PENERAPAN ALGORITMA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* UNTUK ANALISIS SENTIMEN *YOUTUBE* MENGENAI INTENSIF MOBIL LISTRIK

Caswadi [1], Nisa Dienwati [2], Gifthera Dwilestari [3], Fathurrohman [4], Edi Tohidi [5]
[1,2] Teknik Informatika, STMIK IKMI Cirebon
[3] Sistem Informasi, STMIK IKMI Cirebon
[4] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
[5] Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jalan Perjuangan 10B, Karyamulya, Kec.Kesambi, Kota Cirebon, Jawa Barat
caswadip@gmail.com

ABSTRAK

Dampak pemanasan global terhadap perubahan iklim telah menarik perhatian dari berbagai pihak. 27% dari polusi udara disebabkan oleh transportasi. Di berbagai negara, pemerintah telah menerapkan langkah-langkah untuk mengurangi polusi udara dengan mendorong masyarakat untuk beralih menggunakan kendaraan listrik. Namun keberhasilan pemerintah untuk mengkampanyekan teknologi mobil listrik tergantung pada sentimen dan pemahaman masyarakat. Tujuan penelitian ini untuk menganalisa sentimen publik. Dataset digunakan dalam penelitian ini ialah 1517 data komentar di laman *youtube* mengenai intensif mobil listrik. Berdasarkan analisis, sebagian besar komentar *youtube* memiliki sentimen negatif (57,4%), sementara jumlah komentar yang positif (33,3%) dan netral (9,3%). Penelitian ini menerapkan metode *Naïve Bayes* dan *K-Neareast Neighbor*. Hasil penelitian menunjukan bahwa metode *K-Nearest Neighbor* memberikan hasil klasifikasi yang lebih baik dengan *Accuracy* sebesar 93,23%, *precision* 93,91% dan *recall* 91,56%. Sedangkan *Naïve Bayes* memperoleh hasil *Accuracy* 86,95%, *precision* 80,51%, dan *recall* 91,23%.

Kata kunci: *Mobil Listrik, Klasifikasi, Analisis Sentimen, Youtube.*

1. PENDAHULUAN

Dalam upaya untuk mengurangi emisi gas di negara-negara anggota G20, dengan mengacu pada target *Nasional Determined Contribution (NDC)*. *Climate Transparency*, pada laporan tahunan G20 mengenai transparansi iklim 2022 [1] Berdasarkan data tersebut, dapat disimpulkan bahwa transportasi merupakan penyumbang sebanyak 27 persen terhadap polusi udara dunia. Kontributor utama yang signifikan terhadap pencemaran udara adalah transportasi yang menggunakan bahan bakar fosil, termasuk penggunaan alat berat. Sarana yang dapat diadopsi guna mengurangi dan meredam emisi gas melibatkan implementasi kendaraan listrik *(electric vehicle)*, suatu pendekatan yang telah diupayakan secara intensif oleh beberapa negara belakangan ini. Pemerintah Indonesia mengimplementasikan regulasi terkait penggunaan kendaraan listrik dalam konteks transportasi jalan, yang mencakup kendaraan berbaterai. Langkah ini dianggap sebagai bukti konkret dari komitmen pemerintah untuk meningkatkan upaya dalam menekan emisi gas dan mengatasi dampak perubahan iklim.

Pertama kali, mobil listrik diproduksi secara massal dan komersial oleh *General Motors* dirancang untuk digunakan pada durasi waktu yang terbatas dan menempuh jarak yang relatif pendek bila dibandingkan dengan kendaraan konvensional [2]. Tidak hanya terkait dengan aspek teknis, tetapi juga perlu memperhatikan persepsi sosial dalam rangka meningkatkan distribusi komersial kendaraan listrik. Faktor yang mempengaruhi tingkat adopsi mobil listrik di Indonesia mencakup persepsi masyarakat terhadap kendaraan listrik. Adanya korelasi positif, antara perasaan positif saat mengendarai kendaraan listrik dan niat konsumen. Sementara di sisi lain, beberapa konsumen melaporkan pengalaman perasaan negatif seperti “merasa lebih boros” setelah menggunakan kendaraan listrik [3].

Berdasarkan penelitian yang pernah dilakukan oleh Riyadi dkk dengan topik analisa pengukuran sentimen *twitter* mengenai teknologi kendaraan listrik diketahui metode *Naïve Bayes*, *K-Nearest Neighbor* dan *Decision Tree* menghasilkan tingkat *Accuracy* 94% dengan respon positif 53%, negatif sebesar 38% dan nilai respon netral 9%, yang artinya kendaraan mobil listrik di Indonesia tetap dapat diterima oleh masyarakat[3]. Penelitian terkait lainnya yaitu Mengimplementasikan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbor* untuk menganalisis sentimen aplikasi Halodoc [4]. Data yang diambil menggunakan *scraping* pada *Google Play Store*, data yang berhasil terkumpul adalah 500 data dengan ulasan positif dan 500 data dengan ulasan negatif. Hasil penelitian menunjukan algoritma *K-Nearest Neighbor* adalah algoritma terbaik dalam penelitian tersebut dengan tingkat *accuracy* 95.0%.

Penelitian ini berfokus pada analisis sentimen terhadap data komentar *youtube* menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor*. Metode penelitian ini akan mencakup pengumpulan data komentar *youtube* yang relevan dengan topik, *preprocessing* data untuk membersihkan dan mempersiapkan teks, ekstraksi fitur dari teks, dan penggunaan algoritme *Naïve Bayes* dan *K-Nearest*

Neighbor untuk melakukan klasifikasi sentimen. Algoritma *Naïve Bayes* digunakan untuk mengklasifikasikan teks menjadi kategori sentimen berdasarkan probabilitas, sementara *K-Nearest Neighbor* digunakan untuk mengelompokkan *komentar youtube* ke dalam klaster berdasarkan kesamaan mereka.

2. TINJAUAN PUSTAKA

2.1. *Text Mining*

Text mining adalah suatu proses ekstraksi pola-pola yang signifikan dan menarik dari data tekstual sumber dengan tujuan untuk mengeksplorasi pengetahuan. *Text mining* merupakan bidang multi-disiplin yang mendasarkan diri pada pengambilan informasi, penambangan data, pembelajaran mesin, statistik, dan linguistik komputasi. *Text mining* juga dapat dipahami sebagai upaya pencarian data teks yang melibatkan keinginan untuk menemukan pengetahuan yang sebelumnya belum teridentifikasi. Proses ini merupakan elemen esensial dari pencarian data teks yang bertujuan untuk memperoleh pemahaman, mengidentifikasi pola praktis dan merumuskan pengetahuan dari kumpulan data teks atau corpus masif (kumpulan teks yang menangkap penggunaan bahasa dalam bentuk tertulis atau lisan secara utuh dan padat) yang bersifat tidak terstruktur [6].

2.2. *Naïve Bayes*

Naive Bayes merupakan salah satu metode yang dapat diterapkan untuk menganalisis sentimen, dan secara teoritis, metode ini menunjukkan konsistensi data dan ketepatan perhitungan klasifikasi. Penggunaan *Naive Bayes* umumnya terfokus pada teknik klasifikasi, terutama dalam konteks laman seperti *twitter*. Ciri khas dari klasifikasi *Naïve Bayes* terletak pada kemampuannya untuk menghasilkan hipotesis yang kuat terkait suatu kondisi atau kejadian. Perhitungan probabilitas kelompok dalam metode *Naïve Bayes* mengacu pada pendekatan algoritma *Bayesian* yang menggunakan persamaan sebagai dasarnya [7].

2.3. *K-Nearest Neighbor*

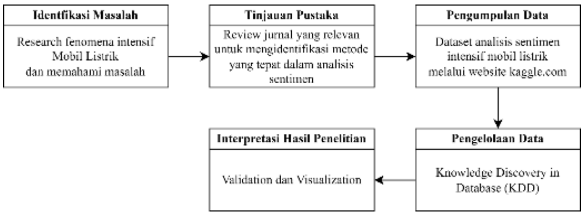
K-Nearest Neighbor adalah algoritma pembelajaran mesin terawasi yang sebagian besar digunakan untuk tujuan klasifikasi. Umumnya algoritma *K-NN* mampu mengklasifikasikan dataset menggunakan model pelatihan yang mirip dengan kueri pengujian dengan mempertimbangkan *K-Nerest* data pelatihan terdekat *(Neighbor)* yang paling dekat dengan kueri yang diuji. algoritma *K-NN* merupakan salah satu bentuk yang paling sederhana dan banyak digunakan dalam tugas-tugas klasifikasi karena memiliki desain yang sangat adaptif dan mudah dipahami [8].

2.4. *Knowledge Discovery in Database* (KDD)

KDD merupakan kegiatan interdisiplin yang merujuk pada metodologi untuk memperoleh pengetahuan dari data [9]. Penambangan data dan KDD sering digunakan secara bergantian untuk menggambarkan proses penggalian informasi tersembunyi dalam database yang luas [10].

3. METODE PENELITIAN

Metode penelitian ini menggunakan metode penelitian eksperimen kuantitatif. Metode eksperimen adalah metode penelitian ilmiah yang dilakukan dengan merancang situasi atau kondisi tertentu untuk mengukur pengaruh variabel tertentu terhadap variabel lain. Dalam metode penelitian eksperimen dapat menguji sebab akibat secara langsung dan dapat mengontrol variabel yang mempengaruhi hasil penelitian. Dapat dilakukan dengan menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor*.



Gambar 1. Penelitian

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. *Data Selection*

Dataset pada penelitian ini didapat dari *website Kaggle.com* dengan operator search *twitter* dan kata kunci Ananlisis Sentimen Mobil Listrik. Pengambilan data ini dilakukan pada 1 Desember 2023. Data yang didapat sebanyak 1517 data *tweets*. Bentuk dataset dapat dilihat pada gambar 1 berikut.

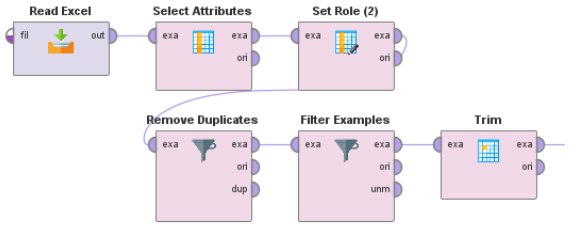
Row No.	Id Komentar	Nama Akun	Tanggal	Text	Sentimen
1	Ugzbll5eyrly3-gdUUJ4AaABAg	Sqn Ldr	2023-08-06 12:54:49+00:00	saran sih bikin harga ionic sama kayak brio insya alloh laris manis	positif
2	UgzEDUiV3OTrV943p8p4AaABAg	lushen ace	2023-08-04 12:16:23+00:00	problem subsidi kualitas diturunin harga dinaikin usaha gitu cari cuan subsidi sebab infl...	negatif
3	UgwqJqu6JMF4EH2CsVV4AaABAg	Fatih Al-Ayyubi	2023-08-04 10:17:57+00:00	baik kualitas kembang dulu baik kualitas motor motor pabrikan jepang	positif
4	UgyYicCMR1rKwuOj2Y14AaABAg	yp office	2023-08-04 08:29:54+00:00	model jelek kwalitas buruk harga mahal croot	negatif
5	UgxKAcLuAwZOQK6es-x4AaABAg	Lembur Kuring	2023-08-04 07:55:37+00:00	syarat ngaco woy anak muda blom punya rumah blom jd umkm bukan serta kur dapet ng...	negatif
6	Ugx-zVY4ktd7JNUB6xV4AaABAg	Syarif Airlangga	2023-08-04 06:58:17+00:00	harga motor mahal masa harga mirip motor beat kualitas bagai langit bumi	positif
7	Ugzu5Wkjya394dPHoq14AaABAg	BajuMax	2023-08-04 06:31:56+00:00	mol keren yah berita plus padahal mol lokal merk batrenya juta garansi tahun anggap bat...	negatif
8	Ugyy3luBOOHQspWjBIR4AaABAg	Putut Parwoto	2023-08-04 01:04:18+00:00	proses kenal produk baru butuh waktu ganti kendara bbm jadi kendara butuh waktu tehn...	negatif
9	Ugw-fxD-u_9LFxhiRTp4AaABAg	Heru Prasetyo	2023-08-03 23:56:55+00:00	subsidi tepat sasar	netral
10	Ugx2k37YO1fch9eEM4d4AaABAg	jonan kick ass hole	2023-08-03 11:25:57+00:00	adil rata terima subsidi jangan jangan pajak pph meh subsidi perintah khianat pancasila...	negatif
11	UgzBFjd8vzYUd7E2bGN4AaABAg	Rendy Ramadhan	2023-08-03 07:20:29+00:00	tepat sasar donk syarat belum tepat subsidi kurang besar jd warga kurang mampu pikir ...	negatif
12	Ugw_wdGd4wffM4N--cs4AaABAg	kohar arief	2023-08-03 05:08:13+00:00	mungkin subsidi baik arah seragam sekolah bila lama guna seragam sekolah hapus ne...	negatif
13	UgzZY-ZoTvfVoOv8Dt4AaABAg	gema	2023-08-03 04:23:25+00:00	kampung sekarang banyak banget bocil sama cewek pake motor	positif
14	Ugxhg359PbSx8yP5yM94AaABAg	F	2023-08-01 12:39:56+00:00	banyak kenderaan buat tiongkok rekam bakar letup amerika rana masalah baterai kende...	negatif
15	UgwzltHDvO9naYvia54AaABAg	Khoirudin 22	2023-07-29 11:08:07+00:00	harga terlalu mahal	positif

ExampleSet (1,517 examples, 0 special attributes, 5 regular attributes)

Gambar 2. Dataset

4.1.2. *Preprocessing*

Untuk meminimalisir adanya *missing value* atau nilai kosong dan data yang tidak konsisten. Pada tahap ini data yang sudah siap diolah akan dihilangkan tanda baca, *URL* dan *noisy*. Berikut tahapan-tahapan *preprocessing* yaitu menggunakan operator sebagai berikut :

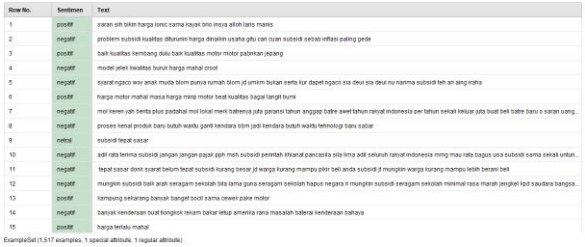


Gambar 3. Proses *Preprocessing*

Adapun penjelasan dari masing-masing operator yaitu:

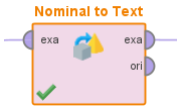
1. *Read Excel*
   Operator *Read Excel* digunakan untuk membuka dan membaca dataset dalam bentuk *file excel*
2. *Selection Atribut*
   Operator selection berfungsi untuk menyeleksi atribut yang akan digunakan dan menghilangkan atribut lain yang tidak digunakan
3. *Set Role*
   Operator *Set Role* digunakan untuk memilih kolom atribut yang akan digunakan sebagai data label
4. *Remove Duplicates*
   Operator ini digunakan untuk menghilangkan atau menghapus data yang dinilai memiliki nilai yang sama
5. *Filter Example*
   Operator ini digunakan untuk menghilangkan atau menghapus data yang mengandung *missing value* atau data yang tidak memiliki nilai.
6. *Trim*
   Operator ini digunakan untuk menghilangkan spasi awal dan akhir dari nilai atribut nominal yang dipilih dan menghilangkan atribut lain yang tidak digunakan.

Berikut hasil dari preprocessing data pada gambar



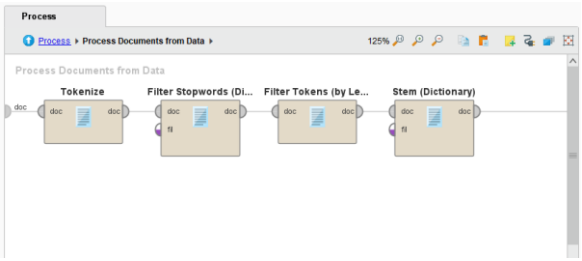
Gambar 3. Hasil *Preprocessing*

4.1.3. *Data Transformation*



Gambar 4. Operator *Nominal to Text*

Proses selanjutnya adalah transformasi data. Operator *Nominal to text* untuk menyesuaikan jenis data yang akan digunakan yaitu atribut text.

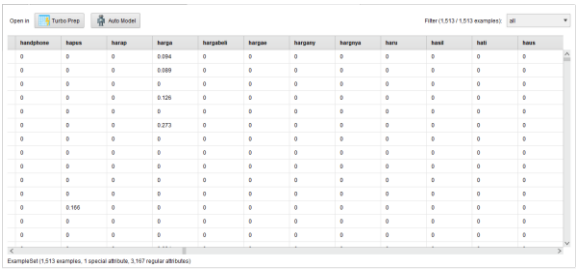


Gambar 5. Operator *Process Documents from Data*

Operator *Process Documents from Data* membutuhkan atribut text dapat dilihat pada gambar 5 untuk memproses semua data menjadi satu dokumen. Terdapat beberapa operator didalamnya yaitu :

1. *Tokenize*
   Operator ini digunakan untuk proses pemecahan teks menjadi bagian kecil yang disebut token sehingga dapat dianalisa.
2. *Stopword*
   Penggunaan stopword merupakan suatu langkah dalam mengeliminasi kata-kata yang tidak relevan dengan topik dokumen, bertujuan untuk meningkatkan *accuracy* dalam proses pengklasifikasian. *Stopword* ini berfungsi untuk menghapus kata-kata yang tidak memiliki makna signifikan ketika digunakan secara mandiri.
3. *Filter Token by Length*
   Adalah proses menghapus kata tidak mengandung makna. Menggunakan parameter paling sedikit 4 huruf dan paling banyak 25 *(default)*
4. *Stem (Dictonary)*
   Operator terakhir pada proses documents from data ini digunakan untuk memproses pemetaan kata dari bentuk kata menjadi bentuk kata dasar.

Proses selanjutnya adalah pembobotan kata. Pada proses ini menggunakan metode *TF-IDF* dimana setiap frekuensi kata akan dihitung kemunculannya pada data komentar *youtube*. Metode *Term Frequency-Inverse Document Frequency (TF-IDF)* adalah suatu teknik transformasi yang mengukur signifikansi kata dalam dokumen dengan memberikan nilai bobot pada masing-masing kata. Contoh kata "*handphone*" dengan nilai 0 artinya kata tersebut tidak muncul pada dokumen pertama sedangkan kata "harga" bernilai 0.094 berarti kata tersebut muncul pada dokumen pertama dan dokumen lainnya.



Gambar 6. Hasil Proses Pembobotan Kata *TF-IDF*

Berikut dokumen yang akan digunakan dalam perhitungan dengan metode *TF-IDF*.

Tabel 1. Dokumen yang digunakan dalam perhitungan TF-IDF

Dokumen	Teks
Dokumen 1 (D1)	Saran sih bikin ionic sama kaya brio insya alloh laris manis
Dokumen 2 (D1)	Problem kualitas diturunin harga dinaikin usaha gitu cari cuan subsidi sebab inflasi paling gede

Perhitungan TF dilakukan pada 2 sampel dokumen. Kata yang muncul diberi nilai 1 dan jika kata tidak muncul diberi nilai 0. Berikut hasil perhitungan *TF (Term Frequency)* dapat dilihat pada tabel berikut.

Tabel 2. Hasil Perhitungan TF

Tokenize	TF			
	D1	D2	D3	D4
saran	1	0	0	0
ionic	1	0	0	0
kaya	1	0	0	0
brio	1	0	0	0
insya	1	0	0	0
alloh	1	0	0	0
laris	1	0	0	0
manis	1	0	0	0
problem	0	1	0	0
subsidi	0	1	0	0
kualitas	0	1	2	1
turun	0	1	0	0
harga	0	1	0	1
naik	0	1	0	0
usaha	0	1	0	0
cuan	0	1	0	0
inflasi	0	1	0	0
gede	0	1	0	0

Kemudian dari hasil perhitungan TF (Term Frequency). Selanjutnya menghitung jumlah dokumen IDF, IDF iaah hasil inverse dari DF

Tabel 3. Hasil Perhitungan IDF

DF	D/Df	IDF (LogD/Df)
1	4	0,6
1	4	0,6
1	4	0,6
1	4	0,6
1	4	0,6
3	1,33	0,6
1	4	0,6
2	2	0,6
1	4	0,6
1	4	0,6
1	4	0,6
1	4	0,6
1	4	0,6

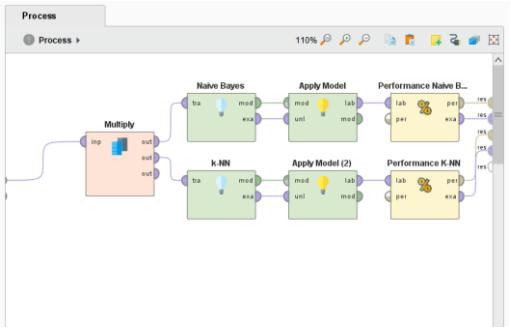
Berikut dapat dihitung hasil TF-IDF dengan mengalikan hasil TF dan IDF. Hasilnya dapat dilihat pada tabel dibawah ini.

Tabel 4. Hasil Perhitungan TF-IDF

W			
D1	D2	D3	D4
0,6	0	0	0
0,6	0	0	0
0,6	0	0	0
0,6	0	0	0
0,6	0	0	0
0,6	0	0	0
0,6	0	0	0
0,6	0	0	0
0	0,6	0	0
0	0,6	0	0
0	0,12	0,24	0,12
0	0	0	0,6
0	0,3	0	0,3
0	0,6	0	0
0	0,6	0	0
0	0,6	0	0
0	0,6	0	0
0	0,6	0	0

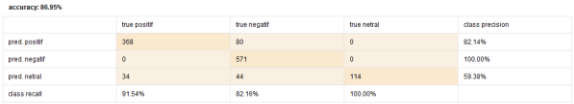
4.1.4. *Data Mining*

Data mining adalah proses mengekstraksi pola atau pengetahuan yang berharga dari sekumpulan data besar. Penelitian ini membagi data training sebanyak 0.8 dan data testing sebanyak 0.2 pada parameter operator *Split Data*. Kemudian operator *Multiply* digunakan untuk membuat salinan objek.



Gambar 7. Proses *Data Mining*

1. *Naïve Bayes*

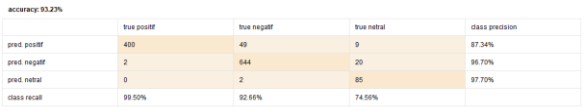

accuracy: 86.95%

	true positif	true negatif	true netral	class precision
pred. positif	368	80	0	82.14%
pred. negatif	0	571	0	100.00%
pred. netral	34	44	114	59.38%
class recall	91.54%	82.16%	100.00%	

Gambar 8. Hasil Naïve Bayes

Berdasarkan hasil Performance *Naïve Bayes* mengenai analisis sentimen pada komentar tentang intensif mobil listrik di laman youtube diperoleh tingkat *accuracy* sebesar 86,95%, dengan nilai *precission* untuk kategori positif sebesar 82,14%, untuk kategori negatif 100,00% dan untuk kategori netral sebesar 59,38%. Lalu nilai *recall* untuk kategori positif sebesar 91,54%, untuk kategori negatif 82,16% dan kategori netral sebesar 100,00%.

2. *K-Nearest Neighbor*


accuracy: 93.23%

	true positif	true negatif	true netral	class precision
pred. positif	400	49	9	87.34%
pred. negatif	2	644	20	96.70%
pred. netral	0	2	85	97.70%
class recall	99.50%	92.66%	74.56%	

Gambar 9. Hasil *K-Nearest Neighbor*

Berdasarkan hasil *performance* mengenai analisis sentimen pada komentar tentang intensif mobil listrik di laman *youtube* menggunakan metode *K-Nearest Neighbor* diperoleh hasil tingkat *accuracy* sebesar 92,23%, dengan nilai *precission* untuk kategori positif sebesar 87,34%, untuk kategori negatif 96,70% dan untuk kategori netral sebesar 97,70%. Lalu nilai *recall* untuk kategori positif sebesar 99,50%, untuk kategori negatif 92,66% dan kategori netral sebesar 74,56%.

4.1.5. *Patern Evaluation*

Setelah proses klasifikasi dengan *Naïve Bayes* dan *K-Nearest Neighbor*. Tahap berikutnya ialah melakukan proses evaluasi yang bertujuan untuk mengetahui nilai *accuracy*, *precission*, dan *recall* dari hasil yang didapatkan. *Confusion Matrix* digunakan dalam evaluasi ini karena berperan sebagai sarana evaluasi untuk mengestimasi objek yang teridentifikasi dengan benar (*True*) dan salah (*False*) pada model klasifikasi. Berikut adalah *Confussion Matrix* dapat dilihat pada tabel 6 di bawah ini:

Tabel 5. *Confusion Matrix* Tiga Kelas

		Kelas Aktual		
		Postif	**Negatif**	**Netral**
Kelas Prediksi	**Positif**	True Positive (TP)	False Negative (FN)	False Neutral (FN)
	Negatif	False Positive (FP)	True Negative (TN)	False Neutral (FN)
	Netral	False Neutral (FN)	False Neutral (FN)	True Netral (TN)

Rumus dan perhitungan untuk mencari nilai *accuracy Naïve Bayes* :

$$Accuracy = \frac{TPositif + TNegatif + TNetral}{TPositif + (FNegatif + FNetral)} \times 100\% \quad (1)$$
$$= \frac{368 + 571 + 114}{1.211} \times 100\%$$
$$= \frac{1.053}{1.211} \times 100\% = \mathbf{86,95\%}$$

Rumus dan perhitungan untuk mencari nilai *precission* positif *Naïve Bayes*

$$Precission\ Positif = \frac{TPositif}{TPositif + (FNegatif + FNetral)} \times 100\% \quad (2)$$
$$Precission\ Positif = \frac{368}{368 + 80} \times 100\%$$
$$= \frac{368}{448} \times 100\% = \mathbf{82,14\%}$$

Rumus dan perhitungan untuk mencari nilai *precission* negatif *Naïve Bayes*

$$Precission\ Negatif = \frac{TNegatif}{TNegatif + (FPositif + FNetral)} \times 100\% \quad (3)$$
$$Precission\ Negatif = \frac{571}{571 + 0} \times 100\%$$
$$= \frac{571}{571} \times 100\% = \mathbf{100,00\%}$$

Rumus dan perhitungan untuk mencari nilai *precission* netral *Naïve Bayes*

$$Precission\ Netral = \frac{TNetral}{TNetral + (FNetral + FNetral)} \times 100\% \quad (4)$$
$$Precission\ Netral = \frac{114}{114 + 78} \times 100\%$$
$$= \frac{114}{192} \times 100\% = \mathbf{59,38\%}$$

Rumus dan perhitungan untuk mencari nilai *recall* positif *Naïve Bayes*

$$Recall\ Positif = \frac{TPositif}{TPositif + (FPositif + FNetral)} \times 100\% \quad (5)$$
$$Recall\ Positif = \frac{368}{368 + 34} \times 100\%$$
$$= \frac{368}{402} \times 100\% = \mathbf{91,54\%}$$

Rumus dan perhitungan untuk mencari nilai *recall* negatif *Naïve Bayes*

$$Recall\ Negatif = \frac{TNegatif}{TNegatif + (TNegatif + FNetral)} \times 100\% \quad (6)$$
$$Recall\ Negatif = \frac{571}{571 + 124} \times 100\%$$
$$= \frac{571}{695} \times 100\% = \mathbf{82,16\%}$$

Rumus dan perhitungan untuk mencari nilai *recall* netral *Naïve Bayes*

$$Recall\ Netral = \frac{TNetral}{TNetral + (TNetral + FNetral)} \times 100\% \quad (7)$$
$$Recall\ Netral = \frac{114}{114 + 0} \times 100\%$$
$$= \frac{114}{114} \times 100\% = \mathbf{100,00\%}$$

Rumus dan perhitungan untuk mencari nilai *accuracy K-Nearest Neighbor*:

$$Accuracy = \frac{TPositif + TNegatif + TNetral}{Jumlah\ Data\ Keseluruhan} \times 100\% \quad (8)$$
$$Accuracy = \frac{400 + 644 + 85}{1.211} \times 100\%$$
$$= \frac{1.129}{1.211} \times 100\% = \mathbf{93,23\%}$$

Rumus dan perhitungan untuk mencari nilai *precission* positif *K-Nearest Neighbor*

$$Precission\ Positif = \frac{TPositif}{TPositif + (FNegatif + FNetral)} \times 100\% \quad (9)$$
$$Precission\ Positif = \frac{400}{400 + 58} \times 100\%$$
$$= \frac{400}{458} \times 100\% = \mathbf{87,34\%}$$

Rumus dan perhitungan untuk mencari nilai *precission* negatif *K-Nearest Neighbor*

$$Precission\ Negatif = \frac{TNegatif}{TNegatif + (FPositif + FNetral)} \times 100\% \quad (10)$$
$$Precission\ Negatif = \frac{644}{644 + 22} \times 100\%$$
$$= \frac{644}{666} \times 100\% = \mathbf{96,70\%}$$

Rumus dan perhitungan untuk mencari nilai *precission* netral *K-Nearest Neighbor*

$$Precission\ Netral = \frac{TNetral}{TNetral + (FNetral + FNetral)} \times 100\% \quad (11)$$
$$Precission\ Netral = \frac{85}{85 + 2} \times 100\%$$
$$= \frac{85}{87} \times 100\% = \mathbf{97,70\%}$$

Rumus dan perhitungan untuk mencari nilai *recall* positif *K-Nearest Neighbor*

$$Recall\ Positif = \frac{TPositif}{TPositif + (TPositif + FNetral)} \times 100\% \quad (12)$$

$$Recall\ Positif = \frac{400}{400 + 2} \times 100\%$$

$$= \frac{400}{402} \times 100\% = \mathbf{99,50\%}$$

Rumus dan perhitungan untuk mencari nilai *recall* negatif *K-Nearest Neighbor*

$$Recall\ Negatif = \frac{TNegatif}{TNegatif + (TNegatif + FNetral)} \times 100\% \quad (13)$$

$$Recall\ Negatif = \frac{644}{644 + 51} \times 100\%$$

$$= \frac{644}{695} \times 100\% = \mathbf{92,66\%}$$

Rumus dan perhitungan untuk mencari nilai *recall* netral *K-Nearest Neighbor*

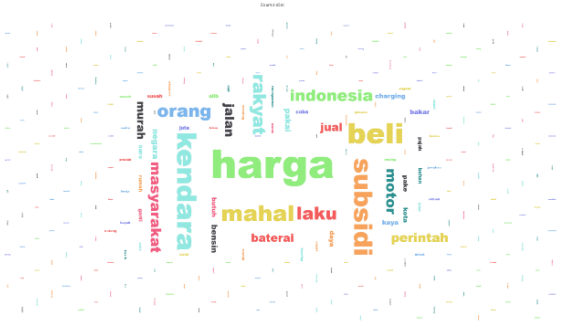
$$Recall\ Netral = \frac{TNetral}{TNetral + (TNetral + FNetral)} \times 100\% \quad (14)$$

$$Recall\ Netral = \frac{85}{85 + 29} \times 100\%$$

$$= \frac{85}{114} \times 100\% = \mathbf{74,56\%}$$

4.1.6. *Knowledge Presentation*

Hasil preproscessing data dikelompokkan dengan tujuan mengidentifikasi frekuensi kemunculan kata-kata yang tinggi. Infromasi frekuensi kata-kata ini akan direpresentasikan secara visual melalui pembuatan *wordcloud*.



Gambar 10. Hasil Visualisasi *Wordcloud*

Dalam gambar 10 frekuensi kemunculan kata yang paling tinggi adalah kata "harga" dengan total kemunculan sebesar 446 kata, disusul oleh kata "beli" dengan total kemunculan sebesar 325 kata. Kata yang yang muncul terlihat lebih kecil maka kata tersebut berarti nilai frekuensi kemunculan katanya lebih rendah.

4.2. Pembahasan Penelitian

Penerapan algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN). Penelitian ini menerapkan dua algoritma *Naïve Bay*es dan *K-Nearest Neighbor*. Peneliti mengamati, berdasarkan jurnal penelitian-penelitian terdahulu yang pernah dibaca oleh peneliti metode *Naïve Bayes* sangat populer digunakan karena performa akurasinya yang sangat baik dan tepat dalam analisis sentimen. Algoritma *Naïve Bayes* dipilih karena keunggulannya dalam proses klasifikasi, dimana hanya membutuhkan sedikit data dari data pelatihan untuk memilih parameter yang diperlukan, juga prosesnya cepat dan efisien [9].

K-Nearest Neighbor juga dipilih karena kelebihannya dalam menangani data pelatihan dengan jumlah noise yang tinggi, cepat, mudah dipahami dan kemampuan untuk menangani data dalam jumlah besar [10]. Algoritma ini terbukti efektif, seperti yang dibuktikan dalam penelitian "Implementasi Algoritma *Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors* Untuk Analisa Sentimen Aplikasi Halodoc" (Indrayuni dkk, 2021)

Hasil penerapan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* didapat hasil klasifikasi adalah sebagai berikut :

Tabel 6. Hasil Klasifikasi Ulasan Sentimen *Youtube*

Algoritma	Prediksi Sentimen		
	Positif	Negatif	Netral
Naïve Bayes	368	571	114
K-Nearest Bayes	400	644	85

Setelah proses klasifikasi dengan *Naïve Bayes* dan *K-Nearest Neighbor*. Menghasilkan nilai confusion matrix berikut ini :

1. *Naïve Bayes*

Tabel 7. *Confusion Matrix Naïve Bayes*

		Kelas Aktual		
		Positif	Negatif	Netral
Kelas Prediksi	Positif	368	80	0
	Negatif	0	571	0
	Netral	34	44	144

Setelah dilakukan pengujian perhitungan manual menggunakan rumus confusion matrix tabel diatas berhasil mengklasifikasikan secara benar dengan nilai (True Positif) sebesar 368, data (True Negatif) 571, dan (True Netral) sebesar 144. Adapun hasil dari perhitungan nilai accuracy, precision dan recall terdapat pada tabel dibawah ini.

Tabel 8. Hasil Perhitungan Manual *Confusion Matrix Naïve Bayes*

	Precision	Recall	Accuracy
Positif	82,14%	91,54%	86,95%
Negatif	100,00%	82,16%	
Netral	59,38%	100,00%	

2. ***K-Nearest Neighbor***

Tabel 9. *Confusion Matrix K-Nearest Neighbor*

		Kelas Aktual		
		Postif	Negatif	Netral
Kelas Prediksi	Positif	400	49	9
	Negatif	2	644	20
	Netral	0	2	85

Setelah dilakukan pengujian perhitungan manual menggunakan rumus *confusion matrix* tabel diatas berhasil mengklasifikasikan secara benar dengan nilai (*True Positif*) sebesar 400, data (*True Negatif*) 644, dan (*True Netral*) sebesar 85. Adapun hasil dari perhitungan nilai *accuracy*, *precision* dan *recall* terdapat pada tabel dibawah ini.

Tabel 10. Hasil Perhitungan Manual *Confusion Matrix K-Nearest Neighbor*

	Precision	Recall	Accuracy
Positif	87,34%	99,50%	93,23%
Negatif	96,70%	92,66%	
Netral	97,70%	74,56%	

Hasil Perbandingan Algoritma *Naïve Bayes* dan *K-Nearest Bayes*

Tabel 11. Hasil Perbandingan Algoritma *Naïve Bayes* dan *K-Nearest Neighbor*

Metode	Accuracy	Precission	Recall
Naïve Bayes	86,95%	80,51%	91,23%
K-Nearest Neighbor	93,23%	93,91%	91,56%

Hasil *accuracy* metode *Naïve Bayes* memperoleh hasil sebesar 86,95% dan metode *K-Nearest Neighbor* memperoleh hasil sebesar 93,23%. Sehingga penulis dapat menyimpulkan dari hasil *accuracy* perbandingan pada penelitian ini klasifikasi *K-Nearest Neighbor* merupakan metode klasifikasi terbaik dalam analisis sentimen mengenai intensif mobil listrik dengan hasil prediksi yang lebih akurat dengan tingkat *accuracy* sebesar 93,23%.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang dilakukan terkait dengan sentimen intensif mobil listrik sebanyak 1517 data komentar pada laman *youtube* menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dan pengujian menggunakan *confussion matrix* algoritma *Naïve Bayes, accuracy* yang diperoleh sebesar 86,95% dan algoritma *K-Nearest Neighbor* memperoleh hasil *accuracy* sebesar 93,23%. Maka dapat disimpulkan bahkan metode *K-Nearest Neighbor* lebih unggul dalam mengklasifikasi *text mining* dengan hasil perolehan *accuracy* 93,23, *precision* 93,91 dan *recall* 91,56. Saran untuk penelitian selanjutnya adalah diharapkan menggunakan dataset lebih banyak mengenai fenomena yang sedang tren. Menerapkan metode algoritma yang tidak banyak peneliti gunakan sehingga dapat mengetahui ferforma metode-metode yang digunakan dan menggunakan sumber data selain *youtube*.

DAFTAR PUSTAKA

[1] S. Wegner, F. Mersmann, and M. G. Grados, "G20 Response To the Energy Crisis: Critical for 1.5°C," 2022. [*Online*]. Available: www.climate-transparency.org

[2] S. Alfarizi and E. Fitriani, "Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization," *Indones. J. Softw. Eng.*, vol. 9, no. 1, pp. 19–27, 2023, [*Online*]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ijse

[3] A. F. Riyadi, F. R. Rahman, M. A. Nofa Pratama, M. K. Khafidli, and H. Patria, "Pengukuran Sentimen Sosial Terhadap Teknologi Kendaraan Listrik: Bukti Empiris di Indonesia," *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 11, no. 2, p. 141, 2021, doi: 10.36448/expert.v11i2.2171.

[4] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc," *Fakt. Exacta*, vol. 14, no. 2, p. 64, 2021, doi: 10.30998/faktorexacta.v14i2.9697.

[5] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.

[6] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, "Sentiment analysis of Covid-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012045, 2021, doi: 10.1088/1757-899x/1088/1/012045.

[7] L. Ardiantoro, S. Zahara, and N. Sunarmi, "Pemanfaatan Knowledge Data Discovery(KDD) Pada Pola Permainan Atlet Bulutangkis," *Explor. IT J. Keilmuan dan Apl. Tek. Inform.*, vol. 11, no. 1, pp. 1–6, 2019, doi: 10.35891/explorit.v11i1.1467.

[8] S. Alam, M. G. Resmi, and N. Masripah, "Classification of Covid-19 vaccine data screening with Naive Bayes algorithm using Knowledge Discovery in database method," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 4, no. 2, pp. 177–185, 2022, doi: 10.47709/cnahpc.v4i2.1584.

[9] A. Erfina and M. R. N. R. Alamsyah, "Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language," *Data Metadata*, vol. 2, pp. 2–11, 2023, doi: 10.56294/dm202345.

[10] I. Tri Julianto, D. Kurniadi, Y. Septiana, and A. Sutedi, "Alternative Text Pre-Processing using Chat GPT Open AI," *J. Nas. Pendidik. Tek. Inform.*, vol. 12, no. 1, pp. 67–77, 2023, doi: 10.23887/janapati.v12i1.59746.