

## ***SENTIMENT ANALYSIS ON THE INCREASE OF FUEL OIL PRICES USING THE K-NEAREST NEIGHBOR ALGORITHM***

### **ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA BAHAN BAKAR MINYAK MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR***

**Chrisley Heltroyce<sup>1</sup>, Gerson Feoh<sup>2\*</sup>, I Made Dwi Ardiada<sup>3</sup>**

<sup>1,2,3</sup>Teknik Informatika, Universitas Dhyana Pura, Bali, Indonesia

(\*) Corresponding Author: [gerson.feoh@undhirabali.ac.id](mailto:gerson.feoh@undhirabali.ac.id)

**Article info**

<b>Keywords:</b>	<b>Abstract</b>
<p><i>Sentiment Analysis, Prices, Fuel Oil, KNN</i></p>	<p><i>Shortly after the implementation of the fuel price increase in early September 2022, the Twitter timeline was flooded with the keywords “BBM naik” (fuel oil is rising); the emergence of many positive and negative opinions to be interesting to study. This study aims to determine the results of sentiment analysis, the results of accuracy precision and recall from data ratio 60:40 and 70:30 using the K-Nearest Neighbor algorithm. This study used the K-Nearest Neighbor (KNN) algorithm because KNN handled training data more accurately. The results of the classification using the KNN algorithm after testing got a percentage of 66.67% for accuracy, 73.02% for negative precision, 48.89% for positive precision, 80.00% for negative recall, and 39.29% for positive recall as a result from a data ratio of 60:40. While at a data ratio of 70:30, the results obtained were 70.31% for accuracy, 76.92% for negative precision, 54.05% for positive precision, 80.46% for negative recall, and 48.78% for positive recall. The results showed that there were 65.7% negative opinions and 34.3% positive opinions, this is because it can harm people who use vehicles in their daily activities.</i></p>
<b>Kata kunci:</b>	<b>Abstrak</b>
<p>Analisis Sentimen, Harga, Bahan Bakar Minyak, KNN</p>	<p>Beberapa saat setelah diberlakukan kenaikan harga BBM pada awal bulan September 2022, <i>timeline</i> Twitter dibanjiri dengan kata kunci BBM naik, timbulnya banyak opini positif dan negatif menjadi menarik untuk diteliti. Penelitian ini bertujuan untuk mengetahui hasil analisis sentiment, hasil <i>accuracy precision</i> dan <i>recall</i> dari perbandingan data 60:40 dan 70:30 menggunakan algoritma <i>K-Nearest Neighbor</i>. Penelitian ini menggunakan algoritma <i>K-Nearest Neighbor</i> (KNN) karena KNN menangani data <i>training</i> dengan lebih akurat. Hasil dari klasifikasi dengan algoritma KNN setelah dilakukan pengujian mendapatkan <i>accuracy</i> 66,67%, <i>precision</i> negatif 73,02%, <i>precision</i> positif 48,89%, <i>recall</i> negatif 80,00% dan <i>recall</i> positif 39,29% dari perbandingan 60:40 sedangkan pada perbandingan 70:30 mendapatkan <i>accuracy</i> 70,31%, <i>precision</i> negatif 76,92%, <i>precision</i> positif 54,05%, <i>recall</i> negatif</p>

80,46%, dan *recall* positif 48,78%. Hasil penelitian menunjukkan terdapat 65,7% opini negatif dan 34,3% opini positif, hal ini dikarenakan dapat merugikan masyarakat yang menggunakan kendaraan dalam kegiatannya sehari-hari.

## PENDAHULUAN

Bahan bakar minyak atau yang dikenal dengan sebutan BBM menjadi salah satu kebutuhan pokok masyarakat, hal ini dilihat pada aktivitas penduduk Indonesia yang banyak menggunakan kendaraan dalam berbagai aktivitas sehari-hari, kenaikan harga BBM sudah tidak asing lagi bagi masyarakat Indonesia mengingat kita menghadapi situasi yang sama beberapa tahun terakhir, pemerintah menerapkan kebijakan perubahan harga sejak tanggal 1 April 2022 selain itu pemerintah juga menetapkan harga BBM pertamax naik dari Rp 9.000-9.400/per liter menjadi Rp 12.500 per liter dan BBM pertalite sebagai BBM subsidi, hal ini juga sejalan dengan kondisi perekonomian di Indonesia yang saat ini masih berada pada ketidakstabilan harga BBM yang terus menerus selama beberapa bulan belakangan ini yang menyebabkan timbulnya berbagai opini publik (Hrp & Aslami, 2022). Penulis tertarik menggunakan algoritma KNN dalam penelitian ini karena menurut Rizki (2019). KNN memiliki kelebihan salah satunya adalah kemampuannya untuk menangani data *Training* dengan lebih akurat yang dapat meningkatkan keakuratan hasil klasifikasi.

Twitter merupakan suatu aplikasi jejaring sosial dan *microblog daring* yang memungkinkan penggunaannya mengirimkan 140 karakter yang disebut dengan *tweet* atau kicauan, aplikasi media sosial Twitter di Indonesia pada Januari 2022 menempati posisi yang cukup tinggi dengan 18,45 juta pengguna dan berada pada posisi kelima negara pengguna Twitter terbesar di dunia (Larasaty, 2022). Media sosial Twitter menjadi tempat masyarakat menyampaikan opininya dan mencurahkan berbagai permasalahan yang dialami dalam kehidupan sehari-hari sehingga menimbulkan berbagai opini, kebijakan pemerintah dalam kenaikan BBM memicu banyaknya kontroversi di media sosial, termasuk pada situs media sosial Twitter, beberapa saat setelah diberlakukan kenaikan harga BBM pada awal bulan September 2022, *timeline* Twitter dibanjiri dengan kata kunci BBM naik, timbulnya banyak opini positif dan negatif menjadi menarik untuk diteliti (Nurhusen et al., 2023). Untuk menentukan hasil positif atau negatif opini pengguna Twitter, penulis memproses opini atau *tweet* ke dalam *RapidMiner*, setelah diproses menggunakan *RapidMiner* akan mengeluarkan hasil analisis sentimen, hasil *accuracy*, *precision*, dan *recall* yang diharapkan dapat menjadi acuan pemerintah untuk mengambil keputusan dalam hal menaikkan harga BBM.

Dalam analisis sentimen yang penulis buat langkah pertama yang dilakukan adalah pengumpulan *dataset* yang berasal dari media sosial Twitter, pengumpulan data menggunakan *tools Jupyter Notebook* untuk melakukan *crawling* data menggunakan kata kunci "bbm naik" pada periode bulan September tahun 2022, dengan limit 10.000 data, menggunakan bahasa Indonesia, setelah mendapatkan 10.000 data *tweet*. Proses dilanjutkan dengan seleksi data dan memberikan label pada data *tweet* berdasarkan emotikon, setelah diberikan label data *tweet* ini akan di *import* ke dalam *RapidMiner* dan masuk pada tahap *Pre-Processing* dalam *RapidMiner*. Tahap berikutnya merupakan klasifikasi KNN yaitu pembuatan data *training* dan data *testing* dimana data *tweet* akan dibagi menjadi 60% dan 70% untuk data *training* dan 40% dan 30% untuk data *testing* dan kemudian data *training* dan *testing* ini akan di *import* ke dalam *RapidMiner* untuk proses klasifikasi dengan metode KNN yang akan menghasilkan *Prediction* positif dan

*Prediction* negatif. Proses selanjutnya pengujian data *training* dan *testing* yang telah dibuat sebelumnya, data akan di *split* dalam proses pengujian ini dimana pengujian pertama menggunakan *split* data 60:40 dengan 60% data *training* dan 40% data *testing* dari data *tweet* keseluruhan kemudian pengujian kedua menggunakan *split* data 70:30 dengan 70% data *training* dan 30% data *testing* dari data *tweet* keseluruhan dan akan keluar hasil *confusion matrix* dimana pada proses ini menghasilkan *accuracy*, *precision* dan *recall*.

Sehingga tujuan penelitian sebagai berikut 1. Untuk mengetahui hasil analisis sentimen terhadap kenaikan harga BBM di media sosial Twitter menggunakan algoritma *K-Nearest Neighbor*. 2. Untuk menghitung tingkat *accuracy*, *precision* dan *recall* dari perbandingan data 60:40 dan 70:30 dengan menggunakan algoritma KNN terhadap analisis sentimen kenaikan harga BBM berdasarkan opini media sosial Twitter.

## METODE

Penelitian ini dimulai dengan mengumpulkan data atau *crawling* data dari Twitter menggunakan kata kunci “bbm naik” pada tahap berikutnya, seleksi data yaitu memisahkan data yang akan digunakan dan tidak digunakan kemudian penulis memberikan label berdasarkan emotikon pada *dataset* dan melanjutkan ke tahapan *Pre-Processing* dan TF-IDF yang bertujuan memberikan bobot pada setiap kata dalam dokumen, pada tahap *Pre-Processing* terdapat lima proses, yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Tujuan dari lima tahap tersebut untuk mengubah dokumen mentah menjadi dokumen representatif yang siap untuk diproses, tahap berikutnya klasifikasi yaitu proses mengklasifikasikan data menggunakan algoritma KNN. Pada tahap terakhir, pengujian dilakukan untuk mendapatkan hasil *accuracy*, *presisi*, dan *recall* dengan menggunakan *confusion matrix*.

## HASIL DAN PEMBAHASAN

Berdasarkan hasil dari pengujian pertama menggunakan perbandingan 60:40 dan perbandingan kedua menggunakan perbandingan 70:30 diketahui :

### Pengujian Dengan Data 60:40

Hasil dari pengujian perbandingan 60:40 dengan data *training* berjumlah 257 dan data *testing* berjumlah 171 dari total data 428

**accuracy: 66.67%**

	true negatif	true positif	class precision
pred. negatif	92	34	73.02%
pred. positif	23	22	48.89%
class recall	80.00%	39.29%	

**Gambar 1. Hasil Confusion Matrix Dengan Data Testing 40%**

Hasil dari *confusion matrix* dengan *split* data 60:40 sebanyak 171 data *testing* meghasilkan *accuracy* 66,67%, *precision* negatif 73,02%, *precision* positif 48,89%, *recall* negatif 80,00% dan *recall* positif 39,29%.

### Pengujian Dengan Data 70:30

Hasil dari pengujian 70:30 dengan berjumlah 300 dan data *testing* berjumlah 128 dari total data 428

accuracy: 70.31%

	true positif	true negatif	class precision
pred. positif	20	17	54.05%
pred. negatif	21	70	76.92%
class recall	48.78%	80.46%	

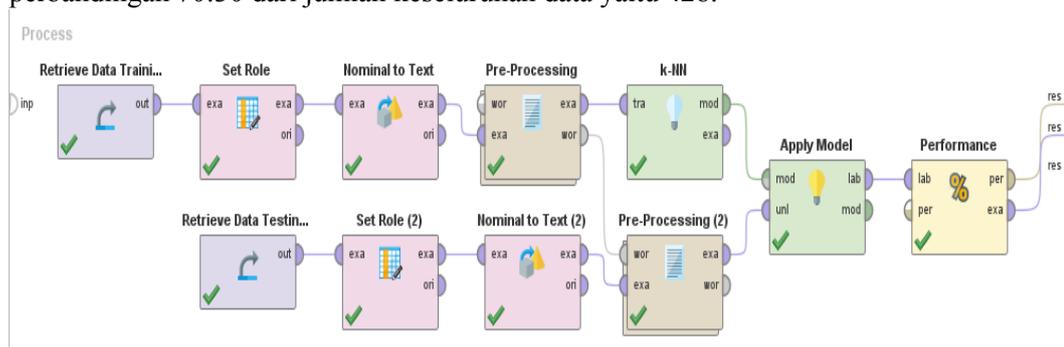
**Gambar 2. Hasil Confusion Matrix Dengan Data Testing 30%**

Hasil dari *confusion matrix* dengan *split* data 70:30 sebanyak 128 data *testing* menghasilkan *accuracy* 70,31%, *precision* negatif 76,92%, *precision* positif 54,05%, *recall* negatif 80,46%, dan *recall* positif 48,78%.

### Pembahasan

Pada penelitian ini pengumpulan data dilakukan dengan cara *crawling* data dengan kata kunci “*bbm naik*”, menggunakan bahasa Indonesia pada tanggal 1 September 2022 hingga 30 September 2022 yang mendapatkan 10.000 data *tweet*, dilanjutkan pada tahapan seleksi data Twitter yang dilakukan sebanyak dua kali yang menghasilkan 428 data *tweet*, 428 data tersebut akan diberi diberikan label secara manual berdasarkan emotikan dan disimpan dalam format *file* CSV, *file* ini sudah siap untuk diproses dalam *RapidMiner*. 428 data *tweet* tersebut di *import* ke dalam *RapidMiner* dan dilanjutkan pada tahapan *Pre-Processing*. Pada tahap *Pre-Processing* terbagi atas lima, yang pertama *cleaning* data *tweet*, guna menghilangkan URL, *mentions*, simbol, RT dan *hashtag*, kedua *case folding*, tahapan *case folding* pada *Pre-Processing* guna menyamaratakan huruf menjadi *lowercase*, dilanjutkan pada tahap ketiga yaitu *tokenizing*, guna untuk merubah kalimat menjadi kata per kata, yang keempat *stopword*, guna untuk menghapus kata yang terdapat dalam kamus *stopword*, dalam penelitian ini jumlah kata dalam kamus *stopword* berjumlah 1.739 kata, lalu tahap yang terakhir adalah *stemming* yang berfungsi untuk mengubah kata berimbuhan menjadi kata dasar, jumlah kata dasar yang digunakan pada penelitian ini berjumlah 452 kata.

Klasifikasi *K-Nearest Neighbors* merupakan tahapan selanjutnya, dilakukan pembagian data menjadi *training* dan *testing* dengan perbandingan 60:40 dan perbandingan 70:30 dari jumlah keseluruhan data yaitu 428.



**Gambar 3. Tahapan Klasifikasi KNN Pada RapidMiner**

Tahapan klasifikasi KNN pada data *testing* dengan menggunakan *RapidMiner* dimana proses awalnya dilakukan dengan menarik data *training* dan data *testing* yang sudah di *import* sebelumnya, dihubungkan ke operator *set role*, operator ini berguna untuk menentukan atribut sentimen dan target *role* sebagai label yang terdapat di dalamnya, selanjutnya dihubungkan ke operator *nominal to text* untuk kedua *set* data, operator ini digunakan untuk mengubah nilai nominal menjadi teks, berikutnya

dihubungkan ke operator *Pre-Processing* dimana pada proses ini terdapat empat tahap yang saling berkaitan yaitu, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* yang sudah dibuat sebelumnya, setelah proses *Pre-Processing* selesai data *training* dihubungkan dengan metode KNN dan tahap selanjutnya KNN dihubungkan dengan operator *apply model* dan *Pre-Processing* pada data *testing* dihubungkan dengan operator *apply model*, yang terakhir *apply model* dihubungkan dengan operator *performance* untuk menampilkan hasil klasifikasi knn dan hasil *accuracy*, *precision* dan *recall* dari data *testing*.

Hasil dari klasifikasi KNN dengan data *testing* 40% yang berjumlah 171 data hasil dari label sentimen positif yang diberi label secara manual berjumlah 56 data sedangkan yang negatif berjumlah 151 data. Data dengan prediksi sentimen KNN menggunakan K2 menghasilkan data positif sebanyak 45 sedangkan data dengan prediksi negatif berjumlah 126. Hasil dari klasifikasi KNN dengan data *testing* 30% yang berjumlah 128 data, Hasil dari label sentimen positif yang diberi label secara manual berjumlah 41 data sedangkan yang negatif berjumlah 87 data. Data dengan prediksi sentimen KNN menggunakan K2 menghasilkan data positif sebanyak 37 sedangkan data dengan prediksi negatif berjumlah 91. Tahapan selanjutnya yaitu pengujian antara *Labeling* manual dengan hasil prediksi sentimen dari metode KNN dimana pada tahap ini akan menghitung banyaknya data *testing* yang sentimennya benar (*True* positif atau *True* negatif) dengan data *testing* hasil prediksi (*Prediction* positif atau *Prediction* negatif) dari hasil tersebut bisa menghasilkan *accuracy*, *precision*, dan *recall* dan tahap ini akan dilanjutkan pada tahapan pengujian. Pada *RapidMiner* pengujian *confusion matrix* menggunakan operator *performance* untuk menampilkan hasil klasifikasi KNN dan hasil *accuracy*, *precision* dan *recall*.

## SIMPULAN

Berdasarkan analisis sentimen yang telah dilakukan dengan topik kenaikan harga BBM pada media sosial Twitter menggunakan algoritma KNN dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan hasil analisis sentimen pengguna Twitter terhadap kenaikan BBM menggunakan algoritma KNN diketahui bahwa banyaknya penolakan terhadap kenaikan BBM tersebut, dikarenakan dapat merugikan masyarakat yang menggunakan kendaraan dalam kegiatannya sehari-hari, namun disisi lain beberapa opini pengguna Twitter tidak merasa terbebani dengan kenaikan BBM tersebut. Hal ini juga didasarkan pada hasil klasifikasi menunjukkan sentimen negatif lebih mendominasi dari sentimen positif, yaitu terdapat 65,7% opini negatif dan 34,3% opini positif.
2. Berdasarkan perhitungan tingkat *accuracy* yang dilakukan, diketahui bahwa hasil dari algoritma KNN, pada penelitian ini dengan perbandingan 60:40 menggunakan proses *split* 60% pada data *training* yang berjumlah 257 dan 40% pada data *testing* berjumlah 171 dari total data 428 menghasilkan hasil *accuracy* 66,67% *accuracy* ini menjelaskan tentang seberapa akurat klasifikasi KNN, *precision* negatif 73,02%, *precision* positif 48,89% *precision* negatif dan positif ini menjelaskan akurasi antar data yang diminta dengan hasil dari prediksi KNN, *recall* negatif 80,00% *recall* positif 39,29% *recall* menjelaskan tentang keberhasilan model, sedangkan dengan perbandingan 70:30 hasil dari proses *split*

70% pada data *training* yaitu berjumlah 300 dan 30% pada data *testing* berjumlah 128 dari total data 428 menghasilkan hasil *accuracy* 70,31% *accuracy* ini menjelaskan tentang seberapa akurat klasifikasi KNN, *precision* negatif 76,92%, *precision* positif 54,05% *precision* negatif dan positif ini menjelaskan akurasi antar data yang diminta dengan hasil dari prediksi KNN, *recall* negatif 80,46% *recall* positif 48,78% *recall* menjelaskan tentang keberhasilan model. Hasil dari dua pengujian diatas dimana data 70:30 mengalami peningkatan pada *accuracy*, *precision* negatif, *precision* positif, *recall* negatif, dan *recall* positif hal ini dikarenakan adanya penambahan pada data *Training* sebanyak 43.

#### DAFTAR PUSTAKA

- Hrp, G. R., & Aslami, N. (2022). Analisis Dampak Kebijakan Perubahan Publik Harga BBM terhadap Perekonomian Rakyat Indonesia. *Jurnal Ilmu Komputer, Ekonomi Dan Manajemen (JIKEM)*, 2(1), 1464–1474. <https://ummaspul.e-journal.id/JKM/article/view/3601/1261>
- Larasaty, D. (2022). Analisis Freedom of Speech di Media Sosial Twitter dengan Kaitannya terhadap Adab Generasi Muda Islam dalam Berinteraksi di Media Sosial. *Al-I'lam: Jurnal Komunikasi Dan Penyiaran Islam*, 6(1), 27. <https://doi.org/10.31764/jail.v6i1.11141>
- Nurhusen, M. R., Indra, J., & Baihaqi, K. A. (2023). Analisis Sentimen Pengguna Twitter Terhadap Kenaikan Harga Bahan Bakar Minyak (BBM) Menggunakan Metode Logistic Regression. *Jurnal Media Informatika Budidarma*, 7, 276–282. <https://doi.org/10.30865/mib.v7i1.5491>
- Rizki, M. M. (2019). Analisis sentimen terhadap produk otomotif dari twitter menggunakan kombinasi algoritma k-nearest neighbor dan pendekatan lexicon (studi kasus: mobil toyota).