

Penerapan Metode Adaptive Boosting Pada Analisis Sentimen Kenaikan BBM Pertamina

Aditya Wahyu Nur Faizi^{*1}, Kristiawan Nugroho²

^{1,2}Unisbank Semarang; Jalan Tri Lomba Juang Semarang

e-mail: ^{*1}adityawahyu659@gmail.com, ²kristiawan@edu.unisbank.ac.id

Abstrak

Masyarakat selalu menentang kenaikan harga BBM. Selain itu, emosi negatif semakin terlihat karena kisah di media sosial sering dipenuhi dengan provokasi yang berlebihan. Dianggap bahwa pengalihan subsidi BBM oleh pemerintah tidak menguntungkan rakyat. Kenaikan harga BBM juga memicu interaksi dan percakapan warganet Indonesia di media sosial. Perbincangan isu kenaikan BBM di media sosial terdapat 403.700 perbincangan dengan 2,2 juta interaksi antar pengguna media sosial. Data penelitian mencakup 560 tweet. Ini dibagi menjadi dua, 500 untuk data latihan dan 60 untuk data uji, yang disimpan dalam format xlsx. Algoritma yang di gunakan adalah AdaBoost dengan klasifikasi sentimen positif atau negatif. Studi ini menghasilkan algoritma AdaBoost memiliki kemampuan untuk mengkategorikan tweet kenaikan BBM Pertamina ke dalam kelas bersentimen positif atau negatif dengan akurasi sebesar 86,8%.

Kata kunci— Analisis Sentimen, AdaBoost, BBM

Abstract

The increase in fuel prices always gets resistance from the public. Moreover, the narrative on social media is often embellished with excessive provocation so that negative sentiment is increasingly visible. The option of diverting fuel subsidies by the government is considered not to be in favor of the interests of the people. The increase in fuel prices also sparked interactions and conversations among Indonesian netizens on social media. There were 403,700 discussions on the fuel price hike issue on social media with 2.2 million interactions between social media users. The data used in this research are 560 tweets. The data is divided into 2, namely 500 training data from tweet data and 60 test data from tweet data stored in xlsx format. The algorithm used is AdaBoost with positive, negative or neutral sentiment classification. The results of the study show that the AdaBoost algorithm can be used to classify a tweet about Pertamina's fuel increase into positive, negative or neutral sentiment classes with an accuracy of 86.8%.

Keywords— AdaBoost, BBM, Sentiment Analysis

1. PENDAHULUAN

Gaya bahasa sangat penting untuk mengartikan maksud dan tujuan opini seseorang melaluinya, baik secara langsung maupun tidak langsung. Menganalisis pendapat seseorang menjadi sulit karena banyak kalimat sindiran yang disampaikan secara langsung tanpa basa-basi, sedangkan orang lain menggunakan pendekatan tidak langsung untuk menyinggung poin yang ingin disampaikan untuk berbagai alasan. Opini sarkasme juga dikenal sebagai komentar yang mengandung sindiran yang pahit dan celaan yang getir. Sarkasmus adalah cara berbicara yang menyinggung orang dengan menggunakan kata-kata yang menyakitkan. Sarkasme yang berbeda makna adalah salah satu jenis yang digunakan. Dalam hal ini, kalimat positif yang memiliki arti negatif digunakan. Seringkali, pendapat sarkasme ini disampaikan secara tidak langsung melalui Twitter.

Twitter adalah tempat yang bagus untuk berbagi ide, mengumpulkan informasi, menginspirasi, atau melihat apa yang dilakukan teman Anda. Pada 21 Maret 2022, Twitter, salah satu platform microblogging terkemuka, genap berusia 16 tahun. Menurut Maulana (2016), media sosial di seluruh dunia memiliki 332 juta pengguna bulanan, dan 500 juta *tweet*, atau kicauan, dikirim setiap hari, dan 200 miliar kicauan dikirim setiap tahunnya [1]. Banyak pengguna Twitter memposting pendapat dan ekspresinya tentang barang, layanan, masalah politik, atau hal lain yang menjadi viral. Twitter digunakan oleh pemerintah di seluruh dunia untuk lebih dekat dengan

masyarakat. Hal ini dapat memberikan wawasan yang mendalam tentang aspirasi masyarakat [2].

Menggunakan media sosial seperti Twitter, pemerintah Indonesia berinteraksi dengan masyarakat. Hal ini didasarkan pada rekomendasi Menpan tentang penyebarluasan informasi melalui media sosial, khususnya Twitter, bagi aparatur sipil negara (ASN) [3]. Indonesia berada di antara negara-negara dengan jumlah pengguna Twitter tertinggi di dunia. Laporan We Are Social melaporkan bahwa jumlah pengguna Twitter Indonesia pada 2022 akan mencapai 18,45 juta, atau 4,23% dari 436 juta pengguna Twitter global. Ini adalah peningkatan 31,3% dari 14,05 juta pengguna tahun sebelumnya [4].

Masyarakat selalu menentang kenaikan harga BBM. Selain itu, emosi negatif semakin terlihat karena kisah di media sosial seting yang penuh dengan provokasi. Dianggap bahwa pengalihan subsidi BBM oleh pemerintah tidak menguntungkan rakyat. Warganet Indonesia juga berbicara di media sosial tentang kenaikan harga BBM. Ada 403.700 diskusi tentang masalah peningkatan BBM di media sosial, dan 2,2 juta orang berinteraksi [5].

Untuk melacak hasil umpan balik masyarakat terhadap kebijakan, Anda dapat menggunakan data Twitter yang berupa keluhan, informasi, atau saran yang terkena dampak kenaikan BBM pada Peraminan. Proses klasifikasi terhadap opini masyarakat digunakan untuk menghasilkan hasil feedback yang jelas. Data opini masyarakat ini dapat diolah dengan algoritma klasifikasi melalui pengajaran mesin dengan metode adaptive booster (AdaBoost). Pada dasarnya, algoritma AdaBoost bekerja secara iteratif untuk menghasilkan pengklasifikasi dan kemudian digabungkan untuk membuat klasifikasi utama.. Studi (Utami & Wahono, 2015) menemukan bahwa metode AdaBoost meningkatkan akurasi analisis sentimen review restoran sebesar 29,5% dari 70 % menjadi 99,5%. Studi ini menggunakan metode ini untuk analisis sentimen kenaikan BBM Pertamina di Twitter [6].

Penelitian sebelumnya terkait dengan analisis sentimen menggunakan algoritma Adaboost dilakukan oleh Qadrini dkk (2021) dengan hasil metode klasifikasi *decision tree* dan AdaBoost memiliki hasil akurasi yang baik yaitu sebesar 94% dan 95% [7]. Penelitian lainnya oleh Byna & Basit (2020) dengan hasil pengujian dengan memberikan akurasi 0,976 dan setelah di optimasi dengan AdaBoost memberikan akurasi 0,981 [8]. Menurut Tanti et al. (2021), penelitian dengan hasil metode Chi-square dan AdaBoost dapat meningkatkan nilai recall, sensitivitas, dan TPR. Oleh karena itu, penerapan metode Chi-square dan AdaBoost dapat meningkatkan kinerja klasifikasi kelas minoritas [9]. Tujuan penelitian berikutnya adalah untuk mengevaluasi nilai akurasi algoritma AdaBoost saat mengklasifikasikan *tweet* kenaikan BBM untuk mengetahui apakah *tweet* tersebut termasuk dalam kelas bersentimen positif atau negatif. 560 *tweet* digunakan dalam penelitian ini. Data dibagi menjadi dua, yaitu 500 untuk data latihan dan 60 untuk data uji, dan disimpan dalam format *xlsx*.

2. METODE PENELITIAN

2.1. Pengumpulan Data *Tweet*

Data yang diperoleh berasal dari data yang diambil secara langsung dari *tweet* yang terkait dengan kenaikan BBM Pertamina dari Twitter. Untuk mengunduh *tweet* kenaikan BBM Pertamina dengan bahasa R, perlu memiliki *library* Twitter, kata sandi API, dan akses token dari Twitter yang telah didaftarkan. *Script crawling* yang digunakan untuk mendapatkan data ini yang diperlihatkan seperti gambar 1.

```
tweet <- search_tweets(q = "kenaikan bbm", n = 1000)
df <- apply(tweet, 2, as.character)
write.xlsx(df, file = "myworkbook2.xlsx", sheetName = "DATA", append = TRUE)
```

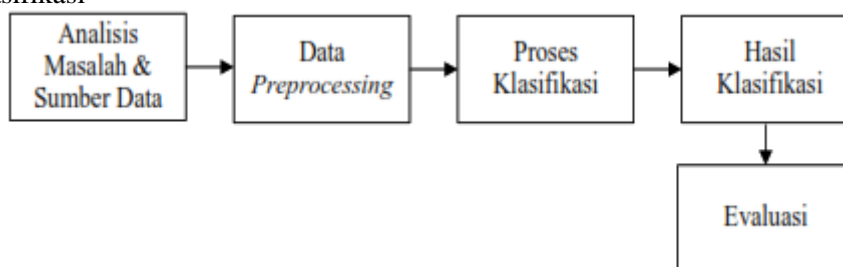
Gambar 1. *Script Crawling*

Hasil *crawling script* menghasilkan data *tweet* kenaikan BBM Pertamina sebanyak 560 *tweet*. Selanjutnya, data *tweet* kenaikan BBM Pertamina dibagi menjadi dua kelompok: 500 data *tweet* latihan untuk melatih model AdaBoost, dan 60 data *tweet* uji untuk menguji akurasi model AdaBoost..

	created_at	id	id_str	full_text	truncated	display
917	2022-12-08 15:14:06	1.600766e+18	160076...	#InformasiTerkini https://t.co/Tw3vgDvoRX	FALSE	
59	2022-12-13 01:28:59	1.602370e+18	160236...	#MakinHits itu kalo elo ikutan demo kenaikan BBM tapi ...	FALSE	
739	2022-12-10 07:29:03	1.601373e+18	160137...	#MakinHits itu kalo elo ikutan demo kenaikan BBM tapi ...	FALSE	
762	2022-12-09 19:38:10	1.601194e+18	160119...	..kepada jasa layanan keagamaan yang tidak memiliki pe...	FALSE	
173	2022-12-12 15:23:05	1.602217e+18	160221...	@aanlegitoto @Paltiwist Saya Kira wajar saja klo Relaw...	FALSE	
887	2022-12-08 19:26:57	1.600829e+18	160082...	@afri_krisni99 @RDriarti @tanyakanri Selain itu, termas...	FALSE	
274	2022-12-11 18:11:16	1.601897e+18	160189...	@akbarfaizal68 Kovit ilang, efek BBM naik msh sangat di...	FALSE	
505	2022-12-10 19:34:52	1.601556e+18	160155...	@AmanSur34526860 @watizsaras98 @Miduk17 Blm mer...	FALSE	
749	2022-12-09 21:34:44	1.601224e+18	160122...	@andre_rosiade @Gerindra @prabowo Bung @andre_r...	FALSE	
748	2022-12-09 21:53:57	1.601229e+18	160122...	@andre_rosiade @prabowo @jokowi @Gerindra Bung ...	FALSE	
925	2022-12-08 14:04:06	1.600748e+18	160074...	@andry19633 @RamlilRizal Bahaya lain akibat Hutang L...	FALSE	
7	2022-12-13 11:23:19	1.602520e+18	160251...	@Anggaradimas69 Dukung pemerintah minimalisir dam...	FALSE	
923	2022-12-08 14:48:06	1.600759e+18	160075...	@bachrum_achmadi @joserasyid Ebong semakin bahagi...	FALSE	
900	2022-12-08 16:43:13	1.600788e+18	160078...	@bachrum_achmadi Alhamdulillah <U+0001F917> smo...	FALSE	
837	2022-12-09 09:58:59	1.601049e+18	160104...	@bachrum_achmadi BBM nom subsidi peruntukan siapa...	FALSE	
918	2022-12-08 15:11:07	1.600765e+18	160076...	@bachrum_achmadi Kadrun kemana mana modal jalan k...	FALSE	
871	2022-12-08 21:29:22	1.600860e+18	160086...	@bachrum_achmadi Ya bahagia lah. Mau BBM naik ya bi...	FALSE	
730	2022-12-10 13:43:59	1.601468e+18	160146...	@Bangsaindo_ Dukung pemerintah minimalisir dampak ...	FALSE	
192	2022-12-12 13:23:20	1.602187e+18	160218...	@Bangsaindo_ Dukung pemerintah minimalisir dampak ...	FALSE	
886	2022-12-08 19:38:24	1.600832e+18	160083...	@BEMUI_Official Woiii... @BEMUI_Official, mmg nya m...	FALSE	
198	2022-12-12 12:59:17	1.602181e+18	160218...	@Cintada16 Setelah ahok tak terima bulian disalahkan ...	FALSE	
68	2022-12-12 22:12:54	1.602321e+18	160232...	@Darmaputrackr @Parland99 @Hasbil Lbs @prastow Yq...	FALSE	

Gambar 2. Data *Tweet* Kenaikan BBM Pertamina

2.2. Tahap Klasifikasi



Gambar 3. Tahap Klasifikasi

Gambar 3 menunjukkan cara penyelesaian klasifikasi *tweet* kenaikan BBM Pertamina menggunakan algoritma AdaBoost:

A. Analisis Masalah dan Sumber Data

Masalah dengan klasifikasi *tweet* kenaikan BBM Pertamina adalah:

- Sejumlah besar pengguna Twitter mengungkapkan pendapat, ekspresi, atau pendapat mereka tentang kenaikan BBM Pertamina. Opini ini dapat digunakan untuk mencari informasi.
- Belum ada alat yang dapat membedakan *tweet* kenaikan BBM Pertamina menjadi positif atau negatif.

Untuk mengatasi masalah *tweet* kenaikan BBM Pertamina, algoritma *text mining* AdaBoost harus digunakan untuk mengklasifikasikan *tweet* tersebut. Sumber data penelitian ini berasal dari *tweet* yang berkaitan dengan kenaikan BBM Pertamina. Data yang dikumpulkan dalam format Microsoft Excel *xlsx*. Setelah data dikumpulkan, langkah selanjutnya adalah mengidentifikasi klasifikasi data secara manual. Sebanyak 560 data *tweet* kenaikan BBM Pertamina dikumpulkan, termasuk 500 *tweet* kenaikan BBM Pertamina data latihan dan 60 *tweet* kenaikan BBM Pertamina data uji.

Saat data *tweet* kenaikan BBM Pertamina melewati tahap *preprocessing*, kemudian akan diklasifikasikan. Dalam penelitian ini, *tweet* kenaikan BBM Pertamina akan diklasifikasikan menjadi dua kategori: positif atau negatif.

B. Data Preprocessing

Pada titik ini, *tweet* yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks, termasuk *case folding*, *tokenizing*, dan *removal of stopwords*. Tujuan pemrosesan teks adalah untuk membuat data yang dikumpulkan lebih terstruktur, sehingga proses pengolahan data menjadi lebih mudah.

C. Proses Klasifikasi

Setelah itu, *tweet* kenaikan BBM Pertamina akan diklasifikasikan, yang akan dilakukan

menggunakan metode algoritma AdaBoost. Ini akan menghitung kemungkinan setiap *tweet* kenaikan BBM Pertamina.

D. Hasil Klasifikasi

Pada tahap ini, harus diketahui hasil klasifikasi menggunakan metode algoritma AdaBoost dari proses klasifikasi, apakah *tweet* kenaikan BBM Pertamina termasuk dalam kategori positif atau negatif.

E. Evaluasi

Pada langkah ini, hasil klasifikasi dievaluasi. Tujuan dari melakukan evaluasi adalah untuk menyesuaikan model yang diperoleh agar sesuai dengan tujuan yang diinginkan. Pada titik ini, model yang dibuat adalah algoritma AdaBoost. Untuk mengetahui nilai akurasi yang diperoleh, algoritma ini akan diuji menggunakan diagram *confusion matrix*. Salah satu cara untuk menilai kinerja metode klasifikasi adalah dengan menggunakan *confusion matrix*, yang pada dasarnya mengandung informasi yang membandingkan hasil klasifikasi sistem dengan hasil klasifikasi yang seharusnya. Pengukuran klasifikasi menggunakan *confusion matrix*, terdapat beberapa langkah yang digunakan yaitu

a. Hitung *True Positive* (TP)

Data yang diprediksi +; *tweet* dengan klasifikasi + juga diklasifikasikan + dari model AdaBoost.

b. Hitung *True Negative* (TN)

Data yang diprediksi -; *tweet* dengan klasifikasi - juga diklasifikasikan - dari model AdaBoost.

c. Hitung *False Positive* (FP)

Data yang diprediksi +; *tweet* dengan klasifikasi + tetapi diklasifikasikan - dari model AdaBoost.

d. Hitung *False Negative* (FN)

Data yang diprediksi -; *tweet* dengan klasifikasi - tetapi diklasifikasikan + dari model AdaBoost.

e. Hitung Akurasi

Setelah mengetahui TP, TN, FP, dan FN, langkah berikutnya adalah menghitung akurasi atau rasio prediksi benar (positif dan negatif) dari metode AdaBoost untuk klasifikasi, yang dilakukan dengan menggunakan rumus akurasi.
$$= \frac{TP+TN}{TP+TN+FP+FN} \times 100 \%$$

3. HASIL DAN PEMBAHASAN

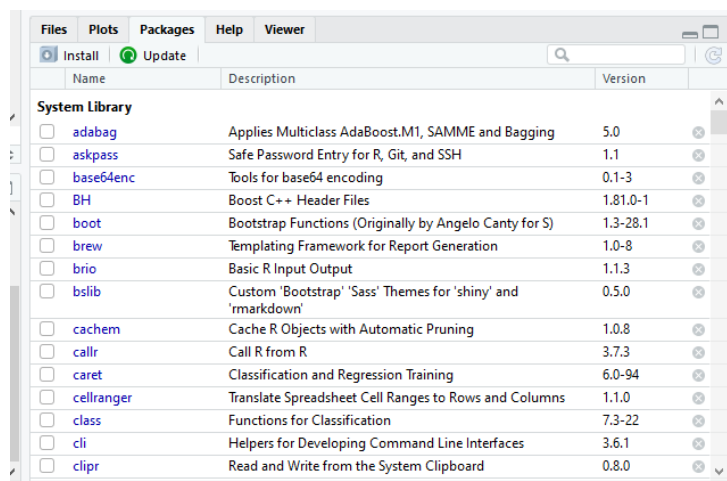
3.1. Instal Packages

Di bahasa R, *packages* adalah kumpulan fungsi R, data, dan kode yang dikompilasi. Lokasi tempat paket disimpan disebut *library*. Untuk bisa menjalankan pemrosesan data *tweet* kenaikan BBM Pertamina positif atau negatif diperlukan *packages* yang berisi fungsi–fungsi yang dibutuhkan untuk proses klasifikasi data *tweet* positif atau negatif. Untuk melakukan pemanggilan *packages* maka harus dilakukan penulisan *script* seperti pada gambar 4.

```
install.packages("readxl")
install.packages("tm")
install.packages("dplyr")
install.packages("caret")
install.packages("wordcloud")
install.packages("adabag")
```

Gambar 4. *Script Instalasi Packages*

Gambar 4 akan memanggil *packages* yang diperlukan yaitu readxl, tm, dplyr, caret, wordcloud, dan adabag. *Packages* yang telah berhasil diinstal dapat dilihat pada menu *packages* seperti yang ada pada gambar 5.

Gambar 5. Menu *Packages*

3.2. Library Packages

Tahap ini adalah memuat semua pustaka yang diperlukan (*packages*), melakukan pemuatan *packages* yang telah diinstal. *Library packages* berisi kumpulan fungsi R, data, dan kode yang digunakan dalam pemrosesan *tweet* kenaikan BBM Pertamina positif atau negatif seperti gambar 6.

```
library("readxl")
library("tm")
library("dplyr")
library("caret")
library("wordcloud")
library("adabag")
```

Gambar 6. *Library Packages*

Gambar 6 menjelaskan *script* pemuatan *packages* yang akan digunakan yaitu readxl, tm, dplyr, caret, wordcloud, dan adabag.

3.3. Persiapan Data

Pada titik ini, data *tweet* kenaikan BBM Pertamina akan dipersiapkan untuk diproses. Sebanyak 560 data tweet akan digunakan, dan rinciannya disajikan dalam tabel 1.

Tabel 1. Data *Tweet*

Kategori	Text
NEGATIF	Kompensasi kenaikan BBM kemaren yang katanya ojol salah satu profesi prioritas aja kayak dapet apa2..
NEGATIF	Setelah pesta meriah yg menurut Yaqut membawa kebahagiaan semua lini. Rakyat diberi hadiah kenaikan BBM.
POSITIF	Dukung pemerintah minimalisir dampak kenaikan BBM #BantuanBBMUntukRakyat #BLTBBMTepatSasaran #SubsidiTepatSasaran #CerdasPakaiBBM #BijakPakaiBBM #BLTBMBantuMasyarakat #BansosBBMRingankanRakyat #SukseskanKebijakanPemerintah https://t.co/1WYM8ZHc8r
POSITIF	Nur Aziz Serahkan Bantuan BLT BBM Dampak Inflasi Kenaikan BBM Kepada Ratusan Driver
POSITIF	BLT Alih Subsidi BBM telah diterima oleh masyarakat yang paling membutuhkan Semoga bansos ini akan membantu meringankan sert...
NEGATIF	Kita tetap kritis kok, Mas...Contoh masalah kenaikan BBM Subsidi yg lalu, sy termasuk yg keberatan jk harga BBM Subsidi naik, lbh baik dibatasi hanya utk roda 2, LCGC, angkutan umum, angkutan barang...Kemudian ternyata tetap naik, ya itu keputusan Pemerintah bersama DPR
NEGATIF	Ahlan wa Sahlan Ramadhan dan Kenaikan BBM semoga semakin sabarrrrrr
NEGATIF	katanya ekonom kelas dunia tp kok sprt rakyat miskin yg ngandalin hutang

	terus. mau di bawa kemana negeri...
NEGATIF	Cuma Di Jaman @Jokowi Banyak Yang Mendapatkan GA. Ada Gubernur, Bupati, Walikota Hingga Letkol Hasil GA. Namun Di Jaman...
POSITIF	Dukung pemerintah minimalisir dampak kenaikan BBM #BantuanBBMUntukRakyat #BLTBBMTepatSasaran #SubsidiTepatSasaran #CerdasPakaiBBM #BijakPakaiBBM #BLTBBMBantuMasyarakat #BansosBBMRingankanRakyat #SukseskanKebijakanPemerintah
NEGATIF	untung nya dulu stafsus2 di kabinet Pres SBY paham akan prosedur..shg walkot ini bebas bicara yg bernada prov...
POSITIF	Kenaikan harga bbm merupakan hal wajar Mengingat yg mengalami kenaikan adalah bbm non subsidi yg harganya fluktuatif mengiku...
POSITIF	Kenaikan harga BBM itu sudah tepat. Harga BBM di Indonesia pun itu bukan harga keekonomian. Dan naiknya harga BBM itu mengi...
POSITIF	Salah satu alasan pertumbuhan bisnis UMKM melambat yakni adalah kenaikan harga BBM bersubsidi. #KinerjaBRIQ3
POSITIF	Intinya, perkembangan UMKM bergantung pada sektor lainnya, seperti bahan baku, kenaikan harga BBM, faktor cuaca, dan lain sebagainya.#KinerjaBRIQ3
NEGATIF	Setelah pesta meriah yg menurut Yaqut membawa kebahagiaan semua lini. Rakyat diberi hadiah kenaikan BBM.
NEGATIF	Ahlan wa Sahlan Ramadhan dan Kenaikan BBM semoga semakin sabarrrrrr #TolakBBMNaik ##tolakBBMnaik
NEGATIF	Yg teriak2 itu jokower sejati,,sambate sundul langit Kemarin ada yg demo tolak kenaikan BBM para celeng nuduh pendemo mental subsidi. Skrg udah pada kena batunya,mewek jamaah Padahal luar jawa udah lama tersiksa susahny dapat SOLAR.
POSITIF	Kenaikan harga bbm merupakan hal wajar Mengingat yg mengalami kenaikan adalah bbm non subsidi yg harganya fluktuatif mengikuti dolar

3.4. Membaca Data *Tweet*

```
df<- read_excel("E:/Rstudio/adaboost/data.xlsx")
glimpse(df)
```

Gambar 7. Membaca Data *Tweet* Excel

Gambar 7 menunjukkan *script* yang digunakan untuk memuat data tweet kenaikan BBM Pertamina ke dalam file Excel.xlsx. Dalam penelitian ini, data yang digunakan adalah kumpulan data positif atau negatif yang telah diberi kategori untuk digunakan dalam proses pelatihan dan pengujian. Data *tweet* positif atau negatif yang digunakan dalam penelitian ini adalah data yang diambil melalui teknik *crawling*.

Gambar 8 menunjukkan hasil membaca data Excel data.xlsx, yang berisi 560 data *tweet* kenaikan BBM Pertamina. Variabel KATEGORI menunjukkan kategori *tweet* kenaikan BBM Pertamina, yang dapat positif atau negatif, sedangkan variabel TEXT menunjukkan kumpulan *tweet* kenaikan BBM Pertamina

	Kategori	TEXT
1	NEGATIF	@PartaiSocmed Hahaha... Kompensasi kenaikan BBM ke...
2	POSITIF	BLT UMKM untuk 12 Juta Pelaku usaha merupakan langk...
3	NEGATIF	Setelah pesta meriah yg menurut Yaqut membawa kebah...
4	POSITIF	@Anggaradimas69 Dukung pemerintah minimalisir dam...
5	POSITIF	Nur Aziz Serahkan Bantuan BLT BBM Dampak Inflasi Kena...
6	POSITIF	RT @catatankaqihati: Pemerintah melalui KemenKopUK...
7	POSITIF	RT @AlshaibaShihab: Pencairan BLT UMKM 2022 kembali...
8	POSITIF	RT @forjakeu: BLT Alih Subsidi BBM telah diterima oleh ...
9	POSITIF	RT @riego_see: Pengalihan subsidi dan kompensasi BB...
10	NEGATIF	@hasyimmah Kita tetap kritis kok, Mas...δ<U+009F> <U...
11	POSITIF	Pengalihan subsidi dan kompensasi BBM menjadi BLT BB...
12	NEGATIF	Rambak Dong mohon pengertiannya ya. Margin kami se...

Showing 1 to 12 of 560 entries

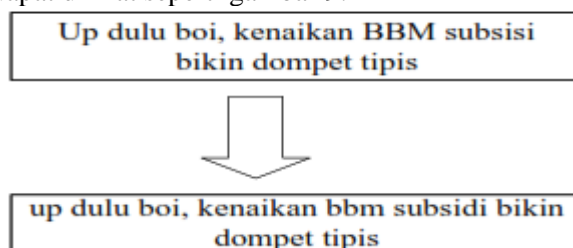
Gambar 8. Hasil Membaca Data *Tweet*

3.5. Text Processing

Proses data *text preprocessing* yang dilakukan yaitu

a. Case Folding

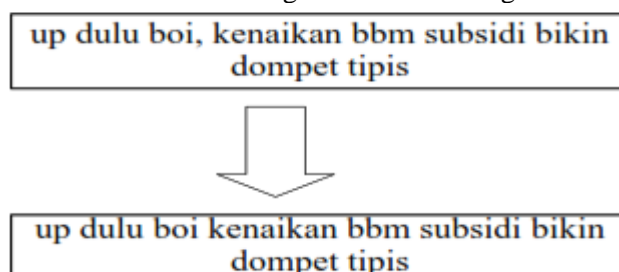
Pada titik ini, semua huruf akan diubah menjadi huruf kecil atau *lowercase*. Dimulai dengan memeriksa setiap karakter dalam *tweet* kenaikan BBM Pertamina dari awal hingga akhir. Jika ada huruf kapital atau huruf kecil, karakter tersebut akan diubah menjadi huruf kecil. Proses penyiapan *case folding* dapat dilihat seperti gambar 9.



Gambar 9 Case Folding

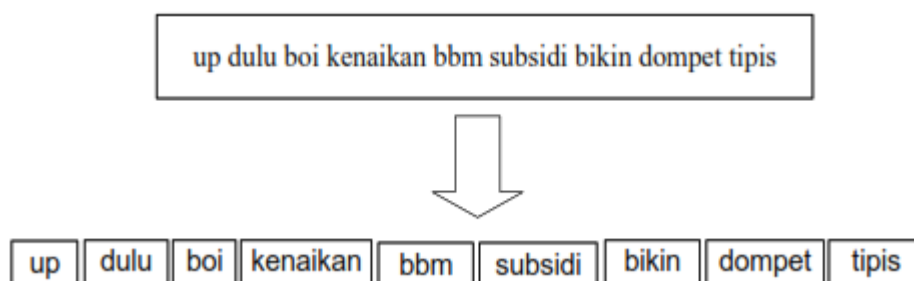
b. Tokenizing

Pada langkah ini, *tweet* kenaikan BBM Pertamina dipotong menjadi token dan menghapus semua karakter, angka, dan tanda baca tertentu seperti @, hashtag (#), titik (.), koma (,), dan tanda baca lainnya. Semua elemen ini dibuang karena tidak mempengaruhi nilai klasifikasi. Untuk memulai proses penghapusan karakter khusus dari setiap *tweet*, pertama-tama perlu memeriksa apakah ada karakter khusus, angka, atau tanda baca dalam *tweet* tersebut; jika ada, karakter tersebut akan secara otomatis dihilangkan. Proses ini digambarkan pada gambar 10.

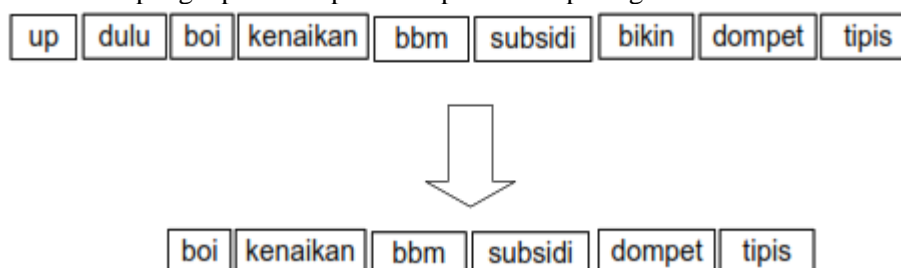


Gambar 10. Proses Penghapusan Karakter Khusus

Setelah menghapus karakter tertentu, proses selanjutnya adalah memotong atau memisahkan *tweet* menjadi kata yang terpisah oleh tanda baca titik (.), koma (,), dan spasi. Setelah mengecek *tweet*, jika ada tanda baca titik (.), koma (,), dan spasi, *tweet* akan dipotong atau dipisahkan berdasarkan tanda baca tersebut. Tokenisasi dapat dilihat pada gambar 11.

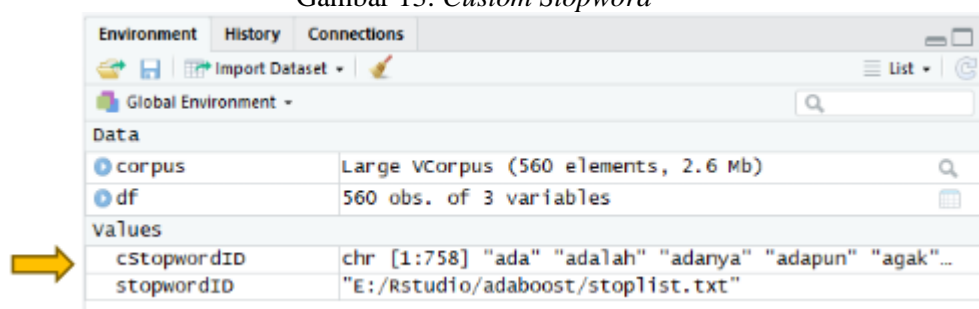
Gambar 11. *Tokenizing*c. *Stopword Removal*

Setelah *tokenizing*, *stopword removal* adalah tahap berikutnya. *Stopword* adalah kosakata yang bukan merupakan kata khusus yang ditemukan dalam dokumen. Misalnya, kata-kata seperti "di", "oleh", "pada", "sebuah", "karena", dan sebagainya. Proses menghapus atau menghilangkan kata-kata umum seperti kata sambung, kata depan, kata ganti, dan kata-kata yang tidak memiliki makna disebut sebagai tahap penghapusan *stopword*. Proses dimulai dengan pencocokan database *stopword*; jika ada kata-kata dalam database, kata tersebut akan dihilangkan. Proses penghapusan *stopword* dapat dilihat pada gambar 12.

Gambar 12. *Stopword Removal*

Untuk menghapus kata sambung, harus digunakan *stopword* dari luar RStudio karena RStudio tidak menyediakan *stopword* dalam format bahasa Indonesia, seperti yang ditunjukkan pada gambar 13.

```
stopwordID <- "E:/Rstudio/adaboost/stoplist.txt"
cStopwordID <- readLines(stopwordID, warn=FALSE)
```

Gambar 13. *Custom Stopword*Gambar 14. Hasil *Custom Stopword*

758 daftar kata termasuk dalam *custom_stopwords* yang diimport, seperti yang ditunjukkan pada gambar 14. Jika ada *tweet* yang menunjukkan kenaikan BBM Pertamina yang positif atau negatif, daftar kata ini akan dihapus. Gambar 15 menunjukkan cara menjalankan tahap *text processing script*.

```
corpus.clean <- tm_map(corpus, content_transformer(tolower))
corpus.clean <- tm_map(corpus.clean, removePunctuation)
corpus.clean <- tm_map(corpus.clean, removeNumbers)
corpus.clean <- tm_map(corpus.clean, removeWords, cStopwordID)
corpus.clean <- tm_map(corpus.clean, stripWhitespace)
```

Gambar 15. *Script Text Processing*

Data *tweet* akan dibagi menjadi data latihan dan uji. Dari 560 *tweet*, 500 akan digunakan sebagai data latihan, dan 60 akan digunakan sebagai data uji.

```
df.train <- df[1:500,]
df.test <- df[501:560,]
dtm.train <- dtm[1:500,]
dtm.test <- dtm[501:560,]
dim(dtm.train)
corpus.clean.train <- corpus.clean[1:500]
corpus.clean.test <- corpus.clean[501:560]
```

Gambar 16. Script Pembagian Data

3.6. Klasifikasi AdaBoost

Menggunakan fungsi AdaBoost dari paket "adabag" dan "caret" untuk melatih model data. Script model AdaBoost dapat ditemukan di gambar 17.

```
model.adaboost <- boosting(quality~., data=corpus.clean.train,
mfinal=5, control=rpart.control(maxdepth=1),
coeflearn='Freund')
pred = predict(model.adaboost, corpus.clean.test)
conf.mat <- table("Predictions" = pred, Actual = df.test$class)
conf.mat <- confusionMatrix(conf.mat)
conf.mat
```

Gambar 17. Model AdaBoost

```
[1] NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF POSITIF
[11] NEGATIF NEGATIF NEGATIF POSITIF NEGATIF NEGATIF NEGATIF POSITIF NEGATIF NEGATIF
[21] NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF POSITIF NEGATIF NEGATIF NEGATIF NEGATIF
[31] NEGATIF NEGATIF POSITIF POSITIF POSITIF NEGATIF POSITIF NEGATIF NEGATIF POSITIF
[41] NEGATIF NEGATIF POSITIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF
[51] POSITIF NEGATIF POSITIF NEGATIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF
[61] POSITIF
```

Gambar 17. Hasil Klasifikasi AdaBoost

Gambar 17 menunjukkan hasil klasifikasi data uji algoritma AdaBoost. Dari pengujian 60 data *tweet*, algoritma AdaBoost menghasilkan hasil 25 *tweet* yang diklasifikasikan positif dan 36 *tweet* yang diklasifikasikan negatif. Proses perhitungan akurasi hasil klasifikasi *tweet* pada data uji akan disesuaikan dengan sentimen yang telah diketahui sebelumnya.

	Actual	
Predictions	NEGATIF	POSITIF
NEGATIF	35	7
POSITIF	1	18

Gambar 18. Hasil Confusion Matrix AdaBoost

Dari gambar 18 hasil pencocokan data sentimen klasifikasi algoritma AdaBoost dengan sentimen sebenarnya menghasilkan:

- True* Negatif menghasilkan 35 *tweet*
- False* Negatif menghasilkan 7 *tweet*
- False* Positif menghasilkan 1 *tweet*
- True* Positif menghasilkan 18 *tweet*

```
Accuracy : 0.8689
95% CI : (0.7578, 0.9416)
No Information Rate : 0.5902
P-value [Acc > NIR] : 2.146e-06
```

Gambar 19. Akurasi Klasifikasi AdaBoost

Pada gambar 19 akurasi dari algoritma AdaBoost dalam melakukan klasifikasi *tweet* dalam penelitian ini sebesar $0,868 = 86,8\%$.

4. KESIMPULAN

Hasil pengujian dan implementasi metode AdaBoost untuk mengklasifikasikan *tweet* kenaikan BBM Pertamina menunjukkan bahwa penelitian ini berhasil menggunakan algoritma ini untuk mengklasifikasikan *tweet* kenaikan BBM Pertamina + atau -, dengan akurasi 86,8%.

DAFTAR PUSTAKA

- [1] A. Maulana, "Twitter Rahasiakan Jumlah Pengguna di Indonesia," 23 March 2016. [Online]. Available: <https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia>. [Accessed 30 Mei 2023].
- [2] R. Arunachalam and S. Sarkar, "The new eye of government: Citizen sentiment analysis in sosial media," In Sixth International Joint Conference on Natural Language Processing, p. 23, 2015.
- [3] Permenpan, Penyebarluasan Informasi Melalui Media Sosial Bagi Aparatur Sipil Negara, Jakarta, 2018.
- [4] M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," 10 August 2022. [Online]. Available: <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>. [Accessed 30 Mei 2023].
- [5] Y. M. Hendarto, "Kompas," 22 September 2022. [Online]. Available: <https://www.kompas.id/baca/riset/2022/09/06/narasi-penggulingan-orde-baru-turut-mewarnai-kenaikan-harga-bbm>.
- [6] L. D. Utami and R. S. Wahono, "Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes," Journal of Intelligent Systems, vol. I, no. 2, pp. 120-126, December 2015.
- [7] L. Qadrini, A. Seppewali and A. Aina, "Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," Jurnal Inovasi Penelitian, vol. II, no. 7, pp. 1959-1966, December 2021.
- [8] A. Byna and M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," SISFOKOM, vol. IX, no. 3, pp. 407-411, November 2020.
- [9] Tanti, P. Sirait and Andri, "Optimalisasi Kinerja Klasifikasi Melalui Seleksi Fitur dan AdaBoost dalam Penanganan Ketidakseimbangan Kelas," Jurnal Media Informatika Budidarma, vol. V, no. 4, pp. 1377-1385, Oktober 2021.