Analisis Sentimen Layanan Rumah Sakit di Purwokerto Berdasarkan Ulasan Google Maps Menggunakan Metode LSTM

1st Fauzi Irfan Syaputra
Sistem Informasi, Fakultas Rekayasa Industri
Universitas Telkom Purwokerto
Purwokerto, Indonesia
fauipang@student.telkomuniversity.ac.id

2nd Sena Wijayanto
Sistem Informasi, Fakultas Rekayasa Industri
Universitas Telkom Purwokerto
Purwokerto, Indonesia
senawijayanto@telkomuniversity.ac.id

***Abstrak* — Persepsi Masyarakat terhadap layanan rumah sakit terlihat dari ulasan yang ditulis secara daring, salah satunya melalui platform Google Maps. Ulasan tidak hanya memberikan informasi bagi calon pasien, tetapi juga mencerminkan pengalaman subjektif yang dapat dimanfaatkan untuk evaluasi layanan secara menyeluruh. Dalam penelitian ini, dilakukan analisis sentimen terhadap ulasan rumah sakit yang ada di wilayah Purwokerto, dengan menggunakan metode Long Short-Term Memory (LSTM). Data yang digunakan berasal dari 17.261 ulasan berbahasa Indonesia yang dikumpulkan melalui proses web scrapping. Tahapan analisis mencakup pre-processing teks, seperti konversi karakter, pembersihan data, normalisasi, tokenisasi, penghapusan stopword, dan stemming, serta pelabelan sentimen menggunakan leksikon SenticNet dan penyeimbangan data dengan metode Synthetic Minority Over-Sampling Technique (SMOTE). Model LSTM yang dibangun dan dievaluasi menggunakan metrik akurasi, precision, recall, dan f1-score. Hasil model LSTM mampu mengklasifikasikan sentimen dengan akurasi sebesar 86%, yang menunjukan model efektif dalam memahami isi ulasan. Penelitian ini menunjukan bahwa deep learning dapat dimanfaatkan untuk menganalisis opini publik, khususnya di sektor kesehatan, dan bisa menjadi bahan pertimbangan dalam upaya peningkatan kualitas layanan rumah sakit.*

***Kata kunci*— analisis sentimen, layanan rumah sakit, LSTM, SMOTE, confusion matrix**

I. PENDAHULUAN

Layanan rumah sakit yang berkualitas sangat penting untuk menjaga kesejahteraan masyarakat. Masyarakat kini banyak membagikan pengalaman layanan melalui ulasan di platform digital seperti Google Maps [1]. Ulasan dapat memuat persepsi, emosi, dan opini pasien yang penting untuk dianalisis [2]. Tingginya volume dan kompleksitas bahasa dalam ulasan membuat analisis manual menjadi tidak efisien. Oleh karena itu, diperlukan pendekatan otomatis menggunakan algoritma analisis sentimen [3].

Penelitian ini mengadopsi metode Long Short-Term Memory (LSTM) untuk melakukan klasifikasi sentimen terhadap ulasan layanan rumah sakit di Purwokerto. LSTM dipilih karena kemampuannya dalam memproses data sekuensial dan memahami konteks jangka panjang pada teks [4].

Penelitian sebelumnya menunjukan efektivitas LSTM dalam menganalisis sentimen ulasan aplikasi dan layanan digital [5][6]. Dalam penelitian ini, digunakan data dari Google Maps, yang dikumpulkan melalui teknik web scraping. Data dianalisis dengan preprocessing, pelabelan menggunakan SenticNet, dan balancing menggunakan SMOTE.

II. KAJIAN TEORI

A. Analisis Sentimen

Analisis sentimen adalah bagian dari NLP untuk mengklasifikasikan opini dalam teks menjadi kategori positif atau negatif [7]. Teknik ini banyak digunakan untuk memahami persepsi publik terhadap suatu layanan atau produk [8].

B. Long Short-Term Memory (LSTM)

LSTM adalah arsitektur jaringan saraf tiruan yang efektif dalam memproses data sekuensial. LSTM memiliki struktur gate, seperti forget gate, input gate, dan output gate, yang memungkinkan jaringan mengingat informasi penting dalam waktu yang lama [9].

1. Forget Gate

$$f_t = \sigma\left(W_f \cdot [a_{t-1}, X_t] + b_f\right) \quad (1)$$

Keterangan
$\sigma = fungsi\ sigmoid$
$W_f = bobot\ yang\ terkait\ dengan\ forget\ gate$
$a_{t-1} = hidden\ state\ sebelumnya$
$X_t = input\ ke\ jaringan$
$b_f = bias\ yang\ terkait\ dengan\ forget\ gate$

2. Input Gate

$$\widetilde{C_t} = tanh(W_C \cdot [a_{t-1}, X_t] + b_C) \quad (2)$$

Keterangan
$\widetilde{C_t} = fungsi\ kandidat$
$W_c = bobot\ yang\ terkait\ dengan\ fungsi\ kandidat$

$a_{t-1} = hidden\ state\ sebelumnya$
$X_t = input\ ke\ jaringan$
$b_c = bias\ yang\ terkait\ dengan\ fungsi\ kandidat$

3. *Output Gate*

$$\mathcal{Y}_t = \sigma(W_y \cdot [a_{t-1}, X_t] + b_y) \quad (3)$$

Keterangan
$\sigma = fungsi\ sigmoid$
$W_y = bobot\ yang\ terkait\ dengan\ output\ gate$
$a_{t-1} = hidden\ state\ sebelumnya$
$X_t = input\ ke\ jaringan$
$b_c = bias\ yang\ terkait\ dengan\ output\ gate$

C. *SenticNet*

SenticNet adalah leksikon semantic yang memberikan nilai polaritas untuk kata-kata, yang digunakan dalam pendekatan leksikal untuk pelabelan sentimen [10].

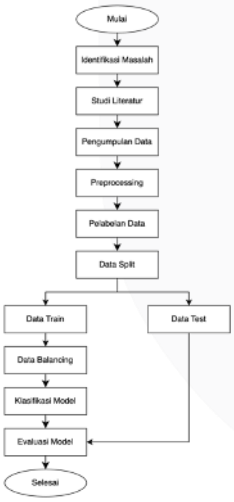
D. *Synthetic Minority Over-Sampling Technique (SMOTE)*

SMOTE adalah metode untuk mengatasi ketidak seimbangan kelas dalam dataset dengan menciptakan sampel sintetis pada kelas minoritas [11].

E. *Google Maps*

Google Maps adalah layanan yang menyediakan fitur ulasan public yang dapat dimanfaatkan sebagai data opini Masyarakat terhadap layanan rumah sakit [12].

III. METODE



GAMBAR 1
(ALUR PENELITIAN)

Penelitian dimulai dengan identifikasi masalah terkait persepsi masyarakat terhadap layanan rumah sakit di Purwokerto berdasarkan ulasan pada Google Maps. Data ulasan dikumpulkan melalui teknik web scraping menggunakan platform Apify, yang menghasilkan 17.261 ulasan dari 16 rumah sakit. Setelah pengumpulan data, dilakukan tahapan teks preprocessing yang mencakup case folding, cleaning, normalization, tokenizing, stemming, dan penghapusan stopword untuk menyiapkan data teks yang siap dianalisis. Data kemudian diberi label sentimen secara otomatis menggunakan pendekatan leksikon dengan kamus SenticNet, berdasarkan skor polaritas kata. Selanjutnya, data dibagi menjadi data latih dan data uji dengan rasio 80:20, dan dilakukan penyeimbangan kelas menggunakan metode Synthetic Minority Over-Sampling Technique (SMOTE). Untuk ekstraksi fitur, digunakan word embedding agar setiap kata direpresentasikan sebagai vektor berdimensi tetap yang dapat digunakan sebagai input ke dalam model. Model klasifikasi dibangun menggunakan arsitektur Long Short-Term Memory (LSTM) dengan 5 skenario eksperimen, yaitu baseline, regularizer, batch normalization, dropout, dan kombinasi dengan ketiga optimasi. Setiap skenario dievaluasi menggunakan metrik akurasi, presisi, recall, dan f1-score, dan skenario terbaik dipilih untuk dianalisislebih lanjut menggunakan confusion matrix.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang terkumpul terdiri dari 17.261 ulasan rumah sakit yang ditulis dalam Bahasa Indonesia. Setelah melalui tahapan preprocessing dan pelabelan, jumlah data yang tersedia menjadi 9.383 data. Selanjutnya, data dibagi menjadi 80% data latih dan 20% data uji. Teknik SMOTE digunakan untuk menyeimbangkan distribusi label pada data latih.

TABEL 1
(HASIL SCRAPING)

PublishedAtDate	Text
2024-11-17T10:01:58.326Z	"Tiga bulan lalu (Agustus) saya lahiran di sini.
2024-11-14T23:55:21.715Z	Gaya excellent ,,ct scan masih numpang ,
2024-10-29T11:35:21.221Z	Awalnya wajah saya mengalami reaksi ketidak cocokan akibat skincare

B. Preprocessing

Tahap preprocessing dilakukan untuk mempersiapkan data sebelum masuk ke tahap analisis. Tahap preprocessing mencakup serangkaian prosedur seperti case folding, cleaning, normalization, stopword, tokenizing, dan stemming. Prosedur preprocessing dilakukan dengan tujuan untuk memaksimalkan kualitas data pada tahap analisis.

a) Case Folding dan Cleaning

Tahap case folding dilakukan kepada seluruh teks ulasan untuk dikonversi menjadi huruh kecil [13]. Proses ini bertujuan untuk menyeragamkan format teks agar tidak terjadi perbedaan interpretasi antar kata. Implementasi dilakukan dengan menggunakan fungsi text.lower() dari library Python.

Tahap cleaning dilakukan untuk membersihkan karakter yang tidak relevan, seperti tagar, URL, tanda baca, angka, ruang kosong, emoji, dan simbol [14]. Dengan melakukan cleaning, data teks ulasan dapat lebih mudah untuk dilakukan analisis lebih lanjut.

TABEL 2
(CASE FOLDING DAN CLEANING)

Sebelum	Sesudah
"Tiga bulan lalu (Agustus) saya lahiran di sini.	tiga bulan lalu agustus saya lahiran di sini

Sebelum	Sesudah
Gaya excellent „ct scan masih numpang ,	gaya excellent ct scan masih numpang
Awalnya wajah saya mengalami reaksi ketidak cocokan akibat skincare	awalnya wajah saya mengalami reaksi ketidak cocokan akibat skincare

b) Normalization

Tahap normalization atau normalisasi dilakukan untuk mengubah kata yang tidak baku menjadi kata yang baku, seperti terdapat kata singkatan pada sebuah kalimat dan juga kata yang tidak lengkap [15]. Proses normalization menggunakan kamus yang dibuat secara manual dengan mencakup semua kata yang tidak lengkap dan singkatan pada dataset.

TABEL 3
(NORMALIZATION)

Sebelum	Sesudah
tiga bulan lalu agustus saya lahiran di sini	tiga bulan lalu agustus saya lahiran di sini
gaya excellent ct scan masih numpang	gaya luar biasa ct scan scan masih menumpang
awalnya wajah saya mengalami reaksi ketidak cocokan akibat skincare	awalnya wajah saya mengalami reaksi ketidak cocokan akibat perawatan kulit

c) Tokenization

Tahap tokenizing dilakukan untuk menguraikan kalimat dalam teks menjadi satuan kata [13]. Proses tokenizing menggunakan library *split()* yang ada pada python untuk memecah teks menjadi kata berdasarkan spasi secara otomatis.

TABEL 4
(TOKENIZATION)

Sebelum	Sesudah
tiga bulan lalu agustus saya lahiran di sini	['tiga', 'bulan', 'lalu', 'agustus', 'saya', 'lahiran', 'di', 'sini']
gaya luar biasa ct scan scan masih menumpang	['gaya', 'luar', 'biasa', 'ct', 'scan', 'scan', 'masih', 'menumpang']
awalnya wajah saya mengalami reaksi ketidak cocokan akibat perawatan kulit	['awalnya', 'wajah', 'saya', 'mengalami', 'reaksi', 'ketidak', 'cocokan', 'akibat', 'perawatan', 'kulit']

d) Stemming

Tahap stemming dilakukan untuk memperbaiki sebuah kata yang memiliki imbuhan diubah menjadi bentuk kata dasar dengan menghilangkan afiksnya [16]. Proses stemming menggunakan fungsi *StemmerFactory()* dari pustaka Sastrawi yang merupakan library python untuk stemming bahasa Indonesia.

TABEL 5
(STEMMING)

Sebelum	Sesudah
['tiga', 'bulan', 'lalu', 'agustus', 'saya', 'lahiran', 'di', 'sini']	['tiga', 'bulan', 'lalu', 'agustus', 'saya', 'lahir', 'di', 'sini
['gaya', 'luar', 'biasa', 'ct', 'scan', 'scan', 'masih', 'menumpang']	['gaya', 'luar', 'biasa', 'ct', 'scan', 'scan', 'masih', 'tumpang']
['awalnya', 'wajah', 'saya', 'mengalami', 'reaksi', 'ketidak', 'cocokan', 'akibat', 'perawatan', 'kulit']	['awal', 'wajah', 'saya', 'alami', 'reaksi', 'tidak', 'cocok', 'akibat', 'menjaga', 'kulit']

e) Stopword Removal

Tahap stopword removal dilakukan untuk mengurangi jumlah kata yang terdiri dari kata hubung tetapi tidak memiliki makna yang signifikan atau memberikan kontribusi yang kurang penting [17]. Proses menghapus kata yang tidak memiliki makna menggunakan daftar kata yang dibuat secara manual dengan mencakup semua kata yang dimemiliki makna.

TABEL 6
(STOPWORD REMOVAL)

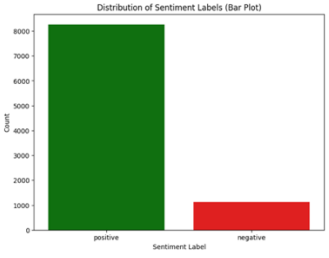
Sebelum	Sesudah
['tiga', 'bulan', 'lalu', 'agustus', 'saya', 'lahir', 'di', 'sini	['tiga', 'bulan', 'lalu', 'lahir', 'sini']
['gaya', 'luar', 'biasa', 'ct', 'scan', 'scan', 'masih', 'tumpang']	['gaya', 'luar', 'biasa', 'ct', 'tumpang']
['awal', 'wajah', 'saya', 'alami', 'reaksi', 'tidak', 'cocok', 'akibat', 'menjaga', 'kulit']	['awal', 'wajah', 'alami', 'reaksi', 'cocok', 'akibat', 'menjaga', 'kulit']

C. Labeling

Tahap selanjutnya setelah dilakukan pelabelan data, dengan melakukan proses labeling sentimen [18]. Proses labeling sentimen menggunakan pendekatan berbasis leksikon *SenticNet.*

TABEL 7
(LABELING)

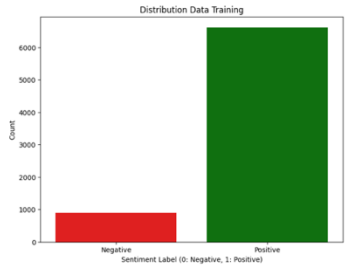
Polarity Skor	Text	Label Sentimen
{'tiga': None, 'bulan': 0.891, 'lalu': None, 'lahir': 0.955, 'sini': None}	2,383	positive
{'gaya': 0.815, 'luar_biasa': 0.574, 'ct_scan': None, 'tumpang': None}	1,389	positive
{'awal': None, 'wajah': None, 'alami': None, 'reaksi': 0.329, 'cocok': 0.881, 'akibat': -0.229, 'menjaga': 0.591, 'kulit': 0.569}	5,515	positive



GAMBAR 2
(DISTRIBUSI DATA SENTIMEN)

D. Data Split

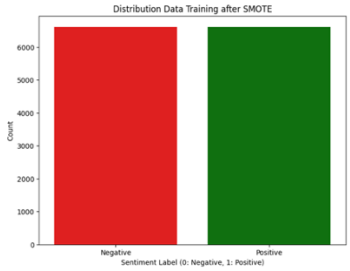
Proporsi pembagian data adalah 80% untuk data latih dan 20% untuk data uji, dengan *random_state=42* untuk menjaga konsistensi hasil pembagian [19]. Paramter *stratify=y* digunakan untuk memastikan distribusi kelas pada data latih dan uji tetap seimbang sesuai proporsi awal.



GAMBAR 3
(DISTRIBUSI DATA LATIH)

E. Balancing Data

Pada tahap selanjutnya, dilakukan proses balancing data menggunakan metode *SMOTE*. Proses balancing penting dilakukan karena pada distribusi pada label sentimen ditemukan ketidakseimbangan pada label positif dan negatif, yang dapat menyebabkan model bias terhadap kelas mayoritas [20].



GAMBAR 3
(DISTRIBUSI DATA SETELAH SMOTE)

F. Klasifikasi Model

Dalam tahap ini dilakukan percobaan terhadap 5 skenario arsitektur model LSTM dengan tujuan mengevaluasi performa pada perbedaan arsitektur model dalam mengklasifikasikan sentimen ulasan rumah sakit.

TABEL 8
(SKENARIO MODEL)

Skenario	Layer
1	Baseline
2	LSTM + Regularizer
3	LSTM + Dropout
4	LSTM + BatchNormalization
5	LSTM + Optimization

Pada skenario 5 dengan layer model LSTM dengan optimasi layer berupa penambahan regularizer, dropout, dan batch normalization menunjukan performa yang optimal, model mampu menjaga keseimbangan antara dua kelas serta memiliki loss yang cukup rendah. Oleh karena itu, skenario 5 digunakan dalam penelitian untuk dilanjutkan pada tahap evaluasi model. Langkah selanjutnya adalah menjelaskan secara rinci arsitektur model LSTM yang digunakan.

TABEL 9
(ARSITEKTUR MODEL)

Layer
LSTM
Regularizer
Batch Normalization
Dropout
Dense

Model dikompilasi menggunakan optimizer Adam dengan learning rate kecil sebesar 0.0005 dengan tujuan proses pelatihan lebih stabil. Fungsi loss yang digunakan adalah categorical_crossentropy, karena untuk memodelkan klasifikasi multi-kelas, dan metrik evaluasi yang digunakan adalah accuracy.

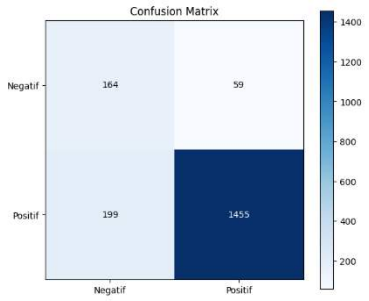
G. Evaluasi Model

Secara keseluruhan, model memperoleh akurasi sebesar 86% dari total 1.877 data uji. Nilai rata-rata makro untuk presisi, recall, dan f1-score masing-masing adalah 0.71, 0.81, dan 0.74, yang mencerminkan performa rata-rata antar kelas tanpa memperhitungkan jumlah data di setiap kelas (sehingga lebih adil terhadap kelas minoritas). Sementara itu, rata-rata tertimbang yang mempertimbangkan proporsi data memberikan nilai precision sebesar 0.90, recall sebesar 0.86, dan f1-score sebesar 0.88. Ini mengindikasikan bahwa model memiliki kinerja keseluruhan yang sangat baik, dengan kecenderungan mendominasi pada kelas mayoritas.

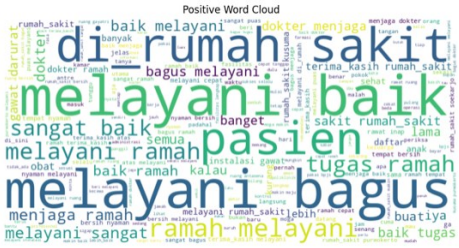
	precision	recall	f1-score	support
0	0.45	0.74	0.56	223
1	0.96	0.88	0.92	1654
accuracy			0.86	1877
macro avg	0.71	0.81	0.74	1877
weighted avg	0.90	0.86	0.88	1877

GAMBAR 4
(METRIK AKURASI)

Evaluasi kinerja model LSTM juga ditunjukan melalui *confusion matrix*, yang menggambarkan hasil prediksi terhadap 2 kelas sentimen, yaitu kelas negatif dan positif. Hasil confusion matrix, model berhasil mengklasifikasikan 164 data kelas negatif dengan tepat, sementara 59 data kelas negatif diprediksi sebagai kelas positif. Untuk data yang termasuk dalam kelas positif, model berhasil mengklasifikasikan 1.455 data dengan benar, dan terdapat 199 data yang seharusnya termasuk kelas positif, namun terklasifikasi sebagai kelas negatif.



GAMBAR 5
(CONFUSION MATRIX)



GAMBAR 6
(WORDCLOUD POSITIF)

GAMBAR 7 :
(WORDCLOUD NEGATIF)

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat disimpulkan, sebagai berikut:

1. Penelitian berhasil melakukan analisis sentimen terhadap ulasan layanan rumah sakit di Purwokerto yang diambil dari platform Google Maps. Hasil klasifikasi sentimen dari data ulasan sebanyak 8.268 ulasan terklasifikasi sebagai sentimen positif, sedangkan 1.115 ulasan terklasifikasi sebagai sentimen negatif. Analisis sentimen menunjukan bahwa sebagian besar masyarakat memberikan tanggapan positif terhadap pelayanan rumah sakit, dengan kata yang sering muncul dalam ulasan positif mencakup ramah, baik, dan bagus. Sebaliknya, meskipun ulasan negatif lebih sedikit, kata yang sering muncul menunjukan adanya perhatian terhadap aspek pelayanan dan kenyamanan pasien.
2. Model klasifikasi Long Short-Term Memory (LSTM) yang diterapkan pada data pelatihan yang telah melalui proses balancing, dengan data positif dan negatif sebanyak 6.614 data menunjukan performa yang baik dalam mengklasifikasikan sentimen ulasan layanan rumah sakit. Hasil evaluasi pada data uji, model mencapai akurasi sebesar 86% dengan nilai precision dan recall yang seimbang pada kedua kelas. Hasil klasifikasi sentimen menunjukan bahwa model mampu mengenali sentimen ulasan secara cukup akurat dan dapat digunakan sebagai alat bantu dalam memahami persepsi masyarakat terhadap layanan rumah sakit secara otomatis.

REFERENSI

[1] D. Maharani, F. Helmiah, and N. Rahmadani, “Penyuluhan Manfaat Menggunakan Internet dan Website Pada Masa Pandemi Covid-19,” *Abdiformatika: Jurnal Pengabdian Masyarakat Informatika*, vol. 1, no. 1, pp. 1–7, 2021, doi: 10.25008/abdiformatika.v1i1.130.

[2] A. Sri Widagdo *et al.*, “Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT,” 2023.

[3] A. C. T Angel and V. H. Pranatawijaya, “Analisis Sentimen dan Emosi dari Ulasan Google Maps untuk Layanan Rumah Sakit di Palangka Raya Menggunakan Machine Learning,” 2024.

[4] H. Bourezk, A. Raji, N. Acha, and H. Barka, “Analyzing Moroccan Stock Market using Machine Learning and Sentiment Analysis,” in *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, IEEE, Apr. 2020, pp. 1–5. doi: 10.1109/IRASET48871.2020.9092304.

[5] J. Serrano-Guerrero, M. Bani-Doumi, F. P. Romero, and J. A. Olivas, “Understanding what patients think about hospitals: A deep learning approach for detecting emotions in patient opinions,” *Artif Intell Med*, vol. 128, p. 102298, 2022, doi: https://doi.org/10.1016/j.artmed.2022.102298.

[6] A. I. A. Rahim, M. I. Ibrahim, K. I. Musa, S. L. Chua, and N. M. Yaacob, “Assessing patient-perceived hospital service quality and sentiment in malaysian public hospitals using machine learning and facebook reviews,” *Int J Environ Res Public Health*, vol. 18, no. 18, Sep. 2021, doi: 10.3390/ijerph18189912.

[7] Yeni Safitri, Rakhmat Kurniawan, and Suhardi, “Analisis Sentimen Mengenai Childfree Menggunakan Metode Naïve Bayes,” *The Indonesian Journal of Computer Science*, vol. 13, no. 4, Jul. 2024, doi: 10.33022/ijcs.v13i4.4136.

[8] E. Salim and M. Syafrullah, “Jakarta Barat Menggunakan Algoritme K-Nearest Neighbor,” *20*, pp. 58–65, 2023.

[9] U. B. Mahadevaswamy and P. Swathi, “Sentiment Analysis using Bidirectional LSTM Network,” in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 45–56. doi: 10.1016/j.procs.2022.12.400.

[10] Ernianti Hasibuan and Elmo Allistair Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *Jurnal Teknik dan Science*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.

[11] S. A. Alex, N. Z. Jhanjhi, M. Humayun, A. O. Ibrahim, and A. W. Abulfaraj, “Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE,” *Electronics (Switzerland)*, vol. 11, no. 17, Sep. 2022, doi: 10.3390/electronics11172737.

[12] “Google Maps.” [Online]. Available: https://support.google.com/maps/answer/144349?hl=en&topic=1687350&visit_id=638843887594750500-2402772628&rd=2

[13] T. T. Widowati and M. Sadikin, “Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine,” *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.

[14] S. Jasmine Putri, Q. Attaqwa, and A. Pratama, “SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Classification Determining Employee Data Work Schedules Using C4.5 and K-Nearest Neighbor Algorithm Klasifikasi Menentukan Jadwal Kerja Data Karyawan Menggunakan Algoritma C4.5 dan K-nearest

Neighbor." [Online]. Available: https://journal.irpi.or.id/index.php/sentimas
[15] E. P. A. Akhmad, "Analisis Sentimen Ulasan Aplikasi DLU Ferry Pada Google Play Store Menggunakan Bidirectional Encoder Representations from Transformers," *Jurnal Aplikasi Pelayaran Dan Kepelabuhanan*, vol. 13, no. 2, pp. 104–112, 2023, doi: 10.30649/japk.v13i2.94.
[16] A. N. Ulfah and M. K. Anam, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 1, pp. 1–10, 2020, doi: 10.35957/jatisi.v7i1.196.
[17] Fauzan Baehaqi and N. Cahyono, "Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan Algoritma Naïve Bayes," *Indonesian Journal of Computer Science*, vol. 13, no. 1, pp. 1051–1063, 2024, doi: 10.33022/ijcs.v13i1.3301.
[18] A. A. Pratiwi and M. Kamayani, "Perbandingan Pelabelan Data dalam Analisis Sentimen Kurikulum Proyek di platform TikTok: Pendekatan Naïve Bayes," *Jurnal Eksplora Informatika*, vol. 14, no. 1, pp. 96–107, Sep. 2024, doi: 10.30864/eksplora.v14i1.1093.
[19] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 273–281, Jan. 2024, doi: 10.57152/malcom.v4i1.1085.
[20] K. Akbar and M. Hayaty, "Data Balancing untuk Mengatasi Imbalance Dataset pada Prediksi Produksi Padi Balancing Data to Overcome Imbalance Dataset on Rice Production Prediction," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 2, no. 02, pp. 1–14, 2020.