

**STRATEGI PERBAIKAN KUALITAS PADA APLIKASI
MYPERTAMINA UNTUK PRODUK BERSUBSIDI BERBASIS
TEXT MINING DAN SWOT**

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1
Program Studi Teknik Industri – Fakultas Teknologi Industri
Universitas Islam Indonesia**



Nama : Reza Dian Shafitri
No. Mahasiswa : 21522111

**PROGRAM STUDI TEKNIK INDUSTRI PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2025**

PERNYATAAN KEASLIAN

PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.

Yogyakarta 09 - 05 - 2025



21522111

SURAT BUKTI PENELITIAN



FAKULTAS
TEKNOLOGI INDUSTRI

Gedung KH. Mas Mansur
Kampus Terpadu Universitas Islam Indonesia
Jl. Kaliurang km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 4100, 4101
F. (0274) 895007
E. fti@uii.ac.id
W. fti.uii.ac.id

SURAT KETERANGAN PENELITIAN

Nomor: 04/Ka.Lab.Datmin/70/Lab.Datmin/V/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Kami yang bertanda tangan dibawah ini, menerangkan bahwa mahasiswa dengan keterangan sebagai berikut :

Nama : Reza Dian Shafitri

No. Mhs : 21522111

Dosen Pembimbing : Dr. Ir. Dwi Handayani., S.T., M.Sc., IPU., ASEAN.Eng.

Telah selesai melaksanakan penelitian yang berjudul " STRATEGI PERBAIKAN KUALITAS PADA APLIKASI MYPERTAMINA UNTUK PRODUK BERSUBSIDI BERBASIS TEXT MINING DAN SWOT" di Laboratorium Data Mining, Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia tercatat mulai tanggal 1 Januari sampai dengan tanggal 31 Maret 2025

Demikian surat keterangan kami keluarkan, agar dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 27 Mei 2025

Kepala Laboratorium
Data Mining

Dr. Dwi Adi Purnama, S.T.

LEMBAR PENGESAHAN PEMBIMBING

**STRATEGI PERBAIKAN KUALITAS PADA APLIKASI
MYPERTAMINA UNTUK PRODUK BERSUBSIDI BERBASIS TEXT
MINING DAN SWOT**



Yogyakarta, 27 Mei 2025

Dosen Pembimbing

(Dr. Dwi Handayani, S.T., M.Sc., IPU., ASEAN.Eng)

LEMBAR PENGESAHAN DOSEN PENGUJI
STRATEGI PERBAIKAN KUALITAS PADA APLIKASI
MYPERTAMINA UNTUK PRODUK BERSUBSIDI BERBASIS TEXT
MINING DAN SWOT
TUGAS AKHIR

Disusun Oleh:

Nama : Reza Dian Shafitri

No. Mahasiswa : 21522111

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 27 Mei 2025

Tim Penguji

**Dr. Dwi Handayani, S.T., M.Sc., IPU.,
ASEAN. Eng**

Ketua

Nashrullah Setiawan, S.T., M.Sc., Ph.D

Anggota I

Dian Janari, S.T., M.T.

Anggota II

Mengetahui,

**Ketua Program Studi Teknik Industri Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia**



Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM
NIK 015220101

HALAMAN PERSEMBAHAN

*Dengan menyebut nama Allah Yang Maha Pengasih Lagi Maha Penyayang
Alhamdulillahirabbil alaamiin, atas segala nikmat yang Engkau berikan, penulis
mempersembahkan skripsi yang telah terselesaikan ini kepada kedua orang tua penulis,*

Papa Hanggono Raharjo

Mama Dwi Herlinawati

*Dan teruntuk keluarga besar serta kerabat penulis yang senantiasa mendoakan dan
memberikan dukungan kepada penulis*

*Dan tidak lupa penulis berterimaasih kepada diri sendiri karena telah berjuang hingga dititik
ini dan selalu berusaha dan bersabar dalam menyelesaikan Tugas Akhir ini.*

MOTTO

“Allah tidak mengatakan hidup ini mudah. Tetapi Allah berjanji, bahwa sesungguhnya Bersama kesulitan ada kemudahan”

(QS. Al – Insyirah 5-6)

“Rasakanlah setiap proses yang kamu tempuh dalam hidupmu, sehingga kamu tau betapa dirimu sudah berjuang sampai detik ini”

“Keberhasilan bukanlah milik orang pintar, tetapi keberhasilan adalah milik mereka yang senantiasa berusaha, apapun hasilnya nanti, setidaknya sudah berusaha”

(Bj Habibie)

KATA PENGANTAR

Bismillahirrohmanirrohim

Assalamualaikum Warahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin, segala puji dan syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini yang berjudul **“Strategi Perbaikan Kualitas Pada Aplikasi Mypertamina Untuk Produk Bersubsidi Berbasis Text Mining Dan Swot”**, sebagai bentuk salah satu syarat untuk mendapatkan gelar sarjana (S1) Jurusan Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia.

Pada Tugas Akhir ini merupakan salah satu syarat yang harus dipenuhi untuk memperoleh gelar sarjana strata-1 pada Program Studi Teknik Industri Universitas Islam Indonesia. Dari Tugas Akhir yang dilaksanakan ini, penulis dapat mengetahui penerapan ilmu yang dipelajari selama kuliah dalam penelitian ilmiah.

Selama pelaksanaan Tugas Akhir ini, tentunya tidak lepas dari adanya dukungan, doa, bantuan, bimbingan, dan nasihat dari berbagai pihak yang telah berkontribusi baik secara langsung maupun tidak langsung. Oleh karena itu, penulis mengucapkan rasa terimakasih kepada:

1. Bapak Prof. Dr. Ir. Hari Purnomo, M.T., IPU., ASEAN.Eng, selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Ir. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM., selaku Ketua Program Studi Teknik Industri Program Sarjana Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Ibu Dr. Ir. Dwi Handayani, S.T., M.Sc., IPU., ASEAN.Eng, selaku dosen pembimbing Tugas Akhir yang telah memberikan bimbingan, arahan, bantuan, kritik dan saran serta meluangkan waktu sehingga penulis mampu menyelesaikan laporan Tugas Akhir.
4. Papa Hanggono Raharjo dan Mama Dwi Herlinawati selaku orang tua saya serta Keluarga besar saya yang selalu memberikan doa, dukungan, dan semangat dalam mengerjakan Tugas Akhir ini.
5. Dan tak lupa untuk Fadhlán Ardhi As'shiddiqy dan Keluarganya, saya ingin mengucapkan Terimakasih telah banyak membantu saya dalam banyak hal dan senantiasa telah memberikan support untuk menyelesaikan tugas akhir ini.
6. Terimakasih teman terbaik selama saya kuliah yaitu Sekar Wijayaning Tyas, Nicky Amanda Putri, Gilang Surya Herlambang yang selalu mengerjakan Bersama dalam progress Tugas Akhir dan mendengarkan saya dalam berkeluh kesah.
7. Serta Annisa Oktaviyana, M. Faizul Ilmi, Rafii Syahanan Ardian, Ghaffar Tuasikal, Ahmad Makki Madani, dan Nabil Ahmad Wiro'i, Terima kasih telah mendukung, menghibur, serta memberikan semangat kepada penulis.
8. Bagian Akademik dan Administrasi Prodi yang membantu dalam memberikan informasi serta kelancaran pengurusan berkas maupun tahapan yang berkaitan dalam penyusunan laporan.
9. Semua pihak yang telah membantu penulis dalam menyelesaikan laporan tugas akhir ini yang tidak dapat disebutkan satu per satu.

Dalam penulisan Tugas Akhir ini penulis menyadari bahwa penulisan masih jauh dari kata sempurna dan banyak kekurangan, oleh karena itu penulis mengharapkan kritik dan saran yang bersifat membangun dari pembaca sehingga laporan ini dapat menjadi lebih baik. Semoga Tugas Akhir ini dapat bermanfaat tidak hanya bagi penulis namun juga bagi pembaca. Diharapkan pula dengan adanya Tugas Akhir ini dapat menambah wawasan ilmu pengetahuan dan dipergunakan sebaik mungkin oleh pembaca.

Alhamdulillahirabbil'alamin.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 9 Mei 2025



Reza Dian Shafitri

NIM 21522111

ABSTRAK

Perkembangan teknologi digital telah mendorong transformasi layanan publik, termasuk dalam distribusi produk bersubsidi seperti bahan bakar minyak (BBM). Salah satu inovasi yang diterapkan adalah aplikasi MyPertamina sebagai sarana pembayaran non-tunai dan registrasi subsidi tepat. Namun, peningkatan jumlah pengguna turut diiringi dengan berbagai keluhan, yang tercermin dari rating rendah dan ulasan negatif di platform aplikasi. Penelitian ini bertujuan untuk mengidentifikasi permasalahan utama yang dihadapi pengguna dan memberikan strategi perbaikan berbasis analisis data ulasan. Metode yang digunakan dalam penelitian ini adalah *text mining* terhadap 6.984 ulasan pengguna MyPertamina di Google Play Store dengan pendekatan *K-Means Clustering* untuk mengelompokkan ulasan berdasarkan kemiripan tema. Pemilihan jumlah cluster ditentukan menggunakan metode Elbow. Selanjutnya, dilakukan analisis SWOT pada masing-masing cluster untuk merumuskan strategi perbaikan yang relevan. Hasil analisis menunjukkan empat cluster utama dengan karakteristik berbeda: dua cluster bersentimen negatif, satu bersifat netral, dan satu menunjukkan respons positif. Strategi perbaikan yang disarankan mencakup *maintenance* sistem, perbaikan interface dan user experience, penambahan metode pembayaran, serta rutin melakukan survei kepuasan dan kebutuhan pengguna agar pengembangan aplikasi lebih tepat sasaran. Penelitian ini memberikan kontribusi praktis bagi pengelola aplikasi MyPertamina dalam merancang langkah perbaikan yang lebih terarah berbasis data, guna meningkatkan kepuasan pengguna dan efektivitas distribusi subsidi BBM.

Kata Kunci: Analisis Sentimen, *Text Mining*, *K-Means*, Metode *Elbow*, Analisis SWOT

DAFTAR ISI

PERNYATAAN KEASLIAN	i
SURAT BUKTI PENELITIAN.....	ii
LEMBAR PENGESAHAN PEMBIMBING	iii
LEMBAR PENGESAHAN DOSEN PENGUJI	iv
HALAMAN PERSEMBAHAN.....	v
MOTTO.....	vi
KATA PENGANTAR	vii
ABSTRAK.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Penelitian	7
1.6 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1 Kajian Literatur	9
2.2 Landasan Teori	15
2.2.1 Aplikasi.....	15
2.2.2 Google Playstore Crawling.....	15
2.2.3 Analisis Sentimen	16
2.2.4 Text Mining	17
2.2.5 Pre-processing Data	18
2.2.6 Pembobotan TF-IDF	20
2.2.7 Clustering.....	21
2.2.8 K-Means Clustering.....	21
2.2.9 Elbow.....	22
2.2.10 WordCloud	23
2.2.11 Analisis SWOT.....	24
BAB III METODE PENELITIAN	25
3.1 Objek dan Subjek Penelitian	25
3.2 Pengumpulan Data	25
3.3 Metode Pengumpulan Data	25
3.4 Jenis dan Sumber Data	26
3.5 Variabel Penelitian	26
3.6 Metode Analisis Data	26
3.7 Alur Penelitian.....	27
BAB IV PENGUMPULAN DAN PENGOLAHAN DATA.....	33
4.1 Pengumpulan Data.....	33
4.2 Pengolahan Data.....	38
4.2.1 Pre-processing Data	38
BAB V PEMBAHASAN	46
5.1 Analisis Karakteristik <i>Cluster</i> Dengan <i>Word Cloud</i>	46

5.1.1	Cluster 1.....	46
5.1.2	Cluster 2.....	49
5.1.3	Cluster 3.....	53
5.1.4	Cluster 4.....	57
5.2	Karakteristik Masing-Masing <i>Cluster</i>	60
5.3	Analisis SWOT Berdasarkan Hasil <i>Clustering</i>	69
BAB VI	PENUTUP	75
6.1	Kesimpulan.....	75
1.7	Saran.....	76
DAFTAR PUSTAKA		78
LAMPIRAN.....		84

DAFTAR TABEL

Tabel 2. 2 Perbandingan Penelitian.....	11
Tabel 2. 3 Contoh Kata <i>Stopword</i>	19
Tabel 2. 4 Kombinasi Aturan Awalan dan Akhiran.....	19
Tabel 4. 1 Bahasa Pemrograman <i>Scrapping Data</i>	34
Tabel 4. 2 Perbedaan Data Setelah Tahap <i>Case Folding</i> dan <i>Tokenizing</i>	40
Tabel 4. 3 Perbedaan Data Setelah Tahap <i>Filerting</i>	41
Tabel 4. 4 Perbedaan Data Setelah Tahap <i>Stemming</i>	41
Tabel 5. 1 Frekuensi dan Pembobotan Kata <i>Cluster 1</i>	47
Tabel 5. 2 Frekuensi dan Pembobotan Kata Gabungan Dua Kata <i>Cluster1</i>	48
Tabel 5. 3 Frekuensi dan Pembobotan Kata <i>Cluster 1</i>	50
Tabel 5. 4 Frekuensi dan Pembobotan Kata Gabungan Dua Kata <i>Cluster 2</i>	52
Tabel 5. 5 Frekuensi dan Pembobotan Kata <i>Cluster 3</i>	54
Tabel 5. 6 Frekuensi dan Pembobotan Kata Gabungan Dua Kata <i>Cluster 3</i>	55
Tabel 5. 7 Frekuensi dan Pembobotan Kata <i>Cluster 4</i>	57
Tabel 5. 8 Frekuensi dan Pembobotan Kata Gabungan Dua Kata <i>Cluster 4</i>	59
Tabel 5. 9 Karakteristik Ulasan Tiap <i>Cluster</i>	60
Tabel 5. 10 Menentukan Setiap SWOT	69
Tabel 5. 11 Hasil dari TOWS <i>Matrix</i>	71

DAFTAR GAMBAR

Gambar 1. 1 Banyaknya Pengguna Aplikasi MyPertamina	2
Gambar 1. 2 Ulasan Aplikasi MyPertamina.....	3
Gambar 2. 1 Proses <i>Text Mining</i>	18
Gambar 2. 2 Contoh Visualisasi <i>WordCloud</i>	23
Gambar 3. 1 Alur Penelitian.....	28
Gambar 3. 2 Alur Penelitian (2)	29
Gambar 4. 1 <i>Google Colaboratory</i>	33
Gambar 4. 2 Aplikasi MyPertamina pada <i>Google Playstore</i>	34
Gambar 4. 3 Hasil <i>Scrapping Data</i>	37
Gambar 4. 4 <i>Scrapping</i> Ulasan Aplikasi MyPertamina 1 Januari 2024 -12 Januari 2025.....	37
Gambar 4. 5 Jumlah Ulasan Tiap Periode (Tahun)	38
Gambar 4. 6 Pemilihan Variabel Data <i>Cleaning</i>	39
Gambar 4. 7 Penghapusan <i>Duplicate Data</i>	40
Gambar 4. 8 Hasil Pembobotan Setiap Ulasan	43
Gambar 4. 9 Grafik Metode <i>Elbow</i>	43
Gambar 4. 10 Grafik <i>Clustering</i> Data Ulasan	44
Gambar 4. 11 Jumlah Data Setiap <i>Cluster</i>	45
Gambar 5. 1 <i>Word Cloud Cluster 1</i>	46
Gambar 5. 2 <i>Word Cloud Cluster 2</i>	50
Gambar 5. 3 <i>Word Cloud Cluster 3</i>	54
Gambar 5. 4 <i>Word Cloud Cluster 4</i>	57

BAB I

PENDAHULUAN

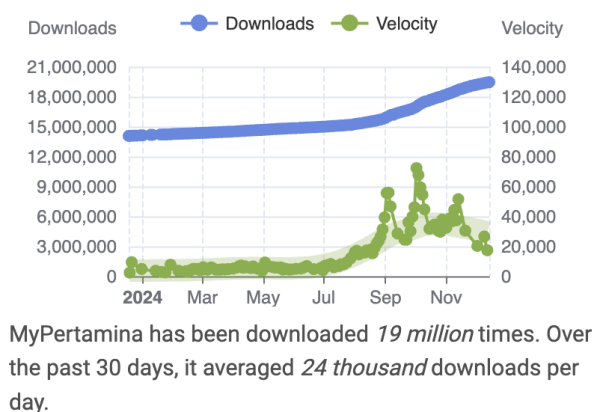
1.1 Latar Belakang

Pada era digital seperti sekarang, transaksi non tunai merupakan kegiatan pembayaran yang mengacu pada mekanisme transaksi tanpa menggunakan uang dalam bentuk fisik (Safa'ah & Indriyanti, 2023). Perkembangan teknologi digital di Indonesia telah mendorong berbagai sektor untuk mengintegrasikan layanan mereka melalui platform digital, termasuk sektor energi (Kusumastuti et al., 2022). Sehingga era digital, atau era teknologi dan informasi yang ditandai dengan teknologi di segala bidang usaha, mengharuskan para pelaku usaha untuk mengadopsinya sebagai bagian dari kegiatan bisnis mereka. Kemudian untuk pemahaman masyarakat saat ini terhadap penggunaan teknologi internet mendukung perusahaan untuk melakukan perubahan ke arah transformasi digital (Saktyadi & Tricahyono, 2024). Aplikasi digital telah menjadi salah satu sarana utama bagi perusahaan dalam menyediakan layanan yang lebih cepat dan efisien kepada pelanggan. Dalam industri energi, aplikasi MyPertamina merupakan salah satu contoh inovasi teknologi yang bertujuan untuk memudahkan pelanggan dalam melakukan pembelian bahan bakar, mengakses promosi, dan mendapatkan informasi terkait layanan Pertamina. Pada tanggal 1 Juli 2022, pemerintah melalui PT Pertamina Persero mewajibkan pengguna BBM bersubsidi untuk mendaftar melalui aplikasi MyPertamina atau *website* Pertamina. Kebijakan yang dilakukan oleh pemerintah bertujuan agar subsidi BBM tepat sasaran, karena data yang didaftarkan di aplikasi akan menunjukkan konsumen mana yang berhak mendapatkan subsidi atau tidak (Maulana et al., 2023).

MyPertamina merupakan program *loyalty* dan *e-payment* yang memberikan *user experience* dari PT Pertamina (Persero) dengan mudah untuk seluruh pelanggan Pertamina (Tualang Ksatria et al., 2023). Aplikasi ini juga merupakan aplikasi layanan keuangan digital dari Pertamina dan anggota Badan Usaha Milik Negara (BUMN) yang terintegrasi dengan aplikasi Link Aja, dimana aplikasi ini digunakan untuk pembayaran bahan bakar minyak (BBM) secara non-tunai di Stasiun Pengisian Bahan Bakar Umum (SPBU) Pertamina yang hadir karena sebagai bentuk apresiasi kepada seluruh pelanggan setia PT. Pertamina yang bekerja sama dengan LinkAja dalam mendukung sistem pembayaran untuk memudahkan pelanggan saat berinteraksi (Apriantoro et al., 2022).

Dilansir dari Kompas.com pada 26 Januari 2023, PT Pertamina melaksanakan uji coba pembelian Solar subsidi dengan pendaftaran melalui fitur Subsidi Tepat di aplikasi MyPertamina dan situs web resmi mereka. Dalam uji coba ini, konsumen yang ingin membeli Solar subsidi atau Biosolar harus menunjukkan QR code saat melakukan pembelian. Konsumen yang belum terdaftar dalam program Subsidi tepat hanya diperbolehkan membeli Solar subsidi sebanyak 20 liter per hari. Sementara itu, konsumen yang telah mendaftar dan memiliki QR code bisa membeli Solar subsidi sesuai dengan ketentuan dalam Surat Keputusan Kepala Badan Pengatur Hilir Minyak dan Gas Bumi (BHP Migas) No. 04/P3JBT/BPH Migas/KOM/2020 (Jatnika et al., 2023). Dengan penerapan fitur Subsidi Tepat dan QR code ini, diharapkan pembelian BBM bersubsidi menjadi lebih terkontrol dan tepat sasaran, serta mendukung efektivitas program subsidi untuk masyarakat yang memenuhi syarat. Selain itu, sistem ini diharapkan dapat meningkatkan efisiensi transaksi pembelian BBM bersubsidi melalui sistem non-tunai di MyPertamina.

Berdasarkan Gambar 1.1 dapat dilihat bahwa aplikasi MyPertamina telah diunduh sebanyak 19 juta kali hingga akhir tahun 2024. Selama 30 haru terakhir, aplikasi ini mencatat rata-rata 24 ribu unduhan per hari dengan rating 4,3 dan ulasan lebih dari 400 ribu pada situs *Google Playstore*. Hal ini dapat menunjukkan adanya tingkat adopsi yang tinggi dari masyarakat terhadap aplikasi MyPertamina, yang disertai denngan rating dan berbagai ulasan bukan berarti bahwa keinginan atau masalah pengguna terpenuhi dan berhenti untuk melakukan evaluasi dalam peningkatan pelayanan terhadap pengguna. Masalah tersebut dapat terjadi karena beberapa pengguna yang memberikan rating tinggi namun ulasannya berupa ulasan yang negatif (Indrayanto et al., 2023).



Gambar 1. 1 Banyaknya Pengguna Aplikasi MyPertamina

Didapatkan bawa pemberian rating aplikasi MyPertamina masih banyak pengguna yang memberikan rating rendah yang dapat dilihat pada Gambar 1.2 sehingga hal ini perlu dilakukan perbaikan untuk meningkatkan *rating* dari aplikasi tersebut. Sebelum dilakukan perbaikan, perlu dicari akar masalah apa saja yang menyebabkan aplikasi MyPertamina memiliki *rating* yang rendah. *Review* dari pengguna aplikasi MyPertamina digunakan sebagai sumber data untuk digali untuk menemukan masalah dari aplikasi tersebut. *Text mining* didefinisikan sebagai salah satu teknik yang dapat diimplementasikan untuk mengklasifikasikan data dengan menemukan pola yang menarik dari kumpulan data teks yang besar (Pratama & Lestika Atmi, 2020). Menurut Antons et al. (2020), *text mining* pada dasarnya memerlukan pendekatan kuantitatif untuk menganalisis data teks tersebut. Implementasi *text mining* pertama kali dilakukan pada pertengahan tahun 1980 dan terus berkembang hingga sekarang merambah ke berbagai bidang, seperti bisnis, ekonomi, dan edukasi (Zanini & Dhawan, 2015). Penelitian ini bertujuan untuk mencari akar masalah yang terdapat pada aplikasi MyPertamina melalui ulasan yang diberikan pengguna pada platform *Google Playstore*.



Gambar 1. 2 Ulasan Aplikasi MyPertamina

Pendapat dan komentar yang diungkapkan oleh seseorang sering kali dipengaruhi oleh bias pribadi terhadap suatu topik, sehingga opini yang disampaikan menjadi subjektif (R. Kumar et al., 2020). Kebiasaan ini sering terlihat dalam interaksi di media sosial, di mana komentar yang dibuat cenderung mencerminkan perasaan atau penilaian pribadi daripada fakta objektif. Hal ini menghadirkan tantangan bagi sistem yang mencoba memahami data berbasis teks, terutama ketika harus berurusan dengan kalimat-kalimat subjektif (Gobel et al., 2024). Seperti halnya dengan aplikasi lainnya, tidak semua aplikasi yang ada dipasaran memiliki fitur dan fungsi yang optimal, sehingga aplikasi pada MyPertamina menerima berbagai ulasan dari penggunaannya di platform seperti *Google Play Store* (Hendi Prastiyo & Lahindah, 2024). Untuk

memahami dan mengatasi permasalahan yang dihadapi pengguna, analisis terhadap data ulasan pelanggan menjadi sangat penting. Salah satu pendekatan yang efektif untuk menganalisis data ulasan adalah dengan menggunakan *text mining*, yaitu proses ekstraksi informasi dari data teks yang tidak terstruktur. Melalui *text mining*, pola dan tema utama dalam ulasan pelanggan dapat diidentifikasi, sehingga memberikan wawasan yang berharga bagi pengambilan keputusan manajemen.

Terdapat beberapa algoritma dalam *clustering* seperti *K-Means Clustering*, DBSCAN, *Affinity Propagation Algorithm*, dan lain-lain. Pendekatan *clustering* sendiri tergolong dalam pendekatan *unsupervised learning* yang merupakan kategori *machine learning* yang tidak membutuhkan label dalam dataset dan tidak membutuhkan identifikasi kelas-kelas (Roihan et al., 2020). Algoritma pada *machine learning* sendiri yaitu *Supervised*, *Unsupervised Learning*, Algoritma klasifikasi seperti *Logistic Regression*, *K-Nearest Neighbor* (KNN), *Decision Tree*, *Random Forest*, *Support Vector Machine* (SVM) dan *Neural Network*. Selain itu, ada beberapa metode *Features Selection* seperti *Lasso*, *Pearson Correlation*, dan juga ada tentang *cross validation* untuk mencegah *Overfitting* serta yang terakhir metode untuk melakukan evaluasi model menggunakan *Confusion Matrix* dan kurva ROC dan AUC (Kristiawan & Widjaja, 2021). *K-Means Clustering* merupakan salah satu metode klasterisasi yang paling sering digunakan dalam analisis data. Algoritma ini berfungsi dengan membagi data ke dalam sejumlah K kelompok yang telah ditentukan sebelumnya, di mana K adalah jumlah kelompok yang diinginkan. Proses klasterisasi dimulai dengan pemilihan acak K titik sebagai pusat kelompok (*centroid*) di ruang data, kemudian setiap titik data dikelompokkan ke dalam kelompok dengan *centroid* terdekat (Yudhistira & Andika, 2023).

Beberapa penelitian mengenai ulasan aplikasi MyPertamina telah dilakukan menggunakan beberapa metode seperti *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) oleh R. R. C. Putra et al. (2023). Penelitian tersebut dilakukan untuk membandingkan 2 algoritma yaitu NBC dan SVM, sehingga didapatkan hasil akhir dari penelitian tersebut bahwa metode SVM memiliki tingkat akurasi yang lebih tinggi sebesar 99,5% dibandingkan dengan NBC yang mencapai 97,6%. Penelitian yang sama juga dilakukan oleh Simbolon et al. (2023) dengan menggunakan algoritma NBC dan SVM. Hasil penelitian ini difokuskan untuk mengetahui hasil dari sentimen dengan 2 algoritma tersebut dan disimpulkan bahwa dengan nilai akurasi (68.50%), precision (70.00%), recall (69.70%), dan F1 score (68.46%) yang lebih tinggi pada model *Support Vector Machine* (SVM) dibandingkan

dengan model *Naïve Bayes* dengan nilai akurasi (63.00%), precision (63.90%), recall (61.34%), dan F1 score (59.55%). Maka dari itu, model Support Vector Machine (SVM) dapat dipilih sebagai model yang lebih optimal.

Dalam penelitian ini, algoritma *K-Means Clustering* digunakan untuk mengelompokkan ulasan pelanggan berdasarkan kesamaan tema. Metode ini dipilih karena kemampuannya dalam mengelompokkan data ke dalam beberapa *cluster* dengan karakteristik yang serupa. Selain itu, untuk menentukan jumlah *cluster* yang optimal, digunakan metode *Elbow*, yang dapat membantu mengidentifikasi titik optimal berdasarkan perubahan inertia dalam proses *clustering*. Untuk meningkatkan akurasi dalam menganalisis teks, pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) akan diterapkan untuk menilai pentingnya kata-kata dalam dokumen secara proporsional.

Analisis SWOT dalam penelitian ini digunakan untuk memberikan gambaran kekuatan dari aplikasi MyPertamina yang nantinya dapat dikembangkan lebih lanjut dalam perbaikan berdasarkan ulasan dari pengguna serta digunakan untuk melakukan evaluasi terhadap suatu organisasi, rencana maupun aktivitas bisnis. Kerangka kerja ini memiliki singkatan dari *strength*, *weakness*, *opportunities*, dan *threats*. Analisis SWOT dapat memberikan informasi mengenai kondisi internal dan eksternal kepada perusahaan sehingga perusahaan nantinya dapat mengembangkan suatu strategi berdasarkan informasi kekuatan, kelemahan, peluang, dan ancaman yang perusahaan punya (Ramadhan et al., 2023).

Dengan kombinasi algoritma *K-Means*, metode *Elbow*, TD-IDF, dan Analisis SWOT penelitian ini bertujuan untuk memberikan strategi perbaikan kualitas yang relevan bagi aplikasi MyPertamina. Hasil dari *clustering* ulasan diharapkan mampu memberikan gambaran menyeluruh tentang kebutuhan dan harapan pelanggan, sehingga perusahaan dapat merancang langkah-langkah yang lebih tepat sasaran dalam meningkatkan kualitas aplikasi. Hal ini pada akhirnya akan mendukung keberlanjutan bisnis Pertamina di era digital.

1.2 Rumusan Masalah

Rendahnya rating aplikasi MyPertamina membutuhkan pendekatan khusus untuk mencari akar permasalahan yang terjadi dan bagaimana upaya yang dapat dilakukan untuk meningkatkan rating aplikasi MyPertamina tersebut dalam upaya untuk meningkatkan kepuasan pengguna. Berdasarkan uraian tersebut, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana pengelompokkan ulasan pengguna aplikasi MyPertamina berdasarkan algoritma *K-Means Clustering*?
2. Bagaimana hasil optimalisasi pengelompokkan ulasan pengguna aplikasi MyPertamina berdasarkan metode *Elbow*?
3. Tindakan apa yang dapat dilakukan untuk meningkatkan rating aplikasi MyPertamina berdasarkan analisis SWOT?

1.3 Tujuan Penelitian

Perkembangan teknologi digital mendorong perusahaan untuk terus berinovasi dalam meningkatkan kualitas layanan, termasuk di sektor energi. Aplikasi MyPertamina yang dikembangkan oleh PT Pertamina bertujuan mempermudah masyarakat dalam melakukan transaksi pembelian bahan bakar. Namun, seiring dengan meningkatnya jumlah pengguna, berbagai keluhan terkait kualitas layanan aplikasi muncul. Adapun berikut tujuan dari penelitian ini dilakukan yaitu sebagai berikut:

1. Mengelompokkan ulasan pengguna aplikasi MyPertamina berdasarkan algoritma *K-Means Clustering*.
2. Mengetahui hasil optimalisasi pengelompokkan ulasan pengguna aplikasi MyPertamina berdasarkan metode *Elbow*.
3. Memberikan tindakan apa saja yang dapat dilakukan untuk meningkatkan rating aplikasi MyPertamina berdasarkan analisis SWOT.

1.4 Manfaat Penelitian

Adapun hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Akademisi
 - a. Memberikan kontribusi terhadap pengembangan metode analisis data tidak terstruktur, khususnya melalui text mining dengan algoritma *K-Means*.
 - b. Menjadi referensi bagi penelitian lanjutan di bidang teknik industri terkait penerapan teknologi informasi dalam meningkatkan kualitas layanan.
 - c. Menambah wawasan tentang integrasi teknologi digital dan pendekatan kuantitatif dalam pemecahan masalah industri.
2. Bagi Perusahaan dan Pembaca

- a. Memberikan panduan untuk memahami persepsi dan kebutuhan pelanggan melalui analisis ulasan pengguna aplikasi.
- b. Membantu mengidentifikasi faktor utama yang memengaruhi kepuasan pengguna aplikasi MyPertamina.
- c. Menyediakan strategi perbaikan yang berbasis data, sehingga dapat meningkatkan kinerja aplikasi dan loyalitas pelanggan.
- d. Mendukung transformasi digital perusahaan dalam era teknologi informasi, sehingga mampu bersaing di industri energi secara lebih efektif.

1.5 Batasan Penelitian

Dalam penelitian ini memiliki batasan dengan tujuan agar penelitian yang dilakukan memiliki arah dan tujuan yang jelas. Adapun batasan pada penelitian ini yaitu sebagai berikut:

1. Data yang digunakan berupa ulasan dari aplikasi MyPertamina yang diambil dari *Google Playstore* tanpa ada pengulangan.
2. Data yang digunakan hanya ulasan aplikasi yang digunakan dari tanggal 1 Januari 2024 hingga 12 Januari 2025
3. Ulasan yang digunakan yaitu dengan menggunakan Bahasa Indonesia.

1.6 Sistematika Penulisan

Penulisan dalam penelitian ini dibuat secara terstruktur agar memudahkan pembaca dalam memahami proses penelitian dari identifikasi masalah hingga pemberian rekomendasi. Sistematika penulisan ini terdiri dari:

BAB I

PENDAHULUAN

Pada bab ini menjelaskan terkait latar belakang permasalahan yang diambil, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II

KAJIAN PUSTAKA

Bab ini berisi tinjauan pustaka yang mencakup teori-teori dasar yang bersumber referensi buku, jurnal, maupun penelitian-penelitian sebelumnya yang berkaitan dengan masalah yang diambil, serta konsep-konsep yang relevan

dengan penelitian. Kajian pustaka ini digunakan sebagai dasar dalam pengolahan data dan analisis.

BAB III

METODOLOGI PENELITIAN

Bab ini menguraikan langkah-langkah sistematis yang ditempuh selama penelitian. Diagram alur digunakan untuk menggambarkan rangkaian proses penelitian, mulai dari pengumpulan data, pengolahan data, hingga metode yang digunakan untuk analisis. Diagram ini juga membantu menjelaskan urutan penggunaan algoritma *K-Means Clustering*, metode *Elbow*, dan pembobotan TF-IDF dalam menganalisis ulasan pengguna aplikasi MyPertamina.

BAB IV

HASIL DAN PENGOLAHAN DATA

Bab ini berisi hasil pengolahan data ulasan pengguna aplikasi MyPertamina, analisis hasil klasterisasi, dan interpretasi data yang telah diolah. Data disajikan dalam bentuk tabel, grafik, dan visualisasi yang relevan.

BAB V

PEMBAHASAN

Bab ini mendiskusikan hasil analisis data yang telah diperoleh, memberikan penjelasan mendalam tentang temuan penelitian, serta merumuskan strategi perbaikan kualitas aplikasi berdasarkan hasil klasterisasi.

BAB VI

KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan yang menjawab rumusan masalah, ringkasan temuan utama, serta saran untuk perbaikan aplikasi MyPertamina dan penelitian di masa mendatang.

DAFTAR PUSTAKA

LAMPIRAN

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Literatur

Dari beberapa penelitian telah dilakukan pada pengelompokan analisis sentimen menggunakan *text mining* sebelumnya. Pada penelitian tersebut sangatlah penting dalam menjalankan penelitian ini sebagai kajian literatur untuk mengetahui adanya keterkaitan antara penelitian terdahulu dengan penelitian yang akan dilakukan, untuk mencegah terjadinya tindakan duplikasi yang dilakukan oleh penulis. Tujuan dari tinjauan untuk mengidentifikasi penelitian terdahulu yang relevan guna memahami perkembangan topik yang sedang diteliti serta menemukan celah atau masalah yang belum terselesaikan. Dengan meninjau penelitian sebelumnya, peneliti juga dapat memilih metode atau pendekatan yang sesuai berdasarkan pengalaman dan hasil yang telah dilakukan oleh peneliti lain.

Pada penelitian ini yang berkaitan untuk meningkatkan rating aplikasi telah banyak dilakukan sebelumnya diberbagai bidang seperti yang dilakukan oleh (Irfani, 2020) dan (Panjaitan & Manurung, 2022) dibidang pendidikan, bidang transportasi (Radiena & Nugroho, 2023) dan (Oktaviani et al., 2021), bidang *e-commerce* (Rejito et al., 2021) dan (Pratmanto et al., 2020), bidang media sosial (Indriyani et al., 2023), bidang layanan (Maulana et al., 2023) dan (R. R. C. Putra et al., 2023), bidang Kesehatan (Rais et al., 2022) dan lain-lain. Pada beberapa penelitian ini mengambil sumber data untuk ulasan yang digunakan dalam menganalisis, diambil dari *Google Playstore* seperti pada penelitian yang dilakukan oleh (Oktaviani et al., 2021) Dimana dari ulasan para pengguna akan dilakukan pengambilan data melalui *text mining*. Kemudian sumber data yang lain seperti yang dilakukan oleh (Rejito et al., 2021) dimana mengambil sumber data dari *Twitter* yang dikumpulkan sebanyak 885 *tweet* yang nantinya akan dilakukan menggunakan *text mining*.

Penelitian – penelitian sebelumnya menunjukkan bahwa penerapan algoritma pada *text mining* seperti *Support Vector Machine* (SVM), *Naïve Bayes Classifier* (NBC), dan *K-Means* yang telah banyak digunakan untuk menganalisis sentiment pengguna terhadap berbagai aplikasi digital. Seperti penelitian yang dilakukan oleh Rais et al. (2022) menggunakan metode *Naïve Bayes Classifier* untuk menganalisis ulasan aplikasi Peduli Lindungi dan mencatat akurasi sebesar 73%, menunjukkan potensi besar pendekatan ini dalam menangkap opini publik. Sementara itu, Indriyani et al. (2023) membandingkan metode SVM dan NBC pada ulasan aplikasi Tiktok dan meemukan bahwa SVM lebih unggul dalam klasifikasi Sentimen

dengan hasil evaluasi menunjukkan akurasi sebesar 79% untuk Naïve Bayes dan 84% untuk SVM. Bahkan studi oleh Simbolon et al. (2023) dan Maulana et al. (2023) secara khusus menerapkan kedua algoritma pada ulasan aplikasi MyPertamina dan menunjukkan bahwa sebagian besar sentimen pengguna bersifat negatif, terutama terkait kendala teknis dan fungsionalitas aplikasi. Namun, kajian-kajian tersebut umumnya hanya berhenti pada pengklasifikasian sentimen dan belum melangkah lebih jauh ke arah penyusunan strategi perbaikan kualitas layanan. Selain itu, belum ada penelitian yang secara komprehensif menggabungkan hasil analisis text mining dengan pendekatan strategis seperti analisis SWOT. Maka dari itu, penelitian ini hadir untuk menjembatani kesenjangan tersebut dengan menyusun strategi perbaikan kualitas aplikasi MyPertamina berbasis data nyata dari ulasan pengguna, guna mendukung peningkatan layanan publik secara lebih tepat sasaran dan berkelanjutan.

Setelah mengkaji beberapa penelitian pada Tabel 2.1 sehingga diketahui terdapat perbedaan penelitian ini dengan penelitian sebelumnya. Penelitian ini menggunakan algoritma *K-Means Clustering*, metode Elbow, dan juga menggunakan metode tambahan seperti Analisis SWOT, yang mana belum pernah dilakukan untuk meneliti ulasan pada aplikasi MyPertamina. Adapun berikut merupakan perbedaannya dapat dilihat pada Tabel 2.2 yaitu:

Tabel 2. 1 Perbandingan Penelitian

No	Penulis	Tahun	Algoritma Penelitian			Metode/ Algoritma Pendukung Lainnya	Jenis Data		Objek Penelitian	
			SVM	Naïve Bayes	K-Means		Data Numerik	Data Kategorik	MyPertamina	Lainnya
1.	Irfani	(2020)	✓					✓		✓ (RuangGuru)
2.	Pratmanto et al.	(2020)		✓		✓ (Partitioning dan TF-IDF)		✓		✓ (Shopee)
3.	Rejito et al.	(2021)			✓			✓		✓ (Tokopedia)
4.	Oktaviani et al.	(2021)		✓		✓		✓		✓

No	Penulis	Tahun	Algoritma Penelitian			Metode/ Algoritma Pendukung Lainnya	Jenis Data		Objek Penelitian	
			SVM	Naïve Bayes	K-Means		Data Numerik	Data Kategorik	MyPertamina	Lainnya
5.	Safitri et al.	(2021)			✓	(Association)		✓		(Traveloka)
6.	Rais et al.	(2022)		✓		✓ (TF-IDF)		✓		(Skytrax)
7.	Panjaitan & Manurung	(2022)	✓					✓		✓ (Google Classroom)
8.	Indriyani et al.	(2023)	✓	✓				✓		✓ (Tiktok)
9.	Maulana et al.	(2023)		✓				✓	✓	
10.	R. R. C. Putra et al.	(2023)	✓	✓				✓	✓	

No	Penulis	Tahun	Algoritma Penelitian			Jenis Data		Objek Penelitian	
			SVM	Naïve Bayes	K-Means	Metode/ Algoritma Pendukung Lainnya	Data Numerik	Data Kategorik	MyPertamina Lainnya
11.	Radiena & Nugroho	(2023)	✓			(Model <i>Cross Industry Standard Process for Data Mining</i>)		✓	✓ (KAI Accses)
12.	Penelitan Selanjutnya	2024			✓	(Metode <i>Elbow</i> , TF-IDF, dan Analisis SWOT)		✓	✓

Berdasarkan rangkuman kajian induktif pada Tabel 2.1 dan Tabel 2.2, dapat disimpulkan bahwa beragam metode yang digunakan dalam melakukan pengelompokkan data yang dilakukan oleh peneliti terdahulu. Penelitian terkait algoritma SVM, *Naïve Bayes*, dan *K-Means* telah dilakukan dengan berbagai pendekatan. Irfani pada tahun 2020 menggunakan algoritma SVM dan *Naïve Bayes* untuk data kategorik. Penelitian Rejito et al. pada tahun 2021 juga melibatkan algoritma SVM untuk analisis data numerik dengan metode pendukung *K-Nearest Neighbor* (KNN). Sedangkan, penelitian Oktaviani et al. pada tahun 2021 lebih berfokus pada asosiasi data kategori menggunakan *Naïve Bayes*.

Pada tahun 2023, Rendy Rian Chrisna Putra et al. masing-masing menggunakan pendekatan berbeda untuk mengelompokkan ulasan aplikasi MyPertamina, dengan Rendy et al. memanfaatkan SVM dan *Naïve Bayes* pada analisis sentimen. Radiena & Nugroho pada tahun 2023 mengintegrasikan model *Cross-Industry Standard Process for Data Mining* dalam menganalisis data numerik dan kategori. Penelitian ini mengindikasikan bahwa metode yang digunakan sangat bergantung pada jenis data, dengan algoritma *K-Means* memungkinkan pengelompokan data secara optimal berdasarkan jumlah *cluster* yang terbentuk. Penelitian ini menggabungkan *K-Means Clustering*, metode *Elbow*, dan pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengelompokkan ulasan pengguna berdasarkan tema yang serupa. Hasil analisis akan diperkuat dengan metode *Strengths, Weaknesses, Opportunities, and Threats* (SWOT) guna mengidentifikasi faktor-faktor yang memengaruhi kualitas aplikasi. Dengan pendekatan ini, diharapkan penelitian dapat memberikan strategi perbaikan yang lebih tepat sasaran, meningkatkan pengalaman pengguna, serta memperkuat daya saing *MyPertamina* di era digital.

2.2 Landasan Teori

2.2.1 Aplikasi

Saat ini penggunaan internet serta penggunaan gadget semakin meningkat pada masyarakat maka dari itu banyak Perusahaan dengan mudah mendekatkan diri dengan konsumen melalui aplikasi. Apalagi sekarang Indonesia menempati peringkat 6 dengan jumlah penggunaan internet sebesar 83,7 juta pengguna, dan akan diramalkan terus bertambah setiap tahunnya (Soraya & Wahyudi, 2021). Menurut Kamus Bahasa Indonesia (2016), aplikasi diartikan sebagai suatu program komputer atau perangkat lunak yang didesain guna mengerjakan tugas-tugas tertentu. Secara umum, aplikasi merupakan alat terapan yang difungsikan secara khusus dan terpadu sesuai kemampuan yang dimilikinya sehingga aplikasi merupakan suatu perangkat komputer yang siap pakai bagi *user* (Fauzi Siregar et al., 2018). Dengan demikian, dapat diketahui bahwa secara harfiah, kata aplikasi memiliki beberapa makna tergantung pada konteks penggunaannya. Dalam penelitian ini, peneliti menggunakan pengertian aplikasi bukan sebagai kata kerja yang berarti penerapan, melainkan sebagai suatu program atau perangkat lunak yang dirancang untuk menyelesaikan masalah-masalah atau tugas tertentu. Menurut Harip Santoso mengartikan istilah aplikasi yaitu suatu kelompok file (*form, class, report*) yang bertujuan untuk melakukan aktivitas tertentu yang saling terkait, misalnya aplikasi *payroll*, aplikasi *fixed asset*, dll) Abdurahman et al. (2014).

Menurut Hengky W. Pramana, pengertian aplikasi adalah satu unit perangkat lunak yang sengaja dibuat untuk memenuhi kebutuhan akan berbagai aktivitas ataupun pekerjaan, seperti aktivitas perniagaan, periklanan, pelayanan Masyarakat, *game*, dan berbagai aktivitas lainnya yang dilakukan oleh manusia (Huda & Priyatna, 2019). Dalam penelitian ini, aplikasi yang digunakan sebagai objek penelitian adalah MyPertamina. Aplikasi ini dirancang untuk memberikan kemudahan bagi pengguna dalam melakukan tugas-tugas spesifik, seperti transaksi non-tunai, akses informasi layanan, dan proses digitalisasi lainnya. Penelitian ini akan berfokus pada bagaimana aplikasi tersebut meningkatkan efisiensi layanan dan memberikan pengalaman yang lebih baik bagi penggunanya.

2.2.2 Google Playstore Crawling

Google Playstore telah menyediakan wadah bagi *user* untuk mendapatkan informasi berharga untuk setiap aplikasi yang diunggah oleh pengembangnya (Asyhari et al., 2023). Wadah tersebut diantaranya adalah informasi terkait aplikasi seperti nama developer, nama aplikasi,

subtitle, jenis aplikasi, deskripsi singkat, deskripsi panjang, *rating* konten, persyaratan sistem, *url* halaman, pembaruan terakhir, penjelasan versi terbaru yang terakhir diunggah, harga untuk aplikasi berbayar, jumlah pengunduhan, keamanan data yang dijamin mulai dari data yang dibagikan, akses data yang digunakan, teknik keamanan datanya, hingga ulasan dari penggunaannya (Strzelecki, 2020).

Google Playstore sendiri merupakan suatu layanan distribusi aplikasi resmi oleh *google* dimana aplikasi ini menyediakan berbagai aplikasi yang bisa diunduh oleh pengguna dan menyediakan layanan bagi pengembang untuk mengunggah aplikasi (Amir Latif et al., 2019). Dalam pengambilan data atau pengumpulan informasi terkait ulasan dalam aplikasi MyPertamina melalui *Google Playstore* digunakan proses *Crawling*. *Crawling* atau perayapan merupakan suatu teknik yang digunakan untuk mengumpulkan informasi dari semua halaman *website* yang tergabung pada situs *website* tersebut (R. P. I. Putra et al., 2020). *Crawling* bekerja secara otomatis, dimana informasi yang dikumpulkan berdasarkan atas kata kunci yang diberikan oleh pengguna (Zuhri & Saputri, 2020).

Google Playstore Crawling adalah suatu Teknik untuk mengumpulkan data secara otomatis melalui platform *Google Playstore* agar mendapatkan informasi dengan penyerapan konten maupun ulasan yang akan diukur dan dianalisis (Bagnasco et al., 2015). Pada penelitian ini didapatkan 12.000 ulasan yang didapatkan dari *google playstore* yang akan dilakukan analisis. Proses *crawling* ini dilakukan dengan menggunakan *python*.

2.2.3 Analisis Sentimen

Semakin berkembangnya dalam era digital ini, internet telah menjadi sumber informasi yang melimpah. Informasi ini dapat membantu tiap individu maupun organisasi dalam pengambilan Keputusan. Namun, kemudahan akses terhadap informasi ini juga menghadirkan tantangan, terutama dalam menganalisis dan memahami opini kolektif dari berbagai sumber. Secara manual, proses ini sangat sulit dilakukan karena memerlukan waktu dan sumber daya yang besar (A. Kumar & Sebastian, 2012). Oleh karena itu, dibutuhkan sebuah teknologi yang mampu mengotomatisasi proses ini. Salah satu teknologi tersebut adalah *sentiment analysis* atau analisis sentimen, yang menjadi solusi untuk memahami opini secara terstruktur dan sistematis.

Analisis sentiment, yang dikenal juga sebagai *opinion mining*, merupakan salah satu bidang dalam pemrosesan bahasa alami (*Natural Language Processing* - NLP) yang berfokus

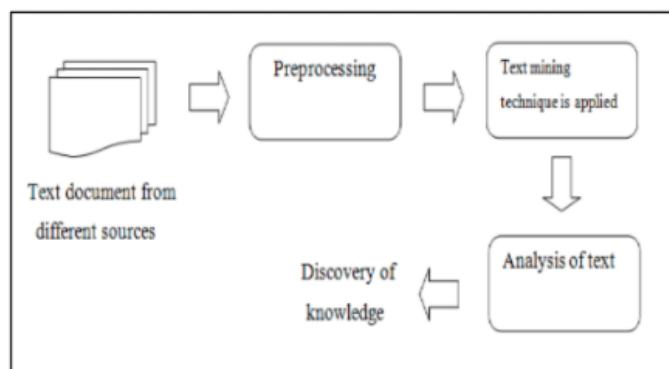
pada analisis opini atau sentimen yang terdapat dalam teks. Studi ini mencakup berbagai aspek, seperti pengenalan emosi dan suasana hati, klasifikasi genre, perhitungan relevansi, hingga pembuatan ringkasan berbasis opini. Analisis sentimen menjadi tren yang menarik dalam media sosial karena memiliki berbagai aplikasi praktis (Pang & Lee, 2008).

Dalam tugas dasarnya, analisis sentimen bertujuan untuk mengelompokkan polaritas dari teks yang terdapat dalam dokumen, kalimat, atau pendapat tertentu. Polaritas tersebut menunjukkan apakah teks tersebut bersifat positif, negatif, atau netral. Selain itu, analisis sentimen juga dapat digunakan untuk mengidentifikasi emosi tertentu seperti sedih, gembira, atau marah. Dengan teknologi ini, kita dapat mencari opini mengenai produk, merek, atau individu, dan menentukan bagaimana opini tersebut dipandang apakah secara positif atau negatif (Amrullah et al., 2020). Dapat disimpulkan bahwa analisis sentimen adalah sebuah ilmu yang mempelajari ulasan, baik positif maupun negatif, yang disampaikan oleh masyarakat atau pengguna. Tujuannya adalah untuk memperoleh masukan yang dapat digunakan sebagai dasar dalam melakukan perbaikan atau pengembangan terhadap hal yang menjadi objek ulasan tersebut.

2.2.4 Text Mining

Dalam penelitian ini, sebelum melakukan analisis sentiment diperlukan *pre-processing* teks dengan metode *text mining* yang bertujuan untuk mengolah data teks agar siap untuk dianalisis. Algoritma *text mining* merupakan algoritma yang digunakan dalam penggalian data supaya dapat melengkapi masalah kebutuhan informasi dengan menerapkan metode *machine learning*, *data mining*, *natural language processing*, manajemen pengetahuan serta pencarian informasi (Ridwansyah, 2022). Metode ini juga digunakan untuk mengubah teks yang tidak terstruktur menjadi format terstruktur guna mengidentifikasi pola, topik, kata kunci, dan atribut lainnya dalam suatu data.

Text mining yaitu proses dalam mengekstraksi informasi dari suatu sumber yang berupa data yang tidak terstruktur, Dimana tujuan dari proses ini adalah untuk menangani masalah *classification*, *clustering*, *information extraction*, dan *information retrieval* (Feldman & Sanger, 2007). Berikut merupakan alur proses dalam melakukan *text mining*.



Gambar 2. 1 Proses *Text Mining*

Berdasarkan ketidak teraturan struktur data teks, maka dalam proses *text mining* memerlukan beberapa tahap awal yang pada intinya mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Dalam penerapannya *text mining* proses yang dilakukan terbagi menjadi tiga tahap utama, yaitu proses awal terhadap teks (*text preprocessing*), transformasi teks (*text transformation*), dan penemuan pola (*pattern discovery*) (Maria Gomez Hidalgo, 2002). Pada sumber yang digunakan dalam *text mining* merupakan sekumpulan teks dengan susunan format yang tidak terstruktur dimana dengan karakteristik dokumen yang digunakan memiliki ukuran database yang relatif besar, dimensi tinggi, banyak mengandung kata yang bias atau ambigu serta mengandung kumpulan kata yang saling terkait antara satu kata dengan kata yang lain memiliki makna yang berbeda (Siregar & Puspabhuana, 2017).

2.2.5 *Pre-processing Data*

Dalam melakukan proses *text mining*, terdapat tahapan penting yang melibatkan serangkaian teknik untuk membersihkan dan mempersiapkan data teks sebelum dilakukan analisis lebih lanjut. Menurut Haddi et al. (2013) *pre-processing data* merupakan proses pembersihan dan persiapan teks untuk dilakukan klasifikasi dimana teks online biasanya banyak berisi noise dan bagian yang tidak informatif seperti *tag* HTML, skrip, dan iklan. Selain itu, pada tingkat kata terdapat banyak kata dalam teks yang tidak berdampak pada orientasi umum teks tersebut.

Terdapat beberapa Langkah dari tahapan *pre-processing* yang dilakukan sesuai kebutuhan dari data tersebut, seperti pada menghilangkan kalimat yang ganda hingga menghapus data-data yang tidak penting dalam proses analisis selanjutnya. Berikut merupakan tahapan *pre-processing data* yang dilakukan dalam menggunakan *text mining*.

1. *Case Folding*

Dalam penelitian *text mining*, *case folding* diperlukan karena proses bersifat *case sensitive* dan tidak semua teks menggunakan huruf kapital secara konsisten. Proses ini akan mengubah seluruh isi dokumen menjadi *lowcase* (Meisya Permata Aulia et al., 2021). Pada text preprocessing proses *case folding* bertujuan untuk mengubah semua huruf pada suatu dokumen teks menjadi huruf kecil, misalnya kata “Wifi” menjadi “wifi” (Rosid et al., 2020).

2. *Tokenizing*

Dokumen dari hasil *case folding* selanjutnya akan dipecah menjadi Kumpulan beberapa kata berdasarkan spasi atau disebut dengan *tokenizing* (Rosid et al., 2020).

3. *Stopword Removal*

Stopword Removal merupakan proses untuk menghilangkan kata yang tidak penting. Tujuannya sendiri untuk mengurangi dimensi teks masukan sehingga proses lebih ringan (Meisya Permata Aulia et al., 2021). Contoh kata-kata tersebut adalah “yang”, ‘dan’, ‘kamu’, ‘sampai’ dihapus dari dokumen karena tidak diukur sebagai kata kunci dalam aplikasi text mining (Rosid et al., 2020). Tahapah ini biasa disebut juga dengan filtering yang mana selain menghapus kata juga menghilangkan spasi yang berlebih pada suatu kalimat yang dianalisis. Berikut merupakan contoh kata yang masuk daftar *stopword*.

Tabel 2. 2 Contoh Kata *Stopword*

<i>Stopword</i> Indonesia	<i>Stopword</i> Manual
Aku	Deh
Di	Halo
Dan	Tuh
Sejak	Wow
Belum	loh

4. *Stemming*

Pada tahap *stemming* merupakan tahap merubah kata yang berimbuhan menjadi akar kata (Mustofa & Novita, 2022). Tujuan dalam proses ini untuk menghapus awalan, sufiks, dan afiks agar mengembalikan kata-kata kembali ke bentuk dasarnya sesuai dengan KBBI (Ramanda MD et al., 2024).

Tabel 2. 3 Kombinasi Aturan Awalan dan Akhiran

Awalan	Akhiran yang Tidak Diizinkan
Se-	-i,, -kan
Me-	-an

Ke-	-i, -kan
Di-	-an
Be-	-i

2.2.6 Pembobotan TF-IDF

Metode TF-IDF (*Term Frequency-Invers Document Frequency*) merupakan suatu metode yang digunakan dalam menentukan seberapa jauh keterhubungan kata (*term*) terhadap dokumen dengan memberikan bobot setiap kata. Metode ini akan menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dari inverse frekuensi dokumen yang mengandung kata tersebut (Andreswari et al., 2024). Tujuan dari penggunaan metode ini yaitu untuk mengevaluasi seberapa penting suatu kata (*term*) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar. TF-IDF terdiri dari dua komponen utama, yaitu.

1. *Term Frequency* (TF)

Term Frequency mengukur seberapa sering suatu kata muncul dalam sebuah dokumen dibandingkan dengan total kata dalam dokumen tersebut (Qaiser & Ali, 2018). Dalam setiap dokumen memiliki panjang yang berbeda-beda maka nilai TF akan dibagi dengan Panjang dokumen (Tri Putra et al., 2023).

$$tf_{t,d} = \frac{n_{t,d}}{(Total\ number\ of\ term\ in\ document)} \quad (2.1)$$

Dimana:

Tf = Frekuensi kemunculan kata pada sebuah dokumen

2. *Inverse Document Frequency* (IDF)

Setelah berhasil menghitung nilai TF akan menghitung nilai IDF yang merupakan nilai untuk mengukur seberapa penting sebuah kata, dengan beracuan semakin kecil nilai IDF maka akan dianggap semakin tidak penting kata tersebut, begitupun sebaliknya (Tri Putra et al., 2023).

$$idf_d = \log \frac{Number\ of\ Document}{(Total\ number\ of\ term\ in\ document)} \quad (2.2)$$

Dimana:

Idf = mengukur penting/tidak sebuah kata dalam dokumen

Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut

menunjukkan seberapa umum kata tersebut (Haris Ahmad Gozali et al., 2020). Untuk rumus pembobotan teks dengan metode TF-IDF ditunjukkan oleh rumus 2.3 berikut.

$$W_{i,j} = tf_{i,j} \cdot \log \left(\frac{N}{df_i} \right) \quad (2.3)$$

2.2.7 Clustering

Menurut Vijay Gaikwad (2014) *Clustering* merupakan suatu metode analisis data yang digunakan untuk mengelompokkan data menjadi grup-grup (*cluster*) berdasarkan Tingkat kesamaan antar elemen data. Dalam konteks pengelompokan dokumen, *clustering* bertujuan untuk menemukan kelompok dokumen dengan konten yang serupa. Hasil dari proses *clustering* berupa partisi yang disebut sebagai *cluster* ($\bar{P}$), di mana setiap *cluster* terdiri dari sejumlah dokumen ($\bar{d}$). Analisis *cluster* adalah suatu teknik statistik *multivariat* yang ditujukan untuk membuat klasifikasi individu individu atau objek-objek ke dalam kelompok-kelompok lebih kecil yang berbeda satu. *Clustering* merupakan proses untuk mengelompokkan suatu sampel, dimana sampel-sampel yang ada dalam tiap grup memiliki suatu kesamaan (Hajar et al., 2022). *Clustering* juga dapat dikatakan sebagai salah satu masalah *unsupervised learning* yang penting, sehingga seperti masalah-masalah lain yang serupa, *clustering* berurusan dengan menemukan struktur dalam suatu kumpulan data.

Dalam Teknik ini akan mencari struktur dalam data tanpa memerlukan label kelas yang telah ditentukan sebelumnya (Syakur et al., 2018). Terdapat berbagai metode dalam *clustering* seperti *K-Means*, *Hierarchical Clustering*, atau DBSCAN, *clustering* memungkinkan pengelompokan data yang serupa berdasarkan pada atribut atau fitur tertentu, sehingga memungkinkan analisis lebih lanjut terhadap kelompok-kelompok ini untuk mengungkap pola-pola yang tersembunyi atau hubungan yang signifikan di antara data (Hendrastuty, 2024).

2.2.8 K-Means Clustering

Metode ini dimulai dengan proses penentuan jumlah *cluster* terlebih dahulu. Salah satu metode non-hierarki adalah *K-means*. Metode *K-means* digunakan sebagai alternatif metode *cluster* untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dibandingkan metode hierarki. Penggunaan K-means untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam *cluster* tertentu berdasarkan rata-rata terdekat (Goreti et al., 2016).

Analisis *cluster* metode *K-means clustering* adalah prosedur analisis *cluster* yang digunakan untuk mengidentifikasi kelompok kasus atau objek secara relatif sama yang didasarkan pada karakteristik-karakteristik yang sudah dipilih. *K-means* merupakan salah satu metode pengelompokan data *non-hierarki* yang berusaha menggabungkan data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini menggabungkan data ke dalam kelompok sehingga data berkarateristik sama dimasukkan ke dalam kelompok yang sama. Proses pengklasifikasian dengan metode *K-means* adalah sebagai berikut (Goreti et al., 2016):

- a. Menentukan besarnya c , yaitu banyaknya *cluster*.
- b. Menentukan *centroid* secara acak (terserah peneliti)
- c. Menghitung jarak tiap objek dengan setiap *centroid* menggunakan jarak *Euclidean*
- d. Mengelompokkan setiap data berdasarkan jarak terdekat data tersebut dengan setiap *centroid*.
- e. Menentukan posisi *centroid* yang baru pada proses iterasi dengan menghitung nilai rata-rata dari data yang ada pada *centroid* yang sama menggunakan persamaan 2.4.

$$C_{hj} = \frac{\sum_{k=1}^n X_{kj}}{n_h} \quad (2.4)$$

- f. Mengulangi langkah 3 jika posisi *centroid* baru tidak sama dengan *centroid* yang sebelumnya.

Pengecekan *konvergensi* dilakukan dengan melihat pada iterasi sebelumnya dengan iterasi yang sedang berjalan. Jika hasilnya sama maka proses *clustering* sudah *konvergen*, tetapi jika berbeda maka belum *konvergen* sehingga perlu dilakukan iterasi berikutnya.

2.2.9 Elbow

Metode *elbow* merupakan suatu metode dimana pada suatu titik tertentu terjadi penurunan yang signifikan dalam grafik, berbentuk lengkungan yang tajam. Penentuan ini dilakukan dengan mengamati persentase perbandingan antar jumlah *cluster* yang divisualisasikan dalam bentuk grafik. Grafik tersebut menunjukkan titik tertentu di mana terjadi penurunan signifikan pada nilai *error* atau *Sum of Square Error* (SSE), membentuk sebuah lengkungan tajam menyerupai siku (*elbow*). Titik ini kemudian digunakan untuk menentukan nilai k , yaitu jumlah *cluster* yang paling baik (Syahfitri et al., 2023).

2.2.11 Analisis SWOT

Analisis SWOT adalah alat strategis untuk menilai kekuatan (*strengths*), kelemahan (*weaknesses*), peluang (*opportunities*), dan ancaman (*threats*) dalam suatu organisasi (Azzahra et al., 2025). Analisis ini diperkenalkan pertama kali oleh Albert Humphrey pada tahun 1960-an, analisis ini juga dapat membantu organisasi atau perusahaan dalam menemukan posisi unik di pasar dan mengembangkan strategi bisnis (Teoli et al., 2025). SWOT juga merupakan kerangka kerja yang dengan tujuan melakukan evaluasi terhadap posisi kompetitif perusahaan dan mengembangkan perencanaan strategis berdasarkan hasil identifikasi faktor internal dan eksternal. Kerangka kerja ini memiliki empat komponen penting yaitu kekuatan, kelemahan, peluang, dan ancaman (Ramadhan et al., 2023).

Menurut Sagala (2013) pada kekuatan (*strengths*) yaitu akan melihat apa saja hal-hal yang menjadi kekuatan sebagai modal yang dapat diandalkan, kelemahan (*weakness*) yaitu melihat hal-hal yang dipandang menjadi kelemahan sehingga membentuk prioritas untuk mengatasi kelemahan tersebut, peluang (*opportunities*) yaitu peluang apa saja yang mungkin dapat diraih untuk mengatasi kelemahan dan mendukung kekuatan dan tantangan, serta ancaman (*threats*) yaitu hal-hal yang dapat menjadi tantangan baik dilihat dari hal positif maupun yang negatif sehingga dapat dijadikan sebagai pemicu dalam meningkatkan strategi yang efisien dan efektif. Matriks TOWS (*Threats, Opportunities, Weaknesses, Strengths*) dibangun berdasarkan dari hasil analisis SWOT yang telah dilakukan. Matriks ini akan membantu dalam mencocokkan kekuatan dan kelemahan internalnya dengan peluang dan ancaman eksternalnya (Rahmawanti & Hastuti, 2019).

BAB III

METODE PENELITIAN

3.1 Objek dan Subjek Penelitian

Pada penelitian objek yang dituju yaitu ulasan yang diberikan oleh pengguna aplikasi MyPertamina pada *database* yang ada di *website Google Playstore*. Ulasan tersebut digunakan untuk menganalisis sentimen dan mendapatkan pola dari pendapat pengguna terhadap aplikasi tersebut.

3.2 Pengumpulan Data

Populasi pada penelitian ini yaitu data historis ulasan pengguna aplikasi MyPertamina yang diambil dari *Google PlayStore* dengan jumlah total 12.000 ulasan, dimana ulasan ini dimulai sejak bulan September 2018 hingga Januari 2025. Akan tetapi, sampel data yang akan digunakan merupakan data ulasan pada tanggal 1 Januari 2024 hingga 12 Januari 2025. Data dalam penelitian ini difokuskan pada ulasan yang diberikan pada rentang waktu Januari 2024 hingga Januari 2025.

Pemilihan rentang waktu ini dilakukan untuk menangkap respons masyarakat terhadap penerapan kebijakan baru pembayaran BBM menggunakan QR Code di SPBU yang mulai disosialisasikan pada tahun 2024. Jumlah ulasan yang digunakan sebanyak 6.984 ulasan yang diperoleh dari Google Play Store. Pemilihan ulasan dalam rentang waktu ini bertujuan untuk menganalisis persepsi pengguna terhadap digitalisasi pembayaran BBM bersubsidi, termasuk dampak sosial dan ekonominya, serta untuk melihat dinamika opini masyarakat selama masa transisi penerapan kebijakan tersebut. Adapun ulasan yang digunakan hanya ulasan yang berbahasa Indonesia untuk memastikan relevansi dan kemudahan analisis.

3.3 Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini menggunakan teknik *web scrapping* dengan menggunakan alat bantu *Google Colaboratory* dengan bahasa pemrograman *python* dalam mengeksplorasi semua ulasan dari aplikasi MyPertamina yang diambil dari *Google Playstore*. *Google Colab* merupakan perangkat lunak berbasis *cloud* yang mirip dengan *Jupyter Notebook* gratis dan dijalankan melalui *browser* seperti *Google Chrome*. *Platform* ini memungkinkan

pengguna untuk menjalankan kode pemrograman *Python* tanpa perlu melakukan instalasi atau pengaturan terlebih dahulu.

3.4 Jenis dan Sumber Data

Jenis data yang digunakan pada penelitian ini yaitu data sekunder. Data sekunder ini diambil dari *review* melalui *Google Playstore* dengan bantuan *web scraping* data ulasan menggunakan *Google Colab*. Sumber data yang digunakan dalam penelitian ini berasal dari aplikasi MyPertamina dengan alamat domain yaitu https://play.google.com/store/apps/details?id=com.dafturn.mypertamina&pcampaignid=web_share. Data yang diperoleh berupa ulasan aplikasi MyPertamina yang diambil pada tanggal 1 Januari 2024 hingga 12 Januari 2025 sebanyak 6.984 ulasan.

Dan juga dalam sumber data sekunder pada penelitian ini mengambil referensi dari beberapa jurnal, berita, artikel, dan informasi yang berhubungan dengan data penunjang penelitian.

3.5 Variabel Penelitian

Dalam data yang diperoleh, terdapat beberapa variabel yang memberikan informasi terkait ulasan. Namun, penelitian ini hanya menggunakan dua variabel utama dari keseluruhan variabel yang tersedia. Variabel tersebut meliputi:

a. *at*

Variabel *at* memberikan informasi mengenai waktu ketika ulasan dibuat oleh pengguna. Variabel ini digunakan sebagai acuan untuk membatasi pengambilan ulasan dalam rentang waktu tertentu yang telah ditetapkan.

b. *Content*

Variabel *content* mencakup teks ulasan yang diberikan oleh pengguna aplikasi. Variabel ini menjadi data utama yang akan dianalisis, diproses, dan dikelompokkan lebih lanjut berdasarkan isi ulasan yang disampaikan.

3.6 Metode Analisis Data

Pada metode analisis data yang digunakan dalam penelitian ini, yaitu menggunakan *Google Colab* dengan bahasa pemrograman *python* dan *Microsoft Excel* dalam melakukan prosesnya. Adapun metode yang digunakan untuk menganalisis data ulasan dalam penelitian ini yaitu.

1. Pengolahan Data

Data yang diperoleh dari *crawling* diproses untuk menghilangkan data duplikat, menghapus simbol atau tanda baca yang tidak relevan, dan membersihkan data menggunakan teknik *preprocessing* seperti tokenisasi, *stemming*, dan *stopword removal*.

2. Analisis Sentimen, hal ini digunakan untuk menilai ulasan pengguna dan kecenderungan sebuah ulasan terhadap suatu topik negatif, positif, maupun netral.

3. Pembobotan Data

Data teks diubah menjadi bentuk numerik menggunakan metode TF-IDF untuk memberikan bobot pada setiap kata.

4. Metode *Machine Learning*

Algoritma *K-Means Clustering* digunakan untuk mengelompokkan ulasan berdasarkan kesamaan isi. Dengan algoritma ini, ulasan pengguna dikelompokkan ke dalam beberapa *cluster*, yang masing-masing mewakili tema atau pola tertentu dalam data ulasan. Tujuannya adalah untuk memahami kategori utama dari umpan balik pengguna.

5. *WordCloud*

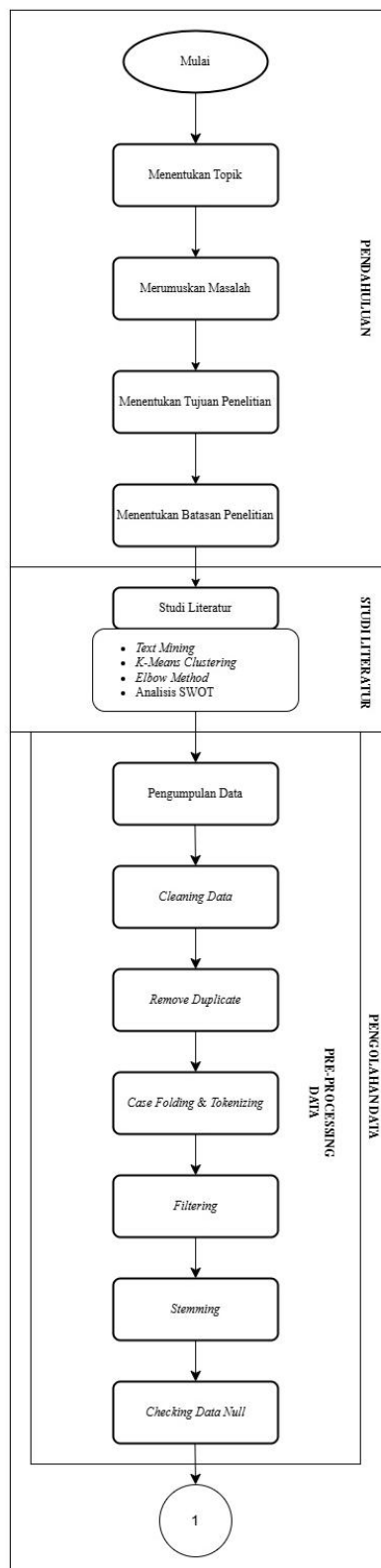
WordCloud digunakan untuk memvisualisasikan kata-kata yang paling sering muncul dalam ulasan pengguna aplikasi MyPertamina. Metode ini membantu peneliti mengidentifikasi kata kunci atau topik dominan yang sering dibahas oleh pengguna.

6. Analisis SWOT

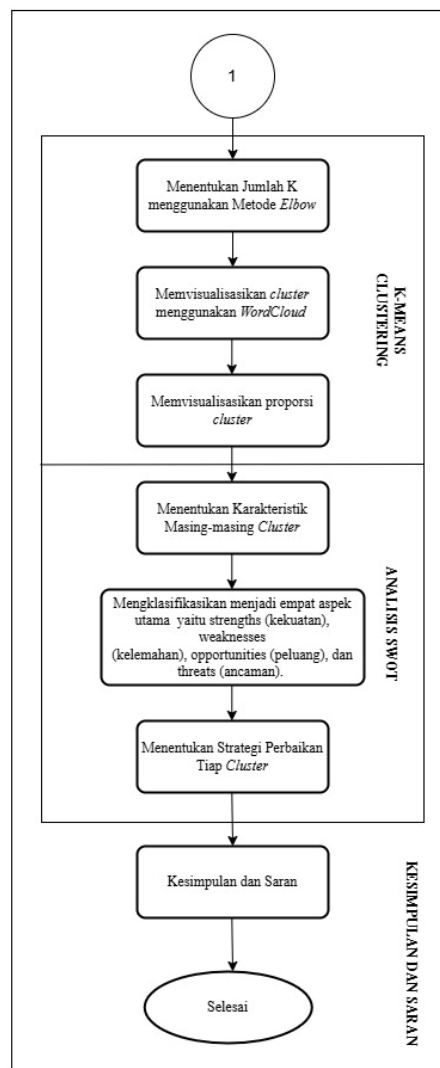
Analisis SWOT diterapkan untuk mengevaluasi berdasarkan faktor *Strengths*, *Weaknesses*, *Opportunities*, dan *Threats* untuk menentukan strategi perbaikan kualitas aplikasi MyPertamina.

3.7 Alur Penelitian

Alur dalam penelitian merupakan rangkaian prosedur yang harus dijalani secara bertahap agar sesuai dengan apa yang menjadi tujuan. Dengan alur penelitian yang telah disusun dengan baik, maka diharapkan segala aktivitas dan kegiatan dalam penelitian dapat terprogram dengan baik pula. Alur penelitian harus dilaksanakan dengan urut dan terstruktur supaya sesuai dengan SOP masing-masing. Dengan alur penelitian yang urut dan terstruktur ini harus diselesaikan mulai dari awal hingga akhir proses. Adapun berikut merupakan tahapan diagram alir dalam melakukan penelitian ini.



Gambar 3. 1 Alur Penelitian



Gambar 3. 2 Alur Penelitian (2)

Berdasarkan dari Gambar 3.1 dan Gambar 3.2 terkait alur penelitian, berikut merupakan penjelasan pada masing-masing tahapannya:

1. Mulai
2. Menentukan Topik

Dalam penelitian ini diawali dengan penentuan topik yang akan dibahas serta merancang Langkah-langkah sistematis untuk analisis tahapan-tahapan selanjutnya. Adapun topik yang diambil dalam pembahasan penelitian ini yaitu terkait pengelompokkan ulasan para pengguna aplikasi MyPertamina serta pemberian strategi yang dapat diusulkan sebagai perbaikan hasil ulasan yang kurang.

3. Identifikasi Masalah

Pada tahap ini dilakukan pengidentifikasian masalah dengan melihat kondisi pada ulasan yang diberikan oleh pengguna aplikasi MyPertamina meliputi *rating* yang diberikan pada aplikasi oleh pengguna, gambaran umum *review* yang diberikan pengguna, dan kenaikan jumlah pengguna dari aplikasi tersebut. Dari identifikasi masalah yang dilakukan, maka dapat disusun rumusan masalah, batasan masalah, dan manfaat dari penelitian yang dilakukan.

4. Merumuskan Masalah

Setelah dilakukan pengidentifikasian masalah dan melakukan riset dari permasalahan terkait topik pada penelitian ini, selanjutnya dapat dibentuk beberapa rumusan permasalahan yang nantinya akan dilanjutkan untuk tujuan dari dilakukannya penelitian ini.

5. Menentukan Tujuan, Batasan, dan Manfaat Penelitian

Pada penentuan tujuan penelitian ini, akan berisikan terkait suatu perancangan agar dapat menjawab pembahasan dari rumusan masalah yang telah dibuat sebelumnya dan juga sebagai acuan hal-hal apa saja yang menjadi capaian dalam penelitian nantinya.

Dalam penelitian ini menggunakan beberapa batasan peneliti yang diperlukan agar saat melakukan penelitiannya dapat lebih fokus pada hal tertentu dan tidak terlalu luas membahas topik sejenisnya. Adapun Batasan masalah yang digunakan yaitu data yang akan diggunakan berupa ulasan dari aplikasi MyPertamina yang diambil dari *Google Playstore* tanpa ada pengulangan. Rentang data ulasan yang digunakan dari tanggal 1 Januari 2024 hingga 12 Januari 2025. Dalam pengambilan ulasan yang digunakan yaitu dengan menggunakan Bahasa Indonesia.

6. Studi Literatur

Pada tahap ini dilakukan studi literatur yang berkaitan dengan topik dari beberapa sumber guna memperkuat informasi-informasi yang relevan dengan topik penelitian. Proses yang dilakukan pada studi literatur ini yaitu melakukan kajian induktif seperti merangkum penelitian-penelitian terdahulu yang selaras dengan topik pada penelitian ini dan landasan teori berupa teori yang digunakan sebagai pendukung pada penelitian yang dilakukan.

7. Pengumpulan Data

Pengumpulan data didapatkan dengan cara *scrapping* melalui ulasan aplikasi MyPertamina pada *Google Playstore*. Data ulasan yang digunakan pada awalnya adalah keseluruhan ulasan tanpa melihat tanggal pembuatannya, sehingga didapatkan 12.000 data ulasan yang menjadi data primer awal. Setelah melakukan pengumpulan data, selanjutnya dilakukan

proses *pre-processing* yang akan membuat data bersih dan nantinya data akan semakin berkurang dari data awal.

8. *Pre-processing Data*

Tahap ini akan melakukan *pre-processing data* sehingga akan menyeleksi data-data yang tidak diperlukan agar menjadi data yang lebih siap digunakan dan lebih terstruktur. Dengan itu akan membuat proses pengolahan lebih jelas dan akurat. Pada *pre-processing* terdapat beberapa aktivitas yaitu:

a. *Cleaning Data*

Pada proses ini, data yang diambil belum bersih sehingga dibutuhkan normalisasi data. Ulasan yang diberikan seseorang tidak semuanya menggunakan bahas baku dan masih banyak menggunakan kata gaul, misalnya “gue”, “lu”, dan lain-lain. Ada juga yang menggunakan potongan kata seperti “yg”, “brp”, “gmh”, dan lain sebagainya. Penelitian ini juga hanya berfokus pada data yang dibutuhkan seperti isi ulasan dan waktu ulasan tersebut dilihat.

b. *Remove Duplicate*

Setelah data ulasan dipilih dengan periode tertentu, kemudian data-data yang *double* seperti ulasan yang dikirim dua kali maupun kesalahan lainnya akan dihapus. Nantinya data yang digunakan hanya ulasan tunggal dari para pengguna.

c. *Case Folding*

Pada proses ini akan dilakukan tahap menyeragamkan ulasan menjadi huruf kecil, menghapus angka, dan menghapus adanya tanda baca.

d. *Tokenizing*

Proses *tokenizing* ini akan memisahkan dalam suatu kalimat sehingga hanya menyisakan kata-kata yang dibutuhkan saja.

e. *Filtering*

Setelah melakukan *tokenizing*, maka tahap selanjutnya yaitu melakukan *filtering*. Proses ini dilakukan dengan pemilihan kata-kata yang mengandung makna yang didapat dari hasil *tokenizing*.

f. *Stemming*

Stemming akan dilakukan setelah melakukan *filtering* dimana tahap ini digunakan untuk membuang kata berimbuhan pada setiap kata disuatu kalimatnya. Sehingga dapat dikatakan juga untuk mendapatkan kata dasar dari setiap katanya.

9. *Clustering*

Data ulasan yang telah melalui tahap *pre-processing data*, kemudian dilakukan proses data ulasan. Adapun pada penelitian ini terbagi menjadi dua tahap yaitu:

a. Menentukan Jumlah K Menggunakan Metode *Elbow*

Pada penentuan jumlah *cluster* dilakukan dengan menggunakan metode *Elbow* sebagai metode bantu agar dapat mengetahui berapa jumlah *cluster* optimum atau yang terbaik dibentuk dari keseluruhan data yang diolah.

b. Memvisualisasikan *Cluster* Menggunakan *WordCloud*

Selanjutnya setelah didapatkan jumlah *cluster* yang terbaik atau optimum, kemudian akan dibentuk kelompok atau *cluster* yang sesuai dengan saran jumlah optimumnya. Dilanjutkan dengan membuat visualisasi data dari masing-masing *cluster* agar dapat dilakukan analisis lebih lanjut.

c. Memvisualisasikan Proporsi *Cluster*

Visualisasi proporsi *cluster* ini membantu untuk memahami lebih dalam terkait hasil *clustering*, mengidentifikasi topik utama, dan mengarahkan strategi pengembangan layanan.

10. Analisis dengan Metode SWOT

Selanjutnya setelah melakukan analisis hasil *clustering* dengan Analisis akan berupa karakter apa yang ada pada setiap kelompok, maka pada analisis dengan metode SWOT ini akan mengelompokkan lagi berdasarkan evaluasi kekuatan (*strengths*), kelemahan (*weaknesses*), peluang (*opportunities*), serta ancaman (*threats*) yang paling mempengaruhi aplikasi pada setiap *cluster*-nya. Kemudian juga diberikan analisis terkait masukan atau Langkah yang seharusnya dilakukan Perusahaan menanggapi hasil sentiment dari ulasan para pengguna dari aplikasi tersebut.

11. Kesimpulan dan Saran

Berdasarkan hasil analisis yang telah dilakukan, maka tahap terakhir pada penelitian ini yaitu dengan memberikan kesimpulan dan saran yang bisa digunakan bagi penelitian selanjutnya, Perusahaan terkait, dan para pembaca lainnya jika ingin melakukan pengembangan lebih lanjut maupun perbaikan.

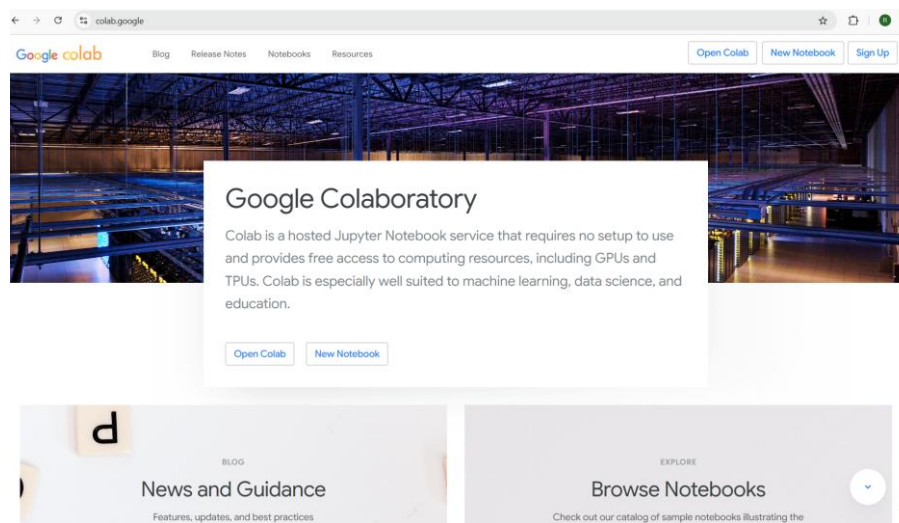
12. Selesai

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

Dalam pengumpulan data ini, data yang dijadikan sebagai bahan utama dalam penelitian ini merupakan data sekunder yang berasal dari ulasan yang diberikan oleh pengguna aplikasi MyPertamina melalui platform *Google Playstore*. Teknik pengambilan data pada penelitian ini yaitu dengan menggunakan *Scraping* melalui *web*. *Web* yang dituju yaitu *Google Playstore* yang fokus pada ulasan aplikasi MyPertamina. Terdapat beberapa *tools* yang bisa digunakan dalam melakukan *scraping data* dimana pada penelitian ini menggunakan *tools* untuk membantu dalam *scraping data* yaitu dengan menggunakan *Google Colaboratory*. *Google Colab* merupakan perangkat lunak berbasis *cloud* yang mirip dengan *Jupyter Notebook* gratis dan dijalankan melalui *browser* seperti *Google Chrome*. *Platform* ini memungkinkan pengguna untuk menjalankan kode pemrograman *Python* tanpa perlu melakukan instalasi atau pengaturan terlebih dahulu.



Gambar 4. 1 *Google Colaboratory*

Kelebihan dari *Google Colaboratory* sendiri yaitu dalam kemudahan aksesibilitas, kolaborasi *real-time*, dan integrasi layanan *Google* lainnya, seperti *Google Drive* dan *Google Docs*. Dan juga karena berbasis *cloud* dan dapat disambungkan dengan *Google Drive* sehingga untuk digunakan dalam pengolahan data yang banyak tidak akan memakan *space* untuk penyimpanannya. Adapun alamat domain yang digunakan untuk pengambilan data melalui

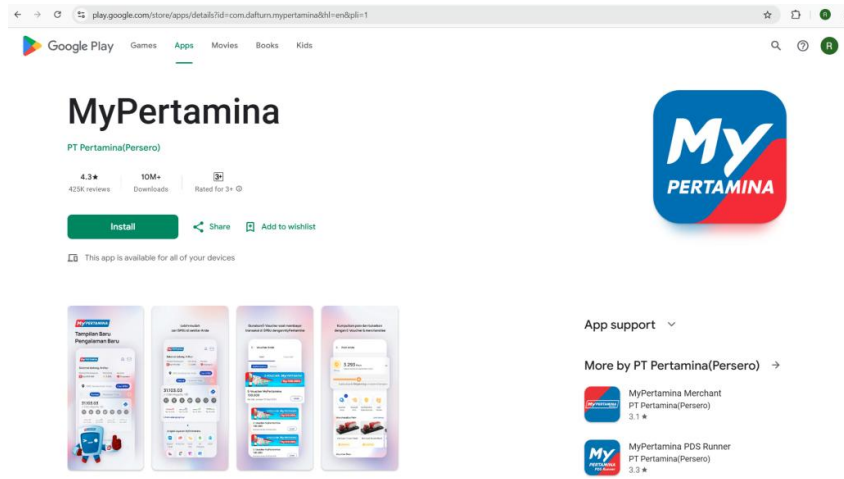
platform

Google

Playstore

yaitu

<https://play.google.com/store/apps/details?id=com.dafturn.mypertamina&hl=en&pli=1>.



Gambar 4. 2 Aplikasi MyPertamina pada Google Playstore

Selanjutnya masuk pada halaman awal *Google Colaboratory* untuk melakukan *scrapping data*. Kemudian menuliskan Bahasa pemrograman *Phyton* “**!pip install -qq google-play-scraper**” untuk meng-*install* agar dapat dilakukan proses *scrapping* pada data yang diambil. Serta *import* juga bahasa pemrograman *Phyton* **Json** yang berfungsi untuk menukar atau mengambil data informasi antara aplikasi atau system yang berbeda serta **Pandas as pd** yang berfungsi untuk membuat *data frame* dan juga membaca *file* csv. Langkah selanjutnya memasukkan bahasa pemrograman *Phyton* seperti pada Tabel 4.1 berikut:

Tabel 4. 1 Bahasa Pemrograman *Scrapping Data*

Phyton	Fungsi
<pre>from google.colab import drive drive.mount('/content/drive', force_remount=True) path = "gdrive/MyDrive/MyPertamina"</pre>	Untuk menyimpan dan mengakses hasil <i>scrapping</i> pada <i>file</i> di folder MyPertamina dari <i>Google Drive</i> langsung melalui <i>Google Colab</i> dengan mengacu pada <i>path</i> tersebut.
<pre>app_packages = ['com.dafturn.mypertamina']</pre>	Untuk menuju pada <i>link</i> aplikasi MyPertamina yang akan diambil ulasannya dari <i>Google Playstore</i> .

Phyton	Fungsi
<pre> app_infos = [] for application in tqdm (app_packages): info = app(application, lang='id', country='id') del info['comments'] app_infos.append(info) </pre>	<p>Untuk mengumpulkan, memahami, dan memantau perubahan data dari waktu ke waktu sehingga dapat dilakukan berbagai analisis. Dan juga memberikan detail dari kriteria data yang diambil seperti data menggunakan Bahasa Indonesia.</p>
<pre> app_reviews=[] for ap in tqdm(app_packages): for score in list(range(1,6)): for sort_order in [Sort.MOST_RELEVANT, Sort.NEWEST]: rvs, _ = reviews(ap, lang='id', country='id', sort=sort_order, count=2000 if score == 3 else 1000, filter_score_with=score) for r in rvs: r['sortOrder']='most_relevant' if sort_order ==Sort.MOST_RELEVANT else 'newest' r['appId'] = ap app_reviews.extend(rvs) app_reviews_df = pd.DataFrame(app_reviews) </pre>	<p>Untuk mengambil dan mengumpulkan hasil <i>review</i> ulasan aplikasi dengan mengambil ulasan dari paling lama hingga terbaru,</p>
<pre> app_reviews_df = app_reviews_df.sort_values(["appId", "at"]) </pre>	<p>Untuk mengubah data ulasan yang awalnya berbentuk <i>list of dictionaries</i> menjadi <i>data frame</i></p> <p>Untuk mengurutkan data ulasan berdasarkan variable tertentu, pada</p>

Phyton	Fungsi
<pre> from google.colab import drive drive.mount('/content/drive', force_remount=True) path = "/content/drive/MyDrive/MyPertamina/" app_reviews_df.to_csv(path+'HasilReviewsAPP.csv', index=None, header=True) import pandas as pd import matplotlib.pyplot as plt drive.mount('/content/drive', force_remount=True) path = "/content/drive/MyDrive/MyPertamina/" app_reviews_df = pd.read_csv(path + 'HasilReviewsAPP.csv') app_reviews_df['at'] = pd.to_datetime(app_reviews_df['at']) app_reviews_df['year'] = app_reviews_df['at'].dt.year reviews_per_year = app_reviews_df.groupby('year')['reviewId'].count() plt.figure(figsize=(10, 6)) plt.bar(reviews_per_year.index, reviews_per_year.values) plt.xlabel('Year') </pre>	<p>penelitian ini data ulasan diurutkan berdasarkan variable “appId” yaitu <i>package name</i> dan variable “at” yaitu menunjukkan tanggal pemberian ulasan.</p> <p>Untuk menghubungkan <i>Google Colab</i> dengan <i>Google Drive</i> dan mengakses <i>file</i> atau <i>folder</i> data secara langsung.</p> <p>Untuk menyimpan <i>Data Frame</i> ke dalam <i>file</i> csv di <i>Google Drive</i>.</p> <p>Untuk menganalisis dan memvisualisasikan jumlah ulasan aplikasi per tahun berdasarkan data ulasan yang disimpan di <i>file</i> csv.</p>

Pyhton	Fungsi
plt.ylabel('Number of Reviews')	
plt.title('Number of Reviews per Year')	
plt.xticks(reviews_per_year.index)	
plt.show()	

Dari Tabel 4.1 kemudian mendapatkan hasil *file* yang berbentuk csv yang dapat diakses pada *Google Drive* sehingga siap digunakan untuk Langkah selanjutnya yaitu pengolahan data. Hasil data tersebut, setelah diekstraksi dapat dilihat pada Gambar 4.3 berikut:

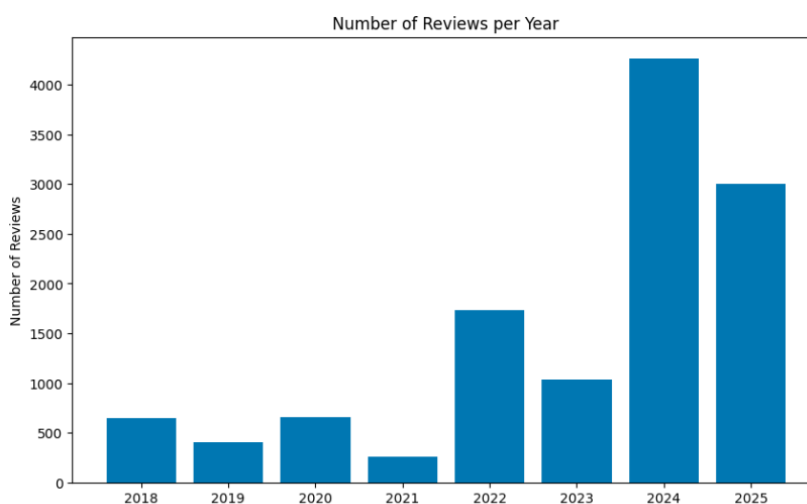
reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
33310a75-5741-4da4-b714-a21707803b38	Pengguna Google	https://play-lh.googleusercontent.com/Ini mau scan s		3	0	1.0.7.3	2018-09-12 9:02:28	Terima kasih at	2018-09-12 9:18:27
c03f3b60-0ed9-49c9-86a7-86e7f1a356e4	Pengguna Google	https://play-lh.googleusercontent.com/Error tolong pe		3	0	1.0.7.3	2018-09-12 12:44:18	Selamat malam	2018-09-12 15:09:34
cff1cd8b-afa7-41b1-839f-7512c229c2f3	Pengguna Google	https://play-lh.googleusercontent.com/Saya tidak dpt		3	0	1.0.7.3	2018-09-13 6:36:12	Hai sobat Hend	2018-09-13 7:59:27
9b1bb46d-423d-4e96-b487-2f321e81a2a7	Pengguna Google	https://play-lh.googleusercontent.com/Bagus buat Pei		4	3	1.0.7.3	2018-09-13 6:54:54	Terima kasih s	2018-09-13 12:11:04
439a1eff-47e0-4a25-a8eb-b64cf28c4d79	Pengguna Google	https://play-lh.googleusercontent.com/Bagus aplikasi		5	0	1.0.7.3	2018-09-13 10:32:09	Terima kasih s	2018-09-13 11:09:16
3aba4c73-719c-4d45-a93d-1073881fa793	Pengguna Google	https://play-lh.googleusercontent.com/Lama banget u		3	0	1.0.7.3	2018-09-13 12:41:54	Selamat malam	2018-09-13 14:52:23
66b7efdb-4420-4736-9067-ec2690033ed	Pengguna Google	https://play-lh.googleusercontent.com/Untuk aplikasi s		4	0	1.0.7.3	2018-09-13 12:43:32	Selamat malam	2018-09-13 14:54:12
4b44ed8d-a71e-4cd9-b5f4-748c1489090a	Pengguna Google	https://play-lh.googleusercontent.com/Dear admin, ur		5	0		2018-09-14 2:39:03	Hai sobat Ali K	2018-09-14 3:01:04
3a7b09a1-5679-4760-ab38-13c28d285982	Pengguna Google	https://play-lh.googleusercontent.com/Hmmmm, saya		3	0	1.0.7.3	2018-09-14 8:41:14	Hai sobat Feni	2018-09-14 9:32:36
02c7d093-c759-4618-85c9-428db729a755	Pengguna Google	https://play-lh.googleusercontent.com/Masih sering ei		4	0	1.0.7.3	2018-09-14 9:30:59	Hai sobat Budi	2019-07-21 3:15:24
a823b040-40bf-4487-8d2a-93ba948a5c54	Pengguna Google	https://play-lh.googleusercontent.com/MudahZan dgn		4	0	1.0.7.3	2018-09-14 9:57:08	Hai sobat Shan	2018-09-14 10:34:42
18f5364c-f53e-4646-ba9f-b6271c707b6	Pengguna Google	https://play-lh.googleusercontent.com/Gmana cara p		4	0	1.0.7.3	2018-09-15 7:57:59	Mohon maaf sc	2018-09-15 8:30:33
dc844c-37-9b99-4805-8f86-52c3a5ed06c3	Pengguna Google	https://play-lh.googleusercontent.com/Aplikasinya an		2	0	1.0.7.3	2018-09-15 9:52:25	Baik sobat Geb	2018-09-15 9:56:46
ee05b595-f9e7-4c44-bd92-10d731232eb8	Pengguna Google	https://play-lh.googleusercontent.com/Tiap mau upgr		3	0	1.0.7.3	2018-09-16 7:50:53	Selamat siang	2018-09-16 8:40:28
d5a937d9-35d4-4413-81c2-9502c3a67de2	Pengguna Google	https://play-lh.googleusercontent.com/Mohon diperba		3	0	1.0.7.3	2018-09-16 8:53:36	Selamat malam	2018-09-16 12:13:46
b749ce86-4e70-4103-830c-c00f163999d4	Pengguna Google	https://play-lh.googleusercontent.com/Kenapa ya say		2	0		2018-09-16 11:37:49	Selamat malam	2018-09-16 12:14:51
588c596c-e590-456d-a69a-a674e0a30bf9	Pengguna Google	https://play-lh.googleusercontent.com/Aplikasi keluar		3	0	1.0.7.3	2018-09-16 11:49:21	Selamat malam	2018-09-16 12:15:45

Gambar 4. 3 Hasil *Scrapping Data*

Hasil dari *scrapping data* didapatkan jumlah ulasan sebanyak 12.000 sejak bulan September 2018 hingga Januari 2025. Kemudian hasil *scrapping data* disaring menjadi hanya ulasan yang diberikan pada tanggal 1 Januari 2024 hingga 12 Januari 2025, sehingga data tersebut berkurang dan menjadi 6.984 ulasan yang ditunjukkan pada Gambar 4.4.

reviewId	userName	userImage	content	score	thumbsUpCo	reviewCreatedVersio	at
2f8964ba-33b9-4924-9d24-486cd54e0ea2	hermawan oktavia	https://play-lh.googleusercontent.com/Pelayanannya	Pelayanannya	4	20	4.2.3	2024-01-01 9:11:56
2f8964ba-33b9-4924-9d24-486cd54e0ea2	hermawan oktavia	https://play-lh.googleusercontent.com/Pelayanannya	Pelayanannya	4	20	4.2.3	2024-01-01 9:11:56
06ae5e15-aefc-48b7-9201-56bc539227dc	Chocolate	https://play-lh.googleusercontent.com/Ini gmn cara dc		4	0	4.1.1	2024-01-01 11:09:37
0e2ed9d8-ca85-4da5-8540-fcd8475626d3	Hartanto Meghalith	https://play-lh.googleusercontent.com/SPBU 344011C		2	1	4.2.3	2024-01-02 4:04:46
0e2ed9d8-ca85-4da5-8540-fcd8475626d3	Hartanto Meghalith	https://play-lh.googleusercontent.com/SPBU 344011C		2	1	4.2.3	2024-01-02 4:04:46
958af132-6717-45d5-a9ec-561cd7094375	Tri MuL	https://play-lh.googleusercontent.com/Setelah update		2	0	4.2.3	2024-01-02 22:28:40
d9cd434d-c880-40a3-8e42-bb5fa3eacd47	taufik repiawan	https://play-lh.googleusercontent.com/Sering terkend		4	0		2024-01-03 5:14:46
12d7fe48-6ebc-4920-a174-dd5e0683e818	Totok Kusumardiya	https://play-lh.googleusercontent.com/Poin ngumpulir		4	0	4.2.3	2024-01-03 7:19:12
a88ac694-645e-4b63-9c5c-04f50d59009a	Rismiadi	https://play-lh.googleusercontent.com/Aplikasinya ba		3	3	4.2.3	2024-01-03 11:24:28
a88ac694-645e-4b63-9c5c-04f50d59009a	Rismiadi	https://play-lh.googleusercontent.com/Aplikasinya ba		3	3	4.2.3	2024-01-03 11:24:28
3c3bf9ad-b3d4-4f7b-a95f-a328bae645bc	rizqy muhamad fird	https://play-lh.googleusercontent.com/Aplikasinya ba		4	2	4.2.3	2024-01-03 13:58:45
3c3bf9ad-b3d4-4f7b-a95f-a328bae645bc	rizqy muhamad fird	https://play-lh.googleusercontent.com/Aplikasinya ba		4	2	4.2.3	2024-01-03 13:58:45
a48484a1-ac1f-447f-8b73-320d67875956	Deddy Mendai Z	https://play-lh.googleusercontent.com/Ribet		3	0	4.2.3	2024-01-03 15:48:59
b562e0c0-bac3-4eef-b1bd-b9155c2cf81a	surisman risman	https://play-lh.googleusercontent.com/Ok		4	0	4.2.3	2024-01-04 0:57:32
b3d1810b-c99a-45cf-a906-733f0b796	Beni Sarumaha	https://play-lh.googleusercontent.com/aplikasi ini cras		3	0	4.2.3	2024-01-04 4:20:38
0a811fa3-a021-4928-9ea1-4b6a50213c59	Abd Rahman	https://play-lh.googleusercontent.com/Tidak diberikan		4	0		2024-01-04 6:53:56
79adb13f-b751-46a1-bd12-c127b533a663	Yuda Rendy Renak	https://play-lh.googleusercontent.com/Banyak SPBU		3	0	4.2.3	2024-01-04 9:52:49
04ccb24b-2651-4afa-8feb-289b7ac49b82	bambang nurdians	https://play-lh.googleusercontent.com/Semoga subsid		4	0	4.2.3	2024-01-05 11:36:01
905b9a75-635a-48ea-8893-7711eef5dc13	Moh Santoso	https://play-lh.googleusercontent.com/Kok gak bisa lil		4	0	4.2.3	2024-01-05 12:06:08
9bd422f3-44e1-475b-8a0f-a703bd34b6c0	Nita Anita	https://play-lh.googleusercontent.com/Sekarang mo d		2	0		2024-01-06 0:52:56
9bd422f3-44e1-475b-8a0f-a703bd34b6c0	Nita Anita	https://play-lh.googleusercontent.com/Sekarang mo d		2	0		2024-01-06 0:52:56
db4be8b6-a655-4f36-b134-9b76ff026e56	Balham Siagian	https://play-lh.googleusercontent.com/OK		4	0	4.2.3	2024-01-06 3:59:46
7dc053fa-c84b-4410-a945-ace5011be542	Agus Agus	https://play-lh.googleusercontent.com/Kesulitan cari s		2	0		2024-01-06 5:20:26

Pengambilan ulasan aplikasi hanya diambil pada tanggal 1 Januari 2024 hingga 12 Januari 2025 dikarenakan perubahan jumlah ulasan mulai meningkat cukup tinggi pada tahun 2024 dan 2025 dibandingkan tahun-tahun sebelumnya yang dapat dilihat pada Gambar 4.5. Hal ini dapat diperkirakan karena peningkatan pengguna aplikasi MyPertamina pada tahun 2024 dan juga disebabkan oleh berbagai faktor, seperti digitalisasi layanan Pertamina, program loyalitas yang menarik, dukungan terhadap UMKM, serta kemudahan transaksi yang terus dikembangkan. Dengan demikian, peningkatan jumlah ulasan dan penggunaannya agar PT. Pertamina tetap bisa mempertahankan para konsumennya maka dari itu perlu dilakukan analisis terkait sudah layak atau belum terkait aplikasi MyPertamina dalam menawarkan fitur-fiturnya.



Gambar 4. 5 Jumlah Ulasan Tiap Periode (Tahun)

4.2 Pengolahan Data

Setelah melakukan pengumpulan data, maka dilakukan pengolahan data. Pada pengolahan data ini terdapat akan dilakukan *pre-processing* terlebih dahulu, beberapa tahapan dalam *pre-processing* ini akan mengubah jumlah ulasan di awal agar sesuai dengan kebutuhan data yang sesuai dengan batasan penelitian yang dilakukan.

4.2.1 Pre-processing Data

Tahap ini akan melakukan pembersihan dan persiapan teks untuk menghilangkan data *noise* atau bagian tidak informatif yang tidak dibutuhkan. Dalam proses ini akan dilakukan dengan menggunakan bantuan *Tools Python* dan *Google Colaboratory*. Berikut akan dijelaskan terkait tahapan-tahapan dalam *pre-processng data*:

a. Cleaning Data

Pada *cleaning data* ini, dilakukan pemilihan variable yang akan difokuskan dalam penelitian ini. Variabel yang digunakan dalam penelitian ini yaitu “*at*” untuk melihat waktu ulasan ditambahkan dan “*content*” untuk melihat informasi ulasan apa yang dituliskan. Ulasan ini yang digunakan hanya ulasan yang diberikan pada tanggal 1 Januari 2024 hingga 12 Januari 2025. Berikut pada Gambar 4.6 merupakan hasil dari *cleaning data* serta untuk code *syntax* dapat dilihat pada bagian halaman lampiran:

	content	at
5018	Sangat membantu Tp kenapa sering minta diupded...	2024-02-09 12:17:55
5019	Sangat membantu Tp kenapa sering minta diupded...	2024-02-09 12:17:55
5020	MyPeetamina is the bes	2024-02-10 09:01:16
5021	Sesudah di update, jadi engga bisa scan barcod...	2024-02-10 23:28:39
5022	Sesudah di update, jadi engga bisa scan barcod...	2024-02-10 23:28:39
...	...	...
11995	susah daftar	2025-04-04 08:28:29
11996	makin sukses terus ya my pertamina	2025-04-04 09:24:07
11997	Bermanfaat	2025-04-04 10:53:10
11998	Alhamdulillah	2025-04-04 11:31:01
11999	sekelas aplikasi bumh yang dimiliki negara wak...	2025-04-04 11:47:24

Gambar 4. 6 Pemilihan Variabel Data *Cleaning*

Dari Gambar 4.6, pemilihan ulasan tersebut didapatkan data sebanyak 6982 data dari 12.001 data. Pada data ke-5018 – 5019 dan data ke-5021 – 5022 terdapat ulasan yang sama. Hal ini akan dilakukan penghapusan salah satu data agar menghilangkan *duplicate* data sehingga tidak terjadi data yang berulang dan semua ulasan dapat dianalisis secara akurat. Dengan hal ini dilakukan penghapusan data duplikasi yang dapat dilihat pda Gambar 4.7 dengan *syntax* yang digunakan.

```
data1 = data1.drop_duplicates(subset="content", inplace=False, keep = "last")
data1
```

	content	at	
5019	Sangat membantu Tp kenapa sering minta diupded...	2024-02-09 12:17:55	
5020	MyPeetamina is the bes	2024-02-10 09:01:16	
5022	Sesudah di update, jadi engga bisa scan barcod...	2024-02-10 23:28:39	
5023	Cukup membantu .	2024-02-11 10:59:16	
5024	O. 🍷 🍷 🍷 🍷 🍷	2024-02-12 08:40:35	
...	...	...	
11995	susah daftar	2025-04-04 08:28:29	
11996	makin sukses terus ya my pertamina	2025-04-04 09:24:07	
11997	Bermanfaat	2025-04-04 10:53:10	
11998	Alhamdulillah	2025-04-04 11:31:01	
11999	sekelas aplikasi burn yang dimiliki negara wak...	2025-04-04 11:47:24	

4650 rows × 2 columns

Gambar 4. 7 Penghapusan *Duplicate Data*

Dapat dilihat dari Gambar 4.7 bahwa setelah dilakukan penghapusan *duplicate data*, jumlah data yang sebelumnya 6892 data menjadi 4650 data.

b. *Case Folding* dan *Tokenizing*

Pada tahap *case folding* akan melakukan pengubahan huruf kapital menjadi huruf non-kapital atau menjadi huruf kecil dengan menggunakan bahasa pemrograman. Sedangkan untuk *tokenizing* akan membuat setiap kata yang di dalam kalimat menjadi sekumpulan kata tiap token. Berikut dapat dilihat pada Tabel 4.2 untuk membandingkan kata sebelum dan sesudah dilakukan *case folding* dan *tokenizing* dengan bahasa pemrograman, serta untuk code *syntax* yang digunakan dapat dilihat pada halaman lampiran.

Tabel 4. 2 Perbedaan Data Setelah Tahap *Case Folding* dan *Tokenizing*

<i>Input</i>	<i>Output</i>
'Sangat membantu Tp kenapa sering minta diupded Setelah diupded mau melihat kode QR malah susah dibuka'	'sangat membantu tp kenapa sering minta diupded setelah diupded mau melihat kode qr malah susah dibuka'
'MyPeetamina is the bes'	'mypeetamina is the bes'

<i>Input</i>	<i>Output</i>
'Sesudah di update, jadi engga bisa scan barcode nya'	'sesudah di update jadi engga bisa scan barcode nya'

c. *Filtering*

Setelah data dilakukan tahap *case folding* dan *tokenizing*, maka selanjutnya dilakukan tahap *filtering*. Proses yang dilakukan pada tahapan ini yaitu melakukan pemilihan kata-kata yang mengandung makna yang didapat dari tahapan sebelumnya. Tahap *filtering* juga disebut sebagai *stopword removal*. Kata-kata yang akan dihilangkan dalam tahapan *filtering* ini seperti kata 'yang', 'dan', 'untuk', 'kamu', 'sampai', dan sebagainya. Hasil perbandingan sebelum dan sesudah yang dilakukan pada tahapan ini dapat dilihat pada Tabel 4.3, serta untuk code *syntax* untuk menampilkan *output* tersebut dapat dilihat pada halaman lampiran.

Tabel 4. 3 Perbedaan Data Setelah Tahap *Filerting*

<i>Input</i>	<i>Output</i>
'sangat membantu tp kenapa sering minta diupded setelah diupded mau melihat kode qr malah susah dibuka'	'membantu tp diupded diupded kode qr susah dibuka'
'mypeetamina is the bes'	'mypeetamina is the bes'
'sesudah di update jadi engga bisa scan barcode nya'	'update engga scan barcode nya'

d. *Stemming*

Pada tahap keempat *pre-processing* ini yaitu *Stemming*. *Stemming* dilakukan agar mengubah kata yang berimbuhan menjadi akar kata atau kata dasar. Proses ini dibutuhkan agar kata-kata dalam ulasan menjadi lebih sederhana sehingga mempermudah dalam melakukan *clustering*. Penelitian ini menggunakan stemmer Sastrawi dimana *library* ini menyediakan *stemming* kata bahasa Indonesia. Berikut hhasil perbandingan sebelum dan sesudah yang dilakukan pada tahapan ini dapat dilihat pada Tabel 4.4, serta untuk code *syntax* yang digunakan dapat dilihat pada bagian halaman lampiran.

Tabel 4. 4 Perbedaan Data Setelah Tahap *Stemming*

<i>Input</i>	<i>Output</i>
'membantu tp diupded diupded kode qr susah di buka'	'bantu tp diupded diupded kode qr susah buka'
'tolong donk team yg ngurus spbu indonesia mewajibkan spbu menyediakan fasilitas pembayaran nont tunai qris dll '	'donk team yg ngurus spbu indonesia wajib spbu sedia fasilitas bayar nont tunai qris'
'kak tolong perbaiki mslah pembayaran kak ngisi saldonya tambahin apk dana kak biar pakek dana mbayarnya kak'	'kak baik mslah bayar kak ngisi saldo tambahin apk dana kak biar pakek dana mbayarnya kak'

e. *Checking Data Null*

Dalam *preprocessing data*, dilakukannya pemeriksaan data terhadap *data null* sangat penting karena data yang kosong dapat menyebabkan hasil analisis menjadi tidak akurat dan tidak memberikan informasi apa pun serta hanya menjadi gangguan (*noise*) dalam proses analisis. Sebelumnya jumlah data ulasan yang didapatkan yaitu sebanyak 4650, kemudian dilakukan pengecekan ada tidaknya data kosong atau yang tidak memiliki makna atau pun tidak terbaca untuk dianalisis. Sehingga didapatkan data baru sebanyak 4595 data yang bersih dan disimpulkan terdapat 54 *data null*. Untuk memastikan pengecekan apakah masih terdapat data kosong digunakan bahasa pemrograman yaitu `text_baru.isnull().sum()`

f. *Clustering dengan Algoritma K-Means Clustering*

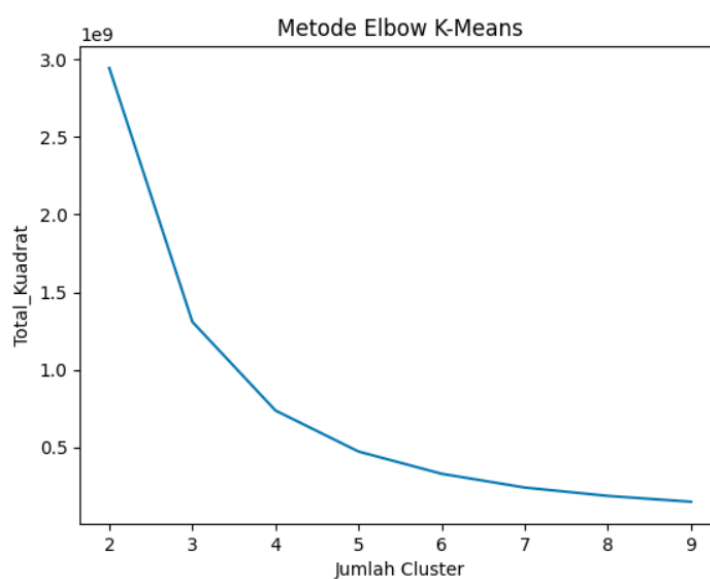
Setelah melakukan tahap *preprocessing data*, langkah selanjutnya melakukan *clustering* dengan menggunakan algoritma *K-Means Clustering* dan bantuan metode *Elbow* untuk menemukan jumlah *cluster* yang optimal. Proses *clustering* ini dibantu menggunakan *Google Colaboratory* dengan bahasa pemrograman yang digunakan yaitu Python. Dalam pemrograman ini jika masih terdapat data yang berbentuk teks atau kalimat maka tidak dapat dilakukan menggunakan metode *Elbow*. Maka dengan ini digunakannya TD-IDF (*Term Frequency & Inverse Document Frequency*) yang berfungsi sebagai pemberian bobot pada kata-kata dalam ulasan pengguna agar dapat diidentifikasi kata-kata yang paling relevan dan mewakili masalah atau keunggulan dalam layanan aplikasi tersebut. Adapun berikut dapat dilihat pada Gambar 4.8 Yang merupakan hasil dari pembobotan pada setiap ulasannya.

```
pre_clustering = pd.DataFrame(count)
pre_clustering
```

0	
0	513
1	4878
2	1331
3	2410
4	3913
...	...
5204	4813
5205	19
5206	1059
5207	4407
5208	5185

Gambar 4. 8 Hasil Pembobotan Setiap Ulasan

Kemudian pembobotan ini dilakukan untuk menentukan jumlah *cluster* yang paling optimal dengan menggunakan bantuan metode *Elbow*. Sehingga untuk analisis awal pada tahap ini yaitu dengan memunculkan dan melihat bentuk grafik dari pembentukan 2 *cluster* hingga 10 *cluster* dimana dalam penentuan jumlah *cluster* yang terbaik dilihat dari pembentukan siku dalam grafik yang dapat dilihat pada Gambar 4.9

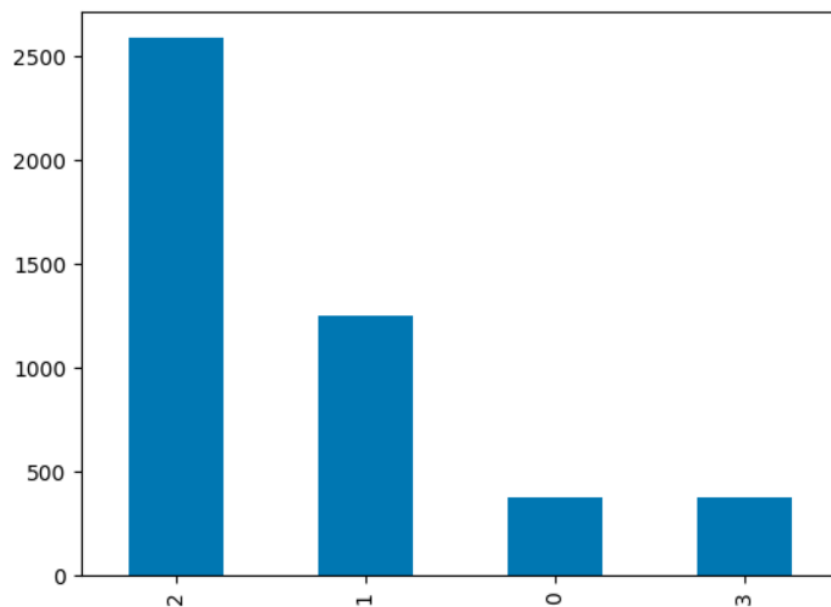


Gambar 4. 9 Grafik Metode *Elbow*

Berdasarkan Gambar 4.9, dapat dilihat hasil perhitungan menggunakan metode *Elbow* dengan metrik *Within Cluster Sum of Squares* (WCSS) bahwa nilai $K = 4$ yang merupakan titik siku yang terdapat pada grafik *Elbow*. Sehingga dapat disimpulkan nilai *cluster* optimal ketika berjumlah 4 dan tahap-tahap selanjutnya akan membentuk *cluster* menjadi 4 kelompok menggunakan algoritma *K-Means Clustering*.

g. *Clustering*

Tahap selanjutnya yaitu melakukan proses *clustering* dengan membentuk 4 kelompok dari hasil bentuk siku pada *cluster* 4. Sehingga nantinya nilai *cluster* optimal Ketika berjumlah 4. Proses *clustering* ini dilakukan dengan bahasa pemograman yaitu Phyton dan hasil *clustering* data ulasan aplikasi MyPertamina dalam bentuk grafik yang dapat dilihat pada Gambar 4.10 dan untuk melihat jumlah data setiap *cluster* dapat dilihat pada Gambar 4.11



Gambar 4. 10 Grafik *Clustering* Data Ulasan

Text_Bersih	
cluster	
0	378
1	1254
2	2585
3	378

Gambar 4. 11 Jumlah Data Setiap *Cluster*

Berdasarkan dari gambar 4.11 dapat dilihat bahwa *cluster* data ulasan yang terbentuk sebanyak 4 *cluster* dengan jumlah ulasan pada tiap *cluster*-nya yaitu *cluster* 1 memiliki jumlah ulasan sebanyak 378, *cluster* 2 memiliki jumlah ulasan sebanyak 1254, *cluster* 3 memiliki jumlah ulsan sebanyak 2585, dan untuk *cluster* 4 memiliki jumlah ulasan sebanyak 378.

Dari hasil pengelompokkan yang telah terbentuk tersebut, langkah selanjutnya akan dilakukan visualisasi tiap *cluster* dengan menggunakan *Word Cloud*. *Word Cloud* ini bertujuan untuk menampilkan informasi-informasi yang berupa kata-kata yang banyak dibicarakan dalam ulasan yang diberikan oleh pengguna aplikasi MyPertamina. Dan juga *word cloud* dapat membantu dalam membandingkan dan membedakan antar satu *cluster* dengan *cluster* yang lainnya. Frekuensi kata dalam keseluruhan data akan mempengaruhi besar atau kecilnya kata tersebut pada *word cloud*. Semakin tinggi frekuensi kata maka ukuran kata tersebut pada *word cloud* akan semakin besar dan begitu juga sebaliknya (Asri & Wahyu, 2021).

Berdasarkan Gambar 5.1 dapat dilihat bahwa hasil analisis menggunakan *word cloud* pada *cluster* 1 memperlihatkan beberapa topik-topik yang sering dibicarakan oleh pengguna aplikasi MyPertamina. *Cluster* 1 memiliki jumlah data ulasan sebanyak 378 ulasan dari 4595 ulasan secara keseluruhan. Beberapa ulasan yang sering dibicarakan pada *cluster* 1 yaitu mengenai ‘aplikasi’, ‘ kirim’, ‘kode’, ‘otp’, ‘masuk’, ‘coba’, ‘ulang’, ‘ribet’, dan lain-lain. Kata-kata yang sering muncul sangat beragam sehingga memiliki sifat cenderung positif, netral, maupun negatif.

b. Perhitungan Frekuensi dan Pembobotan Kata *Cluster* 1

Dari hasil keseluruhan kata-kata yang sering dibicarakan pada *cluster* 1 maka diperoleh hasil analisis frekuensi 20 kata yang paling sering muncul yaitu dilihat dari Tabel 5.1

Tabel 5. 1 Frekuensi dan Pembobotan Kata *Cluster* 1

Kata Sering Muncul	Frekuensi	Bobot	Total Bobot
kode	289	0	0
otp	243	0	0
qr	187	0	0
kirim	138	0	0
aplikasi	132	0	0
masuk	105	0	0
daftar	93	0	0
susah	81	-1	-81
aja	68	0	0
sms	64	0	0
verifikasi	61	0	0
code	57	0	0
wa	49	0	0
isi	48	0	0
muncul	46	0	0
coba	42	0	0
baik	39	+1	+39
ribet	38	-1	-38
sulit	37	-1	-37
kali	37	0	0

Kata Sering Muncul	Frekuensi	Bobot	Total Bobot
Total			-117

Hasil dari pembobotan kata yang didapatkan pada *cluster* 1 ini mendapatkan nilai total sebesar -117 sehingga dapat disebutkan bahwa nilai yang didapatkan cenderung negatif. Adapun beberapa kata-kata yang memiliki nilai positif yaitu ‘baik’ dengan jumlah kata yang dimiliki sebanyak 39. Untuk kata-kata yang memiliki nilai negatif diantaranya yaitu ‘susah’ dengan jumlah kata 81, ‘ribet’ dengan jumlah kata 38, dan ‘sulit’ dengan jumlah kata 37. Sedangkan untuk kata-kata yang memiliki nilai yang tidak berbobot positif maupun negatif atau yang biasa disebut netral antara lain ‘kode’, ‘otp’, ‘qr’, ‘ kirim’, ‘aplikasi’, ‘masuk’, ‘daftar’, ‘aja’, ‘sms’, ‘verifikasi’, ‘code’, ‘wa’, ‘isi’, ‘muncul’, ‘coba’, dan ‘kali’. Dari keluhan ulasan yang dibicarakan oleh pengguna di *cluster* 1 dapat disimpulkan bahwa keluhan tersebut mengenai kerumitan dalam proses login dan verifikasi dan juga saran dalam melakukan pemberian kode OTP yang dikirimkan melalui *whatsapp*.

c. Perhitungan Frekuensi dan Pembobotan Kata Gabungan *Cluster* 1

Analisis pembobotan berdasarkan gabungan dua kata dapat menghasilkan nilai total yang cenderung lebih negatif jika dibandingkan dengan pembobotan kata tunggal. Hal ini mengindikasikan bahwa meskipun analisis diawal telah mencerminkan isi *cluster* 1, tetapi tetap diperlukan validasi tambahan melalui pendekatan gabungan kata untuk memperoleh pemahaman yang lebih tepat terhadap makna sebelumnya. Sebagai contoh, pada analisis awal kata ‘baik’ memiliki bobot positif. Namun setelah dilakukan analisis lebih lanjut pada kombinasi gabungan dua kata, makna tersebut berubah menjadi negatif. Pada Tabel 5.2 akan ditampilkan beberapa kata dari gabungan dua kata sebagai validasi lebih lanjut.

Tabel 5. 2 Frekuensi dan Pembobotan Kata Gabungan Dua Kata *Cluster*1

No	Gabungan Kata		Frekuensi	Bobot	Total Bobot
1.	kode	otp	139	-1	-139
2.	kode	qr	75	-1	-75
3.	qr	code	51	-1	-51
4.	wa	sms	28	-1	-28
5.	otp	kirim	27	-1	-27
6.	kirim	kode	22	-1	-22
7.	kirim	otp	21	-1	-21
8.	otp	masuk	20	-1	-20

No	Gabungan Kata		Frekuensi	Bobot	Total Bobot
9.	kode	verifikasi	19	-1	-19
10.	kirim	kirim	16	-1	-16
Total					-418

Penilaian pada pembobotan kata dari gabungan kata didapatkan berdasarkan identifikasi kalimat ulasan yang terkait pada *cluster* tersebut. Beberapa kalimat ulasan secara utuh dari masing-masing gabungan kata yaitu sebagai berikut:

- 1) Hadehhh susah daftarnya **kode otp** ga masuk" boss
- 2) Sangat membantu Tp kenapa sering minta diupded Setelah diupded mau melihat **kode QR** malah susah dibuka
- 3) Gak bisa buka **qr code** setelah update, error mulu...
- 4) Tolong kirim VERIVIKASI LEWAT **WA**. Jangan **sms**, karena tidak bisa.
- 5) udh verifikasi dibilang gagal, udh beberapa kali kode **otp dikirim** ke gmail, gak bisa, nih aplikasi gmn sih ? tolong perbaiki, ini malah bikin ribet
- 6) Saya pengguna baru, mau daftarkan my pertamina, kok ga bisa ya..diperbaiki dulu, soal **kirim Kode OTP**
- 7) Perbaiki **pengiriman OTP** nya lama sekali, sudah cek email tidak ada, padahal waktu pengisian hanya 3 menit, sudah berulang jadi kadaluarsa OTP nya !!
- 8) Masukin kode **otp** padahal sudah sama dengan yg dikirim di email, tapi kesalahan terus dan ga bisa **masuk**, sistemnya di perbaiki, saya lengkapi lagi, subsidi tepat ketika login bermasalah harus di pencet beberapa kali baru masuk kmudian masukkan kode otp, sudah di masukkan sesuai yg dikirim di email tapi muncul notif kesalahan terus dan coba lagi nanti
- 9) Kenapa akhir2 ini mau masuk ke subsidi pertalite susah ya, nunggu **kode verifikasi** gak kunjung dikirim
- 10) Saya tidak bisa masuk aplikasi minta no verifikasi di wa GK **dikirim kirim**

5.1.2 Cluster 2

a. Visualisasi *Word Cloud Clutser 2*

Selanjutnya dilakukan visualisasi seperti pada *cluster 1* sebelumnya, hal yang sama akan dilakukan pada *cluster 2* ini dalam menampilkan informasi berupa kata-kata yang banyak dibicarakan oleh pengguna mengenai MyPertamina. Berikut dapat dilihat pada Gambar 5.1 terkait *word cloud* yang dihasilkan dari visualisasi *cluster 2*:

Kata Sering Muncul	Frekuensi	Bobot	Total Bobot
salah	213	-1	-213
aja	199	0	0
susah	191	-1	-191
isi	191	0	0
masuk	164	0	0
sistem	163	0	0
bikin	149	0	0
coba	146	0	0
kali	145	0	0
gagal	141	-1	-141
ribet	139	-1	-139
ulang	139	-1	-139
verifikasi	127	0	0
Total			-823

Hasil dari pembobotan kata yang didapatkan pada *cluster 2* ini mendapatkan nilai total sebesar -823 sehingga dapat disebutkan bahwa nilai yang didapatkan cenderung negatif. Pada *cluster 2* tidak didapatkan satupun kata-kata yang memiliki nilai positif. Adapun beberapa kata-kata yang memiliki nilai negatif yaitu ‘sulit’ dengan jumlah kata yang dimiliki sebanyak 230, ‘salah’ dengan jumlah kata yang dimiliki sebanyak 213, ‘susah’ dengan jumlah kata yang dimiliki sebanyak 191, ‘gagal’ dengan jumlah kata yang dimiliki sebanyak 141, ‘ribet’ dengan jumlah kata yang dimiliki sebanyak 139, dan ‘ulang’ dengan jumlah kata yang dimiliki sebanyak 139. Sedangkan untuk kata-kata yang memiliki nilai tidak berbobot positif maupun negative atau yang biasa disebut netral antara lain ‘aplikasi’, ‘daftar’, ‘data’, ‘barcode’, ‘subsidi’, ‘kendara’, ‘aja’, ‘isi’, ‘masuk’, ‘sistem’, ‘bikin’, ‘coba’, ‘kali’, dan ‘verifikasi’. Dari keluhan ulasan yang dibicarakan oleh pengguna di *cluster 2* dapat disimpulkan bahwa keluhan tersebut cenderung ribetnya cara sistem daftarnya dan barcode tidak sesuai dengan data kendaraan atau akun mereka, yang membuat proses verifikasi di SPBU ditolak.

c. Perhitungan Frekuensi dan Pembobotan Kata Gabungan *Cluster 2*

Dari hasil pembobotan pada frekuensi gabungan dua kata yang paling banyak menunjukkan total bobot yang negatif dimana didapatkan selisih perbandingan pada pembobotan kata

tunggal sebesar 489. Hal ini dapat menunjukkan bahwa hasil analisis pembobotan sebelumnya telah mewakili data dari *cluster* 2 yang dapat dilihat pada Tabel 5.4

Tabel 5. 4 Frekuensi dan Pembobotan Kata Gabungan Dua Kata *Cluster* 2

No	Gabungan Kata		Frekuensi	Bobot	Total Bobot
1.	daftar	subsidi	74	-1	-74
2.	salah	sistem	70	-1	-70
3.	data	kendara	54	-1	-54
4.	aplikasi	daftar	44	-1	-44
5.	daftar	aplikasi	44	-1	-44
6.	kendara	daftar	42	-1	-42
7.	aplikasi	aplikasi	41	-1	-41
8.	bikin	aplikasi	40	-1	-40
9.	isi	bbm	38	1	38
10.	isi	data	37	1	37
Total					-334

Penilaian pada pembobotan kata dari gabungan kata didapatkan berdasarkan identifikasi kalimat ulasan yang terkait pada *cluster* tersebut. Beberapa kalimat ulasan secara utuh dari masing-masing gabungan kata yaitu sebagai berikut:

- 1) Aplikasinya ada bug, saya baru daftar udah sesuai semua mau login dikatakan salah pin atau nomer. Harus nunggu 30 menit karna batas masukkan pin salah tapi sama saja habis 30 menit saya coba lagi. Tolong di perbaiki lagi untuk kesiapan aplikasinya, disuruh **daftar** buat **subsidi** tapi aplikasinya saja belum optimal
- 2) Aplikasi gak wajar Daftar Subsidi tepat formulir benar semua jawaban nya ada kesalahan di formulir pas di coba lagi **kesalahan sistem** di coba di aplikasi sama di website sama aja. Kalo gak niat bikin aplikasi jangan ngeribetin gw ribet beli minyak harus pake aplikasi my Pertamina sedangkan aplikasinya aja gak bisa apa apa,pikirin dulu sebelum buat aplikasi jangan ngeribetin!!!
- 3) daftar subsidi tepat semakin di persulit saja,data kendaraan di tolak segera perbaiki **data kendaraan** tapi giliran perbaiki data dan mau kirim data hanya loding saja tidak mau terkirim selalu di ulang²...tolong dong segera di evaluasi jangan menyulitkan rakyat kecil

- 4) **Aplikasinya** ada bug, saya baru **daftar** udah sesuai semua mau login dikatakan salah pin atau nomer. Harus nunggu 30 menit karna batas masukkan pin salah tapi sama saja habis 30 menit saya coba lagi. Tolong di perbaiki lagi untuk kesiapan aplikasinya, disuruh daftar buat subsidi tapi aplikasinya saja belum optimal
- 5) Daftar ga bisa" aneh , tulisannya terjadi kesalah dalam pengisian form , padahal ga ada yg salah, nyuruh doang **daftar** tapi **aplikasi** masih gembel , kenapa si kalo aplikasi pemerintah atau web dan situs pemerintah itu selalu ga beres
- 6) Daftar subsidi tp gabisa di buka, skalinya bisa lagi isi form udh foto stnk dan foto unit tiba2 balik ke beranda, mau daftar lg no **kendaraan** sudah ter**daftar** ga bisa daftar lagi, gimana dong ?
- 7) Minta verifikasi email Tapi pas ngecek email dan mau balik ke **aplikasi**. **Aplikasinya** malah nutup sendiri. Ulang ke awal Jadi butuh 2 device untuk verifikasi doang Padahal. dan setiap nge klik di aplikasinya selalu pemberitahuan ga jelas dan menyuruh melengkapi Padahal data sdh lengkap Mohon ditingkatkan lagi kalau mau diwajibkan Jangan diwajibkan dulu tapi aplikasinya tidak siap
- 8) Aplikasi bodoh, klo ga bisa **bikin aplikasi** jngn bikin daftar aja susah
- 9) Aplikasi mudah di gunakan ..masukan nya tolong dong di mutasi atau di aktifitas harus ada pemberitahuan setiap kita **isi BBM** jadi jika mobil di pinjam kita bisa ta di isi apa gk nya BBM nya...
- 10) Sedikit repot di awal, **isi data** app dan top up saldo. setelah nya jd lebih praktis dan tdk lama ikut antrian, krn punya jalur khusus. boleh lah cukup di apresiasi bumnn nya dgn aplikasi nya.

5.1.3 Cluster 3

a. Visualisasi *Word Cloud Cluster 3*

Berikut merupakan hasil visualisasi *word cloud* pada *cluster 3* dimana pada *cluster* ini pembicaraan yang banyak diberikan oleh pengguna yaitu terkait ‘bayar’. Adapun visualisasi dapat dilihat pada Gambar 5.3 terkait *WordCloud* yang dihasilkan dari *cluster 3*:

Kata Sering Muncul	Frekuensi	Bobot	Total Bobot
baik	102	1	102
masuk	90	0	0
mobil	84	0	0
cepat	79	1	79
coba	77	0	0
kasih	77	1	77
mohon	74	0	0
Total			286

Berdasarkan hasil dari pembobotan kata yang didapatkan pada *cluster* 3 pada Tabel 5.5 mendapatkan nilai total 286 sehingga dapat dikatakan bahwa nilai cenderung positif. Adapun beberapa kata-kata yang memiliki nilai positif yaitu ‘bantu’ dengan jumlah kata yang dimiliki sebanyak 168, ‘bagus’ dengan jumlah kata yang dimiliki sebanyak 144, ‘baik’ sebanyak 102, ‘cepat’ sebanyak 79, dan ‘kasih’ sebanyak 77. Meskipun sentiment negatif lebih dominan, ulasan dari pengguna juga menunjukkan adanya pengakuan positif terhadap beberapa aspek layanan aplikasi MyPertamina. Akan tetapi, penting juga untuk dicermati kembali terkait konteks kalimat lengkap pada ulasan yang diberikan. Karena beberapa kata seperti ‘baik’ bisa bersifat ambigu (misalnya dalam kalimat yang diberikan yaitu ‘mohon diperbaiki’).

c. Perhitungan Frekuensi dan Pembobotan Kata Gabungan *Cluster* 3

Dari hasil pembobotan pada frekuensi gabungan dua kata yang paling banyak menunjukkan total bobot 108 dimana hasil total ini cenderung negatif jika dibandingkan dengan total bobot pada pembobotan kata tunggal. Hal ini dikarenakan nilai total dilihat berdasarkan kalimat ulasan secara lengkap sehingga penilaian tersebut dapat dilihat pada Tabel 5.6

Tabel 5. 6 Frekuensi dan Pembobotan Kata Gabungan Dua Kata *Cluster* 3

No	Gabungan Kata		Frekuensi	Bobot	Total Bobot
1.	isi	bbm	58	-1	-112
2.	metode	bayar	57	-1	-50
3.	aplikasi	bagus	26	1	-20
4.	bayar	pakai	22	1	10
5.	bayar	aplikasi	19	-1	-9
6.	link	aja	18	-1	-9

No	Gabungan Kata		Frekuensi	Bobot	Total Bobot
7.	non	tunai	17	-1	-9
8.	mohon	baik	15	-1	8
9.	kartu	kredit	15	1	8
10.	mohon	bantu	13	1	-7
Total					-108

Penilaian pada pembobotan kata dari gabungan kata ini, didapatkan berdasarkan identifikasi kalimat ulasan yang terkait pada *cluster* tersebut. Adapun berikut merupakan kalimat ulasan secara utuh dari masing-masing gabungan kata yaitu sebagai berikut:

- 1) kuota bbm yg beli tampil aplikasi milik user milik tugas spbu mobil yg gak ngisi **bbm** beli kuota habis ulang oknum tugas spbu mohon baik keamanannnya Tingkat
- 2) Untuk **metode** pembayaran e-money nya kurang lengkap tambahin dana min
- 3) Ini **aplikasi** sebenarnya sudah **bagus**, hanya saja masih banyak SPBU yg tidak menerapkan kebijakan pemerintah, dan masih banyak pengguna AVANZA, INNOVA, Bahkan Fortuner sekaligus masih banyak yg mengisi BBM bersubsidi. Tolong dong diterapkan dengan benar peraturan pemerintah, tindak SPBU yg masih banyak mengacuhkan kebijakan atau memberikan pelayanan kepada selain penerima subsidi.
- 4) MyPertamina ini sangat bermanfaat seperti aplikasi pembayaran biasanya, seperti saya pengguna yang selalu jarang bawa uang tunai/cashless. Tidak perlu turun dari kendaraan untuk **bayar pakai** debit. Terima kasih.
- 5) Tambahkan metode pembayaran **aplikasi** dana lah
- 6) Tlg diperbaiki masalah pembayaran menggunakan kartu BRI. Krn selalu gagal terus transaksinya. Hanya bs menggunakan pembayaran dgn **link aja**
- 7) Banyak spbu yang tidak melayani pembayaran **non tunai**, seperti hari ini minggu, 3 maret 2024 pukul 18.15 di kota palembang tepat nya di spbu hegar alang-alang lebar, antrian pembayaran non tunai di buka tetapi petugas spbu tidak melayani pembayaran non tunai, padahal sudah menunggu antrian sangat lama
- 8) Tidak bisa download barcode. Bugnya terlalu banyak min. **Mohon** diperbaiki. Terimakasih
- 9) Terimakasih, aplikasinya sudah bagus. Tetapi mengapa pembayaran **melalui kartu kredit** sering sekali gagal?

Kata Sering Muncul	Frekuensi	Bobot	Total Bobot
mudah	156	+1	156
aplikasi	122	0	0
upload	113	0	0
daftar	91	0	0
ambil	80	0	0
galeri	80	0	0
kamera	78	0	0
langsung	56	0	0
susah	56	-1	-56
tolak	48	-1	-48
gagal	47	-1	-47
baik	45	1	45
ktp	46	0	0
sulit	40	-1	-40
bagus	29	+1	29
hasil	33	0	0
bikin	23	0	0
bantu	22	+1	22
Total			61

Berdasarkan hasil dari pembobotan kata yang didapatkan pada *cluster* 4 pada Tabel 5.7, mendapatkan hasil yang cenderung positif yaitu dengan nilai total sebesar 61. Adapun beberapa kata-kata yang memiliki nilai positif yaitu ‘mudah’ dengan jumlah kata yang dimiliki sebanyak 156, ‘bagus’ dengan jumlah kata yang dimiliki sebanyak 29, ‘bantu’ dengan jumlah kata yang dimiliki sebanyak 22. Meskipun sentimen negatif lebih dominan, ulasan dari pengguna juga menunjukkan adanya pengakuan positif terhadap beberapa aspek layanan MyPertamina. Hal itu dapat disimpulkan bahwa pada *cluster* 4 ini pengguna membicarakan terkait kepuasan terhadap aplikasi MyPertamina bahwa aplikasi tersebut memiliki kesan baik dan bagus pada fitur tertentu, rasa terbantu dari layanan yang diberikan untuk pengguna serta dapat mudahnya dipahami oleh Sebagian pengguna. Namun, penting juga untuk mencermati kembali terkait konteks kalimat lengkap pada ulasan yang diberikan

karena beberapa kata seperti ‘mudah’ bisa bersifat ambigu (misalnya, dalam kalimat yang diberikan yaitu ‘tidak mudah’).

c. Perhitungan Frekuensi dan Pembobotan Kata Gabungan *Cluster 4*

Dari hasil pembobotan pada frekuensi gabungan dua kata yang paling banyak menunjukkan total bobot 190 dimana hasil total ini cenderung negatif jika dibandingkan total bobot pada pembobotan kata tunggal. Hal ini dikarenakan nilai total bobot dilihat berdasarkan kalimat ulasan secara lengkap sehingga penilaian tersebut dapat dilihat pada Tabel 5.8

Tabel 5. 8 Frekuensi dan Pembobotan Kata Gabungan Dua Kata *Cluster 4*

No	Gabungan Kata		Frekuensi	Bobot	Total Bobot
1.	foto	stnk	112	-1	-112
2.	upload	foto	50	-1	-50
3.	ambil	galeri	20	-1	-20
4.	mudah	transaksi	10	1	10
5.	mudah	cepat	9	-1	-9
6.	mudah	aplikasi	9	-1	-9
7.	mudah	daftar	9	-1	-9
8.	aplikasi	mudah	8	1	8
9.	cepat	mudah	8	1	8
10.	susah	daftar	7	-1	-7
Total					-190

Penilaian pada pembobotan kata dari gabungan kata ini, didapatkan berdasarkan identifikasi kalimat ulasan yang terkait pada *cluster* tersebut. Adapun berikut merupakan kalimat ulasan secara utuh dari masing-masing gabungan kata yaitu seagai berikut:

- 1) Bisa di bantu gak min, akun saya kok verifikasi terus, mau daftar nopol kendaraan di tolak terus, alasannya foto kendaraan tidak sesuai, sudah di ganti dg sesuai saran pengambilan foto, malah **foto stnk** tidak sesuai, di ganti lagi, foto kendaraan tidak sesuai lgi
- 2) SUGGESTION: Pada halaman ketika mengisi biodata (on android) tidak bisa memilih Tahun dan bulan dengan langsung, melainkan harus menggeser ke bulan dan tahun yang dimau karena tanggalnya tenaik ke atas akibat muncul keyboard pada handphone. pada web, alangkah baiknya untuk bisa meng**upload foto** dari galeri (untuk ktp,STNK, dll)

daripada memfoto langsung KTP. Dan alangkah bagusya jika verifikasi data untuk pendaftaran subsidi di percepat verifikasinya.

- 3) kenapa fitur upluod STNK sekarang tidak bisa **ambil** dari **galeri** hp,...! padahal fitur itu sangat penting, ada sebagian hp yg tidak upluod dengan cara ambil foto secara langsung.!
- 4) **kemudahan transaksi**, lanjutkan dan kembangkan.
- 5) verifikasi sangat lama, tolong di **permudah** dan **percepat** proses nya
- 6) kurang **mudah aplikasinya** waktu daftar
- 7) Pada foto stnk aplikasinya tidak berfungsi tolong diperbaiki biar bisa seperti **aplikasi** lainnya seperti **aplikasi** crypto sat set langsung jadi
- 8) Aplikasi mypertamina sangat membantu kami, dan juga sangat **mudah** mendaftarkan kendaraan kami.
- 9) Transaksi lebih **cepat** dan **mudah**
- 10) Template aja upload foto stnk gagal tolak verifikasi aplikasi blur foto baca **susah daftar** unit sebenarnya ga butuh template admin tindak baik baik upload foto galeri

Setelah diketahui hasil perbandingan dari pembobotan kata tunggal dan pembobotan dengan dua kata yang memiliki nilai kecenderungan yang sama, dugaan sentiment berarti sesuai dengan arah tersebut. Seperti contoh pada pembobotan *cluster* 1 yang memiliki nilai negatif dan kedua bernilai negatif sehingga dapat disimpulkan bahwa mayoritas ulasan dalam *cluster* 1 da *cluster* 2 memang mengarah pada sentimen negatif. Konsistensi antara hasil pembobotan kata tunggal dan dua kata mengindikasikan bahwa persepsi pengguna terhadap layanan atau fitur aplikasi yang tergolong dalam *cluster* tersebut cenderung tidak memuaskan. Dengan demikian, hasil pembobotan ini dapat menjadi dasar yang valid untuk mengidentifikasi area yang memerlukan perbaikan, karena arah sentimen yang negatif telah divalidasi melalui dua pendekatan analisis yang berbeda namun menghasilkan kecenderungan yang serupa.

5.2 Karakteristik Masing-Masing *Cluster*

Setelah memvisualisasikan masing-masing *cluster*, selanjutnya akan menjelaskan kemungkinan penyebab dari setiap *cluster* yang didapatkan dari setiap ulasan yang diberikan oleh pengguna. Adapun berikut pada Tabel 5.9 merupakan hasil karakteristik yang didapatkan dari sampel ulasan yang diambil sebelumnya.

Tabel 5. 9 Karakteristik Ulasan Tiap *Cluster*

<i>Cluster</i>	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
<i>Cluster</i> 1	<p>1. Hadehhh susah daftarnya kode otp ga masuk" boss</p> <p>2. Sangat membantu Tp kenapa sering minta diupded Setelah diupded mau melihat kode QR malah susah dibuka</p> <p>3. Gak bisa buka qr code setelah update, error mulu...</p> <p>4. Tolong kirim VERIVIKASI LEWAT WA. Jangan sms, karena tidak bisa.</p> <p>5. udh verifikasi dibilang gagal, udh beberapa kali kode otp dikirim ke gmail, gak bisa, nih aplikasi gmn sih ? tolong perbaiki, ini malah bikin ribet</p> <p>6. Saya pengguna baru, mau daftarkan my pertamina, kok ga bisa ya..diperbaiki dulu, soal kirim Kode OTP</p> <p>7. Perbaiki pengiriman OTP nya lama sekali, sudah cek email tidak ada, padahal waktu pengisian hanya 3 menit, sudah berulang jadi kadaluarsa OTP nya !!</p> <p>8. Masukin kode otp padahal sudah sama dengan yg dikirim di email, tapi ksalahan terus dan ga bisa masuk, sistemnya di perbaiki, saya lengkapi lagi, subsidi tepat</p>	<p>1. Pengguna kesulitan saat menerima atau memasukan kode OTP</p>	<p>Negatif</p> 

<i>Cluster</i>	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
	ketika login bermasalah harus di pencet beberapa kali baru masuk kemudian masukkan kode otp, sudah di masukkan sesuai yg dikirim di email tapi muncul notif kesalahan terus dan coba lagi nanti		
	9. Kenapa akhir2 ini mau masuk ke subsidi pertalite susah ya,nunggu kode verifikasi gak kunjung dikirim	2. Kendala sistem verifikasi saat login atau registrasi	
	10. Saya tidak bisa masuk aplikasi minta no verifikasi di wa GK dikirim kirim		
<i>Cluster</i> 2	1. Aplikasinya ada bug, saya baru daftar udah sesuai semua mau login dikatakan salah pin atau nomer. Harus nunggu 30 menit karna batas masukkan pin salah tapi sama saja habis 30 menit saya coba lagi. Tolong di perbaiki lagi untuk kesiapan aplikasinya, disuruh daftar buat subsidi tapi aplikasinya saja belum optimal 2. Aplikasi gak wajar Daftar Subsidi tepat formulir benar semua jawaban nya ada kesalahan di formulir pas di coba lagi kesalahan sistem di coba di aplikasi sama di website sama aja. Kalo gak niat bikin aplikasi jangan ngeribetin gw	1. Terdapat banyak bug dan error saat pengisian formulir maupun login aplikasi.	Negatif 

<i>Cluster</i>	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
	<p>ribet beli minyak harus pake aplikasi my Pertamina sedangkan aplikasinya aja gak bisa apa apa,pikirin dulu sebelum buat aplikasi jangan ngeribetin!!!</p> <p>3. daftar subsidi tepat semakin di persulit saja,data kendaraan di tolak segera perbaiki data kendaraan tapi giliran perbaiki data dan mau kirim data hanya loding saja tidak mau terkirim selalu di ulang²...tolong dong segera di evaluasi jangan menyulitkan rakyat kecil</p> <p>4. Aplikasinya ada bug, saya baru daftar udah sesuai semua mau login dikatakan salah pin atau nomer. Harus nunggu 30 menit karna batas masukkan pin salah tapi sama saja habis 30 menit saya coba lagi. Tolong di perbaiki lagi untuk kesiapan aplikasinya, disuruh daftar buat subsidi tapi aplikasinya saja belum optimal</p> <p>5. Daftar ga bisa" aneh , tulisannya terjadi kesalah dalam pengisian form , padahal ga ada yg salah, nyuruh doang daftar tapi aplikasi masih gembel , kenapa si kalo</p>		

<i>Cluster</i>	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
	aplikasi pemerintah atau web dan situs pemerintah itu selalu ga beres		
	6. Daftar subsidi tp gabisa di buka, skalinya bisa lagi isi form udh foto stnk dan foto unit tiba2 balik ke beranda, mau daftar lg no kendaraan sudah ter daftar ga bisa daftar lagi, gimana dong ?	2. Pengguna mengalami kesulitan saat proses pendaftaran subsidi, validasi data kendaraan, dan verifikasi email.	
	7. Minta verifikasi email Tapi pas ngecek email dan mau balik ke aplikasi . Aplikasinya malah nutup sendiri. Ulang ke awal Jadi butuh 2 device untuk verifikasi doang Padahal. dan setiap nge klik di aplikasinya selalu pemberitahuan ga jelas dan menyuruh melengkapi Padahal data sdh lengkap Mohon ditingkatkan lagi kalau mau diwajibkan Jangan diwajibkan dulu tapi aplikasinya tidak siap		
	8. Aplikasi bodoh, klo ga bisa bikin aplikasi jngn bikin daftar aja susah		
	9. Aplikasi mudah di gunakan ..masukan nya tolong dong di mutasi atau di aktifitas harus ada pemberitahuan setiap kita isi BBM jadi jika mobil di pinjam kita bisa ta di isi apa gk nya BBM nya...	3. Sebagian pengguna mengapresiasi fitur setelah berhasil mendaftar, namun proses awal cukup menyulitkan.	
	10. Sedikit repot di awal, isi data app dan top up saldo. setelah nya jd		

<i>Cluster</i>	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
	lebih praktis dan tdk lama ikut antrian, krn punya jalur khusus. boleh lah cukup di apresiasi bumh nya dgn aplikasi nya.		
<i>Cluster</i> 3	<p>1. kuota bbm yg beli tampil aplikasi milik user milik tugas spbu mobil yg gak ngisi bbm beli kuota habis ulang oknum tugas spbu mohon baik keamanannnya Tingkat</p> <p>2. Untuk metode pembayaran e-money nya kurang lengkap tambahin dana min</p> <p>3. Ini aplikasi sebenarnya sudah bagus, hanya saja masih banyak SPBU yg tidak menerapkan kebijakan pemerintah, dan masih banyak pengguna AVANZA, INNOVA, Bahkan Fortuner sekaligus masih banyak yg mengisi BBM bersubsidi. Tolong dong diterapkan dengan benar peraturan pemerintah, tindak SPBU yg masih banyak mengacuhkan kebijakan atau memberikan pelayanan kepada selain penerima subsidi.</p> <p>4. MyPertamina ini sangat bermanfaat seperti aplikasi pembayaran biasanya, seperti saya pengguna yang selalu jarang bawa</p>	<p>1. Pengguna sudah merasa puas dengan layanan aplikasi, hanya saja cenderung mengeluhkan pada metode pembayaran.</p>	<p>Positif</p> 

<i>Cluster</i>	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
	uang tunai/cashless. Tidak perlu turun dari kendaraan untuk bayar pakai debit. Terima kasih.		
	5. Tambahkan metode pembayaran aplikasi dana lah		
	6. Tlg diperbaiki masalah pembayaran menggunakan kartu BRI. Krn selalu gagal terus transaksinya. Hanya bs menggunakan pembayaran dgn link aja		
	7. Banyak spbu yang tidak melayani pembayaran non tunai , seperti hari ini minggu, 3 maret 2024 pukul 18.15 di kota palembang tepat nya di spbu hegar alang-alang lebar, antrian pembayaran non tunai di buka tetapi petugas spbu tidak melayani pembayaran non tunai, padahal sudah menunggu antrian sangat lama		
	8. Tidak bisa download barcode. Bugnya terlalu banyak min. Mohon diperbaiki. Terimakasih	2. Masih adanya bug teknis dan fitur yang belum optimal.	
	9. Terimakasih, aplikasinya sudah bagus. Tetapi mengapa pembayaran melalui kartu kredit sering sekali gagal?	3. Keluhan yang diberikan pengguna cenderung memberikan saran metoede-metoede	
	10. Untuk aplikasi nya bagus tapi tidak bisa menambahkan pembayaran	pembayaran yang	

Cluster	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
	gopay ataupun dana..dan sering kali error. mohon bantuan dan perbaikan nya. terimakasih	diinginkan pengguna aplikasi	
Cluster 4	<p>1. Bisa di bantu gak min, akun saya kok verifikasi terus, mau daftar nopol kendaraan di tolak terus, alasannya foto kendaraan tidak sesuai, sudah di ganti dg sesuai saran pengambilan foto, malah foto stnk tidak sesuai, di ganti lagi, foto kendaraan tidak sesuai lgi</p> <p>2. SUGGESTION: Pada halaman ketika mengisi biodata (on android) tidak bisa memilih Tahun dan bulan dengan langsung, melainkan harus menggeser ke bulan dan tahun yang dimau karena tanggalnya tenaik ke atas akibat muncul keyboard pada handphone. pada web, alangkah baiknya untuk bisa mengupload foto dari galeri (untuk ktp,STNK, dll) daripada memfoto langsung KTP. Dan alangkah bagusnya jika verifikasi data untuk pendaftaran subsidi di percepat verifikasinya.</p> <p>3. kenapa fitur upluod STNK sekarang tidak bisa ambil dari galeri hp,..! padahal fitur itu sangat penting, ada sebagian hp yg</p>	<p>1. Pengguna masih cenderung mengeluhkan verifikasi pendaftaran yang berjalan lambat</p>	<p>Netral</p> 

<i>Cluster</i>	Sampel Ulasan	Kemungkinan Penyebab	Sentimen
	tidak upluod dengan cara ambil foto secara langsung.!		
	4. kemudahan transaksi , lanjutkan dan kembangkan.		
	5. verifikasi sangat lama, tolong di permudah dan percepat proses nya		
	6. kurang mudah aplikasinya waktu daftar		
	7. Pada foto stnk aplikasinya tidak berfungsi tolong diperbaiki biar bisa seperti aplikasi lainnya seperti aplikasi crypto sat set langsung jadi		
	8. Aplikasi mypertamina sangat membantu kami, dan juga sangat mudah mendaftarkan kendaraan kami.	2. Meskipun terdapat ulasan negative yang diberikan pengguna akan tetapi seimbang dengan pemberian ulasan positifnya	
	9. Transaksi lebih cepat dan mudah		
	10. Template aja upload foto stnk gagal tolak verifikasi aplikasi blur foto baca susah daftar unit sebenarnya ga butuh template admin tindak baik baik upload foto galeri	3. Perlu fleksibilitas input data (kamera atau galeri).	

Setelah menentukan karakteristik ulasan tiap *cluster* dalam perbaikan kedepannya dan mendapatkan kemungkinan-kemungkinan penyebab yang didapatkan, maka dilakukan analisis SWOT dari kemungkinan penyebab tersebut serta diberikan strategi dan Solusi dari permasalahan yang ditimbulkan.

5.3 Analisis SWOT Berdasarkan Hasil *Clustering*

Setelah dilakukan pengelompokan ulasan pengguna aplikasi MyPertamina ke dalam empat *cluster* menggunakan pendekatan *text mining* dan *K-Means Clustering*, selanjutnya dilakukan analisis SWOT untuk mengetahui kondisi internal (kekuatan dan kelemahan) serta eksternal (peluang dan ancaman) dari masing-masing *cluster*.

Analisis ini bertujuan untuk membantu perusahaan dalam merumuskan strategi yang tepat guna memperbaiki kualitas layanan aplikasi berdasarkan persepsi dan pengalaman nyata dari pengguna, dimana setiap *cluster* mewakili karakteristik opini yang berbeda. *Cluster* 1 dan 2 didominasi sentimen negatif dan menunjukkan masalah fungsional aplikasi seperti kendala pendaftaran dan metode pembayaran, sedangkan *cluster* 3 mencerminkan kepuasan pengguna dan loyalitas. *Cluster* 4 bersifat netral namun tetap mengandung potensi perbaikan dari sisi teknis.

Dengan melakukan analisis SWOT, perusahaan dapat mengidentifikasi kekuatan yang perlu dipertahankan, kelemahan yang harus diperbaiki, peluang yang bisa dimanfaatkan, serta ancaman yang harus diantisipasi. Hasil analisis ini menjadi dasar penyusunan strategi perbaikan aplikasi MyPertamina secara lebih terarah dan sesuai kebutuhan pengguna. Adapun penjelasan analisis SWOT dapat dilihat pada Tabel 5.10 berikut:

Tabel 5. 10 Menentukan Setiap SWOT

<i>Cluster</i>	Strengths (Kekuatan)	Weaknesses (Kelemahan)	Opportunities (Peluang)	Threats (Ancaman)
<i>Cluster 1</i> <i>(Sentimen Negatif)</i>	<ul style="list-style-type: none"> Aplikasi sudah dikenal luas Dukungan digitalisasi dari pemerintah 	<ul style="list-style-type: none"> Masalah OTP & verifikasi lambat Error pada QR code Pengalaman login buruk 	<ul style="list-style-type: none"> Dukungan masyarakat terhadap digitalisasi BBM Program Subsidi Tepat 	<ul style="list-style-type: none"> Pengguna beralih ke alternatif Citra negatif karena pengalaman awal buruk
<i>Cluster 2</i> <i>(Sentimen Negatif)</i>	<ul style="list-style-type: none"> Kemampuan integrasi 	<ul style="list-style-type: none"> Bug saat input data 	<ul style="list-style-type: none"> Masih ada kebutuhan pengguna 	<ul style="list-style-type: none"> Kepercayaan publik menurun

<i>Cluster</i>	Strengths (Kekuatan)	Weaknesses (Kelemahan)	Opportunities (Peluang)	Threats (Ancaman)
	dengan data subsidi	<ul style="list-style-type: none"> Error pengisian form Sistem pendaftaran tidak optimal 	terhadap layanan berbasis aplikasi	karena sistem gagal
	<ul style="list-style-type: none"> Fitur cukup lengkap secara konsep 			<ul style="list-style-type: none"> Persepsi buruk terhadap aplikasi pemerintah
Cluster 3 (<i>Sentimen Positif</i>)	<ul style="list-style-type: none"> Sistem pembayaran cashless membantu pengguna Pengalaman pengguna relatif baik 	<ul style="list-style-type: none"> Keterbatasan metode pembayaran Kendala teknis seperti error saat download barcode 	<ul style="list-style-type: none"> Potensi meningkatkan loyalitas pengguna Perluasan fitur pembayaran 	<ul style="list-style-type: none"> Persaingan aplikasi pembayaran lainnya Tidak semua SPBU melayani sistem
Cluster 4 (<i>Sentimen Netral</i>)	<ul style="list-style-type: none"> Proses pendaftaran dianggap cukup mudah oleh sebagian pengguna Tampilan cukup membantu 	<ul style="list-style-type: none"> Masalah upload foto & form Verifikasi data lambat UI/UX tidak optimal 	<ul style="list-style-type: none"> Meningkatkan efisiensi pendaftaran subsidi Perbaikan sistem foto dan form input 	<ul style="list-style-type: none"> Tingkat frustrasi pengguna dapat meningkat jika tidak diperbaiki Ketidakpercayaan terhadap proses verifikasi digital

Setelah melihat hasil analisis SWOT tiap *cluster*, langkah selanjutnya adalah merumuskan strategi perbaikan dengan pendekatan TOWS yang akan disesuaikan dengan kekuatan dan tantangan masing-masing *cluster*. Berikut Tabel 5.11 adalah strategi perbaikan berdasarkan TOWS *Matrix* dari hasil analisis SWOT masing-masing *cluster* ulasan aplikasi MyPertamina:

Tabel 5. 11 Hasil dari TOWS *Matrix*

<i>Cluster</i>	Strategi S-O (Strength– Opportunity)	Strategi W-O (Weakness– Opportunity)	Strategi S-T (Strength– Threat)	Strategi W-T (Weakness– Threat)
Cluster 1 (<i>Negatif Error Verifikasi & OTP</i>)	- Maksimalkan promosi fitur aplikasi yang mendukung program “Subsidi Tepat” untuk meningkatkan adopsi	- Tingkatkan sistem backend (OTP, QR code) dan lakukan audit berkala untuk mencegah error	- Manfaatkan kekuatan branding MyPertamina untuk mengatasi persepsi awal pengguna	- Perbaiki sistem verifikasi & login agar pengguna tidak beralih ke alternatif dan menjaga reputasi
Cluster 2 (<i>Negatif Bug Pendaftaran</i>)	- Kembangkan edukasi digital untuk bantu pengguna memahami proses pendaftaran subsidi	- Sederhanakan form input, perbaiki bug, dan tambah fitur upload otomatis berbasis AI	- Gunakan citra aplikasi pemerintah untuk menjaga kepercayaan pengguna	- Evaluasi sistem front-end & back-end agar bug tidak berulang dan pengguna tidak kehilangan kepercayaan
Cluster 3 (<i>Positif Cashless & Efisien</i>)	- Perluas fitur pembayaran (e.g. GoPay, Dana) dan jadikan USP (Unique Selling Proposition)	- Tambah opsi pembayaran & UI agar pengalaman pengguna makin praktis	- Jadikan testimoni pengguna puas sebagai alat branding & peningkatan engagement	- Perkuat server dan layanan pelanggan untuk mencegah kesalahan kecil menjadi

<i>Cluster</i>	Strategi S-O (Strength– Opportunity)	Strategi W-O (Weakness– Opportunity)	Strategi S-T (Strength– Threat)	Strategi W-T (Weakness– Threat)
				ulasan negatif
<i>Cluster 4</i> (<i>Netral Verifikasi Foto & Form</i>)	- Luncurkan update UI/UX baru dengan fitur upload dari galeri, dan panduan foto otomatis	- Permudah <i>form</i> verifikasi dengan sistem deteksi kesalahan dan progress bar	- Promosikan kemudahan proses daftar kepada calon pengguna baru	- Bangun sistem validasi otomatis untuk hindari frustrasi karena pengulangan proses pendaftaran

Berdasarkan hasil analisis pengelompokan ulasan pengguna aplikasi MyPertamina, ditemukan bahwa terdapat empat *cluster* dengan karakteristik sentimen yang berbeda. *Cluster 1* dan *cluster 2* memiliki dominasi sentimen negatif, *cluster 3* menunjukkan sentimen positif, sedangkan *cluster 4* cenderung bersifat netral. Dari keempat *cluster* tersebut, prioritas utama perbaikan diarahkan pada *cluster 1* karena memiliki jumlah ulasan terbanyak dengan keluhan utama mengenai pengguna kesulitan saat menerima atau memasukan kode OTP dan kendala sistem verifikasi saat login atau registrasi. Oleh karena itu, strategi perbaikan yang dapat diterapkan antara lain sebagai berikut:

1) Memaksimalkan Promosi Fitur-Fitur Aplikasi

Perusahaan dapat memaksimalkan promosi fitur-fitur aplikasi melalui *sharing session* yang mendukung kemudahan dalam proses verifikasi, seperti fitur “Subsidi Tepat”, untuk meningkatkan adopsi dan kepercayaan pengguna terhadap aplikasi. Promosi ini bisa dilakukan melalui media sosial, iklan digital, serta edukasi langsung kepada pengguna agar lebih familiar dengan fitur yang disediakan (Kusuma & Diana, 2023).

2) *Maintenance* Sistem

Strategi perbaikan untuk mengatasi keluhan pada Cluster 1, khususnya terkait error verifikasi dan OTP, adalah dengan melakukan maintenance sistem secara berkala. Langkah ini mencakup audit backend (seperti OTP dan QR code), peningkatan stabilitas server, serta penggunaan provider OTP yang lebih andal. Selain itu, penting untuk menerapkan sistem monitoring real-time dan pengujian sebelum pembaruan, guna mencegah error berulang dan memastikan pengalaman pengguna tetap optimal (Gede Yoga Shamgita et al., 2020).

Untuk *cluster 2* ulasan yang diberikan pengguna banyak mengeluhkan terkait banyak bug dan error saat pengisian formulir maupun login aplikasi, pengguna mengalami kesulitan saat proses pendaftaran subsidi, validasi data kendaraan, dan verifikasi email, dan juga memberikan apresiasi pada fitur setelah berhasil mendaftar, namun proses awal cukup menyulitkan. Dari hasil ulasan yang diberikan tersebut, dapat diberikan strategi perbaikan bagi perusahaan antara lain sebagai berikut:

1) Desain Ulang *Interface & UX*

Perusahaan perlu melakukan *redesign* tampilan aplikasi dan alur penggunaan (*user experience/UX*) agar lebih intuitif dan *user-friendly*, terutama pada tahap awal pendaftaran, pengisian data kendaraan, dan *login*. Langkah ini bertujuan agar pengguna tidak kesulitan dalam memahami alur pendaftaran maupun proses validasi. Dan juga kecepatan dalam pengaplikasian suatu aplikasi juga dipengaruhi oleh *interface* aplikasi (Rama Putra et al., 2025).

2) Penambahan Fitur Bantuan Cepat (*Quick Help*)

Menyediakan panduan langkah demi langkah dan fitur chatbot atau tombol bantuan langsung pada halaman registrasi untuk membantu pengguna menyelesaikan kendala secara *real-time* (Aryani et al., 2022).

Selanjutnya pada *cluster 3*, menunjukkan sentimen positif dari pengguna, di mana sebagian besar merasa terbantu dengan kemudahan transaksi, fitur pembayaran non-tunai, serta kenyamanan saat menggunakan aplikasi MyPertamina. Sehingga kesalahan yang terdapat pada aplikasi tidak begitu menjadi masalah tetapi pengguna banyak untuk memberikan saran. Terdapat beberapa saran yang berkaitan dengan keterbatasan metode pembayaran dan kendala minor teknis seperti *error* saat unduh *barcode* atau pembayaran gagal. Berdasarkan ulasan tersebut, strategi perbaikan dan penguatan yang dapat diterapkan adalah Perluasan Metode Pembayaran Digital. Dimana menambahkan lebih banyak opsi *e-wallet* dan metode pembayaran digital yang populer seperti DANA, OVO, GoPay, dan kartu kredit tambahan.

Dengan memenuhi preferensi pengguna, aplikasi dapat menjangkau lebih luas segmen pengguna *cashless* di Indonesia (Audina et al., 2021). Dan juga Maintenance sistem secara berkala sangat penting dilakukan untuk memastikan performa aplikasi tetap stabil, terutama karena mayoritas pengguna di Cluster 3 telah merasa puas dengan fitur yang ada. Perusahaan perlu rutin memantau dan mengatasi bug minor seperti kegagalan unduh barcode atau error transaksi, yang meskipun tidak dominan, tetap dapat mengganggu kenyamanan pengguna. Selain itu, pemeliharaan ini juga mencakup pembaruan keamanan, optimalisasi performa server, serta penyesuaian fitur agar tetap relevan dengan perkembangan teknologi dan preferensi pengguna (Fairnando Augusto et al., 2019).

Dan terakhir untuk *cluster* 4 mayoritas ulasan yang diberikan oleh pengguna menunjukkan sentimen netral, di mana keluhan yang muncul berkaitan dengan masalah teknis ringan seperti kendala saat mengunggah foto STNK atau KTP, sistem verifikasi yang lambat, dan aplikasi yang kadang tertutup sendiri atau kembali ke beranda. Meskipun begitu, beberapa pengguna juga menyampaikan apresiasi terhadap kemudahan penggunaan aplikasi secara umum. Berdasarkan karakteristik tersebut, strategi perbaikan yang dapat dilakukan adalah dengan memperbaiki sistem unggah dokumen, khususnya dengan menambahkan opsi unggah dari galeri serta memperluas toleransi terhadap kualitas dan format gambar agar pengguna tidak mengalami kegagalan saat proses pendaftaran (Paramartha et al., 2018). Selain itu, Selain itu, perusahaan juga sebaiknya secara rutin melakukan survei atau analisis terhadap kebutuhan, keinginan, dan harapan para pengguna aplikasi MyPertamina. Langkah ini penting agar perusahaan bisa mengetahui secara langsung apa yang benar-benar dibutuhkan oleh pengguna, sehingga pengembangan aplikasi bisa lebih tepat sasaran. Dengan memahami masukan dari pengguna, perusahaan bisa memberikan perbaikan yang sesuai dan layanan yang lebih baik sejak awal. Hal ini diharapkan dapat mencegah munculnya keluhan atau ulasan negatif yang serupa di kemudian hari (Muharsyah & Ekawati, 2022).

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan hasil pengolahan data dan analisis yang telah dilakukan, maka dapat ditarik Kesimpulan dari penelitian ini yaitu sebagai berikut:

1. Hasil pengelompokkan ulasan pengguna aplikasi MyPertamina terbentuk sebanyak 4 *cluster* dengan jumlah ulasan pada tiap *cluster*-nya yaitu *cluster* 1 memiliki jumlah ulasan sebanyak 378, *cluster* 2 memiliki jumlah ulasan sebanyak 1254, *cluster* 3 memiliki jumlah ulasan sebanyak 2585, dan untuk *cluster* 4 memiliki jumlah ulasan sebanyak 378. Masing-masing *cluster* menunjukkan karakteristik sentimen yang berbeda, yakni dua *cluster* dengan sentimen negatif, satu *cluster* positif, dan satu *cluster* netral. *Cluster* 1 berisi keluhan mengenai verifikasi dan *login* (khususnya kode OTP), *cluster* 2 memuat keluhan tentang bug sistem dan proses pendaftaran, *cluster* 3 menunjukkan kepuasan pengguna terhadap kemudahan penggunaan, sedangkan *cluster* 4 mencerminkan ulasan netral yang disertai dengan masukan tentang perbaikan kualitas layanan.
2. Proses optimalisasi jumlah *cluster* dilakukan dengan menggunakan metode Elbow, dan hasilnya menunjukkan bahwa empat *cluster* merupakan jumlah yang paling ideal. Hal ini ditunjukkan dengan titik tekuk (*elbow*) yang signifikan pada grafik inertia. Ini berarti bahwa empat kelompok tersebut sudah cukup mewakili variasi sentimen dan topik yang muncul dalam ulasan pengguna, tanpa membuat kelompok yang terlalu banyak atau tumpang tindih.
3. Berdasarkan hasil analisis SWOT yang dilakukan pada setiap *cluster* ulasan pengguna aplikasi MyPertamina, terdapat beberapa strategi perbaikan yang dapat diterapkan untuk meningkatkan rating aplikasi. Pada *cluster* 1 dan 2 yang didominasi oleh sentimen negatif, fokus utama strategi diarahkan pada perbaikan sistem verifikasi (khususnya OTP), penguatan backend, perbaikan bug dan error saat pengisian data, serta desain ulang tampilan dan alur penggunaan aplikasi agar lebih mudah digunakan oleh pengguna baru. Selain itu, perlu ditambahkan fitur bantuan cepat agar pengguna dapat menyelesaikan kendala secara real-time. Sementara itu, pada *cluster* 3 yang menunjukkan sentimen positif, strategi diarahkan untuk mempertahankan dan meningkatkan kepuasan pengguna melalui penambahan opsi metode pembayaran digital yang lebih variatif (seperti DANA, GoPay, OVO, kartu kredit), serta pemeliharaan aplikasi secara berkala agar tetap stabil dan nyaman digunakan. Sedangkan pada *cluster* 4 yang bersifat netral, perbaikan lebih difokuskan pada

sisi teknis seperti sistem unggah dokumen, optimasi kualitas dan format gambar, serta penyesuaian sistem verifikasi agar lebih cepat dan tidak menyulitkan pengguna. Perusahaan juga disarankan untuk rutin melakukan survei kepuasan dan kebutuhan pengguna agar pengembangan aplikasi lebih tepat sasaran. Dengan menerapkan strategi berbasis hasil SWOT dari tiap cluster ini, perusahaan dapat meningkatkan kualitas layanan dan pengalaman pengguna secara menyeluruh, yang pada akhirnya akan berdampak positif terhadap peningkatan rating aplikasi MyPertamina di platform digital.

1.7 Saran

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, peneliti memberikan beberapa saran sebagai berikut:

1. Bagi PT Pertamina (Persero) Khususnya yang Menangani Teknologi Informasi

Sehubungan dengan empat *cluster* yang didapatkan dari *K-meas clustering* dan metode elbow dengan *cluster* 1 berisi keluhan mengenai verifikasi dan *login* (khususnya kode OTP), *cluster* 2 memuat keluhan tentang bug sistem dan proses pendaftaran, *cluster* 3 menunjukkan kepuasan pengguna terhadap kemudahan penggunaan, sedangkan *cluster* 4 mencerminkan ulasan netral yang disertai dengan masukan tentang perbaikan kualitas layanan. Perusahaan diharapkan dapat lebih responsif dalam menanggapi ulasan pengguna, khususnya yang berkaitan dengan proses verifikasi OTP, pendaftaran subsidi, serta bug dan error dalam aplikasi. Disarankan untuk melakukan *maintenance* sistem secara berkala, memperkuat server, dan melakukan evaluasi rutin terhadap fitur-fitur yang bermasalah. Selain itu, perusahaan dapat meningkatkan pengalaman pengguna dengan memperbaiki tampilan antarmuka aplikasi agar lebih intuitif dan mudah dipahami, terutama pada tahap awal pendaftaran. Perlu juga dilakukan penambahan fitur bantuan cepat seperti chatbot serta perluasan metode pembayaran digital untuk meningkatkan kenyamanan pengguna. Dengan memperhatikan setiap klaster ulasan yang terbentuk, PT Pertamina dapat menyusun strategi perbaikan yang lebih tepat sasaran dan berdampak positif terhadap persepsi serta rating aplikasi MyPertamina di masa mendatang.

2. Bagi Penelitian Selanjutnya

Berdasarkan penelitian yang telah dilakukan, saran dari peneliti terhadap penelitian selanjutnya yaitu:

- a. Penelitian ini dapat dijadikan acuan metodologi untuk analisis ulasan aplikasi berbasis *text mining* dan *clustering*. Peneliti selanjutnya dapat melanjutkan studi ini dengan menambahkan tahapan lanjutan seperti analisis sentimen secara mendalam atau pengujian strategi perbaikan yang telah diusulkan.
- b. Disarankan untuk membandingkan metode *clustering* (seperti K-Means) yang digunakan dalam penelitian ini dengan metode lain seperti klasifikasi agar mendapatkan suatu analisis baru yang lebih tepat dan dapat melihat akurasi dan relevansi hasil analisis strategi per klaster.
- c. Sebagai tambahan, akan sangat menarik jika penelitian lanjutan menggunakan pendekatan evaluatif dengan mengukur dampak implementasi strategi yang telah diusulkan terhadap peningkatan rating dan kepuasan pengguna aplikasi.

DAFTAR PUSTAKA

- Abdurahman, H., Riswaya, A. R., & Id, A. (2014). Aplikasi Pinjaman Pembayaran Secara Kredit Pada BANK YUDHA BHAKTI STMIK Mardira Indonesia, Bandung Email: asepe_ririh@stmik-mi. *Jurnal Computech & Bisnis*, 8, 61–69.
- Amir Latif, R. M., Talha Abdullah, M., Aslam Shah, S. U., Farhan, M., Ijaz, F., & Karim, A. (2019). Data Scraping from Google Play Store and Visualization of its Content for Analytics. *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET)*, 1–8. <https://doi.org/10.1109/ICOMET.2019.8673523>
- Amrullah, A. Z., Sofyan Anas, A., Adrian, M., & Hidayat, J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal*, 2(1). <https://doi.org/10.30812/bite.v2i1.804>
- Andreswari, D., Suranti, D., & Aulia Trianggara, D. (2024). Penerapan Text Mining Dalam Pengelompokan Topik Skripsi Menggunakan Metode TF-IDF Berdasarkan Abstrak Skripsi. *Jurnal Komputer Indonesia*, 3(2), 69–78. <https://doi.org/10.37676/jki.v3i2>
- Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R&D Management*, 50(3), 329–351. <https://doi.org/10.1111/radm.12408>
- Apriantoro, M. S., Sabila, R. F., & Rosyadhi, R. I. (2022). Islamic Law Perspective in the Application of My Pertamina as a Non-Cash Payment System and Control of Fuel Subsidy Flow. *Journal of Transcendental Law*, 4(1), 71–80. <https://doi.org/10.23917/jtl.v4i1.19976>
- Aryani, R., Marzal, J., & Weni, I. (2022). Implementasi Aplikasi Bantuan Cepat Daring untuk Peningkatan Layanan TIK di Universitas Jambi. *Jurnal Ilmiah Media Sisfo*, 16(1), 41–53. <https://doi.org/10.33998/mediasisfo.2022.16.1.1119>
- Asri, J. S., & Wahyu, S. (2021). *Analisis Sentimen Menerapkan Lexicon-Learning Based Untuk Melihat Opini Masyarakat Mengenai Protokol Kesehatan Dan Perkembangan Vaksin Covid-19 Di Indonesia Menggunakan Dataset Twitter*.
- Asyhari, M. Y., Juwari, J., Hapsari, E. D., & Yulianto, S. (2023). Pendekatan Metode Kolokasi untuk Text Processing Ulasan Aplikasi Android Surveilans Penyebaran Covid-19 di Indonesia. *Journal Information System Development (ISD)*, 8(1), 33–42. <https://doi.org/10.19166/isd.v8i1.586>
- Audina, M., Andriana, I., Ekonomi, F., & Sriwijaya, U. (2021). *Faktor Yang Mempengaruhi Behavioral Intention Transaksi Keuangan Digital (E-Wallet) Pada Generasi Milenial Di Kota Palembang*. <https://doi.org/10.25105/mrbm.v2i12.12895>
- Azzahra, S. F., Setiawan, N. Y., & Purnomo, W. (2025). *Analisis Quality in Use dan SWOT Produk Skincare Lokal dan Global Berdasarkan Sentimen Pengguna* (Vol. 9, Issue 2). <http://j-ptiik.ub.ac.id>
- Bagnasco, S., Berzano, D., Guarise, A., Lusso, S., Masera, M., & Vallero, S. (2015). Monitoring of IaaS and scientific applications on the Cloud using the Elasticsearch ecosystem. *Journal of Physics: Conference Series*, 608, 012016. <https://doi.org/10.1088/1742-6596/608/1/012016>
- Fairnando Augusto, Y., Rachmadi, A., & Herlambang, A. D. (2019). *Analisis Kesuksesan Aplikasi Mobile Pemesanan Tiket Bioskop M-Tix Cinema 21 Berdasarkan Perspektif Pengguna di Kota Malang Menggunakan Pendekatan Delone and McLean Success Model*. 3(1), 7–12. <http://j-ptiik.ub.ac.id>

- Fauzi Siregar, H., Handika Siregar, Y., & Jend Ahmad Yani Kisaran Sumatera Utara, J. (2018). Perancangan Aplikasi Komik Hadist Berbasis Multimedia. *Jurnal Teknologi Informasi*, 2(2).
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*.
- Gede Yoga Shamgita, I., Gusti Lanang Agung Raditya, I., & Gede Juliana Eka Putra, I. (2020). Analisis Dan Evaluasi Tata Kelola Teknologi Informasi USSI Software Menggunakan Framework COBIT 5 Pada PT. BPR Naga. 9, 67–74.
- Gobel, R. K., Laksmono, B. S., Huseini, M., & Siscawati, M. (2024). Equity and Efficiency: An Examination of Indonesia's Energy Subsidy Policy and Pathways to Inclusive Reform. *Sustainability*, 16(1), 407. <https://doi.org/10.3390/su16010407>
- Goreti, M., Novia N, Y., & Wahyuningsih, S. (2016). Perbandingan Hasil Analisis Cluster dengan Menggunakan Metode Single Linkage dan Metode C-Means (Studi Kasus: Data Tingkat Kualitas Udara Ambien pada Perusahaan Perkebunan di Kabupaten Kutai Barat Tahun 2014). *Jurnal EKSPONENSIAL*, 7(1).
- Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26–32. <https://doi.org/10.1016/j.procs.2013.05.005>
- Hajar, M., Novia Nasution, Y., Prangga, S., Matematika, J., Matematika, F., Ilmu, D., & Alam, P. (2022). *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya Terbitan II*.
- Haris Ahmad Gozali, Mochamad Alfian Rosid, & Sumarno. (2020). Classification of Student Complaints with the Naive Bayes and Literature Methods. *JOINCS (Journal of Informatics, Network, and Computer Science)*, 3(1), 22–26. <https://doi.org/10.21070/joincs.v3i0.711>
- Hendi Prastiyo, R., & Lahindah, L. (2024). Customer Requirements on Digital Application in Energy Sectors for B2B Customers (a Part of Quality Function Deployment Analysis on My Pertamina for Business). *Interdisciplinary Journal and Hummanity (INJURITY)*, 3(1), 31–48. <https://doi.org/10.58631/injury.v3i1.160>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Huda, B., & Priyatna, B. (2019). Penggunaan Aplikasi Content Management System (CMS) Untuk Pengembangan Bisnis Berbasis E-commerce. *SYSTEMATICS*, 1(2), 81. <https://doi.org/10.35706/sys.v1i2.2076>
- Indrayanto, C. G., Ratnawati, D. E., & Rahayudi, B. (2023). Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(3), 1131–1139. <http://j-ptiik.ub.ac.id>
- Indriyani, F. A., Fauzi, A., & Faisal, S. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine. *TEKNOSAINS: Jurnal Sains, Teknologi Dan Informatika*, 10(2), 176–184. <https://doi.org/10.37373/tekno.v10i2.419>
- Irfani, F. F. (2020). Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine. *JBMI (Jurnal Bisnis, Manajemen, Dan Informatika)*, 16(3), 258–266. <https://doi.org/10.26487/jbmi.v16i3.8607>
- Jatnika, R. F., Kaniawulan, I., & Singasatia, D. (2023). Analisis Penerimaan Aplikasi MyPertamina Menggunakan Metode Technology Acceptance Model (TAM). 14(2). <http://ejurnal.provisi.ac.id/index.php/JTIKP>
- KBBI Daring. (2016).

- Kristiawan, K., & Widjaja, A. (2021). Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1). <https://doi.org/10.28932/jutisi.v7i1.3182>
- Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, 4(10), 1–14. <https://doi.org/10.5815/ijisa.2012.10.01>
- Kumar, R., Sarddar, D., Sarkar, I., Bose, R., & Roy, S. (2020). A Literature Survey On Sentiment Analysis Techniques Involving Social Media And Online Platforms. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, 9, 5. www.ijstr.org
- Kusuma, I. G. N. A. E. T., & Diana, N. W. I. (2023). Optimalisasi Penggunaan Aplikasi My Pertamina Untuk Penjualan Bbm Subsidi Melalui Sharing Session Di Spbu Gatsu Timur. *Prosiding Seminar Nasional Pengabdian Masyarakat (SENEMA)*, 2(1), 359–365. <https://e-journal.unmas.ac.id/index.php/seminarfeb/article/view/6494>
- Kusumastuti, R. D., Nurmala, N., Rouli, J., & Herdiansyah, H. (2022). Analyzing the factors that influence the seeking and sharing of information on the smart city digital platform: Empirical evidence from Indonesia. *Technology in Society*, 68, 101876. <https://doi.org/10.1016/j.techsoc.2022.101876>
- Maria Gomez Hidalgo, J. (2002). *Tutorial: Text Mining and Internet Content Filtering*. <https://www.researchgate.net/publication/237250353>
- Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC. *Jurnal Teknologi Terpadu*, 9(1), 42–48. <https://doi.org/10.54914/jtt.v9i1.609>
- Meisya Permata Aulia, T., Jamaludin, A., & Nur Padilah, T. (2021). Extractive Text Summerization Pada Berita Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 727–735.
- Muharsyah, A., & Ekawati, R. K. (2022). Analisis Pengaruh Kualitas Informasi, Kualitas Sistem dan Kualitas Layanan Terhadap Kepuasan Pengguna Pada Aplikasi Tokopedia Dengan Model Delone And Mclean Di Kota Palembang. *JuSiTik : Jurnal Sistem Dan Teknologi Informasi Komunikasi*, 4(2), 20–27. <https://doi.org/10.32524/jusitik.v4i2.505>
- Mustofa, A., & Novita, R. (2022). Klasifikasi Sentimen Masyarakat Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Text Mining Pada Twitter. *Building of Informatics, Technology and Science (BITS)*, 4(1). <https://doi.org/10.47065/bits.v4i1.1628>
- Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. *Journal of Physics: Conference Series*, 1361(1), 012015. <https://doi.org/10.1088/1742-6596/1361/1/012015>
- Oktaviani, V., Warsito, B., Yasin, H., Santoso, R., & Suparti. (2021). Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method. *Journal of Physics: Conference Series*, 1943(1), 012147. <https://doi.org/10.1088/1742-6596/1943/1/012147>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/15000000011>
- Panjaitan, N., & Manurung, N. (2022). Sentiment Analysis of Google Classroom Application Using Support Vector Machine. *2022 6th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, 117–120. <https://doi.org/10.1109/ELTICOM57747.2022.10038262>

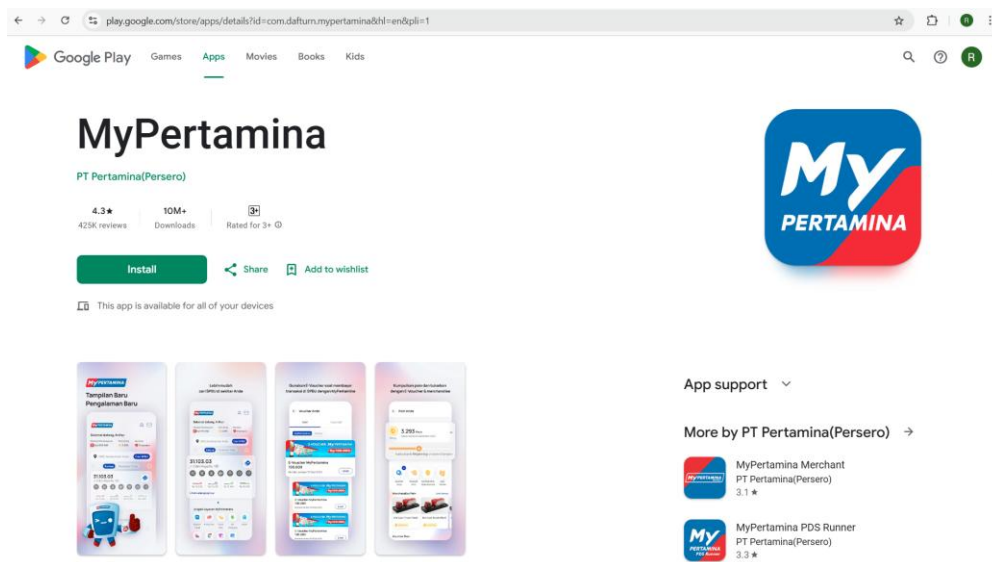
- Paramartha, A. A. G. Y., Sindu, I. G. P., & Permana, A. A. J. (2018). Usability Testing Pada Office Add-Ins Dalam Mengintegrasikan Sistem Rekonstruksi Materi Dan Sistem Manajemen Pembelajaran. *Jurnal Teknologi Informasi Dan Komputer*, 4(2). <https://doi.org/10.36002/jutik.v4i2.550>
- Pratama, E. E., & Lestika Atmi, R. (2020). A Text Mining Implementation Based on Twitter Data to Analyse Information Regarding Corona Virus in Indonesia. *Journal of Computers for Society*, 1. <https://ejournal.upi.edu/index.php/JCS>
- Pratanto, D., Rousyati, R., Wati, F. F., Widodo, A. E., Suleman, S., & Wijianto, R. (2020). App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm. *Journal of Physics: Conference Series*, 1641(1), 012043. <https://doi.org/10.1088/1742-6596/1641/1/012043>
- Putra, R. P. I., Akbar, M., & Amalia, R. (2020). Analisis Sentimen Masyarakat Terhadap Kinerja Persatuan Sepakbola Seluruh Indonesia Menggunakan Metode Backpropagation. *Journal of Information Technology Ampera*, 1(2), 106–118. <https://doi.org/10.51519/journalita.volume1.issuue2.year2020.page106-118>
- Putra, R. R. C., Perkasa, E. B., Sugihartono, T., Alkayess, A. P., Sandro, I. D., & Indallah, R. (2023). Komparasi Naive Bayes dengan Support Vector Machine dalam Analisis Sentimen Aplikasi MyPertamina. *SATIN - Sains Dan Teknologi Informasi*, 9(2), 90–99. <https://doi.org/10.33372/stn.v9i2.1042>
- Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181(1), 25–29. <https://doi.org/10.5120/ijca2018917395>
- Radiena, G., & Nugroho, A. (2023). Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine. *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, 6(1), 1–10. <https://doi.org/10.37792/jukanti.v6i1.836>
- Rahmawanti, M., & Hastuti. (2019, August 30). Analisis SWOT dan TOWS Sebagai Landasan dalam Perencanaan Strategis Sistem Informasi Akuntansi Persediaan di Kabupaten Bandung Barat.
- Rais, Z., Hakiki, F. T. T., & Aprianti, R. (2022). Sentiment Analysis of Peduli Lindungi Application Using the Naive Bayes Method. *SAINSMAT: Journal of Applied Sciences, Mathematics, and Its Education*, 11(1), 23–29. <https://doi.org/10.35877/sainsmat794>
- Rama Putra, G., Forca, A. J., Sardjono, W., & Nursetiaji, O. (2025). Re-Design User Interface (UI) Aplikasi Mobile Domino's Pizza Berdasarkan Hasil Analisis User Experience (UX). *Jurnal Teknologi Dan Informasi*, 15. <https://doi.org/10.34010/jati.v15i1.13657>
- Ramadhan, R. F., Hadi Wijoyo, S., & Saputra, M. C. (2023). Penerapan Metode K-Means Clustering pada Ulasan Perumahan PT XYZ di Google Maps untuk Formulasi Strategi Bisnis dengan Analisis SWOT. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(6), 2879–2888. <http://j-ptiik.ub.ac.id>
- Ramanda MD, Reza Dwi Restiyan, & Hafidz Irsyad. (2024). Analisis Sentimen Masyarakat terhadap Perilaku Lawan Arah yang Diunggah pada Media Sosial Youtube Menggunakan Naïve Bayes. *BANDWIDTH: Journal of Informatics and Computer Engineering*, 2(2), 75–83. <https://doi.org/10.53769/bandwidth.v2i2.706>
- Rejito, J., Atthariq, A., & Abdullah, A. S. (2021). Application of text mining employing k-means algorithms for clustering tweets of Tokopedia. *Journal of Physics: Conference Series*, 1722(1), 012019. <https://doi.org/10.1088/1742-6596/1722/1/012019>
- Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan

- Naïve Bayes Classifier. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 2(5), 178–185. <https://doi.org/10.30865/klik.v2i5.362>
- Roihan, A., Abas Sunarya, P., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
- Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, 874(1), 012017. <https://doi.org/10.1088/1757-899X/874/1/012017>
- S O Mosioi, H. B., & Mailoa, E. (2021). Analisa Sentimen Publik Terkait Otonomi Khusus (OTSUS) di Papua dengan Pendekatan Sains Data. *Prosiding SISFOTEK*, <https://seminar.iaii.or.id/index.php/SISFOTEK/issue/view/6>, 153–156. <https://academic.droneempriit.id/#/search/view/id/1104>
- Safa'ah, S. N., & Indriyanti, A. D. (2023). Analisis Kepuasan Pengguna Layanan Aplikasi MyPertamina Berbasis Mobile Dengan Metode EUCS (End User Satisfaction with Computing). *Journal of Emerging Information Systems and Business Intelligence*, 4.
- Safitri, S. I., Suhery, C., & Bahri, S. (2021). Implementasi Algoritma K – Means Untuk Clustering Sentimen Pada Opini Kualitas Pelayanan Jasa Penerbangan. *Coding Jurnal Komputer Dan Aplikasi*, 9(02), 186. <https://doi.org/10.26418/coding.v9i02.47377>
- Sagala, S. (2013). *Manajemen strategik dalam peningkatan mutu pendidikan : pembuka ruang kreativitas, inovasi dan pemberdayaan potensi sekolah dalam sistem otonomi sekolah*. Alfabet.
- Saktyadi, A., & Tricahyono, D. (2024). Examining Customer Behavior Influences on MyPertamina App Adoption: A UTAUT2 Model Study in South Banten Gas Stations. *International Journal of Economics and Business Administration*, XII(Issue 2), 113–131. <https://doi.org/10.35808/ijeba/845>
- Simbolon, O. S., Manullang, M. E., Alvarez, S., Brutu, L. F. M., & Indra, E. (2023). Sentiment Analysis Of Mypertamina Application Using Support Vector Machine And Naïve Bayes Algorithms. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 7(1), 177–189. <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v7i1.4078>
- Siregar, A. M., & Puspabhuana, A. (2017). *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group.
- Soraya, A., & Wahyudi, A. D. (2021). Rancang Bangun Aplikasi Penjualan Dimsum Berbasis Web (Studi Kasus: Kedai Dimsum Soraya). *Jurnal Teknologi Dan Sistem Informasi (JTSTI)*, 2(4), 43–48. <http://jim.teknokrat.ac.id/index.php/JTSTI>
- Strzelecki, A. (2020). Application of Developers' and Users' Dependent Factors in App Store Optimization. *International Journal of Interactive Mobile Technologies (IJIM)*, 14(13), 91. <https://doi.org/10.3991/ijim.v14i13.14143>
- Syahfitri, N., Budianita, E., Nazir, A., & Afrianty, I. (2023). Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(3), 1668–1675. <https://doi.org/10.30865/klik.v4i3.1525>
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336, 012017. <https://doi.org/10.1088/1757-899X/336/1/012017>
- Teoli, D., Sanvictores, T., & An, J. (2025). *SWOT Analysis*.

- Tri Putra, K., Amin Hariyadi, M., & Crysdian, C. (2023). *Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm*. 1449.
- Tualang Ksatria, D., Yunefri, Y., & Lhaura Van, L. F. (2023). *Analisis Sentimen Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore menggunakan K-Nearest Neighbor dan Naïve Bayes* (Vol. 2, Issue 1).
- Vijay Gaikwad, S. (2014). Text Mining Methods and Techniques. In *International Journal of Computer Applications* (Vol. 85, Issue 17).
- Yudhistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 20–28. <https://doi.org/10.58602/jaiti.v1i1.22>
- Zuhri, K., & Saputri, N. A. O. (2020). Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier. *Journal of Computer and Information Systems Ampera*, 1(3), 185–199. <https://doi.org/10.51519/journalcisa.v1i3.45>

LAMPIRAN

Lampiran 1. Tampilan Web Google Play Store MyPertamina



Lampiran 2. Scrapping Data

```
[1] !pip install -qq google-play-scraper
```

50.2/50.2 kB 1.6 MB/s eta 0:00:00

```
[2] import json
import pandas as pd

import numpy as np
import matplotlib.pyplot as plt

from tqdm import tqdm

from pygments import highlight
from pygments.lexers import JsonLexer
from pygments.formatters import TerminalFormatter

from google_play_scraper import Sort, reviews, app
```

```
[3] from google.colab import drive
drive.mount('/content/drive', force_remount=True)
path = "gdrive/MyDrive/MyPertamina"
```

Mounted at /content/drive

```
[4] app_packages = ['com.dafturn.mypertamina']
```

```
[5] app_infos = []

for application in tqdm(app_packages):
    info = app(application, lang='id', country='id')
    del info['comments']
    app_infos.append(info)
```

100%|██████████| 1/1 [00:00<00:00, 5.47it/s]

```
[6] app_reviews=[]

for ap in tqdm(app_packages):
    for score in list(range(1,6)):
        for sort_order in [Sort.MOST_RELEVANT, Sort.NEWEST]:
            rvs, _ = reviews(
                ap,
                lang='id',
                country='id',
                sort=sort_order,
                count=2000 if score == 3 else 1000,
                filter_score_with=score
            )
            for r in rvs:
                r['sortOrder'] = 'most_relevant' if sort_order == Sort.MOST_RELEVANT else 'newest'
                r['appId'] = ap
            app_reviews.extend(rvs)
```

100%|██████████| 1/1 [00:13<00:00, 13.25s/it]

```
[7] app_reviews_df = pd.DataFrame(app_reviews)
```

```
[8] app_reviews_df = app_reviews_df.sort_values(["appId", "at"])
```

```
[9] from google.colab import drive
drive.mount('/content/drive', force_remount=True)
path = "/content/drive/MyDrive/MyPertamina/"
```

Mounted at /content/drive

```
[10] app_reviews_df.to_csv(path+'HasilReviewsAPP.csv', index=None, header=True)
```

Lampiran 3. Pemanggilan Data dan *Import Library*

```
[1] from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
[2] !pip install Sastrawi
```

Collecting Sastrawi
 Downloading Sastrawi-1.0.1-py2.py3-none-any.whl.metadata (909 bytes)
 Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
 209.7/209.7 kB 3.9 MB/s eta 0:00:00
 Installing collected packages: Sastrawi
 Successfully installed Sastrawi-1.0.1

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re
import nltk
nltk.download('punkt')
from nltk import word_tokenize
import sys
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

```

```

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.

```

```
[4] path = '/content/drive/My Drive/MyPertamina/'
```

```
[5] data1 = pd.read_csv(path + 'HasilReviewsAPP.csv')
data1
```

	reviewId	userName	userImage	content	score
	33310a75-5741-4da4-b714-a21707803b38	Pengguna Google	lh.googleusercontent.com/EGemol2N...	Ini mau scan struk..!tpi masuk aj susah...udah...	3
	c03f3b60-0ed9-49c9-86a7-86e7f1a356e4	Pengguna Google	lh.googleusercontent.com/EGemol2N...	Error tolong perbaiki... konten berkahenergi.	3
	cff1cd8b-afa7-41b1-839f-7512c229c2f3	Pengguna Google	lh.googleusercontent.com/EGemol2N...	Saya tidak dpt sms balik untuk kode otp,jd gk ...	3
	9b1bb46d-d23d-4e96-b487-2f321e81a2a7	Pengguna Google	lh.googleusercontent.com/EGemol2N...	Bagus buat Pertamina telah membuat apk My Pert...	4
	439a1eff-47e0-4a25-a8eb-b64cf28c4d79	Pengguna Google	lh.googleusercontent.com/EGemol2N...	Bagus aplikasi nya,. Semoga bermanfaat.. Tapi ...	5
	...	...	...	...	...
5	7b98515f-445d-43c5-9075-97f34a372b65	M Ilham	lh.googleusercontent.com/a/ACg8oc...	tolong di perbaiki lagi jangan dipersulit lagi	1

Lampiran 4. Pre-processing Data

```

[6] import numpy as np # Import numpy as np
data=pd.read_csv(path+'HasilReviewsAPP.csv', usecols = ['content','at'],)
data.sort_index(inplace=True)

```

```
[7] data1 = data.loc[5018:11999]
data1
```



	content	at
5018	2024 saya coba pakai lagi MyPertamina sudah cu...	2024-01-11 04:19:40
5019	Memudahkan	2024-01-11 04:25:16
5020	dalam sebulan lebih dari sekali update ga bagu...	2024-01-11 21:29:27
5021	Tiap Minggu minta update paksa nih apk	2024-01-11 22:45:33
5022	Masa dikit" update.. Fitur oke tp kl dikit" up...	2024-01-12 01:44:12
...	...	...
11995	tolong di perbaiki lagi jangan dipersulit lagi	2025-05-07 09:59:30
11996	bagus	2025-05-07 11:52:48
11997	barcode tidak tersimpan di aplikasi, payah	2025-05-07 12:04:59
11998	Kapan pembayaran kk BRI kurang responsif diban...	2025-05-07 13:26:02
11999	saya mau ambil kode QR tidak bisa selalu di bi...	2025-05-07 13:55:22

6982 rows × 2 columns

```
[8] data1 = data1.drop_duplicates(subset="content", inplace=False, keep = "last")
data1
```



	content	at
5018	2024 saya coba pakai lagi MyPertamina sudah cu...	2024-01-11 04:19:40
5019	Memudahkan	2024-01-11 04:25:16
5020	dalam sebulan lebih dari sekali update ga bagu...	2024-01-11 21:29:27
5021	Tiap Minggu minta update paksa nih apk	2024-01-11 22:45:33
5022	Masa dikit" update.. Fitur oke tp kl dikit" up...	2024-01-12 01:44:12
...	...	...
11995	tolong di perbaiki lagi jangan dipersulit lagi	2025-05-07 09:59:30
11996	bagus	2025-05-07 11:52:48
11997	barcode tidak tersimpan di aplikasi, payah	2025-05-07 12:04:59
11998	Kapan pembayaran kk BRI kurang responsif diban...	2025-05-07 13:26:02
11999	saya mau ambil kode QR tidak bisa selalu di bi...	2025-05-07 13:55:22

4703 rows × 2 columns

```

import os
import re
import json
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary

# Meliputi emotikon, duplikat data, URLS, tag
def remove_dirty(text):
    # Ekspresi reguler untuk mencocokkan karakter emotikon
    emoticon_pattern = re.compile("[
        u\"\\U0001F600-\\U0001F64F\" # emoticon umum
        u\"\\U0001F300-\\U0001F5FF\" # simbol & emoticon lainnya
        u\"\\U0001F680-\\U0001F6FF\" # emoticon transportasi & ikon
        u\"\\U0001F1E0-\\U0001F1FF\" # bendera (ios)
        u\"\\U00002500-\\U00002BEF\" # karakter CJK Extension A
        u\"\\U00002702-\\U000027B0\"
        u\"\\U00002702-\\U000027B0\"
        u\"\\U000024C2-\\U0001F251\"
        u\"\\U0001f926-\\U0001f937\"
        u\"\\U00010000-\\U0010ffff\"
        u\"\\u2640-\\u2642\"
        u\"\\u2600-\\u2B55\"
        u\"\\u200d\"
        u\"\\u23cf\"
        u\"\\u23e9\"
        u\"\\u231a\"
        u\"\\ufe0f\" # variasi penggunaan emoji
        u\"\\u3030\"

    \"]+", flags=re.UNICODE)

    return emoticon_pattern.sub(r'', text) # Menghapus karakter emotikon dari teks

```

```

# Fungsi untuk menghapus duplikat dari teks
def remove_duplicates(text):
    # Membagi teks menjadi baris-baris unik
    unique_lines = set(text.split('\n'))
    # Menggabungkan baris-baris unik kembali menjadi teks
    clean_text = '\n'.join(unique_lines)
    return clean_text

def case_folding(text):
    # Lowercase text
    text = text.lower()

    # Remove numbers
    text = re.sub(r'\d+', '', text)

    # Remove punctuation
    text = re.sub(r'^\w\s', '', text)

    # Remove whitespace
    text = text.strip()

    return text

```

```

def remove_stopword(text):
    factory = StopWordRemoverFactory()
    stopwords = factory.get_stop_words()
    stopword = factory.create_stop_word_remover()

    text = stopword.remove(text)
    return text

# menghapus slangword
def remove_slangword(inputstream):
    file_dict = "/kaggle/input/json-slangword/_json_colloquial-indonesian-lexicon.txt"
    with open(file_dict, 'r') as f:
        data = f.read()
        f.close()
    slangdict = json.loads(data)

    output = "stopandslangremoved.txt"

    with open(output, "w") as f:
        text = inputstream.splitlines()
        inputstream.close()

        for line in text:
            res = " ".join(slangdict.get(ele, ele) for ele in line.split())

            f.write(str(res) + "\n")

```

```

def stemming(text):
    # Inisialisasi stemmer dari Sastrawi
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()

    # stemmed_text = stemmer.stem(text)
    new_text = []
    textiterable = text.splitlines()
    for line in textiterable:
        if(line != ''):
            stemmed_text = stemmer.stem(line)
            new_text.append(stemmed_text)
        continue

    return "\n".join(new_text)

```

```

[10] fSlang = '/content/drive/My Drive/MyPertamina/Slangwords.txt'
      bahasa = 'id'
      sw = open(fSlang, encoding='utf-8', errors = 'ignore', mode = 'r');SlangS =sw.readlines();sw.close()
      SlangS = {slang.strip().split('|')[0]:slang.strip().split('|')[0]for slang in SlangS}

```

```
[11] Docs = []
df = open('/content/drive/My_Drive/MyPertamina/stopword_dokumen.csv', "r", encoding = "utf-8", errors='replace')
Docs.append(df.readlines());df.close()
stops= set([t.strip() for t in Docs[0]])
for x in ['my','pertamina']:
    stops.add(x)
```

```
[12] def formalndanstop(t):
    t=word_tokenize(t)
    for i,x in enumerate(t):
        if x in SlangS.keys():
            t[i]=SlangS[x]
    return''.join(' '.join(x for x in t if x not in stops))
```

```
[13] review=[x for x in data1['content']]
```

```
review_prepoc1=[]
for x in review:
    if isinstance(x, str):
        pattern = re.compile(r'http[s]?://(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[*\(\)])|(?:%[0-9a-fA-F][0-9a-fA-F]))+')
        pattern1 = re.compile(r'pic.twitter.com/(?:[a-zA-Z]|[0-9]|[$-_@.&+]|[*\(\)])|(?:%[0-9a-fA-F][0-9a-fA-F]))+')
        x = re.sub(pattern, ' ',x)
        x = re.sub(pattern1, ' ', x)
        x=x.lower()
        x=re.sub('((www\.[^\s]+)|(https?:?[^^\s]+))',' ',x)
        x=re.sub(r'[^\.,a-zA-Z0-9\n\.\.]', ' ',x)
        x=x.replace(' ','').replace('.', ' ')
        x=re.sub('@+', '',x)
        x=re.sub('[\s]+', ' ',x)
        x=re.sub(r'#([^\s]+)',r'\1',x)
        x=x.strip('\n')
        review_prepoc1.append(str(x))
review_prepoc1
```

['2024 saya coba pakai lagi mypertamina sudah cukup baik karena ada beberapa tambahan pilihan fitur pembayaran sayang bayar lewat my pertamina tolong pertamina perbaiki infrastruktur digital di setiap spbu biar kita lebih praktis bayar 'memudahkan', 'dalam sebulan lebih dari sekali update ga bagus ', 'tiap minggu minta update paksa nih apk', 'masa dikit update fitur oke tp kl dikit update nyusahin orang mw pake', 'kl update jangan langsung tiba2 tidak bisa dipakai aplikasinya kasih kesempatan 2 3 x update karena aplikasi dibuk'

```
[14] !pip install nltk # installing NLTK to ensure the latest version, for safety
import nltk
nltk.download('punkt_tab') # downloading the required punkt_tab to your environment
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt_tab.zip.
True
```

```

review_prepoc2=list(map(formaldanstop,review_prepoc1))
review_prepoc2

'ga masukin otp uda nyoba 10x',
'daftarin kartu debit gak eror wae',
'developer grafik bagus bug yg perbaiki',
'puas',
'udh coba bbrp kali msh gagal salah pin udah gua plototin tuh pin bnr msh aja salah susah hidup beli bensin aja kayak ngerampok',
'ribet susah log in',
'gua gak barcode nya salah ribet',
'mendaftar subsidi kendaraan roda 4 terbaca sistem kendaraan roda 2 tolong sistem aplikasi nya perbaiki aplikasi nya penggunaan',
'susah utk verifikasi aja gimana bagus pelayanan nya mengulang',
'mohon pembayaran tambahkan aplikasi dana',
'sdh lbh tingkatan',
'daftar pemberitahuan kesalahan pengisian form',
'aplikasi error pengisian form salah gimana sistemnya tololl',
'spbu pakai wilayah jakarta membantu pembayaran mypertamina',
'aplikasi sy sempurna n ribet org login sj mesti pake otp akun giliran sdh ngisi otp error kesalahan verifikasi otp klaim qr code klaim trjadi error data yg isi sesuai dgn data yg dipake pas daftar anak muda mudah mengurusnya gmn klo yg tua2',
'ribet masuk keaplikasinya',

```

```

from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

stopword = StopWordRemoverFactory().create_stop_word_remover()
stemmer=StemmerFactory().create_stemmer()

clean_review = []

for i,kalimat in enumerate (review_prepoc2):
    stop = stopword.remove(kalimat)
    stem=stemmer.stem(stop)
    print('loading kalimat ke:', i+1, 'dari', len(review_prepoc2))
    clean_review.append(stem)

```

```

loading kalimat ke: 4645 dari 4702
loading kalimat ke: 4646 dari 4702
loading kalimat ke: 4647 dari 4702
loading kalimat ke: 4648 dari 4702
loading kalimat ke: 4649 dari 4702
loading kalimat ke: 4650 dari 4702
loading kalimat ke: 4651 dari 4702
loading kalimat ke: 4652 dari 4702
loading kalimat ke: 4653 dari 4702
loading kalimat ke: 4654 dari 4702
loading kalimat ke: 4655 dari 4702
loading kalimat ke: 4656 dari 4702
loading kalimat ke: 4657 dari 4702
loading kalimat ke: 4658 dari 4702
loading kalimat ke: 4659 dari 4702
loading kalimat ke: 4660 dari 4702
loading kalimat ke: 4661 dari 4702
loading kalimat ke: 4662 dari 4702

```

```

clean_review

'ga masukin otp uda nyoba 10x',
'daftarin kartu debit gak eror wae',
'developer grafik bagus bug yg baik',
'puas',
'udh coba bbrp kali msh gagal salah pin udah gua plototin tuh pin bnr msh aja salah susah hidup beli bensin aja kayak ngerampok',
'ribet susah log in',
'gua gak barcode nya salah ribet',
'daftar subsidi kendara roda 4 baca sistem kendara roda 2 sistem aplikasi nya baik aplikasi nya guna',
'susah utk verifikasi aja gimana bagus layan nya ulang',
'mohon bayar tambah aplikasi dana',
'sdh lbh tingkat',
'daftar pemberitahuan salah isi form',
'aplikasi error isi form salah gimana sistem tololl',
'spbu pakai wilayah jakarta bantu bayar mypertamina',
'aplikasi sy sempurna n ribet org login sj mesti pake otp akun gilir sdh ngisi otp error salah verifikasi otp klaim qr code ribet trjadi error data yg isi sesuai dgn data yg dipake pas daftar anak muda mudah urus gmn klo yg tua2',
'ribet masuk aplikasi',
'daftar ribet apl',
'update data hilang barcode hilang',
'tingkat',
'baik isi data aja gak suruh muat ulang mulu',
'install isi gopay spbu alas mesin edc rusak mohon perhati spbunya terima kasih',
'apk nih kode otp masuk',
'bantu',
'sempurna aplikasi bbrp kurang aplikasi paksa',
'layan admin cepat',
'aplikasi klu daftar gk hrs kantor trdekat',
'qr nya tampil ui nya halaman profil ga input data lupa download jilang galery qr terus nya',

```

```
[19] hasil = pd.DataFrame(clean_review, columns=['Text_Bersih'])
      hasil.to_csv(path + 'Final_Data.csv', index=False)
```

```
▶ text_baru= pd.read_csv(path + 'Final_Data.csv')
   text_baru
```

 **Text_Bersih** 

0	2024 coba pakai mypertamina tambah pilih fitur...
1	mudah
2	bulan update ga bagus
3	minggu update paksa nih apk
4	dikit update fitur oke tp kl dikit update nyus...
...	...
4697	baik sulit
4698	bagus
4699	barcode simpan aplikasi payah
4700	bayar kk bri responsif banding gopay
4701	ambil kode qr bilang nik daftar tahun daftar k...

4702 rows × 1 columns

```
[21] text_baru.isnull().sum()
```

 **0**

Text_Bersih 57

dtype: int64

data_baru=text_baru.dropna()
data_baru

↔

	Text_Bersih
0	2024 coba pakai mypertamina tambah pilih fitur...
1	mudah
2	bulan update ga bagus
3	minggu update paksa nih apk
4	dikit update fitur oke tp kl dikit update nyus...
...	...
4697	baik sulit
4698	bagus
4699	barcode simpan aplikasi payah
4700	bayar kk bri responsif banding gopay
4701	ambil kode qr bilang nik daftar tahun daftar k...

4645 rows × 1 columns

Next steps: [Generate code with data_baru](#) [View recommended plots](#) [New interactive sheet](#)

```
[23] data_baru.isnull().sum()
```

↔

0

Text_Bersih 0

dtype: int64

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
vectorizer = CountVectorizer()
tf_transformer = TfidfTransformer()

datavector=vectorizer.fit_transform(data_baru['Text_Bersih'])
datafidf = tf_transformer.fit_transform(datavector)
datavector.shape
```

↔ (4645, 5177)

```
[25] Count_vect = CountVectorizer()
Count_vect.fit(data_baru['Text_Bersih'])
print(Count_vect.vocabulary_)
```

↔ : 3284, 'bbm': 587, 'pertamax': 3701, 'baik': 476, 'infrastruktur': 1941, 'digital': 1242,

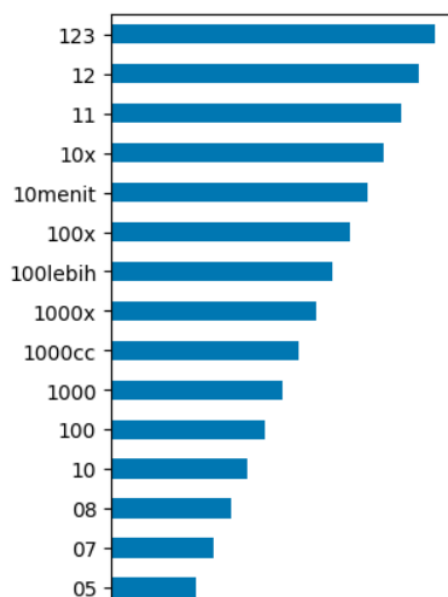
```

count_vectorizer = CountVectorizer()
count_vectorizer.fit(data_baru['Text_Bersih'])
dictionary = count_vectorizer.vocabulary_.items()
vocab = []
count = []
for key, value in dictionary:
    vocab.append(key)
    count.append(value)
vocab_bef_stem = pd.Series(count, index=vocab)
vocab_bef_stem = vocab_bef_stem.sort_values(ascending=True)
top_vacab = vocab_bef_stem.head(20)
top_vacab.plot(kind = 'barh', figsize=(3,7))

```



<Axes: >



```
pre_clustering = pd.DataFrame(count)
pre_clustering
```



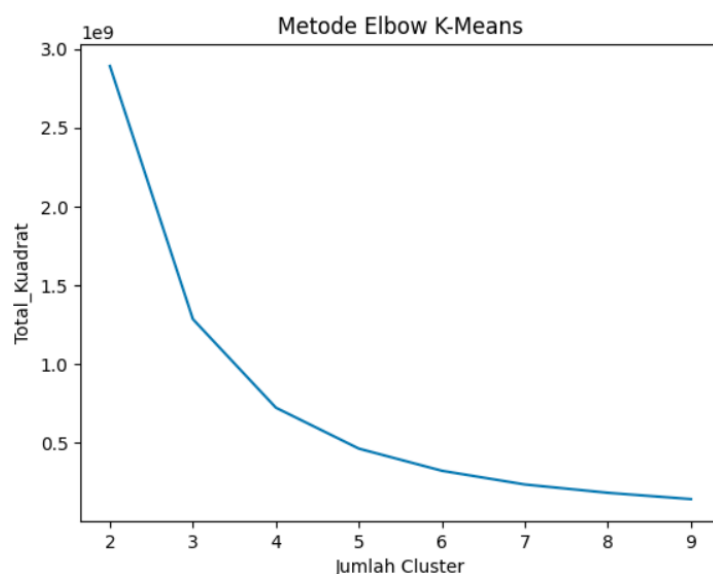
	0
0	57
1	1033
2	3527
3	3153
4	4594
...	...
5172	2202
5173	1699
5174	4288
5175	5055
5176	4547



5177 rows × 1 columns

Lampiran 5. *Elbow Method*

```
from sklearn.cluster import KMeans
wcss=[]
for i in range (2,10):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, random_state= 42)
    kmeans.fit(pre_clustering)
    wcss.append(kmeans.inertia_)
plt.plot(range (2,10), wcss)
plt.title('Metode Elbow K-Means')
plt.xlabel('Jumlah Cluster')
plt.ylabel('Total_Kuadrat')
plt.show()
```



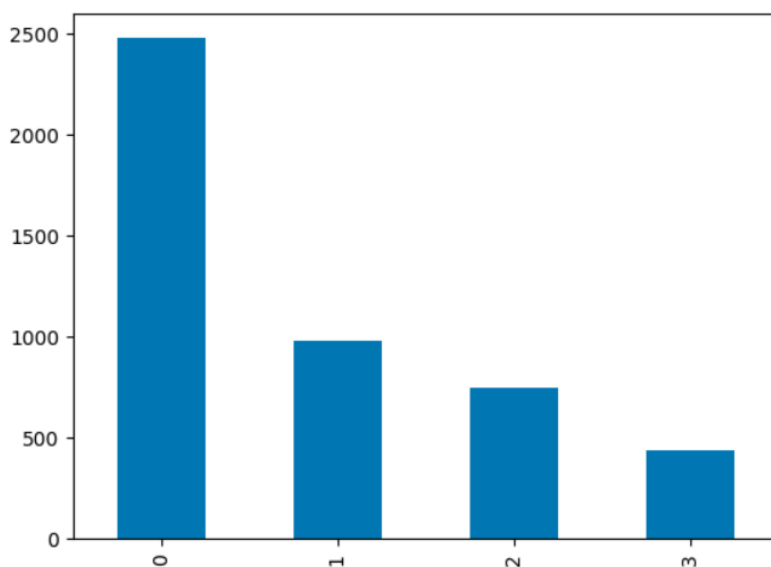
Lampiran 6. K-Means Clustering

```
[32] from sklearn.cluster import KMeans
      kmeans = KMeans(n_clusters=4, random_state=42).fit(datafidf)
      cluster = kmeans.labels_
      len(cluster)
```

↗ 4645

```
▶ hist=pd.value_counts(cluster)
  hist.plot(kind='bar')
```

↗ <ipython-input-33-6288d10c7559>:1: FutureWarning: pandas.value_counts is deprecated
hist=pd.value_counts(cluster)
<Axes: >



```
[34] data_baru['Cluster'] = cluster
      print(data_baru.shape)
      data_baru.head()
```

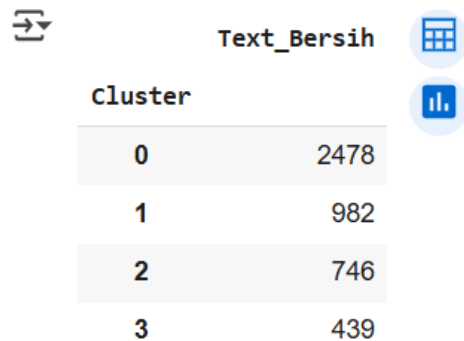
↗ (4645, 2)
<ipython-input-34-8100866e312a>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-d>
data_baru['Cluster'] = cluster

	Text_Bersih	Cluster
0	2024 coba pakai mypertamina tambah pilih fitur...	3
1	mudah	0
2	bulan update ga bagus	0
3	minggu update paksa nih apk	0
4	dikit update fitur oke tp kl dikit update nyus...	0

```
[35] output=pd.DataFrame(data_baru, columns=['Text_Bersih','Cluster'])
      output.to_csv(path+'Data_Cluster.csv', index=False)
```

```
[▶] data_baru.groupby('Cluster').count()
```



	Text_Bersih
Cluster	
0	2478
1	982
2	746
3	439

```
[37] kelas1 = data_baru[data_baru.Cluster==0]
      kelas2 = data_baru[data_baru.Cluster==1]
      kelas3 = data_baru[data_baru.Cluster==2]
      kelas4 = data_baru[data_baru.Cluster==3]
```

Lampiran 7. Pembentukan *Word Cloud*

```
[38] from wordcloud import WordCloud, STOPWORDS

comment_words = ''
stopwords = set(STOPWORDS)
mystopword = ['nya','yg','gk','gmn','tdk','ga','gak','udah','beli','banget','apk','pas','app','tp','hp','sih','kalo','klu','sj','agk']

for val in kelas1['Text_Bersih']:
    val = str(val)
    tokens = val.split()
    for i in range(len(tokens)):
        tokens[i] = tokens[i].lower()

    comment_words += " ".join([word for word in tokens if word not in mystopword])+" "

wordcloud = WordCloud(width=800, height= 800,
                      background_color = 'white',
                      stopwords = stopwords,
                      min_font_size=10).generate(comment_words)
```

```
[▶] from collections import Counter

id_words= comment_words.split()
Counter(id_words).most_common(10)
```

```
[▶] plt.figure(figsize=(6, 6), facecolor="Green")
      plt.imshow(wordcloud)
      plt.axis("off")
      plt.tight_layout(pad=0)
      plt.show()
```

