JURNAL FASILKOM
Volume 13 No. 3 | Desember 2023: 511-517

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Analisis Sentimen Kebijakan Pemberian Subsidi Motor Listrik Menggunakan Metode Support Vector Machine

Sriani[1], Suhardi[2], Irsan Frianda Gultom[3]
[123]Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara
[1]sriani@uinsu.ac.id, [2]suhardi@uinsu.ac.id, [3]irsan.frianda@uinsu.ac.id

Abstract

The development of electric vehicles is an innovation and technology that continues to develop and change. With the presence of electric vehicles, the Indonesian government is trying to encourage people to switch from fuel powered vehicles to electric powered vehicles. Therefore, the government through the Coordinating Ministry for Maritime Affairs and Investment issued a policy of subsidizing electric motorbikes. This study aims to determine the level of accuracy and application of the Support Vector Machine method in processing sentiment data. With the dataset used as many as 700 tweets obtained from 1000 filtered tweet data from social media Twitter. By going through the preprocessing process and labeling each tweet using the Lexicon dictionary. The next stage is classification with a Support Vector Machine which uses a linear kernel as a hyperplane determinant. From the use of the Support Vector Machine method, an accuracy value of 86.43%, a precision of 83.33%, a recall of 87.30%, and an f1-score of 85.27% are obtained. With this high accuracy obtained, the use of the SVM method is considered to be one of the appropriate methods for calculating accuracy values in sentiment analysis.

Keywords: sentiment analysis, electric motor subsidies, support vector machine.

Abstrak

Perkembangan kendaraan listrik merupakan sebuah inovasi dan teknologi yang terus berkembang dan berubah. Dengan hadirnya kendaraan listrik tersebut pemerintah Indonesia berupaya untuk mendorong masyarakat agar beralih dari kendaraan bertenaga bahan bakar minyak ke kendaraan bertenaga listrik. Oleh karena itu pemerintah melalui Kementrian Koordinator Bidang Kemaritiman dan Investasi (Menko Marves) mengeluarkan kebijakan pemberian subsidi motor listrik. Pada penelitian ini bertujuan untuk mengetahui tingkat akurasi dan penerapan metode *Support Vector Machine* dalam mengolah data sentimen. Dengan dataset yang digunakan sebanyak 700 *tweet* yang diperoleh dari 1000 data *tweet* yang telah disaring dari media sosial *Twitter*. Dengan melewati proses *preprocessing* dan memberikan label pada setiap *tweet* menggunakan kamus *Lexicon*. Tahap selanjutnya adalah klasifikasi dengan *Support Vector Machine* yang menggunakan kernel *linear* sebagai penentu *hyperplane*. Dari penggunaan metode *Support Vector Machine* tersebut diperoleh nilai *accuracy* sebesar 86.43%, *precision* sebesar 83.33%, *recall* sebesar 87.30%, dan *f1-score* sebesar 85.27%. Dengan perolehan akurasi yang cukup tinggi tersebut maka penggunaan metode SVM dianggap menjadi salah satu metode yang tepat untuk menghitung nilai akurasi pada analisis sentimen.

Kata kunci: analisis sentimen, subsidi motor listrik, *support vector machine*

1. Pendahuluan

Text mining dapat didefinisikan sebagai sebuah proses yang dilakukan untuk mengekstrak pengetahuan implisit yang tersembunyi pada sebuah data tekstual. Pengetahuan implisit yang diperoleh dari hasil ekstraksi *text mining* perlu dikelola secara mendalam karena memiliki output yang berbeda dibanding dengan pengelolaan pada tipe data yang lain sehingga perlu analisis secara terpisah. Dalam *text mining* terdapat beberapa proses yang membedakan dengan pengolahan *data mining* lainnya yaitu kategori teks, ekstraksi konsep atau entitas, *text clustering*, *sentiment analysis*, produksi taksonomi granular, penyimpulan dokumen serta pemodelan relasi entitas.[1]

Klasifikasi pada proses *text mining* dapat dikerjakan dengan menggunakan berbagai jenis metode klasifikasi, yang salah satunya yaitu metode *Support Vector Machine* (SVM). *Support Vector Machine* adalah teknik regresi, metode pengklasifikasian data berdasarkan data sebelumnya, dan pemodelan dilakukan terlebih dahulu. SVM termasuk dalam tipe pengklasifikasi linear biner dan tidak probabilistik. *Support Vector Machine* menggunakan batasan keputusan untuk menentukan klasifikasi data latih sehingga dapat dibangun model linier atau *hyperplane* terbaik untuk mengklasifikasikan data tersebut.[2]

Tingkat akurasi model yang dihasilkan oleh proses transisi SVM sangat bergantung pada fungsi kernel dan parameter yang digunakan. Metode SVM dibedakan menjadi dua menurut karakteristiknya yaitu *linear* dan *non linear*. *Linear* SVM merupakan data yang dipisahkan secara *linear* yaitu memisahkan dua kelas pada *hyperplane* dengan *soft margin*. Sedangkan fungsi trik kernel *non linier* berada pada ruang berdimensi tinggi. SVM sangat cepat dan efisien dalam

P-ISSN : 2089-3353
E-ISSN : 2808-9162

menyelesaikan masalah klasifikasi teks, secara geometris klasifikasi *biner* dapat dilihat sebagai *hyperplane* ruang fitur yang memisahkan titik-titik yang mewakili *instance* positif dari kelas-kelas yang mewakili keadaan negatif. [3]

Dengan penggunaan metode tersebut dapat digunakan pada topik keputusan Pemerintah Indonesia dalam hal subsidi motor listrik ini tentu menjadi bahasan yang menarik sehingga menimbulkan beragam reaksi komentar pro dan kontra dari para pengguna *Twitter*. Pembatasan pengguna dalam menyaring informasi tersebut tentu akan menghambat kemampuan masyarakat atau bahkan pemerintah dalam menganalisa informasi yang beredar di masyarakat.

Mengutip dari media berita CNBC Indonesia Selasa (13/12/2022), ekonom Indef Eko Listiyanto menilai kebijakan tersebut bukanlah tujuannya. Sebab, ia menilai pemberian insentif sebesar itu kepada kelompok kelas menengah menunjukkan ketidakadilan. Menurutnya, 6,5 juta rubel lebih cocok untuk digunakan masyarakat miskin.[4]

CNN Indonesia pada *plattform Twitter* yang membahas tentang kebijakan pemberian subsidi motor listrik, menunjukkan terdapat 868 komentar yang berisi komentar pro dan kontra atas kebijakan tersebut. Dengan banyaknya sentimen yang muncul, maka diperlukan suatu solusi yang dapat mempengaruhi kebijakan pemerintah. Analisis sentimen merupakan salah satu pendekatan yang memungkinkan untuk meredam sentimen masyarakat yang tinggi terhadap suatu kebijakan.

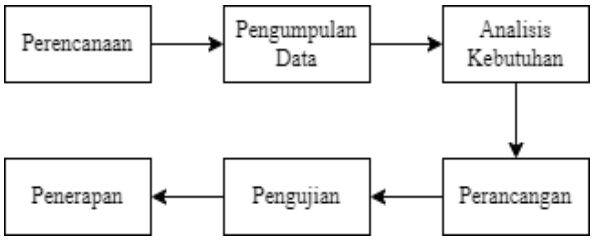
Analisis Sentimen adalah bidang penelitian yang menganalisis opini, penilaian, evaluasi, sikap, dan perasaan masyarakat tentang keseluruhan, seperti produk, layanan, organisasi, individu, isu, peristiwa, atau topik. Analisis ini digunakan untuk mengekstrak informasi spesifik dari dataset yang ada. Analisis sentimen berfokus pada pengolahan opini dengan polaritas, yaitu. pendapat positif atau negatif.[3]

Untuk menganalisis sentimen tersebut diperlukan sebuah metode, salah satu metode yang dapat digunakan ialah metode Support Vector Machine. Metode *Support Vector Machine* menghasilkan tingkat akurasi yang cukup tinggi. Secara keseluruhan, hasil perolehan *FI-score* dengan algoritma *Support Vector Machine* memberikan hasil yang cukup tinggi. SVM memiliki keungulan penting dalam pendekatan teorinya yang dibenarkan atas masalah *overfitting* yang memungkinkannya bekerja dengan baik.[3]

Berdasarkan hal tersebut penulis ingin menganalisis tingkat akurasi dan efektivitas metode *Support Vector Machine* dalam mengolah data sentimen. Tujuan yang ingin dicapai dari penelitian ini untuk mengetahui bagaimana penerapan metode klasifikasi *Support Vector Machine* dalam mengolah data sentimen. Analisis ini nantinya dibuat untuk mengklasifikasikan opini dari media sosial *twitter*. Hasil penelitian ini adalah model klasifikasi untuk analisis sentimen yang diharapkan dapat membantu memberi informasi tentang sentimen yang terdapat pada opini positif atau negatif yang telah diberikan oleh pengguna terhadap kebijakan tersebut.

2. Metode Penelitian

Metodologi penelitian adalah langkah-langkah yang dilakukan untuk membuat dan menyelesaikan penelitian.[5] Berikut adalah diagram metodologi pada penelitian ini.



Gambar 1. Alur Penelitian

2.1. Perencanaan

Perencanaan merupakan tahapan awal serta pedoman papa penelitian ini. Tanpa adanya perencanaan yang baik, penelitian tidak akan berjalan dengan lancar.

2.2. Pengumpulan Data

Pada penelitian ini dilakukan pengumpulan data opini mengenai kebijakan pemberian subsidi motor listrik berupa *tweet* dengan tahapan sebagai berikut:

1. Melakukan *crawling data* menggunakan *library python* yaitu *snscrape*
2. Data yang yang dikumpulkan berupa tweet yang berisi opini masyarakat mengenai pemberian kebijakan subsidi motor listrik pada *plattform twitter* sebanyak 1000 data *tweet.*
3. Pengumpulan *tweet* tersebut dilakukan dengan memasukkan *query* tertentu yaitu #subdisimotorlistrik, #motorlistrik, subsidi motor listrik, dan motor listrik.
4. Data yang sudah diperoleh disimpan dalam bentuk ekstensi .csv.

@ mister_felix14 View 30 November 2022. Setelah semua kendaraan berlistrik, tarif listrik kemudian naik hehehe
@_pln_id View 1 Januari 2023. Langkah ini untuk mendukung program transisi energi bersih guna mencapai target Net Zero Emission (NZE) pada 2060

Gambar 2. Contoh Data *Tweet* Dari Proses *Crawling*

2.3. Analisis Kebutuhan

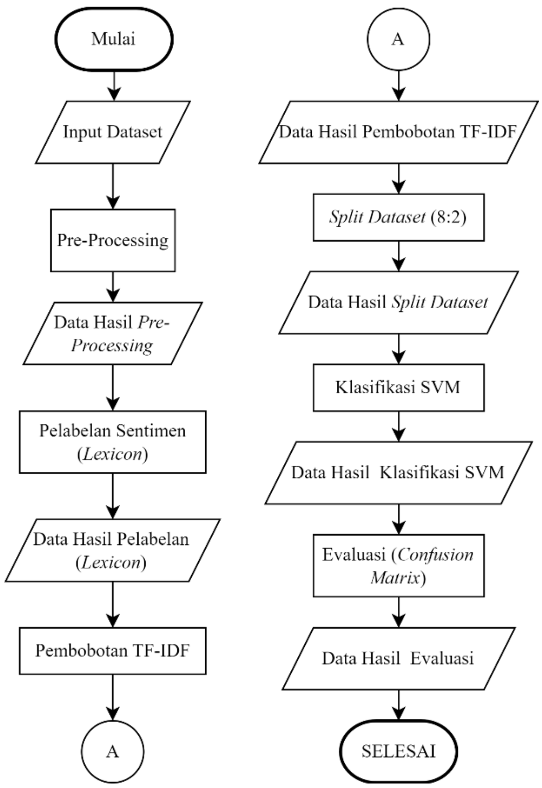
Pada penelitian ini peneliti menggunakan kombinasi pembagian data 80%:20%. Sehingga dari 700 data yang di dapat dibagi 2 menjadi 560 untuk *data training* dan 140 untuk *data testing*. *Data training* digunakan

P-ISSN : 2089-3353
E-ISSN : 2808-9162

untuk membentuk model analisis sentimen yang diberi label secara manual yaitu positif dan negatif sesuai dengan sentimen dari data yang diperoleh. Sedangkan, data testing digunakan sebagai data uji yang akan dilabeli oleh sistem secara otomatis.

2.4. Perancangan

Proses perancangan sistem pada penelitian ini bertujuan agar sistem yang dihasilkan sesuai dengan kebutuhan, dan dapat membantu pada saat penerapan rancangan ini menjadi sebuah sistem sehingga dapat menghemat waktu pembuatan. Proses perancangan dari sistem analisis sentimen yang akan dibangun terdiri dari perancangan *flowchart*. Berikut disajikan *flowchart* dari perancangan sistem ini:



Gambar 3. *Flowchart System* Analisis Umum

2.5. Pengujian

Proses pengujian pada sistem yang dibangun memiliki tujuan untuk memastikan bahwa sistem yang dibangun telah berjalan sebagaimana mestinya. Proses pengujian pada sistem yang dibangun untuk melakukan analisis sentimen masyarakat dilakukan dengan mengumpulkan semua data yang diperlukan, dilanjutkan dengan tahap *preprocessing* lalu disimpan dalam format csv. Setelah itu semua data akan melewati proses pemberian label dengan menggunakan kamus *lexicon*, selanjutnya dilakukan proses klasifikasi pada data tersebut dengan menggunakan metode *Support Vector Machine*.

2.6. Penerapan

Dataset yang berjumlah 1000 data dikumpulkan dari *plattform twitter* menggunakan metode *web scrapping* dengan menggunakan algoritma *phyton snscrape.* Dataset tersebut disaring untuk menghilangkan data *tweet* yang mengandung kata sindiran atau sarkasme ataupun ironi secara manual sehingga tersisa 700 data. Selanjutnya *dataset* akan melalui proses *preprocessing* sehingga menghasilkan *dataset* bersih. Kemudian data hasil *preprocessing* diberi label positif atau negatif menggunakan kamus *lexicon.*

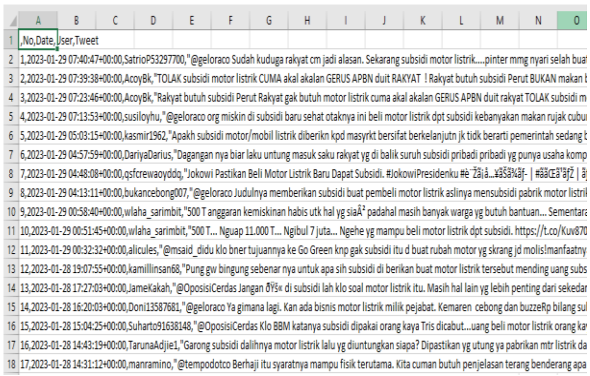
Kemudian data diberi bobot dengan metode TF-IDF. Selanjutnya, sistem membagi *dataset* menjadi *dataset training* dan *dataset testing* dengan rasio 8:2 sehingga *dataset training* berjumlah 560 data dan *dataset testng* berjumlah 140 data. Berikutnya, sistem memproses *dataset training* dan menjadikan *dataset training* sebagai model pembelajaran dalam mengklasifikasi sentimen berbasis kata. Setelah sistem sudah selesai di *training* dilanjutkan dengan sistem memprediksi sentimen data *testing* berbasis kata. Setelah sistem selesai melakukan proses data, selanjutnya sistem menampilkan hasil *preprocessing* yaitu data bersih dan prediksi sentimen tiap kata. Sistem diharapkan memiliki tingkat akurasi prediksi yang tinggi.

3. Hasil dan Pembahasan

Pada bab ini akan disajikan hasil survei analisis opini masyarakat di Twitter mengenai pembiayaan kendaraan listrik roda dua dengan metode *Support Vector Machine* yang dilanjutkan pada bagian yang telah dijelaskan pada bab sebelumnya..

3.1 Pengumpulan Dataset

Pengumpulan *dataset* berupa tanggapan atau opini masyarakat Indonesia terhadap kebijakan pemberian subsidi motor listrik yang berasal dari Twitter. Total dataset yang dikumpulkan 1000 tweet. Proses pengumpulan dataset menggunakan teknik *web scraping* dengan memanfaatkan *library python* yaitu *snscrape*. Dataset yang berhasil dikumpulkan melalui *web scraping* kemudian diunduh dalam bentuk *excel* dengan menggunakan format ".csv". Berikut ini tampilan dataset yang telah didapatkan dari proses *web scraping* dengan menggunakan *library python.*



	A
1	,No,Date,User,Tweet
2	1,2023-01-29 07:40:47+00:00,SatrioP53297700,"@geloraco Sudah kuduga rakyat cm jadi alasan. Sekarang subsidi motor listrik.....pinter mmg nyari selah buat
3	2,2023-01-29 07:39:38+00:00,AcoyBk,"TOLAK subsidi motor listrik CUMA akal akalan GERUS APBN duit RAKYAT ! Rakyat butuh subsidi Perut BUKAN makan
4	3,2023-01-29 07:23:46+00:00,AcoyBk,"Rakyat butuh subsidi Perut Rakyat gak butuh motor listrik cuma akal akalan GERUS APBN duit rakyat TOLAK subsidi m
5	4,2023-01-29 07:13:53+00:00,susiloyhu,"@geloraco org miskin di subsidi baru sehat otaknya ini beli motor listrik dpt subsidi kebanyakan makan rujak cubu
6	5,2023-01-29 05:03:15+00:00,kasmir1962,"Apakh subsidi motor/mobil listrik diberikn kpd masyrkt bersifat berkelanjutn jk tidk berarti pemerintah sedang b
7	6,2023-01-29 04:57:59+00:00,DariyaDarius,"Dagangan nya biar laku untung masuk saku rakyat yg di balik suruh subsidi pribadi pribadi yg punya usaha komp
8	7,2023-01-29 04:48:08+00:00,qsfcrewaoyddq,"Jokowi Pastikan Beli Motor Listrik Baru Dapat Subsidi. #JokowiPresidenku #è¨Žãj å…¥ãŠã¾ãƒ- | #ãäŒã¨ãƒŽ | ã
9	8,2023-01-29 04:13:11+00:00,bukancebong007,"@geloraco Judulnya memberikan subsidi buat pembeli motor listrik aslinya mensubsidi pabrik motor listri
10	9,2023-01-29 00:58:40+00:00,wlaha_sarimbit,"500 T anggaran kemiskinan habis utk hal yg siaÂ² padahal masih banyak warga yg butuh bantuan... Sementara
11	10,2023-01-29 00:51:45+00:00,wlaha_sarimbit,"500 T... Nguap 11.000 T... Ngibul 7 juta... Ngehe yg mampu beli motor listrik dpt subsidi. https://t.co/Kuv870
12	11,2023-01-29 00:32:32+00:00,alicules,"@msaid_didu klo bner tujuannya ke Go Green knp gak subsidi itu d buat rubah motor yg skrang jd molis!manfaatny
13	12,2023-01-28 19:07:55+00:00,kamillinsan68,"Pung gw bingung sebenar nya untuk apa sih subsidi di berikan buat motor listrik tersebut mending uang subs
14	13,2023-01-28 17:27:03+00:00,JameKakah,"@OposisiCerdas Jangan ðŸ˜« di subsidi lah klo soal motor listrik itu. Masih hal lain yg lebih penting dari sekedar
15	14,2023-01-28 16:20:03+00:00,Doni13587681,"@geloraco Ya gimana lagi. Kan ada bisnis motor listrik milik pejabat. Kemaren cebong dan buzzeRp bilang sul
16	15,2023-01-28 15:04:25+00:00,Suharto91638148,"@OposisiCerdas Klo BBM katanya subsidi dipakai orang kaya Tris dicabut...uang beli motor listrik orang ka
17	16,2023-01-28 14:43:19+00:00,TarunaAdjie1,"Garong subsidi dalihnya motor listrik lalu yg diuntungkan siapa? Dipastikan yg utung ya pabrikan mtr listrik da
18	17,2023-01-28 14:31:12+00:00,manramino,"@tempodotco Berhaji itu syaratnya mampu fisik terutama. Kita cuman butuh penjelasan terang benderang apa

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Gambar 4. Hasil *Crawling* Data Dalam Bentuk .csv

Berdasarkan data yang diperoleh dari hasil *crawling* sebanyak 1000 data, digunakan 700 data yang sudah disaring secara manual untuk menghilangkan kalimat sarkasme, sindiran, serta duplikat. Data tersebutlah yang akan digunakan dalam perhitungan analisis sentimen dengan metode *Support Vector Machine* meggunakan *Jupyter notebook.* Berikut contoh data yang sudah disaring dan dipilih untuk penerapan perhitungan manual *Support Vector Machine.*

3.2 *Preprocessing* Data

Pada langkah ini, data yang sebelumnya diambil dari proses pengindeksan mengalami *preprocessing* data, karena data tersebut masih berisi teks tidak terstruktur yang banyak mengandung *noise*.. Oleh karena itu data tersebut perlu dibersihkan dari segala elemen yang tidak diperlukan tersebut. Berikut contoh data yang akan diproses : "@kompascom Saya pingin motor listrik tp mahal Alhamdulilah d subsidi".

a. *Cleaning*

Tahap pertama dalam proses *pre-processing* yaitu *cleaning*, tahap ini bertujuan untuk membersihkan atau menghapus karakter yang tidak perlu dari data *tweet* seperti tanda baca, *numeric, url, username, mention hashtag* dan *retweet.* Hasil *cleaning* : "Saya pingin motor listrik tp mahal Alhamdulilah subsidi".

b. *Case Folding*

Tahapan *case folding* dilakukan untuk mengkonversi atau mengubah huruf kapital ke huruf kecil (*lower case*) pada semua data yang terdapat dalam dokumen. Hasil *Case Folding* : "saya pingin motor listrik tp mahal alhamdulilah subsidi".

c. *Tokenizing*

Langkah *tokenizing* dilakukan untuk mengekstrak *string* atau memecah kalimat menjadi kata untuk mendapatkan nilai kata. Hasil *Tokenizing* : "['saya', 'pingin', 'motor', 'listrik', 'tp', 'mahal', 'alhamdulilah', 'subsidi']".

d. *Normalization*

Pada tahap normalisasi, kata-kata yang disingkat atau tidak baku diubah dan dikoreksi menjadi kata-kata yang mempunyai makna yang sama berdasarkan KBBI, sehingga menjadi informasi yang mudah diolah misalnya "utk" menjadi "untuk", "yg" menjadi "yang" dan sebagainya. Hasil *Normalization* : "saya ingin motor listrik tapi mahal alhamdulillah subsidi".

e. *Stopword Removal*

Tahapan ini untuk menghilangkan kata yang tidak memiliki makna ataupun tidak penting pada data. Pada tahapan ini menerapkan algoritma *stoplistt* atau *stopword* dalam *library sasttrawi*. *Stopword* berisi kata-kata yang biasa dipakai tetapi tidak jelas dan dapat dibuang seperti "dari", "yang", "untuk", "dan", "di" dan sebagainya. Hasil *Stopword Removal* : "['motor', 'listrik', 'mahal', 'alhamdulillah', 'subsidi']".

f. *Stemming*

Tahapan ini berfungsi untuk menghapuskan seluruh kata imbuhan yang terdapat pada data tweet seperti prefix, suffix dan konfix. Hasil *Stemming* : "motor listrik mahal alhamdulillah subsidi".

3.3 Pelabelan

Dataset yang telah diperoleh sebelumnya, selanjutnya akan dilakukan proses pemberian label (kelas) terhadap data tersebut. Pada pelabelan ini data diolah secara otomatis dengan melakukan perhitungan nilai skor berdasarkan kamus *lexicon.* Pelabelan pada penelitian ini akan dibagi dalam dua kelas yaitu kelas positif dan negatif.

Pada penelitian ini penulis akan menggunakan sumber daya kamus *lexicon* yaitu InSet (*Indonesian Sentiment Lexicon*) yang dikembangkan oleh Fajri Koto dan Gemala Y. Rahmaningtyas yang berjudul "InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs". InSet merupakan kamus *lexicon* yang pembobotannya dikerjakan manual oleh dua orang penutur asli Indonesia yang sudah diberikan instruksi yang sama sebelum melakukan penilaian.[6]

3.4 Pembobotan TF-IDF

TF-IDF (*Term Frekuensi-Invers Dokumen Frekuensi*) adalah metode yang digunakan untuk menghitung nilai bobot setiap kata yang diekstraksi. Metode ini digunakan untuk menghitung kata-kata umum dalam pengambilan informasi. Model pembobotan TF-IDF merupakan metode yang mengintegrasikan model frekuensi kata (TF) dan model aliran teks (IDF). Frekuensi kemunculan (TF) adalah metode penghitungan jumlah kemunculan kata-kata yang dianggap umum dan tidak relevan dalam sebuah teks.[7]

Pada tahap ini, sistem melakukan pembobotan pada *term* untuk mengetahui bobot dari setiap kata yang terdapat pada opini masyarakat yang berupa *tweet* mengenai kebijakan pemberian subsidi motor listrik pada *plattform twitter.* Perhitungan pembobotan diperoleh dari jumlah skor perkata pada data yang sudah dipilih. Berikut rumus pembobotan TF-IDF.

$$IDF = \left(\frac{|D|+1}{DF+1}\right) + 1 \quad (1)$$

3.5 *Split Dataset*

Split data merupakan tahapan untuk membagi dataset menjadi dua bagian yang terdiri dari data latih dan data uji dimana komposisi pembagiannya akan lebih banyak jumlah data latih dibandingkan data uji. Dalam penelitian ini akan diterapkan split data dengan perbandingan 8:2 yang artinya 80% dari total dataset

akan digunakan sebagaidata latih dan 20% lainnya akan digunakan sebagai data uji. Presentase perbandingan 8:2 dipilih dikarenakan mendapatkan hasil akurasi tertinggi setelah dilakukan beberapa percobaan dengan nilai perbandingan lainnya.

3.6 Klasifikasi *Support Vector Machine*

Support Vector Machine bisa menjadi strategi pilihan yang membandingkan parameter standar dari sekumpulan nilai diskrit yang disebut himpunan kandidat, dan memilih salah satu yang memiliki presisi klasifikasi paling baik. [8]

Inti dari strategi SVM adalah menentukan hyperplane (garis isolasi antar kelas). Buat pengaturan yang ideal untuk mempercepat siklus dibandingkan dengan strategi biasa. [9]

Adapun tahapannya adalah sebagai berikut.

1. Melakukan perhitungan matriks dengan kernel linear.

$$K(x, y) = x.y \quad (2)$$

Keterangan :
$K(x, y)$ = kernel linear
x = data
y = kelas data

2. Inisialisasi parameter yang digunakan, antara lain λ, γ, C, dan iterasi maksimum..
3. Setelah itu, α=0 diinisialisasi dan *matriks Hessian* dihitung..

$$D_{ij} = y_i y_j\left(K(x_i.x_j\right) + \lambda^2 \quad (3)$$

Untuk *i,j* = 1,...,n

Keterangan :
x_i = data ke-*i*
x_j = data ke-*j*
y_i = kelas data ke-*i*
y_j = kelas data ke-*j*
n = jumlah data
λ = lamda
$\left(K(x_i.x_j\right)$ = fungsi kernel

4. Membuat 3 perhitungan sebagai berikut (sampai batas iterasi)

$$E_i = \sum_{J=1}^{n} \alpha_i \, D_{iJ} \quad (4)$$

Keterangan :
α_i = alfa ke-i
D_{ij} = matriks *Hessian*
E_i = *error rate*

$$\delta\alpha_i = \min\{\max[\,\gamma(1 - E_i), -\alpha_i]\,, C - \alpha_i\} \quad (5)$$

Keterangan :
α_i = alfa ke-i
γ = konstanta gamma
E_i = *error rate*
C = variabel *slack*
γ = *learing rate* =
$max_{\{i\}}D_{ii}$ = nilai maks dari matriks *Hessian*

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (6)$$

Keterangan :
α_i = alfa ke-i
$\delta\alpha_i$ = delta alfa ke-i

5. Kemudian melakukan perhitungan nilai bias yang ditunjukkan persamaan di bawah ini

$$b = -\tfrac{1}{2}[\textstyle\sum_{i=1}^{m} \alpha_i \, y_i K(x_i, x^+) + \sum_{i=1}^{m} \alpha_i \, y_i K(x_i, x^-)] \quad (7)$$

Keterangan :
α_i = alfa ke-i
y_i = kelas datake-i
$K(x_i, x^+)$ = kernel kelas positif
$K(x_i, x^-)$ = kernel kelas negatif

6. Melakukan perhitungan fungsi $f(x)$

$$f(x) = w.x + b \quad (8)$$
$$f(x) = \textstyle\sum_{i=1}^{m} \alpha_i \, y_i K(x_i.x) + b \quad (9)$$

Keterangan :
$\sum_{i=1}^{m} \alpha_i \, y_i$ = jumlah nilai bobot w
$K(x_i.x)$ = nilai kernel
b = nilai bias

7. Menentukan kelas positif atau negatif dengan menggunakan ketentuan sebagai berikut untuk menentukan kelas data uji.

Data akan bernilai positif apabila :

$$w.x + b = 1 \quad (10)$$

Data akan bernilai negatif apabila :

$$w.x + b = -1 \quad (11)$$

Setelah melalui proses diatas didapatkan hasil nilai bias pada klasifikasi penelitian *Support Vector Machine* sebagai berikut :

$$b = -\frac{1}{2}[w.x^+ + w.x^-]$$

$$b = -\tfrac{1}{2}(0{,}11587 - 0{,}093452)$$

$$b = -0{,}10465$$

Maka diperoleh nilai *hyperplane*

$$f(x) = w.x_i + b$$

$$f(x) = 1{,}9572\,.\,0{,}8903 + (-0{,}10465) = 1{,}6378$$

Dan didapatkan :

P-ISSN : 2089-3353
E-ISSN : 2808-9162

$$Fungsi\ klasifikasi = sign(1{,}6378)$$
$$= 1$$

Dapat diketahui bahwa pada fungsi klasifikasi data uji mendapatkan nilai berupa 1 sehingga terklasifikasi sebagai data kelas 1 dimana termasuk dalam kelas **positif**.

3.7 Evaluasi *(Confusion Matrix)*

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. [10] Setelah selesai melakukan proses analisis data, maka diperoleh hasil dari analisis data berupa nilai atau label pada data yang dipilih sebagai data *testing* dengan menggunakan metode *Support Vector Machine* sesuai dengan sistem analisis sentimen yang dibangun berdasarkan proses pembelajaran yang telah dilakukan. Pada penelitian ini, *confusion matrix* digunakan sebagai evaluasi hasil dari sistem yang dibangun. Dari *confusion matrix* dapat dihitung nilai *accuracy*, *precision*, *recall*, dan *f1-score* sebagai hasil dari sistem analisis sentimen dan pengklasifikasian menggunakan metode *Support Vector Machine*. Berikut disajikan tabel *confusion matrix* pada penelitian ini.

Tabel 1. *Confusion Matrix* Hasil Klasifikasi

ij		**Kelas Prediksi (*j*)**	
		Negatif	**Positif**
Kelas Aktual (*i*)	**Negatif**	66	11
	Positif	8	55

Dari Tabel 1 maka dapat dihitung seberapa besar nilai *accuracy*, *precision*, *recall*, dan *f1-score* sebagai berikut:

$$Accuracy = \frac{55+66}{55+8+66+11} \times 100\% = 86.43\%$$

$$Precision = \frac{55}{55+11} \times 100\% = 83.33\%$$

$$Recall = \frac{55}{55+8} \times 100\% = 87.30\%$$

$$f1score = \frac{2 \times 87{,}30 \times 83{,}33}{87{,}30 + 83{,}33} \times 100\% = 85.27\%$$

Dari nilai *confusion matrix* yang diperoleh dari proses klasifikasi yang dapat digunakan untuk menentukan nilai dari *precision*, *recall*, *f1-score* serta untuk mengetahui tingkat akurasi yang dimiliki sistem yang telah dibangun. Secara keseluruhan nilai-nilai diatas dapat disajikan dalam *classification report*. Berikut disajikan *classification report* dari *confusion mastrix* atas pengujian data uji dengan menggunakan metode *Support Vector Machine* :

```
Confusion Matrix:

[[66 11]
 [ 8 55]]


Test Accuracy    : 86.43%
Test Recall      : 87.30%
Test precision   : 83.33%
Test F1-Score    : 85.27%


Classification Report:

              precision    recall  f1-score   support

     Negatif       0.89      0.86      0.87        77
     Positif       0.83      0.87      0.85        63

    accuracy                           0.86       140
   macro avg       0.86      0.87      0.86       140
weighted avg       0.87      0.86      0.86       140
```

Gambar 5 *Classification Report*

Dari hasil perhitungan diatas dapat dilihat bahwa jumlah data uji sebanyak 140 data dan didapat nilai *accuracy* sebesar 86.43%, *precision* sebesar 83.33%, *recall* sebesar 87.30%, dan *f1-score* sebesar 85.27%.

4. Kesimpulan

Hasil dari penelitian ini yaitu Analisis sentimen memiliki jumlah sentimen negatif lebih banyak dibanding dengan sentimen positif yaitu dengan sentimen negatif sebanyak 361 *tweet* dan sentimen positif sebanyak 339 *tweet*, menjelaskan bahwa kebijakan pemerintah masih mendapat pandangan negatif dari masyarakat.

Tingkat akurasi yang dihasilkan oleh metode *Support Vector Machine* dalam melakukan klasifikasi sentimen (opini masyarakat) dapat dikatakan sangat baik. Hal ini dapat dilihat dari tingkat akurasi yang dihasilkan pada sistem dengan menerapkan metode *Support Vector Machine*. Pada penelitian ini digunakan dataset yang diperoleh dengan teknik crawling sebanyak 1000 data. Setelah dilakukan penyaringan secara manual dengan cara menghilangkan kalimat sarkasme, SARA, serta data duplikat maka diperoleh 700 data tweet (opini) dengan penyeragaman kata kunci "subsidi motor listrik", "motor listrik", "#motorlistrik" dan "#subsidimotorlistrik". Dataset yang berjumlah 700 data dengan perbandingan data latih dan data uji sebesar 8:2 yaitu 560 data *training* dan 140 data *testing* diperoleh nilai *accuracy* sebesar 86.43%, *precision* sebesar 83.33%, *recall* sebesar 87.30%, dan *f1-score* sebesar 85.27%.

Daftar Rujukan

[1] Nugraha, F., Harani, N., & Habibi, R. 2020. *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif Industri Nusantara.

[2] Akbar, M. N., Hasanahlmar'iyah Rusydi, N., Hasrul, M., & Ramadhanti, S. 2022. *Sentiment Analysis Terhadap Review Aplikasi Maxim di Google Play Store Menggunakan Support Vector Machine (SVM)*. *2*(2), 1.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

[3] Rahman, A., Indra, A., Alita, D., & Satya, N. 2021. Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM. *JDMSI*, *2*(1), 31–37.

[4] Sopiah, A. 2022. *Subsidi Motor Listrik Rp6,5 Juta Salah Sasaran, Lho Kok Bisa?* CNBC Indonesia. https://www.cnbcindonesia.com/news/20221214071228-4-396648/subsidi-motor-listrik-rp65-juta-salah-sasaran-lho-kok-bisa [Accessed 14 Mei 2023]

[5] Prasetya, Y. N., Winarso, D., & Syahril. 2021. Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twitter Terhadap Isu Covid-19. *Jurnal FASILKOM*, *11*(2), 97–103.

[6] Koto, F., & Rahmaningtyas, G. Y. 2018. Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017, 2018

[7] Luqyana, W. A., Cholissodin, I., & Perdana, R. S. 2018. Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(11), 4704–4713.

[8] Rahmawati, S., Alfa, M., Fitri, N., Rizkie, Y., & Nooraeni, R. 2020. Analisis Sentimen Pengguna Instagram Terhadap Kebijakan Kemdikbud Mengenai Bantuan Kuota Internet dengan Metode Support Vector Machine (SVM). *Jurnal Matematika Dan Statistika Serta Aplikasinya*, *8*(2).

[9] Alvianda, F. 2018. Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM). Universitas Brawijaya.

[10] Widoyoko, L., Silva, D., & Faflillah, H. 2020. Analisis Sentimen E-Commerce Traveloka Pada Twitter Menggunakan Metode Naive Bayes Classifier.