



Analisis sentimen review aplikasi mypertamina pada twitter menggunakan metode *naïve bayes classifier*

Nabilla Saumi Putri¹, Elin Haerani², Muhammad Fikry³, Elvia Budianita⁴

Email: ¹11950125147@students.uin-suska.ac.id, ²elin.haerani@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id,

⁴elvia.budianita@uin-suska.ac.id

¹²³⁴Teknik Informatika, UIN Sultan Syarif Kasim Riau

Diterima: 31 Maret 2023 | Direvisi: 14 April 2023 | Disetujui: 28 Mei 2023

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

BPH Migas atau Badan Pengatur Hilir dan Gas Bumi dengan PT. Pertamina meluncurkan aplikasi mypertamina untuk BBM atau bahan bakar minyak bersubsidi seperti solar serta pertalite yang kemudian akan digunakan sebagai layanan keuangan digital yang mengintegrasikan dengan aplikasi link aja guna menjamin tahap disalurkan nya BBM subsidi sesuai pada target. Masalah setelah peluncuran aplikasi mypertamina banyak menuai pro dan kontra salah satunya aplikasinya yang sulit digunakan, rumit untuk digunakan seperti yang dibuktikan dalam komentar di *Twitter*. Dari masalah ini dapat digunakan sebagai pengkajian analisis sentimen untuk mengetahui komentar positif serta negatif dalam *review* aplikasi mypertamina. Metode *naïve bayes classifier* dipakai pada pengkajian ini untuk tahap klasifikasi. Pada pengkajian ini memakai sejumlah 1001 data *tweet* dari proses pengumpulan data bersumber dari twitter yang terdiri dari 494 positif dan sebanyak 507 negatif, dalam proses klasifikasi *naïve bayes classifier* nilai akurasi 71% pada 10 % data uji serta data latih 90%.

Kata kunci: *aplikasi mypertamina, twitter, analisis sentimen, naïve bayes classifier*

Sentiment Analysis of MyPertamina Application Reviews on Twitter Using the Naïve Bayes Classifier Method

Abstract

Downstream and Natural Gas Regulatory Agency (BPH Migas) Pertamina launched the mypertamina application for subsidized fuel oil (BBM) such as pertalite and diesel which will then be used as a digital financial service that integrates with the link aja application to ensure the subsidised fuel distribution process is truly on target. The problem after the launch of the mypertamina application harvested many pros and cons, one of which was that the application was difficult to use, complicated to use as evidenced in the comments on *Twitter*. From this problem it can be used as a sentiment analysis research to find out positive comments and negative comments in the mypertamina application review. The *naïve Bayes classifier* method was used in this study for the classification process. In this study using as many as 1001 tweet data from the data collection process sourced from *Twitter* consisting of 494 positive and 507 negative, in the *naïve Bayes classifier* classification process the accuracy value is 71% with 90% training data and 10% test data.

Keywords: *mypertamina application, twitter, sentiment analysis, naïve bayes classifier*

1. PENDAHULUAN

Pertumbuhan teknologi komunikasi serta informasi yang kian maju mempengaruhi warga untuk mendapatkan informasi dan komunikasi yang semakin mudah. Ada beberapa media sosial diantaranya *Twitter*, *Facebook*, dan lain-lain [1]. *Twitter* adalah komunikasi populer saat ini. Pada tahun 2022, berdasarkan survei yang diambil dari dataindonesia.id mencatat bahwa Indonesia ialah suatu negara pada pemakai *twitter* terbesar di dunia dengan jumlah pengguna *twitter* Indonesia mencapai 18,45 juta [2].

Belakangan ini ketika ada berita tentang kenaikan BBM pemerintah melalui BPH Migas atau Badan Pengatur Hilir dan Gas Bumi dengan PT. Pertamina dilaporkan sedang merangkai aturan mengenai disalurkannya BBM atau bahan bakar minyak serta menciptakan aplikasi Mypertamina guna mensubsidi berupa solar serta pertalite. Aplikasi ini sebagai pelayanan financial digital melalui Pertamina serta peserta BUMN atau Badan Usaha Milik Negara mengintegrasikan pada aplikasi Link Aja guna menjamin bahwa tahap disalurkannya BBM subsidi sesuai target[3]. Masalah yang terjadi tentang peluncuran aplikasi mypertamina bersubsidi tepat sasaran oleh PT. Pertamina sempat menjadi trending di *twitter* yang menuai pro dan kontra salah satunya aplikasinya yang sulit digunakan, rumit dalam pemakaiannya tetapi ada juga yang memberikan komentar yang baik tentang aplikasi my Pertamina. Dari permasalahan ini bisa dijadikan sebagai penelitian analisis sentimen.

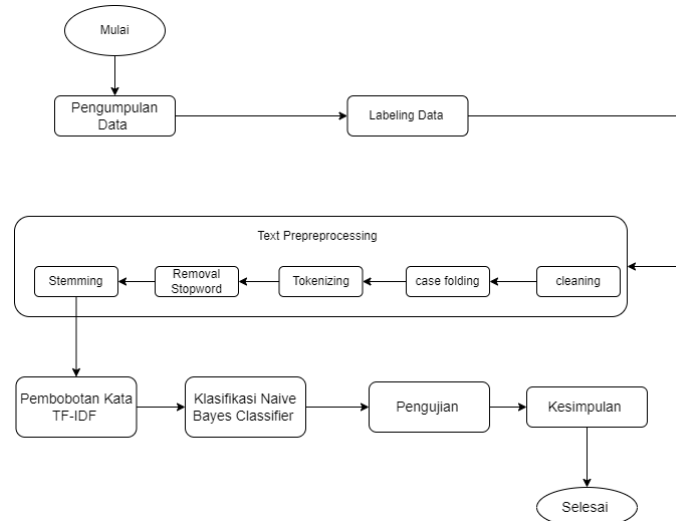
Analisis sentimen yang tujuannya untuk menganalisis, mengolah, mengekstraksi data tekstual yang berupa tanggapan suatu entitas dan suatu topik guna memperoleh sebuah informasi atau kecenderungan tanggapan terhadap sebuah masalah apakah termasuk sentimen positif serta negatif[4].

Terdapat sebagian pengkajian yang mengkaji mengenai analisa sentimen menggunakan *naïve bayes classifier*, berupa pengkajian yang dilaksanakan Dewi Aryanti berjudul “Analisis Sentimen Ibukota Negara Terbaru Memakai Cara Naïve Bayes Classifier” memperoleh nilai presisi 0,87, akurasi 94,18% serta *recall* 0,96[5]. Berikut pengkajian yang berjudul “Analisa Sentimen Ulasan Konsumen Pada Pelayanan Grab Indonesia Memakai Multinomial Naïve Bayes Classifier” memperoleh nilai akurasi dalam tahap ulasan yang diperoleh guna sentimen negatif memperoleh nilai *recall* 67% *precision* 57%, serta *f1-score* 62%. Melainkan guna sentimen positif memperoleh *recall* 95% *precision* 97% serta *f1-score* 96% yang dilakukan oleh Firman Noor Hasan, Monica Dwijayanti[6].

Pada pengkajian ini memakai metode *naïve bayes classifier* ialah suatu tehnik untuk mengklasifikasi *teorema bayes* dengan probabilitas sederhana yang mengkategorikan probabilitas dan mengamsusikan ketidakketergantungan (independen) yang tinggi [7]. Penelitian bertujuan untuk melihat sentimen *review* aplikasi mypertamina sesuai dengan komentar positif dan komentar negatif serta untuk mengetahui nilai akurasi dari proses klasifikasi metode *naïve bayes classifier*.

2. METODE PENELITIAN

Metode pengkajian ini dilaksanakan secara 5 tahapan bisa diamati terhadap Gambar 1.



Gambar 1 Desain Penelitian

2.1 Pengumpulan Data

Dalam pengkajian ini data didapati pada sosial media *twitter* menggunakan *library snsrape* untuk mendapatkan data dengan kata kunci “MyPertamina” memakai bahasa pemrograman *pyhton* pada *notebook Google Colaboratory*. *Library snsrape* adalah sebuah *library* yang dipakai guna mengambil data melalui sosial media *twitter* dan bahasa yang dipakai dalam proses pengambilan data di *twitter* yaitu bahasa Indonesia.

2.2 Labeling Data

Selanjutnya tahapan labeling data. Proses labeling data dilakukan secara manual dengan memberikan label positif dan label negatif ke data yang telah diperoleh. Apabila sebuah kalimat mengandung rasa ketidakpuasaan dengan aplikasi mypertamina

maka diberi label negatif serta apabila tak terdapat penolakan dari aplikasi mypertamina sehingga diberi label positif[8]. Dalam melakukan labeling data juga dibantu oleh seorang ahli bahasa untuk membantu dalam melabelkan.

2.3 Text Preprocessing

Hal ini ialah tahapan penseleksian data yang hendak ditindak lanjuti tiap berkasnya. Hal ini dibutuhkan guna membersihkan data pada kata-kata yang tak dibutuhkan, pada target guna langkah mencantumkan pada metode *naïve bayes classifier*[9]. Didalam *text preprocessing* terdapat beberapa tahap, yaitu:

1. *Cleaning*
Tahap ini bertarget guna membuang 1001 data *tweet* yang tak digunakan atau menghilangkan tanda baca, simbol, angka, URL, hastag (#), dan emoticon.
2. *Case Folding*
Tahap ini bertarget guna merubah seluruh huruf terhadap komentar sebagai huruf kecil (lowercase).
3. *Tokenizing*
Tahap ini bertarget guna memisahkan kalimat sebagai potongan kata
4. *Removal Stopword*
Tahap ini bertarget guna menghapus kata yang tidak mempunyai arti
5. *Stemming*
Tahap ini bertarget guna merubah kata yang mengumpul menjadi suatu landasan kata yang akan mengandung sebuah makna.

2.4 Pembobotan kata TF-IDF

Hal ini ialah cara guna membagikan nilai terhadap frekuensi kemunculan kata dalam dokumen teks[10]. *Term Frequency* (TF) untuk menghitung berapa kali muncul kata pada setiap dokumen, memberikan bobot guna mengukur intensitas timbulan melalui suatu kata dikatakan DF atau *Document Frequency*[11]. Nilai TF-IDF didapat sesudah nilai IDF serta TF dikalkulasi awalnya.

$$IDF(w) = \log\left(\frac{N}{DF(w)}\right) \quad (1)$$

$$Wdt = TF.IDF \quad (2)$$

$$TF - IDF(w, d) = \frac{TF-IDF(w, d)}{\sqrt{\sum_{w=1}^n TF-IDF(w, d)}} \quad (3)$$

Keterangan:

IDF(w) = kata diberi bobot terhadap dokumen

W = kata

TF(w,d) = total intensitas timbulan kata w pada kolom d,

IDF(w) = invers DF melalui kata w

N = total semua dokumen

DF(w) = total dokumen yang terkandung kata w

2.5 Klasifikasi Naïve Bayes Classifier

Naïve bayes Classifier adalah algoritma klasifikasi yang memproses melalui nilai kalkulasi probabilitas. Dengan algoritma ini, disebutkan jika sebuah atribut sifatnya independen serta tak berkaitan maka tak normalan pada nilai data bukan sebagai konflik. Landasan konsep yang dipakai *naïve bayes* ialah *Teorema bayes* ialah yang dipakai statistika guna mengkalkulasi probabilitas perkelas melalui data masuk yang dibagikan serta menetapkan kelas yang mempunyai nilai probabilitas terbesar[12]. Kelebihan *Naïve Bayes Classifier* ialah percepatan guna pengkalkulasian, algoritma sederhana serta presisi. Klasifikasi ini sangat akurat diaplikasi dalam data yang banyak[13]. Rumus bisa diamati terhadap persamaan 4.

$$P(H|X) = \frac{P(H|X).PX}{P(X)} \quad (4)$$

Keterangan:

X = Data pada class yang tak dikenal

H = Hipotesis data X ialah sebuah class spesifik

P(H|X) = Probabilitas hipotesis H melalui keadaan X (posterior probability)

P(H) = Probabilitas hipotesis H (prior probability)

P(X|H) = Probabilitas X melalui keadaan terhadap hipotesis H

P (X) = Probabilitas X

Terdapat tahapan algoritma *Naïve Bayes Classifier* tersebut adalah

1. Mengkalkulasi total kelas/label " $P(H)$ "
2. Mengkalkulasi total kelas per kelas " $P(X|H)$ "
3. Kalikan seluruh variabel kelas " $P(X|H) * P(H)$ "
4. Bedakan perolehan perkelas

2.6 Pengujian

Pada tahapan pengujian yang akan dilakukan untuk mengukur nilai akurasi yang dihasilkan oleh model klasifikasi terhadap data *tweet* sebagai referensi guna menjamin jika pengkajian ini dilaksanakan selaras pada target serta keinginan, uji ini hendak dilaksanakan memakai *confusion matrix*[14][15]. Hal tersebut digambarkan dengan tabel yang menyertakan klasifikasi total data uji yang tepat serta total data uji yang tak tepat[16]. Klasifikasi yang diidentifikasi *confusion matrix* mencakup 2 output label yaitu positif serta negatif. Bisa diamati terhadap Tabel 1.

Tabel 1 *Confusion Matrix*

	Prediction Positive	Prediction Negative
Actual Positive	TP (True Positive)	FN (False Negative)
Actual Negative	FP (False Positive)	TN (True Negative)

Confusion Matrix adalah parameter yang digunakan menghitung untuk akurasi, presisi, dan *recall*. Nilai akurasi adalah jumlah perbandingan prediksi yang benar(5), presisi adalah jumlah perbandingan teks yang selaras terkontrol pada seluruh dokumen teks yang dipilihkan pada sistem(6), serta *recall* adalah total perbandingan dokumen teks yang selaras terkontrol pada seluruh dokumen teks selaras yang berada dikoleksi (7)[17]. Berikut rumus untuk akurasi, presisi, dan *recall* dibawah ini.

$$\text{Akurasi} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (5)$$

$$\text{Presisi} = \frac{TP}{(TP+FP)} \times 100\% \quad (6)$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \times 100\% \quad (7)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pada pengambilan data pengkajian ini awalnya secara penarikan data pada sosial media *twitter* dengan kata kunci “mypertamina” rentang bulan Juni 2022 hingga bulan Januari 2023 menggunakan bantuan *library snsrape*. Proses pengambilan data menggunakan bahasa pemrograman *pyhton* dan proses pengolahan dilakukan dengan *notebook Google Colaboratory*. Setelah mendapatkan sebanyak 1001 data *tweet* setelah itu data disimpan ke format excel diberi nama file *crawling_data_mypertamina*.

2022-11-2	1,6E+18	GesurID	b'Mitha Gencarkan Sosialisasi Penggunaan Aplikasi MyPertamina https://t.co/kTWk9RpITd
2022-11-2	1,6E+18	beritabuatmyu	b'Anggota DPR RI Ini Serukan Warga Brebes Daftarkan Kendaraan di https://t.co/7VHoA7gRk0
2022-11-2	1,6E+18	portaportuna	b'Anggota DPR RI Ini Serukan Warga Brebes Daftarkan Kendaraan di https://t.co/5zcmvmbvgI
2022-11-2	1,6E+18	Suarajelata1	b'Berharap Subsidi BBM Tepat Sasaran, Paramitha Widya Kusuma Gelar Sosialisasi MyPertamina di https://t.co/ScUnCodiQB
2022-11-2	1,6E+18	MyPertaminal	b@sadderninja Hai, Sobat Aluna. Mengenal hal tersebut sudah kami bantu sampaikan melalui DM. Tks. -Tedi-
2022-11-2	1,6E+18	SyariahLinkAja	b'Isi BBM via MyPertamina, beli 20L dapat ekstra berhak 1L dengan bayar pakai LinkAja Syariah. Yuk transaksi sekarang! https://t.co/vUbP6hcwcl
2022-11-2	1,6E+18	MyPertaminal	b@sadderninja Pastikan sobat telah follow akun kami dan akun tidak di-private ya. Mohon maaf atas ketidaknyamanannya. Terima kasih. -Adi-
2022-11-2	1,6E+18	MyPertaminal	b@sadderninja Hai, Sobat Aluna. Mohon maaf atas ketidaknyamanannya yang dialami. Mengenal perihal tersebut, silakan disampaikan detail permasalahan melalui DM. Terima k
2022-11-2	1,6E+18	mantiq_media	b'Paramitha Serukan Warga Brebes Daftarkan Kendaraan di https://t.co/qUSFPpIPN8
2022-11-2	1,6E+18	MyPertaminal	b@sadderninja Hai, Sobat Aluna. Silakana ada yang dapat kami bantu ? Tks. -Tedi-
2022-11-2	1,6E+18	MyPertaminal	b@notthingsisr0al0ai, Sobat Ujar. 021-85500135 adalah dari Pertamina Call Center 135. Pertamina Call Center 135 menghubungi untuk menawarkan BBM, LPG dan Pelumas untuk
2022-11-2	1,6E+18	pertamaxserie	b'STOP AND POST! https://t.co/nCoba dong post foto pas kamu lagi melakukan perjalanan paling jauh menggunakan motor! Jangan lupa tujuan dan pengalamannya juga ya! https://t.co/nLsgsoooh
2022-11-2	1,6E+18	MyPertaminal	b'ydu, mulai beralih ke #LebiBaikPertamax sekarang! https://t.co/jadiPilihanPas #PerfectionInPerformance #MyPertamina #Call135
2022-11-2	1,6E+18	MyPertaminal	b'Jaki, coganet banget deh baru kamu lebih tenang dan nyaman selama bepergian atau jalan-jalan di malam minggu. Ekhem semoga yang ini Pertamax hubungannya juga awet kaye
2022-11-2	1,6E+18	MyPertaminal	b'Kendaraan kamu haus? PERTAMAX SOLUSINYA! https://t.co/nSebelum malam minggu, isi dulu kendaraanmu pakai Pertamax ya! Soalnya, bahan bakar keren yang satu ini punya RON 92
2022-11-2	1,6E+18	brightgas_LPG	b'Kamu bisa mendapatkan berbagai macam variasi Bright Gas sesuai kebutuhan mulai dari Minimarket, Supermarket, Bright Store, Kios Matic, dan Aplikasi MyPertamina dengan Pe
2022-11-2	1,6E+18	dwarinto	b'_@MyPertaminalo o min, pembelian pertalite ke mypertamina ga bisa pakai OVO https://t.co/ovo_id ya pembayarannya?
2022-11-2	1,6E+18	pertamaxseries	b'So, jangan lupa untuk ajak keluarga kamu baru seru-serunya di Pertamina Eco Run Festival 2022 HARI INI! See you, Sob https://t.co/x9fX92x9Jn https://t.co/nPnPasUntukSemua #EnergizingYou #
2022-11-2	1,6E+18	brightgas LPG	b'Sampai berjumpa di sana ya https://t.co/x9fX91x9Jn https://t.co/nPasUntukSemua #EnergizingYou #MyPertamina #Call135 #PertaminaEcoRunFest'

Gambar 2 Hasil Data Tweet

3.2 Labeling Data

Proses *Labeling* data dilaksanakan secara manual dengan dibantu oleh seorang ahli bahasa sebanyak 1001 data *tweet* yang memperoleh 494 komentar positif dan 507 komentar negatif. Setelah diberi label data disimpan dalam format excel diberi nama file pertamina.

GesurilD	1 Mitha Gencarkan Sosialisasi Penggunaan Aplikasi MyPertamina https://t.co/kTWk9RpITd	positif
Suarajelata1	2 Berharap Subsidi BBM Tepat Sasaran, Paramitha Widya Kusuma Gelar Sosialisasi MyPertamina di Bumiayu https://t.co/ScUnCodlqB	positif
SyariahLinkAja	3 Isi BBM via MyPertamina, beli 20L dapat ekstra berkah 1L dengan bayar pakai LinkAja Syariah. Yuk transaksi sekarang! #MakinBerkah https://t.co/vUbP6hcwv	positif
MyPertaminaID	4 Kendaraan kamu haus? PERTAMAX SOLUSINYA! \nSebelum malam minggu, isi dulu kendaraanmu pakai Pertamina ya! Soalnya, bahan bakar keren yang sat	positif
dambadwsm	5 @hematnaps @psychicmew2 @Edqbal @sopir_idiot Bisa juga, tapi pertamina mengharuskan kita untuk download mypertamina tujuan utamanya bukan itu	positif
irmatsamdo	6 @MyPertaminaID saran saya ke MyPertamina agar didata lagi, dievaluasi dan direview lagi, SPBU yg benar-benar bisa bayar pakai aplikasi krn banyak SPBU yg	positif
ardian_mp	7 @remosisbrian @d_prayogi @solehsolihun Untuk mypertamina, info saja ada kawan yg mau beli pake aplikasi, tak bisa digunakan, ternyata info dari petugas	negatif
irmatsamdo	8 MyPertaminaID saran saya ke MyPertamina agar didata lagi, dievaluasi dan direview lagi, SPBU yg benar-benar bisa bayar pakai aplikasi krn banyak SPBU yg	positif
aryanto79	9 @kulo_zhargopro @d_prayogi @solehsolihun Itu yg selalu saya temui...dan akhirnya duit mengendap di mypertamina dan link aja pun bangkrut	negatif
kulo_zhargopro	10 @d_prayogi @solehsolihun Percuma, kemarin mau bayar pake mypertamina juga gak bisa.. gangguan mulu bilanganya.. mana udah isi top link aja banyak is	negatif
pertamaxseries	11 @dewalova_jeiend Hi, Sobat. Ayo segera tukarkan poin MyPertamina kamu ke voucher, merchandise dan hadiah menarik lainnya. Jangan sampai poin kamu	positif
ekoce	12 untung ga pernah niat daftar mypertamina https://t.co/38EnDtzY8O	negatif
harisNO	13 Woi data mypertamina bocor	negatif
Wisnu_L_K	14 Terimakasih MyPertamina, pelayanannya sangat lama dibanding pake cash :)Saran saya apus aja itu aplikasi, buang buang anggaran APBN saja :)	negatif
rimasi	15 Udah gua duga daftarin mobil di MyPertamina pasti cuma buat nyari data orang2 middle class&	negatif
detikinet	16 Kebocoran data masih saja terjadi, seperti dugaan pada platform MyPertamina dan PeduliLindungi. Lalu, siapa yang seharusnya bertanggungjawab apabila	negatif
itsmeidr	17 Saat ini, Program subsidi tepat MyPertamina adalah program yang dengan tujuan agar memberikan subsidi BBM pada kalangan yang tepat, karena untuk saa	negatif
nolantv_com	18 [Software Aplikasi] Bjorka Diduga Retas 44 Juta Recordsdata dari Software Aplikasi MyPertamina! Apa Langkah Pemerintah? https://t.co/PSwk3zblidP #Video #	negatif
dialtudyas	19 Ngisi BBM pakai #MyPertamina banyak SPBU di Jogja yang belum menggunakan, trus pie je kon download aplikasinya tapi kebanyakan SPBU nya gak suport#	negatif
andhikatsu	20 untung MyPertamina gw cuma instal, ga sampe daftar	negatif
MyPertaminaID	21 Langsung aja tukerin Poin MyPertamina kamu sekarang juga dan vote jagoanmu sebanyak-banyaknya untuk raih kesempatan jadi salah satu pemenang TOP \	positif
SuryaPringas	22 Halah kampret lah, katanya di batasi, ujung nya pake mypertamina, baru install pas nanya tukang ngisi bensinnya, ternyata utk roda 4 dan 2 ga usah, nanya : negatif	negatif
KompasData	23 Hingga saat ini pemerintah belum berikan keterangan terkait dengan dugaan kebocoran data dari aplikasi MyPertamina dan PeduliLindungi. Komisi I DPR, m	negatif
taheoo	24 Udah kali kedua daftar di mypertamina di tolak, gak tau sih alesanya apa. Masa iya setiap beli bahan bakar terbaca solar terus. Mohon petunjuk @MyPertam	negatif
Wiznu_Prayoga	25 Katanya sih udah dibocorkan terus dijual.Tp ga masuk akalnya katanya 44juta data pelanggan, sedangkan pelanggan baru MyPertamina baru 3jtan.3jt dg tan	negatif
MBE_AJAA	26 Abis ini pertamina minta anggaran buat perkuat sistem mypertamina abis itu bjorka ngilang lagi https://t.co/GWtu4XUp3a	negatif

Gambar 3 Hasil Pelabelan Data

3.3 Text Preprocessing

Setelah data *tweet* di labelkan, selanjutnya dilakukan *text preprocessing* dengan tujuan membersihkan data mentah yang didapatkan dari proses pengumpulan data. Berikut hasil proses *text preprocessing* pada Tabel 2.

Tabel 2 Proses *Text Preprocessing*

@kulo_zhargopro @d_prayogi @solehsolihun Itu yg selalu saya temui...dan akhirnya duit mengendap di mypertamina dan link aja pun bangkrut	
Proses	Hasil
<i>Cleaning</i>	Itu yg selalu saya temui... dan akhirnya duit mengendap di mypertamina dan link aja pun bangkrut
<i>Case Folding</i>	itu yg selalu saya temui dan akhirnya duit mengendap di mypertamina dan link aja pun bangkrut
<i>Tokenizing</i>	itu, yg, selalu, saya, temui, dan, akhirnya, duit, mengendap, di, mypertamina, dan, link, aja, pun, bangkrut
<i>Stopword Removal</i>	temui, akhirnya, duit, mengendap, mypertamina, link, bangkrut
<i>Stemming</i>	temui, akhir, duit, endap, mypertamina, link, bangkrut

Selanjutnya, dilakukan pengujian dengan menerapkan teknik *preprocessing* yang berbeda-beda. Pada tahapan ini dilakukan 2 kali pengujian yaitu seluruh tahapan *preprocessing* dan melakukan teknik *preprocessing* tanpa *stopword removal*. Berikut Tabel 3 teknik preprocessing

Tabel 3 teknik preprocessing

Preprocessing	Akurasi
Full <i>Preprocessing</i> 70% data latih dan 30 % data uji	69%
Full <i>Preprocessing</i> 80% data latih dan 20 % data uji	68%
Full <i>Preprocessing</i> 90% data latih dan 10 % data uji	71%
Tanpa <i>Stopword removal</i> 70% data latih dan 30% data uji	72%
Tanpa <i>Stopword removal</i> 80% data uji dan 20% data uji	71%
Tanpa <i>Stopword removal</i> 90% data latih dan 10% data uji	70%

Terhadap Tabel 3 terlihat kinerja terbaik di peroleh ketika menggunakan teknik *preprocessing* tanpa *stopword removal* terhadap skenario pembagian data uji 30% serta data latih 70% yang menghasilkan akurasi sebesar 72%. Hal tersebut terjadi disebabkan karena ketika dilakukan *stopword removal* terdapat kata-kata di dalam kamus *stopword* bahasa yang apabila dihilangkan dapat mengurangi informasi dari kalimat tersebut[18].

3.4 Pembobotan TF-IDF

Selanjutnya dalam proses ini menggunakan library python yaitu *TfidfVectorizer* dan *CountVectorizer*.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
```

Gambar 4 library python pembobotan TF-IDF

Dari *library python* diatas dapat dihasil penggalan data dari perolehan pembobotan TF- IDF terhadap Tabel 4.

Tabel 4 Penggalan Data Hasil Pembobotan TF-IDF

Kata	id_weigths
ada	0,15399
akhir	0,230394
akun	0,413518
voucher	0,256058

3.5 Klasifikasi *Naïve Bayes Classifier*

Data yang telah melalui proses *preprocessing* dan *tf-idf*, tahap selanjutnya klasifikasi memakai tehnik *Naïve Bayes Classifier* guna memperoleh model yang bisa mengklasifikasi data terbaru tak dengan pelabelan sendiri. Pada proses klasifikasi pada pengkajian ini dilakukan tiga skenario terbaginya data latih dan data uji. Skenario pertama yaitu pembagian 30% data uji serta data latih 70%, kedua menggunakan pembagian 20% data uji serta data latih 80%, lalu ketiga menggunakan pembagian data uji 10% serta data latih 90%. Pengujian ini memakai *confusion matrix* secara mengkalkulasi presisi, akurasi serta *recall*.

	precision	recall	f1-score	support
0	0.71	0.72	0.72	143
1	0.00	0.00	0.00	19
2	0.68	0.81	0.74	131
3	0.00	0.00	0.00	8
accuracy			0.69	301
macro avg	0.35	0.38	0.36	301
weighted avg	0.63	0.69	0.66	301

Gambar 5 hasil klasifikasi pembagian data 70%:30%

Selanjutnya klasifikasi dibaginya data 80% :20%.

	precision	recall	f1-score	support
0	0.68	0.76	0.72	91
1	0.00	0.00	0.00	13
2	0.69	0.76	0.72	90
3	0.00	0.00	0.00	7
accuracy			0.68	201
macro avg	0.34	0.38	0.36	201
weighted avg	0.61	0.68	0.65	201

Gambar 6 hasil klasifikasi pembagian data 80%:20%

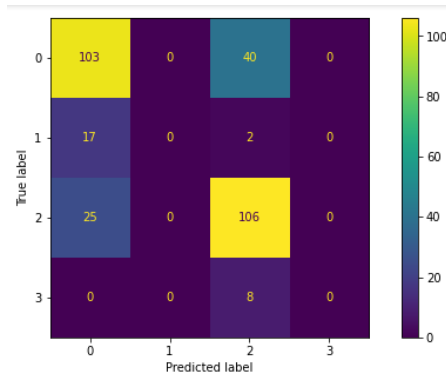
Selanjutnya klasifikasi dibaginya data 90%:10%.

	precision	recall	f1-score	support
0	0.69	0.80	0.74	44
1	0.00	0.00	0.00	7
2	0.74	0.79	0.76	47
3	0.00	0.00	0.00	3
accuracy			0.71	101
macro avg	0.36	0.40	0.37	101
weighted avg	0.64	0.71	0.68	101

Gambar 7 hasil klasifikasi pembagian data 90%:10%

3.6 Pengujian

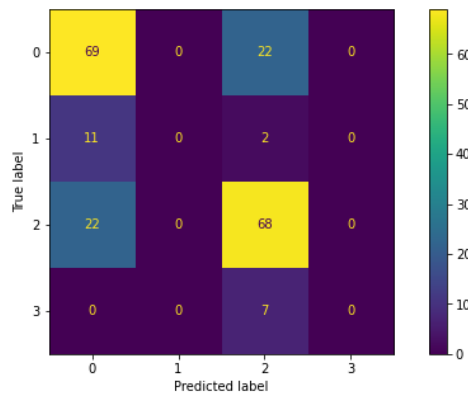
Dalam pengkajian ini pengujian dilakukan berdasarkan dibaginya data uji serta latih yang berjumlah 1001 data. Pembagian data pada pengkajian ini memakai perbandingan 3 skenario yaitu 30% data uji: 70% data latih, 20% data uji: 80% data latih serta 10% data uji: 90% data latih.



Gambar 8 hasil pengujian 70%:30%

$$\begin{aligned} \text{Perhitungan Akurasi} &= \frac{103+106}{103+40+17+2+25+106+0+8} \times 100\% \\ &= \frac{209}{301} \times 100\% \\ &= 69\% \end{aligned}$$

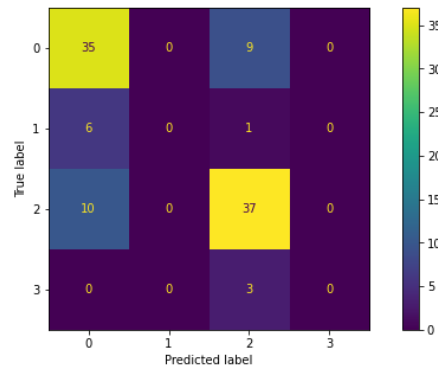
Selanjutnya hasil pengujian *confusion matrix* 80%:20%.



Gambar 9 hasil pengujian 80%:20%

$$\begin{aligned} \text{Pengujian akurasi} &= \frac{69+68}{69+22+11+2+22+68+0+7} \times 100\% \\ &= \frac{137}{201} \times 100\% \\ &= 68\% \end{aligned}$$

Selanjutnya hasil pengujian *confusion matrix* 90%:10%.



Gambar 10 hasil pengujian 90%:10%

$$\begin{aligned} \text{Pengujian akurasi} &= \frac{35+37}{35+9+6+1+10+37+0+3} \times 100\% \\ &= \frac{72}{101} \times 100\% \\ &= 71\% \end{aligned}$$

4. KESIMPULAN

Melalui perolehan pengkajian pada analisis sentimen review aplikasi mypertamina terhadap *twitter* dalam proses klasifikasi memakai metode *naïve bayes classifier* dengan jumlah 1001 tweet terdiri dari 494 komentar positif dan 507 komentar negatif. Hasil pengujian pada penelitian ini pada tahapan *preprocessing* dengan dua kali pengujian kinerja sistem terbaik diperoleh ketika teknik *preprocessing* tanpa *stopword removal* secara dibaginya 30 % serta 70% menghasilkan akurasi sebesar 72%. Hal ini disebabkan karena ketika dilakukan *stopword removal* terdapat kata-kata yang apabila dihilangkan dapat mengurangi informasi dari kalimat tersebut. Selanjutnya pengujian klasifikasi memakai metode *naïve bayes classifier* secara tiga skenario pembagian data uji serta data latih yaitu pembagian data latih 70%:30% data uji memperoleh *recall* 69%, akurasi 69%serta presisi 77% untuk pembagian data latih 80%:20% data uji menghasilkan nilai akurasi 68%, nilai *recall* 68% dan nilai presisi 76%. Terakhir untuk terhadap dibaginya data latih 90%:10% data uji menghasilkan nilai akurasi 71%, nilai *recall* 71%, nilai presisi 79%. Dimana nilai akurasi tertinggi memakai pembagian data latih 90%:10% data uji menghasilkan nilai akurasi 71%, nilai *recall* 71%, nilai presisi 79%. Saran untuk penelitian selanjutnya dapat mengambil data di *Google playstore* agar data yang di dapat lebih bervariasi dan dapat menggunakan metode klasifikasi yang lain sehingga diperoleh klasifikasi yang lebih spesifik dan lebih baik diperoleh.

DAFTAR PUSTAKA

- [1] I. F. Rozi, E. N. Hamdana, and M. B. I. Alfahmi, "PENGEMBANGAN APLIKASI ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (Studi Kasus SAMSAT Kota Malang)," *Jurnal Informatika Polinema*, pp. 149–154, 2018.
- [2] M. A. Rizaty, "Indonesia menjadi salah satu negara dengan pengguna Twitter terbesar di dunia. Berdasarkan laporan We Are Social, jumlah pengguna Twitter di Indonesia mencapai 18,45 juta pada 2022.," *dataindonesia.id*, 2022. <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022> (accessed Mar. 11, 2023).
- [3] Zulfikar and Hardiansyah, "Apa Itu MyPertamina yang Bakal Dipakai buat Beli Pertalite dan Solar?," *KOMPAS.com*, 2022. <https://tekno.kompas.com/read/2022/06/06/12150027/apa-itu-mypertamina-yang-bakal-dipakai-buat-beli-pertalite-dan-solar?page=all> (accessed Nov. 19, 2022).
- [4] T. Yulianita, T. W. Utami, and M. Al Haris, "Analisis Sentimen Dalam Penanganan Covid-19 di Indonesia Menggunakan Naive Bayes Classifier," *Seminar Nasional Variansi (Venue Artikulasi-Riset, Inovasi, Resonansi-Teori, dan Aplikasi Statistika)*, pp. 235–243, 2020.
- [5] D. Aryanti, "Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier," *Journal of Information System Research*, vol. 3, no. 4, pp. 524–531, 2022, doi: 10.47065/josh.v3i4.1944.
- [6] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier," *Jurnal Linguistik Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [7] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 234–241, Dec. 2022, doi: 10.37859/coscitech.v3i3.4343.
- [8] A. Q. Surbakti, R. Hayami, and J. Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 2, no. 2, pp. 91–97, 2021, doi: 10.37859/coscitech.v2i2.2851.

- [9] E. Salehudin Basryah, A. Erfina, and C. Warman, "Analisis Sentimen Aplikasi Dompot Digital Di Era 4.0 Pada Masa Pandemi Covid-19 Di Play Store Menggunakan Algoritma Naive Bayes Classifier," *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, pp. 189–196, 2021.
- [10] I. Novitasari, T. B. Kurniawan, and D. A. Dewi, "Analisis Sentimen Masyarakat Terhadap Tweet Ruang Guru Menggunakan Algoritma Naive Bayes Classifier (NBC)," *Institute of Computer Science (IOCS)*, vol. 6, no. 3, 2022.
- [11] E. B. Susanto, P. A. Christianto, M. R. Maulana, and S. W. Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 234–241, 2022, doi: 10.37859/coscitech.v3i3.4343.
- [12] W. Dwitiyasa, "Klasifikasi Jenis Gangguan pada Base Transceiver Station (BTS) menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes, dan Genetic Algorithm," 2018.
- [13] A. A. Permana *et al.*, *Machine Learning*. Global Eksekutif Teknologi, 2023.
- [14] A. Shiddicky and S. Agustian, "Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Metode Logistic Regression," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 99–106, 2022, doi: 10.37859/coscitech.v3i2.3836.
- [15] M. N. Rusardi, B. Rahayudi, and P. P. Adikara, "Analisis Sentimen Masyarakat terhadap Isu New Normal Scenario berdasarkan Opini dari Twitter menggunakan Algoritma Naive Bayes Classifier," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 3, pp. 1434–1440, 2022.
- [16] M. Christianto, J. Andjarwirawan, and A. Tjondrowiguno, "Aplikasi analisa sentimen pada komentar berbahasa Indonesia dalam objek video di website YouTube menggunakan metode Naïve Bayes classifier," *Jurnal Infra*, vol. 8.1, pp. 255–259, 2020.
- [17] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indonesian Journal of Applied Statistics*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.
- [18] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.