menghasilkan 6 *clustering* yang diinterpretasi menjadi nama aspek sebagai berikut : aspek pengisian baterai, produksi, perkembangan, subsidi, wilayah, dan aspek merk. Kemudian hasil performa klasifikasi *naïve bayes* menghasilkan nilai *accuracy* sebesar 53%, *precision* sebesar 52%, *recall* 53%, dan *F1-score* 53%. Hasil performa yang tidak begitu tinggi menunjukkan bahwasanya *naive bayes classifier* sangat bergantung dengan boosting dan treatment khusus sehingga mampu memberikan performa yang tinggi.

2. Analisis sentimen berbasis aspek terhadap kehadiran mobil listrik di Indonesia dilakukan menghasilkan 6 aspek. Pada ***aspek pengisian baterai*** terdapat 112 tweet dengan 80 tweet memiliki sentimen positif, 24 tweet memiliki sentimen negatif, dan 8 tweet memiliki sentimen netral. Pada ***aspek produksi*** terdapat total 136 tweet dengan 46 tweet bersentimen positif, 55 tweet bersentimen negatif, dan 35 tweet bersentimen netral. Pada ***aspek perkembangan*** terdapat total 72 tweet dengan 30 tweet bersentimen positif, 29 tweet bersentimen negatif, dan 13 tweet bersentimen netral. Pada ***aspek subsidi*** terdapat 332 total tweet dengan 101 tweet bersentimen positif, 174 tweet negatif, dan 57 tweet netral. Pada ***aspek wilayah*** terdapat 276 total tweet dengan 150 tweet bersentimen positif, 70 tweet negatif, dan 56 tweet netral. Pada ***aspek merk*** terdapat 284 total tweet dengan 112 tweet bersentimen positif, 94 tweet negatif, dan 78 tweet sentimen netral.

### 3. Saran

1. Menggunakan *feature extraction* selain TF-IDF untuk mengklasifikasi *naïve bayes* seperti; *bag of word*, *word embedding*, *WF-IDF*, *Doc2vec*, dan lain lain.
2. Menambahkan *stopword list* pada tahap *preprocessing* sehingga data yang diproses dapat lebih bersih.
3. Menambahkan *boosting* pada algoritma klasifikasi *naïve bayes* untuk meningkatkan akurasi, presisi, recall, dan f1-score.



UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR PUSTAKA

- Albrecht, J., Ramachandran, S., & Winkler, C. (2020). *Blueprints for Text Analytics Using Python*. O'Reilly Media, Inc.
- Alghunaim, A. (2015). *A vector space approach for aspect-based sentiment analysis*. Massachusetts Institute of Technology.
- Ali, T., Omar, B., & Soulaïmane, K. (2022). Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. *MethodsX*, 9, 101894.
- Al-Talib, G. A., & Hassan, H. S. (2013). A study on analysis of SMS classification using TF-IDF Weighting. *International Journal of Computer Networks and Communications Security*, 1(5), 189–194.
- Ariadi, M. F. (2022). *Analisis sentimen perspektif masyarakat Indonesia pada media sosial Twiter terhadap Dark Jokes menggunakan metode naïve bayes classifier dengan boosting adaboost*. UIN Sunan Ampel Surabaya.
- Arslina, A., & Liebenlito, M. (2019). Sequential topic modelling: A case study on indonesian lgbt conversation on twitter. *InPrime: Indonesian Journal of Pure and Applied Mathematics*, 1(1), 17–31.
- Asnawi, M. H., Firmansyah, I., Novian, R., & Pontoh, R. S. (2021). Perbandingan Algoritma Naïve Bayes, K-NN, dan SVM dalam Pengklasifikasian Sentimen Media Sosial. *E-Prosiding Seminar Nasional Statistika/ Departemen Statistika FMIPA Universitas Padjadjaran*, 10, 20–20.
- Astuti, S. P. (2020). *Analisis sentimen berbasis aspek pada aplikasi tokopedia menggunakan lda dan naïve bayes*. Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah.
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4. 5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640–651.
- Aziz, M., Marcellino, Y., Rizki, I. A., Ikhwanuddin, S. A., & Simatupang, J. W. (2020). Studi Analisis Perkembangan Teknologi dan Dukungan Pemerintah Indonesia Terkait Mobil Listrik. *TESLA: Jurnal Teknik Elektro*, 22(1), 45. <https://doi.org/10.24912/tesla.v22i1.7898>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
- Bustami, B. (2013). Penerapan algoritma Naive Bayes untuk mengklasifikasi data nasabah asuransi. *TECHSI-Jurnal Teknik Informatika*, 5(2).
- Byna, A., & Basit, M. (2020). Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(3), Article 3. <https://doi.org/10.32736/sisfokom.v9i3.1023>
- Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, 81(3–4), 429–450.

- Chai, K. M. A., Chieu, H. L., & Ng, H. T. (2002). Bayesian online classifiers for text classification and filtering. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 97–104.
- DataIndonesia.id. (2022). *Penjualan Mobil Listrik di Indonesia*. DataIndonesia.id. <https://dataIndonesia.id/sektor-riil/detail/penjualan-mobil-listrik-di-indonesia-capai-15437-unit-pada-2022>
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511802843>
- Febrianta, M. Y., Widiyanesti, S., & Ramadhan, S. R. (2021). Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation. *Journal of Animation and Games Studies*, 7(2), 117–144.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge university press.
- Ganesan, K. (2019). All you need to know about text preprocessing for NLP and Machine Learning. *Published Feb, 23*, 1–14.
- Giffari, M. R. A. (2022). *Analisis sentimen berbasis aspek pada ulasan aplikasi tangerang live menggunakan latent dirichlet allocation dan naïve bayes*. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
- Gunawan, B., Sastypratiwi, H., & Pratama, E. E. (2019). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 4(2), 113–118.
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed). Elsevier ; Morgan Kaufmann.
- Hartono, H., Furqan, M., Abdullah, D., Hartama, D., Roslina, R., Zarlis, M., & Situmorang, Z. (2016). *Sentiment Analysis Using Context Based Fuzzy Linguistic Hedges*.
- Herdhianto, A. (2020). *Sentiment analysis menggunakan Naïve Bayes Classifier (NBC) PADA tweet tentang zakat*. Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah.
- Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1).
- Jha, N., & Mahmoud, A. (2019). Mining non-functional requirements from app store reviews. *Empirical Software Engineering*, 24, 3659–3695.
- Karyono, G., & Utomo, F. S. (2012). Temu balik informasi pada dokumen teks berbahasa Indonesia dengan metode vector space retrieval model. *Semantik*, 2(1).
- Kementrian Agama Republik Indonesia. (2023). *Al-Qur'an dan terjemahan*. <https://quran.kemenag.go.id/surah/2>
- Latif, S. (2018). *Text Mining Untuk Klasifikasi Konten Abstrak Jurnal Bahasa Inggris Menggunakan Metode Reduksi Dimensi Dan Naive Bayes*.
- Lee, J., Kang, J.-H., Jun, S., Lim, H., Jang, D., & Park, S. (2018). Ensemble Modeling for Sustainable Technology Transfer. *Sustainability*, 10, 2278. <https://doi.org/10.3390/su10072278>

- Listari. (2019). *Inisiasi Natural Language Processing (NLP) dan Klasifikasi Jenis Wisata Kuliner untuk Program Chatbot*.
- Manning, C. D. (2008). *Introduction to information retrieval*. Syngress Publishing,.
- Mardiana, L., Kusnandar, D., & Satyahadewi, N. (2022). Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak. *Bimaster: Buletin Ilmiah Matematika, Statistika Dan Terapannya*, 11(1).
- Maybank Finance. (2023, March 7). *Elon Musk Bukan yang Pertama, Inilah Asal Mula Mobil Listrik!* Maybank Finance. <https://www.maybankfinance.co.id/artikel/sejarah-mobil-listrik>
- Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. O'Reilly Media, Inc.
- Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. (2021). Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *Sistemasi: Jurnal Sistem Informasi*, 10(1), Article 1. <https://doi.org/10.32520/stmsi.v10i1.1179>
- Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. *Proceedings of the 2nd International Conference on Knowledge Capture*, 70–77.
- Permana, M. E., Ramadhan, H., Budi, I., Santoso, A. B., & Putra, P. K. (2020). Sentiment Analysis and Topic Detection of Mobile Banking Application Review. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–6.
- Prakosa, H. A., & Nasiroh, S. (2021). Analisis Sentimen dan Pemodelan Topik Untuk Mengidentifikasi Topik Pandemi Covid-19 Pada Media Sosial Twitter menggunakan Naive Bayes Classifier dan Latent Dirichlet Allocation. *JNANALOKA*, 73–78.
- Putra, K. B., & Kusumawardani, R. P. (2017). Analisis topik informasi publik media sosial di surabaya menggunakan pemodelan latent dirichlet allocation (LDA). *Jurnal Teknik ITS*, 6(2), A446–A450.
- Putu, N. L. P. M., & Amrullah, A. Z. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 123–131.
- Rianti, R. (2020). *Kontruksi Kalimat Bahasa Inggris*. Alprin.
- Salsabila, N. A., Winatmoko, Y. A., Septiandri, A. A., & Jamal, A. (2018). Colloquial indonesian lexicon. *2018 International Conference on Asian Language Processing (IALP)*, 226–229.
- Sina, D. R., Kusrorong, N. K. S. B., & Rumlaklak, N. D. (2019). Kajian Machine Learning dengan Komparasi Klasifikasi Prediksi Dataset Tenaga Kerja Non-aktif. *Jurnal Komputer Dan Informatika*, 7(1), 37–49. <https://doi.org/10.35508/jicon.v7i1.880>
- Sugiyono. (2018). *Metode Penelitian Kuantitatif*. Alfabeta.
- Suprihatin, D. (2017). *Pokok-Pokok Bahasa Indonesia*. Absolute Media.
- Wikipedia. (2023). Mobil listrik. In *Wikipedia bahasa Indonesia, ensiklopedia bebas*.

[https://id.wikipedia.org/w/index.php?title=Mobil\\_listrik&oldid=23069080](https://id.wikipedia.org/w/index.php?title=Mobil_listrik&oldid=23069080)

- Yudha Pratomo. (2021, April 14). *Sejarah Twitter, Jejaring Sosial yang Terinspirasi dari SMS Halaman all*. KOMPAS.com. <https://tekno.kompas.com/read/2021/04/14/20420077/sejarah-twitter-jejaring-sosial-yang-terinspirasi-dari-sms>
- Yunanto, I., & Yulianto, S. (2022). Twitter Sentiment Analysis Pedulilindungi Application Using Naïve Bayes and Support Vector Machine. *Jurnal Teknik Informatika (Jutif)*, 3(4), 807–814.



UIN SUNAN AMPEL  
S U R A B A Y A