Analisis Sentimen pada Komentar Aplikasi MyPertamina dengan Metode Multinomial Naïve Bayes

Putu Alvita Wagiswari R D[1], Indah Susilawati[2], Arita Witanti[3]
[1, 2, 3]Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta
[1]alvitawagiswari@gmail.com, [2]indah@mercubuana-yogya.c.id, [3]arita@mercubuana-yogya.ac.id

ABSTRACT

The government's decision to allow users of the MyPertamina app to buy subsidized fuel starting in July 2022 has generated both support and opposition in the local community. On the Google Play Store, users of the MyPertamina app leave a variety of comments. In order to ascertain the degree of precision, accuracy, recall, and f1-score as an evaluation result, this study will analyze the sentiments of each user's opinion of MyPertamina. The Naive Bayes method was employed in the classification process to analyze 1370 pieces of data. The accuracy rate, precision, recall, and f1-score are all 81%, 82%, 79%, and 80% for the distribution of 30 percent testing data and 70 percent training data. While the accuracy rate, precision, recall, and f1-score are all 79%, 81%, 78%, and 79% for the distribution of 20 percent testing data and 80 percent training data.

Keyword: *Classification, Naïve Bayes, Sentiment Analysis, MyPertamina*

1. Introduction

Setelah terjadinya revolusi industri, perkembangan teknologi informasi dan komunikasi melaju dengan pesat. Berbagai macam sektor dalam tatanan kehidupan manusia menerima dampak dari teknologi tersebut. Salah satu sektor penting yang menerima dampak dari adanya revolusi industri adalah sektor pemerintahan. Dalam bidang pemerintahan, proses digitalisasi disebut juga dengan e-*government*. E-*government* dapat dikatakan sebagai penggunaan TIK dalam penyelenggaraan pemerintahan yang menggunakan elektronik [1]. Dalam segi praktis, keberadaan e-*government* memungkinkan pelaksanaan seluruh proses pemerintahan secara efektif dan efisien sehingga dapat mengurangi praktik-praktik yang tidak benar dalam administrasi negara [2].

Salah satu bentuk digitalisasi dalam sektor pemerintahan adalah pemakaian aplikasi MyPertamina untuk pembelian BBM bersubsidi. Tujuan digunakannya aplikasi MyPertamina adalah agar proses dokumentasi pembelian BBM bersubsidi tepat sasaran di masyarakat [3]. Namun, kebijakan pemerintah mengenai penggunaan aplikasi MyPertamina memberikan komentar pro dan kontra di masyarakat. Akibatnya, aplikasi MyPertamina dihujani ribuan komentar oleh pengguna Android. Agar dapat mengetahui jenis sentimen yang dominan dalam komentar aplikasi MyPertamina, maka diperlukan analisis sentimen untuk mengklasifikasi jenis-jenis sentimen yang dituliskan oleh masyarakat.

Metode klasifikasi Naïve Bayes akan digunakan untuk melakukan analisis sentimen. Metode Naïve Bayes adalah salah satu metode yang sering digunakan dalam proses klasifikasi karena metode ini dianggap memiliki tingkat akurasi yang lebih tinggi dibanding metode lainnya [4]. Dengan metode ini, kecenderungan opini pengguna dapat diketahui sehingga dapat menjadi evaluasi ke depannya untuk meningkatkan kualitas pelayanan aplikasi MyPertamina agar semakin baik.

2. Research Method

Penelitian ini dilakukan dengan beberapa tahapan. Tahapan tersebut akan dibagi menjadi beberapa tahapan. Berikut merupakan tahapan dalam proses penelitian ini.

2.1. Pengumpulan Data (*Scraping Dataset*)

Proses pengumpulan data menggunakan *library Google Play Scrap* yang dimiliki oleh Python. Dalam proses pengumpulan data, data yang diambil berupa data komentar, rating, dan juga tanggal dibuatnya komentar tersebut.

2.2. Labeling Dataset

Proses *labeling* menggunakan bantuan dari Microsoft Excel. Data dengan rating satu dan dua akan dilabeli negatif sedangkan data dengan rating empat dan lima akan dilabeli positif.

2.3. Deteksi Bahasa

Proses deteksi bahasa *menggunakan library language detection* (*langdetect*) yang dimiliki oleh Python. Data yang diambil hanyalah data yang menggunakan Bahasa Indonesia.

2.4. Preprocessing

Proses *preprocessing* dilakukan dengan beberapa tahapan. Tahapan tersebut diantaranya adalah proses *data cleaning*, *case folding*, *tokenization*, normalisasi data, *stopwords removal*, dan *stemming*. Proses *preprocessing* melibatkan transformasi data teks mentah menjadi format yang lebih gampang untuk dimengerti. Pada saat proses *preprocessing*, data mentah akan diproses lebih lanjut sehingga dapat menyelesaikan masalah yang ada [5].

2.5. Pembobotan Kata

Proses ekstraksi fitur dalam penelitian ini menggunakan metode TF-IDF. Proses pembobotan menggunakan *library Scikit-learn* yang dimiliki oleh Python dengan bantuan algoritma *CountVectorizer*. *Term Frequency-Invers Document Frequency* (TF-IDF) merupakan salah satu metode pembobotan yang dapat digunakan dalam proses pembobotan [6]. Persamaan TF-IDF adalah sebagai berikut.

$$\mathrm{IDF_A} = \log \frac{\mathrm{N}}{\mathrm{df_A}} \qquad (1)$$

$$\mathrm{TF-IDF} = \mathrm{TF_A} \times \mathrm{IDF_A} \qquad (2)$$

Dengan N adalah jumlah keseluruhan dokumen. df_A adalah frekuensi dokumen dengan kata A. TF_A adalah frekuensi kata A.

2.6 Klasifikasi

Dalam proses klasifikasi, metode Naïve Bayes digunakan dengan memanfaatkan *library Scikit-learn*. Algoritma yang digunakan adalah Multinomial Naïve Bayes. Penerapan teorema ini sering digunakan karena independensinya yang kuat [7]. Dengan kata lain, Naïve Bayes merupakan aturan asumsi yang tidak memiliki ketergantungan antar fitur satu dengan lainnya. Prediksi dengan Teorema Naïve Bayes memiliki persamaan sebagai berikut.

$$\mathrm{P(H|E)} = \frac{\mathrm{P(E|H)P(H)}}{\mathrm{P(E)}} \qquad (3)$$

Dengan P(H|E) adalah Probabilitas H terjadi dengan bukti E, yang juga dikenal sebagai probabilitas kondisional. P(E|H) adalah Terjadinya bukti E yang memengaruhi hipotesis H melalui probabilitas bukti E yang terjadi. P(H) adalah Tanpa mempertimbangkan hipotesis atau bukti lainnya, probabilitas awal (priori) dari terjadinya bukti H. P(E) adalah Tanpa mempertimbangkan hipotesis atau bukti lainnya, probabilitas awal (priori) dari terjadinya bukti E.

Dengan menggunakan Naïve Bayes dalam proses klasifikasi, nilai polaritas akan digunakan untuk mengumpulkan pendapat yang memiliki arti yang sama [8]. Namun, dalam perhitungan *Natural Language Processing* (NLP), hal pertama yang harus dilakukan sebelum perhitungan Naïve Bayes adalah proses ekstraksi fitur. Jika dengan cara biasa, kalimat akan dilihat secara keseluruhan. Ketika data tidak muncul dalam dataset, maka probabilitas akan bernilai nol [9]. Sedangkan dalam penerapan NLP, setiap kata akan dianggap berdiri sendiri [10]. Maka, proses perhitungan probabilitas dapat dilihat pada persamaan berikut.

$$\widehat{\theta \mathrm{y}} = \frac{\mathrm{Nyi} + \mathrm{a}}{\mathrm{Ny} + \mathrm{an}} \qquad (4)$$

Dengan Nyi adalah kemunculan fitur i pada kelas y. Sementara a adalah probabilitas kemunculan yang disepakati bernilai sama dengan satu. Ny adalah jumlah fitur kelas y. an adalah jumlah keseluruhan kata.

Dari proses klasifikasi, akan dihasilkan confusion matrix sebagai hasil dari evaluasi model. *Confusion matrix* merupakan metode yang sering dipakai dalam proses *text mining*. Metode ini lebih baik digunakan untuk mengevaluasi kinerja *classifier*. Untuk mengukur kinerja *classifier* dengan *confusion matrix*, maka harus menghitung jumlah kelas X yang diklasifikasikan sebagai kelas Y [11].

	Positive	*Negative*
Positive	TP	FP
Negative	FN	TN

Dengan TP yang merupakan prediksi data positif yang akurat. FP adalah data yang seharusnya negatif ternyata diidentifikasi sebagai data positif. FN adalah Data yang seharusnya positif ternyata diidentifikasi sebagai data negatif. TN adalah data negatif yang diprediksi akurat.

Confusion matrix dapat digunakan untuk melakukan pengkuran terhadap tingkat presisi, akurasi, *recall*, dan f1-*score* [12]. Rumus yang dapat digunakan adalah sebagai berikut.

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

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

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

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (8)$$

3. Result and Analysis

3.1. Pengumpulan Data (Scraping Dataset)

Dalam proses pengumpulan data, terdapat sejumlah 323.085 komentar yang dimiliki oleh aplikasi MyPertamina. Namun, dalam lingkup penelitian ini, hanya data yang diambil dari bulan Juli hingga Desember 2022 yang digunakan. Didapatkan data sejumlah 1600 yang akan digunakan dalam proses klasifikasi.

3.2. Deteksi Bahasa

Dalam proses deteksi bahasa ditemukan sekitar 1370 data yang menggunakan Bahasa Indonesia. Rincian bahasa dalam dataset adalah sebagai berikut.

Tabel 1 Daftar Jenis Bahasa

Bahasa	Jumlah Komentar
Indonesia	1370
Tagalog	76
Romanian	21
Somali	21
English	17
German	16
Africans	15
Slovenian	10
Estonian	8

Croatian	7

3.3 Preprocessing

Proses *preprocessing* dibagi menjadi beberapa tahapan. Langkah-langkah preprocessing adalah sebagai berikut.

3.3.1 Data Cleaning dan Case Folding

Proses pembersihan data merupakan salah satu tahapan yang sangat krusial dalam proses pra-pemrosesan data [13]. *Data cleaning* digunakan untuk menghilangkan simbol-simbol yang tidak diperlukan dalam data, seperti *emoticon*, tanda baca, dan juga *emoji* [14]. Proses *data cleaning* menggunakan bantuan *regular expression* (regex) untuk menghilangkan simbol-simbol yang tidak diperlukan. Sementara itu, *case folding* merupakan proses mengubah data yang memiliki huruf kapital menjadi huruf kecil (*lowercase*) [15]. Hasil *data cleaning* dan *case folding* dapat dilihat pada tabel berikut.

Tabel 2. Hasil Data Cleaning dan Case Folding

Komentar	Data Cleaning dan Case Folding
Membantu saat dibutuhkan.	membantu saat dibutuhkan
Semoga kedepan lebih bagus	semoga kedepan lebih bagus

3.3.2 Tokenisasi

Proses tokenisasi merupakan sebuah proses untuk memisahkan frasa atau kalimat menjadi unit yang lebih kecil [16]. Hasil tokenisasi dapat dilihat pada tabel berikut.

Tabel 3. Hasil Tokenisasi

Data Cleaning dan Case Folding	Tokenisasi
membantu saat dibutuhkan	[membantu, saat, dibutuhkan]
semoga kedepan lebih bagus	[semoga, kedepan, lebih, bagus]

3.3.3 Normalisasi Data

Normalisasi data dilakukan untuk mengoreksi kata-kata agar sesuai dengan aturan yang terdapat di Kamus Besar Bahasa Indonesia (KBBI) [17]. Hasil normalisasi data dapat dilihat pada tabel berikut.

Tabel 4. Hasil Tokenisasi

Komentar	Normalisasi Data
Harus ny beli bbm di permudah	Harus nya beli bbm di permudah
Aplikasi yg menyenangkan	Aplikasi yang menyenangkan

3.3.4 *Stopwords Removal*

Proses *stopwords removal* merupakan sebuah proses menghapus kata umum yang sering muncul dalam data [18]. Proses ini menggunakan bantuan *library Natural Language Toolkit* (NLTK) untuk menghilangkan kata umum tersebut. Hasil *stopwords removal* dapat dilihat pada tabel berikut.

Tabel 5. Hasil Stopwords Removal

Tokenisasi	Stopword Removal
[mantap, bisa, buat, beli, bensin]	[mantap, beli, bensin]
[bikin, susah, rakyat, saja]	[susah, rakyat]
[salam, sukses, buat, pertamina, semoga, semakin, jaya]	[salam, sukses, pertamina, semoga, jaya]

3.3.5 *Stemming*

S*temming* merupakan proses untuk menghilangkan awalan dan akhiran dalam kata sehingga hasilnya berupa kata dasar saja [19]. Dalam penelitian ini digunakan *library* Sastrawi untuk proses *stemming*. Hasil proses *stemming* dapat dilihat pada tabel berikut.

Tabel 6. Hasil Stemming

Kata Imbuhan	Kata Dasar
Permudah	Mudah
Pembayaran	Bayar
Mengunduh	Unduh
Kebijakan	Bijak

Membantu	Bantu
Pembenahan	Benah
Mengurangi	Kurang

3.4 Pembobotan

Term Frequency - Inverse Document Frequency (TF-IDF) merupakan salah satu cara yang dapat dilakukan untuk proses ekstraksi fitur dalam penggunaan metode Naïve Bayes [20]. Proses pembobotan kata dilakukan dengan memanfaatkan *library Scikit-learn* yang memiliki algoritma *CountVectorizer CountVectorizer* mampu mengubah teks menjadi vektor sehingga mempermudah proses perhitungan Hasil representasi vektor menggunakan *CountVectorizer* mendapatkan hasil sebanyak 1.370 angka yang memiliki 1.945 kata. Contoh perhitungan TF-IDF dapat dilihat pada tabel berikut.

Tabel 7. Contoh Data

No	**Komentar**	**Sentimen**
D1	Aplikasi susah rakyat	Negatif
D2	Aplikasi Mypertamina bagus	Positif
D3	Ribet susah rakyat	Negatif

Perhitungan *Term Frequency* adalah sebagai berikut.

Tabel 8. Perhitungan Term Frequency

Kata	TF			DF
	D1	D2	D3	
Aplikasi	1	1	0	2
Susah	1	0	1	2
Rakyat	1	0	1	2
Ribet	0	0	1	1
MyPertamina	0	1	0	1
Bagus	0	1	0	1

Setelah melakukan perhitungan *Term Frequency*, selanjutnya dilakukan perhitungan terhadap *Invers Document Frequency*. Hasil *Invers Document Frequency* adalah sebagai berikut.

$$IDF_{Aplikasi} = \log\frac{3}{2} = 0.238$$

$$IDF_{Susah} = \log\frac{3}{2} = 0.238$$

$$IDF_{Rakyat} = \log\frac{3}{2} = 0.238$$

$$IDF_{Ribet} = \log\frac{3}{1} = 0.477$$

$$IDF_{Ribet} = \log\frac{3}{1} = 0.477$$

$$IDF_{Ribet} = \log\frac{3}{1} = 0.477$$

Untuk mendapatkan hasil dari proses *Term Frequency-Invers Document Frequency*, maka diperlukan proses perkalian antara TF dengan IDF. Hasil TF-IDF adalah sebagai berikut.

Tabel 9. Perhitungan TF-IDF

TF-IDF		
D1	D2	D3
0.238	0.238	0
0.238	0	0.238
0.238	0	0.238
0	0	0.477
0	0.477	0

0	0.477	0

3.5 Klasifikasi

Proses klasifikasi dengan Naïve Bayes menggunakan *library Scikit-learn* yang telah disediakan oleh Python. Pada proses klasifikasi digunakan data uji sebesar 20% dengan data latih 80% dan juga data uji 30% dengan data latih 70% dari data keseluruhan. Data yang digunakan dengan pembagian data uji 20% dan 30% adalah sebagai berikut.

Tabel 10. Pembagian Data Uji dan Data Latih

Rasio	Data Uji	Data Latih
20:80	274	1096
30:70	411	959

Proses klasifikasi memanfaatkan algoritma Multinomial Naïve Bayes. Hasil klasifikasi berupa visualisasi dengan *confusion matrix*. Hasil *confusion matrix* dengan data uji 20% dan data latih 80% adalah sebagai berikut.

Tabel 11. Confusion Matrix Data Uji 20%

		Actual Class	
		Positif	Negatif
Predict Class	Positif	108	25
	Negatif	30	111

Dari *confusion matrix* tersebut, dapat dihitung nilai presisi, *recall*, akurasi, dan f1-*score* dengan persamaan berikut.

$$\text{Akurasi} = \frac{108 + 111}{108 + 25 + 111 + 30} = 0.79$$

$$\text{Presisi} = \frac{108}{108 + 25} = 0.81$$

$$Recall = \frac{108}{108 + 30} = 0.78$$

$$F1 - Score = 2 \times \frac{0.81}{0.78} = 0.79$$

Sementara itu, untuk pembagian data uji 30% dan data latih 70% menghasilkan *confusion matrix* sebagai berikut.

Tabel 12. Confusion Matrix Data Uji 30%

		Actual Class	
		Positif	Negatif
Predict Class	Positif	164	35
	Negatif	43	169

Dari *confusion matrix* tersebut, nilai presisi, akurasi, *recall*, dan f1-*score* yang dihasilkan adalah sebagai berikut.

$$\text{Akurasi} = \frac{164 + 169}{164 + 35 + 169 + 43} = 0.81$$

$$\text{Presisi} = \frac{164}{164 + 35} = 0.82$$

$$Recall = \frac{164}{164 + 43} = 0.79$$

$$F1 - Score = 2 \times \frac{0.82}{0.79} = 0.80$$

3.6 Aplikasi NVIVO

Selain *confusion matrix*, dihasilkan pula *output* dengan menggunakan aplikasi NVIVO. Aplikasi NVIVO merupakan aplikasi yang digunakan untuk pengolahan data kualitatif. Hasil dengan menggunakan aplikasi NVIVO adalah sebagai berikut.

3.6.1 Wordcloud

Gambar 1 Wordcloud

Dari *wordcloud* tersebut, kata aplikasi merupakan kata paling besar diantara yang lainnya. Kata yang paling besar tersebut merupakan kata yang kemunculannya paling banyak dalam data.

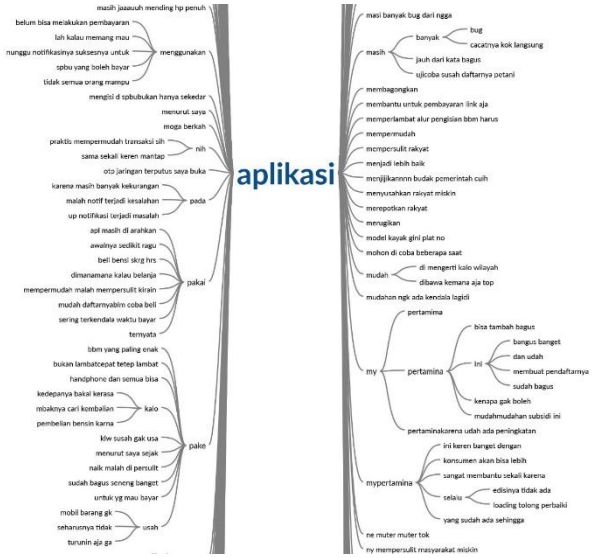
3.6.2 *Word Frequency*

Tabel 13 Word Frequency

Kata	Panjang Kata	Jumlah Kemunculan	Persentase
Aplikasi	8	277	1.82%
Pertamina	9	137	0.90%
Sangat	6	112	0.74%
Semoga	6	92	0.60%
Dengan	6	82	0.54%

Dari fitur *word frequency query* yang dimiliki oleh NVIVO, kata aplikasi memiliki frekuensi pencarian sejumlah 1.82%. Dilanjutkan dengan kata Pertamina dengan frekuensi sebesar 0.90%.

3.6.3 *Word Tree* Kata Aplikasi



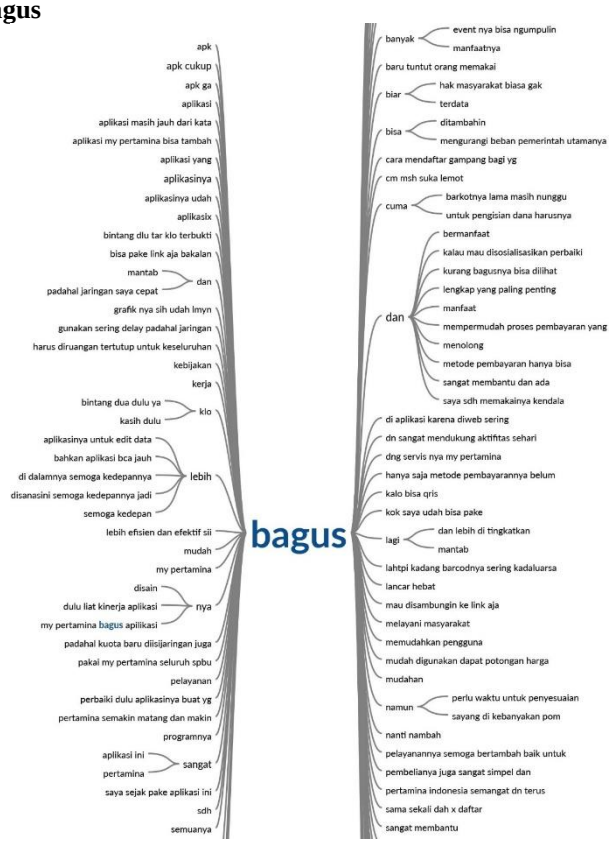
Gambar 2 Word Tree Kata Aplikasi

Dari *word tree* yang telah dihasilkan, kata aplikasi lebih banyak menerima sentimen negatif dari pengguna aplikasi MyPertamina. Para pengguna mengeluhkan aplikasi yang masih memiliki *bug* ketika digunakan. Selain itu, beberapa dari pengguna kesulitan untuk menerima *barcode* dan *One Time Password* (OTP) ketika proses pembelian Bahan Bakar Minyak (BBM) bersubsidi di SPBU dengan menggunakan aplikasi MyPertamina. Beberapa orang juga mengeluhkan bahwa tidak semua masyarakat memiliki

smartphone sehingga mereka tidak dapat mengakses aplikasi MyPertamina. Dengan kesulitan tersebut, pembelian BBM bersubsidi menjadi lebih sulit dari biasanya. Akibatnya, antrean menjadi semakin panjang ketika harus menggunakan aplikasi MyPertamina di SPBU. Pengguna aplikasi juga mempertanyakan keamanan dan keselamatan ketika menggunakan aplikasi di SPBU dikarenakan aturan dari stasiun pengisian bahan bakar yang mengharuskan pembeli untuk tidak menggunakan *handphone* ketika mengisi bensin. Risiko kebakaran akan mengintai ketika *handphone* aktif digunakan. Ini dikarenakan sinyal elektrik dari *smartphone* dapat menimbulkan percikan api dan berbahaya bagi proses pengisian BBM.

Dari permasalahan yang diutarakan oleh pengguna, sebaiknya pengelola aplikasi MyPertamina dapat dengan cepat dan tanggap menghadapi permasalahan tersebut. Pengelola sebaiknya memperbaiki *bug* yang terdapat dalam aplikasi sehingga pengguna semakin mudah dalam menggunakannya. Selain itu, perlu dipertimbangkan lagi penggunaan aplikasi MyPertamina ketika melakukan pembelian BBM di SPBU. Demi kenyamanan dan keselamatan bersama, sebaiknya pengelola terlebih dahulu melakukan penelitian mengenai keamanan penggunaan *handphone* di SPBU. Pemerintah yang telah membuat kebijakan mengenai pemakaian aplikasi MyPertamina juga harus mulai mempertimbangkan kebijakan terbaru yang dibuatnya. Hal ini dikarenakan oleh masyarakat Indonesia yang belum seluruhnya melek akan teknologi. Akibatnya, beberapa dari mereka tidak dapat membeli BBM bersubsidi karena tidak memiliki akses dalam menggunakan aplikasi tersebut.

3.6.4 *Word Tree* Kata Bagus



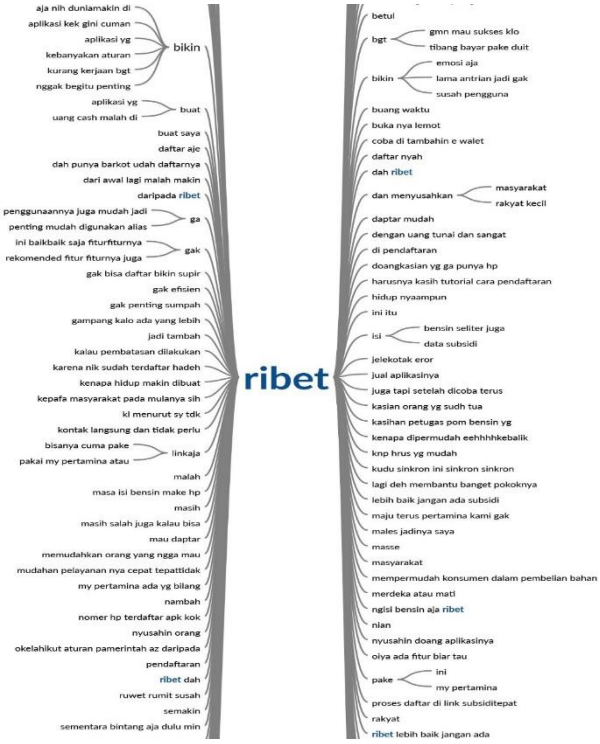
Gambar 3 Word Tree Kata Bagus

Dalam *word tree* kata bagus, opini masyarakat mengharapkan agar aplikasi MyPertamina dapat digunakan ketika *bug* telah diperbaiki oleh pihak pengelola. Selain itu, pengguna juga menginginkan agar aplikasi MyPertamina terhubung ke e-*wallet* lainnya selain LinkAja. Menurut pengguna, penggunaan aplikasi MyPertamina dapat membantu ketika pembelian BBM karena sistem aplikasi ini yang *cashless* sehingga pengguna tidak perlu mengeluarkan uang tunai. Dengan opini pengguna tersebut, mereka berharap bahwa aplikasi MyPertamina dapat meningkatkan layanannya sehingga penggunaan aplikasi ini dapat tepat sasaran sesuai dengan tujuan dibuatnya.

3.6.5 *Word Tree* Kata Ribet

Menurut hasil *word tree* yang telah dibentuk oleh NVIVO, kata ribet merujuk pada penggunaan aplikasi MyPertamina yang menurut pengguna sangat menyusahkan masyarakat ketika pembelian BBM bersubsidi. Pengguna mengeluhkan data mereka yang telah terdaftar begitu saja dalam aplikasi ini. Padahal beberapa dari mereka masih kesulitan untuk masuk ke dalam sistem. Selain itu, pengguna juga mengeluhkan

bahwa aplikasi MyPertamina harus terus menyinkronkan berbagai macam data ketika proses pendaftaran. Hal ini tentu tidak efisien dan membuang-buang waktu bagi beberapa masyarakat yang memiliki banyak kesibukan.



Gambar 4 Word Tree Kata Ribet.

Dari masalah yang banyak dikeluhkan oleh pengguna, pihak pengelola aplikasi MyPertamina sebaiknya mempermudah proses yang harus dilewati oleh pengguna. Penggunaan *user interface* yang simpel sangat dibutuhkan untuk menjawab permasalahan ini. Selain itu, keamanan server perlu diperkuat sehingga data masyarakat yang telah terdaftar dalam aplikasi ini dapat terjaga dengan baik.

4. Conclusion

Berdasarkan hasil pembahasan dari penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naïve Bayes berhasil digunakan untuk proses analisis sentimen. Performa yang dihasilkan dengan data uji 30% dan data latih 70% menghasilkan akurasi sebesar 81%, presisi 82%, *recall* 79%, dan f1-*score* 80%. Sementara itu, untuk pembagian data uji 20% dan data latih 80% didapatkan hasil akurasi sebesar 79%, presisi 81%, *recall* 78%, dan f1-*score* 79%. Metode Naïve Bayes dalam klasifikasi ini menghasilkan performa yang cukup baik sehingga cocok digunakan dalam proses analisis. Selanjutnya dibutuhkan data yang lebih besar agar proses klasifikasi menghasilkan performa yang lebih akurat.

5. Acknowledgment

Terima kasih untuk program studi Informatika UMBY.

References

[1] J. Napitupulu, Darmawan; Lubis, Muhammad Ridwan; Revida, Erika; Putra, Surya Hendra; Saputra, Syifa; Jamaludin; Negara, Edi Surya; Simarmata, *E-Government: Implementasi, Strategi dan Inovasi*, vol. 21, no. 1. Medan: Yayasan Kita Menulis, 2020.

[2] C. A. Cholik, "Perkembangan Teknologi Informasi Komunikasi / Ict dalam Berbagai Bidang," vol. 2, no. 2, pp. 39–46, 2021.

[3] E. Purnomohadi, *Hiswana Migas : Mengalirkan Energi Membangun Negeri*. Bogot: Bypass, 2019.

[4] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.

[5] A. Kulkarni and A. Shivananda, *Natural Language Processing Recipes*. 2019. doi: 10.1007/978-1-4842-4267-4.

[6] R. Sebastian and M. Vahid, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rd Edition*. Packt Publishing, 2019.

[7] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Sleman: Deepublish, 2020.
[8] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
[9] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
[10] S. Joko, *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo, 2019.
[11] A. Géron and R. Russell, "Machine learning step-by-step guide to implement machine learning algorithms with Python," *O'Reilly Media, Inc*, p. 106, 2019.
[12] T. K. Ronald, *Practical Deep Learning A Python-Based Introduction*. No Starch Press, 2021.
[13] N. P. G. Naraswati, R. Nooraeni, D. C. Rosmilda, D. Desinta, F. Khairi, and R. Damaiyanti, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sistemasi*, vol. 10, no. 1, p. 222, 2021, doi: 10.32520/stmsi.v10i1.1179.
[14] F. Sidik, I. Suhada, A. H. Anwar, and F. N. Hasan, "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, 2022, doi: 10.26418/jlk.v5i1.79.
[15] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019, doi: 10.26418/jp.v5i3.34368.
[16] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
[17] M. R. Fais Sya' bani, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.
[18] L. O. Sihombing, H. Hannie, and B. A. Dermawan, "Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algortima Naïve Bayes Classifier," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
[19] M. Hudha, E. Supriyati, and T. Listyorini, "Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 1, pp. 1–6, 2022, doi: 10.33387/jiko.v5i1.3376.
[20] D. F. Zhafira, B. Rahayudi, and I. Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55–63, 2021, doi: 10.25126/justsi.v2i1.24.