

ANALISIS SENTIMEN PADA OPINI KENAIKAN HARGA BAHAN BAKAR MINYAK PADA MEDIA SOSIAL TWITTER

Windah Maria Sonia Nadiyah Hutagalung*¹, Tony², Novario Jaya Perdana³
¹Sistem Informasi, Universitas Tarumanagara,
²Sistem Komputer, Universitas Tarumanagara,
³Teknik Informatika, Teknik Informatika Institut Teknologi Sepuluh Nopember,
Jl. Letjen S. Parman No. 1, Jakarta Barat 11440, Indonesia
e-mail: ^{1*}windahnadiyah@gmail.com, ²tony@fti.untar.ac.id, ³novariojp@fti.untar.ac.id

Bahan bakar minyak (BBM) adalah salah satu kebutuhan masyarakat di Indonesia. Pada awal September 2022, pemerintah Indonesia mengeluarkan pernyataan penyesuaian terhadap 3 jenis bahan bakar minyak. Pernyataan tersebut mendapat respon yang kurang baik oleh masyarakat di Indonesia, baik penyampaian secara langsung maupun penyuaran opini melalui media sosial. Twitter adalah salah satu media sosial tempat masyarakat menyuarkan opini, dengan topik kenaikan harga bahan bakar minyak ini akan dilakukan sebuah penelitian analisis sentimen. Tweet akan diklasifikasikan menjadi sentimen positif atau negatif, proses klasifikasi ini menggunakan algoritma Naïve Bayes. Crawling data dilakukan dengan aplikasi Rapid Miner, lalu pemrosesan data dan analisis dilakukan dengan bahasa pemrograman Python dengan bantuan Google Colaboratory sebagai compiler secara online. Hasil dari penelitian ini ditemukan sebanyak 78,13% tweet bersifat negatif dan 21,87% tweet bersifat positif, dengan kata "harga" dan "bbm" adalah kata yang paling sering muncul dalam Tweet yang berbeda.

Kata Kunci: Analisis Sentimen, Twitter, Kenaikan Harga BBM, RapidMiner, Naïve Bayes, Python, Google Colaboratory

I. PENDAHULUAN

Bahan bakar minyak (BBM) adalah salah satu bagian dari kebutuhan primer yang dibutuhkan oleh masyarakat di Indonesia dikarenakan memiliki peranan penting dalam membantu seluruh aktivitas kehidupan masyarakat di Indonesia seperti untuk bahan bakar kendaraan bermotor, maupun sebagai pemenuhan bahan bakar alat industri. Pada awal September 2022, pemerintah Indonesia mengeluarkan pernyataan bahwa akan ada penyesuaian harga pada 3 jenis bahan bakar minyak yaitu pertalite, pertamax, dan solar bersubsidi. Kenaikan harga bahan bakar minyak tersebut disebabkan oleh beberapa faktor, yaitu harga minyak mentah yang cenderung tinggi, membesarnya anggaran subsidi, dan turunnya nilai tukar USD terhadap Rupiah. Pernyataan tersebut tentu saja mendapat respon yang kurang baik oleh masyarakat Indonesia, baik penyuaran opini secara langsung maupun

melalui media sosial. Twitter merupakan salah satu media sosial yang paling banyak digunakan oleh pengguna media sosial, dikarenakan penggunaan Twitter yang sangat mudah untuk saling berbagi informasi sehingga setiap orang yang membaca dapat saling terhubung [1]. Tweet pengguna media sosial Twitter akan dijadikan sebagai data penelitian yang akan dilakukan menggunakan aplikasi RapidMiner. Tujuan dari analisis sentimen adalah untuk mengetahui bagaimana tanggapan masyarakat di Indonesia tentang kenaikan harga bahan bakar minyak dengan dua klasifikasi sentimen yaitu positif dan negatif. Pengumpulan data dilakukan dengan periode waktu selama 1 bulan yaitu dari bulan September s/d Oktober dan hasil data yang diperoleh untuk di analisa sebanyak 1.029 data.

II. LANDASAN TEORI

A. Website

Website adalah kumpulan dari beberapa halaman yang memiliki isi berupa data digital seperti video, gambar, suara, teks, ataupun gabungan dari semua hal tersebut yang dapat diakses oleh semua orang yang memiliki koneksi dengan internet [2]. Untuk mengakses website dapat menggunakan berbagai macam platform penunjang seperti *Google Chrome*, *Mozilla Firefox*, *Windows Edge*, dan *browser* lainnya. Penggunaan website mempunyai kelebihan seperti tidak membutuhkan aplikasi tersendiri untuk mengakses informasi yang diperlukan, hanya memerlukan koneksi internet yang stabil agar dapat menggunakan platform dengan maksimal.

B. Analisis Sentimen

Analisis sentimen adalah alat klasifikasi teks paling umum yang menganalisis pesan masuk dan memberitahu apa sentimen yang mendasari teks tersebut bersifat positif, negatif, atau netral [3]. Contoh dari analisis sentimen yaitu dapat menganalisis sentimen masyarakat pengguna media sosial Twitter, sehingga dapat memberikan penilaian berdasarkan tweet masyarakat di Indonesia.

C. Naïve Bayes

Naive Bayes adalah metode machine learning untuk

probabilitas. Dalam katalain Naive Bayes merupakan metode untuk klasifikasi text dengan kecepatan pemrosesan yang tinggi jika dalam skala data besar [4]. Naive Bayes mempunyai ciri-ciri yaitu asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi. Naive Bayes adalah bentuk paling sederhana dari pengklasifikasi jaringan Bayesian. Dalam Naive Bayes, setiap node fitur memiliki simpul kelas sebagai induknya, tetapi tidak memiliki orang tua dari node fitur lainnya.

D. Rapid Miner

Rapid Miner adalah salah satu aplikasi *open source* yang terkenal dalam dunia data mining, dan sangat mendukung dalam proses *text mining* dan teknik *data mining* yang dalam pengaplikasiannya dapat dikombinasikan dengan *text mining*. Dalam aplikasi *rapid miner* ada banyak jenis fitur tambahan terpisah yang dapat di *install* sesuai dengan kebutuhan pengguna, seperti untuk *text processing*, *wekaextension*, *parallel processing*, *web mining*, *reporting extension*, *series processing*, *predictive model markup language (PMML)*, *community*, dan *R extension packages*. [5]. Pada penelitian ini, *operators* yang digunakan yaitu *search Twitter*, *remove duplicate*, *select attribute*, *subprocess*, dan *write csv*. Untuk tampilan *Rapid Miner* dapat dilihat pada Gambar 1.



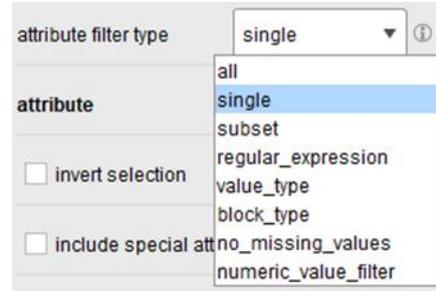
Gambar 1 Tampilan Rapid Miner.

Fitur Search Twitter dapat dilihat pada Parameters. Query digunakan untuk keyword pencarian komentar Twitter. Keyword yang digunakan disini adalah kenaikan bbm, sehingga semua hasil data yang diperoleh adalah tentang kenaikan harga bbm. Limit disini berfungsi sebagai parameter mengukur seberapa banyak data yang ingin didapatkan, semakin banyak data yang diambil, prosesnya juga akan semakin lama. Language adalah parameter untuk menentukan bahasa yang digunakan disini yaitu bahasa Indonesia. Tampilan parameter dapat dilihat pada Gambar 2.

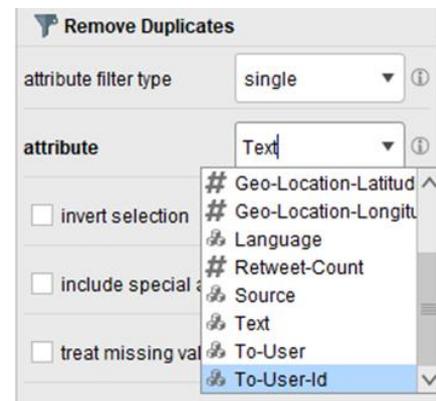


Gambar 2 Search Twitter

Fitur *remove duplicates* adalah untuk menghapus komentar atau tweets yang berulang agar mempermudah proses analisis yang akan dilakukan. Pada fitur *remove duplicates* terdapat beberapa parameters yaitu *attribute filter type* dan *attribute*. Untuk tampilan dari fitur filter *type* dapat dilihat pada Gambar 3 sedangkan untuk fitur *attribute* dapat dilihat pada Gambar 4.

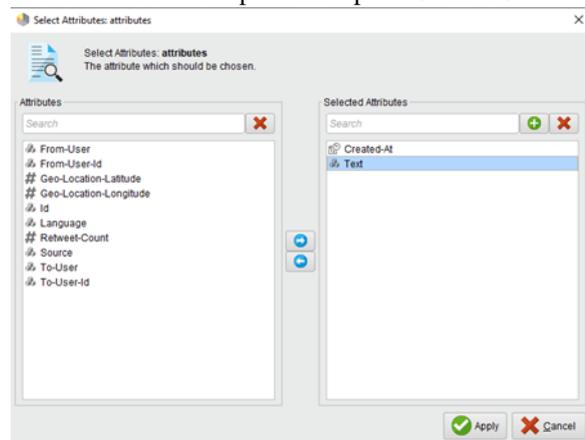


Gambar 3 attribute filter type



Gambar 4 attribute

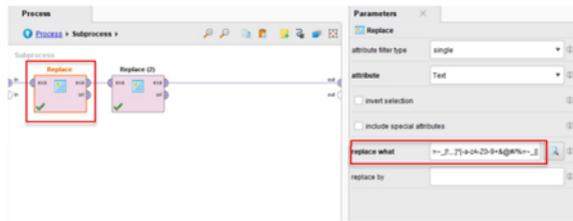
Fitur *select attributes* berfungsi untuk memilih kolom mana saja yang datanya ingin diambil. Terdapat beberapa kolom yang dapat dipilih yaitu *Row No.*, *Id*, *Created-at*, *From-User*, *From-User-Id*, *To-User*, *To-User-Id*, *Language*, *Source*, *Text*, *Geo-Location-Latitude*, *Geo-LocationLongitude*, *Retweet-Count*. Untuk tampilan dari fitur *select attributes* dapat dilihat pada Gambar 5.



Gambar 5 Select attribute

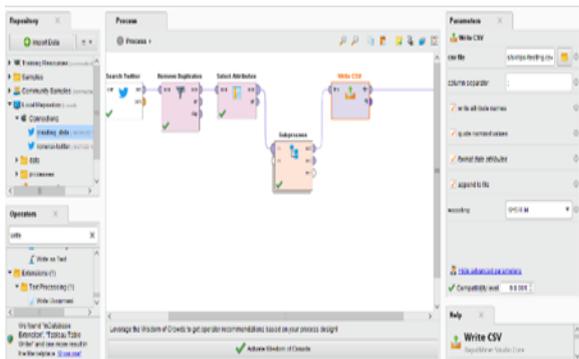
Fitur *replace* pada *RapidMiner* berfungsi untuk menghilangkan url link dan juga hastag. Fitur *replace* sendiri memiliki beberapa parameters yaitu *attribute filter type*, *attribute*, *replace what* dan *replace by*. Fungsi dari *replace what* adalah untuk menghilangkan text yang ditulis pada fungsi. Pada penelitian ini, text yang akan

dihilangkan ada 2 jenis yaitu text yang mengandung url link dan juga text yang mengandung hastag beserta isinya. Untuk tampilan dari fitur *replace* dapat dilihat pada Gambar 6.



Gambar 6 *replace*

Fitur *Write CSV* untuk menyimpan hasil dari proses *crawling* data dan pengolahan data kedalam suatu *file* dengan *format* CSV. Nama *file* serta tempat penyimpanannya dapat dipilih sendiri oleh *user*. Pada fitur *write CSV* terdapat beberapa *parameters* yang digunakan yaitu *csv file* yang merupakan tempat menyimpan *filenya*, kemudian ada *date format* yaitu untuk mengatur *format* tanggal yang digunakan yaitu *yyyy-MM-dd HH:mm:ss*. Untuk tampilan dari fitur *Write CSV* dapat dilihat pada Gambar 7.



Gambar 7 *write CSV*

E. *Twitter*

Twitter adalah sebuah situs jejaring sosial yang sedang berkembang pesat saat ini karena pengguna dapat berinteraksi dengan pengguna lainnya dari komputer maupun *smartphone* mereka dari manapun dan kapanpun[6]. Dalam proses analisis sentimen *Twitter* ini, penulis menggunakan trending topik pembicaraan oleh pengguna *Twitter* dengan topik kenaikan harga *bbm* agar dapat diambil datanya untuk di proses. Proses pengambilan data ini melalui website *Twitter Developer*, yaitu website pengembangan dari *Twitter* yang terdapat data API *Twitter* didalamnya.

F. *Google Colaboratory*

Google Colab adalah layanan yang disediakan oleh *Google* dan bersifat gratis. Untuk membantu memanfaatkan *Python* dan *Pandas* dengan mudah, diperlukan juga *integrated development environment* yang paling mudah diperoleh. Keunggulan yang dimiliki *google colab* adalah pengguna tidak perlu menginstall *Python* dan *Pandas* kedalam komputer, serta tidak membutuhkan editor yang terinstall. Karena yang dibutuhkan hanyalah akun *Google* aktif [7].

G. *Confusion Matrix*

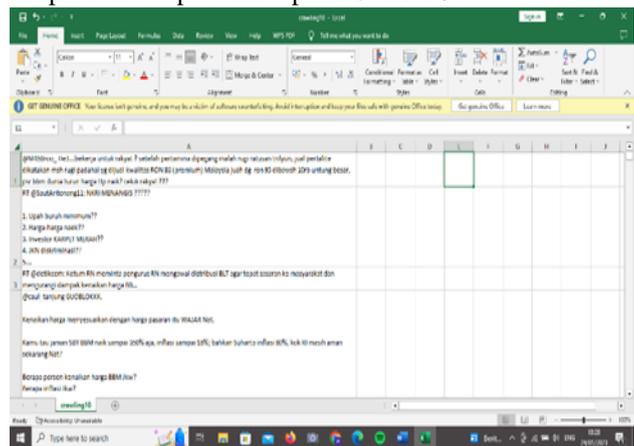
Confusion Matrix adalah tabel yang berguna untuk menguji seberapa akurat proses klasifikasi yang sudah dilakukan [8]. Untuk tabel *confusion matrix* dapat dilihat pada Gambar 8.

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	<p>TP (True Positive)</p>	<p>FP (False Positive) <i>Type I Error</i></p>
	0 (Negative)	<p>FN (False Negative) <i>Type II Error</i></p>	<p>TN (True Negative)</p>

Gambar 8 *confusion matrix*

H. *Data Penelitian*

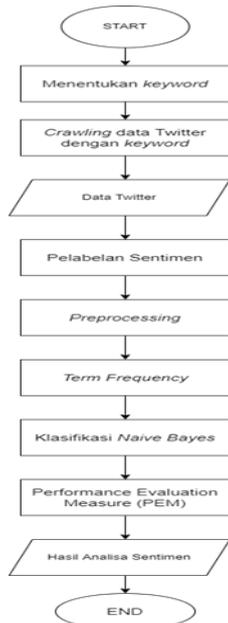
Data penelitian berupa data komentar pengguna media sosial *Twitter* dimana *crawling data* dilakukan dengan bantuan aplikasi *Rapid Miner*. Data hasil *crawling* disimpan kedalam file dengan format tipe data *.CSV*, tampilan file dapat dilihat pada Gambar 9.



Gambar 9 Data penelitian

III. HASIL DAN PEMBAHASAN

Alur pengerjaan menggambarkan langkah-langkah proses analisis sentimen dimulai dari proses *crawling* data sampai dengan menampilkan hasil. Alur pengerjaan dapat dilihat pada Gambar 10. Penelitian dimulai dari pengambilan data dengan menggunakan aplikasi *Rapid Miner* yang kemudian data yang telah didapatkan akan disimpan didalam file dengan format *.CSV* kemudian data akan diproses pada tahapan *preprocessing*. Tahapan-tahapan dalam *preprocessing* ada *cleaning*, *case folding*, *stopword removal*, *tokenizing*, dan *stemming*. Kemudian setelah data dibersihkan, akan dilakukan proses klasifikasi sebanyak dua kali, yang pertama klasifikasi sentimen kemudian klasifikasi *naïve bayes*. Klasifikasi tanpa metode adalah proses klasifikasi tanpa menggunakan metode yang proses klasifikasinya dilakukan secara otomatis menggunakan library *Python*. Proses terakhir adalah *confusion matrix* yang berfungsi untuk mencari tahu seberapa akurat.

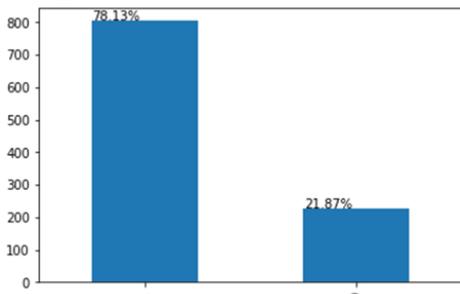


Gambar 10 Alur pengerjaan.

Terdapat beberapa hasil yang diperoleh setelah menjalankan proses analisis sentimen sebagai berikut :

A. Bar Chart

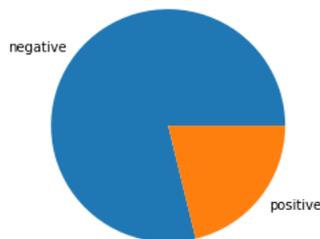
Bar chart berikut menggambarkan presentase dari hasil pelabelan data *tweet*, sebanyak 78,13% merupakan *tweet* yang bersifat negatif dan sebanyak 21,87% merupakan *tweet* yang bersifat positif. Tampilan gambar *bar chart* dapat dilihat pada Gambar 11.



Gambar 11 Bar Chart

B. Pie Chart

Pie chart berikut menunjukkan visualisasi dari presentase pelabelan data, dimana warna biru lebih mendominasi dengan sentimen negatif, dan warna orange dengan sentiment positif. Tampilan *pie chart* dapat dilihat pada Gambar 12.



Gambar 12 Pie Chart

C. Klasifikasi Sentimen

Pada proses ini dilakukan proses klasifikasi sentimen untuk menentukan sentimen positif, dan negatif.

Klasifikasi tanpa metode adalah proses klasifikasi tanpa menggunakan metode yang proses klasifikasinya dilakukan secara otomatis menggunakan library Python dengan aplikasi *google colab*. Setelah proses tersebut dilakukan maka didapatkan hasil sentimen positif sebanyak 225 dan sentimen negatif sebanyak 804 dengan total data yang diklasifikasi sebanyak 1029 data. Untuk tampilan hasil klasifikasi tanpa metode dapat dilihat pada Gambar 13.

```

1    804
0    225
Name: label, dtype: int64
  
```

Gambar 13 hasil pelabelan data

D. Akurasi Naive Bayes

Akurasi naive bayes didapatkan dengan cara membandingkan klasifikasi tanpa metode dengan klasifikasi dengan menggunakan metode naive bayes menggunakan confusion matrix. Jadi dapat disimpulkan bahwa klasifikasi dengan menggunakan metode naive bayes mempunyai tingkat akurasi sebesar 85%. Untuk tampilan akurasi dapat dilihat pada Gambar 14.

```

NB Accuracy Score -> 85.65891472868216
  
```

Gambar 14 Akurasi score

E. Wordcloud

Word cloud berfungsi untuk menampilkan data text secara visual. Semakin besar kata tersebut maka jumlah frekuensi kata tersebut semakin banyak. Dapat diketahui bahwa jumlah kata yang paling sering muncul adalah harga disusul dengan kata *bbm*. Selain kata yang paling banyak muncul secara keseluruhan, word cloud juga bisa menampilkan kata positif yang paling banyak muncul dan juga kata negatif yang paling banyak muncul. Untuk word cloud dan frekuensi kata positif dapat dilihat Gambar 18 dan kata negatif dapat dilihat pada Gambar 19.



Gambar 18 Wordcloud Positif



Gambar 19 Wordcloud Negatif

IV. KESIMPULAN

Berdasarkan analisis sentimen yang telah dilakukan, maka kesimpulan yang dapat ditarik adalah sebagai berikut :

1. Pada penelitian ini ditemukan sebanyak 78,13% sentimen bersifat negatif, dan 21,87% sentimen bersifat positif. Jadi, hasil yang didapatkan adalah pengguna media sosial *Twitter* lebih banyak menuliskan *tweet* yang bersifat positif dengan *query* pencarian “kenaikan harga bbm”.
2. Pada penelitian ini nilai akurasi yang fungsinya untuk mengukur seberapa akurat proses analisis sentimen yang telah dilakukan, nilai akurasi yang didapatkan pada penelitian ini adalah 85.6%.
3. Dari data yang digunakan dapat diketahui kata yang paling sering diucapkan, atau dengan kata lain sering muncul. Kata yang paling sering muncul pada penelitian ini adalah kata “harga” sebanyak 1200 kali, kemudian kata “bbm” menjadi kata yang paling sering digunakan selanjutnya, yaitu sebanyak 1150 kali.

DAFTAR PUSTAKA

- [1] Maireder, A. &. (2014). Political discourses on Twitter: networking topics, objects and people. *Twitter and Society*, 305-318.
- [2] Abdulloh, R. 2018. 7 in 1 Pemrograman web untuk pemula. Elex Media Komputindo.
- [3] Park, S. &. (2019, March 29). Gender Classification Using Sentiment Analysis and Deep Learning in Health Web Forum. *Applied Sciences*, 9(6), pp. 85-86.
- [4] Fitriyyah, S. N. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi & Penelitian Informatika*, 5(3).
- [5] Hoffman M. (2015). *Data Mining*. Chapman and Hall.
- [6] Maireder, A. &. (2014). Political discourses on Twitter: networking topics, objects and people. *Twitter and Society*, 305-318.
- [7] Jubilee Enterprise. (2021). *Belajar Sendiri Mengolah Data dengan Python dan Pandas*. Elex Media Komputindo.
- [8] Normawati, D., & Prayogi, S. A. (2021). Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 5(2), 697-711