

# PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP KEBIJAKAN SUBSIDI KENDARAAN LISTRIK

Aga Syabana Putra<sup>1\*</sup>, Dian Anubhakti<sup>2</sup>, Rusdah<sup>3</sup>, Lauw Li Hin<sup>4</sup>

<sup>1,2,3,4</sup> Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: <sup>1\*</sup>agasyabana@gmail.com, <sup>2</sup>dian.anubhakti@budiluhur.ac.id, <sup>3</sup>rusdah@budiluhur.ac.id, <sup>4</sup>lihin@budiluhur.ac.id  
(\* : corresponding author)

**Abstrak-** Ancaman krisis energi di masa depan mendorong pemerintah Indonesia untuk berperan aktif dalam mendukung peralihan penggunaan kendaraan dari kendaraan bermotor berbahan bakar minyak ke kendaraan listrik berbasis baterai dengan mengeluarkan peraturan terkait pemberian subsidi kendaraan listrik. Peran media sosial seperti *Twitter* dapat menjadi wadah dan sarana pengumpulan data yang dapat digunakan untuk mengetahui reaksi publik terhadap kebijakan pemberian subsidi kendaraan listrik yang dilakukan oleh pemerintah. Opini masyarakat terhadap kebijakan subsidi kendaraan listrik di media sosial *Twitter* cenderung negatif, sebab kebijakan tersebut dianggap tidak adil karena lebih memihak kepada masyarakat golongan menengah ke atas. Namun demikian belum terdapat penelitian yang mengklasifikasikan sentimen opini masyarakat di *Twitter* terhadap kebijakan tersebut. Penelitian ini bertujuan untuk mencari nilai positif dan negatif atas kebijakan pemberian subsidi kendaraan listrik menggunakan algoritma *K-Nearest Neighbor* dengan mengimplementasikan *Cross-Industry Standard Process for Data Mining (CRISP-DM)* menggunakan kata kunci “Subsidi Kendaraan Listrik”, “Subsidi Mobil Listrik”, dan “Mobil Listrik”. Proses pengumpulan data dilakukan dengan menggunakan bantuan tools *Rapidminer*. Data awal yang dikumpulkan berjumlah 7,518 data *tweet* namun setelah melewati tahap *data preprocessing* dan membuang data bersentimen netral, maka jumlah data yang siap digunakan berkurang menjadi 1,107 *data tweet* dimana 298 data berlabel positif dan 809 data berlabel negatif, sehingga dapat diketahui 73.08% dari seluruh data *tweet* yang digunakan dalam penelitian ini berisi sentimen negatif. Berdasarkan penerapan model menggunakan algoritma *K-Nearest Neighbor* dan metode *Split Data*, diperoleh nilai *accuracy* sebesar 73.42%, *precision* sebesar 81.99%, dan *recall* sebesar 81.48%.

**Kata Kunci:** Subsidi Kendaraan Listrik, *Text Mining*, Analisis Sentimen, *K-Nearest Neighbor*

## APPLICATION OF K-NEAREST NEIGHBOR ALGORITHM FOR SENTIMENT ANALYSIS OF TWITTER USERS ON ELECTRIC VEHICLE SUBSIDY POLICIES

**Abstract-** The threat of an energy crisis in the future has prompted the Indonesian government to play an active role in supporting the transition in vehicle use from oil-fueled motorized vehicles to battery-based electric vehicles by issuing regulations regarding the provision of subsidies for electric vehicles. The role of social media such as *Twitter* can be a forum and means of collecting data that can be used to find out the public's reaction to the government's policy of subsidizing electric vehicles. Public opinion towards the electric vehicle subsidy policy on *Twitter* social media tends to be negative, this policy is considered unfair because it is more beneficial to the middle and upper economic class. However, there is no research that classifies the sentiment of public opinion on *Twitter* towards this policy. This study aims to find positive and negative values for the policy of subsidizing electric vehicles using the *K-Nearest Neighbor* algorithm by implementing the *Cross-Industry Standard Process for Data Mining (CRISP-DM)* using the keywords "Electric Vehicle Subsidies", "Electric Car Subsidies", and "Electric Cars". The data collection process was carried out using the help of *Rapidminer* tools. The initial data collected amounted to 7,518 *tweet* data but after going through the *data preprocessing* stage and removing neutral sentiment data, the amount of data ready for use was reduced to 1,107 *tweet* data where 298 data were labeled as positive and 809 data were labeled as negative, thus it can be seen that 73.08% of all *tweet* data used in this study contains negative sentiments. Based on the application of the model using the *K-Nearest Neighbor* algorithm and the *Split Data* method, the *accuracy* value is 73.42%, *precision* is 81.99%, and *recall* is 81.48%.

**Keywords:** *Electric Vehicle Subsidies*, *Text Mining*, *Sentiment Analysis*, *K-Nearest Neighbor*

## 1. PENDAHULUAN

Dengan tujuan mendorong percepatan adopsi kendaraan listrik dan mempercepat menuju transisi yang berkelanjutan, kemudian pemerintah Indonesia melalui Kementerian Keuangan mengeluarkan regulasi subsidi mobil listrik dan bus listrik. Untuk Kendaraan Bermotor Tenaga Baterai (KLBBB) yang memenuhi persyaratan,

konsumen hanya membayar 1% Pajak Pertambahan Nilai (PPN), dari 11% yang seharusnya dibayar. Relaksasi ini tertuang dalam Peraturan Menteri Keuangan Nomor 38 Tahun 2023 tentang Pajak Pertambahan Nilai atas penyediaan KBLBB roda empat tertentu dan bus KBLBB tertentu yang dibiayai pemerintah. Syarat untuk mendapatkan subsidi ini adalah kendaraan harus memiliki Tingkat Komponen Nasional (TKDN) minimal 40%. Namun kebijakan tersebut banyak menuai tanggapan pro dan kontra di kalangan masyarakat, khususnya pengguna *Twitter* dimana kebijakan tersebut dianggap lebih menguntungkan masyarakat golongan ekonomi menengah ke atas, sementara sebagian masyarakat masih kesulitan memenuhi kebutuhan dasar. Selain itu, Belum adanya penelitian yang mengklasifikasikan sentimen opini masyarakat di *Twitter* terhadap kebijakan subsidi kendaraan listrik dengan menggunakan algoritma *K-Nearest Neighbor*. Rumusan permasalahan dari penelitian ini adalah bagaimana penerapan algoritma dan nilai akurasi yang diperoleh algoritma *K-Nearest Neighbor* dalam melakukan analisis sentimen masyarakat Indonesia terhadap kebijakan pemerintah tersebut. Tujuan penelitian ini adalah untuk memperoleh hasil dari analisis sentimen dengan metode *K-Nearest Neighbor* sebagai bahan evaluasi dan referensi bagi pemerintah agar dikemudian hari bisa lebih berhati-hati dan tepat sasaran dalam menentukan kebijakan.

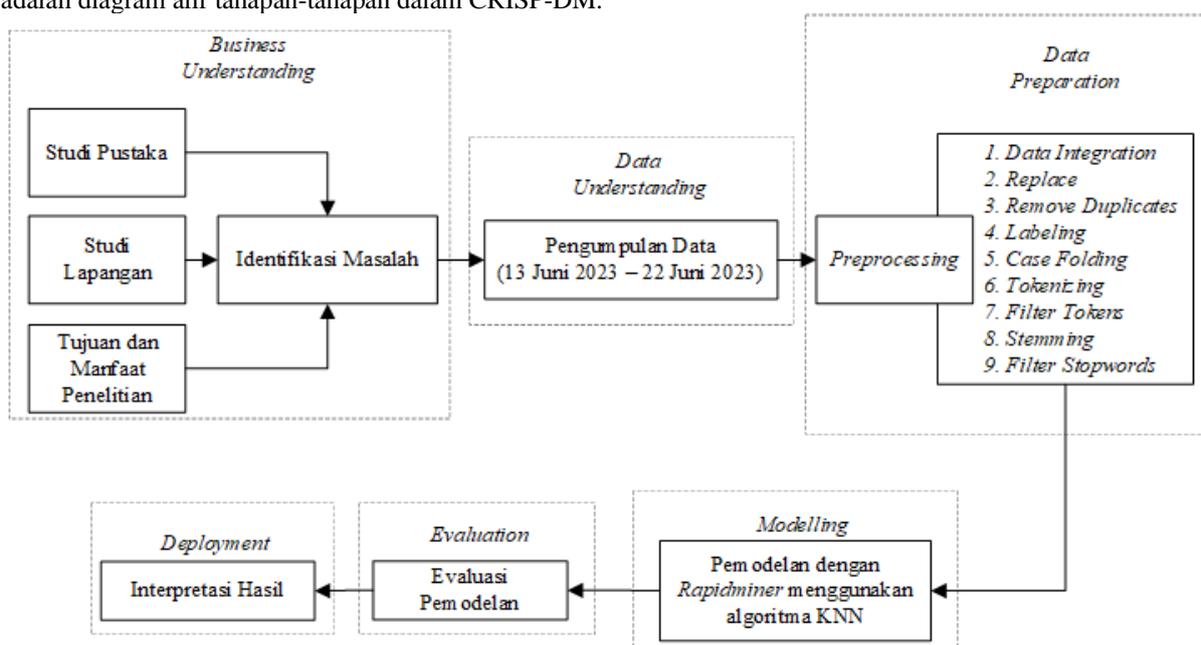
Data *tweet* yang telah dikumpulkan selanjutnya dilakukan analisis lebih lanjut untuk mengetahui secara umum bagaimana tanggapan masyarakat terhadap kebijakan tersebut. Salah satu cara untuk meperoleh data-data yang dibutuhkan untuk melakukan analisis sentimen yaitu dengan melakukan *crawling* menggunakan *tools rapidminer*[2]. *Tweet* dapat di analisis sentimennya dengan cara manual untuk menentukan sentimen positif ataupun negatif, akan tetapi apabila jumlah *tweet* yang digunakan berjumlah ribuan maka akan memerlukan waktu yang sangat lama. Oleh karena itu digunakan metode CRISP-DM yang akan membuat proses analisis menjadi lebih terstruktur dan sistematis [3].

Penelitian terdahulu telah dilakukan analisis sentimen sosial terhadap teknologi kendaraan listrik menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbor*, dan *Decision Tree* [4]. Telah dilakukan juga penelitian mengenai sentimen terhadap penghapusan subsidi BBM dengan algoritma *Naïve Bayes* [5]. Serta penelitian tentang opini penggunaan kendaraan listrik menggunakan algoritma *Naïve Bayes* [6]. Maka pada penelitian ini akan menggunakan metode text mining dan analisis sentimen menggunakan algoritma KNN untuk mengetahui opini masyarakat terhadap kebijakan subsidi kendaraan listrik sehingga dapat digunakan sebagai data analisis opini publik [7]. Kemudian dari hasil pengukuran yang didapat akan dilakukan pengujian menggunakan *confusion matrix* [8].

## 2. METODE PENELITIAN

### 2.1 Metode CRISP-DM

Tahapan penelitian ini mengacu pada metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk memudahkan peneliti dalam proses data mining agar lebih lebih terstruktur dan sistematis. Berikut ini adalah diagram alir tahapan-tahapan dalam CRISP-DM.



Gambar 1. Tahapan Penelitian Dengan CRISP-DM

a. *Business Understanding*

Dari permasalahan yang terjadi tentang subsidi kendaraan listrik penulis mencoba untuk memahami permasalahan tersebut dengan melakukan pencarian informasi dan mempelajari artikel yang berkaitan dengan peraturan subsidi kendaraan listrik yang dikeluarkan oleh kementerian keuangan yang dituangkan dalam Peraturan Menteri Keuangan Republik Indonesia Nomor 38 Tahun 2023. Selain itu juga dilakukan studi pustaka dengan mencari dan membaca jurnal serta buku terkait penelitian yang dilakukan demi memperkaya pengetahuan dan memahami teori yang dapat membantu proses penelitian.

b. *Data Understanding*

Data yang diambil adalah data tweet yang diunggah pada periode 13 Juni 2023 sampai 22 Juni 2023 “Kendaraan Listrik”, “Subsidi Mobil Listrik”, dan “Mobil Listrik”. Dari hasil proses pengambilan data tersebut maka didapatkan sebanyak 7,518 data tweet.

c. *Data Preparation*

Pada fase ini dilakukan preprocessing data. Data yang sudah didapat dilakukan pemrosesan seperti *replace*, *remove duplicates*, *labeling*, *case folding*, *tokenizing*, *filter token*, *stemming*, dan *filter stopwords*.

d. *Modeling*

Pada fase ini dilakukan pemodelan terhadap data yang telah melalui fase data preprocessing menggunakan metode split data dan *cross validation*. Pada split data, pembagian data *training* dan data *testing* dilakukan dengan menggunakan rasio 60:40, 70:30, dan 80:20. Sementara pada *cross validation* menggunakan *numbers of fold* 10.

e. *Evaluation*

Pada tahap ini penulis akan melakukan evaluasi terhadap data yang telah melalui fase *modelling* menggunakan metode *confusion matrix*.

f. *Deployment*

Pada tahap penyebaran ini hasil penelitian akan ditemukan sesuai tujuan dari penelitian. Sehingga dapat diambil kesimpulan dari seluruh proses penelitian.

## 2.2 K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) mengklasifikasikan objek berdasarkan data pembelajaran yang memiliki jarak terdekat dengannya. Pada perhitungan algoritma KNN, perhitungan jarak yang paling umum digunakan adalah dengan menggunakan persamaan *cosine similarity* sebagai berikut.

$$\cos(\theta_{ij}) = \frac{\sum k (d_{ik}d_{jk})}{\sqrt{\sum k d_{ik}^2} \sqrt{\sum k d_{jk}^2}} \quad (1)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Proses pengumpulan data dilakukan dengan cara mengambil data *tweet* menggunakan *tools Rapidminer* dengan mengambil *tweet* yang mengandung kata kunci ‘Subsidi Kendaraan Listrik’ dan ‘Subsidi Mobil Listrik’, dan ‘Mobil Listrik’. Data *tweet* yang diambil adalah *tweet* yang diunggah mulai tanggal 13 Juni 2023 sampai 22 Juni 2023. *Tweet* yang diambil hanya *tweet* dengan Bahasa Indonesia dan mempunyai limit data dengan jumlah sepuluh ribu data. Jumlah data *tweet* yang telah didapat dari proses *crawling* sebanyak 7,518 dan tersimpan dalam file *csv*. Jumlah hasil data *crawling* untuk masing-masing kata kunci dapat dilihat pada tabel dibawah ini.

**Tabel 1.** Data Hasil Crawling Berdasarkan Kata Kunci

Kata Kunci	Jumlah Tweet
Subsidi Kendaraan Listrik	1,415
Subsidi Mobil Listrik	939
Kendaraan Listrik	5,164
Total	7,518

### 3.2 Data Preprocessing

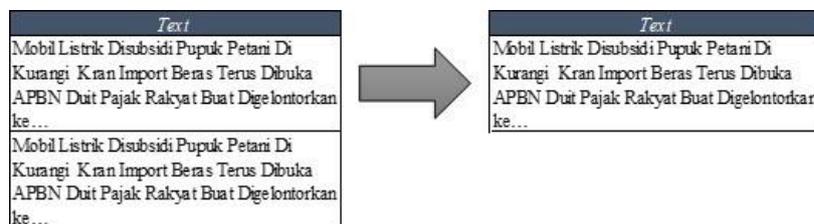
Proses ini meliputi beberapa tahapan, antara lain *data integration*, *replace*, *remove duplicates*, pelabelan sentimen, *case folding*, *tokenize*, *filter tokens*, *stemming*, dan *filter stopwords* [9]. Berikut ini merupakan tahapan-tahapannya.

- a. **Replace**, bertujuan untuk menghapus link, mention, hashtag, dan karakter khusus yang terdapat pada *tweet*. Tahapan ini dilakukan dengan menggunakan operator *replace*.

Tabel 2. Contoh Proses *Replace*

Sebelum Proses <i>Replace</i>	Setelah Proses <i>Replace</i>
BETUL !???? TOLAK SUBSIDI MOBIL LISTRİK !? https://t.co/4DUSzUr2Xg	BETUL TOLAK SUBSIDI MOBIL LISTRİK

- b. **Remove duplicates**, yaitu proses menghilangkan data yang terduplikasi. Proses ini dilakukan dengan menggunakan operator *Remove Duplicates*. Dari proses ini jumlah *tweet* berkurang dari 7,518 menjadi 1,329 *tweet*.



Gambar 2. Hasil Proses *Remove Duplicates*

- c. **Pelabelan Sentimen**, proses ini dilakukan oleh pakar dengan menganalisa setiap *tweet* yang sudah didapatkan. Label sentimen yang diberikan pada *tweet* berupa sentimen positif, netral, dan negatif.

Tabel 3. Contoh *Tweet* Setelah Diberikan Label Sentimen

<i>Text</i>	Label
BETUL TOLAK SUBSIDI MOBIL LISTRİK	Negatif
Akulah si penikmat subsidi mobil listrik maafkeun	Positif
Respon Anindya Bakrie soal Beli Bus Listrik Bakal Dapat Subsidi dari Pemerintah	Netral

Dari total 1,329 *tweet* yang telah diberikan label sentimen positif, negatif, dan juga netral, maka jumlah data *tweet* berkurang menjadi 1,107 data *tweet*. Hal ini dikarenakan terdapat beberapa *tweet* yang berlabel netral sebanyak 222 *tweet*. Sehingga dalam dataset terdapat 26.92% *tweet* dengan sentimen positif dan 73.08% *tweet* dengan sentimen negatif.

- d. **Case folding**, merupakan suatu proses yang dilakukan untuk menyeragamkan semua bentuk huruf. Pada penelitian ini bentuk huruf yang digunakan adalah huruf kecil. Sehingga semua bentuk huruf kapital akan diubah menjadi huruf kecil.

Tabel 4. Contoh Data Hasil *Case Folding*

Sebelum <i>Transform Cases</i>	Sebelum <i>Transform Cases</i>
BETUL TOLAK SUBSIDI MOBIL LISTRİK	betul tolak subsidi mobil listrik

- e. **Tokenize**, *tokenizing* merupakan proses penguraian yang semula berupa kalimat menjadi kata [10]. Dalam proses ini data *tweet* yang masih berupa kalimat akan dipecah menjadi kata per kata, Sehingga dari kalimat-kalimat tersebut menghasilkan sebanyak 4,514 atribut kata. Proses ini juga menghasilkan pembobotan untuk masing-masing kata berdasarkan frekuensi kemunculannya.

Tabel 5. Contoh Hasil Proses *Tokenize*

Sebelum <i>Tokenize</i>	Sesudah <i>Tokenize</i>
normal tuh kaya apa subsidi pupuk di cabut mobil listrik di subsidi	"normal", "tuh", "kaya", "apa", "subsidi", "pupuk", "di", "cabut", "mobil", "listrik", "di", "subsidi"

- f. **Filter tokens**, Pada langkah ini data yang dihasilkan dari proses *tokenize* difilter berdasarkan panjang karakter atau jumlah huruf terkecil dalam sebuah kata. Ini dilakukan dengan *Rapidminer* menggunakan operator *filter*

*token* berdasarkan panjang, yang mengubah jumlah minimum menjadi empat karakter. Jumlah atribut kata turun dari 4,514 menjadi 4,043 sebagai hasil dari proses ini.

**Tabel 6.** Contoh Data Hasil Proses *Filter Token*

Sebelum Proses <i>Filter Token</i>	Setelah Proses <i>Filter Token</i>
"normal", "tuh", "kaya", "apa", "subsidi", "pupuk", "di", "cabut", "mobil", "listrik", "di", "subsidi"	"normal", "kaya", "subsidi", "pupuk", "cabut", "mobil", "listrik", "subsidi"

- g. **Stemming**, pada tahap *stemming*, kalimat-kalimat yang sudah ditokenisasi akan dilakukan proses *stemming* untuk mengubah kata-kata berimbuhan menjadi kata dasar. Proses ini juga mengurangi atribut kata sehingga jumlahnya menjadi 2,218 atribut kata.

**Tabel 7.** Contoh Data Sebelum dan Sesudah *Stemming*

Sebelum Proses <i>Stemming</i>	Setelah Proses Proses <i>Stemming</i>
"karena", "subsidi", "untuk", "setiap", "pembelian", "mobil", "listrik", "lumayan", "besar"	"karena", "subsidi", "untuk", "tiap", "beli", "mobil", "listrik", "lumayan", "besar"

- h. **Filter Stopwords**, dalam proses ini kata-kata yang dering muncul dan tidak memiliki makna yang berhubungan dengan topik penelitian akan dieliminasi dari dataset.

**Tabel 8** Contoh Data Sebelum dan Sesudah *Filter Stopwords*

Sebelum Proses <i>Filter Stopword</i>	Setelah Proses <i>Filter Stopword</i>
"malah", "mobil", "listrik", "subsidi", "orang", "kaya", "lagi", "yang", "dapat", "subsidi"	"mobil", "listrik", "subsidi", "orang", "kaya", "subsidi"

### 3.3 Penentuan Data Latih dan Data Uji

Data akan dibagi menjadi data training dan data testing menggunakan beberapa rasio perbandingan. Penentuan data latih dan data uji dibagi dengan menggunakan rasio 60:40, 70:30, dan 80:20.

**Tabel 9** Rasio Perbandingan Data Latih dan Data Uji

Perbandingan 60:40			
Sentimen	<i>Training</i>	<i>Testing</i>	Total
Positif	179	119	298
Negatif	485	324	809
Total	664	443	1107
Perbandingan 70:30			
Sentimen	<i>Training</i>	<i>Testing</i>	Total
Positif	209	89	298
Negatif	566	243	809
Total	775	332	1107
Perbandingan 80:20			
Sentimen	<i>Training</i>	<i>Testing</i>	Total
Positif	238	60	298
Negatif	647	162	809
Total	885	222	1107

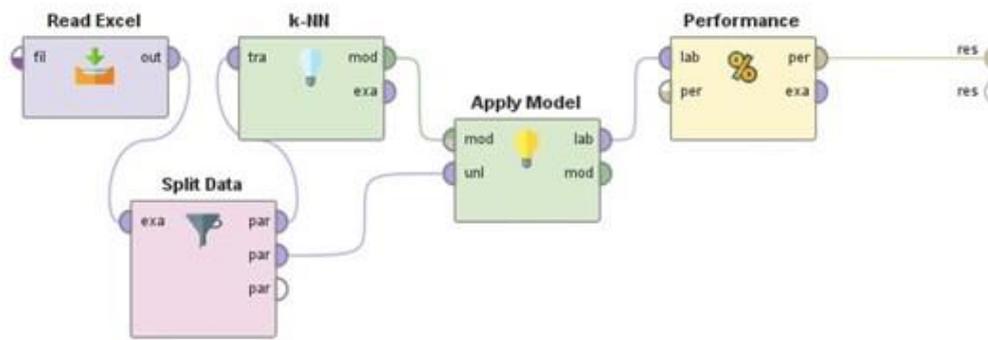
### 3.4 Modeling

Pemodelan dilakukan dengan menggunakan algoritma *K-Nearest Neighbor*. Selain itu juga akan ditampilkan hasil pemodelan dengan menggunakan algoritma *Naïve Bayes* dan *Decision Tree* sebagai pembandingan.

#### 3.4.1 Hasil Komparasi Model

- a. **Modeling Dengan Split Data**

Proses modeling dilakukan dengan menggunakan teknik split data dengan rasio perbandingan data latih dan data uji sebesar 60:40, 70:30, dan 80:20. selain itu pada proses ini juga menggunakan *k-fold cross validation* dengan 10-fold *cross validation*, yang berarti membagi data keseluruhan menjadi 10 bagian. Sedangkan penggunaan teknik pembagian data dan porsi data yang berbeda bertujuan untuk melihat konsistensi model dalam menghasilkan prediksi pada berbagai kondisi data.



Gambar 3. Proses Modeling Menggunakan Split Data Di Rapidminer

Tabel 10. Hasil Perbandingan Nilai Accuracy Menggunakan Split Data

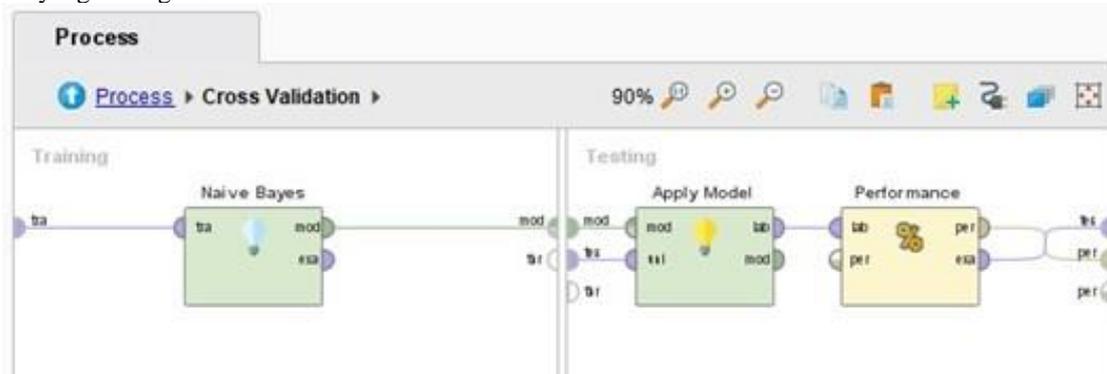
Rasio Perbandingan (Accuracy)	Naïve Bayes	Algoritma K-NN	Decision Tree
60:40	72.23%	73.36%	77.43%
70:30	71.08%	71.99%	76.51%
80:20	70.27%	<b>73.42%</b>	75.23%

Tabel 11. Hasil Perbandingan Nilai AUC Menggunakan Split Data

Rasio Perbandingan (nilai AUC)	Naïve Bayes	Algoritma K-NN	Decision Tree
60:40	0.624	0.756	0.500
70:30	0.610	0.737	0.500
80:20	0.600	<b>0.756</b>	0.500

b. Modeling Dengan Cross Validation

Pada modeling menggunakan metode ini digunakan *number of folds* 10. Pemilihan jumlah fold tersebut adalah salah satu yang direkomendasikan untuk pemilihan model terbaik karena dapat memberikan estimasi akurasi yang kurang bias.



Gambar 4. Proses Modeling Menggunakan Cross Validation Di Rapidminer

Dari hasil pemodelan menunjukkan dengan menggunakan *split data* dengan rasio 80:20 menghasilkan nilai *accuracy* dan AUC paling besar dengan nilai *accuracy* 73.42% dan nilai AUC sebesar 0.756.

Tabel 12. Perbandingan Nilai Accuracy Menggunakan Cross Validation

Algoritma	Accuracy
Naïve Bayes	64.50%
K-NN	66.83%
Decision Tree	75.34%

Tabel 13 Perbandingan Nilai AUC Menggunakan Cross Validation

Algoritma	Nilai AUC
Naïve Bayes	0.512

K-NN	0.749
Decision Tree	0.500

### 3.4.2 Penyajian Model Terbaik

Algoritma KNN menggunakan *split data* dengan rasio 80:20 digunakan sebagai model prediksi sentimen dari *tweet* yang membahas tentang subsidi kendaraan listrik dengan hasil luaran *confusion matrix* sebagai berikut:

**Tabel 14.** Luaran Model Terbaik

accuracy: 73.42%			
	true negatif	true positif	class precision
pred. negatif	132	29	81.99%
pred. positif	30	31	50.82%
class recall	81.48%	51.67%	

### 3.5 Perhitungan Manual Algoritma Terpilih

Perhitungan manual algoritma terpilih menggunakan *K-Nearest Neighbor* dengan sample *dataset* yang telah diberikan label sentimen di tahap sebelumnya. Data yang diambil sebanyak 5 data *tweet* secara acak.

**Tabel 15.** Sample Tweet

Sample	Text	Label
D1	Negeri yg lucu subsidi listrik untuk rakyat kecil dicabut trus kendaraan listrik malah	Negatif
D2	Stop subsidi kendaraan listrik	Negatif
D3	Saatnya beralih ke kendaraan listrik yg ramah lingkungan	Positif
D4	Keren mudah2an lebih fokus ke transportasi umum daripada mobil listrik	Positif
D5	Malah subsidi mobil listrik yg notabeneanya orang kaya	?

Berdasarkan sample *tweet* yang digunakan, dimana pada S5 belum terdapat label sentimennya maka akan dilakukan pemberian label dengan menghitung frekuensi kemunculan kata dan bobot setiap kata dengan menghitung TF-IDF. Setelah mendapatkan bobot TF-IDF, langkah selanjutnya adalah menghitung perkalian skalar antara data uji (D5) dan data pelatihan (D1, D2, D3, dan D4). Kemudian jumlahkan hasil perkalian tersebut pada masing-masing data pelatihan dan kemudian menghitung kemiripan panjang vektor setiap tweet dengan mengkuadratkan bobot setiap term, kemudian jumlahkan nilai kuadrat tersebut dan diakarkan sehingga hasilnya didapat sebagai berikut.

**Table 16.** Panjang Vektor Tiap Sample

Panjang Vektor				
D1	D2	D3	D4	D5
0.0283	0.0343	0.0592	0.0532	0.0236
0.1682	0.1853	0.2432	0.2308	0.1535

Setelah mendapatkan panjang vektor selanjutnya menerapkan *cosine similarity* untuk mengetahui kemiripannya.

$$\begin{aligned} \text{Cos}(D5, D1) &= \frac{0.0047}{0.1535 \times 0.1682} = 0.1820 \\ \text{Cos}(D5, D2) &= \frac{0.0019}{0.1535 \times 0.1853} = 0.0668 \\ \text{Cos}(D5, D3) &= \frac{0.0002}{0.1535 \times 0.2432} = 0.0054 \\ \text{Cos}(D5, D4) &= \frac{0.0029}{0.1535 \times 0.2308} = 0.0819 \end{aligned}$$

Setelah mendapatkan hasil perhitungan dengan menggunakan persamaan *cosine similarity*, maka langkah selanjutnya adalah mengurutkan hasil dari perhitungan berdasarkan jarak dari nilai terbesar ke terkecil.

**Tabel 17.** Urutan Jarak dari Nilai Terbesar ke Terkecil

D1	D4	D2	D3
----	----	----	----

0.1820	0.0819	0.0668	0.0053
--------	--------	--------	--------

Karena dalam perhitungan manual ini menggunakan 5 *sample* data maka diambil nilai *k* sebanyak (*k*=3) dari yang paling tinggi tingkat kemiripannya dengan D5.

**Tabel 18.** Sentimen Terbanyak Berdasarkan Nilai *k* (*k*=3)

D1	D4	D2
Negatif	Positif	Negatif

Maka dapat disimpulkan bahwa *tweet* D5 merupakan *tweet* yang memiliki sentimen negatif. Karena jumlah sentimen yang kemunculannya paling banyak pada nilai *k* (*k*=3) adalah sentimen negatif.

### 3.6 Pengujian (Evaluation)

Pada penelitian ini dilakukan *evaluation test* dengan *confusion matrix* dimana nilai yang digunakan didapat dari tahap pemodelan dengan algoritma *K-Nearest Neighbor* yang mendapat akurasi tertinggi menggunakan rasio *split data* 80:20.

**Tabel 19** *Confusion Matrix K-Nearest Neighbor* dengan perbandingan 80:20

	<i>True Negatif</i>	<i>True Positif</i>	<i>Class Precision</i>
<i>pred. Negative</i>	132	29	81.99%
<i>pred. Positive</i>	30	31	50.82%
<i>Class Recall</i>	81.48%	51.67%	

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{31 + 132}{31 + 132 + 30 + 29} = 73.42\%$$

$$Precision = \frac{TN}{TN + FN} = \frac{132}{132 + 29} = 81.99\%$$

$$Recall = \frac{TN}{TN + FP} = \frac{132}{132 + 30} = 81.48\%$$

Berdasarkan pengujian dengan *confusion matrix* diatas maka dapat diketahui nilai *precision* dan *recall* menunjukkan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan untuk *accuracy* menunjukkan nilai keakuratan dalam memprediksi suatu sentimen dengan sebuah algoritma. Setelah dilakukan pengujian menggunakan *confusion matrix* maka diperoleh nilai *accuracy* 73.42%, dengan nilai *precision* 81.99% untuk prediksi negatif dan 50.82% untuk prediksi positif, serta nilai *recall* 81.48% untuk kelas negatif dan 51.67% untuk kelas positif.

## 4. KESIMPULAN

Dari hasil yang didapat setelah dilakukan penelitian, maka dapat diambil kesimpulan untuk memprediksi klasifikasi opini masyarakat Indonesia terhadap kebijakan pemberian subsidi kendaraan listrik yaitu mayoritas opini yang dibagikan oleh masyarakat mengenai kebijakan subsidi kendaraan listrik di *Twitter* berisi sentimen negatif dengan jumlah persentase 73.08%. Penerapan algoritma *K-Nearest Neighbor* terbukti dapat mengklasifikasikan sentimen masyarakat terhadap kebijakan pemberian subsidi kendaraan listrik yang ditetapkan oleh pemerintah Indonesia ditunjukkan oleh nilai *accuracy* yang didapat, yaitu sebesar 73.42%. Tingkat presisi untuk prediksi label negatif cukup baik yaitu sebesar 81.99%, dimana model berhasil memprediksi *tweet* negatif dengan benar sebanyak 132 *tweet* dari keseluruhan 162 *tweet* negatif. Sedangkan pada prediksi positif tingkat keberhasilan prediksi rendah yaitu sebesar 50.82%, dimana model hanya mampu memprediksi *tweet* positif dengan benar sebanyak 31 *tweet* dari keseluruhan 60 *tweet* positif. Hal ini disebabkan pada kedua label sentimen memiliki banyak kemunculan kata yang sama sehingga kata-kata yang sering digunakan pada *tweet* positif juga sering muncul di *tweet* negatif.

## DAFTAR PUSTAKA

- [1] Kementerian Keuangan Republik Indonesia, “Peraturan Pemerintah Tahun Nomor 38 Tahun 2023 Tentang Pajak Pertambahan Nilai Atas Penyerahan Kendaraan Bermotor Listrik Berbasis Baterai Roda Empat Tertentu Dan Kendaraan Bermotor Listrik Berbasis Baterai Bus Tertentu Yang Ditanggung Pemerintah Tahun Anggaran 2023,” 2023.
- [2] G. Fajriansyah, G. Forda Nama, Y. Mulyani, J. Sumantri, B. No, and B. Lampung, “ANALISIS DAFTAR PEMILIH TETAP PADA HASIL REKAPITULASI KPU BERDASARKAN USIA MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS : KOTA BANDAR LAMPUNG),” *Electrician : Jurnal Rekayasa dan Teknologi Elektro*, vol. 15, no. 1, pp. 39–53, Jan. 2021, doi: 10.23960/ELC.V15N1.2147.
- [3] N. Sucahyo, L. Nurlaela, and R. R. Waryono, “Analisis Sentimen Masyarakat Jakarta Terhadap Kebijakan Perluasan Dan Perpanjangan Ganjil Genap di Media Sosial Twitter,” *Jurnal Teknologi Informatika dan Komputer*, vol. 7, no. 1, pp. 97–111, Mar. 2021, doi: 10.37012/jtik.v7i1.506.
- [4] 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: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 11, no. 2, p. 141, Dec. 2021, doi: 10.36448/expert.v11i2.2171.
- [5] Y. Sergio, V. Putranta, B. Rahayudi, and W. Purnomo, “Analisis Sentimen Masyarakat terhadap Kebijakan Penghapusan Subsidi BBM pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier dengan Ekstraksi Fitur N-Gram TF-IDF,” 2023. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] A. Agustian, Tukino, and F. Nurapriani, “PENERAPAN ANALISIS SENTIMEN DAN NAIVE BAYES TERHADAP OPINI PENGGUNAAN KENDARAAN LISTRIK DI TWITTER,” *Jurnal Tika*, vol. 7, no. 3, pp. 243–249, 2022.
- [7] A. Luthfika Fairuz, R. Dias Ramadhani, N. Annisa, and F. Tanjung, “Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter,” 2021. [Online]. Available: <http://journal.ittelkom-pwt.ac.id/index.php/dinda>
- [8] C. Anam and H. B. Santoso, “Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa,” *Jurnal Ilmiah Ilmu-Ilmu Teknik*, vol. 8, no. 1, pp. 2088–4591, 2018.
- [9] O. Somantri and D. Dairoh, “Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining,” *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 2, pp. 191–196, Aug. 2019, doi: 10.26418/JP.V5I2.32661.
- [10] O. Y. Findawati, M. M. Muhammad, A. Rosid, S. Kom, and M. Kom, *BUKU AJAR TEXT MINING*, vol. 1. UMSIDA PRESS, 2021.