

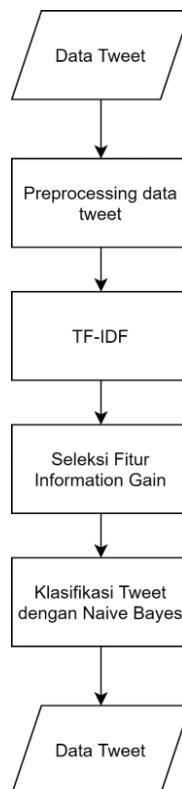
BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan

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

3.2 Deskripsi Umum

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



Gambar III- 1 Alur Program

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

3.3 Dataset

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

Tabel III- 1 Dataset

	<i>Tweet</i>	Klasifikasi
1	Pertanyaan saya, apakah mungkin produk mobil listrik digunakan untuk perusahaan sejenis ekspedisi? Berapa biaya dikeluarkan?	Netral
2	Mantap BAYAR LISTRIK PASTI SEMAKIN MAHAL .. TOKEN LISTRIK PASTI BARU 3 HARI AUTO BUNYI	Negatif
3	Semoga nanti kedepannya mobil dan motor listrik harganya akan lebih terjangkau alias murah	Positif
4	@dimsdalle @mouldie_sep Nunggu baterai murah. Bakalan terjadi kalo makin byk org pake kendaraan listrik juga	Negatif
5	Mungkin butuh waktu agak lama cari yg murah dan sporty untuk yg listrik	Netral
6	Kalau pake kendaraan listrik gaperlu pajak mahal #PakaiMolis #ZonaUang #ZonaJajan https://t.co/FoWdDr4FGH	Positif

7	Klo produsennya banyak bgt Ya perang harga Awal awal saja mahal karna blom ada persaingan yg ketat	Netral
8	Untung nya apa utk rakyat?biaya hidup Bs turun ga?	Positif
9	Subsidi kalangan tengah ke atas.. Mending gratisin biaya berobat lebih manfaat	Negatif
10	Knpa ga bikin mobil listrik produk dalam negri saja.	Netral
...		

3.4 Pengumpulan data

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

3.5 Preprocessing Data

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

3.5.1 Case Folding

Pada proses ini semua huruf kapital pada dokumen *tweet* diubah menjadi huruf kecil. Dengan tujuan untuk menghilangkan duplikasi data yang hanya berbeda pada hurufnya saja. Sebagai gambaran dari *case folding*, berikut hasil dari proses.

Tabel III- 2 Case Folding

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

3.5.2 *Cleaning*

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

Tabel III- 3 *Cleaning*

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

3.5.3 *Tokenizing*

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

Sebagai gambaran dari *Tokenizing*, berikut hasil dari proses

Tabel III- 4 *Tokenizing*

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

3.5.4 Normalisasi

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

Tabel III- 5 Normalisasi

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

3.5.5 *Stopword Removal*

Ditahap ini, merupakan tahap untuk menghilangkan kata yang tidak sesuai dengan topik. Kata yang akan dihilangkan dikumpulkan kedalam *data stopwords*. Jika dalam suatu dokumen *tweet* terdapat kata yang sesuai dengan kata yang berada dalam *stopword* maka kata tersebut akan

dihilangkan dan diubah dengan karakter *space*. Sebagai gambaran dari *Stopword Removal*, berikut hasil dari proses

Tabel III- 6 *Stopword Removal*

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

3.5.6 *Stemming*

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

Tabel III- 7 *Stemming*

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

3.6 *TF-IDF*

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

Tabel III- 8 Tabel TF

kata	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
produk	1/7	0	0	0	0	0	0	0	0	1/4
mobil	1/7	0	1/8	0	0	0	0	0	0	1/4

listrik	1/7	1/6	1/8	1/6	1/5	1/6	0	0	0	1/4
usaha	1/7	0	0	0	0	0	0	0	0	0
jenis	1/7	0	0	0	0	0	0	0	0	0
biaya	1/7	0	0	0	0	0	0	1/5	1/6	0
keluar	1/7	0	0	0	0	0	0	0	0	0
mantap	0	1/6	0	0	0	0	0	0	0	0
bayar	0	1/6	0	0	0	0	0	0	0	0
mahal	0	1/6	0	0	0	1/6	1/7	0	0	0
token	0	1/6	0	0	0	0	0	0	0	0
bunyi	0	1/6	0	0	0	0	0	0	0	0
mpga	0	0	1/8	0	0	0	0	0	0	0
motor	0	0	1/8	0	0	0	0	0	0	0
Harga	0	0	1/8	0	0	0	1/7	0	0	0
jangkau	0	0	1/8	0	0	0	0	0	0	0
alias	0	0	1/8	0	0	0	0	0	0	0
murah	0	0	1/8	1/6	1/5	0	0	0	0	0
nunggu	0	0	0	1/6	0	0	0	0	0	0
baterai	0	0	0	1/6	0	0	0	0	0	0
pakai	0	0	0	1/6	0	1/6	0	0	0	0
kendara	0	0	0	1/6	0	1/6	0	0	0	0
butuh	0	0	0	0	1/5	0	0	0	0	0
cari	0	0	0	0	1/5	0	0	0	0	0
sporty	0	0	0	0	1/5	0	0	0	0	0
gaperlu	0	0	0	0	0	1/6	0	0	0	0
pajak	0	0	0	0	0	1/6	0	0	0	0
produsen	0	0	0	0	0	0	1/7	0	0	0
perang	0	0	0	0	0	0	1/7	0	0	0
karna	0	0	0	0	0	0	1/7	0	0	0
saing	0	0	0	0	0	0	1/7	0	0	0
ketat	0	0	0	0	0	0	1/7	0	0	0
untung	0	0	0	0	0	0	0	1/5	0	0
rakyat	0	0	0	0	0	0	0	1/5	0	0
hidup	0	0	0	0	0	0	0	1/5	0	0
turun	0	0	0	0	0	0	0	1/5	0	0
subsidi	0	0	0	0	0	0	0	0	1/6	0
mending	0	0	0	0	0	0	0	0	1/6	0
gratis	0	0	0	0	0	0	0	0	1/6	0
obat	0	0	0	0	0	0	0	0	1/6	0
manfaat	0	0	0	0	0	0	0	0	1/6	0
negeri	0	0	0	0	0	0	0	0	0	1/4

Tabel III- 9 Tabel DF

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

1	3.321928095
1	3.321928095
1	3.321928095
1	3.321928095
1	3.321928095

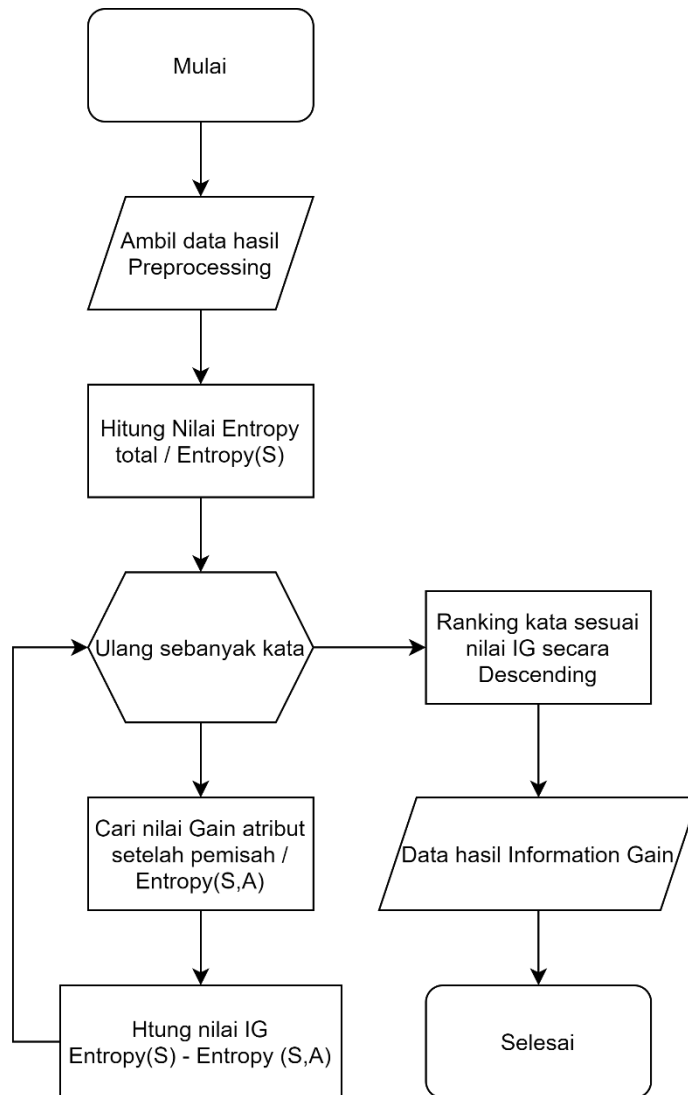
Tabel III- 10 TF-IDF

kata	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
produk	1/3	0	0	0	0	0	0	0	0	4/7
mobil	1/4	0	2/9	0	0	0	0	0	0	3/7
listrik	0	0	0	0	1/9	0	0	0	0	1/8
usaha	1/2	0	0	0	0	0	0	0	0	0
jenis	1/2	0	0	0	0	0	0	0	0	0
biaya	1/4	0	0	0	0	0	0	1/3	2/7	0
keluar	1/2	0	0	0	0	0	0	0	0	0
mantap	0	5/9	0	0	0	0	0	0	0	0
bayar	0	5/9	0	0	0	0	0	0	0	0
mahal	0	2/7	0	0	0	2/7	1/4	0	0	0
token	0	5/9	0	0	0	0	0	0	0	0
bunyi	0	5/9	0	0	0	0	0	0	0	0
mpga	0	0	2/5	0	0	0	0	0	0	0
motor	0	0	2/5	0	0	0	0	0	0	0
Harga	0	0	2/7	0	0	0	1/3	0	0	0
jangkau	0	0	2/5	0	0	0	0	0	0	0
alias	0	0	2/5	0	0	0	0	0	0	0
murah	0	0	2/9	2/7	1/3	0	0	0	0	0
nunggu	0	0	0	5/9	0	0	0	0	0	0
baterai	0	0	0	5/9	0	0	0	0	0	0
pakai	0	0	0	2/5	0	2/5	0	0	0	0
kendara	0	0	0	2/5	0	2/5	0	0	0	0
butuh	0	0	0	0	2/3	0	0	0	0	0
cari	0	0	0	0	2/3	0	0	0	0	0
sporty	0	0	0	0	2/3	0	0	0	0	0
gaperlu	0	0	0	0	0	5/9	0	0	0	0
pajak	0	0	0	0	0	5/9	0	0	0	0
produsen	0	0	0	0	0	0	1/2	0	0	0
perang	0	0	0	0	0	0	1/2	0	0	0
karna	0	0	0	0	0	0	1/2	0	0	0
saing	0	0	0	0	0	0	1/2	0	0	0
ketat	0	0	0	0	0	0	1/2	0	0	0

untung	0	0	0	0	0	0	0	2/3	0	0
rakyat	0	0	0	0	0	0	0	2/3	0	0
hidup	0	0	0	0	0	0	0	2/3	0	0
turun	0	0	0	0	0	0	0	2/3	0	0
subsidi	0	0	0	0	0	0	0	0	5/9	0
mending	0	0	0	0	0	0	0	0	5/9	0
gratis	0	0	0	0	0	0	0	0	5/9	0
obat	0	0	0	0	0	0	0	0	5/9	0
manfaat	0	0	0	0	0	0	0	0	5/9	0
negeri	0	0	0	0	0	0	0	0	0	5/6

3.7 *Information gain*

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



Gambar III- 2 *Information gain*

Pembahasan berikut merupakan proses penilaian maanual dari proses *Information gain*.

Dengan menghitung sesuai rumus pada persamaan II-1.

Tabel III- 11 *Data Training*

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

Tabel III- 12 Perhitungan *Entropy* data

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

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

Tabel III- 13 Perhitungan *Entropy* Kata

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

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

Tabel III- 14 Perhitungan *Entropy* negasi Kata

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

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

Terakhir, hitung nilai Gain dengan persamaan II-3.

Tabel III- 15 Nilai Gain

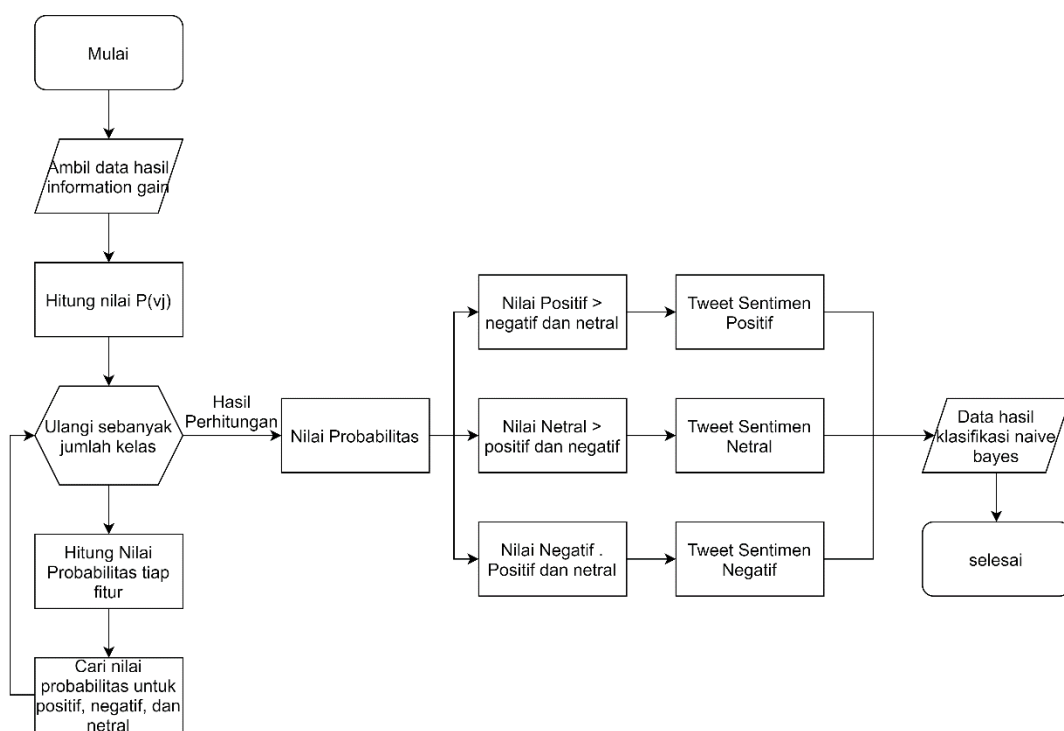
Gain
0.446439345
0.281290899
0.019973094
0.193506843
0.193506843
0.005802149
0.193506843
0.193506843
0.193506843
0.193506843
0.005802149
0.193506843
0.193506843
0.144484344
0.144484344
0.121928095
0.144484344
0.144484344
0.005802149
0.193506843
0.193506843
0.121928095
0.121928095
0.193506843
0.193506843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

0.193506843
0.193506843
0.193506843
0.193506843

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

3.8 Naïve Bayes Classifier

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



Gambar III- 3 Naïve Bayes

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

3.6.1 Training

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

Tabel III- 16 Nilai n

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

Tabel III- 17 Nilai n1

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

Tabel III- 18 Nilai P(ai|vj)

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

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

3.6.2 *Testing*

Selanjutnya merupakan perhitungan untuk data *testing*.

Tabel III- 19 Data *Testing Naïve Bayes*

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

Proses *testing* dilakukan dengan menghitung nilai kemungkinan tertinggi.

Tabel III- 20 Fitur setelah *Information gain*

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

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

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

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

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

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

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

Tabel III- 21 Hasil Klasifikasi *Naïve Bayes*

<i>Tweet</i>	Fitur	Klasifikasi
--------------	-------	-------------

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

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

3.9 Pengembangan Perangkat Lunak

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

1. Requirement

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

2. System Design

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

3. Coding & Testing

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

4. Implementation

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

5. Maintenance

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

3.10 Menejemen Proyek Penelitian

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

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

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

Membuat kelas analisis, <i>seuence Diagram</i> , dan <i>activity Diagram</i>	14 hari	17/04/24	29/04/24
Coding & Testing	28 hari	30/04/24	28/05/24
Membuat <i>Class Diagram</i>	5 hari	30/04/24	04/05/24
Membuat kode program	20 hari	06/05/24	25/05/24
Melakukan <i>testing</i>	3 hari	26/05/24	28/05/24
Implementasi	12 hari	29/05/24	10/06/24
Melakukan uji coba implementasi	7 hari	29/05/24	05/06/24
Deploy sistem pada server	2 hari	06/06/24	07/06/24
Pelatihan pengguna	1 hari	08/06/24	08/06/24
Tersedia dokumen bab pengembangan perangkat lunak	2 hari	09/06/24	10/06/24
Maintenance	-	11/06/24	-
Melaksanakan pengujian penelitian	14 hari	11/06/24	25/06/24
Membuat rancangan hasil pengujian	7 hari	11/06/24	17/06/24
Melakukan pengujian menggunakan perangkat lunak yang telah dibuat	4 hari	18/06/24	21/06/24
Tersedia dokumen bab hasil dan analisis penelitian	3 hari	22/06/24	25/06/24
Analisis hasil pengujian dan membuat kesimpulan	15 hari	26/06/24	10/07/24

3.11 Kesimpulan

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

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