

## ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP ULASAN APLIKASI MYPERTAMINA MENGGUNAKAN SUPPORT VECTOR MACHINE

**Ikram Maulana<sup>1\*</sup>, Winda Apriandari<sup>2</sup>, Agung Pambudi<sup>3</sup>**

<sup>1,2,3</sup>Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi, Sukabumi, Indonesia

Email: <sup>1</sup>ikram075@ummi.ac.id, <sup>2</sup>winda.apriandari@ummi.ac.id, <sup>3</sup>agungpambd@ummi.ac.id

(\* : corresponding author)

**Abstrak**-PT. Pertamina (Persero), sebagai BUMN terbesar di bidang minyak dan gas bumi di Indonesia, memiliki tanggung jawab untuk menyalurkan BBM bersubsidi secara tepat sasaran dan sesuai kuota yang ditetapkan oleh pemerintah. Sejak 1 Juli 2022, aplikasi MyPertamina menjadi syarat untuk pembelian BBM Peralite dan Biosolar. Dengan lebih dari 10 juta unduhan dan peringkat 2,5 di Google Play Store berdasarkan data pada Oktober 2022, penelitian ini bertujuan untuk mengevaluasi aplikasi MyPertamina dengan mengelompokkan ulasan ke dalam dua kelas sentimen dan tiga kelas aspek. Metode yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Database* (KDD) dengan menerapkan algoritma *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa MyPertamina dinilai membantu pengguna dalam pembelian BBM, meskipun terdapat kendala yang dirasakan pengguna. Kendala tersebut meliputi kesulitan dalam mendaftar akun dan sering mengalami kegagalan *login* pada aspek *Bug*, kerumitan dalam penggunaan pada aspek kegunaan, serta kadang-kadang tidak muncul *barcode* pada aspek pembayaran. Evaluasi model klasifikasi sentimen dan aspek menghasilkan tingkat akurasi rata-rata sebesar 92% dan 96% secara berturut-turut. Dengan demikian, dapat disimpulkan bahwa model yang telah dikembangkan cukup andal dalam melakukan klasifikasi pada data ulasan aplikasi MyPertamina.

**Kata Kunci:** Pertamina, BBM Subsidi, Aplikasi MyPertamina, Analisis Sentimen, *Support Vector Machine* (SVM)

**Abstract**-PT. Pertamina (Persero), as the largest state-owned enterprise in the oil and gas sector in Indonesia, has the responsibility to distribute sub-sidized fuel (BBM) accurately and according to the designated quota set by the government. Since July 1, 2022, the MyPertamina application has become a requirement for purchasing peralite and biosolar fuels. With over 10 million downloads and a rating of 2.5 on the Google Play Store based on data from October 2022, this research aims to evaluate the MyPertamina application by categorizing reviews into two sentiment classes and three aspect classes. The research methodology employed in this study is *Knowledge Discovery in Databases* (KDD), utilizing the *Support Vector Machine* (SVM) algorithm. The findings indicate that MyPertamina is perceived to assist users in purchasing BBM, although users encounter certain challenges. These challenges include difficulties in creating an account and frequent login failures in the Bug aspect, complexity in usability in the usability aspect, and the occasional absence of barcodes in the payment aspect. The evaluation of the sentiment and aspect classification models resulted in an average accuracy rate of 92% and 96%, respectively. Therefore, it can be concluded that the developed models are sufficiently reliable in classifying the reviews of the MyPertamina application.

**Keywords:** Pertamina, Subsidized BBM, MyPertamina Application, Sentiment Analysis, *Support Vector Machine* (SVM)

### 1. PENDAHULUAN

Minyak bumi memainkan peran penting dalam kehidupan manusia sebagai salah satu komoditas pertambangan. Minyak bumi berfungsi sebagai bahan baku dalam produksi Bahan Bakar Minyak (BBM) dan menjadi sumber energi yang sangat dibutuhkan untuk mendukung aktivitas sosial di seluruh dunia. Selain itu, BBM juga memiliki peran yang krusial dalam menentukan harga di berbagai sektor dan dapat memicu terjadinya inflasi [1]. Di Indonesia, BBM menjadi bahan bakar yang paling dominan digunakan untuk kendaraan seperti mobil dan sepeda motor, dikarenakan tingginya jumlah kendaraan yang beroperasi. Hal ini mengakibatkan permintaan BBM di Indonesia tergolong sebagai yang terbesar di Asia Tenggara dan menempati peringkat kelima di Asia-Pasifik [2].

Dalam industri BBM di Indonesia, saat ini telah banyak terdapat perusahaan retail yang beroperasi, di antaranya Pertamina, Vivo, Shell, dan perusahaan lainnya. Sebagai Badan Usaha Milik Negara (BUMN) terbesar di sektor minyak dan gas bumi di Indonesia, Pertamina diberi mandat oleh pemerintah untuk menyalurkan BBM bersubsidi agar tepat sasaran dan sesuai dengan kuota yang ditetapkan [3]. BBM subsidi merupakan jenis bahan bakar minyak yang mendapatkan subsidi dari pemerintah untuk membantu masyarakat dengan penghasilan rendah. Namun sayangnya, program subsidi BBM saat ini dianggap masih belum tepat sasaran dan sebagian besar dinikmati oleh kalangan mampu [4].

Pada tanggal 1 Juli 2022, PT. Pertamina (Persero) telah menerapkan mekanisme baru untuk memastikan penyaluran BBM bersubsidi dapat terlaksana sesuai dengan target dan kuota yang telah ditentukan. Melalui

mekanisme baru tersebut, pengguna bensin Ptalite dan Biosolar subsidi diwajibkan untuk mendaftarkan kendaraannya pada aplikasi MyPertamina. MyPertamina adalah aplikasi layanan keuangan digital yang dimiliki Pertamina dan anggota BUMN, yang terintegrasi dengan aplikasi LinkAja [5]. Tujuan dari hadirnya aplikasi MyPertamina adalah untuk membantu masyarakat dalam menjalankan transaksi yang lebih efektif dan efisien, juga memberikan keuntungan bagi pelanggan seperti poin yang dapat ditukarkan dengan berbagai hadiah melalui aplikasi [6]. Uji coba aplikasi MyPertamina pertama kali dilakukan di beberapa wilayah dan kabupaten di Indonesia, seperti Kabupaten Agam, Kota Bandung, Kota Banjarmasin, Kota Bukittinggi, Kabupaten Ciamis, Kota Manado, Kota Padang Panjang, Kota Sukabumi, Kabupaten Tanah Datar, Kota Tasikmalaya, dan Kota Yogyakarta [7].

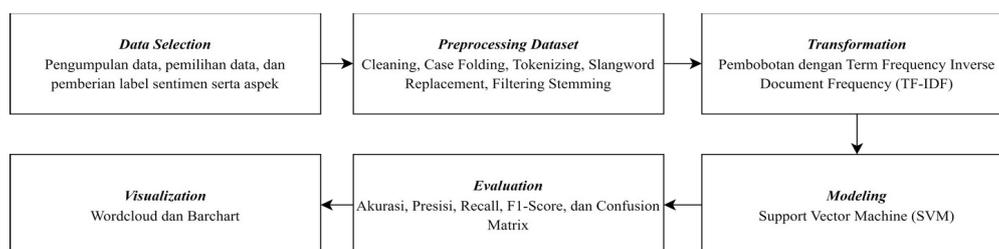
Aplikasi MyPertamina dapat diakses secara gratis melalui situs Google Play Store dan dapat diunduh oleh pengguna menggunakan *smartphone*. Berdasarkan data terakhir dari Google Play Store pada 8 Oktober 2022, MyPertamina memiliki peringkat 2,5 serta lebih dari 10 juta kali unduhan. Namun, semakin rendah peringkat sebuah aplikasi di Google Play Store menandakan bahwa kinerja aplikasi tersebut masih kurang memuaskan bagi para pengguna [8]. Mengingat banyaknya jumlah ulasan pengguna yang tidak terstruktur dan tidak terkategori, diperlukan suatu teknik untuk memperoleh data evaluasi, yaitu dengan melakukan analisis sentimen terhadap kumpulan data ulasan pengguna.

Analisis sentimen merupakan suatu proses komputasi yang dilakukan untuk menemukan informasi berharga dari kalimat pendapat atau opini yang tertulis dalam bentuk teks secara otomatis [9]. Dampak dan manfaat yang signifikan dari analisis sentimen telah mengakibatkan pesatnya perkembangan penelitian berbasis analisis sentimen, bahkan banyak perusahaan saat ini menggunakan analisis sentimen sebagai sarana untuk memperoleh informasi dari kumpulan data pendapat [10].

Penelitian ini merujuk pada beberapa referensi jurnal terkait analisis sentimen yang memiliki kesamaan topik dengan penelitian ini. Dalam penelitian yang dilakukan oleh [11], analisis sentimen terhadap data ulasan aplikasi Grab dilakukan dengan menggunakan algoritma *Support Vector Machine* (SVM). Evaluasi dilakukan pada lebih dari 1000 data ulasan pengguna yang diperoleh dari Google Play Store, dan hasil penelitian menunjukkan tingkat akurasi sebesar 85,54%. Kata "ovo" teridentifikasi sebagai kata yang sering muncul pada ulasan positif, sementara kata "driver" sering muncul dalam ulasan negatif. Penelitian lain yang melibatkan analisis sentimen pada ulasan berbahasa Indonesia di Google Play Store, yang dilakukan oleh [12], membandingkan kinerja metode SVM dan *Naïve Bayes*. Hasil penelitian tersebut menunjukkan bahwa metode SVM memiliki tingkat akurasi tertinggi, yaitu 81,46%, dibandingkan dengan metode *Naïve Bayes* yang mencapai 75,41%. Penelitian lainnya, yang melibatkan analisis sentimen terhadap data ulasan aplikasi Zoom *Cloud Meetings*, dilakukan oleh [9]. Metode SVM dalam penelitian ini mencapai tingkat akurasi 81,22%, sementara metode *Naïve Bayes* hanya mencapai 74,37%. Penelitian menggunakan metode *K-Fold Cross Validation* dengan sepuluh kali iterasi dan menggunakan label sentimen positif dan negatif.

Penelitian [11] menjadi dasar untuk melakukan pelabelan sentimen positif dan negatif pada data ulasan. Namun, penelitian tersebut tidak memilah data ulasan ke dalam kategori aspek yang spesifik untuk sentimen negatif atau positif, serta tidak melakukan penggantian kata slang untuk meningkatkan keakuratan hasil. Sehubungan dengan itu, penelitian ini akan difokuskan pada analisis sentimen berbasis aspek terhadap ulasan pengguna aplikasi MyPertamina di Google Play Store menggunakan SVM. Algoritma SVM digunakan karena dianggap sebagai algoritma yang paling akurat dalam studi kasus klasifikasi teks karena kemampuannya dalam memisahkan data non-linear berdimensi tinggi [13]. Penelitian ini membagi kelas sentimen menjadi dua kelompok, yaitu sentimen positif dan negatif. Penelitian ini memiliki batasan hanya pada *dataset* terkait ulasan pengguna aplikasi MyPertamina pada periode Agustus 2022 hingga Maret 2023. Jumlah total data ulasan yang digunakan dalam penelitian ini adalah 19562. Adapun aspek yang akan digunakan dalam penelitian ini terbatas pada 3 aspek, yaitu *bug*, kegunaan, dan pembayaran.

## 2. METODE PENELITIAN



Gambar 1. Tahapan Metode Penelitian

Dalam penelitian ini digunakan metode *Knowledge Discovery in Database* (KDD) dengan menerapkan algoritma SVM. KDD merupakan suatu proses yang melibatkan penggunaan algoritma untuk mengekstraksi informasi penting dari kumpulan data yang ada, sehingga pengetahuan yang sebelumnya tidak diketahui dapat ditemukan [14]. Penelitian ini terdiri dari beberapa tahap, yaitu data *selection*, *preprocessing dataset*, *transformation*, *modeling*, *evaluation*, dan *visualization*. Detail tahapan penelitian dapat dilihat pada Gambar 1.

### 2.1. Data Selection

Dalam penelitian ini, langkah awal yang dilakukan adalah melakukan seleksi pada data. Peneliti mengumpulkan data ulasan pengguna aplikasi MyPertamina dari situs Google Play Store pada periode bulan Agustus 2022 hingga Maret 2023, dengan jumlah data ulasan sebanyak 19562. Data ulasan tersebut diperoleh melalui teknik *scraping* dengan menggunakan bahasa pemrograman Python dan *library* Google Play Scraper, kemudian disimpan dalam format berkas *parquet* untuk dilakukan pelabelan sentimen dan aspek pada tahap selanjutnya.

Setelah data berhasil diperoleh, langkah berikutnya adalah melakukan pelabelan sentimen pada data ulasan. Pelabelan sentimen dilakukan berdasarkan *rating* yang terdapat pada setiap data ulasan. Untuk ulasan dengan *rating* 1 dan 2, diberikan label sentimen negatif, sementara ulasan dengan *rating* 4 dan 5 diberikan label sentimen positif. Kemudian, dilanjutkan dengan pelabelan aspek pada data ulasan. Penelitian ini menggunakan tiga aspek yaitu *Bug*, kegunaan, dan pembayaran yang ditentukan setelah peneliti membaca data ulasan dan menemukan aspek-aspek yang sering dibahas. Untuk melakukan pelabelan aspek pada data ulasan, digunakan metode *Cosine Similarity* dengan menggunakan kata kunci yang ditentukan oleh peneliti. Detail Kata kunci yang digunakan dalam penelitian dapat dilihat pada Tabel 1.

**Tabel 1.** Daftar Kata Kunci dari Setiap Aspek

No	Aspek	Kata Kunci
1	Bug	Bug, Masalah, Berhenti, Error, Login, Lag, Lemot, Registrasi, Password, Salah, Dibenahi, Daftar, Jaringan, Gangguan, Lelet, Download, Gagal, Program, Verifikasi, Verifikator, Tersendat, OTP, Delay, Stuck, Ngestuck, Stuk, Ngestuk, Kekurangan, Lambat
2	Kegunaan	Kegunaan, Fitur, Manfaat, Fungsionalitas, Pengguna, Nyaman, Fungsi, Berguna, Pelayanan, Kemampuan, Keuntungan, Kelebihan, Paksa, Terpaksa, Susah, Pusing, Ribet, Rumit, Disuruh, Maksimal, Guna, Gunakan, Masyarakat, Solusi
3	Pembayaran	Pembayaran, Sistem, Kesulitan, Sulit, Cashless, Bayar, Menyusahkan, Topup, Saldo, Uang, Linkaja, Gopay, Ovo, Barcode

### 2.2. Preprocessing Dataset

Langkah selanjutnya setelah tahap data *selection* dalam penelitian ini akan dilakukan *text preprocessing* pada data ulasan. *Text preprocessing* merupakan tahap awal dalam pengolahan teks sebelum dianalisis atau diproses lebih lanjut, dan memiliki peran penting dalam berbagai aplikasi seperti *text mining*, analisis sentimen, dan ekstraksi informasi dari dokumen teks yang tidak terstruktur, untuk menghindari kesalahan dalam ekstraksi fitur atau atribut yang dapat secara signifikan mengurangi kinerja analisis sentimen [15]. Dalam penelitian ini, akan dilakukan enam tahap *text preprocessing* di antaranya *cleaning*, *case folding*, *tokenizing*, *slang word replacement*, *filtering*, dan *stemming*.

### 2.3. Transformation

Pada umumnya tahap *transformation* merupakan tahap yang dilakukan setelah tahap *text preprocessing* selesai. Tahap transformasi dilakukan untuk mengubah data ulasan menjadi representasi numerik agar lebih mudah diproses. Metode yang sering digunakan adalah pembobotan kata menggunakan metode *Term Frequency Inverse Document Frequency* (TF-IDF). Metode ini merupakan metode yang populer karena dapat menentukan hubungan antara kata dengan dokumen melalui pemberian bobot pada setiap kata [16]. Dalam penelitian ini, metode TF-IDF dapat membantu peneliti dalam menganalisis dokumen teks dengan lebih efektif dan efisien.

#### 2.3.1. Term Frequency (TF)

Dalam menghitung nilai TF-IDF, nilai dari TF didapatkan dengan cara menghitung jumlah kemunculan fitur atau kata *t* pada sebuah dokumen (1).

$$TF_t = (t, d) \tag{1}$$

Keterangan: d: Dokumen ke-d  
t: Kata ke-t dari kata kunci

### 2.3.2. Inverse Document Frequency (IDF)

Dalam perhitungan TF-IDF, nilai IDF didapatkan melalui perhitungan logaritma (2) dari jumlah total dokumen n dibagi jumlah dokumen df yang mengandung fitur atau kata t.

$$IDF_t = \log \frac{n}{df(t)} \quad (2)$$

Keterangan: n: Total dokumen

df(t): Banyak dokumen yang mengandung kata t

### 2.3.3. Term Frequency Inverse Document Frequency (TF-IDF)

Dalam perhitungan TF-IDF, nilai akhir diperoleh dari hasil kali antara nilai TF dan IDF (3)

$$W_t = TF_t \times IDF_t \quad (3)$$

Keterangan: TF : Banyaknya kata yang dicari pada sebuah dokumen

IDF : *Inverse Document Frequency*

## 2.4. Modeling

Dalam penelitian ini, digunakan algoritma SVM untuk melakukan klasifikasi data ulasan. Algoritma ini dipilih karena telah terbukti efektif untuk memecahkan masalah klasifikasi teks dan telah banyak digunakan dalam penelitian lain. SVM adalah algoritma *supervised learning* yang bekerja dengan mencari *hyperplane* optimal yang dapat memisahkan dua buah kelas data. *Hyperplane* optimal ini ditentukan berdasarkan jarak margin terbesar antara *hyperplane* dan *support vector* [17]. Data ulasan akan dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji, sebelum dilakukan proses pemodelan. Dalam penelitian ini, digunakan *kernel* linear sebagai *kernel* SVM. Persamaan perhitungan *hyperplane* klasifikasi SVM linear dapat dilihat pada (4).

$$f(x) = w \cdot x + b \quad (4)$$

Keterangan: w : Parameter bobot

x : Vektor input

b : bias

## 2.5. Evaluation

Langkah berikutnya setelah tahap modeling pada penelitian ini adalah tahap evaluasi, yang bertujuan untuk mengevaluasi hasil dari model yang telah dibangun pada tahap sebelumnya. Model akan diuji pada data testing yang telah dipersiapkan sebelumnya. Tahap evaluasi sangat penting dilakukan untuk mengetahui seberapa baik model dapat menangkap pola yang ada dalam data ulasan. Evaluasi model akan menggunakan beberapa metrik seperti akurasi, presisi, *recall*, dan *F1-Score* yang merupakan hasil perhitungan dari *confusion matrix*. *Confusion matrix* merupakan metode ukur klasik dalam algoritma *supervised learning*, di mana hasilnya digunakan untuk menghitung nilai akurasi, presisi, *recall*, dan *F1-Score*.

## 2.6. Visualization

Tahap terakhir dalam metode KDD adalah visualisasi, yang bertujuan untuk menampilkan model secara visual agar lebih mudah untuk dipahami. Pada tahap ini, data ulasan akan diolah menjadi frekuensi kata atau kata yang paling sering muncul berdasarkan kategori aspek, kemudian divisualisasikan dalam bentuk diagram atau grafik untuk mempermudah analisis data ulasan. Beberapa jenis visualisasi yang sering digunakan meliputi *bar chart*, *fishbone* diagram, atau *wordcloud*. Tahap ini sangat penting karena dapat membantu pihak terkait dalam mengambil keputusan berdasarkan evaluasi model yang telah dibangun.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Data Selection

Dalam penelitian ini, data ulasan pengguna aplikasi MyPertamina diperoleh melalui teknik *scraping* dengan bantuan *library* Google Play Scraper. Data yang berhasil diperoleh terdiri dari 19562 baris dan tiga kolom yang terdiri dari tanggal, ulasan, dan *rating*. Kemudian, dilakukan pelabelan pada data ulasan untuk membedakan sentimen positif dan negatif. Sentimen berlabel positif diberikan pada ulasan dengan *rating* 4 dan 5, sementara sentimen berlabel negatif diberikan pada ulasan dengan *rating* 1 dan 2. Dengan demikian, terdapat 5783 ulasan bersentimen positif dan 13779 ulasan bersentimen negatif. Selanjutnya, setiap data ulasan diberikan

label aspek dengan kategori *Bug*, kegunaan, dan pembayaran. Proses pelabelan dilakukan menggunakan metode *Cosine Similarity* untuk menemukan kemiripan antara ulasan dengan kata kunci pada aspek. Hasilnya, didapatkan 3807 ulasan dengan aspek *Bug*, 3379 ulasan beraspek kegunaan, dan 12376 ulasan beraspek pembayaran.

### 3.2. Preprocessing Dataset

Pada tahap *preprocessing dataset*, data yang telah diperoleh melalui teknik *scraping* dan disimpan dalam format *parquet* diolah kembali agar dapat diproses dengan lebih mudah menggunakan algoritma pemrosesan bahasa natural. Pada tahap ini, terdapat penggunaan *library* pada bahasa pemrograman *python3* seperti *Sastrawi*, *NLTK*, atau *library* lainnya. Proses *preprocessing* pada penelitian ini mencakup *cleaning*, *case folding*, *tokenizing*, penggantian slang word, *filtering*, dan *stemming*.

#### 3.2.1. Cleaning

Tahap ini merupakan tahap awal dalam proses *preprocessing dataset* yang bertujuan untuk membersihkan data ulasan dari karakter-karakter yang tidak relevan atau disebut dengan *noise*. Pada tahap ini, dilakukan penghapusan tanda baca dengan menggunakan *Regular Expression* (Regex). Hasil dari proses *cleaning* data ulasan yang telah dilakukan dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil Proses *Cleaning*

Dokumen	Input Process	Output Process
D1	Aplikasi kurang membantu, sudah daftar masuk kembali tidak bisa, no hp atau password salah tetus	Aplikasi kurang membantu sudah daftar masuk kembali tidak bisa no hp atau password salah tetus
D2	Sangat memuaskan, dan bisa di buat transaksi apa aja, jozzzzz	Sangat memuaskan dan bisa di buat transaksi apa aja jozzzzz
D3	Untuk apa aplikasi ini dibuat. Gak membantu sama sekali. Gak bergunaaaaaaaaaaaaaaaaaaaaa	Untuk apa aplikasi ini dibuat Gak membantu sama sekali Gak bergunaaaaaaaaaaaaaaaaaaaaa

#### 3.2.2. Case Folding

Tahap ini merupakan langkah kedua dalam proses *preprocessing dataset* yang berfungsi untuk mengubah semua huruf pada setiap dokumen menjadi huruf kecil atau huruf besar. Tujuannya adalah untuk mempermudah proses selanjutnya karena huruf kapital dan huruf kecil dapat dianggap sama dalam analisis. Hasil dari proses *case folding* terhadap data ulasan dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil Proses *Case Folding*

Dokumen	Input Process	Output Process
D1	Aplikasi kurang membantu sudah daftar masuk kembali tidak bisa no hp atau password salah tetus	aplikasi kurang membantu sudah daftar masuk kembali tidak bisa no hp atau password salah tetus
D2	Sangat memuaskan dan bisa di buat transaksi apa aja jozzzzz	sangat memuaskan dan bisa di buat transaksi apa aja jozzzzz
D3	Untuk apa aplikasi ini dibuat Gak membantu sama sekali Gak bergunaaaaaaaaaaaaaaaaaaaaa	untuk apa aplikasi ini dibuat gak membantu sama sekali gak bergunaaaaaaaaaaaaaaaaaaaaa

#### 3.2.3. Tokenizing

**Tabel 4.** Hasil Proses *Tokenizing*

Dokumen	Input Process	Output Process
D1	aplikasi kurang membantu sudah daftar masuk kembali tidak bisa no hp atau password salah tetus	['aplikasi', 'kurang', 'membantu', 'sudah', 'daftar', 'masuk', 'kembali', 'tidak', 'bisa', 'no', 'hp', 'atau', 'password', 'salah', 'tetus']
D2	sangat memuaskan dan bisa di buat transaksi apa aja jozzzzz	['sangat', 'memuaskan', 'dan', 'bisa', 'di', 'buat', 'transaksi', 'apa', 'aja', 'jozzzzz']
D3	untuk apa aplikasi ini dibuat gak membantu sama sekali gak bergunaaaaaaaaaaaaaaaaaaaaa	['untuk', 'apa', 'aplikasi', 'ini', 'dibuat', 'gak', 'membantu', 'sama', 'sekali', 'gak', 'bergunaaaaaaaaaaaaaaaaaaaaa']

Tahap *tokenizing* memiliki tujuan untuk memecah teks pada setiap dokumen menjadi kata-kata atau *token* yang lebih kecil. Proses ini sering kali dilakukan menggunakan bantuan *library* NLTK. Hasil dari tahap *tokenizing* pada data ulasan dapat dilihat pada Tabel 4.

### 3.2.4. Slang Word Replacement

Tahap ini merupakan tahap keempat dalam proses *preprocessing dataset* dilakukan untuk mengganti kata-kata slang dengan bentuk kata yang baku atau umum. Proses penggantian kata-kata slang dilakukan dengan menggunakan kamus kata-kata slang yang sudah disediakan. Hasil dari proses penggantian kata-kata slang pada data ulasan dapat dilihat pada Tabel 5.

**Tabel 5.** Hasil Proses Slang Word Replacement

Dokumen	Input Process	Output Process
D1	['aplikasi', 'kurang', 'membantu', 'sudah', 'daftar', 'masuk', 'kembali', 'tidak', 'bisa', 'no', 'hp', 'atau', 'password', 'salah', 'tetus']	['aplikasi', 'kurang', 'membantu', 'sudah', 'daftar', 'masuk', 'kembali', 'tidak', 'bisa', 'nomor', 'handphone', 'atau', 'password', 'salah', 'terus']
D2	['sangat', 'memuaskan', 'dan', 'bisa', 'di', 'buat', 'transaksi', 'apa', 'aja', 'jozzzzz']	['sangat', 'memuaskan', 'dan', 'bisa', 'di', 'buat', 'transaksi', 'apa', 'saja', 'jozzzzz']
D3	['untuk', 'apa', 'aplikasi', 'ini', 'dibuat', 'gak', 'membantu', 'sama', 'sekali', 'gak', 'bergunaaaaaaaaaaaaaaaaaaaaa']	['untuk', 'apa', 'aplikasi', 'ini', 'dibuat', 'tidak', 'membantu', 'sama', 'sekali', 'tidak', 'berguna']

### 3.2.5. Filtering

Tahap *filtering* pada proses *preprocessing dataset* berfungsi untuk mengeliminasi kata-kata yang tidak relevan dari suatu dokumen teks, seperti kata hubung atau yang lebih dikenal dengan istilah *stopword*. Hasil dari tahap *filtering* pada data ulasan dapat dilihat pada Tabel 6.

**Tabel 6.** Hasil Proses Filtering

Dokumen	Input Process	Output Process
D1	['aplikasi', 'kurang', 'membantu', 'sudah', 'daftar', 'masuk', 'kembali', 'tidak', 'bisa', 'nomor', 'handphone', 'atau', 'password', 'salah', 'terus']	['aplikasi', 'membantu', 'daftar', 'masuk', 'tidak', 'nomor', 'handphone', 'password', 'salah']
D2	['sangat', 'memuaskan', 'dan', 'bisa', 'di', 'buat', 'transaksi', 'apa', 'saja', 'jozzzzz']	['memuaskan', 'transaksi']
D3	['untuk', 'apa', 'aplikasi', 'ini', 'dibuat', 'tidak', 'membantu', 'sama', 'sekali', 'tidak', 'berguna']	['aplikasi', 'tidak', 'membantu', 'tidak', 'berguna']

### 3.2.6. Stemming

Tahap terakhir dari *preprocessing dataset* berguna untuk menemukan kata dasar dengan cara menghilangkan imbuhan pada setiap kata yang ada dalam dokumen. Tahap ini dikenal dengan nama *stemming*. Hasil dari proses *stemming* pada data ulasan dapat dilihat pada Tabel 7.

**Tabel 7.** Hasil Proses Stemming

Dokumen	Input Process	Output Process
D1	['aplikasi', 'membantu', 'daftar', 'masuk', 'tidak', 'nomor', 'handphone', 'password', 'salah']	['aplikasi', 'bantu', 'daftar', 'masuk', 'tidak', 'nomor', 'handphone', 'password', 'salah']
D2	['memuaskan', 'transaksi']	['muas', 'transaksi']
D3	['aplikasi', 'tidak', 'membantu', 'tidak', 'berguna']	['aplikasi', 'tidak', 'bantu', 'tidak', 'guna']

## 3.3. Transformation

Pada tahap ini, dilakukan proses penghitungan bobot kata dalam setiap dokumen menggunakan metode TF-IDF. Nilai TF akan menunjukkan seberapa sering kata tertentu muncul dalam suatu dokumen, sementara nilai IDF akan menunjukkan seberapa sering kata tersebut muncul dalam seluruh dokumen. Kedua nilai tersebut kemudian akan dikalikan untuk menghasilkan bobot kata pada setiap ulasan. Proses pembobotan kata menggunakan metode TF-IDF dihitung dengan menggunakan persamaan (1), (2), dan (3). Hasil perhitungan pembobotan kata dapat dilihat pada Tabel 8.

### 3.4. Evaluation

Setelah tahap modeling selesai, langkah selanjutnya adalah melakukan evaluasi untuk mengukur efektivitas model dalam memprediksi hasil yang diinginkan. Evaluasi dilakukan dengan menampilkan *confusion matrix*, kemudian menghitung nilai akurasi, presisi, *recall*, dan *F1-Score*. Hasil evaluasi untuk klasifikasi sentimen akan ditampilkan pada Tabel 9.

**Tabel 8.** Hasil Perhitungan Pembobotan Kata

Term	TF1	TF2	TF3	DF	IDF	TF-IDF1	TF-IDF2	TF-IDF3
Aplikasi	1	0	1	2	0.176	0.176	0	0.176
Bantu	1	0	1	2	0.176	0.176	0	0.176
Daftar	1	0	0	1	0.477	0.477	0	0
Masuk	1	0	0	1	0.477	0.477	0	0
Tidak	1	0	2	2	0.176	0.176	0	0.352
Nomor	1	0	0	1	0.477	0.477	0	0
Handphone	1	0	0	1	0.477	0.477	0	0
Password	1	0	0	1	0.477	0.477	0	0
Salah	1	0	0	1	0.477	0.477	0	0
Muas	0	1	0	1	0.477	0	0.477	0
Transaksi	0	1	0	1	0.477	0	0.477	0
Guna	0	0	1	1	0.477	0	0	0.477

**Tabel 9.** Confusion Matrix Model Sentimen

	Positif	Negatif
Positif	2623	265
Negatif	82	775

Selanjutnya, diperoleh nilai performa model klasifikasi untuk masing-masing kelasnya. Detail pengukuran performa model klasifikasi sentimen dapat dilihat pada Tabel 10.

**Tabel 10.** Nilai Performa Model Sentimen Setiap Kelas

Kelas	Akurasi	Presisi	Recall	F1-Score
Positif	0.91	0.97	0.91	0.94
Negatif	0.91	0.75	0.90	0.82

Berdasarkan Tabel 11, dapat disimpulkan bahwa model yang telah dibangun mempunyai nilai performa yang baik, dengan akurasi mencapai 91% untuk kelas positif maupun negatif. Selanjutnya, dilakukan pengujian menggunakan metode *K-Fold Cross Validation* dengan 5 iterasi, yang dilakukan dengan bantuan *library Scikitlearn Kfold*. Dari hasil pengujian menggunakan metode *K-Fold Cross Validation* dapat disimpulkan bahwa model memiliki performa yang baik dengan nilai akurasi sebesar 92% untuk kelas positif dan negatif. Selain evaluasi untuk model klasifikasi sentimen, dilakukan juga evaluasi untuk model klasifikasi aspek. Hasil evaluasi untuk model klasifikasi aspek akan ditampilkan pada Tabel 11.

**Tabel 11.** Confusion Matrix Model Aspek

	Bug	Kegunaan	Pembayaran
Bug	729	11	24
Kegunaan	8	638	14
Pembayaran	29	24	2268

Setelahnya, diperoleh nilai performa dari model klasifikasi untuk setiap kelas aspek. Detail hasil pengukuran performa model klasifikasi aspek dapat dilihat pada Tabel 12.

**Tabel 12.** Nilai Performa Model Aspek Setiap Kelas

Kelas	Akurasi	Presisi	Recall	F1-Score
Bug	0.97	0.95	0.95	0.95
Kegunaan	0.97	0.95	0.97	0.96
Pembayaran	0.97	0.98	0.98	0.98

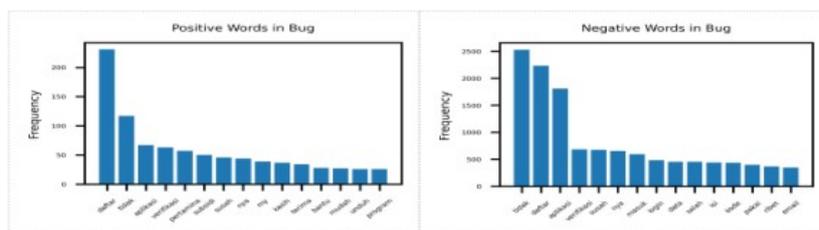
Dari Tabel 12 dapat disimpulkan bahwa model klasifikasi aspek yang telah dibangun memiliki nilai performa yang baik, dengan akurasi mencapai 97% untuk masing-masing kelas aspek, yaitu *Bug*, kegunaan, dan pembayaran. Selanjutnya, dilakukan pengujian menggunakan metode *K-Fold Cross Validation* dengan 5 iterasi dan menghasilkan nilai akurasi sebesar 96% untuk masing-masing kelas aspek tersebut.

### 3.5. Visualization

Dalam rangka memudahkan pemahaman terhadap data, data akan disajikan dalam bentuk visual atau grafik pada tahap ini. Pendekatan yang digunakan adalah dengan mengelompokkan visualisasi berdasarkan sentimen dari setiap aspek. Oleh karena itu, pada tahap visualisasi sentimen, data dibagi menjadi tiga bagian, yaitu aspek *Bug*, kegunaan, dan pembayaran, masing-masing dengan sentimen positif dan negatif.

#### 3.5.1. Aspek Bug

Berikut ini adalah hasil visualisasi data untuk aspek *Bug* dan sentimen positif serta negatif, yang ditampilkan dalam grafik *barchart* dapat dilihat dalam Gambar 2.



Gambar 2. Barchart Visualisasi Sentimen Aspek Bug

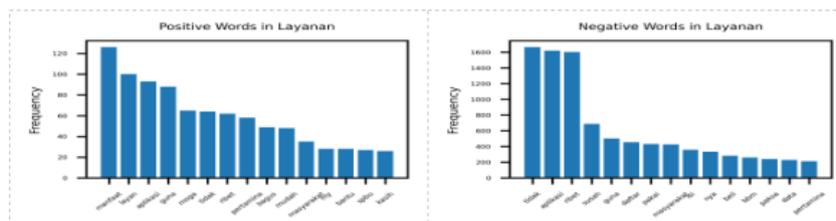
Berdasarkan hasil visualisasi *barchart* pada data ulasan dengan aspek *Bug*, didapatkan daftar kata-kata yang sering muncul. Pada data ulasan dengan aspek *Bug* dan sentimen positif, kata-kata yang sering muncul adalah daftar, tidak, aplikasi, verifikasi, Pertamina, subsidi, susah, nya, my, kasih, terima, bantu, mudah, unduh, dan program. Sementara pada data ulasan dengan aspek *Bug* dan sentimen negatif, kata-kata yang sering muncul adalah tidak, daftar, aplikasi, verifikasi, susah, nya, masuk, login, data, salah, isi, pakai, ribet, dan email. Selanjutnya, kata-kata yang memiliki frekuensi kemunculan tinggi dari kedua data tersebut ditampilkan dalam bentuk diagram *wordcloud*, yang dapat dilihat pada Gambar 3.



Gambar 3. Wordcloud Visualisasi Sentimen Aspek Bug

#### 3.5.2. Aspek Kegunaan

Berikut ini terdapat hasil visualisasi data dalam bentuk *barchart* untuk sentimen positif dan negatif pada aspek kegunaan, yang dapat dilihat pada Gambar 4.



Gambar 4. Barchart Visualisasi Sentimen Aspek Kegunaan

Berdasarkan hasil visualisasi *barchart* pada data ulasan dengan aspek kegunaan, ditemukan daftar kata-kata yang sering muncul. Pada data ulasan dengan aspek kegunaan dan sentimen positif, beberapa

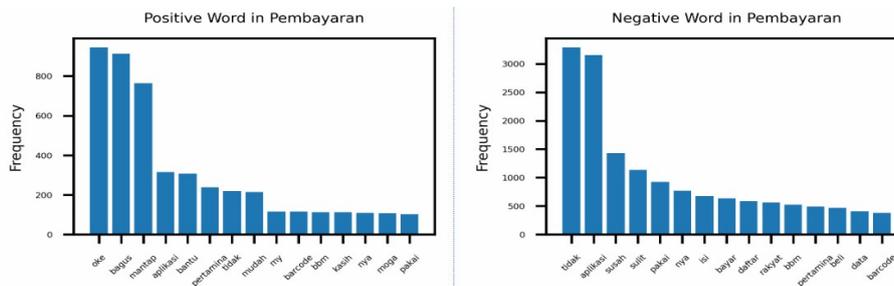
kata yang sering muncul meliputi manfaat, layanan, aplikasi, guna, harapan, tidak, ribet, Pertamina, bagus, mudah, masyarakat, milik saya, membantu, SPBU, dan kasih. Sementara pada data ulasan dengan aspek kegunaan dan sentimen negatif, kata-kata yang sering muncul antara lain tidak, aplikasi, ribet, sulit, guna, daftar, menggunakan, masyarakat, isi, milik saya, membeli, BBM, dipaksa, data, dan Pertamina. Selanjutnya, kata-kata yang memiliki frekuensi kemunculan tinggi dari kedua data tersebut ditampilkan dalam bentuk diagram *wordcloud* yang dapat dilihat pada Gambar 5.



Gambar 5. Wordcloud Visualisasi Sentimen Aspek Kegunaan

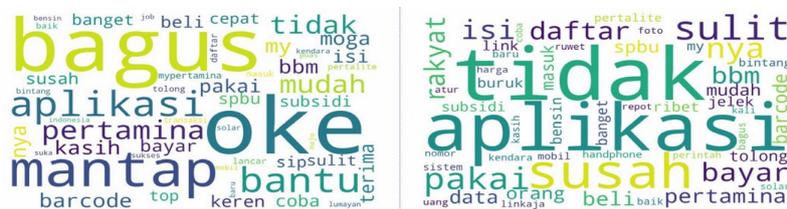
### 3.5.3. Aspek Pembayaran

Berikut adalah hasil visualisasi data untuk aspek pembayaran, yang dibagi menjadi sentimen positif dan negatif, yang ditampilkan dalam bentuk *barchart* pada Gambar 6.



Gambar 6. Barchart Visualisasi Sentimen Aspek Pembayaran

Berdasarkan hasil visualisasi *barchart* pada data ulasan dengan aspek pembayaran, ditemukan kumpulan kata-kata yang sering muncul. Pada data ulasan dengan aspek pembayaran dan sentimen positif, terdapat beberapa kata yang sering muncul seperti oke, bagus, mantap, aplikasi, bantu, Pertamina, tidak, mudah, my, barcode, BBM, kasih, nya, moga, dan pakai. Di sisi lain, pada data ulasan dengan aspek pembayaran dan sentimen negatif, kata-kata yang sering muncul antara lain tidak, aplikasi, susah, sulit, pakai, nya, isi, bayar, daftar, rakyat, BBM, Pertamina, beli, data, dan barcode. Selanjutnya, kata-kata yang memiliki frekuensi kemunculan tinggi dari kedua data tersebut ditampilkan dalam bentuk diagram *wordcloud* yang dapat dilihat pada Gambar 7.



Gambar 7. Wordcloud Visualisasi Sentimen Aspek Pembayaran

## 4. KESIMPULAN

Dalam penelitian ini, SVM digunakan sebagai algoritma untuk mengklasifikasikan data ulasan dan TF-IDF sebagai metode pembobotan kata. Tahap awal yang dilakukan adalah melakukan pengambilan data ulasan pengguna aplikasi MyPertamina dari Google Play Store dengan menggunakan *library* Google Play Scraper. Setelah itu, data yang telah berhasil diambil diberikan label dalam 2 kelas sentimen dan 3 kelas aspek. Totalnya terdapat 19562 data ulasan yang dikumpulkan dari rentang waktu Agustus 2022 hingga Maret 2023.

Dalam setiap aspek, aplikasi MyPertamina dinilai membantu pengguna dalam pembelian BBM, namun terdapat kendala yang dirasakan pengguna. Pada aspek *Bug*, pengguna mengalami kesulitan dalam mendaftar akun dan sering mengalami kegagalan *login*. Pada aspek kegunaan, beberapa pengguna mengalami kesulitan, kebingungan, dan kerumitan dalam penggunaannya. Sementara pada aspek pembayaran, terdapat keluhan tentang terbatasnya pilihan metode pembayaran dan kadang-kadang tidak muncul *barcode*.

Setelah melakukan evaluasi dengan menghasilkan rata-rata *K-Fold*, ditemukan bahwa model klasifikasi sentimen yang dibangun memiliki tingkat akurasi sebesar 92%. Sedangkan untuk model klasifikasi aspek, memiliki rata-rata tingkat akurasi sebesar 96%. Dengan demikian, dapat disimpulkan bahwa model yang telah dibuat cukup andal dalam melakukan klasifikasi pada data ulasan aplikasi MyPertamina.

## DAFTAR PUSTAKA

- [1] R. Islamia, I. R. Al Faizy, A. Aqilla, R. F. Ahmad, A. Z. P. Arum, and G. Pratama, "Dampak Kenaikan Harga Bahan Bakar Minyak (Bbm) Terhadap Sembilan Bahan Pokok (Sembako) Di Toko Sani Kabupaten Cirebon," *J. Ekon. Manaj.*, vol. 17, no. 2, pp. 1–7, 2022, [Online]. Available: <http://oaj.stiecirebon.ac.id/index.php/jem>
- [2] D. Rahayuningtyas, R. Laksmono, and Y. D. Kuncjoro, "Analisis Pemanfaatan Coral Reef Sebagai Penyimpanan Cadangan Strategis Energi Untuk Ketahanan Energi Nasional," *J. Ketahanan Energi*, vol. 7, pp. 44–59, 2021, [Online]. Available: <https://jurnalprodi.idu.ac.id/index.php/KE/article/view/1070>
- [3] A. Hendra, "Pelaksanaan Praktik Industri (PI) Mahasiswa PTK UNM di PT. Pertamina (Persero) Marketing Operation Region VII Sulawesi," <https://tik.ft.unm.ac.id/>, 2019. <https://tik.ft.unm.ac.id/index.php/berita/index/Pelaksanaan-Praktik-Industri-x01PIx02-Mahasiswa-PTK-UNM-di-PT.-Pertamina-x01Perserox02-Marketing-Operation-Region-VII-Sulawesi> (accessed Oct. 08, 2022).
- [4] Darmayuda, E. A. Pailis, A. F. Aulia, and M. Hasnan, "Analisis Pergeseran Pola Konsumsi Energi dan Dampak Ekonominya Pada Saat Covid-19 di Provinsi Riau," *J. Ekon. dan Bisnis*, vol. 10, no. 1, pp. 203–216, 2022.
- [5] N. Andari, "MyPertamina, Kenali Aplikasi dan Fungsinya Selain Buat Beli Bensin," <https://www.carmudi.co.id/>, Jun. 30, 2022. <https://www.carmudi.co.id/journal/my Pertamina-kenali-aplikasi-dan-fungsinya-selain-buat-beli-bensin/> (accessed Oct. 08, 2022).
- [6] R. M. Ibrahim and N. N. K. Moeliono, "Persepsi Konsumen Pada My Pertamina (Studi Pada Penggunaan My Pertamina Kota Bandung)," *J. Ilm. Mhs. Ekon. Manaj.*, vol. 5, no. 2, pp. 396–413, 2020, [Online]. Available: <http://jim.unsyiah.ac.id/ekm>
- [7] Syamsir, A. Lutfi, A. A. Fitriani, I. Ramadani, N. A. Putri, and Y. S. Nelsi, "Efektivitas Penggunaan Aplikasi My Pertamina Di Era Kenaikan Bbm Bersubsidi," *Pros. Semin. Nas. Pendidikan, Bahasa, Sastra, Seni, dan Budaya*, vol. 1, no. 2, pp. 245–253, 2022.
- [8] S. Fransiska, Rianto, and A. Irham Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Sci. J. Informatics*, vol. 7, no. 2, pp. 2407–7658, 2020, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>
- [9] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [10] M. Izunahdi, G. Aburrahman, and A. E. Wardoyo, "Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google PlayStore Menggunakan Metode Multinomial Naive Bayes Sentiment Analysis on KAI Access Application Review Data on Google PlayStore Using Multinomial Naive Bayes Method," vol. 1, no. 1, pp. 1–7, 2021.
- [11] R. Wahyudi and G. Kusumawardana, "Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [12] L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.
- [13] I. Santoso, Windu Gata, and Atik Budi Paryanti, "Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 364–370, 2019, doi: 10.29207/resti.v3i3.1084.
- [14] N. W. Utami and I. G. J. E. Putra, "TEXT MINING CLUSTERING UNTUK PENGELOMPOKAN TOPIK DOKUMEN PENELITIAN MENGGUNAKAN ALGORITMA K-MEANS DENGAN COSINE SIMILARITY," *JINTEKS (Jurnal Inform. Teknol. dan Sains)*, vol. 4, p. 66, 2022, [Online]. Available: <http://www.jurnal.uts.ac.id/index.php/JINTEKS/article/view/1907/1021>
- [15] 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)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [16] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 6, no. 2, p. 129, 2021, doi: 10.30998/string.v6i2.10133.
- [17] Y. Julianto, D. H. Setiabudi, and S. Rostianingsih, "Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine," *J. Infra*, vol. 10, no. 1, 2022.