

Analisis Sentimen Ketidakstabilan Harga Gabah Berbasis Data Twitter

Dedy Sugiarto, Reyhan Dwi Putra, Wahyu Hidayat, Ema Utami, Ainul Yaqin

Program Studi PJJ S2 Teknik Informatika

Universitas Amikom Jogjakarta

Jogyakarta, Indonesia

dedy_sugiarto@students.amikom.ac.id, reyhan.p@students.amikom.ac.id,

20.55.1384_wahyuhidayat@students.amikom.ac.id, ema.u@amikom.ac.id, ainulyaqin@amikom.ac.id

Abstract-Public opinion, whether positive, negative, or neutral, regarding a particular policy or phenomenon in society, is a valuable thing to analyze through a method known as sentiment analysis. The case in this study is the decline in grain prices in early to mid-2021. This study aims to determine the percentage of sentiment polarity that appears when associated with the keyword price of grain and determine the level of accuracy of sentiment class predictions using the Naïve Bayes method. The results showed that the largest percentage of sentiment was negative as much as 46.30%, neutral 32.70% and positive as much as 20.99%. The results of the wordcloud also show that twitter users link the issue of grain prices to rice imports, the role of the government and fertilizers. The results of the classification show a fairly good accuracy value of 67.32%.

Keywords: Sentiment Analysis, Grain Prices, Naïve Bayes

Abstrak-Pendapat masyarakat baik bersifat positif, negatif maupun netral terkait sebuah kebijakan tertentu atau fenomena di masyarakat ini menjadi hal yang berharga untuk dianalisis melalui sebuah metode yang disebut sebagai analisis sentimen. Kasus dalam penelitian ini adalah turunnya harga gabah pada awal sampai dengan pertengahan tahun 2021. Penelitian ini bertujuan untuk mengetahui persentase polaritas sentiment yang muncul bila dikaitkan dengan kata kunci harga gabah serta menentukan tingkat akurasi prediksi kelas sentiment menggunakan metode Naïve Bayes. Hasil penelitian menunjukkan persentase sentiment terbesar adalah bersifat negative sebanyak 46.30%, netral 32,70 % dan positif sebanyak 20,99 %. Hasil wordcloud juga menunjukkan pengguna twitter mengkaitkan persoalan harga gabah ini dikaitkan dengan impor beras, peran pemerintah serta pupuk. Hasil klasifikasi menunjukkan nilai akurasi yang cukup baik yaitu sebesar 67,32 %.

Kata Kunci: Analisis Sentiment, Harga Gabah, Naïve Bayes

1. Pendahuluan

Saat ini masyarakat banyak menggunakan media sosial twitter untuk berbagi pendapat atau opini terhadap suatu isu termasuk dalam hal isu pertanian. Pendapat masyarakat atau opini publik terkait sebuah kebijakan tertentu atau fenomena di masyarakat ini menjadi hal yang berharga untuk dilacak dan dianalisis. Hasil analisis dapat menjadi informasi dan pengetahuan bagi para pengambil keputusan terkait baik dari pihak pemerintah, swasta maupun komunitas masyarakat. Analisis sentimen atau penambangan opini saat ini juga merupakan topik yang paling aktif diteliti di bidang pemrosesan bahasa alami. Meskipun menjadi alat yang sangat kuat, namun tidak banyak digunakan di sektor pertanian [1]. Salah sektor pertanian yang dapat diangkat dalam riset analisis sentimen adalah pertanian pangan yang memasok kebutuhan pangan bagi manusia dimana pelaku utamanya

antara lain adalah para petani sebagai produsen. Namun petani seringkali terkendala dalam hal berproduksi manakala harga jual gabah anjlok atau turun drastis maupun harga pupuk mahal. Hasil pemantauan harga gabah yang bersumber dari Badan Pusat Statistik menunjukkan adanya tren penurunan harga pada tahun 2021 seperti dapat dilihat pada Gambar 1.

Riset yang menganalisis opini publik mengenai kebijakan di sektor pertanian dengan memanfaatkan data twitter berbahasa Indonesia antara lain pernah dilakukan oleh [2] dimana disimpulkan bahwa perubahan harga susu, telur dan bawang merah adalah tiga komoditas yang memiliki kontribusi tertinggi dalam munculnya sentimen negatif tweet. Hasil penelitian juga menunjukkan bahwa klasifikasi dengan Support Vector Machine (SVM) memberikan tingkat akurasi yang paling tinggi

Vol.13 no.1 | Juni 2022

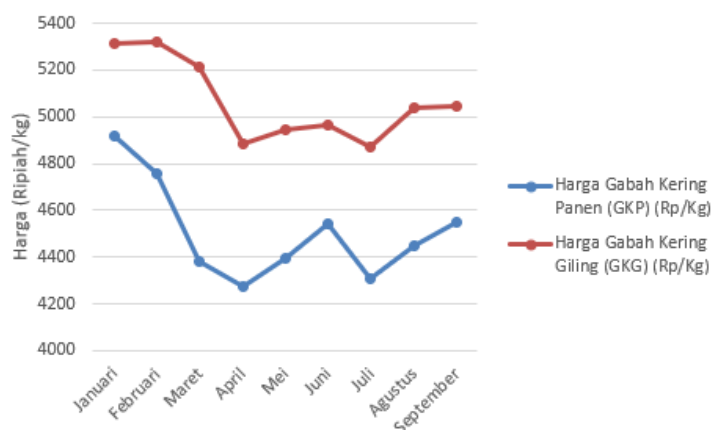
EXPLORE : ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: <http://dx.doi.org/10.36448/jsit.v13i1.2197>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

dibandingkan metode Naïve Bayes dan Decision Tree yaitu sebesar 75,19% (tanpa stemming) dan 80,35 % (dengan stemming). Penelitian oleh [3] menghasilkan kesimpulan bahwa bisa terjadi wacana kebijakan impor beras akan merugikan petani dimana bila harga beras lokal dinilai lebih mahal dari harga beras impor. Hal ini tentunya akan mengakibatkan juga pada turunnya harga gabah. Pengguna twitter masih beranggapan bahwa pasokan beras dalam negeri masih mencukupi. Penelitian lain yang memanfaatkan media sosial twitter juga pernah dilakukan untuk mengetahui persepsi masyarakat di

Indonesia terkait polemik impor beras namun lebih ditekankan pada pengembangan metodologi yaitu penambahan fitur pembobotan jumlah retweet untuk analisis sentimen [4]. Metode yang digunakan adalah metode SVM dengan akurasi yang dihasilkan sebesar 50 % serta bila dengan pembobot jumlah retweet adalah sebesar 48,95 % dan dinilai tidak terdapat perbedaan yang signifikan. Analisis sentiment juga pernah dilakukan terkait kebijakan dibukanya keran ekspor benih lobster [5].



Gambar 1. Pergerakan harga gabah pada tahun 2021 (Sumber : BPS)

Secara umum penelitian terkait analisis sentimen dapat dibagi menjadi tiga pendekatan yaitu berbasis berbasis pembelajaran mesin, berbasis leksikon dan gabungan keduanya [6]. Penelitian analisis sentimen yang menggunakan pendekatan baik pembelajaran mesin maupun leksikon antara lain pernah dilakukan oleh [7], [8]. Kedua penelitian tersebut menggunakan kamus leksikon untuk menentukan sebuah sentiment sebagai positif atau tidak namun menggunakan cara perhitungan skor yang berbeda untuk penetaan sentimennya. Berbeda

dengan beberapa penelitian sebelumnya, penelitian ini bermaksud untuk melakukan analisis sentiment dalam kasus anjloknya harga gabah yang dinilai akan merugikan petani menggunakan pendekatan gabungan baik pembelajaran mesin (metode naïve bayes) serta pendekatan kamus leksikon untuk proses pelabelan sentiment. Data latih yang diberikan diberikan label menggunakan pendekatan leksikon merujuk pada [9] sedangkan data uji yang digunakan untuk mengevaluasi kinerja pengklasifikasi naïve bayes merujuk pada [10].

2. Metodologi

A. Data

Data dari penelitian ini didapatkan dari media sosial Twitter dengan menggunakan library sncrape dari bahasa pemrograman python. Data diambil dari tanggal 1 Januari 2021 hingga 30 Juni 2021 sesuai tren penurunan harga gabah (Gambar 1) dengan kata kunci “harga gabah”. Jumlah data yang berhasil diperoleh sebanyak 4067 tweet.

B. Tahapan Penelitian

Tahapan penelitian dapat dilihat pada Gambar 2. Masalah yang dianalisis dan dirumuskan adalah bagaimana pendapat atau opini masyarakat terkait harga gabah yang anjlok di awal tahun 2021 sampai dengan pertengahan tahun yang dikemas dalam persentase kata

yang paling sering muncul, persentase sentiment positif, negative dan netral serta klasifikasi dengan metode Naïve Bayes. Hasil klasifikasi dapat digunakan untuk memprediksi bila ada tweet baru terkait harga gabah sehingga pemerintah atau pihak terkait akan lebih mudah mengantisipasi.

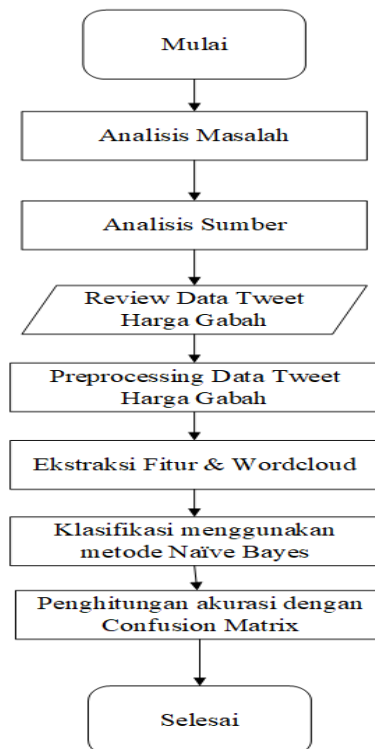
Sumber data berasal dari tweet masyarakat umum yang diharapkan memberikan gambaran mengenai keadaan harga gabah saat penelitian ini dilakukan. Data dikumpulkan dari media sosial Twitter karena pengguna Twitter umumnya lebih mudah dalam mengekspresikan pendapatnya dimana bisa dilakukan dengan username tertentu tanpa mengungkapkan identitas asli pembuat



tweet. Review Data Tweet Harga Gabah. Tahapan preprocessing dan selanjutnya dilakukan menggunakan bahasa pemrograman R. Dalam tahapan ini dilakukan tahapan seperti menghilangkan kata tidak baku (*slang*), mengubah semua tweet menjadi huruf kecil (*case folding*), menghilangkan karakter khusus seperti titik, koma (*cleansing*), menghilangkan kata-kata yang tidak perlu (*stopwords*), mengembalikan tiap kata menjadi bentuk baku (*stemming*). Tahap akhir dalam preprocessing adalah pemberian label pada sentiment. Pelabelan merupakan kegiatan memberikan label terhadap teks yang terdapat dalam data yang sudah di preprocessing. Pelabelan ini dilakukan secara otomatis dengan memberikan bobot pada teks. Bobot tersebut diberikan berdasarkan persamaan kata yang terdapat dalam teks dengan kata-kata yang terdapat dalam kamus kata positif dan kamus kata negatif. Kata yang sama berdasarkan kamus tersebut mendapatkan nilai 1. Hasilnya, jumlah kata positif - jumlah kata negatif yang menghasilkan bobot pada teks.

Setelah bobot sudah diberikan maka dilanjutkan dengan pelabelan berdasarkan bobot yang ada. Jika bobot < 0 maka hasilnya negatif. Jika bobot $= 0$ maka labelnya netral dan jika bobot > 0 maka labelnya bernilai positif.

Pada tahapan ekstraksi fitur dilakukan penghitungan frekuensi tiap kata (*word count*) pada tweet untuk menjadi fitur yang digunakan untuk memprediksi kelas. Metode ini dikenal masuk dalam keluarga *bag of word* atau BoW [11]. Library yang digunakan adalah *DocumentTermMatrix* di R serta fitur terpilih adalah yang frekuensi kemunculannya lebih dari 10. Dalam tahapan ini juga dilakukan visualisasi menggunakan diagram batang yang tersusun secara menurun (*descending*) berdasarkan frekuensi kemunculan serta word cloud. *Word cloud* yaitu membuat sekumpulan kata yang terdapat dalam data berkumpul membentuk kumpulan kata yang memiliki perbedaan ukuran dan warna berdasarkan jumlah frekuensi kemunculannya. Selain *Word Cloud*, visualisasi data juga dapat menggunakan histogram untuk mengetahui jumlah frekuensi terhadap klasifikasi.



Gambar 2. Tahapan Penelitian

C. Metode

Naive Bayes merupakan salah satu metode pengklasifikasian statistik yang menggunakan probabilitas sederhana secara independent atau tidak bergantung. Metode ini didasari oleh Bayes Rule yaitu memprediksikan sebuah peluang masa depan dari pembelajaran di masa sebelumnya [12]. Naïve Bayes merupakan sebuah algoritma pembelajaran sederhana

terkait pengklasifikasian yang memanfaatkan *Bayes Rule*. Misalkan terdapat contoh yang akan diklasifikasikan dijelaskan oleh $x = (x_1, \dots, x_n)$ maka prinsip algoritma tersebut adalah sebagai berikut [13]:

1. Untuk setiap x_1 pada setiap kelas c_j hitung probabilitas bersyarat, $P(x_1|c_j)$, sebagai frekuensi relative dari x_1 di antara data latih yang termasuk dalam c_j .

2. Untuk setiap kelas, c_j , lakukan dua langkah berikut:
 - i) menduga nilai $P(c_j)$ sebagai frekuensi relatif kelas ini dalam set data latih;
 - ii) hitung peluang bersyarat, $p(x_i|c_j)$, dengan menggunakan asumsi "naive" dari kejadian yang saling bebas dengan rumus (1)

$$P(x|c_j) = \prod_{i=1}^n P(x_i|c_j) \dots \dots \dots (1)$$

3. Pilih nilai tertinggi dari $P(c_j) \cdot P(x|c_j)$

Setelah data diklasifikasi menggunakan metode Naïve Bayes, data yang sudah diolah ditampilkan dalam Confusion Matrix untuk penghitungan akurasi analisis sentimen. Persentase data latih dan data uji yang digunakan adalah 90% dan 10 %.

Datetime	Tweet.Id	Tweet
2021-06-29 23:20:37+00:00	1.410015e+18	@Titiektitiek2 @ND45MUpetal @jokowi Ini juga artikel dari ...
2021-06-29 23:13:50+00:00	1.410014e+18	Harga pupuk naik dari 3x lipat sementara harga gabah anjlo...
2021-06-29 23:06:41+00:00	1.410012e+18	@Titiektitiek2 @jokowi Biasanya kalau harga gabah anjlok b...
2021-06-29 23:02:13+00:00	1.410011e+18	Harga gabah petani di Cirebon masih rendah - ANTARA Ne...
2021-06-29 23:00:23+00:00	1.410010e+18	@Titiektitiek2 @ND45MUpetal @jokowi Tidak Ada artikel te...
2021-06-29 22:48:48+00:00	1.410007e+18	@Titiektitiek2 @ND45MUpetal @jokowi Sekarang seluruh d...
2021-06-29 22:24:31+00:00	1.410001e+18	@zackariemt @jokowi Makanya, sudah baca UU otonomi d...
2021-06-29 22:22:08+00:00	1.410001e+18	@Titiektitiek2 @jokowi Pem daerah bisa apa, ini bukan banji...
2021-06-29 22:17:40+00:00	1.410000e+18	@Titiektitiek2 @jokowi Bukan banjir ini harga gabah seluru...
2021-06-29 17:52:18+00:00	1.409933e+18	@Syahrul_YL harga gabah alias padi berapa pak?.. seimbang...
2021-06-29 15:23:46+00:00	1.409895e+18	NAIKKAN HARGA GABAH!!! PPKM DARURAT

Gambar 3. Sampel dataset hasil scraping

```
> print(data$Tweet[29])
[1] "TWK = Singkirkan Penyidik Senior KPK Yg lagi Tangani Kasus Korupsi Besar,\nLive Service : Janji2 Parlente Jalan Terus, Janji Tak Import, Fakta\nImport Beras Ugal2an Harga Gabah Petani Anjlog ! http s://t.co/zFqOKN6VCA"
> print(aya[29])
[1] "TWK = Singkirkan Penyidik Senior KPK yang lagi Tangani Kasus Korupsi Besar,\nLive Service : Janji 2 Parlente Jalan Terus, Janji Tak Import, Fakta\nImport Beras Ugal2an Harga Gabah Petani Anjlog ! http s://t.coatauzFqOKN6VCA"
```

Gambar 4. Perbedaan sebelum dan sesudah Normalisasi

```
> inspect(aya[4])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] Harga gabah petani di Cirebon masih rendah - ANTARA News Jawa Barat https://t.coatauaawh3meDh6
> aya <- tm_map(aya, tolower)
Warning message:
In tm_map.SimpleCorpus(aya, tolower) : transformation drops documents
> inspect(aya[4])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] harga gabah petani di cirebon masih rendah - antara news jawa barat https://t.coatauaawh3medh6
```

Gambar 5. Pengubahan teks menjadi huruf kecil

```
[1] twk = singkirkan penyidik senior kpk yang lagi tangani kasus korupsi besar,\nlive service : janji2 parlente jalan terus, janji tak import, fakta\nimport beras ugal2an harga gabah petani anjlog ! http s://t.coatauzfqokn6vca
> aya <- tm_map(aya, removeURL)
Warning message:
In tm_map.SimpleCorpus(aya, removeURL) : transformation drops documents
> inspect(aya[29])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] twk = singkirkan penyidik senior kpk yang lagi tangani kasus korupsi besar,\nlive service : janji2 parlente jalan terus, janji tak import, fakta\nimport beras ugal2an harga gabah petani anjlog !
```



Gambar 6. Penghapusan URL

```
[1] @titiektitiek2 @nd45mupeta1 @jokowi ini juga artikel dari situbondo, mohon bantuannya sekarang musim panen para tengkulak juga tidak berani membeli harga gabah \n
> aya <- tm_map(aya,removeAT)
Warning message:
In tm_map.SimpleCorpus(aya, removeAT) : transformation drops documents
> inspect(aya[1])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] ini juga artikel dari situbondo, mohon bantuannya sekarang musim panen para tengkulak juga tidak berani membeli harga gabah \n
```

Gambar 7. Sesudah dan sesudah penghapusan mention

```
[1] gabah saya dibeli dengan harga rp3.600 per kilogram. #hargagabah #petani #karawang #padi
> aya <- tm_map(aya,removehash)
Warning message:
In tm_map.SimpleCorpus(aya, removehash) : transformation drops documents
> inspect(aya[39])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] gabah saya dibeli dengan harga rp3.600 per kilogram.
```

Gambar 8. Sebelum dan sesudah penghapusan tagar

```
[1] kata rakyat kecil petani gurem wongndeso di juga benar janji2 parlente petani makmur, harga pupuk subsidi dinaikkan, import beras ugalkan disaat panen raya harga gabah petani jatuh
> aya <- tm_map(aya, removeNL)
Warning message:
In tm_map.SimpleCorpus(aya, removeNL) : transformation drops documents
> inspect(aya[41])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] kata rakyat kecil petani gurem wongndeso di juga benar janji2 parlente petani makmur, harga pupuk subsidi dinaikkan, import beras ugalkan disaat panen raya harga gabah petani jatuh
```

Gambar 9. Sebelum dan Sesudah Penghapusan tanda “\n”

```
> inspect(aya[271])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] rt bps_statistics: selama mei 2021, rata-rata harga gkp di tingkat petani rp4.398,- per kg atau naik 2,87 persen dan di tingkat penggilingan rp4.498,- per kg atau naik 2,27 persen dibandingkan harga gabah kualitas yang sama pada bulan sebelumnya.
> removeRT <- function(y) gsub("rt", "", y)
> aya <- tm_map(aya, removeRT)
Warning message:
In tm_map.SimpleCorpus(aya, removeRT) : transformation drops documents
> inspect(aya[271])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] bps_statistics: selama mei 2021, rata-rata harga gkp di tingkat petani rp4.398,- per kg atau naik 2,87 persen dan di tingkat penggilingan rp4.498,- per kg atau naik 2,27 persen dibandingkan harga gabah kualitas yang sama pada bulan sebelumnya.
```

Gambar 10. Sebelum dan Sesudah Penghapusan “rt”

```
> inspect(aya[2])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] harga pupuk naik dari 3x lipat sementara harga gabah anjlok. harga kopi turun. buat mereka itu cuma statistik. kan kalian punya lahan masih bisa makan... jangan salahkan ya kalau petani ngomong ke anaknya sekolah yang pintar buat jadi pegawai negeri setelah lulus, hidupnya enak.
> aya <- tm_map(aya, removePunctuation)
Warning message:
In tm_map.SimpleCorpus(aya, removePunctuation) :
transformation drops documents
> inspect(aya[2])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] harga pupuk naik dari 3x lipat sementara harga gabah anjlok harga kopi turun buat mereka itu cuma statistik kan kalian punya lahan masih bisa makan jangan salahkan ya kalau petani ngomong ke anaknya sekolah yang pintar buat jadi pegawai negeri setelah lulus hidupnya enak
```



Gambar 11. Sebelum dan Sesudah Penghapusan Tanda Baca

```
> inspect(aya[221])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] dari dulu cuman 14 masih blom stabil harga jual barang lokal masih murah sepei gabah dan hasil
    bumi lainnya
> aya <- tm_map(aya, removeNumbers)
Warning message:
In tm_map.SimpleCorpus(aya, removeNumbers) : transformation drops documents
> inspect(aya[221])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] dari dulu cuman masih blom stabil harga jual barang lokal masih murah sepei gabah dan hasil bu
    mi lainnya
```

Gambar 12. Sebelum dan Sesudah Penghapusan Angka

```
> inspect(aya[25])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] politik lah paling benar harga pupuk dari politik harga beli gabah dari politik impor beras da
    ri politik kalau aturan yang dibuat tidak pro rakyat jangan mimpi
> cStopwordID <- readLines(stopwordID)
Warning message:
In readLines(stopwordID) :
  incomplete final line found on 'C:/Users/ibrahim/Documents/1Riset/stopword_id.txt'
> aya <- tm_map(aya, removeWords, cStopwordID)
Warning message:
In tm_map.SimpleCorpus(aya, removeWords, cStopwordID) :
  transformation drops documents
> inspect(aya[25])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] politik harga pupuk politik harga beli gabah politik impor beras politik aturan pro
    rakyat mimpi
```

Gambar 13. Sebelum dan Sesudah Penghapusan Stopword

```
> inspect(aya[25])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] politik harga pupuk politik harga beli gabah politik impor beras politik aturan pro rakyat
    mimpi
> aya <- tm_map(aya, stripwhitespace)
Warning message:
In tm_map.SimpleCorpus(aya, stripwhitespace) :
  transformation drops documents
> inspect(aya[25])
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content: documents: 1

[1] politik harga pupuk politik harga beli gabah politik impor beras politik aturan pro rakyat mimpi
```

Gambar 14. Sebelum dan Sesudah penghapusan spasi berlebihan.

```
> library(katadasar)
> stemming <- function(x){
+   paste(lapply(x,katadasar),collapse = " ")
+ }
> library(tokenizers)
> tweets <- lapply(tokenize_words(stem[]), stemming)
> stem[25]
[1] " politik harga pupuk politik harga beli gabah politik impor beras politik aturan pro rakyat mimpi"
> tweets[25]
[[1]]
[1] "politik harga pupuk politik harga beli gabah politik impor beras politik atur pro rakyat mimpi"
```

Gambar 15. Proses Stemming

3. Hasil dan Pembahasan

Dataset yang didapatkan dari hasil scraping dapat dilihat pada Gambar 3. Terdapat 4 kolom dalam dataset berbentuk dataframe di R Studio dimana kolom tersebut berisi waktu pengguna twitter menyampaikan *tweet*, *tweet_id*, teks yang berisi *tweet* serta *user name*. Tahapan pertama preprocessing adalah dengan melakukan normalisasi. Gambar 4 memperlihatkan contoh

perbedaan sebelum dan sesudah normalisasi yang mengganti kata “Yg” menjadi “yang” dapat dilihat pada Gambar 4. Selanjutnya dilakukan pengubahan ukuran kalimat menjadi huruf kecil semua seperti pada Gambar 5 dimana kata “Harga” dan kata “Cirebon” sudah dimulai dengan kecil. Setelah diubah menjadi huruf kecil semua, dilakukan lagi preprocessing untuk menghilangkan URL

Vol.13 no.1 | Juni 2022

EXPLORE : ISSN: 2087-2062, Online ISSN: 2686-181X / DOI: <http://dx.doi.org/10.36448/jsit.v13i1.2197>



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

seperti pada Gambar 6. Penghapusan *mention* seperti “@titiectitiek2” yang digunakan oleh pengguna dapat dilihat pada Gambar 7. Penghapusan tagar dalam dokumen seperti #hargagabah, #petani, #karawang juga telah dapat dihilangkan seperti terlihat pada Gambar 8. Selanjutnya melakukan penghapusan tanda”\n” dalam dokumen seperti dapat dilihat pada Gambar 9.

Penghapusan kata “rt” yang biasanya digunakan pengguna untuk mengulang postingan dari pengguna lain seperti pada Gambar 10. Penghapusan tanda baca dan angka dapat dilihat pada Gambar 11 dan Gambar 12.

Selanjutnya melakukan Penghapusan Stopword atau kata-kata yang tidak begitu penting seperti pada Gambar 13 dan penghapusan terkait spasi berlebihan yang terdapat dalam dokumen seperti pada Gambar 14. Proses *stemming* atau mengubah kata menjadi kata dasar seperti pada kata “aturan” telah berubah menjadi kata “atur” seperti dapat dilihat pada Gambar 15.

Tahap yang terakhir yaitu masuk ke tahap pelabelan yaitu melakukan pelabelan menggunakan metode *Lexicon-Based* [9] dengan hasil seperti pada Gambar 16.

score	text	klasifikasi
0	aikel situbondo mohon bantu musim panen tengkulak bera...	Netral
-1	harga pupuk x lipat harga gabah anjlok harga kopi turun sta...	Negatif
-1	harga gabah anjlok bulog beli gabah petan mekanisme bu ...	Negatif
1	harga gabah petan cirebon rendah news jawa barat	Positif
1	aikel terbaru harga gabah daerah fakta lapang	Positif
1	daerah harga gabah anjlok ngegas mohon sungguh hormat...	Positif
-3	baca uu otonomi daerah paham desentralisasi wewenang k...	Negatif
-1	pem daerah banjir harga gabah anjlok daerah bulog turun b...	Negatif
0	banjir harga gabah daerah turin turun tang bulog serap pe...	Netral
0	harga gabah alias padi imbang naik harga pupuk	Netral

Gambar 16. Pelabelan terhadap dokumen

Nilai skor pada label yang bernilai negative menunjukkan bahwa sentiment tersebut bersifat negatif dengan hasilnya sebanyak 46.30 % dari seluruh dataset sentiment yang digunakan dalam penelitian ini. Sentimen netral atau memiliki nilai nol adalah sebanyak 32,70 % dan

sentiment positif sebanyak 20,99 %. Berdasarkan hasil ini dapat diketahui bahwa kasus harga gabah lebih disikapi dengan sentiment negative. Hal ini tentunya perlu menjadi perhatian serius bagi pihak-pihak terkait seperti Bulog dan Kementerian terkait.

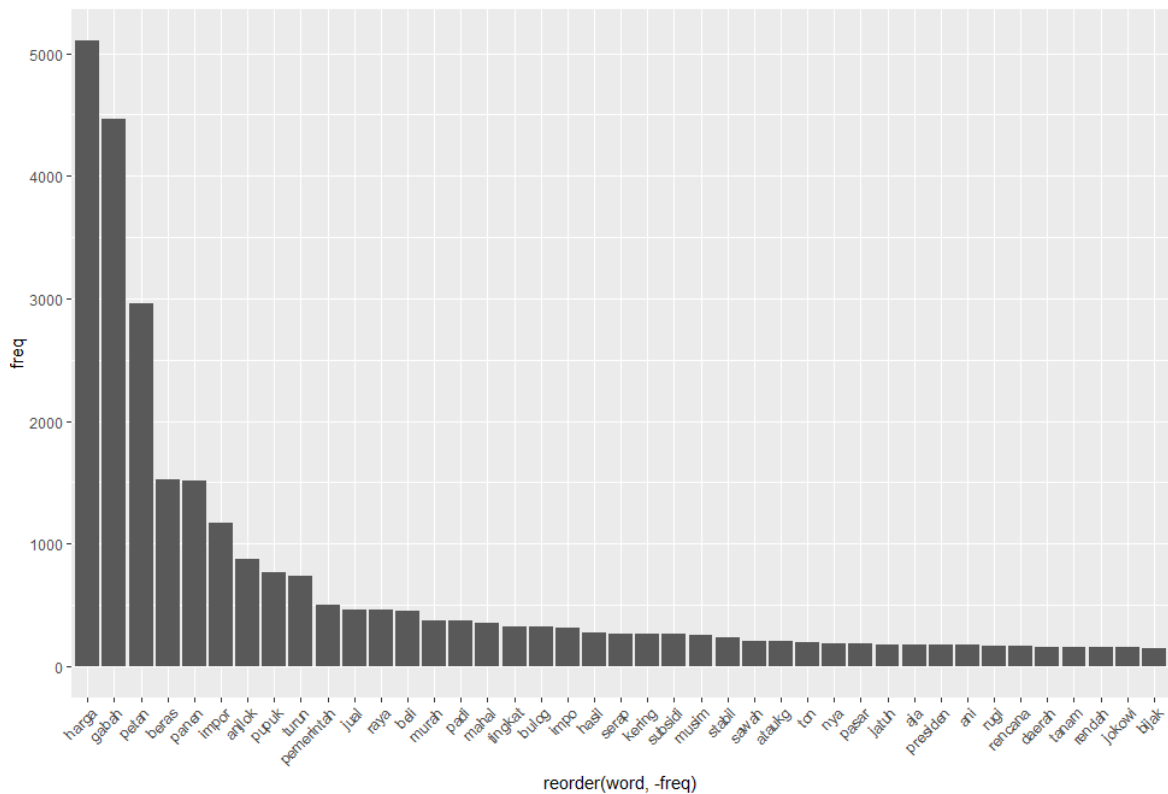
Docs	anjlok	beras	gabah	harga	impor	panen	pemerintah	petan	pupuk	turun
1007	0	1	2	1	0	2	0	3	0	0
1021	1	0	2	4	0	1	2	1	0	0
1296	0	1	2	2	0	0	0	3	2	0
2147	0	2	2	3	1	1	2	1	0	1
2633	0	4	1	3	0	0	2	3	0	0
2964	1	3	1	2	2	1	0	2	0	1
3092	1	0	1	2	3	1	1	2	1	0
3197	1	1	1	1	1	3	2	1	0	0
3295	1	1	1	1	1	2	1	2	0	0
3354	0	1	2	3	1	2	0	4	0	1

Gambar 17. Sampel frekuensi kemunculan kata pada tiap dokumen



```
> freq <- colSums(as.matrix(dtms))
> freq
bantu      beli      gabah      harga      mohon      musim      panen      tengkulak      anak      anjlok
69          454      4466      5107      59          253      1515      135          51          877
hidup      lahan      neger      petan      pupuk      salah      turun      bulog      moga      jawa
56          61          60          2961      768        93          740      322          58          85
rendah     daerah     lapang     butuh     debat     tolong     baca      jual     pemerintah     presiden
154        160        55          94          63          79          48          463        503          176
serap      naik      padi      amp      jerit     kering     pasar     aja      beras      impor
270        81          370        96          76          267        185        179        1529       1167
isu        langsung   wacana     menteri    biaya     hasil     jokowi    ataukg    jatuh     terjun
112        94          104        122        140        279        153        202        181        63
hati       desa      mas        hancur     impo      langka     murah     sawah     pro       rakyat
55         77         56         116        317       112        373        205        45        134
banding    laku      blm        hama       juta      modal     pakai     subsidi   masuk     raya
92         47         47         41         142       65         58         266       68        458
stok       tanam     jalan     bayar     ribu     keluh     gudang     rugi     hpp      kilogram
116        159       44         45         57         117        51         168       128       67
saat      rasa      tan        dengar     bagus     kalo      mahal     pengaruh  produksi  gimana
49         42        120       44         82         147       353        61        148       83
main      obat     ton       jelang     mendag    pangan     stabil     akibat    bawah     tingkat
90         51        193       58         51         116       234        79        48        324
kampung    orang     berita    cari       kerja     susah     udah     kabupaten  nasib     ani
72         78         56         42         70         100       116        69        73        173
nya       pei       untung    banget     hujan     pas       sulit     indonesia  solusi     suai
187        87        137       104        56         98         70         136       49        97
mafia     dagang    tani      sih        bebas     bijak     sejahtera  sebab     utk       gkp
68         44        80        73         55         151       76         45        85        79
nilai     kisar     tau       jamin     negara    harap     tambah     gitu     rencana  giling
50         46        63        121       90         69         47         63        163       96
kualitas  bilang    kali      april     kilo      limpah    drastis    biar     jga       gerak
82         60        58        45         45         56         55         76        134       69
basah     dampak    alami     lokal     indramayu  tuju     henti     masyarakat  ekonomi   produk
83         53        47        92         53         43         47         50        44        51
senang    kait     coba     kementan  kasihan   dapat     upaya     rosot     47
52         44        45        111       62         44         65         47
```

Gambar 18. Frekuensi kemunculan kata dalam seluruh dokumen



Gambar 19. Diagram batang kata yang paling sering muncul





Gambar 20. Visualisasi wordcloud

Selanjutnya sampel dari hasil perhitungan frekuensi kemunculan kata pada tiap dokumen dapat dilihat pada Gambar 17. Sedangkan frekuensi kemunculan kata pada seluruh dokumen dapat dilihat pada Gambar 18. Dapat diketahui bahwa gabah dan kata harga adalah dua kata yang paling sering muncul. Hal ini wajar karena memang kata kunci yang digunakan untuk pencarian di awal ketika scraping adalah menggunakan kedua kata kunci tersebut. Kata berikutnya yang memiliki frekuensi terbesar selanjutnya adalah petan (berasal dari kata petani), beras, panen, impor dan anjlok. Hal ini menunjukkan bahwa memang masyarakat pengguna twitter mengkaitkan kasus harga gabah dengan terjadi penurunan harga yang turun drastis dimana kata anjlok banyak digunakan dan ini kaitannya dengan beras, petani, panen dan juga impor. Grafik diagram batang lebih jelas lagi memperlihatkan frekuensi kemunculan kata yang tersusun secara menurun (Gambar 19). Visualisasi berdasarkan wordcloud juga menampilkan informasi yang mudah dipahami dimana beberapa kata muncul dengan ukuran huruf yang lebih besar setelah kata

kunci harga gabah yaitu ‘petani’, ‘panen’, ‘impor’, ‘anjlok’ dan ‘beras’. Kemudian dilanjutkan dengan kata pupuk, beli dan pemerintah menunjukkan sentiment harga gabah juga dikaitkan dengan persoalan pupuk serta kata pemerintah dan kata beli (Gambar 20).

Hasil pengujian dari algoritma Naïve Bayes untuk klasifikasi sentiment dapat dilihat pada Gambar 21. Tampilan confusion matrix menunjukkan kesesuaian antara kelas pada bagian predicted dan bagian actual atau referensi. Nilai akurasi model didapatkan sebesar 67,32 % yang didapatkan dari penjumlahan angka pada diagonal kanan dengan total. Persentase jumlah sentiment negative, netral dan positif relatif berimbang pada data uji sehingga nilai akurasi bisa kuat menggambarkan kebaikan dari model prediksi yang digunakan. Hal ini juga diperkuat dengan nilai sensitifitas dan spesifitas yang berada pada kisaran 60 % sampai dengan 90%. Sampel dari sentiment dengan nilai referensi atau aktual dengan hasil prediksinya dapat dilihat pada Gambar 22.



5. Daftar Pustaka

- [1] J. Novák, P. Benda, E. Šilerová, J. Vaněk, and E. Kánská, "Sentiment Analysis in Agriculture," *Agris On-line Pap. Econ. Informatics*, vol. 13, no. 1, pp. 121–130, 2021, doi: 10.7160/aol.2021.130109.
- [2] I. Surjandari, M. S. Naffisah, and M. I. Prawiradinata, "Text Mining of Twitter Data for Public Sentiment Analysis of Staple Foods Price Changes," *J. Ind. Intell. Inf.*, vol. 3, no. 3, pp. 253–257, 2014, doi: 10.12720/jiii.3.3.253-257.
- [3] Z. M. Basuki, R. J. Putra Hidayat, P. S. Asih, and I. T. Sihidi, "Analisis Kebijakan Impor Beras 1 Juta Ton Oleh Pemerintah Indonesia: Data Dan Analisis Media Sosial," *J. Acad. Praja*, vol. 4, no. 2, pp. 474–495, 2021, doi: 10.36859/jap.v4i2.485.
- [4] R. A. Nandini, Y. A. Sari, and P. P. Adikara, "Analisis Sentimen Impor Beras 2018 Pada Twitter Menggunakan Metode Support Vector Machine dan Pembobotan Jumlah Retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 4, pp. 3396–3406, 2019.
- [5] K. Sulastri, "Klasifikasi Naïve Bayes pada Analisis Sentimen atas Penolakan Dibukanya Larangan Ekspor Benih Lobster," vol. 1, no. 2, pp. 68–75, 2020.
- [6] D. Sharma, M. Sabharwal, V. Goyal, and M. Vij, *Sentiment Analysis Techniques for Social Media Data: A Review*. Springer Singapore, 2020.
- [7] S. Yuliyanti, T. Djatna, and H. Sukoco, "Sentiment mining of community development program evaluation based on social media," *Telkomnika (Telecommunication Comput. Electron. Control)*, vol. 15, no. 4, pp. 1858–1864, 2017, doi: 10.12928/TELKOMNIKA.v15i4.4633.
- [8] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.
- [9] N. Nigam and D. Yadav, "Lexicon-Based Approach to Sentiment Analysis of Tweets Using R Language: Second International Conference, ICACDS 2018, Dehradun, India, April 20-21, Lexicon-based approach to Sentiment Analysis of tweets using R language," no. February 2019, pp. 372–380, 2018, doi: 10.1007/978-981-13-1810-8.
- [10] S. Aftab, M. Ahmad, and S. Aft, "Hybrid Tools and Techniques for Sentiment Analysis: A Review Related papers Tools and Techniques for Lexicon Driven Sentiment Analysis: A Review Machine Learning Techniques for Sentiment Analysis: A Review," *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 4, 2017, [Online]. Available: www.ijmse.org.
- [11] Z. Liu, Y. Lin, and M. Sun, *Representation Learning for Natural Language Processing*. 2020.
- [12] C. Sammut and G. . Webb, *Encyclopedia of Machine Learning and Data Mining*. Springer, 2011.
- [13] M. Kubat, *An Introduction to Machine Learning*. 2017.

