

PROSIDING
Seminar Nasional Teknologi,
Kualitas dan Aplikasi 2022



SEMINAR NASIONAL
TEKNOKA 7
Seminar Nasional Teknologi, Kualitas dan Aplikasi

KOLABORASI SAINS DAN TEKNOLOGI DALAM MENGATASI PROBLEMATIKA PANGAN DAN ENERGI DI INDONESIA

ISSN: 2502-8782 e-ISSN: 2580-6408
Memiliki Digital Object Identifier (DOI)



SABTU, 3 DESEMBER 2022
08.00 s.d 15.45 WIB



PENYELENGGARA

FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA

📍 Jl. Tanah Merdeka No.6 Kp. Rambutan, Ciracas, Jakarta Timur
☎️ (021) 8400941 📠 (021) 87782739
✉️ teknoka@uhamka.ac.id
🌐 www.teknoka.uhamka.ac.id

Vol. 4 Vol. 3 Vol. 3 Vol. 4 Vol. 1 Vol. 6 Vol. 5 Vol. 2 Vol. 7 QEEG Engin Engin Vol 6 V x +

https://journal.uhamka.ac.id/index.php/teknoka/issue/view/444

Analisa Visualisasi Data Kematian Yang Disebabkan Oleh Penyakit Hiv Dan Malaria Diseluruh Dunia Dengan Metode Business Intelligence Menggunakan Dashboard Tableau
Mutiar Zahra Arifin, Bagas Kembar Rezkyllah, Fadli Hardiyanto Putra, Rizky Ramdhani, Firman Noor Hasan 63-70
[PDF \(Bahasa Indonesia\)](#)

Implementasi Business Intelligence Untuk Menvisualisasi Data Kekerasan Di Provinsi Jawa Barat Menggunakan Tableau
Farhan Bias Purnama Putra, Rizki Alamsyah, Mohammad Akhdaan Juliandra, Isnan Wisnu Prastiyo, Firman Noor Hasan 85-94
[PDF \(Bahasa Indonesia\)](#)

Analisis Sentimen Kenaikan Harga BBM pada Media Sosial Twitter
Iqbal Musyaffa, Mia Kamayani 124-134
[PDF \(Bahasa Indonesia\)](#)

Perancangan Sistem Pembukuan dan Transaksi Berbasis Website (Studi Kasus: Kios Pelangi)
Fahmi Triyadi Fauzi, Nunik Pratiwi 106-113
[PDF \(Bahasa Indonesia\)](#)

Klasifikasi Komentar Abusive Dan Hate Speech Teks Twitter Menggunakan Metode Convolutional Neural Network
Indri Pangestuti, Surya Agustian 23-30

https://journal.uhamka.ac.id/index.php/teknoka/article/view/11244

30° Search ENG 20:25 22/10/2023

Analisis Sentimen Kenaikan Harga BBM pada Media Sosial Twitter

Iqbal Musyaffa, Mia Kamayani

Universitas Muhammadiyah Prof. Dr. Hamka

Jl. Tanah Merdeka No.6 Kota Jakarta Timur 13830, (021) 87782739, Fax. (021)7261226

Website: www.ft.uhamka.ac.id, E-mail: ft@uhamka.ac.id

Abstrak

Berdasarkan informasi dalam jumpa pers di Istana Merdeka pada hari Sabtu, 03 September 2022 pukul 14.30 WIB, Presiden Joko Widodo menetapkan bahwa harga BBM resmi dinaikkan. Hal tersebut menjadi beban bagi pengguna jalan terutama pengguna kendaraan roda dua. Saat ini, banyak masyarakat yang menggunakan situs media sosial untuk menyampaikan keluhan terkait topik tersebut. Salah satunya media sosial Twitter. Oleh karena itu, dilakukan analisis sentimen dengan menggunakan 2 metode yaitu Naïve Bayes Classifier dan Decision Tree terhadap 1100 tweet yang didapat dari kata kunci "bbm naik". Hasil pengujian bahwa performa terbaik dengan menggunakan algoritma Naïve Bayes Classifier, dengan menghasilkan nilai dengan akurasi 94.91% dan untuk algoritma Decision Tree hanya mendapatkan akurasi 62.57%. Hasil sentimen yaitu, positif berjumlah 68 data, netral berjumlah 20 data, dan negatif berjumlah 301 data. Hasil sentimen negatif yang lebih banyak menunjukkan bahwa kenaikan harga bbm di Indonesia belum dapat diterima oleh masyarakat Indonesia pada media sosial Twitter.

Kata kunci: : RapidMiner, Naïve Bayes Classifier, Decision Tree, Analisis sentimen, BBM Naik

Abstract

Based on information in a press conference at the Merdeka Palace on Saturday, 03 September 2022 at 14.30 WIB, President Joko Widodo has decreed that the price of fuel be officially raised. This is a burden for road users, especially users of two-wheeled vehicles. Currently, many people use social media sites to submit complaints regarding this topic. One of them social media Twitter. Therefore, sentiment analysis was carried out using 2 methods, namely the Naïve Bayes Classifier and the Decision Tree on 1100 tweets obtained from the keyword "bbm up". The test results show that the best Performance is using the Naïve Bayes Classifier algorithm, which produces values with an accuracy of 94.91% and for the Decision Tree algorithm only gets an accuracy of 62.57%. The results of sentiment are, positive totaling 68 data, neutral totaling 20 data, and negative totaling 301 data. The results of more negative sentiment show that the increase in fuel prices in Indonesia has not been accepted by the Indonesian people on social media Twitter.

Keyword: RapidMiner, Naïve Bayes Classifier, Decision Tree, Sentiment Analysis, Fuel Oil Rises

1 PENDAHULUAN

Dampak besar untuk masyarakat Indonesia, baik konsumsi langsung dan tidak langsung adalah Bahan Bakar Minyak (BBM). Karena dampak dari harga BBM yang tidak stabil memberikan pengaruh pada distribusi, transportasi, biaya produksi sehingga beberapa harga barang yang lain ikut terpengaruh. Kenaikan harga BBM untuk masyarakat menimbulkan pro dan kontra khususnya kalangan menengah ke bawah yang merasa dirugikan dimana dampak yang dialami yaitu pengeluaran anggaran rumah tangga mengalami peningkatan, pengeluaran pada sektor transportasi dan

pasti biaya untuk membeli bahan pokok ikut bertambah [1].

Berdasarkan informasi dalam jumpa pers di Istana Merdeka pada hari Sabtu, 03 September 2022 pukul 14.30 WIB, Presiden Joko Widodo menetapkan bahwa harga bahan bakar minyak (BBM) seperti Pertalite, Solar, dan Pertamina resmi dinaikkan. Hal tersebut menjadi beban bagi pengguna jalan terutama bagi pengguna kendaraan roda dua. Selain itu, tarif angkutan umum pun juga ikut menaik karena penyebab kenaikan harga BBM ini salah satunya adalah subsidi yang membengkak maka secara tidak langsung tarif angkutan umum pun akan ikut

menaik.

Saat ini, banyak masyarakat yang menggunakan situs media sosial untuk menyampaikan tanggapan atau keluhan terkait topik yang sedang dibicarakan. Salah satunya adalah media sosial Twitter. Tanggapan pada Twitter adalah data yang memiliki bentuk teks, maka *text mining* digunakan dalam melaksanakan analisis. Pada *text mining* sebuah kelas dibutuhkan klasifikasi. Klasifikasi adalah sebuah objek data untuk dimasukkan ke dalam kelas tertentu. *Tweet* atau tanggapan akan dikelompokkan untuk memasukan ke dalam kelas tanggapan positif, negatif, dan netral yang akan menjadi bahan pertimbangan agar diperbaiki. Sebelum melakukan klasifikasi diperlukan tahap *preprocessing* yaitu *case folding*, *stopwords*, *tokenizing*, dan *stemming* [2].

Penelitian sejenis yang berhubungan dengan analisis sentimen dan juga menjadi acuan terhadap peneliti yang sekarang yaitu penelitian dari [3] dengan topik “Analisis Sentimen Jasa Transportasi *Online* pada Twitter Menggunakan Metode *Naïve Bayes Classifier*” pada penelitian ini peneliti mengambil data komentar sebanyak 2000 data yang kemudian dilakukan klasifikasi untuk menghasilkan sentimen positif dan negatif serta akurasi sebesar 86.80%. Pada penelitian selanjutnya yaitu penelitian dari [4] dengan topik “Analisis Sentimen Pengguna E- Money pada Twitter Menggunakan Algoritma C4.5 dan *Naïve Bayes*” peneliti mengambil data komentar sebanyak 300 data yang kemudian dilakukan klasifikasi untuk menghasilkan sentimen positif dan sentimen negatif serta akurasi sebesar 84%.

Pada penelitian selanjutnya yaitu penelitian dari [6] dengan topik “Perbandingan Metode *Naïve Bayes*, KNN dan *Decision Tree* Terhadap Analisis Sentimen Transportasi KRL Commuter Line” pada penelitian ini peneliti mengambil data komentar sebanyak 150 data yang kemudian dilakukan klasifikasi untuk menghasilkan sentimen positif dan sentimen negatif. Pada perbandingan tiga algoritma didapatkan tingkat akurasi masing-masing

yaitu untuk *Naïve Bayes* sebesar 80%, KNN sebesar 80% dan *Decision Tree* sebesar 100%.

Berdasarkan penelitian tersebut, penulis akan melakukan perbandingan dengan mengimplementasikan dua metode yaitu, metode *Naïve Bayes Classifier* dan *Decision Tree* untuk analisis setimen. Harapan dari hasil tugas akhir terkait analisis sentimen dan visualisasi dengan pengambilan data tanggapan Twitter adalah penulis dapat memberikan informasi tentang tanggapan masyarakat terhadap kenaikan harga BBM di Indonesia.

2 LANDASAN TEORI

A. Analisis Sentimen

Analisis sentimen adalah metode yang dipakai untuk melakukan pemahaman, mengekstrak data opini, dan melakukan pengolahan data tekstual agar memperoleh sentiment yang mengandung sebuah opini secara otomatis. Terdapat 3 jenis opini, yaitu negatif, positif dan netral, oleh sebab itu respon terhadap sebuah produk atau pelayanan agar perusahaan terkait dapat mengetahuinya melalui feedback masyarakat [3].

B. RapidMiner

Rapid Miner ialah *software* yang dirancang oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* dan Ralf Klinkenberg dengan tampilan *Graphical User Interface* untuk memberi kemudahan *user* dalam memakai *software* ini.

C. Naïve Bayes Classifier

Naïve Bayes Classifier merupakan pengklasifikasian statistik yang dipakai ketika melakukan prediksi probabilitas sebuah class. Berdasarkan pada teorema Bayes, *Naïve Bayes Classifier* mampu melakukan klasifikasi sesuai dengan neural network dan *decision tree*. *Naïve Bayes Classifier* memiliki akurasi dan kecepatan yang tinggi jika data yang diimplementasikan pada database sangat besar.

D. Decision Tree

Decision tree ialah model prediksi yang bisa dipakai untuk prediksi dan klasifikasi. Proses pada *decision tree* ialah sebuah data tabel diubah ke dalam model *tree*. Model *tree* yang dihasilkan adalah rule dan disederhanakan.

3 METODOLOGI PENELITIAN

A. Pengumpulan Data

Pada perangkat lunak *RapidMiner* operator search twitter digunakan untuk pengambilan data di Twitter. Data tweet diambil dari beberapa pendapat masyarakat mengenai kenaikan harga BBM.

B. Klasifikasi Sentimen

Melakukan klasifikasi sentiment dengan pelabelan analisis sentiment untuk menentukan respon yang tersedia pada data (komentar tweet), semua respon ini dibagi menjadi respon positif dan negatif.

C. Preprocessing

1. Melakukan perubahan kalimat yang semula huruf kapital menjadi huruf kecil.
2. Membersihkan karakter-karakter yang tidak digunakan pada kalimat.
3. Menggunakan kosakata *stopwords* untuk menghapus kata yang tidak perlu digunakan pada kalimat.
4. Melakukan pemisahan kalimat menjadi potongan-perkata.
5. Merubah kalimat teks korpus yang mengandung nilai frekuensi banyaknya kata yang sering digunakan.
6. Membuat visualisasi data kalimat ke dalam bentuk frekuensi kemunculan kata.

D. Visualisasi Wordcloud

Melakukan visualisasi data berdasarkan teks yang sudah dilakukan *preprocessing* untuk menampilkan kata yang populer atau yang sering muncul pada teks dengan melihat frekuensi ukuran dari kata tersebut.

E. Klasifikasi *Naïve Bayes Classifier*

Mempersiapkan model data yaitu data uji dan data latih berdasarkan klasifikasi sentiment dengan perbandingan 80:20

dengan algoritma *Naïve Bayes Classifier*.

F. Klasifikasi *Decision Tree*

Mempersiapkan model data yaitu data uji dan data latih berdasarkan klasifikasi sentiment dengan perbandingan 80:20 dengan algoritma *Decision Tree*.

G. Analisis Hasil Penelitian

Menentukan hasil penelitian klasifikasi sentimentulasan tweet kenaikan harga BBM dengan metode *Decision Tree* dan *Naïve Bayes Classifier*. Hasil penelitian berupa analisis sentiment dari jenis sentiment yang didapatkan, tingkat akurasi dan visualisasi teks yang dihasilkan.

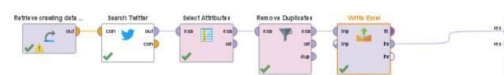
H. Pembuatan Laporan Hasil Penelitian

Membuat laporan berdasarkan hasil olah data yang dilakukan oleh penulis tentang analisis sentiment terhadap kenaikan harga BBM.

4 HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Operator search twitter di perangkat lunak *RapidMiner* digunakan untuk proses pengambilan data pada twitter. Data tweet diambil dari beberapa pendapat masyarakat mengenai kenaikan harga BBM dengan kata kunci “bbm naik”. Berikut ini adalah visualisasi proses pengambilan data twitter dengan *software RapidMiner* pada gambar 1.



Gambar 1 Proses pengambilan data

Keterangan :

a. Search Twitter

Operator yang memiliki fungsi untuk mengambil data dari media sosial Twitter. *Operator search twitter* berupa sejumlah Parameter yaitu: *connection*, *result type*, *query*, *limit*, *max id*, *since id*, *language*, dan *local*.

b. Select Attributes

Operator yang memiliki fungsi untuk melakukan penyaringan atribut yang ingin dipakai. Operator *select attributes* berupa sejumlah parameter yaitu: *attribute attribute*, *filter type*, *include special*

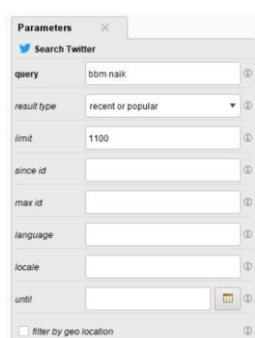
attributes dan *invert selection*.

c. *Remove duplicate*

Remove duplicate memiliki fungsi untuk menghapus data yang terduplikat. *Operator remove duplicate* mempunyai sejumlah parameter yakni: *include special attributes*, *attribute filter type*, *treat missing values as duplicates* dan *invert selection*.

d. *Write Excel*

Operator yang memiliki fungsi untuk melakukan *export* data ke *file excel*. Beberapa operator *write excel* yaitu: *file format*, *excel file*, *encoding*, *date format*, *sheet name* dan *numberformat*.



Gambar 2 Operator write excel

Pada gambar 2 di atas adalah parameter yang ada pada operator *search twitter* meliputi parameter *query*, *connection*, *result type*, *language* dan *limit*. Proses *crawling* data pada penelitian ini memakai kata kunci *bbm naik*. Jenis pencarian yang dipakai ialah *popular or recent*, pengambilan data untuk analisis sentimen berjumlah 1100 data tweet dengan bahasa Indonesia.

Row No.	ID	Text
1	157327202	Apakah benar Pertamina telah menaikkan kembali harga BBM naik? Ada yang merasakan pertambahan??
2	157189904	Masa akan yang tergolong dari Mahasiswaa Muhammadiyah (MAMU) Jakarta melakukan demo menolak ...
3	157129473	Sejauh PCIP Hasil KKR/KBK mendiri PKC yang selakangan virus mendiri kenaikan harga BBM. Hasil memo...
4	157331360	RT @baluyaini: Rupiah udah diburu Rp 15.000/USD. Apa dampaknya? 1 Ancaman PKK akan Merganda
5	157331428	RT @baluyaini: Tesas ini kean seati
6	157331428	@somet3 He ah oi ma, sebelum kenaikan BBM masih harga normal, itu bisa 1 liter ke lebih ut.
7	157330102	Bilangnya harga BBM naik, mencekuk tabung monev, tapi volume kendaraan di jalan makin parah zaididid.
8	157329881	RT @h3Phe: Peningkatan harga Gas @peramparawo "Di boleh ada yg ganggu pelaksanaan demo kompasan...
9	157329880	Ya sudah, mas lah into yang and to pada naik semua harga nya. Ya di bisa yg pernah walang "tapi naik ...
10	157329214	RT @tadellidong11: 4. Tapi lebih naik.
11	157327197	RT @waidididid: @salsalididid harga BBM melonjak, warga mendapatkan naik terlahi dawa kean in.
12	157329822	RT @Pamemidid: Bunga bank diprediksi naik.
13	157329866	Harga fuel Land Mula naik, Imbas Kenaikan harga BBM https://t.co/ndydygkUg

Gambar 3 Hasil pengambilan data

Setelah mendapatkan data tersebut pada gambar 3, kemudian dibagi menjadi 2 data yaitu data latih 80% dan data uji 20%. Pada 1100 data tweet, data latih berjumlah 880 data dan data uji berjumlah 220 data.

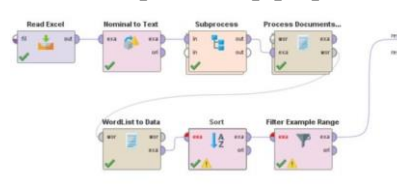
Kemudian dilakukan pemberian label pada data latih secara manual oleh peneliti dan dibantu 2 orang lainnya dengan menggunakan aplikasi Microsoft Excel sebagai penentuan respon yang ada dalam data (komentar tweet), semua respon pada penelitian ini terbagi menjadi respon positif, negatif, dan netral. Setelah melakukan pelabelan secara manual pada data latih mendapatkan total 389 data yang telah disortir, terdapat data dengan sentimen positif berjumlah 68 data, negatif berjumlah 301 data, dan netral berjumlah 20 data. Seperti yang ada pada tabel4.

Tabel 1 Data yang telah diberi label

No.	Teks	Sentimen
1	BBM harga naik, tapi upah kerja nggak naik. Kadang bingung mau nyalahin siapa ??????? https://t.co/LHNu3KjEXw	Negatif
2	Mahasiswa & masyarakat Pulau Jawa gerah harga BBM naik....kami pulau Kalimantan sudah lama merasakan, udah puluhan tahun dengan harga 18 ribuan perliter ...mana itu pemberataan OMDO z....!!! ?? https://t.co/UIFu6PIGfu	Positif
3	Pasca harga BBM naik, pemerintah akan kurangi peredaran LPG 3 kg subsidi. #hargabbm #SubsidiTepatSasaran https://t.co/APIpJqrkMx	Netral

B. Preprocessing

Preprocessing adalah tahap pertama dimana data yang bersih agar bisa berjalan pada tahap selanjutnya, preprocessing pada penelitian ini tahap preprocessing terdiri dari *filtering*, *cleansing* dan *tokenizing*. Pada gambar 4 di bawah ini adalah visualisasi alur pada tahap preprocessing.



Gambar 4 Alur tahap preprocessing

Keterangan :

a. Read excel

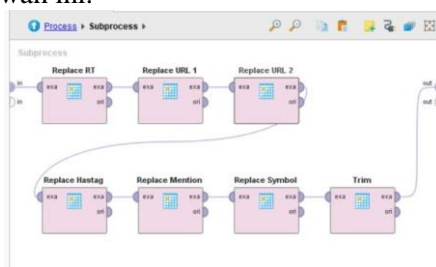
Read excel memiliki fungsi untuk membaca file dataset pada saat pengambilan data. Operator read excel mempunyai sejumlah parameter yaitu: *excel file, import configuration wizard, sheet number, sheetsselection, imported cell range, sheet name, annotations, first row as names, date format, locale, time zone, data set meta data information, read all values as polynomial, data management* dan *read not matching values as missings*.

b. Nominal Text

Dikarenakan tidak semua data yang diambil dari tweet berbentuk teks. Data yang penulis ubah merupakan teks dari tweet maka pada parameter *Operator nominal to text* pada *Attribute filter type* diubah menjadi *single* dan atributnya diubah menjadi *text*.

c. Subprocess

Pada tahap ini digunakan untuk menghilangkan tanda baca dan komponen khusus yang terdapat pada Twitter seperti, @, #, RT, URL, dan lain-lain. Untuk menjalankan tahapan ini peneliti menggunakan *operator Replace* dan *Subprocess*. Yang dimana *operator replace* ini akan dimasukkan ke dalam *operator subprocess*. Seperti gambar 5 di bawah ini.



Gambar 5 Proses operator replace pada operatorsubproces

Pada parameter operator *replace* *attribute filter type* diubah menjadi *single* dan atributnya diubah menjadi *text* setelah itu *replace what* dimasukkan dengan komponen regular *expression* seperti *http\S+,http\S+, #\S+, @\S+,RT,* dan *icon* seperti *[!"#\\$%'()*+,-./:;<=>t;?@[\\|_`{|}~]* dan untuk kolom *replace by* dikosongkan saja. Fungsi dari regular

expression digunakan untuk menghapus kata-kata yang mengandung komponen di atas.

d. Process Document

Pada tahap ini dilakukan pemberian bobot terhadap setiap kata yang terdapat pada tweet dengan melakukan *preprocessing*. Dalam pemberian bobot ini menggunakan metode TF-IDF, bobot setiap kata akan digunakan untuk proses pemeringkatan. Rumus umum untuk proses TF-IDF seperti di bawah ini.

$$W_{dt} = TF_{dt} * IDF_t \quad (3)$$

Di mana:

d = dokumen ke- d

t = kata ke- t dari kata kunci

W = bobot dokumen ke- d terhadap kata ke- t

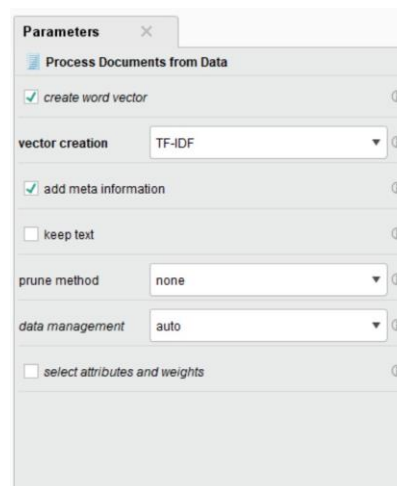
tf = banyaknya kata yang dicari pada sebuah dokumen

IDF = *Inverse Document Frequency*

D= total dokumen

df = banyak dokumen yang mengandung kata yang dicari

Pada tahap TF-IDF ini tetap menggunakan *operator Process Documents from Data* dan pada parameternya terdapat *checkbox* pilihan *create word vector* kemudian ceklis dan jangan lupa pada bagian *vector creation* pilih TF-IDF. Seperti pada gambar 6 di bawah ini.



Gambar 6 Process Documents from data

1) Transform Cases

Transform case memiliki tujuan untuk mengubah huruf Kapital (*Uppercase*) menjadi huruf kecil, seperti pada tabel 5 di bawah ini.

Tabel 2 Hasil Transform Case

Sebelum	Sesudah
@geloraco IYA LEBIH BAIK HUTANGNYA DAN RAKYAT TERBEBAN KAN SEMUA HARGA NAIK, BBM NAIK, TDL NAIK..	iya lebih baik hutangnya dan rakyat terbeban kan semua harga naik bbm naik tdl naik
RT @Quvvatt: Ternyata setelah harga BBM naik Mereka mau ganti mobil lama dg mobil listrik dan mau menaikan gaji perusahaan produsen BBM	ternyata setelah harga bbm naik mereka mau ganti mobil lama dg mobil listrik dan mau menaikan gaji perusahaan produsen bbm
@ruhutsitompul Menolak harga BBM naik tapi tetap antrian Beli BBM subsidi	menolak harga bbm naik tapi tetap antrian beli bbm subsidi

2) Tokenizing

Pada fase ini sebuah kalimat dipenggal menjadi kata tunggal, seperti contoh pada tabel 6 di bawah ini.

Tabel 3 Hasil Tokenizing

Sebelum	Sesudah
@geloraco IYA LEBIH BAIK HUTANGNYA DAN RAKYAT TERBEBAN KAN SEMUA HARGA NAIK, BBM NAIK, TDL NAIK..	“lebih” “baik” “rakyat” “semua” “harga” “naik” “bbm”
RT @Quvvatt: Ternyata setelah harga BBM naik Mereka mau ganti mobil lama dg mobil listrik dan mau menaikan gaji perusahaan produsen BBM	“ternyata” “setelah” “harga” “bbm” “naik” “ganti” “mobil” “lama” “mobil” “listrik” “menaikan” “gaji” “perusahaan” “produsen”
@ruhutsitompul Menolak harga BBM naik tapi tetap antrian Beli BBM subsidi	“menolak” “harga” “bbm” “naik” “tetap” “antrian” “beli” “bbm” “subsidi”

3) Stemming

Tahap *stemming* memiliki fungsi untuk mengubah kata menjadi kata dasar, seperti contoh pada tabel 7 di bawah ini.

Tabel 4 Hasil Stemming

Sebelum	Sesudah
@geloraco IYA LEBIH BAIK HUTANGNYA DAN RAKYAT TERBEBAN KAN SEMUA HARGA NAIK, BBM NAIK, TDL NAIK..	lebih baik hutang dan rakyat beban semua harga naik
RT @Quvvatt: Ternyata setelah harga BBM naik Mereka mau ganti mobil lama dg mobil listrik dan mau menaikan gaji perusahaan produsen BBM	Setelah harga bbm naik mobil lama dengan mobil baru
@ruhutsitompul Menolak harga BBM naik tapi tetap antrian Beli BBM subsidi	Menolak harga bbm

4) Filter Token By Length

Filter Token By Length berfungsi untuk memfilter jumlah kata yang akan digunakan dengan ketentuan minimal 4 karakter dan maksimal 25 karakter, seperti pada tabel 8 di bawah ini.

Tabel 5 Hasil Filter Token By Length

Sebelum	Sesudah
@geloraco IYA LEBIH BAIK HUTANGNYA DAN RAKYAT TERBEBAN KAN SEMUA HARGA NAIK, BBM NAIK, TDL NAIK..	lebih baik hutang dan rakyat beban semua harga naik
RT @Quvvatt: Ternyata setelah harga BBM naik Mereka mau ganti mobil lama dg mobil listrik dan mau menaikan gaji perusahaan produsen BBM	Setelah harga bbm naik mobil lama dengan mobil baru
@ruhutsitompul Menolak harga BBM naik tapi tetap antrian Beli BBM subsidi	Menolak harga bbm

e. *WordList to Data*

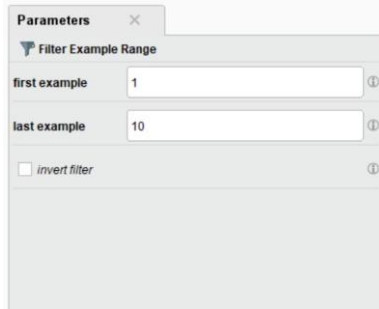
Wordlist to Data digunakan untuk menghitung bobot pada setiap kata yang akan muncul dalam dokumen.

f. *Sort*

Berfungsi untuk mendapatkan data berdasarkan jumlah terbanyak yang muncul hingga jumlah terkecil yang muncul.

g. *Filter Example Range*

Filter Example Range berfungsi untuk memfilter data dari yang paling atas hingga terakhir sesuai dengan ketentuan yang kita pilih dalam parameter *filter example range*. Di sini menggunakan 10 data sebagai nilai maksimal, maka untuk bagian *first example* diisi 1 untuk nilai paling atas atau sedikit dan *last example* diisi 10 untuk nilai terakhir atau yang paling banyak.



Gambar 7 Filter Example Range

Setelah semua operator saling terhubung, maka selanjutnya kita akan menjalankan proses ini, dengan cara menekan atau klik *button* segitiga berwarna biru. Dan proses yang dijalankan akan muncul 10 data teratas yang telah dipilih.

Row No.	word	in documents	total
1	harga	373	459
2	turun	66	69
3	listrik	42	55
4	rakyat	41	54
5	minyak	41	48
6	pemerintah	40	45
7	subsidi	31	36
8	jokowi	24	29
9	butuh	24	27
10	masyarakat	27	27

Gambar 8 Kata yang paling banyak muncul dari data latih

Pada gambar di atas ialah term(kata) yang telah melewati tahapan *preprocessing*, *in documents* merupakan jumlah kata yang muncul dalam setiap

dokumen.

C. *Visualisasi Wordcloud*

Setelah dilakukan tahap *preprocessing* kemudian dibuatkan visualisasi untuk dapat menampilkan kata yang populer atau yang sering muncul pada teks dengan melihat frekuensi ukuran dari kata tersebut yang bernama *wordcloud*. *Wordcloud* dibuat dengan bantuan RStudio menggunakan library (*wordcloud*) dan library (*RColorBrewer*) untuk memberikan warna pada visualisasi.

Gambar 9 Script untuk membuat wordcloud

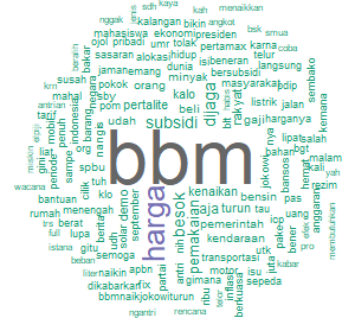
Hasil dari visualisasi *wordcloud* pada R-Studio didapatkan bahwa kata yang sering muncul atau ukuran frekuensi kata

```

00 cojola = plemel'buJ( ' ,b9K5,.)
01 nolqCJonq(ceK' 'wp' nolqE = J20' 'loz'bel = 0' 2' 'lnduqo'olqel = hVg2E'
02 cExE <- 92'cHlSccEel(qz9p9e2rjE)
03 JprEwY(kcojovlEmel)
04 nolqCJonq(ceK' 'wp' nolqE = J20' 'loz'bel = 0' 2' 'lnduqo'olqel = hVg2E'
05 JprEwY(kcojovlEmel)
06 nolqCJonq(ceK' 'wp' nolqE = J20' 'loz'bel = 0' 2' 'lnduqo'olqel = hVg2E'
07 nolqCJonq(ceK' 'wp' nolqE = J20' 'loz'bel = 0' 2' 'lnduqo'olqel = hVg2E'

```

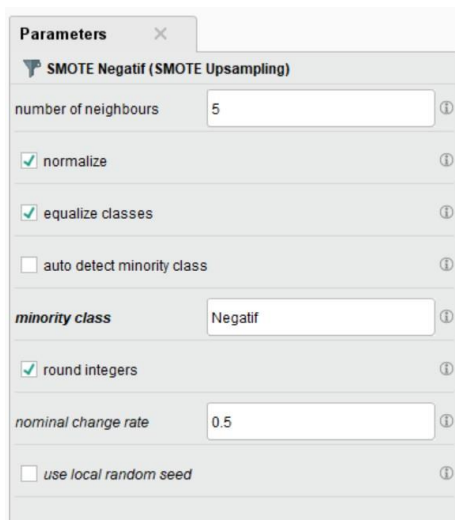
yang besar ada pada kata “bbm”, “harga”, “subsidi”.



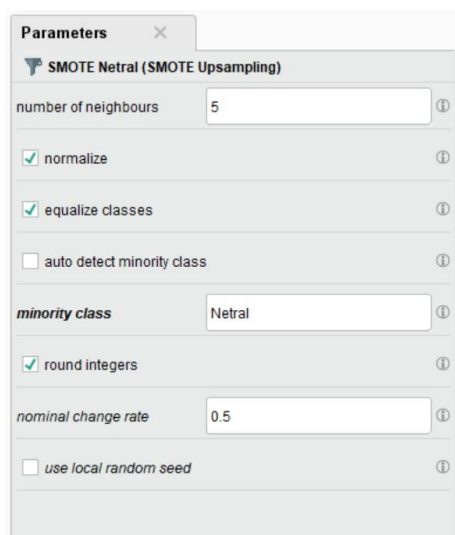
Gambar 10 Visualisasi Wordcloud

D. SMOTE Up Sampling

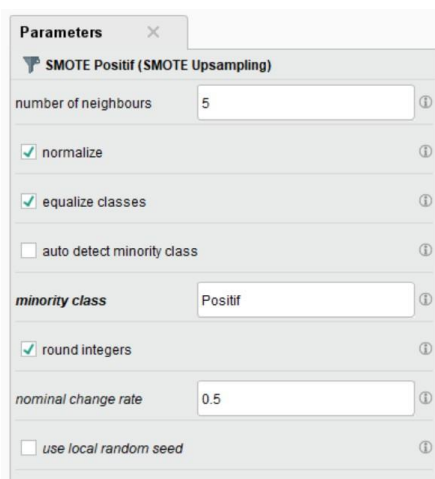
Setelah proses *preprocessing* pada dataset. Maka tahap berikutnya ialah meningkatkan kinerja metode memakai *Synthetic Minority Oversampling Technique* atau SMOTE. Proses ini dilaksanakan sebanyak tiga kali dengan operator yang sama dikarenakan terdapat tiga kelas sentimen yang berbeda dan memiliki label yang tidak seimbang, pada pilihan parameter SMOTE Up Sampling terdapat pilihan “*Minority Class*” ditulis dengan Positif, Netral dan Negatif. Untuk pilihan auto detect minority class jangan kita pilih. Seperti gambar 11- 13 di bawah ini.



Gambar 11 Smote Negatif



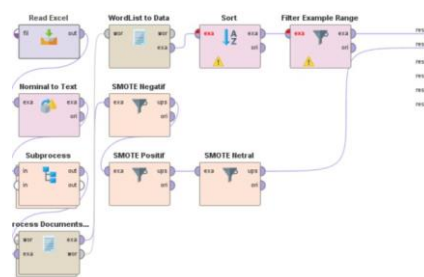
Gambar 12 Smote Netral



Gambar 13 Smote Positif

Setelah menentukan parameter SMOTE UP Sampling yang digunakan

kemudian dilanjut ke proses menyambungkan semua operator halaman process seperti gambar 14 di bawah ini.



Gambar 14 Proses menentukan parameter Smote UpSampling

Setelah semua operator terhubung, kemudian proses selanjutnya yaitu menjalankan semua proses akan didapatkan hasil seperti pada gambar yang merupakan hasil *smote up* sampling dari perintah yang telah dijalankan rinciannya seperti gambar 15 berikut ini :

Row No.	Sentimen	acara	adalah	adil	apak	apar	ajak	akal
1	Positif	0	0	0	0	0	0	0
2	Positif	0	0	0	0	0	0	0
3	Positif	0	0	0	0	0	0	0
4	Positif	0	0	0	0	0	0	0
5	Positif	0	0	0	0	0	0	0
6	Positif	0	0	0	0,258	0	0	0
7	Positif	0	0	0	0	0	0	0
8	Positif	0	0	0	0	0	0	0
9	Positif	0	0	0	0	0	0	0
10	Positif	0	0	0	0	0	0	0
11	Positif	0	0	0	0	0	0	0

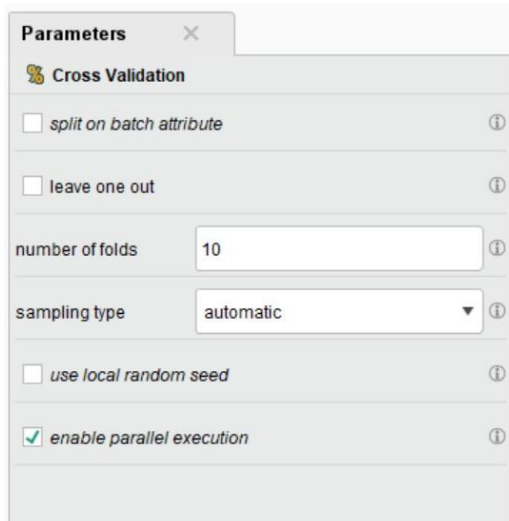
Gambar 15 Hasil smote up sampling

Dari hasil gambar 15 di atas terlihat bahwa sebelumnya data berjumlah 389 data dan setelah dilakukan proses SMOTE data bertambah menjadi 903 Dengan keterangan semua data menjadi seimbang mengikuti data negatif sebagai total jumlah data terbanyak yaitu 301 data, data positif yang sebelumnya sebanyak 68 data menjadi 301 data dan data netral yang sebelumnya 20 data menjadi 301 data.

E. Uji Validitas Data

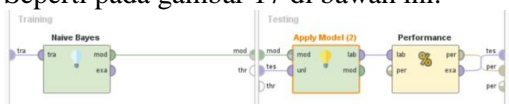
Setelah dilakukan peningkatan kinerja maka berikutnya adalah dengan menggunakan *samplingtype automatic* dan operator *Cross Validation* dengan k=10 fold validation yang ada pada *RapidMiner Studio* untuk proses validasi algoritma, operator ini digunakan untuk menguji akurasi dari data.

Parameter yang digunakan seperti pengaturan default dan pada bagian kolom “number of fold” diisi dengannilai “10” dan sampling type dipilih “automatic”. Seperti pada gambar 16 di bawah ini.



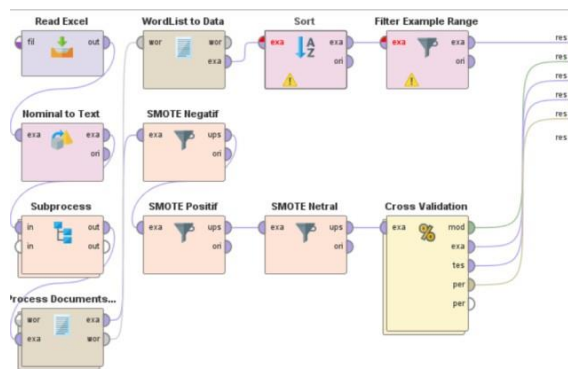
Gambar 16 tampilan dari parameter Cross Validation

Setelah menentukan parameter yang digunakan pada Cross Validation, selanjutnya open operator Cross Validation dengan cara “Klik 2X”. Pada bagian dalam operator Cross Validation digunakan untuk menunjukkan proses dari data uji dan data latih yang dipakai untuk melihat akurasi data dengan menggunakan 2 metode yaitu Naïve Bayes Classification dan Decision Tree pada RapidMiner. Pada tahap ini metode Naïve Bayes Classification menerapkan model data latih yang dihubungkan dengan data uji dengan menggunakan operator “Performance” dan “Apply Model” pada operator Performance dipakai untuk mengevaluasi data yang terdapat pada halaman process data uji. Seperti pada gambar 17 di bawah ini.



Gambar 17 Tampilan process Cross Validation dengan Naïve Bayes

Setelah menentukan parameter dan proses yang terdapat pada operator Cross Validation kemudian semua operator saling dihubungkan.



Gambar 18 Tampilan Cross Validation

Setelah semua operator saling terhubung, maka proses selanjutnya yaitu menjalankan semua proses dengan cara menekan tombol segitiga berwarna biru dan hasil yang didapat seperti pada tabel 9 berikut ini yang merupakan hasil dari Performance vector.

Accuracy : 94.91% +/- 2.83% (micro average:94.91%)

Tabel 1 Hasil akurasi Naïve Bayes Classifier

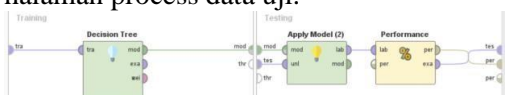
	True Positif	True Netral	True Negatif	Class precision
Pred. Positif	296	0	27	91.64%
Pred. Netral	1	301	14	95.25%
Pred. Negatif	4	0	260	98.48%
Class recall	98.34%	100.00%	86.38%	

Dari hasil perhitungan klasifikasi dengan metode Naïve Bayes Classification diperoleh hasil accuracy: 94.91% +/- 2.83% (micro average: 94.91%). Dari gambar di atas ini memperlihatkan hasil dari Performance Vector yang terdapat pada RapidMiner. Untuk melihat hasil pada data view dapat dilihat seperti pada gambar 19 yang merupakan hasil pelabelan pada data view.

Row No.	Sentimen	predictionS...	confidence...	confidence...	confidence...	acara	adalah	adi
1	Positif	Negatif	0	0	1	0	0	0
2	Negatif	Negatif	0	0	1	0	0	0
3	Negatif	Negatif	0	0	1	0	0.212	0
4	Negatif	Negatif	0	0	1	0	0	0
5	Negatif	Negatif	0	0	1	0	0	0
6	Negatif	Positif	0	1	0	0	0	0
7	Negatif	Negatif	0	0	1	0	0	0
8	Netral	Netral	1	0	0	0	0	0
9	Netral	Netral	1	0	0	0	0	0
10	Netral	Netral	1	0	0	0	0	0
11	Negatif	Negatif	0	0	1	0	0	0

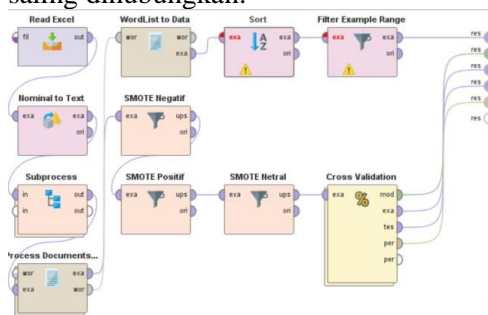
Gambar 19 Hasil pada data view dari metode NaiveBayes

Selanjutnya tahap ini menggunakan metode *Decision Tree* menerapkan model data latih yang dihubungkan dengan data uji dengan memakai operator “Performance” dan “Apply Model” pada operator *Performance* dipakai untuk mengevaluasi data yang terdapat pada halaman process data uji.



Gambar 20 Tampilan process Cross Validation dengan Decision Tree

Setelah menentukan parameter dan proses yang terdapat pada operator *Cross Validation* kemudian semua operator saling dihubungkan.



Gambar 21 Tampilan proses Cross Validation

Setelah semua operator saling terhubung, maka proses selanjutnya yaitu menjalankan semua proses dengan cara menekan tombol segitiga berwarna biru dan hasil yang didapat seperti pada tabel 10 berikut ini yang merupakan hasil dari *Performance vector*.

Accuracy : 62.57% +/- 2.19%
(micro average:62.57%)

Tabel 2 Tabel Hasil akurasi dengan Decision Tree

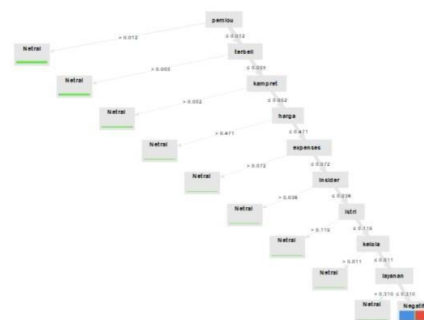
	True Positif	True Netral	True Negatif	Class precision
Pred. Positif	87	14	89	52.62%
Pred. Netral	2	271	5	71.00%
Pred. Negatif	212	16	207	47.59%
Class recall	28.90%	90.03%	68.77%	

Dari hasil perhitungan klasifikasi dengan menggunakan metode *Decision Tree* diperoleh hasil *accuracy*: 62.57% +/- 2.19% (*micro average*: 62.57%). Dari gambar di atas ini memperlihatkan hasil dari *Performance Vector* yang terdapat pada *Rapid Miner*. Untuk melihat hasil pada data view dapat dilihat seperti pada gambar 22 yang merupakan hasil pelabelan pada data view.

Row No.	Sentimen	predictionS...	confidence...	confidence...	confidence...	acara	adalah	adi	agak
1	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
2	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
3	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
4	Negatif	Positif	0	1	0	0	0	0	0
5	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
6	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
7	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
8	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
9	Negatif	Negatif	0.456	0.087	0.456	0	0	0	0
10	Netral	Negatif	0.456	0.087	0.456	0	0	0	0
11	Netral	Negatif	0.456	0.087	0.456	0	0	0	0

Gambar 22 Hasil pelabelan pada data view dari Decision Tree

Dan dari hasil metode *Decision Tree* terdapat juga hasil gambar pohon keputusan. Berikut gambar 23 yang merupakan gambar dari hasil analisis menggunakan metode *Decision Tree*.



Gambar 23 Hasil Tree dari metode Decision Tree

Berdasarkan hasil dari pengujian menggunakan algoritma *Naïve Bayes* dan *Decision Tree* dapat dilihat bahwa nilai akurasi 94.91% merupakan akurasi tertinggi yang didapatkan oleh algoritma *Naïve Bayes*. Sedangkan algoritma *Decision Tree* hanya mendapatkan nilai akurasi 62.57%.

Hasil sentimen negatif dengan jumlah 301 yang lebih banyak dibandingkan sentimen positif dengan jumlah 68 dan netral dengan jumlah 20 tersebut dapat disimpulkan bahwa BBM naik di Indonesia belum dapat diterima oleh masyarakat Indonesia pada mediasosial Twitter.

KEPUSTAKAAN

- [1] B. C. Sandy, D. Manongga and A. Iriani, "Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak (Bbm) Pada Media Online," *Prosiding Semmau*, 2015.
- [2] E. F. U. Latifah, "Perbandingan Kinerja Machine Learning Berbasis Algoritma Support Vector Machine Dan Naive Bayes," 2018.
- [3] B. M. Pintoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier," 2018.
- [4] Y. S. Mahardhika and E. Zuliarso, "Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier," 2018.
- [5] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree-Knn-Naïve Bayes," 2020.
- [6] N. T. Romadloni, I. Santoso and S. Budilaksono, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line," 2019.
- [7] T. S. Boenga NurCitra, "Dampak Pembelajaran Jarak Jauh Dan Physical Distancing PADA Tingkat Kecemasan Mahasiswa Fakultas Kedokteran Universitas Pembangunan Nasional "Veteran" Jakarta," *Journal of Bourneo Holistic Health*, vol. 3, p. 2, 2020.
- [8] S. K. Dzikra Rafik Putra, "Sistem Pendukung Keputusan Untuk Diagnosis Banding Gangguan Somatoform Berbasis PPDGJ III," *JURNAL FASILKOM*, vol. Volume 10 No. 2, p. 2, 2020.
- [9] T. V. d. P. P. O. S. M. Rizkiana Prima Rahmadina, "Visualisasi Data Jumlah Penderita Diare yang Dilayani dan Jumlah Desa/Kelurahan yang Melaksanakan Sanitasi Total Berbasis Masyarakat (STBM) Tahun 2016 dan 2017 di Indonesia Menggunakan Software R-Studio," pp. 2-8, 2019.
- [10] N. M. M. Komang Trisnadewi, "Pembelajaran Daring di Masa Pandemi Covid-19," in *COVID-19: Perspektif Pendidikan*, Bali, Yayasan Kita Menulis, 2020, pp. 48-49.