

# Penerapan Text Mining Dengan Algoritma Naïve Bayes Untuk Mengklasifikasikan Sentimen Rakyat Terhadap Minyak Goreng Subsidi Pemerintah

Rion Aztin<sup>1\*</sup>, Krisna Adiyarta<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budiluhur, Jakarta, Indonesia

Email: <sup>1\*</sup>1611500693@student.budiluhur.ac.id, <sup>2</sup>krisna.adiyarta@budiluhur.ac.id

(\* : corresponding author)

**Abstrak-**Kenaikan harga minyak goreng terjadi di Indonesia secara mendadak di Indonesia. Cukup mengagetkan mengingat Indonesia merupakan negara penghasil minyak sawit terbesar di dunia. Banyak masyarakat Indonesia yang mengeluhkan kenaikan harga minyak goreng yang tinggi ini. Pemerintah hadir ditengah tingginya harga minyak goreng dipasaran dengan melakukan subsidi untuk menstabilkan kembali harga minyak goreng yang tinggi di masyarakat. Harga minyak goreng memang sudah disubsidi, namun stok dan keberadaan minyak goreng subsidi pemerintah di gerai dan minimarket banyak yang tidak tersedia. *Social media twitter* tempat menjadinya trending topik #Minyak Goreng. Penelitian ini bertujuan untuk melihat respon masyarakat di social media terkait kebijakan minyak goreng subsidi pemerintah. Pada penelitian ini dilakukan klasifikasi pada data twitter untuk mengklasifikasikan sentiment masyarakat terhadap minyak goreng subsidi pemerintah. Kontribusi penelitian ini dapat membantu pemerintah terkait kebijakan minyak goreng subsidi terdapat masalah selama kebijakan ini berlaku dengan mengetahui respon masyarakat. Penelitian ini diterapkan Algoritma *Naïve Bayes* untuk mengklasifikasikan sentiment masyarakat Indonesia terhadap subsidi minyak goreng pemerintah ditemukan bahwa tingkat akurasi yang didapatkan yakni sebesar 69.4% dengan *precision yang dihasilkan sebesar 65,3%* sedangkan *recall yang dihasilkan sebesar 31.6%* dan *error rate sebesar 30,6%*

**Kata Kunci:** klasifikasi, minyak goreng, data mining, pemerintah, twitter, subsidi, naïve bayes, confusion matrix, API

## *Application Of Text Mining With Naive Bayes Algorithm To Classificate People's Sentiment On Government Subsidy Cooking Oil*

**Abstract-***The increase in cooking oil prices occurred in Indonesia suddenly in Indonesia. It is quite surprising considering that Indonesia is the largest palm oil producing country in the world. Many Indonesians have complained about the high price of cooking oil. The government is present in the midst of high cooking oil prices in the market by providing subsidies to stabilize the high cooking oil prices in the community. The price of cooking oil has indeed been subsidized, but the stock and availability of government subsidized cooking oil at outlets and minimarkets is not readily available. Social media twitter is the trending topic for #Fried Oil. This study aims to see the public's response on social media related to the government's subsidized cooking oil policy. In this study, classification of twitter data was carried out to classify public sentiment towards government subsidized cooking oil. The contribution of this research can help the government regarding the subsidized cooking oil policy, there are problems as long as this policy is in effect by knowing the community's response. This study applied the Naïve Bayes algorithm to classify Indonesian people's sentiments towards government cooking oil subsidies. It was found that the level of accuracy obtained was 69.4% with a precision of 65.3%, while the recall produced was 31.6% and the error rate was 30.6. %*

**Keywords:** classification, cooking oil, data mining, government, twitter, subsidy, naïve bayes, confusion matrix, API

## 1. PENDAHULUAN

Beberapa bulan terakhir, terjadi kenaikan harga minyak goreng di seluruh wilayah di Indonesia. Warga dan masyarakat Indonesia di *social media* mengeluhkan tingginya harga minyak goreng yang terjadi secara mendadak ini, kebijakan pemerintah pun dipertanyakan. Di Tengah tingginya harga minyak goreng, pemerintah mencoba untuk menstabilkan harga minyak goreng dengan melakukan subsidi harga. Namun cara itu ternyata masih belum cukup. Harga minyak goreng memang sudah disubsidi, namun stok dan keberadaan minyak subsidi pemerintah di gerai dan minimarket banyak yang tidak tersedia

*Data Mining* merupakan proses maupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi–informasi yang nantinya dapat digunakan. *Data mining* merupakan teknik yang menggabungkan teknik analisis data dan menemukan pola-pola yang penting pada data. *Data mining* tersebut akan menjadi tolak ukur ataupun acuan untuk mengambil keputusan[1]. Berangkat dari permasalahan tersebut tujuan dari penulisan tugas akhir adalah untuk membangun suatu sistem untuk mengetahui reaksi serta sentiment masyarakat di tengah kelangkaan minyak goreng subsidi menggunakan metode *text mining*.

Kontribusi penelitian ini dapat membantu pemerintah terkait kebijakan yang mereka buat lancar atau tidaknya kebijakan ini diberlakukan. Respon masyarakat sangatlah penting untuk mengetahui kebijakan minyak goreng subsidi ini berjalan dengan lancar atau mengalami masalah.

## 2. METODE PENELITIAN

### 2.1 Data Mining

*Data Mining* adalah proses mengumpulkan data berukuran besar, kemudian mengekstraksi data tersebut menjadi informasi yang nantinya dapat digunakan. *Data mining* merupakan teknik menggabungkan teknik analisis data dan menemukan pola-pola yang penting pada data. *Data mining* tersebut akan menjadi tolak ukur ataupun acuan untuk mengambil keputusan [1]

Dalam *data mining*, data disimpan secara elektronik kemudian diproses secara otomatis oleh komputer menggunakan perhitungan tertentu. *Data mining* tersebut akan menjadi tolak ukur ataupun acuan untuk mengambil keputusan [1]

Tidak semua pencarian informasi dinyatakan sebagai *data mining*. Tahapan proses *data mining* diawali dari seleksi data sumber ke data target, tahap *preprocessing* memperbaiki kualitas data, *transformasi*, *data mining* serta tahap *interpretasi* dan evaluasi yang menghasilkan output pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut:

#### a. Data Selection

Pemilihan data (seleksi) dari sekumpulan data operasional diperlukan untuk pengalihan informasi dalam KDD dimulai

#### b. Pre-processing

*Pre-processing* pada data yang menjadi fokus KDD dimulai, proses *pre-processing* mencakup antara lain membuang duplikat data, memeriksa data yang inkonsisten dan memperbaiki data tersebut

#### c. Transformation

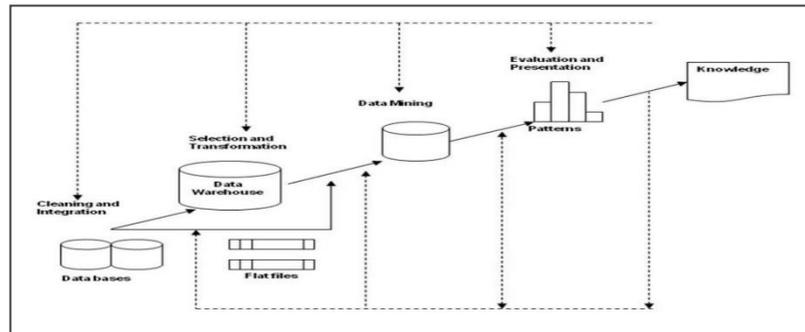
*Coding* proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai dengan data mining

#### d. Data mining

Proses mencari pola yang menarik dalam data yang terpilih dengan metode tertentu

#### e. Interpretation /Evaluation

Pola ini yang dihasilkan dari *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh user



Gambar 1. Tahapan Data Mining[2]

### 2.2 Naïve Bayes

Algoritma *Naïve Bayes* adalah metode klasifikasi menggunakan probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naïve Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Ciri utama dari *Naïve Bayes* ini adalah asumsi yg sangat kuat (*Naïve*) akan independensi dari masing-masing kondisi atau kejadian. *Naïve Bayes* berasumsi bahwa efek dari nilai variabel pada kelas tertentu independen dari nilai-nilai variabel yang lain. Hal tersebut dibuat untuk penyederhanaan perhitungan dan dalam hal ini disebut juga “*Naïve*”[3]. Secara umum Algoritma *Naïve Bayes* dapat dirumuskan sebagai berikut:

$$P(H|X) = \frac{P(X|H)}{1P(X)} \cdot P(H) \quad (1)$$

Keterangan :

X : Kelas data belum diketahui

H : Hipotesis data X suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan dengan kondisi X (*posteriori probability*)

P(H) : Probabilitas hipotesis H (*prior probability*)

P(X|H) : Probabilitas X berdasarkan kondisi hipotesis H

P(X) : Probabilitas X

Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yg diperlukan dalam proses pengklasifikasian.

Karena yg diasumsikan dijadikan variabel independen, jadi hanya varians dari variabel sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan.

### 2.3 Confusion Matrix

*Confusion Matrix* merupakan metode yang digunakan untuk melakukan perhitungan akurasi pada *data mining*. Presisi atau *confidence* adalah proposisi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Nilai yang dihasilkan melalui metode *Confusion Matrix* multiple class adalah berupa evaluasi sebagai berikut:

*Accuracy* adalah persentase jumlah record data yang diklasifikasikan (prediksi) secara benar oleh Algoritma

$$\frac{\sum_i tpi + tni}{l} \quad (2)$$

$$Accuracy = \frac{\sum_{l=1}^l tpi + fni + fpi + tni}{l}$$

*Precision* merupakan rasio item relevan yang dipilih terhadap semua *item* yang terpilih. *Precision* dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut

$$\frac{\sum_i tpi}{\sum_i tpi + fpi} \quad (3)$$

$$Precision = \frac{\sum_{l=1}^l tpi}{\sum_{l=1}^l (tpi + fpi)}$$

*Recall* merupakan rasio dari *item* relevan yang dipilih terhadap total jumlah *item* relevan yang tersedia

$$\frac{\sum_i tpi}{\sum_{l=1}^l (tpi + fpi)} \quad (4)$$

$$Recall = \frac{\sum_i tpi}{\sum_{l=1}^l (tpi + fpi)}$$

*Error Rate* merupakan persentase jumlah *record* data yang diklasifikasikan secara salah oleh algoritme

$$\frac{\sum_i fpi + fni}{l} \quad (5)$$

$$Error Rate = \frac{\sum_{l=1}^l fpi + fni}{l}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Lingkungan Percobaan

Sebelum sistem diimplementasikan dan bisa diuji coba, maka terlebih dahulu harus dilakukan persiapan sebelum implementasi ini bekerja dan berjalan baik sesuai yang diharapkan dan direncanakan. Untuk persiapan implementasi terdiri dari beberapa langkah sebagai berikut:

#### a. Hardware

Pada saat aplikasi ini dibuat, test dilakukan menggunakan *webservice local*. Dibawah ini merupakan spesifikasi *hardware* yang digunakan untuk menjalankan aplikasi ini :

Processor AMD Ryzen 5 3600  
RAM 16 GB 3200 MHz DDR4  
VGA NVIDIA RTX 2060 6GB  
SSD berkapasitas 1 TB

#### b. Software

Dalam penelitian ini menggunakan beberapa aplikasi perangkat lunak sebagai berikut:

Windows 10  
Notepad ++ sebagai *code editor*

XAMPP sebagai database dan application server



Gambar 2. Deployment Program

### 3.1.1 Implementasi Metode

Pada tahap ini akan dilakukan klasifikasi data menggunakan algoritme *Naive Bayes*. Metode tersebut merupakan metode yang digunakan untuk menghitung dan mengklasifikasikan data untuk memperoleh hasil terhadap dataset yang labelnya telah ditentukan sebelumnya.

Tabel 1 . Contoh *sample* pemberian label

ID	Text	Label
1537730271599923200	@tvOneNews Harus massa minyak goreng sudah disubsidi masih mahal contohnya di wilayah saya sampai saat ini masih di kisaran minyak kemasan 24000/liter sedangkan minyak curah 16000/liter mana keadilan sosial bagi seluruh rakyat Indonesia???????	Negatif
1537355266949513217	Kemenperin mengakselerasi perluasan pasar ekspor minyak sawit mentah atau crude palm oil (CPO) dan minyak goreng sawit ke Pakistan. Ini dilakukan setelah dibukanya kran ekspor CPO dan turunannya harga minyak goreng curah saat ini. @jokowi Gaspol Pemulihan Ekonomi <a href="https://t.co/35WYQAE6fG">https://t.co/35WYQAE6fG</a>	Positif
1537084937673011201	Para Babinsa jajaran Koramil 12/Margorejo melaksanakan Komunikasi Sosial (Komsos) sebagai sarana membangun keakraban dengan masyarakat sekaligus monitoring harga sembako dan Minyak goreng curah di agen dan pasar tradisional di Wilayah Koramil 12/Margorejo.Rabu,(15/06/2 022). <a href="https://t.co/c5uX4qzVoz">https://t.co/c5uX4qzVoz</a>	Netral
1537647723771744256	Such a bullshit! Kebijakan beliau ttg larangan ekspor cpo	Negatif

Proses yang dilakukan pada awal klasifikasi adalah mencari nilai probabilitas tiap klasifikasi. Ini dilakukan dengan menghitung probabilitas setiap klasifikasi. Data yang kita miliki masih dalam format teks, sehingga *pre processing* perlu dilakukan terhadap data yang kita miliki agar dapat menghitung jumlah kata

Tabel 2 .Contoh *sample* data untuk diuji

Text	Pre-processing	Label
@abu_waras Emang ada yg mengeluh harga minyak Goreng saat ini...? Warga Indonesia pasti mampu beliin, walau beda harga dari sebelumnya,. Yg penting minyak goreng ada beserta stoknya banyak dan tak mengalami kelangkaan kayak sebelumnya.... Masalah harga minyak goreng trus yg di bahas kwkwkw???????	emang harga goreng beda minyak kayak bahas	Negatif

Tabel 2 menunjukan *sample* data yang diklasifikasikan dan diuji oleh program

**Tabel 3.** Contoh *Sample* data untuk training table

<b>Text</b>	<b>Pre-processing</b>	<b>Label</b>
<p>Kemenperin mempercepat perluasan pasar ekspor minyak sawit mentah atau crude palm oil (CPO) dan minyak goreng sawit ke Pakistan. Ini dilakukan setelah dibukanya kran ekspor CPO dan turunannya harga minyak goreng curah saat ini. @jokowi Gaspol Pemulihan Ekonomi  <a href="https://t.co/35WYQAE6fG">https://t.co/35WYQAE6fG</a></p>	<p>pasar ekspor minyak sawit goreng istan harga curah ekonomi</p>	<p>Positif Minyak goreng wilayah curah luruh rakyat Negatif</p>
<p>Para Babinsa jajaran Koramil 12/Margorejo melaksanakan Komunikasi Sosial (Komsos) sebagai sarana membangun keakraban dengan masyarakat sekaligus monitoring harga sembako dan Minyak goreng curah di agen dan pasar tradisional di Wilayah Koramil 12/Margorejo.Rabu,(15/06/2022).  <a href="https://t.co/c5uX4qzVoz">https://t.co/c5uX4qzVoz</a></p>	<p>komunikasi sarana masyarakat harga minyak goreng curah agen pasar wilayah</p>	<p>Netral</p>
<p>Such a bullshit! Kebijakan beliau ttg larangan ekspor cpo kemarin bener" bikin petani sawit susah. Harga sawit anjlok (sedangkan pupuk mahal) tp harga minyak goreng ttp ga turun. Skrg ekspor usah buka lg tp cpo terlanjur menumpuk di pabrik kadang tutup, jd bingung jual sawit kemana  <a href="https://t.co/q0ndY66c4X">https://t.co/q0ndY66c4X</a></p>	<p>ekspor bikin sawit harga anjlok pupuk minyak goreng turun buka pabrik tutup bingung jual</p>	<p>Negatif</p>
<p>@tvOneNews Harus massa minyak goreng sudah disubsidi masih mahal contohnya di wilayah saya sampai saat ini masih di kisaran minyak kemasan 24000/liter sedangkan minyak curah 16000/liter mana keadilan sosial bagi seluruh rakyat Indonesia????????</p>		

Tabel 3 menunjukkan data *sampel* yang diklasifikasikan program dan di *training* oleh program

**Tabel 4.** Contoh Jumlah Data

Jumlah kata pada label positif	Jumlah kata pada label negatif	Jumlah kata pada label netral	Total seluruh kata
497	610	396	1503

Untuk menghitung nilai *prior probability* kita bisa menghitung nilai *prior probability* dengan menggunakan persamaan sebagai berikut :

$P(H) = \text{Jumlah data} / \text{Jumlah seluruh data}$  Keterangan :

$P = \text{Prior probability}$

$H = \text{Hipotesis atau Klasifikasi dalam penelitian ini}$

Untuk menghitung peluang nilai *prior probability*

klasifikasi positif dilakukan sebagai berikut :  $P(\text{Positif}) = 497 / 1503$

$P(\text{Positif}) = 0.33067$

Berikutnya pada klasifikasi negatif dapat dihitung peluang kemunculannya sebagai berikut :

$P(\text{Negatif}) = 610 / 1503$

$P(\text{Negatif}) = 0.40585$

Terakhir pada klasifikasi netral peluang kemunculannya adalah sebagai berikut :

$P(\text{Netral}) = 396 / 1503$

$P(\text{Netral}) = 0.26348$

Setelah kita mengetahui nilai *prior probability* pada setiap label, selanjutnya kita mencari nilai *prior probability* kata dari masing-masing label atau klasifikasi dengan menggunakan rumus.

$$P(XI|VJ) = \frac{nk + 1}{n + \text{Kosakata}} \quad (6)$$

Mencari nilai probabilitas kata data uji pada data latih dengan dengan label atau klasifikasi Positif

$P(\text{emang} | \text{Positif}) = 0+1 / 497 + 1503 = 0.0005$

$P(\text{harga} | \text{Positif}) = 26+1 / 497 + 1503 = 0.0135$

$P(\text{goreng} | \text{Positif}) = 55+1 / 497 + 1503 = 0.028$

$P(\text{beda} | \text{Positif}) = 0+1 / 497 + 1503 = 0.0005$

$P(\text{minyak} | \text{Positif}) = 55+1 / 497 + 1503 = 0.0028$

$P(\text{kayak} | \text{Positif}) = 0+1 / 497 + 1503 = 0.0005$

$P(\text{bahas} | \text{Positif}) = 0+1 / 497 + 1503 = 0.0005$

Mencari nilai probabilitas kata data uji pada data latih dengan dengan label atau klasifikasi Negatif:

$P(\text{emang} | \text{Negatif}) = 2+1 / 610+1503=0.00143$

$P(\text{harga} | \text{Negatif}) = 22+1 / 610+1503=0.01041$

$P(\text{goreng} | \text{Negatif}) = 62+1 / 610+1503=0.02982$

$P(\text{beda} | \text{Negatif}) = 0+1 / 610+1503=0.00047$

$P(\text{minyak} | \text{Negatif}) = 64+1 / 610+1503=0.03076$

$P(\text{kayak} | \text{Negatif}) = 0+1 / 610+1503=0.000473$

$P(\text{bahas} | \text{Negatif}) = 0+1 / 610+1503=0.000473$

Mencari nilai probabilitas kata data uji pada data latih dengan dengan label atau klasifikasi Netral

$P(\text{emang} | \text{Netral}) = 0+1 / 396+1503=0.00526$

$P(\text{harga} | \text{Netral}) = 16+1 / 396+1503=0.089521$

$P(\text{goreng} | \text{Netral}) = 51+1 / 396+1503=0.027383$

$P(\text{beda} | \text{Netral}) = 1+1 / 396+1503=0.001053$

$P(\text{minyak} | \text{Netral}) = 52+1 / 396+1503=0.027909$

$P(\text{kayak} | \text{Netral}) = 0+1 / 396+1503=0.000526$

$$P(\text{bahas} | \text{Netral}) = 0 + 1/396 + 1503 = 0.000526$$

**Tabel 5. Perhitungan Kata**

Label	P(Xi Vj) P(Vj)	Hasil
Positif	0.33067 * (0.0005 * 0.0135 * 0.028 * 0.0005 * 0.028 * 0.0005 * 0.0005)	2.1873952095808E-19
Negatif	0.40585 * (0.00143 * 0.01041 * 0.02982 * 0.00047 * 0.03076 * 0.000473 * 0.000473)	5.8327346134674E-19
Netral	0.26348 * (0.000526 * 0.0089521 * 0.027383 * 0.001053 * 0.027909 * 0.000526 * 0.000526)	2.7721682325857E-19

Dari hasil perhitungan didapat bahwa hasil perhitungan untuk *tweet* didapat *skor* untuk klasifikasi negatif paling tinggi. Maka *tweet* tersebut diklasifikasikan sebagai *tweet* negatif

Setelah perhitungan klasifikasi sudah selesai dilakukan, langkah selanjutnya adalah melakukan pengujian terhadap metode klasifikasi tersebut. Evaluasi terhadap metode ini merupakan tahap menentukan apakah metode *Naïve Bayes* dapat memprediksi *tweet* dan mempunyai performa yang bagus atau tidak. Untuk melakukan pengujian terhadap hasil klasifikasi digunakan sebuah metode bernama *Confusion Matrix* dengan *Multiple Class*

**Tabel 6. Confusion Matrix**

Aktual Prediksi	Positif	Negatif	Netral
Positif	6	7	2
Negatif	1	16	0
Netral	0	1	3

Peforma metode klasifikasi dapat dilihat melalui perhitungan *accuracy*, *precision*, *recall* dan *error rate*. *Accuracy* merupakan persentase jumlah *record* data yang diklasifikasikan secara benar oleh Algoritma.

$$Accuracy = \frac{6+12+3}{6+7+2+1+16+0+0+1+3} \quad (7)$$

$$= \frac{21}{30} = 70\%$$

Setelah mendapatkan nilai *accuracy*, maka selanjutnya adalah menghitung *precision*. *Precision* menggambarkan proporsi dari klasifikasi yang dianggap benar berdasarkan kategori yang diprediksi kemungkinannya.

$$Precision = \frac{6+16+3}{9+1+1} \quad (8)$$

$$= \frac{21}{11} = 190.9\%$$

Setelah nilai *accuracy* dan *precision* didapatkan selanjutnya adalah menghitung *recall*. *Recall* menggambarkan proporsi kategori dari hasil klasifikasi yang teridentifikasi dengan benar

$$Recall = \frac{6+16+3}{1+8+2} \quad (9)$$

$$= \frac{21}{11} = 190.9\%$$

Setelah diketahui nilai *accuracy*, *precision* dan *recall* maka selanjutnya adalah menghitung *error rate*. *Error rate* menggambarkan persentase jumlah *record* data yang diklasifikasikan secara salah oleh Algoritma

$$Error\ Rate = 1 - Accuracy = 1 - 0.7 = 0.3 = 30\% \quad (10)$$

$$\begin{aligned} &6+7+2+1+16+0+0+1+3 \\ &\cdot 11 \\ &36 \\ &=30.6\% \end{aligned}$$

#### 4. KESIMPULAN

Dengan menerapkan Algoritma *Naïve Bayes* pada prediksi klasifikasi yang telah dicoba, dari hasil perhitungan menggunakan *confusion matrix* terhadap hasil prediksi *Naïve Bayes*, ditemukan bahwa tingkat akurasi yang didapatkan yakni sebesar 69.4% dengan *precision* yang dihasilkan sebesar 65.3% sedangkan *recall* yang dihasilkan sebesar 31.6% dan terakhir *Error Rate* sebesar 30.6%. dan hal mengklasifikasi sentimen masyarakat terhadap kebijakan pemerintah terkait minyak goreng subsidi dengan tingkat akurasi sebesar 69.4%.

Penelitian ini masih banyak kekurangan. Penulis menyarankan beberapa hal berikut untuk penelitian dan pengembangan di masa yang akan datang dengan dapat menambahkan data dari media *social Twitter*, karena di Indonesia sangatlah beragam seperti *Facebook, Instagram*

Penelitian selanjutnya juga diharapkan dapat menerapkan algoritma lainnya atau menggabungkan algoritma lainnya agar dapat meningkatkan tingkat akurasi, serta memberikan hasil klasifikasi

#### DAFTAR PUSTAKA

- [1] S. Yakub, A. Fitri Boy, I. Mariami, W. Stmik, dan T. Dharma, “*J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Penerapan Data Mining Pengaturan Pola Tata Letak Barang Pada Berkah Swalayan Untuk Strategi Penjualan Menggunakan Algoritma Apriori*,” v, vol. 69, no. 1, pp. 69–75, 2019.
- [2] R. Buaton, A. Sihombing, F. D. Aritonang, dan C. R. Wijaya, “*Data Mining Untuk Menentukan Korelasi (Confidence Dan Support) Jurusan Siswa Pada Tingkat Sekolah Menengah Terhadap Indeks Prestasi Kumulatif (IPK) Di Perguruan Tinggi Sebagai Solusi Tepat Pemilihan Program Studi Di Perguruan Tinggi*,” *Jurnal Sistem Informasi Kaputama (JSIK)*, vol. 1, no. 2, 2017,
- [3] M. Ridho Handoko, “*Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web*,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 1, pp. 50–58, 2021,
- [4] U. Banten Jaya, S. Royan, and A. Yulian, “*Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu*,” vol. 5, no. 2, 2021.
- [5] M. H. Rifqo and A. Wijaya, “*Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit*,” 2017.
- [6] S. Lestari, M. Badrul, P. Studi Sistem Informasi, dan S. Nusa Mandiri Jakarta, “*Implementasi Klasifikasi Naive Bayes untuk Prediksi Kelayakan Pemberian Pinjaman Pada Koperasian ugerah Bintang Cemerlang*,” vol. 7, no. 1, 2020.
- [7] R. A. Anggraini, G. Widagdo, A. Setya Budi, dan M. Qomaruddin, “*Penerapan Data Mining Classification untuk Data Blogger Menggunakan Metode Naïve Bayes*,” vol. 7, no. 1, 2019.
- [8] J. Avegad et al., “*Data Mining Klasifikasi Untuk Memprediksi Status Keberlanjutan Polis Asuransi Kesehatan Dengan Algoritme Naïve Bayes*,” *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K (SeNTIK)*, vol. 3, no. 1, 2019.
- [9] C. Selvi, D. Sembiring, L. Hanum, dan S. Parsaoran Tamba, “*Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus FTIK UNPRI)*,” *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.
- [10] E. Febrivani, R. Winanjaya, P. *Data Mining Asosiasi Pada Persediaan Obat, S. Tinggi Ilmu Komputer Tunas Bangsa, and I. Dosen Sekolah Tinggi Ilmu Komputer Tunas Bangsa*, “*JIKOMSI [Jurnal Ilmu Komputer dan Sistem Informasi] Penerapan Data Mining Asosiasi Pada Persediaan Obat*,” vol. 4, no. 1, pp. 25–36, 2021.