

ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN HARGA BBM DI MEDIA SOSIAL TWITTER MENGUNAKAN METODE *SUPPORT VECTOR MACHINE*

Muazim Rahman Fauzan, Harma Oktafia Lingga Wijaya, Satrianansyah,
Joni Karman

*Program Studi Sistem Informasi, Universitas Bina Insan, Lubuklinggau

Correspondence Author: mrahmanfauzan9@gmail.com¹, harmoaktafialingga@gmail.com²,
Satrianansyah@univbinainsan.ac.id³, joni_karman@univbinainsan.ac.id⁴

Abstract

Melalui media sosial twitter, masyarakat bisa berbagi cuitan tentang kesehariannya, berbagi foto, berita terbaru, cuaca, politik, ataupun menyampaikan pendapat tentang suatu hal yang sedang trending topic. Salah satu isu yang sedang trending topic di twitter pada bulan september 2022 di indonesia adalah harga BBM naik. Indonesia secara resmi menaikkan harga BBM dikarenakan meroketnya harga minyak dunia. Sehingga pemerintah tidak dapat lagi menjual harga BBM yang sama seperti sebelumnya kepada masyarakat, hal ini dilakukan pemerintah untuk mengurangi pengeluaran negara yang semakin membesar. Penelitian dilakukan untuk mengetahui sentimen masyarakat terhadap kenaikan harga BBM pada tweets pengguna media sosial twitter menggunakan metode SVM. Data yang di ambil adalah 5000 data tweets yang menggunakan kata kunci "harga bbm naik". Pada tahap preprocessing dari 5000 data, tersisa menjadi 4209 data. Lebeling data dilakukan otomatis menggunakan library TextBlob. Hasil labelling data menggunakan library TextBlob dibagi menjadi 3 sentimen yaitu sentimen positif sebanyak 1275, sentimen negatif sebanyak 1287, dan sentimen netral sebanyak 1647. Klasifikasi dan evaluasi dilakukan menggunakan metode SVM. Berdasarkan nilai perbandingan data training dan testing sebesar 9:1, 8:2, dan 7:3 diperoleh hasil yang paling tinggi menggunakan metode SVM adalah perbandingan data training dan testing 9:1 dengan tingkat akurasi sebesar 76%.

Keyword: Analisis Sentimen, harga BBM, *Support Vector Machine*, *Twitter*.

1. PENDAHULUAN

Melalui *twitter*, pengguna bisa berbagi cuitan tentang kesehariannya, berbagi foto, berita terbaru, cuaca, politik, ataupun menyampaikan pendapat tentang suatu permasalahan yang sedang trending topic.[1] Salah satu isu yang sedang trending topic di *twitter* pada bulan september 2022 adalah tentang harga BBM naik di indonesia.

Indonesia secara resmi menaikkan harga BBM dikarenakan meroketnya harga minyak dunia. Sehingga pemerintah tidak dapat lagi menjual harga BBM yang sama seperti sebelumnya kepada masyarakat, hal ini dilakukan pemerintah untuk mengurangi pengeluaran negara yang semakin membesar.[2] Harga BBM yang naik akan memberikan pengaruh yang besar bagi masyarakat Indonesia, seperti harga barang ikut menaik, manainya harga jasa, berkurangnya produksi barang, menurunkan minat daya beli masyarakat dan meningkatkan tarif transportasi.[2] Akibatnya banyak terjadi demo yang dilakukan masyarakat karena menolak kebijakan pemerintah tentang harga BBM yang naik. Dari permasalahan itu, penulis ingin untuk melakukan penelitian terkait analisis sentimen dari *Tweets* pengguna media sosial *Twitter* mengenai kebijakan pemerintah indonesia tentang harga BBM yang naik. Penelitian ini dilakukan untuk mengetahui opini masyarakat mengenai kebijakan pemerintah indonesia tentang harga BBM yang naik pada bulan september 2022.

Analisis sentimen merupakan suatu analisis yang memiliki tujuan guna menentukan isi dari suatu data yang berbentuk teks untuk menentukan apakah data tersebut termasuk dalam sentimen netral, positif, atau negatif.[3] Analisis sentimen memerlukan metode untuk mengklasifikasi teks, pada penelitian ini penulis memilih metode *Support Vector Machine* (SVM). SVM merupakan salah satu metode yang sangat baik untuk mengklasifikasikan data dalam jumlah besar, terutama untuk analisis sentimen.[4] Oleh karena itu, penulis memilih menggunakan metode SVM dalam penelitian ini.

Penelitian yang terkait analisis sentimen yaitu penelitian yang dilakukan oleh Anna Baita dkk, pada tahun 2021 tentang "Analisis sentimen mengenai vaksin sinovac menggunakan *Algoritma Support Vector*

Mechine (SVM) dan *K-Nearest Neighbor* (KNN). Hasil penelitian yang dilakukan oleh Anna Baita dkk, menggunakan metode SVM tersebut memiliki tingkat akurasi mencapai 70% dan akurasi metode KNN sebesar 56% pada proses pengujian menggunakan *Confusion Matrix*. [5] Penelitian lain yang membahas tentang analisis sentimen adalah penelitian yang dilakukan oleh Lusiana Lestari, pada tahun 2018 tentang “Analisis sentimen Twitter dengan menggunakan *Algoritma Support Vector Machine* (SVM). Hasil penelitian yang dilakukan oleh Lusiana Lestari, menggunakan metode SVM menghasilkan akurasi sebesar 73,6% pada proses pengujian menggunakan *Confusion Matrix*. [6]

Berdasarkan kondisi dan ide yang ada penulis ingin untuk melakukan penelitian untuk tugas akhir dengan judul penelitian yaitu “Analisis Sentimen Masyarakat Terhadap Kenaikan Harga BBM di Media Sosial Twitter Menggunakan Metode *Support Vector Machine*”.

2. METODE PENELITIAN

Dalam penelitian ini, penulis menggunakan beberapa metode untuk memperoleh data dan informasi yang digunakan untuk memecahkan masalah yang ada. Berikut adalah gambar yang menjelaskan tahapan dalam penelitian yang telah penulis lakukan ini untuk menyelesaikan permasalahan yang ada.



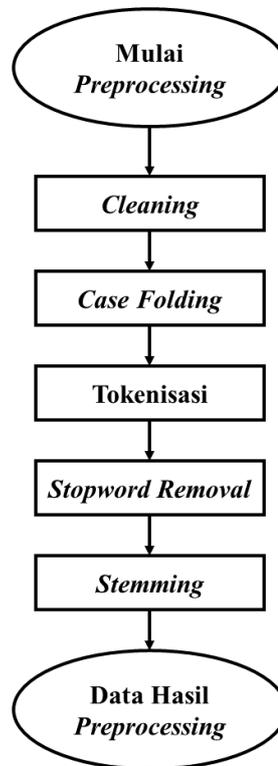
Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Proses pengumpulan data dilakukan menggunakan bantuan dari *library* yang ada pada bahasa pemrograman *python* yaitu *snsrape*. [7] Dalam penelitian ini data yang penulis peroleh yaitu data dari *tweets* pengguna *twitter*. Data yang penulis dapatkan merupakan *tweets* yang mencantumkan kata kunci tentang trending topic yang terjadi di Indonesia atas kebijakan pemerintah yang menaikkan harga BBM pada September 2022. Dalam penelitian ini kata kunci pencarian yang penulis gunakan untuk mengambil data *tweets* pengguna *twitter* adalah “harga BBM naik”. Data yang berhasil dikumpulkan penulis sebanyak 5000 data *tweets* dari bulan September 2022 - Oktober 2022.

2.2. Preprocessing

Tahap *Preprocessing* dilakukan menggunakan bahasa pemrograman *Python*, adapun tahapan-tahapan dalam *Preprocessing* dapat dilihat pada Gambar 2 berikut ini :



Gambar 2. Tahapan *Preprocessing*

Pada gambar 2 di atas menjelaskan tahapan dalam *preprocessing* yang digunakan untuk mengubah data yang penulis peroleh menjadi sesuai dengan format yang penulis butuhkan menggunakan bahasa pemrograman *python*. [8] Berikut penjelasan mengenai tahapan tahapan dalam *preprocessing*.

- a) *Cleaning*
Cleaning adalah tahapan yang digunakan untuk menghapus kata yang tidak berguna dari data *tweets*, hal ini dilakukan untuk mengurangi kesalahan. Kata yang dihapus dalam data *tweets* yaitu HTML, emoji, hashtag (#), username (@username), angka (1,2...), url (<https://situs.com>) dan lain lain. [8]
- b) *Case Folding*
Case folding adalah tahapan ini melakukan penyamaan *case* dari data *tweets*. Data *tweets* yang kita dapatkan tidak semua teks nya konsisten dengan huruf kapital. Oleh karena itu dibutuhkan proses *case folding* untuk mengkonversikan data *tweets* menjadi bentuk *lowercase* (huruf kecil). [9]
- c) *Tokenisasi*
Tokenisasi adalah tahapan yang digunakan untuk memisahkan deretan kata yang ada di dalam data *tweets* menjadi kata tunggal. Pada proses tokenisasi kita juga membuang karakter atau tanda baca yang ada pada data *tweets*. Proses tokenisasi ini menghasilkan kumpulan kata saja. [9]
- d) *Stopword Removal*
Stopword Removal adalah tahapan yang digunakan untuk membersihkan kata yang tidak berguna dari tokenisasi. Pada penelitian ini penulis menggunakan *stopword* bahasa Indonesia yang di dapatkan dari *library* NLTK yang terdapat pada bahasa pemrograman *Python*. [9]
- e) *Stemming*
Stemming adalah tahapan yang digunakan untuk menghapus awalan dan akhiran yang terdapat pada kata kata dasarnya. Proses *Stemming* dilakukan menggunakan bahasa pemrograman *python* di google colab dan *library* yang penulis gunakan yaitu *sastrawi*. [4]

2.3. Labelling Data

Labelling data dilakukan menggunakan *library TextBlob* yang ada pada bahasa pemrograman *python*. [7] *TextBlob* hanya bisa membaca data yang berbahasa *Inggris*, oleh karena itu data harus di *translate*

ke bahasa *Inggris* terlebih dahulu. *TextBlob* akan menghitung nilai dari *polarity* dan *subjectivity*. *Polarity* adalah fungsi yang ada dalam *TextBlob* yang digunakan untuk melihat kecondongan sentimen sebuah teks. Sedangkan *subjectivity* merupakan fungsi yang ada dalam *TextBlob* yang digunakan untuk melihat value dari sebuah teks.[5] Pada penelitian ini kita akan mencari nilai *polarity* menggunakan *TextBlob*, karena semakin tinggi nilai *polarity* itu menandakan teks tersebut bersentimen positif.

2.4. Visualisasi Data

Visualisasi data merupakan tahap memvisualisasikan data yang sudah di label agar lebih mudah dibaca, serta membuat tampilannya lebih menarik. Pada penelitian ini penulis menggunakan *library matplotlib* yang terdapat pada bahasa pemrograman *python* untuk melakukan visualisasi data kedalam bentuk *pie chart*.[7]

2.5. Klasifikasi dan Evaluasi

Metode SVM merupakan salah satu metode yang bagus digunakan untuk melakukan prediksi, baik itu kasus klasifikasi maupun kasus regresi.[10] Langkah awal proses klasifikasi dan evaluasi metode SVM adalah memisahkan data menjadi *training* dan *testing*. Pada penelitian ini penulis menggunakan tiga nilai perbandingan dari data *training* dan *testing*, yaitu:

- Training* 90% dan *testing* 10%
- Training* 80% dan *testing* 20%
- Training* 70% dan *testing* 30%

Setelah dilakukan pembagian data *training* dan data *testing* maka penulis mengubah data tersebut menjadi bentuk data vector. Pengubahan data kedalam bentuk vector ini dilakukan dengan *library CountVectorizer* yang terdapat di bahasa pemrograman *Python*. *Library CountVectorizer* akan menghitung frekuensi kata dalam dokumen yang kemudian akan mengubah data tersebut menjadi representasi vector.[11] Setelah itu metode *Support Vector Machine* akan melakukan pelatihan dengan cara mendeteksi ciri-ciri data pada tiap sentimen yang dibentuk pada data training. Hasil pelatihan pada metode *Support Vector Machine* kemudian di evaluasi menggunakan data testing. Evaluasi dilakukan menggunakan *library* yang terdapat pada metode *Support Vector Machine* yaitu *confusion matrix*.[9] Berikut adalah tabel *confusion matrix*.

Tabel 1. *Confusion Matrix*

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Berdasarkan nilai dari *confusion matrix* tersebut, maka akan diperoleh nilai akurasi, presisi dan *recall*.

a) Akurasi

Akurasi adalah nilai yang menggambarkan tingkat akurasi metode yang digunakan dalam klasifikasi dengan benar berdasarkan semua data yang di klasifikasi.[9] Berikut adalah rumus untuk mencari nilai Akurasi.

$$Akurasi = \frac{TN + TP}{TN + TP + FN + FP}$$

b) Presisi

Presisi adalah nilai yang menggambarkan jumlah data kategori sentimen positif yang diklasifikasi dengan benar di bagi dengan semua data yang diklasifikasi sebagai sentimen positif.[9] Berikut adalah rumus untuk mencari nilai Presisi.

$$Presisi = \frac{TP}{TP + FP}$$

c) Recall

Recall merupakan suatu rasio prediksi yang benar-benar sentimen positif dibagi dengan semua data yang benar-benar sentimen positif.[9] Berikut adalah rumus untuk mencari nilai *Recall*.

$$Recall = \frac{TP}{TP + FN}$$

3. HASIL DAN ANALISA

3.1. Pengumpulan Data

Menggunakan *library snsrape* yang ada pada bahasa pemrograman *python* penulis berhasil mengumpulkan data *tweets* pengguna *twitter* sebanyak 5000 data. Data yang telah berhasil dikumpulkan kemudian disimpan dalam format csv. Gambar 3 berikut ini merupakan contoh hasil pengumpulan data menggunakan *library snsrape*.

Gambar 6 dibawah ini adalah komposisi hasil pelebelan data.

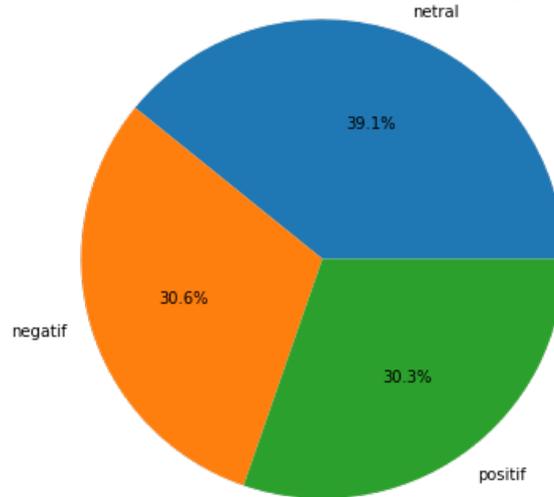
```

netral    1647
negatif   1287
positif   1275
Name: TextBlob, dtype: int64
    
```

Gambar 6. Komposisi Lebellng Data

3.4. Visualisasi Data

Penulis melakukan visualisasi data agar data lebih mudah dibaca, serta membuat tampilannya lebih menarik. Gambar 7 berikut ini merupakan hasil dari visualisasi data menggunakan *Library Matplotlib*.



Gambar 7. Hasil Visualisasi Data

3.5. Klasifikasi dan Evaluasi

Pada penelitian ini klasifikasi dan evaluasi dilakukan menggunakan metode SVM. Langkah pertama proses klasifikasi dan evaluasi yaitu memisahkan data menjadi *training* dan *testing*. Pada penelitian ini pembagian data dilakukan menggunakan 3 nilai perbandingan data yaitu :

- a) *Training* 90% dan *Testing* 10%

Tabel 2 dibawah ini merupakan hasil dari pembagian data *training* 90% dan *testing* 10%

Tabel 2. Hasil pembagian data 9:1

Data	Persen	Total
<i>Training</i>	90%	3788
<i>Testing</i>	10%	421
Jumlah	100%	4209

Sesudah data dibagi menjadi 2 yakni, *training* dan *testing* maka tahapan berikutnya yaitu klasifikasi dan evaluasi menggunakan *confusion matrix* yang terdapat pada metode SVM. Gambar 8 berikut ini merupakan hasil klasifikasi dan evaluasi menggunakan *confusion matrix* pada perbandingan data 90% : 10%.

```

confusion matrix
[[ 88 23 18]
 [ 2 151 12]
 [ 20 25 82]]
    
```

	precision	recall	f1-score	support
negatif	0.80	0.68	0.74	129
netral	0.76	0.92	0.83	165
positif	0.73	0.65	0.69	127
accuracy			0.76	421
macro avg	0.76	0.75	0.75	421
weighted avg	0.76	0.76	0.76	421

Gambar 8. Hasil SVM perbandingan 9:1

Berdasarkan gambar 8 di atas dapat dilihat bahwa hasil klasifikasi menggunakan metode SVM dengan data *training* 90% dan data *testing* 10% memperoleh nilai akurasi yaitu 76%, nilai presisi sentimen negatif 80% dan recall sentimen negatif 68%.

b) *Training* 80% dan *Testing* 20%.

Tabel 3 dibawah ini merupakan hasil dari pembagian data *training* 80% dan *testing* 20%

Tabel 3. Hasil pembagian data 8:2

Data	Persen	Total
<i>Training</i>	80%	3367
<i>Testing</i>	20%	842
Jumlah	100%	4209

Sesudah data dibagi menjadi 2 yakni, *training* dan *testing* maka tahapan berikutnya yaitu klasifikasi dan evaluasi menggunakan *confusion matrix* yang terdapat pada metode SVM. Gambar 9 berikut ini merupakan hasil klasifikasi dan evaluasi menggunakan *confusion matrix* pada perbandingan data 80% : 20%.

```

confusion matrix
[[168 43 46]
 [ 7 294 29]
 [ 32 62 161]]

```

	precision	recall	f1-score	support
negatif	0.81	0.65	0.72	257
netral	0.74	0.89	0.81	330
positif	0.68	0.63	0.66	255
accuracy			0.74	842
macro avg	0.74	0.73	0.73	842
weighted avg	0.74	0.74	0.74	842

Gambar 9. Hasil SVM perbandingan 8:2

Dari gambar 9 tersebut dapat dilihat bahwa hasil klasifikasi menggunakan metode SVM dengan data *training* 80% dan data *testing* 20% memperoleh nilai akurasi yaitu 74%, nilai presisi sentimen negatif 81% dan recall sentimen negatif 65%.

c) *Training* 70% dan *Testing* 30%

Tabel 4 dibawah ini merupakan hasil pembagian data *training* 70% dan *testing* 30%

Tabel 4. Hasil pembagian data 7:3.

Data	Persen	Total
<i>Training</i>	70%	2946
<i>Testing</i>	30%	1263
Jumlah	100%	4209

Sesudah data dibagi menjadi 2 yakni, *training* dan *testing* maka tahapan berikutnya yaitu klasifikasi dan evaluasi menggunakan *confusion matrix* yang terdapat pada metode SVM. Gambar 10 berikut ini merupakan hasil klasifikasi dan evaluasi menggunakan *confusion matrix* pada perbandingan data 70% : 30%.

```

confusion matrix
[[243 65 78]
 [ 14 440 40]
 [ 47 92 244]]

```

	precision	recall	f1-score	support
negatif	0.80	0.63	0.70	386
netral	0.74	0.89	0.81	494
positif	0.67	0.64	0.66	383
accuracy			0.73	1263
macro avg	0.74	0.72	0.72	1263
weighted avg	0.74	0.73	0.73	1263

Gambar 10. Hasil SVM perbandingan 7:3

Dari gambar 10 tersebut dapat dilihat bahwa hasil klasifikasi menggunakan metode SVM dengan data *training* 70% dan data *testing* 30% memperoleh nilai akurasi yaitu 73%, nilai presisi sentimen negatif 80% dan recall sentimen negatif 63%.

4. KESIMPULAN

Adapun hasil dari penelitian yang sudah penulis lakukan mengenai analisis sentimen masyarakat terhadap kenaikan harga BBM di media sosial twitter menggunakan metode *support vector machine* adalah dengan menggunakan *library snsrape* penulis memperoleh data *tweets* sebanyak 5000 data, yang kemudian dilakukan preprocessing sehingga menyisakan sebanyak 4209 data yang bisa diolah. Data yang penulis ambil yaitu data *tweets* dari pengguna media sosial *twitter* dari jangka waktu bulan september 2022 – oktober 2022. Hasil labelling data menggunakan *library TextBlob* dibagi menjadi 3 sentimen yaitu sentimen positif 1275 data, sentimen negatif 1287 data, dan sentimen netral 1647 data. Berdasarkan nilai perbandingan data *training* dan *testing* sebesar 70% : 30%, 80% : 20% dan 90% : 10% diperoleh hasil yang paling tinggi menggunakan metode SVM adalah perbandingan data 90%:10% dengan tingkat akurasi sebesar 76%, nilai presisi sentimen negatif 80% dan recall sentimen negatif 68%.

ACKNOWLEDGEMENTS

Paper ini merupakan hasil penelitian tugas akhir mahasiswa Universitas Bina Insan, Lubuklinggau tahun 2023.

DAFTAR PUSTAKA

- [1] S. D. Harijiatno, "Analisis Sentimen Pada Twitter Menggunakan Multinomial Naive Bayes," pp. 5–10, 2019.
- [2] G. R. Hrp and N. Aslami, "Analisis Dampak Kebijakan Perubahan Publik Harga BBM terhadap Perekonomian Rakyat Indonesia," *J. Ilmu Komputer, Ekon. dan Manaj.*, vol. 2, no. 1, pp. 1464–1474, 2022.
- [3] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015.
- [4] P. M. Nirmala Dharmapatni and N. L. P. Merawati, "Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan," *J. Bumigora Inf. Technol.*, vol. 2, no. 2, pp. 105–112, 2020, doi: 10.30812/bite.v2i2.904.
- [5] A. Baita, Y. Prityanto, and N. Cahyono, "Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN)," *Inf. Syst. J.*, vol. 4, no. 2, pp. 42–46, 2021, [Online]. Available: <https://jurnal.amikom.ac.id/index.php/infos/article/view/687>.
- [6] L. Lusiana, "Analisis Sentimen Twitter Dengan Menggunakan Algoritma Support Vector Machine (SVM) (Studi Kasus: 3556 Data Tweets dengan Kata Kunci Cadar dan Hijab)," 2018.
- [7] J. Enterprise, *Otodidak Pemrograman Python*. Elex Media Komputindo, 2017.
- [8] F. Syadid, "Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf-Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network," *Skripsi Univ. Islam Negeri Syarif Hidayatullah Jakarta*, p. 72, 2019.
- [9] A. Alwasi'a, "Analisis Sentimen Pada Review Aplikasi Berita Online Menggunakan Metode Maximum Entropy (Studi Kasus: Review Detikcom pada Google Play 2019)," 2020.
- [10] A. S. Handayani, S. Soim, T. E. Agusdi, R. Rumiasih, and A. Nurdin, "Klasifikasi Kualitas Udara Dengan Metode Support Vector Machine," *J. Inform. dan Rekayasa Elektron.*, vol. 3, no. 2, pp. 187–199, 2020.
- [11] S. Y. R. Munawar, "Sistem Pendeteksi Berita Hoax Di Media Sosial Dengan Teknik Data Mining Scikit Learn," vol. 4, pp. 173–179, 2019.