

Analisis Sentimen Menggunakan Algoritma *Logistic Regression* dan LDA(*Latent Dirichlet Allocation*) pada Ulasan Aplikasi PLN Mobile Di Play Store

1st Ranya Fauzi Bawazir

Department Of Telecommunication
Engineering
Telkom University
Bandung, Indonesia
ranyafauzibawazir@student.telkomuniversity.ac.id

2nd Rita Purnamasari

Department Of Telecommunication
Engineering
Telkom University
Bandung, Indonesia
ritapurnamasari@student.telkomuniversity.ac.id

3rd Yulinda Eliskar

Department Of Telecommunication
Engineering
Telkom University
Bandung, Indonesia
yulindaeliskar@student.telkomuniversity.ac.id

Abstrak — Pertumbuhan ekosistem digital mendorong peningkatan volume ulasan aplikasi di platform distribusi perangkat lunak, termasuk PLN Mobile milik PT PLN (Persero). Penelitian ini bertujuan menganalisis sentimen dan mengidentifikasi topik dominan pada ulasan PLN Mobile tahun 2022. Data sebanyak 134.325 ulasan diperoleh melalui web scraping dari Google Play Store, kemudian difilter menjadi 40.043 ulasan relevan. Tahapan meliputi pelabelan sentimen berdasarkan skor ulasan, pra-pemrosesan teks, pemisahan data, serta pembangunan model analisis sentimen menggunakan *Logistic Regression*. Model dilatih selama 15 *epoch* dengan representasi fitur TF-IDF dan menghasilkan akurasi 86%, *Precision* 98%, *Recall* 85%, dan *F1-Score* 91%. Selain itu, metode *Latent Dirichlet Allocation* (LDA) digunakan untuk pemodelan topik dan menghasilkan empat topik utama: Transaksi, Layanan, Aplikasi, dan Pengaduan. Hasil integrasi kedua model menunjukkan proporsi sentimen positif yang dominan pada semua topik, dengan kisaran 85,9%–86,3%. Temuan ini memberikan wawasan strategis bagi PT PLN (Persero) untuk meningkatkan kualitas layanan, kinerja aplikasi, dan respons terhadap pengaduan pelanggan.

Kata kunci — *Analisis sentimen, Logistic Regression, Latent Dirichlet Allocation, pln mobile.*

I. PENDAHULUAN

Pertumbuhan ekosistem digital menyebabkan *volume* data daring meningkat secara signifikan, termasuk pada ulasan aplikasi di platform distribusi perangkat lunak. PT PLN (Persero) melalui aplikasi PLN *Mobile* memfasilitasi pelanggan untuk mengakses layanan kelistrikan. Namun, tingginya jumlah ulasan pengguna di *Google Play Store* pada tahun 2022 menghasilkan big data yang sulit dianalisis secara manual.

Analisis sentimen berbasis machine learning menjadi salah satu pendekatan yang efektif untuk memproses data

tersebut, memungkinkan klasifikasi ulasan menjadi sentimen positif atau negatif [1]. Pendekatan serupa juga telah banyak digunakan di berbagai bidang untuk memahami pola, tren, dan persepsi pengguna, seperti dalam kajian pemasaran perpustakaan melalui systematic review dan analisis tren [2].

Salah satu algoritma yang banyak digunakan adalah *Logistic Regression*, yang sederhana, efisien, dan memiliki interpretabilitas tinggi pada kasus klasifikasi biner [3]. Berbagai penelitian lain juga membuktikan efektivitas supervised learning pada domain berbeda, misalnya klasifikasi tutupan lahan menggunakan citra satelit dengan Google Earth Engine [4].

Selain mengetahui polaritas sentimen, penting pula untuk mengidentifikasi topik pembahasan yang dominan pada ulasan. *Latent Dirichlet Allocation* (LDA) digunakan sebagai metode topic modeling untuk mengelompokkan dokumen ke dalam topik-topik utama secara unsupervised [5]. Kombinasi *Logistic Regression* dan LDA tidak hanya menyediakan informasi mengenai sentimen, tetapi juga memberikan konteks pembahasan, yang dapat dimanfaatkan sebagai insight bisnis strategis oleh PT PLN (Persero). Penelitian ini memanfaatkan data ulasan PLN *Mobile* selama satu tahun, menggunakan pendekatan pre-processing teks, pembelajaran mesin, serta visualisasi hasil. Pemodelan sentimen dilakukan dengan *Logistic Regression*, sedangkan pemodelan topik menggunakan LDA. Kinerja metode diukur menggunakan metrik evaluasi seperti akurasi, presisi, *Recall*, dan *F1-Score*. Hasil yang diperoleh diharapkan dapat mendukung pengambilan keputusan strategis untuk peningkatan layanan.

II. KAJIAN TEORI

A. PT PLN (Persero)

PT PLN (Persero) adalah badan usaha milik negara yang memegang mandat penyediaan listrik di seluruh Indonesia. PLN bertanggung jawab atas perencanaan, pembangunan, pengoperasian, dan pemeliharaan infrastruktur kelistrikan, serta memastikan ketersediaan energi listrik yang andal, terjangkau, dan berkelanjutan. Dalam beberapa tahun terakhir, PLN gencar melakukan transformasi digital untuk meningkatkan kualitas layanan dan efisiensi operasional [7].

B. Aplikasi PLN Mobile

PLN Mobile merupakan aplikasi resmi dari PT PLN (Persero) yang dirancang untuk mempermudah interaksi pelanggan dengan layanan perusahaan. Fitur utama aplikasi ini meliputi pengecekan tagihan, pembayaran online, pelaporan gangguan, pengajuan pasang baru, dan pelacakan progres layanan. Keberadaan PLN Mobile diharapkan dapat meningkatkan pengalaman pelanggan melalui pelayanan yang cepat, efisien, dan transparan [8].

C. Natural Language Processing (NLP)

NLP adalah cabang dari kecerdasan buatan (artificial intelligence) yang berfokus pada interaksi antara komputer dan bahasa manusia. Bidang ini mencakup berbagai teknik untuk memproses dan menganalisis bahasa alami, mulai dari tokenisasi, stemming, lemmatization, hingga analisis semantik. Dalam penelitian ini, NLP digunakan untuk mengolah teks ulasan pelanggan sebelum dilakukan analisis sentimen dan pemodelan topik [9].

D. Analisis Sentimen

Analisis sentimen adalah metode komputasional yang bertujuan untuk mengidentifikasi dan mengklasifikasikan pendapat atau emosi yang terkandung dalam teks. Pendekatan ini bermanfaat untuk memahami persepsi publik terhadap produk, layanan, atau kebijakan tertentu. Analisis sentimen banyak digunakan dalam manajemen reputasi, riset pasar, dan pengambilan keputusan strategis [1].

E. Logistic Regression

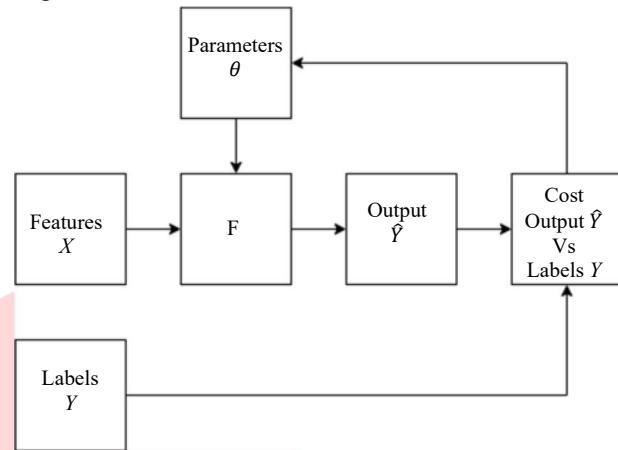
Logistic Regression adalah algoritma pembelajaran mesin yang digunakan untuk memprediksi probabilitas suatu peristiwa biner, seperti ya/tidak, benar/salah, atau setuju/tidak setuju [6]. Berbeda dengan regresi linier yang menghasilkan output tanpa batas, *Logistic Regression* menggunakan fungsi sigmoid untuk membatasi output pada rentang 0 hingga 1. Algoritma supervised learning ini juga telah terbukti efektif diterapkan pada berbagai domain lain, misalnya pada klasifikasi tutupan lahan menggunakan citra satelit dengan *Google Earth Engine* [3]. Fungsi hipotesisnya dapat dituliskan sebagai:

$$h(X) = \frac{1}{1 + e^{-X}} \quad (1)$$

Keterangan variabel:

- $h(X)$ = merupakan nilai output dari model, yang diinterpretasikan sebagai probabilitas bahwa suatu data termasuk ke dalam kelas positif.
- (X) = merupakan hasil dari kombinasi linier antara parameter model dan fitur input, yang dirumuskan sebagai: $X = \beta_0 + \beta_1 z_1 + \dots + \beta_n z_n$.

- β_0 = intersep (bias) dari model.
- $\beta_0 + \beta_1 z_1 + \dots + \beta_n z_n$ = koefisien regresi yang merepresentasikan bobot masing fitur.
- $z_1 + z_2 + \dots + z_n$ = fitur-fitur input dari data.
- e = bilangan eksponensial (sekitar 2.718), basis dari logaritma natural.



GAMBAR 1
Transformasi Fungsi $F(\cdot)$

Pada gambar 2.1 menunjukkan transformasi dari Fungsi $F(\cdot)$ yang beroperasi pada parameter θ dan vector fitur $X^{(i)}$ untuk menghasilkan output \hat{Y} yang diketahui juga sebagai fungsi hipotesis [6].

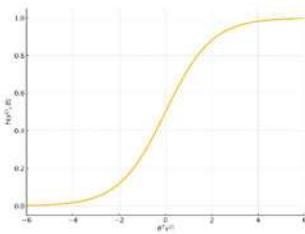
Secara spesifik, $h(x^{(i)}; \theta)$ merepresentasikan fungsi hipotesis, di mana indeks i mengacu pada pengamatan atau data ke- i . Dalam *Logistic Regression*, fungsi *sigmoid* ini mengubah input menjadi nilai keluaran dalam rentang (0,1) yang dapat diinterpretasikan sebagai probabilitas [6]. Oleh karena itu, bentuk umum dari fungsi regresi logistik dapat dituliskan sebagai berikut:

$$h(x^{(i)}; \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}} \quad (2)$$

Keterangan variabel:

- $h(x^{(i)}; \theta)$ = fungsi hipotesis yang memberikan probabilitas keluaran untuk data ke- i .
- $x^{(i)}$ = vektor fitur dari data ke- i .
- θ = vektor parameter model yang berisi bobot dan bias.
- $\theta^T x^{(i)}$ = perkalian titik (*dot product*) antara parameter model dan vektor fitur.

Secara lebih rinci, Gambar 2 memperlihatkan representasi visual dari fungsi sigmoid, di mana nilainya akan mendekati 0 apabila hasil perkalian $\theta^T x^{(i)}$ menuju ke arah $-\infty$ dan akan semakin mendekati 1 apabila perkalian $\theta^T x^{(i)}$ menuju ke arah ∞ [6].



GAMBAR 2
Sigmoid Function

Dalam proses klasifikasi, biasanya digunakan sebuah ambang batas (*threshold*) untuk menentukan hasil prediksi. Secara umum, nilai threshold ini ditetapkan sebesar 0.5, yang setara dengan kondisi ketika hasil dari perkalian $\theta^T x^{(i)}$ bernilai 0. Berdasarkan nilai tersebut, prediksi klasifikasi ditentukan sebagai berikut:

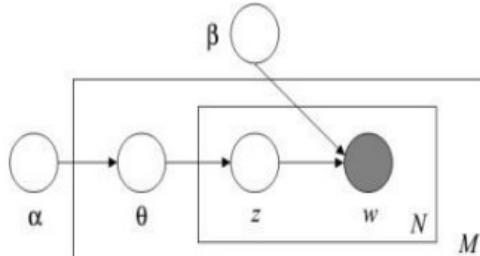
- Prediksi klasifikasi bernilai 1 jika $\theta^T x^{(i)} \geq 0$.
- Prediksi klasifikasi bernilai 0 jika $\theta^T x^{(i)} < 0$.

Dengan demikian, prediksi bernilai 1 diinterpretasikan sebagai kelas positif, sedangkan nilai 0 merepresentasikan kelas negatif.

F. Topic Modeling Latent Dirichlet Allocation

Topic modeling adalah metode unsupervised learning yang bertujuan menemukan struktur tersembunyi berupa kumpulan topik dalam korpus teks. Menurut Blei et al., komponen utama dalam topic modeling mencakup “kata”, “dokumen”, dan “korpus” [10]. Topik direpresentasikan sebagai distribusi kata dengan probabilitas tertentu, dan sebuah dokumen dapat mengandung beberapa topik dengan proporsi berbeda.

Latent Dirichlet Allocation (LDA) adalah salah satu algoritma topic modeling yang memodelkan dokumen sebagai campuran topik, dan setiap topik sebagai distribusi kata [10]. Parameter utamanya adalah:



GAMBAR 3
Diagram Model LDA

Keterangan variabel:

- α = parameter untuk distribusi topic yang berada dalam tahapan corpus. Berfungsi untuk memutuskan distribusi topik pada sebuah dokumen. Apabila nilai α pada suatu dokumen semakin besar maka menandakan semakin banyak campuran topik di dalam dokumen tersebut [10].
- β = parameter untuk mendapatkan nilai distribusi kata pada suatu topik. Semakin besar nilai dari β artinya makin banyak kata yang menyusun suatu topik [10].

$$\Phi_k \sim \text{Dirichlet}(\beta) \quad (3)$$

- θ = variabel yang berada pada tingkat dokumen. Semakin tinggi nilai θ menandakan dokumen-dokumen tersebut memiliki topik yang banyak [10].

$$\theta_d \sim \text{Dirichlet}(\alpha) \quad (4)$$

- zN = variabel ini merepresentasikan suatu topik dari sebuah kata di dalam dokumen [10].

$$zd, n \sim \text{Multinomial}(\theta d) \quad (5)$$

- wN = variabel yang digunakan untuk merepresentasikan kata yang terkait kepada suatu topik tertentu di dalam dokumen [10].

$$wd, n \sim \text{Multinomial}(\Phi zd, n) \quad (6)$$

Proses generatif pada model LDA seperti yang ditunjukkan pada Gambar 2.6 dapat diformulasikan secara matematis sebagai berikut:

$$P(W, Z, \theta, \varphi; \alpha, \beta) = \prod_{j=1}^M P(\theta_j; \alpha) \prod_{i=1}^K P(\varphi_i; \beta) \prod_{t=1}^N P(Z_{j,t} | \theta_j) P(W_{j,t} | \varphi_{j,t}) \quad (7)$$

Keterangan variabel:

- $P(W, Z, \theta, \varphi; \alpha, \beta)$ = probabilitas gabungan dari kata W , topik Z , distribusi topik per dokumen θ , distribusi kata per topik φ , dan parameter hyper α, β .

III. METODE

A. Scraping Data

Data ulasan aplikasi PLN Mobile diperoleh dari kolom komentar Google Play Store menggunakan metode web scraping dengan library *google_play_scraper* pada *Google Colab*. Proses dimulai dengan menghubungkan *Google Colab* ke Google Drive, menginstal library yang diperlukan, dan menentukan parameter seperti ID aplikasi (*com.icon.pln123*), bahasa (*lang=id*), dan negara (*country=id*). Ulasan yang berhasil diambil kemudian disimpan dalam format .csv di *Google Drive*. Hasil *scraping* menghasilkan total 134.325 entri ulasan yang digunakan untuk tahap analisis berikutnya.

B. Labeling Data

Proses filter dilakukan untuk mengambil ulasan aplikasi PLN Mobile sepanjang tahun 2022 dengan variabel yang relevan, yaitu *content*, *at*, dan *score*. Tahapannya meliputi penghapusan kolom yang tidak diperlukan, pemilihan data berdasarkan rentang tanggal 1 Januari – 31 Desember 2022, dan penghapusan kolom tanggal setelah penyaringan. Data hasil *filter* kemudian disimpan dalam format .csv. Proses ini menghasilkan 40.043 ulasan yang siap digunakan pada tahap analisis.

C. Pra-Processing Data

Tahap pra-pemrosesan dilakukan untuk mengubah dataset ulasan yang tidak terstruktur menjadi format terstruktur agar siap digunakan dalam analisis. Proses ini mencakup:

- *Data Cleansing* - menghapus tanda baca, simbol, dan karakter yang tidak diperlukan;
- *Case Folding* - mengubah seluruh teks menjadi huruf kecil dan menghapus karakter selain a-z;
- *Tokenizing* - memisahkan teks menjadi unit kata;

- Stopword Removal dan Normalisasi - menghapus kata-kata umum yang tidak relevan dan mengganti kata tidak baku menjadi bentuk baku sesuai kamus normalisasi.

D. Data Splitting

Dataset dibagi menjadi data latih dan data uji dengan perbandingan 70:30 menggunakan fungsi `train_test_split` dari `scikit-learn`. Parameter `stratify` digunakan untuk menjaga distribusi label sentimen seimbang antara kedua subset. Data fitur (X) berupa teks ulasan yang telah melalui tahap pra-pemrosesan, sedangkan label (y) adalah hasil pelabelan berdasarkan skor ulasan (positif ≥ 3 , negatif < 3).

E. Pembangunan Model

Model *Logistic Regression* dibangun menggunakan pustaka `scikit-learn` dengan pendekatan supervised learning untuk klasifikasi biner (positif dan negatif). Data teks diubah menjadi representasi numerik menggunakan `TfidfVectorizer` dengan `ngram_range` (1, 2) dan batas maksimal 5000 fitur. Model diinisialisasi dengan `class_weight='balanced'` untuk menangani ketidakseimbangan kelas, `max_iter=1` dan `warm_start=True` untuk memungkinkan pelatihan bertahap per *epoch*. Proses pelatihan dilakukan selama 15 *epoch*, dengan pemantauan akurasi pada setiap interval 5 *epoch*.

F. Training Model

Model *Logistic Regression* dilatih selama 15 *epoch* menggunakan data vektor TF-IDF hasil pra-pemrosesan. Pelatihan dilakukan secara bertahap dengan konfigurasi `warm_start=True` dan `max_iter=1`, sehingga bobot diperbarui satu iterasi per *epoch* untuk meniru proses pembelajaran bertahap seperti pada deep learning. Evaluasi dilakukan di setiap *epoch* dengan menghitung *akurasi*, *presisi*, *Recall*, dan *F1-Score* pada data latih dan uji. Berikut ditampilkan tabel hasil pada setiap *epoch*. Pada tabel tersebut, digunakan singkatan sebagai berikut: Epoch (E), Recall (R), Precision (P), F1-Score (F1), dan Accuracy (A):

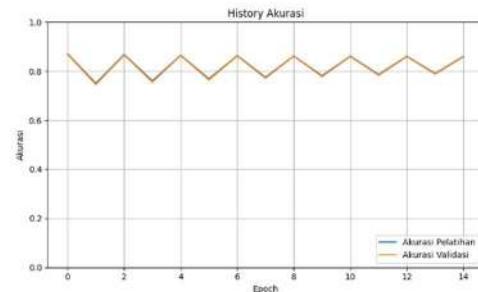
TABEL 1

Data Training dan Data Validasi Logistic Regression.

E	Data Training				Data Validasi			
	R	P	F1	A	R	P	F1	A
1	87%	98%	92%	87%	87%	98%	92%	87%
2	71%	99%	83%	75%	71%	99%	83%	74%
3	86%	98%	92%	87%	86%	98%	92%	87%
4	73%	99%	84%	76%	72%	72%	84%	76%
5	86%	98%	91%	87%	86%	98%	92%	86%
6	74%	99%	85%	77%	74%	99%	84%	76%
7	86%	87%	92%	86%	86%	98%	92%	86%
8	75%	99%	85%	78%	74%	99%	85%	77%
9	86%	98%	91%	86%	85%	98%	91%	86%
10	75%	99%	86%	78%	75%	99%	85%	78%
11	85%	98%	91%	86%	85%	98%	91%	86%
12	76%	99%	86%	79%	76%	99%	86%	78%
13	85%	98%	91%	86%	85%	98%	91%	86%
14	77%	99%	86%	79%	76%	99%	86%	79%
15	85%	98%	91%	86%	85%	98%	91%	86%

Hasil menunjukkan pola akurasi yang stabil dengan fluktuasi periodik. Akurasi tertinggi tercapai pada *epoch* tertentu (1, 3, 5, 7, 9, 11, 13, 15) dengan nilai sekitar 86–87% untuk data latih dan uji, sedangkan pada *epoch* lainnya sedikit menurun ke kisaran 75–79%. Performa kedua subset relatif

seimbang, menunjukkan model mampu mempertahankan generalisasi.



GAMBAR 4
Hasil Training dan Accuracy Logistic Regression

Grafik hasil pelatihan menunjukkan perkembangan akurasi model *Logistic Regression* selama 15 *epoch* pada data latih dan uji. Kedua kurva memiliki pola mirip dan saling berdekatan, mengindikasikan performa yang seimbang pada kedua subset data. Nilai akurasi berfluktuasi antara 77% – 86% akibat konfigurasi `warm_start=True` dan `max_iter=1` yang membatasi pembaruan bobot hanya satu iterasi per *epoch*. Meskipun demikian, tren akurasi tetap stabil pada level tinggi, menunjukkan kemampuan model dalam mempelajari pola data dan mempertahankan generalisasi.

G. Topic Modeling

Latent Dirichlet Allocation (LDA) digunakan untuk mengidentifikasi topik utama dalam ulasan PLN Mobile. Proses dimulai dengan pra-pemrosesan data, meliputi data *cleansing* (menghapus teks tidak relevan), *tokenizing* (memecah teks menjadi kata), *POS tagging* (mengambil kata penting seperti kata benda, kerja, sifat, dan keterangan), serta *lemmas* (menggabungkan kembali kata terpilih menjadi teks siap olah).

Evaluasi model diawali dengan membentuk bigram dan trigram menggunakan *Gensim*, lalu membuat corpus BoW (*Bag-of-Words*) sebagai input LDA. Model dilatih menggunakan *LdaMulticore* dengan parameter 5 topik, `passes=8`, `iterations=500`, dan pengaturan alpha serta eta otomatis. Keluaran model dianalisis melalui kata kunci tiap topik dan dua metrik: *perplexity* (makin kecil makin baik) dan *Coherence Score* (0–1, makin tinggi makin baik).

Hasil evaluasi menunjukkan *Coherence Score* tertinggi diperoleh pada 4 topik, yaitu Transaksi, Aplikasi, Layanan, dan Pengaduan, sehingga jumlah topik ini dipilih sebagai representasi terbaik struktur semantik ulasan.

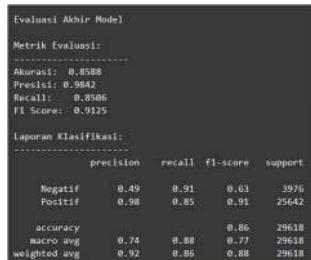
IV. HASIL DAN PEMBAHASAN

A. Pengujian Data Logistic Regression

Dataset ulasan yang telah dilabeli (positif ≥ 3 , negatif < 3) dibagi menjadi data latih dan uji dengan rasio 70:30. Fitur diekstraksi menggunakan TF-IDF *Vectorizer* (maksimal 5000 fitur, `ngram_range` (1, 2)), kemudian model *Logistic Regression* diinisialisasi dengan `class_weight='balanced'` dan `warm_start=True` untuk pelatihan bertahap selama 15 *epoch*. Performa dipantau melalui metrik akurasi pada setiap *epoch*.



GAMBAR 5
Visual Confusion Matrix Logistic Regression



GAMBAR 6
Hasil Pengujian Logistic Regression

Evaluasi model terhadap data uji menggunakan metrik *accuracy*, *Precision*, *Recall*, dan *F1-Score*, serta visualisasi confusion matrix. Hasil menunjukkan nilai *Recall* 85%, *Precision* 98%, *F1-Score* 91%, dan akurasi 85%, menandakan model mampu mengklasifikasikan sentimen ulasan dengan performa tinggi dan seimbang pada kedua kelas.

B. Pengujian Topic Modeling

Pengujian dilakukan dengan memprediksi topik dominan pada setiap ulasan menggunakan metode *dominant topic*. Setiap ulasan sebenarnya mengandung semua topik, namun topik dengan nilai probabilitas tertinggi ditetapkan sebagai label utama. Proses ini dilakukan dengan memetakan nilai pada kolom *Dominant_Topic* ke daftar topik terdefinisi. Hasil menunjukkan bahwa model mampu mengidentifikasi topik dominan secara konsisten sesuai distribusi kata pada ulasan.

Document	Topic	Per. (Probabilitas)	Topic	Per. (Probabilitas)
0	0	0.5157	ulik, nyek, bantul, pah, sepihak, ya, tan, neg...	
1	0	0.5777	ulik, nyek, bantul, pah, sepihak, ya, tan, neg...	
2	0	0.4713	ulik, nyek, bantul, pah, sepihak, ya, tan, neg...	
3	7	0.8555	masih, bantul, sepihak, sepihak, bantul, pah...	
4	0	0.7720	ulik, nyek, bantul, pah, sepihak, ya, tan, neg...	
5	2	0.8069	masih, bantul, sepihak, sepihak, bantul, pah...	
6	3	0.8114	masih, bantul, sepihak, sepihak, bantul, pah...	
7	2	0.8019	masih, bantul, sepihak, sepihak, bantul, pah...	
8	0	0.8559	ulik, nyek, bantul, pah, sepihak, ya, tan, neg...	
9	2	0.5168	masih, bantul, sepihak, sepihak, bantul, pah...	
10	7	0.3749	masih, bantul, sepihak, sepihak, bantul, pah...	
11	9	0.5249	masih, bantul, sepihak, sepihak, bantul, pah...	
12	1	0.4559	cepat, pelayanan, bantul, sepihak, sepihak, ya...	
13	3	0.4513	masih, bantul, sepihak, sepihak, bantul, pah...	
14	0	0.5884	ulik, nyek, bantul, pah, sepihak, ya, tan, neg...	

GAMBAR 7
Hasil Dominant Topic

Pengujian dilakukan dengan memprediksi topik dominan pada setiap ulasan menggunakan metode *dominant topic*. Setiap ulasan sebenarnya mengandung semua topik, namun topik dengan nilai probabilitas tertinggi ditetapkan sebagai label utama. Proses ini dilakukan dengan memetakan nilai pada kolom *Dominant_Topic* ke daftar topik terdefinisi. Hasil menunjukkan bahwa model mampu mengidentifikasi topik dominan secara konsisten sesuai distribusi kata pada ulasan.

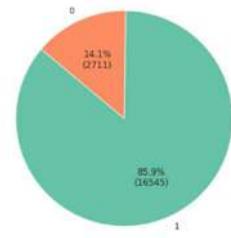
C. Perbandingan Sentimen

Hasil klasifikasi LDA menghasilkan empat topik utama: Transaksi, Layanan, Aplikasi, dan Pengaduan. Analisis sentimen menunjukkan proporsi sentimen positif yang tinggi pada seluruh topik:

- Topik Transaksi

Distribusi sentimen pada topik Transaksi ditunjukkan pada Gambar 8. Dari total 19.256 ulasan, 85,9% (16.545 ulasan) memiliki sentimen positif dan 14,1% (2.711 ulasan) negatif. Persentase positif yang tinggi mengindikasikan bahwa sebagian besar pengguna merasa puas terhadap proses transaksi di PLN Mobile, yang kemungkinan dipengaruhi oleh kemudahan penggunaan, kecepatan, dan keamanan layanan. Meski demikian, proporsi sentimen negatif menunjukkan adanya masalah tertentu yang masih dialami sebagian pengguna, seperti kegagalan transaksi atau kendala teknis lainnya.

Perbandingan Sentimen untuk Topik: TRANSAKSI

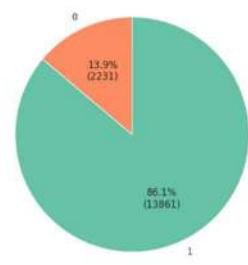


GAMBAR 8
Perbandingan Sentimen Positif dan Negatif Topik Transaksi

- Topik Layanan

Dari 16.092 ulasan, 86,1% (13.861) bersentimen positif dan 13,9% (2.231) bersentimen negatif. Mayoritas penilaian positif menggambarkan kualitas layanan yang memuaskan, termasuk pelayanan pelanggan yang responsif dan kemudahan akses informasi. Namun, ulasan negatif menandakan perlunya peningkatan pada kecepatan respons atau kualitas solusi yang diberikan.

Perbandingan Sentimen untuk Topik: LAYANAN



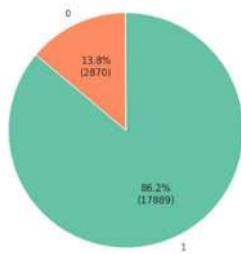
GAMBAR 9
Perbandingan Sentimen Positif dan Negatif Topik Layanan

- Topik Aplikasi

Topik Aplikasi didominasi sentimen positif sebesar 86,2% (17.889) dari total 20.759 ulasan, sedangkan 13,8% (2.870) bersentimen negatif. Penilaian positif dipengaruhi oleh antarmuka yang ramah pengguna, stabilitas aplikasi, dan fitur yang mendukung kebutuhan pelanggan.

Ulasan negatif biasanya berkaitan dengan bug, gangguan teknis, atau masalah kompatibilitas perangkat.

Perbandingan Sentimen untuk Topik: APLIKASI



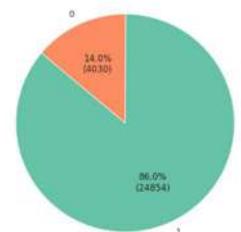
GAMBAR 10

Perbandingan Sentimen Positif dan Negatif Topik Aplikasi

- Topik Pengaduan

Dari 28.884 ulasan, 86,0% (24.854) memiliki sentimen positif, sedangkan 14,0% (4.030) negatif. Meski pengaduan identik dengan keluhan, tingginya proporsi positif mengindikasikan kepuasan pengguna terhadap respons dan penyelesaian masalah oleh PLN Mobile. Sentimen negatif menggarisbawahi perlunya peningkatan konsistensi dalam penanganan kasus.

Perbandingan Sentimen untuk Topik: PENGADUAN



GAMBAR 11

Perbandingan Sentimen Positif dan Negatif Topik Pengaduan

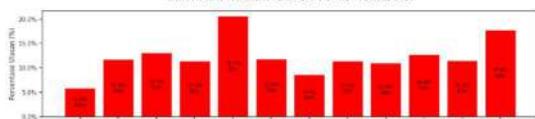
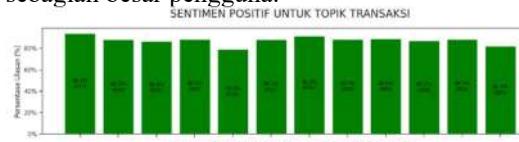
D. Analisis Tren Bulanan

Hasil klasifikasi LDA menghasilkan empat topik utama, yaitu Transaksi, Layanan, Aplikasi, dan Pengaduan. Analisis sentimen terhadap masing-masing topik menunjukkan bahwa proporsi sentimen positif secara umum lebih dominan dibandingkan sentimen negatif. Meskipun terdapat fluktuasi pada bulan-bulan tertentu, terutama di bulan Mei yang menjadi titik dengan sentimen negatif tertinggi pada hampir semua topik, tren sentimen positif tetap konsisten mendominasi sepanjang tahun 2022. Hal ini menggambarkan bahwa pengalaman pengguna PLN Mobile secara keseluruhan dinilai baik, baik dalam aspek transaksi, layanan, kinerja aplikasi, maupun penanganan pengaduan. Berikut akan di tunjukkan hasil untuk setiap topik:

- Topik Transaksi

Tren sentimen positif untuk topik Transaksi relatif stabil sepanjang tahun 2022, meski mengalami beberapa fluktuasi. Puncak sentimen positif tercatat pada Januari (94,1%) dan terendah pada Mei (79,3%), yang juga menjadi bulan dengan sentimen negatif tertinggi (20,7%). Setelah itu, tren positif kembali naik hingga 91,3% di Juli sebelum

sedikit menurun menjelang akhir tahun. Secara umum, ulasan positif tetap mendominasi, menandakan pengalaman transaksi dinilai baik oleh sebagian besar pengguna.



GAMBAR 12

Grafik Analisis Tren Bulanan Topik Transaksi

- Topik Layanan

Tren sentimen positif pada topik Layanan yang cenderung tinggi dan stabil. Nilai tertinggi terjadi pada Januari (94,2%) dan Agustus (94,3%), sementara titik terendah ada di Mei (82,5%) yang juga diiringi sentimen negatif tertinggi (17,5%). Walaupun terdapat beberapa penurunan, dominasi ulasan positif sepanjang tahun menunjukkan layanan PLN Mobile umumnya dinilai memuaskan.

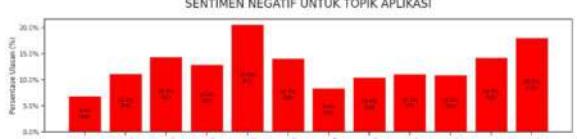


GAMBAR 13

Grafik Analisis Tren Bulanan Topik Layanan

- Topik Layanan

Tren sentimen positif pada topik Layanan yang cenderung tinggi dan stabil. Nilai tertinggi terjadi pada Januari (94,2%) dan Agustus (94,3%), sementara titik terendah ada di Mei (82,5%) yang juga diiringi sentimen negatif tertinggi (17,5%). Walaupun terdapat beberapa penurunan, dominasi ulasan positif sepanjang tahun menunjukkan layanan PLN Mobile umumnya dinilai memuaskan.



GAMBAR 14

Grafik Analisis Tren Bulanan Topik Aplikasi

- Topik Layanan

Tren sentimen positif pada topik Pengaduan (Gambar 5.17) memperlihatkan dinamika yang cukup variatif. Januari mencatatkan nilai tertinggi (93,6%), sedangkan Mei menjadi bulan dengan sentimen positif terendah (80,5%) dan negatif tertinggi (19,5%). Setelah itu, sentimen positif kembali stabil di kisaran 84,9%–90,1% hingga November, sebelum sedikit menurun pada Desember (85,0%). Data ini menunjukkan bahwa meskipun respon pengaduan umumnya dinilai baik, terdapat periode tertentu di mana pengguna merasa kurang puas terhadap tindak lanjut yang diberikan.



GAMBAR 15
Grafik Analisis Tren Bulanan Topik Pengaduan

V. KESIMPULAN

Berdasarkan hasil pengujian model analisis sentimen menggunakan algoritma *Logistic Regression*, diperoleh akurasi sebesar 86%. Meskipun tingkat akurasi ini menunjukkan kinerja yang cukup baik dalam mengklasifikasikan sentimen positif dan negatif pada ulasan aplikasi PLN Mobile, performanya masih menunjukkan fluktuasi di beberapa periode pelatihan. Hal ini disebabkan oleh keterbatasan *Logistic Regression* dalam menangani pola data teks yang kompleks, khususnya dalam menangkap konteks dan hubungan antar kata pada ulasan pengguna.

Pada hasil pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) yang dipadukan dengan hasil klasifikasi *Logistic Regression*, diperoleh perbandingan sentimen positif dan negatif pada setiap topik sebagai berikut:

- Transaksi: 85,9% positif dan 14,1% negatif.
- Layanan: 86,1% positif dan 13,9% negatif.
- Aplikasi: 86,2% positif dan 13,8% negatif.
- Pengaduan: 86,0% positif dan 14,0% negatif.

REFERENSI

- [1] Kumar, M., Khan, L., dan Chang, H.-T. 2025. Evolving Techniques in Sentiment Analysis: A Comprehensive Review. Department of Computer Science, Mir Chakar Khan Rind University, Sibi, Balochistan, Pakistan; IBADAT International University Islamabad, Pakpattan Campus, Pakistan; Chang Gung University, Taoyuan, Taiwan.
- [2] Thaha, A. R., & Purnamasari, R. (2021). Trends and pattern on library marketing: A systematic review. Library Philosophy and Practice, (6211), 1–18. University of Nebraska-Lincoln.
- [3] Afasel, D., Purnamasari, R., & Edwar. (2022). Klasifikasi tutupan lahan menggunakan supervised machine learning pada citra satelit menggunakan Google Earth Engine. e-Proceeding of Engineering, 8(6), 3278–3286.
- [4] Mahira, S. A., Sukoco, I., Barkah, C. S., Jamil, N., Novel, A., dan Bisnis, J. A. 2023. Teknologi Artificial Intelligence dalam Analisis Sentimen: Studi Literatur pada Perusahaan Kata.Ai. Bulan Agustus Tahun, 6(2), 139–148.
- [5] Wahyudin. 2020. Aplikasi Topic Modeling pada Pemberitaan. Seminar Nasional Official Statistics: Pengembangan Official Statistics dalam Mendukung Implementasi SDG's, 309–318.
- [6] Novirianto, I. F. 2023. Analisis Sentimen Berbasis Aspek dalam Ulasan Aplikasi MyBluebird dengan Implementasi N-Gram dan Algoritma Logistic Regression. Skripsi Sarjana, Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Jakarta.
- [7] PLN (Perusahaan Listrik Negara). 2021. Company Profile. PT PLN (Persero). [Online]. Tersedia: <https://web.pln.co.id/statics/uploads/2021/08/Company-Profile-PLN-082021.pdf>
- [8] Yulistiar, F. W., dan Kriswibowo, A. 2024. Evaluation of Customer Service through PLN Mobile Application in the Framework of Electronic Government Maturity Model in Sidoarjo Customer Service Implementation Unit. East Java "Veteran" National Development University. <https://web.pln.co.id/statics/uploads/2021/08/Company-Profile-PLN-082021.pdf>
- [9] Shad, R., Potter, K., dan Gracias, A. 2024. Natural Language Processing (NLP) for Sentiment Analysis: A Comparative Study of Machine Learning Algorithms.
- [10] Pradana, R. O. 2023. Analisis Tren Topik Sistem Informasi di Indonesia dari Perspektif Topic Modeling Menggunakan LDA (Latent Dirichlet Allocation). Skripsi.