ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN BAHAN BAKAR MINYAK PADA MEDIA SOSIAL YOUTUBE DENGAN METODE K-NEAREST NEIGHBOR DAN SUPPORT VECTOR MACHINE

Muhammad Zaidan Ramadhan, Roy Mubarak
Teknik Informatika, Universitas Mercu Buana
Jl. Menteng Raya No.29, Indonesia
Zaidanramadhan09@gmail.com

ABSTRAK

Penggunaan bahan bakar minyak (BBM) makin pesat. Tingkat kebutuhan akan bahan bakar minyak tersebut menjadi sangat tinggi, dan dapat dikatakan bahwa bbm memiliki kontribusi penting dalam seluruh kegiatan ekonomi. Tak hanya untuk kegiatan ekonomi, bahan bakar minyak dibutuhkan untuk kendaraan pribadi yang setiap harinya digunakan oleh masyarakat. Masyarakat yang kesehariannya menggunakan kendaraan pribadi, sangat khawaitir dengan harga bbm yang tidak stabil. Masyarakat yang tidak puas dengan dengan harga bbm yang tidak stabil bahkan selalu naik ini, memberikan kritikan-kritikan yang menurutnya dengan menaikkan harga bbm tersebut dapat membuat hidupnya sengsara. Kritikan tersebut mengarah kepada rezim kepresidenan era Jokowi Dodo. Salah satu *platform* yang digunakan dalam menuangkan kritikan mereka adalah *platform* Youtube. Untuk memahami pola sentimen yang diungkapkan masyarakat, diperlukan analisis ini untuk selanjutnya digunakan sebagai kajian pemerintah dalam menaikkan harga bbm yang sedang ramai diperbincangkan. Analisis ini menggunakan metode *machine learning*, yaitu *K-Nearest Neighbor* dan *Support Vector Machine.* Data yang digunakan untuk analisis ini sebanyak 1469 data terkait kenaikan bahan bakar minyak. Pada hasil analisis, model menunjukkan keakuratan sebesar 88%, *recall* 88%, *precision* 90% dan *f1-score* 85% untuk model *K-nearest Neighbor* dan untuk model SVM memiliki keakuratan sebesar 89%, *recall 89%,f1-score 84% dan precision 90%*. Hasil untuk pelabelan menunjukkan bahwa sebanyak 920 merupakan komentar negatif dan 123 komentar untuk positif. Analisis tersebut memberikan gambaran yang jelas mengenai opini atau sentimen masyarakat terhadap kenaikan bahan bakar minyak.

Kata kunci : *BBM, K-Nearest Neighbor, Support Vector Machine, Analisis Sentimen, Youtube.*

1. PENDAHULUAN

Bahan bakar minyak (BBM) merupakan bahan baku yang memiliki kontribusi penting dalam seluruh sudut kegiatan ekonomi. Tidak hanya menimbulkan beban yang lebih besar bagi masyarakat umum, tetapi juga berdampak pada sektor usaha [1].

Bahan bakar minyak (BBM) sangat dibutuhkan pada saat ini, apalagi pada saat ini volume kendaraan semakin membesar. Indonesia merupakan negara keempat di dunia dengan jumlah penduduk paling tinggi mencapai lebih dari 200 juta jiwa. Bahkan, penggunaan bahan bakar minyak sudah menjadi bahan pokok bagi para Masyarakat. Tak hanya masyarakat yang sudah bekerja, namun para mahasiswa yang belum bekerja pun juga menggunakan bahan bakar minyak untuk kendaraan yang dipakainya ke kampus sehari-hari. Kenaikan harga bahan bakar minyak ini, membuat para mahasiswa resah. Sebab, ongkos yang mereka punya sudah habis banyak pada saat membeli bahan bakar minyak. BBM merupakan salah satu dari industri minyak dan gas bumi yang memiliki peran vital dalam sektor perekonomian Indonesia yaitu sebagai sumber energi yang menggerakan berbagai sektor kehidupan dan juga sebagai sumber pendapatan APBN negara. Penyebab kenaikan bahan bakar minyak ini, biasanya disebabkan oleh cadangan minyak yang kian menipis, tak hanya itu meningkatnya permintaan serta kurangnya kemampuan negara-negara pemasok dalam meningkatkan pasokan minyak tersebut [2].

Pada Oktober 2023 harga bahan bakar minyak kembali naik lagi sejak diturunkan kembali harga bahan bakar minyak beberapa bulan lalu. di lansir dari laman mypertamina, harga pertamax di Jabodetabek naik sebesar Rp700 perliter dari harga awal Rp13.300 perliter menjadi Rp14.000. Harga pertamax *green* juga naik sebesar Rp1000 dari harga jual 15.000 perliter menjadi Rp1600 perliter. Dari kenaikan harga tersebut, tentu saja menjadi perdebatan oleh masyarakat yang kesehariannya menggunakan kendaraan untuk bekerja. Salah satu aplikasi yang jadi wadah perdebatan yaitu, aplikasi *Youtube*. Banyak Masyarakat yang memberikan komentar tidak puas pada berita kenaikan harga bbm tersebut. *Youtube* merupakan sebuah wadah video terbesar di dunia dengan jumlah pengguna mencapai sebanyak 1,5 miliar pada tahun 2018. Terdapat sebanyak 93,8% pengguna *Youtube* dari jumlah populasi di Indonesia pada tahun 2021, dilansir dari hasil survei yang dilakukan oleh *Hootsuite* (*We Are Social*,2021) tentang *Indonesian Digital Report*. Peningkatan jumlah pengguna berbanding lurus dengan peningkatan konten yang diunggah ke *Youtube*. Saat ini *Youtube* tidak hanya menjadi sarana hiburan bagi penggunanya, tetapi juga sarana memberikan informasi terkini.

Dalam perdebatan yang dituangkan kedalam wadah media sosial *Youtube* tersebut, pasti ada positif dan negatif suatu komentar atau perkataan. Oleh karena itu, dibutuhkan penelitian ini agar dapat mengetahui tren dan pola yang dalam ada pada data tersebut. Penelitian ini melibatkan dua buah metode *machine learning* untuk dibandingkan performanya, yaitu metode *K-Nearest Neighbor* dan *Support Vector Machine* (SVM). K-NN dan SVN ini banyak dipakai dalam penelitian untuk analisis teks, seperti analisis sentimen.

Hasil yang diharapkan pada penelitian ini adalah dapat memberikan kontribusi besar terhadap sektor *data science* dalam hal peningkatan performa dari model *machine learning* yang dibuat untuk menganalisis data mengenai berita kenaikan harga bahan bakar minyak, dan sebagai bahan kajian pemerintah untuk membuat keputusan dalam menaikkan harga bahan bakar minyak yang bisa membuat masyarakat resah.

2. TINJAUAN PUSTAKA

2.1. Youtube

Youtube merupakan sebuah aplikasi yang berisi beberapa video yang diupload oleh pengguna dan dapat pakai oleh berbagai kalangan dari yang muda hingga kalangan yang tua. Pada masa sekarang ini media sosial *Youtube* merupakan jaringan media sosial yang paling sering dikunjungi [3].

Rata-rata pengunjung dalam sehari mencapai jutaan orang untuk mengupload, menonton, dan mencari video pada aplikasi ini [4].

Beriringan dengan berkembangnya zaman, Youtube juga memiliki kekurangan selain memiliki kelebihan, apalagi dengan Y*outube*, semua hal negatif dengan mudah dapat ditayangkan pada media tersebut. Selain itu penggunaan kuota yang boros juga menjadi salah satu kekurangan *Youtube* tersebut [5].

Walaupun begitu, *Youtube* selalu ramai dikunjugi oleh pengguna media sosial dan tidak pernah sedikit pengunjung. Pengguna *Youtube* di Internet bukan hanya untuk mendapatkan hiburan, tetapi juga untuk belajar atau mendapatkan informasi. Informasi di atas menjadi kajian menarik untuk diteliti dan dikembangkan.

2.2. Bahan Bakar Minyak

Bahan bakar minyak (BBM) Adalah sebuah produk utama dari hasil penyulingan Minyak Bumi. Menurut Undang-Undang Nomor 22 Tahun 2001 tentang Minyak dan Gas Bumi pada pasal 1 ayat 4 mengenai Bahan Bakar Minyak (BBM) yakni bahan bakar yang bersumber dari minyak bumi [6].

BBM mempunyai peranan yang sangat penting dalam segala aspek pada sektor ekonomi. BBM ini banyak digunakan oleh Masyarakat untuk keperluan dalam usaha maupun keperluan pribadi [7].

Bahan Bakar Minyak (BBM) ini disalurkan dan dijual oleh PERTAMINA yang terdiri dari beberapa jenis, yaitu Avigas, Avtur, Bensin Super, Bensin Premium,dll.

2.3. Machine Learning

Machine Learning merupakan sebuah cabang ilmu komputer yang mempelajari algoritma yang dapat mempelajari atau beradaptasi pada pola data tanpa melakukan program secara eksplisit [8] .

Machine Learning merupakan sebuah pemikiran yang didasarkan bahwa komputer dapat belajar dari data dan dapat mengidentifikasi suatu pola dalam data, serta memungkinkan untuk membuat indentifikasi pola dalam data, sehingga dapat membuat suatu keputusan tanpa banyak keterlibatan dari manusia.

Menurut Sidik & Ansawarman [9] Terdapat 3 bagian dari *Machine Learning*:

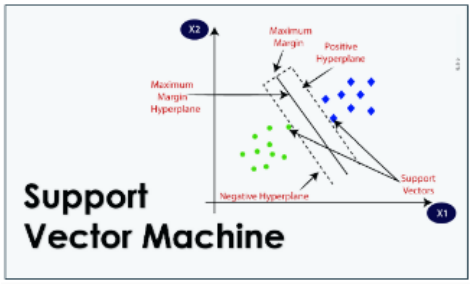
a. *Supervised Learning*, merupakan sebuah model yang proses pembelajarannya menggunakan data yang memiliki label atau jawaban.
b. *Unsupervised Learning*, merupakan sebuah model yang tidak memiliki label atau jawaban pada datanya pada proses pembelajaran, sehingga model ini melakukan pelabelan sendiri atau mengelompokkan data dengan sendirinya.
c. *Reinforcement Learning*, adalah model yang dapat memungkinkan sebuah agent perangkat lunak beroperasi secara mandiri untuk menentukan tindakan optimal dengan tujuan memaksimalkan kinerja sebuah algoritma. Contoh dari model ini adalah pengembangan kecerdasan buatan dalam permainan.

Metode *data mining* yang dikenal sebagai klasifikasi memungkinkan model untuk memprediksi kelas atau kategori objek yang ada dalam *database*. Proses klasifikasi terdiri dari dua tahap yaitu pelatihan dan klasifikasi. Pada tahap pelatihan, algoritma klasifikasi akan membangun model klasifikasi dengan membangun analisis data pelatihan. Selain itu, langkah pembelajaran juga dapat disebut sebagai langkah pelatihan fungsi atau pemetaan Y=F(X), dimana Y adalah kelas yang akan diprediksi dan X adalah kumpulan data yang kelasnya ingin diprediksi. Selanjutnya, langkah klasifikasi menggunakan model yang dihasilkan untuk melakukan klasifikasi. Klasifikasi yaitu merupakan proses menemukan sebuah pola yang menggambarkan dan membedakan kelas data dikenal dengan klasifikasi. Tujuan dari klasifikasi tersebut adalah agar pola-pola ini dapat digunakan untuk memprediksi kelas objek yang kelasnya tidak tidak diketahui [10].

2.4. Support Vector Machine

Support Vector Machine (SVM) merupakan metode yang berguna untuk memprediksi untuk kasus regresi atau klasifikasi. SVM termasuk kedalam *supervised learning* yang pengimplementasiannya dibutuhkan tahapan pelatihan menggunakan squential training SVM dan diikuti proses pengujian metode tersebut [11]. Penggunaan SVM dengan set data latih pada bentuk (xi, yi) *Xi* merupakan tuple dan yi

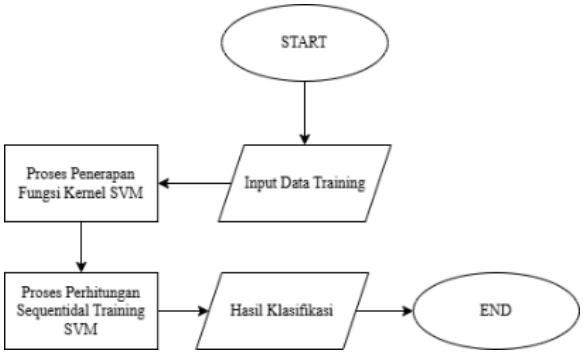
merupakan sebuah label kelas dengan $i = 1 \ldots . N, Xi \in R^d$ dan $yi \in \{-1,1\}$. SVM memiliki tujuan untuk dapat membentuk classifier seperti persamaan dibawah ini.



Gambar 1. Algoritma svm

$$f(xi) = \{\geq 0, yi = +1 \ < 0, yi = -1 \quad (1)$$

Terdapat sebuah *flowchart* yang menggambarkan proses dari algoritma *SVM*.



Gambar 2. Flowchart algoritma svm

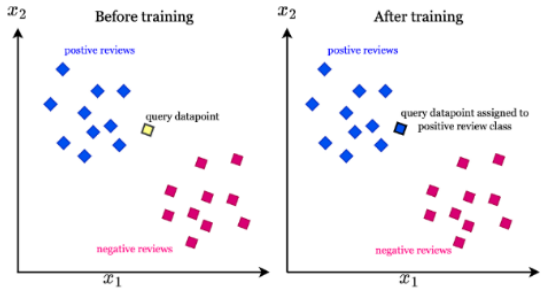
Penjelasan *flowchart* sebagai berikut:

a. Menginput data training
b. Memproses fungsi pada kernel SVM.
c. Memproses perhitungan *sequentidal* pada Training SVM.
d. Keluaran hasil Klasifikasi.

2.5. Algoritma K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* merupakan sebuah algoritma yang memiliki kaitan dengan proses klasifikasi berdasarkan jarak. Data pembelajaran yang didapatkan akan disajikan ke beberapa bagian ruang berdimensi. Ruang ini terpisah dan dijadikan sebagai klasifikasi data rujukan. Kelas c sebagai sebuah titik dalam ruang ini, kelas cadalah klasifikasi paling banyak ditemukan pada k-buah tetangga yang terdekat pada titik tersebut. Algoritma K-NN merupakan sebuah algoritma yang lumayan popular. Metode K-NN merupakan sebuah teknik lazy learning. Artinya, untuk membuat model tidak menggunakan titik data training. Singkatnya, tidak ada fase training pada algoritma K-NN ini, kalaupun ada mungkin juga bisa sangat minim. Semua data training akan digunakan pada tahapan testing. Inilah yang dapat membuat proses training menjadi sangat cepat dan tahap testing menjadi sangat lambat serta cenderung mahal atau menggunakan banyak cost dari sisi waktu dan memori [12].

Algoritma K-NN menjelaskan bahwa sesuatu yang sama akan ada pada jarak yang berdekatan atau bertetangga. Artinya data-data yang cenderung mirip akan dekat satu sama lain.

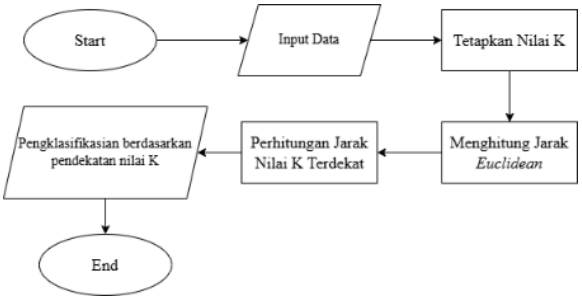


Gambar 3. Algoritma k-nn

Persamaan K-NN Biasanya dihitung menggunakan jarak *Euclidean* menurut legito [13] untuk menentukan kedekatan atau jauhnya lokasi. rumusnya sebagai berikut:

$$d(x,y) = d(x,y) = \ sqrt{(\sum_{i=1}^{N}(x_i - \ y_i)^2)} \quad (2)$$

Terdapat sebuah *flowchart* yang menggambarkan proses dari algoritma *K-Nearest Neighbor*.



Gambar 4. Flowchart algoritma knn

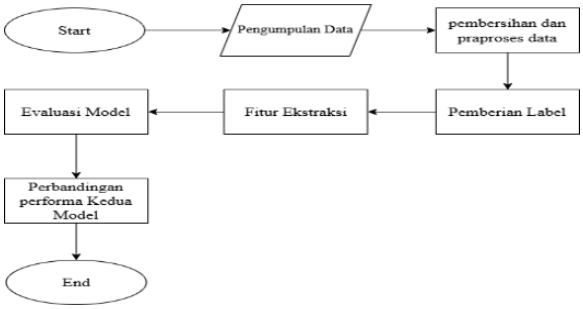
Penjelasan *flowchart* sebagai berikut:

a. Menginput data yang akan dianalisis
b. Menetapan nilai K.
c. Mengitung jarak dari *Euclidean* setelah menginput nilai K.
d. Menghitung jarak nilai K terdekat setelah mendapatkan jarak dari *Euclidean*
e. Mendapatkan hasil pengklasifikasian berdasarkan pendekatan pada nilai K.

3. METODE PENELITIAN

Penelitian ini menggunakan penelitian eksperimen yang memiliki tujuan untuk menganalisis sentimen mengenai kenaikan harga bahan bakar minyak pada media sosial *Youtube* serta membandingkan performa dari kedua model *machine learning* yang akan digunakan. Pada penelitian ini, akan menggunakan model machine learning seperti *K-Nearest Neighbor* dan *Support Vector Machine* untuk

mengklasifikasikan sentimen pada sosial media *Youtube* terkait kenaikan harga bahan bakar minyak. Tahapan penelitian ini meliputi:



Gambar 5. Alur Penelitian

3.1. Pengumpulan Data

Teknik yang cocok untuk mengambil data pada kolom komentar youtube adalah teknik *scraping*. Setelah berhasil mengambil data dengan teknik tersebut, hasil data yang diambil tersebut masih belum bersih, sehingga dibutuhkan suatu metode untuk membuat data tersebut menjadi bersih dan menjadi lebih berstruktur. Pengambilan data menggunakan *API* Youtube yang tersedia melalui *Google Cloud*.

3.2. Pembersihan dan Praproses Data

Dalam pembersihan dan praproses data, kita akan melalui 5 tahapan, sebagai berikut:

a. *Cleaning* atau Pembersihan Data
   Pada Tahapan ini, data akan dibersihkan dari ketidaksesuaian kata yang dapat mempengaruhi kualitas dan integritas data. Dalam kolom komentar yang terdapat emoticon atau karakter spesial, tanda baca seperti titik, koma, tanda tanya dan simbol-simbol seperti lainnya akan dihapus. Dikarenakan, karakter-karakter ini tidak memiliki makna penting dan dapat menghasilkan *noise* yang dapat mengganggu analisis data.

b. *Case Folding*
   Pada tahapan ini, seluruh data akan diseragamkan kedalam model huruf kecil (*lowercase).* Tujuan dari *case folding* ini adalah untuk menyamakan pola huruf yang ada pada teks sehingga menghasilkan konsistensi dalam pengolahan teks

c. *Stopwords*
   Stopwords merupakan sebuah teknik untuk menghilangkan kata-kata yang umum digunakan serta tidak mempunyai makna didalamnya. Contoh kata-kata *stopword* diantaranya adalah "yang", "dan", "di", "tidak" dan sebagainya.

d. *Tokenizing*
   Merupakan sebuah proses dimana teks akan dipecah menjadi unit-unit kecil yang biasa disebut token. Token ini berupa kata, frasa, atau simbol tergantung pada tujuan dari pemrosesan itu sendiri. *Tokenizing* merupakan langkah awal yang penting dalam menganalisis teks untuk konteks *Natural language processing* (NLP).

e. *Stemming*
   Stemming merupakan sebuah tahapan yang dimana setiap kata akan diubah dari kata berimbuhan menjadi kata dasar. Biasanya proses *stemming* menggunakan *library* sastrawi [14].

3.3. Pelabelan Data

Pelabelan dalam analisis sentimen merupakan sebuah pemberian penanda pada setiap kolom komentar dengan kategori tertentu, seperti "positif", "negatif", atau "netral". Proses pelabelan data ini menggunakan *library Lexicon Based class* 3.

3.4. Fitur Ekstraksi

Mengekstraksi fitur teks menggunakan TF-IDF yang berfungsi untuk mengonversi teks menjadi representasi vektor numerik. TF-IDF adalah sebuah fitur pembobotan yang memiliki akurasi serta recall yang tinggi. Pembobotan menggunakan fitur ini sangat penting karena apabila kata lebih sering muncul maka nilai kontribusinya akan semakin besar [15].

TF (*Term Frequency*) merupakan sebuah jumlah kemunculan kata dalam sebuah dokumen, sedangkan IDF (*Inverse Document Frequency*) merupakan sebuah nilai yang dipakai dalam mengukur seberapa pentignya sebuah kata dalam kumpulan dokumen. Semakin besar nilai IDF, maka kata tersebut hanya muncul sedikit pada dokumen, sebaliknya, jika nilai IDF semakin kecil, berarti kata tersebut semakin sering muncul dalam banyak dokumen. Rumus TF sebagai berikut.

$$\text{TF(t,d)} = \frac{\text{Jumlah kemunculan t dalam d}}{\text{Jumlah total kata dalam d}} \quad (3)$$

TF(t,d) = *term frequency* kata t dalam d.
|D| = jumlah total dokumen.
Sehingga rumus TF-IDF adalah sebagai berikut.

$$\text{TFIDF (t,d,D)} = \text{TF(t,d)} \times \text{IDF(t,D)} \quad (4)$$

3.5. Evaluasi Model

Di tahap ini, akan menggunakan data uji untuk menguji performa dari model klasifikasi yang digunakan. Evaluasi model disini akan menggunakan metrik evaluasi seperti akurasi, presisi, recal, *F1-score*, dan *confusion matrix* untuk mengetahui kinerja dari model K-NN dan *Support Vector Machine* (SVM) yang digunakan dalam analsis teks untuk analisis sentimen. Rumus Evaluasi Matriks :

a. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
b. $Precision = \frac{TP}{TP+FP}$
c. $Recall = \frac{TP}{TP+FN}$
d. $F1-Score = \frac{Precision\ .\ Recall}{TN+FP}$

Keterangan:
TP : *True Positive*
TN : *True Negative*
FP : *False Positive*
FN : *False Negatuve*

Accuracy merupakan prediksi yang benar dari total keseluruhan data, *precision* merupakan prediksi positif yang benar dibandingkan dengan semua prediksi positif, *Recall* merupakan proporsi prediksi positif yang benar dibandingkan dengan totak data yang sebenarnya positif, dan *f1-score* merupakan harmonic mean dari *precision* dan *recall.*

3.6. Perbandingan Kedua Algoritma

Pada tahapan ini, setelah model melewati tahap evaluasi, yang dilakukan selanjutnya adalah membuat perbandingan performa yang dihasilkan dari kedua model algoritma tersebut.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dataset yang digunakan untuk penelitian analisis sentimen ini adalah komentar Youtube yang menayangkan berita kenaikan harga bahan bakar minyak sepanjang tahun 2024. Berikut gambar dari hasil pengambilan komentar berformat csv menggunakan teknik scrapping dengan *API Youtube* pada video mengenai kenaikan harga bahan bakar minyak sebanyak 1469 data:

Gambar 6. Dataset

Dataset tersebut nantinya akan digunakan sebagai bahan penelitian untuk mencari tren atau pola sentimen. Dataset tersebut juga digunakan untuk melatih model machine learning yang digunakan, yaitu K-Nearest Neighbor dan SVM. Dengan dataset tersebut, diharapkan dapat memberikan wawasan lebih lanjut mengenai sentimen yang diberikan masyarakat terhadap kenaikan bahan bakar minyak yang terjadi pada periode tersebut.

4.2. Pembersihan dan Praproses

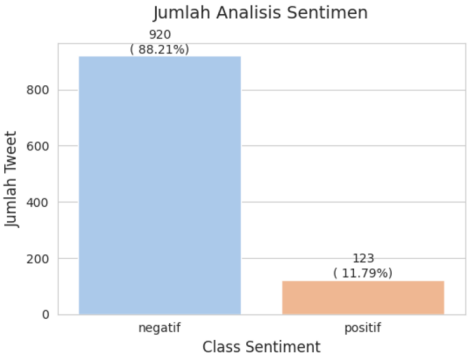
Tahapan ini merupakan tahapan yang vital dalam analisis sentimen sebelum dataset tersebut digunakan pada implementasi pemodelan machine learning tersebut. Pada tahap ini banyak proses yang dilakukan seperti, pembersihan data, *case folding, stopwords, tokenizing,* dan *stemming.* Proses ini dibuat untuk menghilangkan informasi yang tidak memiliki makna atau kurang relavan dengan penelitian, sehingga data tersebut memiliki keakuratan pada penelitian ini. Proses yang dilakukan tahapan ini seperi menghilangkan emoji, simbol-simbol khusus, angka, dan tanda baca. Semua kata akan diseragamkan menjadi huruf kecil atau *lowercase.* Setelah itu, kata-kata yang tidak memiliki makna akan dihilangkan seperti kata penghubung, contohnya “dari”, “dan”,dan lain-lain yang termasuk kata penghubung. Selanjutnya semua kata akan dipecah menjadi sebuah token. Proses pembersihan terakhir yaitu mengubah semua kata menjadi bentuk dasarnya.

Tabel 1. Proses preproses

Tahap	Hasil
Komentar Awal	Di rezim ini Sudah 2x Harga BBM naik.. Harga pangan2 mahal Maka dari itu..Saatnya perubahan Bersama pak Anies Baswedan
Pembersihan Data	Di rezim ini Sudah x Harga BBM naik Harga pangan mahal Maka dari itu Saatnya perubahan Bersama pak Anies Baswedan
Case Folding	di rezim ini sudah x harga bbm naik harga pangan mahal maka dari itu saatnya perubahan bersama pak anies baswedan
Stopwords	rezim harga bbm naik harga pangan mahal perubahan anies baswedan
Tokenizing	['rezim', 'harga', 'bbm', 'naik', 'harga', 'pangan', 'mahal', 'perubahan', 'anies', 'baswedan']
stemming	rezim harga bbm naik harga pangan mahal ubah anies baswedan

4.3. Pelabelan

Pada analisis ini, untuk pelabelan menggunakan bantuan *library Lexicon Based class 3. Library* lexicon akan menentukan kategori suatu kata berdasarkan kamus atau daftar kata yang disediakan pada file *positive.tsv* dan *negative.tsv.* Setelah pelabelan selesai, langkah selanjutnya adalah membuat visualisasi data untuk menampilkan berapa banyak komentar positif dan negatif dari dataset tersebut. berikut adalah visualisasi hasil pelabelan yang telah diberikan :



Gambar 7. Jumlah Label Dataset

Seperti yang ditunjukkan pada gambar visualisasi diatas, untuk data dengan label negatif mendapatkan jumlah sebanyak 920 data, sedangan untuk positif mendapatkan sebanyak 123 data.

4.4. Fitur Ekstraksi

Fitur Eksraksi disini menggunakan metode TF-IDF. TF-IDF digunakan untuk mengubah dataset menjadi representatif numerik.

```
# TF-IDF
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(df['stemmed_text'])
y = df['Sentimen']
```

Gambar 8. Penggunaan tf-idf

4.5. Evaluasi Model

Pada tahap ini, ditampilkan hasil yang telah dilakukan dalam analisis sentimen mengenai kenaikan bahan bakar minyak, sekaligus mengevaluasi hasil prediksi model *K-Nearest Neighbor* dan *Support Vector Machine* menggunakan *accuracy*, *classification report*, dan *confusion matrix. Accuracy* merupakan pemahaman mengenai metrik evaluasi yang mengukur proporsi prediksi dengan predikat benar dari seluruh data yang ada.

Sementara itu, *classification report* merupakan ringkasan dari metrik evaluasi, laporan ini biasanya melipuri *Accuracy, Precision, Recall,* dan *F1-Score.* Dan yang terakhir ada *Confusion Matrix* yang akan memberikan informasi tambahan yang lebih rinci, berikut 4 kategorinya adalah *True Positive, True Negative, False Positive,* dan *False Negative.* Dengan Data Latih 80 dan Data Uji 20 :

1. *K-Nearest Neighbor*

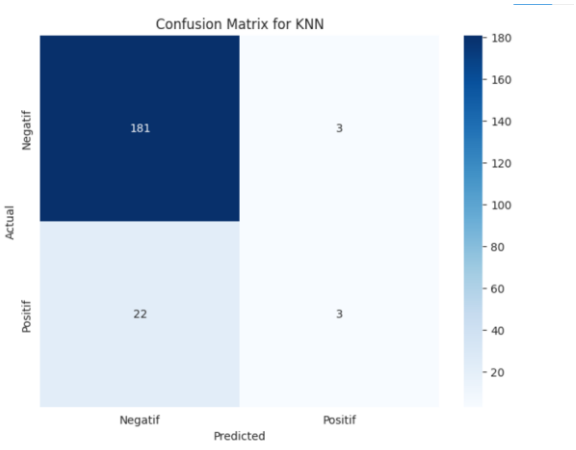
```
K-Nearest Neighbors Classification Report:
              precision    recall  f1-score   support

     negatif       0.89      0.98      0.94       184
     positif       0.50      0.12      0.19        25

    accuracy                           0.88       209
   macro avg       0.70      0.55      0.56       209
weighted avg       0.84      0.88      0.85       209

K-Nearest Neighbors Accuracy: 0.8803827751196173
```

Gambar 9. K-nn report



Gambar 10. Matriks confusion K-NN

Gambar diatas merupakan hasil yang diperoleh dari pengujian model Machine Learning menggunakan algoritma K-Nearest Neighbor, yang dimana data tersebut dibagi menjadi dua bagian, yaitu 80% data latih dan 20% uji. Dapat dilihat dari gambar diatas, diperoleh nilai dari *accuracy* sebesar 88%, untuk *precision* sebesar 84%, *recall* sebesar 88%, dan *f1-score* sebesar 85%. Untuk perhitungan akurasi lebih jelasnya, sebagai berikut :

a. Akurasi model dengan K-Nearest Neighbor = $\frac{Prediksi\ Benar}{Total\ Data}$.
b. Kelas negatif : Recall = 0.98 x 184 = 180.32 (dibulatkan menjadi 180).
c. Kelas positif : Recall = 0.12 x 25 = 3.
d. Total prediksi benar = 180 +3 = 183.
e. Akurasi model dengan K-Nearest Neighbor = $\frac{183}{209}$ = 0.8756 atau 87.56% dibulatkan menjadi 88%.

Berikut merupakan perhitungan untuk total *true positives* (tp) dari masing-masing kelas sesuai dengan laporan yang diberikan diatas :

a. Kelas negatif :
   TP negatif = Recall negatif x Support negatif = 0.98 x 184 = 180, yang berarti nilai untuk TP adalah 180 untuk kelas negatif.
b. Kelas Positif :
   TP positif = Recall positif x Support positif = 0.12 x 25 = 3, yang berarti nilai untuk TP adalah 3 untuk kelas positif.

2. *Support Vector Machine*

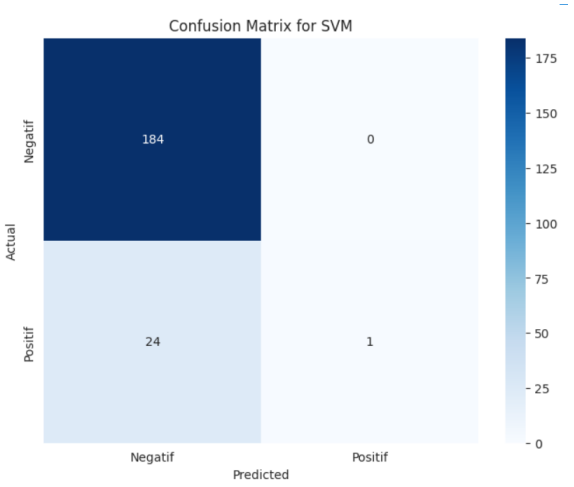
```
SVM Classification Report:
              precision    recall  f1-score   support

     negatif       0.88      1.00      0.94       184
     positif       1.00      0.04      0.08        25

    accuracy                           0.89       209
   macro avg       0.94      0.52      0.51       209
weighted avg       0.90      0.89      0.84       209

SVM Accuracy: 0.8851674641148325
```

Gambar 11. Svm report



Gambar 12. Matriks confusion SVM

Gambar (10) merupakan hasil yang diperoleh dari pengujian algoritma SVM dengan pembagian data

sebanyak 80:20, 80% untuk data latih dan 20% untuk data uji. Dari hasil yang diperoleh, mendapatkan nilai *accuracy* 89%, untuk *precision* sebesar 90%, *recall* sebesar 89%, dan *f1-score* sebesar 84%. . Untuk perhitungan akurasi lebih jelas, sebagai berikut :

a. Akurasi model dengan SVM = $\frac{Prediksi\ Benar}{Total\ Data}$.
b. Kelas negatif : Recall = 1.00 x 184 = 184.
c. Kelas positif : Recall = 0.04 x 25 = 1.
d. Total prediksi benar = 184 +1 = 185.
e. Akurasi model dengan K-Nearest Neighbor = $\frac{185}{209}$ = 0.8852 atau 88.52% dibulatkan menjadi 89%.

Berikut merupakan perhitungan untuk total *true positives* (tp) dari masing-masing kelas sesuai dengan laporan yang diberikan diatas :

a. Kelas negatif :
   TP negatif = Recall negatif x Support negatif = 1.00 x 184 = 184, yang berarti nilai untuk TP adalah 184 untuk kelas negatif.
b. Kelas Positif :
   TP positif = Recall positif x Support positif = 0.04 x 25 = 1, yang berarti nilai untuk TP adalah 1 untuk kelas positif.

4.6. Perbandingan Kedua Model

Berikut merupakan tabel perbandingan dari kedua algoritma yang digunakan untuk analisis sentimen mengenai kenaikan harga bahan bakar minyak pada media sosial *Youtube*.

Tabel 2. Perbandingan Model

	Accuracy	***Precision***	***Recall***	***F1-score***
K-Nearest Neighbor	88%	84%	88%	85%
Support Vector Machine	89%	90%	89%	84%

Dari hasil perbandingan , didapatkan bahwa model dengan metode K-NN mendapatkan hasil akurasi yang lebih rendah sebesar 88% dan *recall* 88% dibandingkan dengan metode SVM sebesar 89% dan *recall* sebesar 89%. Namun pada model yang dihasilkan, untuk model K-NN hanya bisa mendeteksi 180 untuk data yang benar pada kelas negatif, sehingga model lebih baik dalam mendeteksi kelas negati. Sedangkan untuk kelas positif hanya dapat mendeteksi sebesar 3 data, sehingga model kesulitan dalam mendeteksi kelas positif. Sedangkan, untuk metode SVM hanya bisa mendeteksi 184 untuk data yang benar pada kelas negatif, sehingga model lebih baik dalam mendeteksi kelas negatif , sedangkan kelas positif hanya dapat mendeteksi sebesar 1 data saja, oleh karenanya model kesulitan dalam mendeteksi kelas positif. Hal ini dikarenakan data yang tidak seimbang karena lebih banyak dataset yang berlabel negatif dan dataset dengan label positif sangat sedikit , dengan demikian model hanya dapat mendeteksi data mayoritas yaitu negatif pada analisis sentimen ini.

5. KESIMPULAN DAN SARAN

Dari dataset yang dihasilkan dengan melakukan *scraping* pada berita kenaikan harga bahan bakar minyak, diketahui bahwa komentar Negatif masyarakat lebih tinggi daripada komentar Positif terhadap berita tersebut. Untuk komentar Negatif memperoleh sebanyak 920 komentar, untuk positif sebanyak 123 komentar. Hal ini menunjukkan bahwa, banyak masyarakat yang tidak setuju terhadap kenaikan harga bahan bakar minyak yang dinilai memberatkan masyarakat untuk membeli bahan bakar minyak tersebut, dan tidak sedikit pula masyarakat yang mengecam keras pemerintah serta menginginkan agar harga bahan bakar minyak kembali diturunkan.

Penelitian dengan menerapkan model *machine learnig* dengan algoritma K-NN dan SVM mengenai analisis sentimen mengenai kenaikan harga bahan bakar minyak, mendapatkan hasil *accuracy* untuk model K-NN sebesar 88% dan untuk SVM mendapatkan hasil sebesar 89% dengan pembagian data sebanyak 80:20.

Hal tersebut menunjukkan bahwa, untuk model K-NN cenderung lebih banyak dalam memprediksi kelas negatif dengan *recall* 88%, *precision 84%*, dan untuk *f1-score* sebesar 85%. Sedangkan, untuk model SVM, memiliki *accuracy* sebesar 89% dengan pembagian data 80:20. Model SVM juga cenderung lebih banyak dalam memprediksi kelas negatif dengan memiliki *recall* sebesar 89%, untuk *precisison* sebesar 90% dan mendapatkan nilai *f1-score* sebesar *84%*. Untuk penelitian selanjutnya mungkin bisa menggunakan algoritma lain seperti *Naïve Bayes* atau *Logistic Resression* untuk dibandingkan pada model yang ada pada penelitian ini. Dan bisa untuk menggunakan teknik smote untuk menyeimbang data dalam hal meningkatkan kelas minoritas.

DAFTAR PUSTAKA

[1] A. N. Badri, N. Noviandi, F. Anastya, and M. Roland, "SENTIMENT ANALISIS UNTUK IDENTIFIKASI KEPUASAN MASYARAKAT TERHADAP KENAIKAN BBM MENGGUNAKAN ALGORITMA NAÏVE BAYES," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 2, p. 287, Sep. 2023, doi: 10.26798/jiko.v7i2.873.

[2] G. P. Nabila, M. Sultan Mubarok, M. T. Abadi, U. K. H. Abdurrahman, and W. Pekalongan, "DAMPAK KENAIKAN HARGA BAHAN BAKAR MINYAK (BBM) TERHADAP ANGGARAN KELUARGA DI BATANG", Sahmiyya: Jurnal Ekonomi dan Bisnis, pp. 393-401, 2023.

[3] J. Komunikasi, H. Pemikiran, D. Penelitian, and H. Mujianto, "PEMANFAATAN YOUTUBE SEBAGAI MEDIA AJAR DALAM MENINGKATKAN MINAT DAN MOTIVASI BELAJAR," *Jurnal Komunikasi Hasil Pemikiran dan Penelitian*, vol. 5, no. 1, pp. 135–

159, 2019. Available: www.journal.uniga.ac.id135.

[4] S. Suwarto, A. Muzaki, and M. Muhtarom, "Pemanfaatan Media YouTube sebagai Media Pembelajaran pada Siswa Kelas XII MIPA di SMA Negeri 1 Tawangsari," *Media Penelitian Pendidikan : Jurnal Penelitian dalam Bidang Pendidikan dan Pengajaran*, vol. 15, no. 1, pp. 26–30, Jun. 2021, doi: 10.26877/mpp.v15i1.7531.

[5] F. Nur Setiyana and A. B. Kusuma, "EduMatSains Jurnal Pendidikan, Matematika dan Sains POTENSI PEMANFAATAN YOUTUBE DALAM PEMBELAJARAN MATEMATIKA," 2021. [Online]. Available: http://ejournal.uki.ac.id/index.php/edumatsains.

[6] M. A. Sahudi *et al.*, "Penimbunan Bahan Bakar Minyak (BBM) Tanpa Izin yang Disubsidi Pemerintah dalam Persperktif Pidana," Jurnal Litigasi Amsir, vol. 10, no. 2, pp.108-121,2023.

[7] H. Lukman, "Respon Harga Saham Transportasi Darat Terhadap Kenaikan Bahan Bakar Minyak Bersubsidi 2022," vol. 15, no. 1, pp. 91–103, 2023.

[8] R. Kurniawan, P. B. Wintoro, Y. Mulyani, and M. Komarudin, "IMPLEMENTASI ARSITEKTUR XCEPTION PADA MODEL MACHINE LEARNING KLASIFIKASI SAMPAH ANORGANIK," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 2, Apr. 2023, doi: 10.23960/jitet.v11i2.3034.

[9] A. D. Sidik and A. Ansawarman, "Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning," *Formosa Journal of Multidisciplinary Research*, vol. 1, no. 3, 2022, doi: 10.55927/fjmr.v1i3.745.

[10] Heliyanti Susana, "PENERAPAN MODEL KLASIFIKASI METODE NAIVE BAYES TERHADAP PENGGUNAAN AKSES INTERNET," *Jurnal Riset Sistem Informasi dan Teknologi Informasi (JURSISTEKNI)*, vol. 4, no. 1, 2022, doi: 10.52005/jursistekni.v4i1.96.

[11] V. Kevin, S. Que, : Analisis, S. Transportasi, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization)," 2020. [Online]. Available: www.tripadvisor.com.

[12] F. A. Rohmansyah, B. Bintoro, I. Santoso, S. Cipta, K. Informatika, and U. Jakarta, "ANALISIS SENTIMEN TERHADAP PENERAPAN SISTEM GANJIL GENAP MENGGUNAKAN METODE K-NEAREST NEIGHBOR (KNN)." [Online]. Available: www.gataframework.com/textmining.

[13] L. Legito *et al.*, "Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 324–330, Nov. 2023, doi: 10.57152/malcom.v3i2.893.

[14] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," vol. 6, no. 3, 2021, [Online]. Available: http://situs.com

[15] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 422, Apr. 2021, doi: 10.30865/mib.v5i2.2845.