JAISEN: **Journal of Advanced Information Systems and Engineering**
Vol. 1, 1 (August, 2025), pp 9-17
ISSN: - EISSN: - , DOI: -

UTILIZATION OF MACHINE LEARNING ALGORITHMS FOR CUSTOMER COMPLAINT CLASSIFICATION AT PERUMDAM

Poly Reksi[1*], Susandri Susandri[2], M. Syawalludin[3], Feldiansyah[4]

[1] Magister Ilmu Komputer, Universitas Lancang Kuning; polyreksi@unilak.ac.id
[2] Magister Ilmu Komputer, Universitas Lancang Kuning; susandri@unilak.ac.id
[3] Magister Ilmu Komputer, Universitas Lancang Kuning; m.syawalluddin@unilak.ac.id
[4] Magister Ilmu Komputer, Universitas Lancang Kuning ; feldiansyah@unilak.ac.id

ARTICLE INFO

Keywords:

Machine Learning;
Customer Complaint;
Classification;
TF-IDF;
Logistic Regression.

Article history:

Received 2025-08-06
Revised 2025-08-09
Accepted 2025-08-13

ABSTRACT

In today's digital era, managing customer complaints poses a significant challenge for public service providers, such as the Regional Drinking Water Company (PERUMDAM). With the increasing number of customers and the complexity of complaints, manual methods are no longer adequate to efficiently handle large volumes of data. This study focuses on the application of machine learning algorithms, namely Logistic Regression, LightGBM, and Random Forest, to classify customer complaints based on patterns in the data. The research stages include dataset collection, data preprocessing such as cleaning, casefolding, tokenization, filtering, normalization, and stemming, as well as feature extraction using TF-IDF, SMOTE features, and hyperparameter tuning. Evaluation is conducted based on metrics such as accuracy, precision, recall, and F1-score. The results indicate that the LightGBM algorithm provides the best performance in terms of precision and recall, while Random Forest achieves the highest accuracy, and Logistic Regression serves as an efficient model for data with linear relationships. Therefore, LightGBM is recommended for further implementation in managing customer complaints data quickly and accurately. This study contributes to the development of technology-based solutions to enhance the efficiency of customer complaint management in the public service sector.

Corresponding Author:
Poly Reksi
Magister Ilmu Komputer, Universitas Lancang Kuning; polyreksi@unilak.ac.id

1. INTRODUCTION

Customer complaint management is a crucial element in improving service quality and guiding strategic decision-making, particularly for public service providers such as the Regional Public Water Utility Company (Perusahaan Umum Daerah Air Minum, PERUMDAM) (Hutauruk, Nabila, & Furqan, 2023). As the primary provider of clean water, PERUMDAM receives a wide range of customer complaints, covering technical issues, administrative matters, and service quality concerns. The growing number of customers, coupled with increasing complexity in complaints, has rendered manual methods less effective in managing large volumes of data efficiently. Complaints collected through customer service centers, digital applications, and social media further increase the challenge of providing timely and accurate responses.

Delays or inaccuracies in handling complaints not only lead to customer dissatisfaction but may also harm the company's reputation. This situation calls for a technology-driven solution capable of automatically categorizing complaints and determining follow-up priorities more effectively. Machine learning offers a modern approach to addressing these challenges by enabling automatic processing and classification of complaint data based on identifiable patterns. Such an approach can improve operational efficiency, accelerate decision-making, and allow the company to respond to customers more effectively.

This study focuses on three widely adopted machine learning algorithms—Logistic Regression, LightGBM, and Random Forest. Logistic Regression is a simple yet effective model for linear relationships; LightGBM excels in processing large-scale datasets at high speed; and Random Forest is highly reliable for identifying complex data patterns (Alsubayhin, Ramzan, & Alzahrani, 2024). Previous research has demonstrated that these algorithms deliver competitive results in text classification, sentiment analysis, and data grouping. However, their application in classifying customer complaints for PERUMDAM remains limited.

The research uses a dataset comprising 4,466 customer complaint records collected between 2019 and 2024. Data preprocessing stages include cleaning, case folding, tokenization, filtering, normalization, and stemming to ensure optimal data quality. Feature extraction is conducted using the TF-IDF method, while data imbalance is addressed through the Synthetic Minority Oversampling Technique (SMOTE). The Logistic Regression, LightGBM, and Random Forest models are trained with hyperparameter optimization to enhance performance. Model evaluation is carried out using accuracy, precision, recall, and F1-score metrics.

The aim of this study is to identify the most effective algorithm for customer complaint management at PERUMDAM. By leveraging machine learning technologies, the study is expected to make a significant contribution to developing technology-based solutions that improve operational efficiency and service quality in the public sector. The findings also provide practical recommendations for more effective implementation, supporting faster and more accurate customer responses.

2. METHODS

The research workflow and analytical stages undertaken in this study are outlined below. This investigation employs three machine learning algorithms—Logistic Regression, LightGBM, and Random Forest—to classify customer complaint data from PERUMDAM. The figure below presents a systematic overview of the approach adopted in this research, providing a clear understanding of the workflow and methodology implemented.

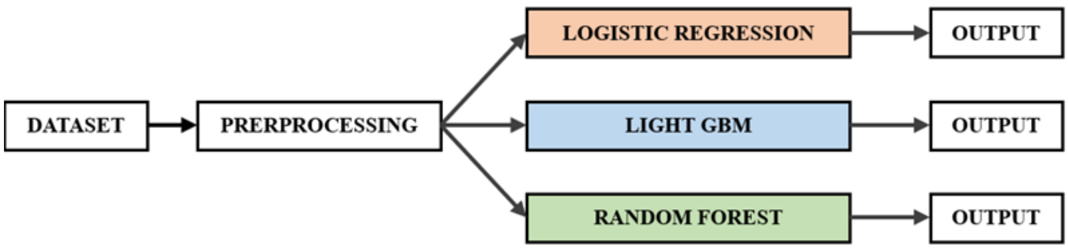


Figure 1. Research Workflow

2.1 Dataset

The dataset utilized in this study consists of 4,466 customer complaint records from PERUMDAM. Data collection was carried out by compiling complaint reports from internal PERUMDAM records, covering the period from 2019 to 2024. These reports were then entered into an Excel format to facilitate further processing. The selected time span was intended to provide a comprehensive representation of various types of customer complaints. The dataset exhibits an imbalanced class distribution, necessitating the application of additional techniques to address the disproportion among categories.

NORUT	PENGADUAN	LABEL
151	Tekanan Air Kecil	Tekanan Air Kecil
152	Ada Kebocoran Air Dirumah Dinas No 4	Instalasi Bocor
153	Air Tidak Jalan	Air Tidak Jalan
154	Sambung Kembali	Buka Kembali
155	Posisi Angka Meteran 558,28 minta tolong ditera angka meter	Kalibrasi Meter Air
156	Air Tidak Jalan	Air Tidak Jalan
157	Minta Ganti Meteran	Meteran Rusak
158	Stoc Kran Rusak Dekat amper	Kran Tidak Berfungsi
159	Pindah Line dari Pipa 3 ke Pipa 6 inc	Pindah Jalur Pipa
160	Air Tidak Jalan	Air Tidak Jalan
161	Minta Cek Meteran Di Segel	Cek Instalasi
162	Minta Tolong Meteran di Semen	Instalasi Tidak Sempurna
163	Tekanan Air Kecil	Tekanan Air Kecil
164	Tekanan Air Kecil	Tekanan Air Kecil
165	Air Keruh	Air Kotor

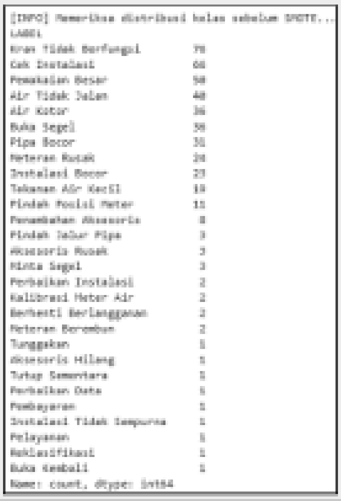
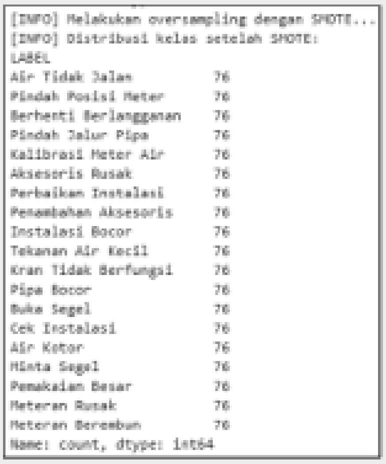
Figure 2. Sample from the Dataset

2.2 Data Preprocessing

The data preprocessing stage aims to ensure that the dataset meets the required quality standards for subsequent analysis.

Table 1. Data Preprocessing

NO	Preprocessing Data	Keterangan	Tampilan
1	Cleaning	menghapus data duplikat, data kosong, karakter tidak relevan seperti simbol atau tautan.	NORUT PENGADUAN LABEL 0 1 Tidak Bisa Terbaca Meteran Berembun 1 2 Tidak Bisa Terbaca Meteran Berembun 2 3 Tidak Bisa Terbaca Meteran Berembun 3 4 Tidak Bisa Terbaca Meteran Berembun 4 5 Tidak Bisa Terbaca Meteran Berembun ... ... 4462 4463 Bocor sebelum meteran Instalasi Bocor 4463 4464 Kran merah los, meteran mutar Kran Tidak Berfungsi 4464 4465 Kran merah los Kran Tidak Berfungsi 4465 4466 Buka segel Buka Segel 4466 4467 Cek meteran minta ganti Cek Instalasi
2	Casefolding	semua teks diubah menjadi huruf kecil untuk memastikan konsistensi.	PENGADUAN casefolding 0 Tidak Bisa Terbaca Tidak Bisa Terbaca 1 Tidak Bisa Terbaca Tidak Bisa Terbaca 2 Tidak Bisa Terbaca Tidak Bisa Terbaca 3 Tidak Bisa Terbaca Tidak Bisa Terbaca 4 Tidak Bisa Terbaca Tidak Bisa Terbaca ... ... 4462 Bocor sebelum meteran Bocor sebelum meteran 4463 Kran merah los, meteran mutar Kran merah los meteran mutar 4464 Kran merah los Kran merah los 4465 Buka segel Buka segel 4466 Cek meteran minta ganti Cek meteran minta ganti
3	Tokenisasi	mengurai teks pengaduan menjadi unit kata menggunakan pustaka NLP seperti NLTK.	casefolding tokenizing 0 Tidak Bisa Terbaca [Tidak, Bisa, Terbaca] 1 Tidak Bisa Terbaca [Tidak, Bisa, Terbaca] 2 Tidak Bisa Terbaca [Tidak, Bisa, Terbaca] 3 Tidak Bisa Terbaca [Tidak, Bisa, Terbaca] 4 Tidak Bisa Terbaca [Tidak, Bisa, Terbaca] ... ... 4462 Bocor sebelum meteran [Bocor, sebelum, meteran] 4463 Kran merah los meteran mutar [Kran, merah, los, meteran, mutar] 4464 Kran merah los [Kran, merah, los] 4465 Buka segel [Buka, segel] 4466 Cek meteran minta ganti [Cek, meteran, minta, ganti]

4	Filtering	menghapus kata-kata yang tidak memiliki makna signifikan (Stopword), seperti kata "dan", "atau", "yang", dll.	
5	Normalisasi	mengubah kata-kata tidak baku menjadi bentuk standar, seperti kata "tdk" menjadi "tidak"	
6	Stemming	mengubah kata ke bentuk dasarnya dengan menggunakan algoritma stemming bahasa indonesia seperti sastrawi.	
7	SMOTE	penanganan kelas data tidak seimbang, menggunakan metode SMOTE (*Synthetic Minority Oversampling Technique*) menangani ketidakseimbangan distribusi data antar kelas, sehingga model lebih mampu mengenali semua kategori pengaduan. Dengan penerapan SMOTE, kelas minoritas yang sebelumnya kurang terwakili meningkat secara proporsional. Grafik menunjukkan peningkatan signifikan dalam akurasi dan recall pada LightGBM setelah SMOTE dibandingkan dengan model yang dilatih tanpa SMOTE. Visualisasi distribusi dataset sebelum dan sesudah SMOTE menunjukkan bahwa ketidakseimbangan antar kelas dapat diminimalkan. Hal ini penting untuk memastikan semua kelas memiliki representasi yang memadai selama pelatihan model.  [INFO] Memeriksa distribusi kelas sebelum SMOTE... LABEL Kran Tidak Berfungsi 76 Cek Instalasi 66 Pemakaian Besar 58 Air Tidak Jalan 48 Air Kotor 36 Buka Segel [illegible] Pipa Bocor 31 Meteran Rusak 26 Instalasi Bocor 23 Tekanan Air Kecil 19 Pindah Posisi Meter 11 Penambahan Aksesoris 8 Pindah Jalur Pipa 3 Aksesoris Rusak 3 Minta Segel 3 Perbaikan Instalasi 2 Kalibrasi Meter Air 2 Berhenti Berlangganan 2 Meteran Berembun 2 Tunggakan 1 Aksesoris Hilang 1 Tutup Sementara 1 Perbaikan Data 1 Pembayaran 1 Instalasi Tidak Sempurna 1 Pelayanan 1 Reklasifikasi 1 Buka Kembali 1 Name: count, dtype: int64  [INFO] Melakukan oversampling dengan SMOTE... [INFO] Distribusi kelas setelah SMOTE: LABEL Air Tidak Jalan 76 Pindah Posisi Meter 76 Berhenti Berlangganan 76 Pindah Jalur Pipa 76 Kalibrasi Meter Air 76 Aksesoris Rusak 76 Perbaikan Instalasi 76 Penambahan Aksesoris 76 Instalasi Bocor 76 Tekanan Air Kecil 76 Kran Tidak Berfungsi 76 Pipa Bocor 76 Buka Segel 76 Cek Instalasi 76 Air Kotor 76 Minta Segel 76 Pemakaian Besar 76 Meteran Rusak 76 Meteran Berembun 76 Name: count, dtype: int64	

2.3 Feature Extraction and Tuning.

In this study, the feature extraction process was carried out to transform unstructured customer complaint text data into numerical representations that could be processed by machine learning algorithms. One of the methods employed was Term Frequency–Inverse Document Frequency (TF-IDF), a widely used technique for text representation that assigns weights to words based on their occurrence in a specific document relative to the entire document collection. TF-IDF consists of two primary components: Term Frequency (TF), which measures how often a word appears within a document, and Inverse Document Frequency (IDF), which determines how rare the word is across all documents. The combination of these components produces a weight that reflects the relative importance of a word within a given document.

```
# Step 10: TF-IDF Vectorization
print("[INFO] Mengonversi teks menjadi fitur TF-IDF...")
tfidf = TfidfVectorizer(ngram_range=(1, 3), max_features=2000)
X = tfidf.fit_transform(data['cleaned_text']).toarray()
y = data['LABEL']
```

Figure 3. TF-IDF Code

In this study, the implementation of TF-IDF was configured with specific parameters to enhance the quality of data representation. The parameter ngram_range=(1, 3) was applied to capture word relationships at the unigram, bigram, and trigram levels, enabling the algorithm to detect more complex inter-phrase patterns. Furthermore, the number of features was limited to the top 2,000 highest-weighted terms (max_features=2000) to maintain computational efficiency without compromising the relevance of the information. Prior to applying TF-IDF, stopwords such as “yang,” “dan,” and “atau” were removed to ensure that only meaningful terms were analyzed.

The TF-IDF transformation produced a sparse matrix in which each row represents a document (customer complaint) and each column corresponds to significant words identified within the dataset. This matrix served as the primary input for machine learning algorithms such as Logistic Regression, LightGBM, and Random Forest. TF-IDF enabled these algorithms to better recognize patterns in customer complaint text data, thereby improving predictive accuracy and overall model performance.

The primary strength of TF-IDF lies in its ability to assign higher weights to relevant terms, thus improving the quality of feature representation for subsequent analysis. This technique is also computationally efficient, as it can reduce data dimensionality without losing critical information. By integrating TF-IDF with text preprocessing techniques such as stemming and normalization, this study successfully generated high-quality data for machine learning models. These findings reaffirm the importance of TF-IDF as a foundational approach in classifying unstructured text data.

3. FINDINGS AND DISCUSSION

The implementation results of the three machine learning algorithms applied in this study—Logistic Regression, LightGBM, and Random Forest—are evaluated based on their performance in classifying PERUMDAM customer complaints. The evaluation employed standard metrics, including accuracy, precision, recall, and F1-score, to assess the effectiveness of each algorithm. The assessment followed a systematic approach, beginning with data preprocessing and continuing through to the final stage of analyzing prediction outcomes. All three algorithms were tested using the same dataset to ensure a fair and consistent comparison of their classification performance.

3.1 Logistic Regression Algorithm

Table 2. Logistic Regression Model Evaluation

MODEL	Data Set				Proses Data				Evaluasi Model					Hasil Klasifikasi			
	Dataset	data uji	data latih		Stemming (Menit)	CrossValid	Randomized Search CV (Menit)		Akurasi	Precision	Recall	F1-Score		Data	Benar	Salah	Akurasi
LOGISTIC REGRESSION	0.1	10%	90%		3.77	0.96	36.42		0.90	0.78	0.80	0.77		447	401	46	89.71%
	0.2	20%	80%		3.77	0.97	26.83		0.90	0.79	0.79	0.77		893	801	92	89.70%
	0.3	30%	70%		3.77	0.96	19.28		0.90	0.72	0.72	0.69		1340	1203	137	89.78%
	0.4	40%	60%		3.77	0.96	15.77		0.89	0.67	0.65	0.65		1786	1591	195	89.08%
	0.5	50%	50%		3.77	0.96	13.78		0.89	0.66	0.62	0.63		2233	1989	244	89.07%
	0.6	60%	40%		3.77	0.96	10.59		0.88	0.64	0.60	0.62		2679	2358	321	88.02%
	0.7	70%	30%		3.77	0.98	8.00		0.88	0.64	0.59	0.60		3126	2741	385	87.68%
	0.8	80%	20%		3.77	0.98	3.67		0.90	0.62	0.57	0.59		3572	3217	355	90.06%
	0.9	90%	10%		3.77	0.98	1.28		0.89	0.53	0.48	0.50		4019	3558	461	88.53%

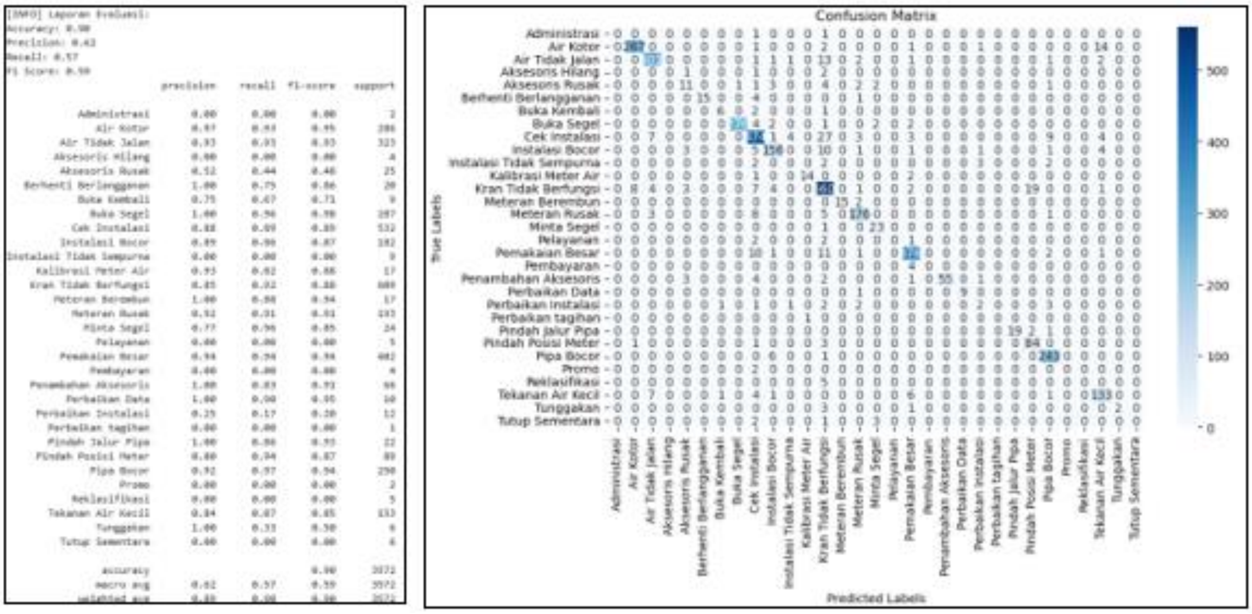


Figure 4. Accuracy and Confusion Matrix

3.2 LightGBM Algorithm

Table 3. LightGBM Model Evaluation

MODEL	DataSet			Proses Data			Evaluasi Model				Hasil Klasifikasi			
	Dataset	data uji	data latih	Stemming (Menit)	CrossValid	Randomized Search CV (Menit)	Akurasi	Precision	Recall	F1-Score	Data	Benar	Salah	Akurasi
LightGBM	0.1	10%	90%	3.57	0.96	3.78	0.89	0.72	0.77	0.72	447	396	51	88.59%
	0.2	20%	80%	3.57	0.96	3.67	0.88	0.75	0.75	0.71	893	786	107	88.02%
	0.3	30%	70%	3.57	0.96	2.48	0.88	0.68	0.69	0.66	1340	1181	159	88.13%
	0.4	40%	60%	3.57	0.95	2.39	0.87	0.66	0.64	0.64	1786	1560	226	87.35%
	0.5	50%	50%	3.57	0.95	2.08	0.88	0.64	0.63	0.62	2233	1955	278	87.55%
	0.6	60%	40%	3.57	0.95	1.65	0.86	0.65	0.57	0.59	2679	2304	375	86.00%
	0.7	70%	30%	3.57	0.97	1.32	0.87	0.66	0.58	0.60	3126	2710	416	86.69%
	0.8	80%	20%	3.57	0.97	0.84	0.88	0.61	0.54	0.56	3572	3148	424	88.13%
	0.9	90%	10%	3.57	0.96	0.48	0.87	0.51	0.48	0.49	4019	3510	509	87.34%

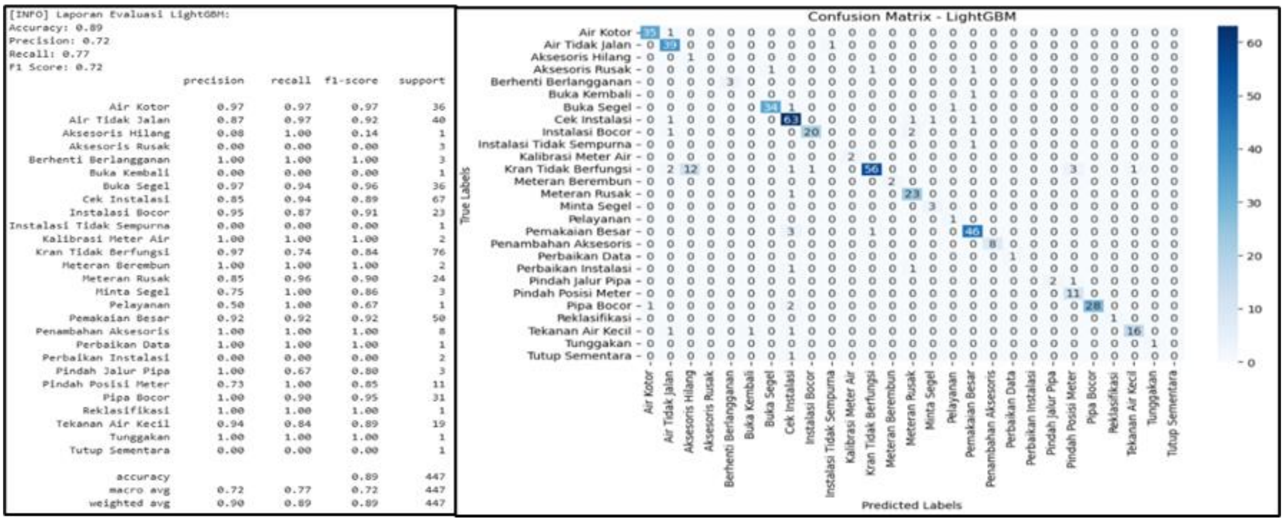


Figure 5. Accuracy and Confusion Matrix

3.3 Random Forest Algorithm

Table 4. Random Forest Model Evaluation

MODEL	Data Set				Proses Data				Evaluasi Model					Hasil Klasifikasi			
	Dataset	data uji	data latih		Stemming (Menit)	CrossValid	Randomized Search CV (Menit)		Akurasi	Precision	Recall	F1-Score		Data	Benar	Salah	Akurasi
RANDOM FOREST	0.1	10%	90%		3.52	0.97	6.01		0.89	0.74	0.75	0.74		447	400	47	89.49%
	0.2	20%	80%		3.52	0.97	5.20		0.89	0.78	0.73	0.74		893	798	95	89.36%
	0.3	30%	70%		3.52	0.97	3.99		0.90	0.72	0.66	0.68		1340	1203	137	89.78%
	0.4	40%	60%		3.52	0.96	3.31		0.90	0.69	0.64	0.66		1786	1603	183	89.75%
	0.5	50%	50%		3.52	0.96	2.48		0.89	0.67	0.61	0.62		2233	1992	241	89.21%
	0.6	60%	40%		3.52	0.96	1.92		0.87	0.66	0.58	0.60		2679	2341	338	87.38%
	0.7	70%	30%		3.52	0.98	1.35		0.88	0.62	0.57	0.58		3126	2747	379	87.88%
	0.8	80%	20%		3.52	0.98	0.62		0.90	0.62	0.56	0.58		3572	3218	354	90.09%
	0.9	90%	10%		3.52	0.98	0.16		0.87	0.49	0.47	0.48		4019	3514	505	87.43%

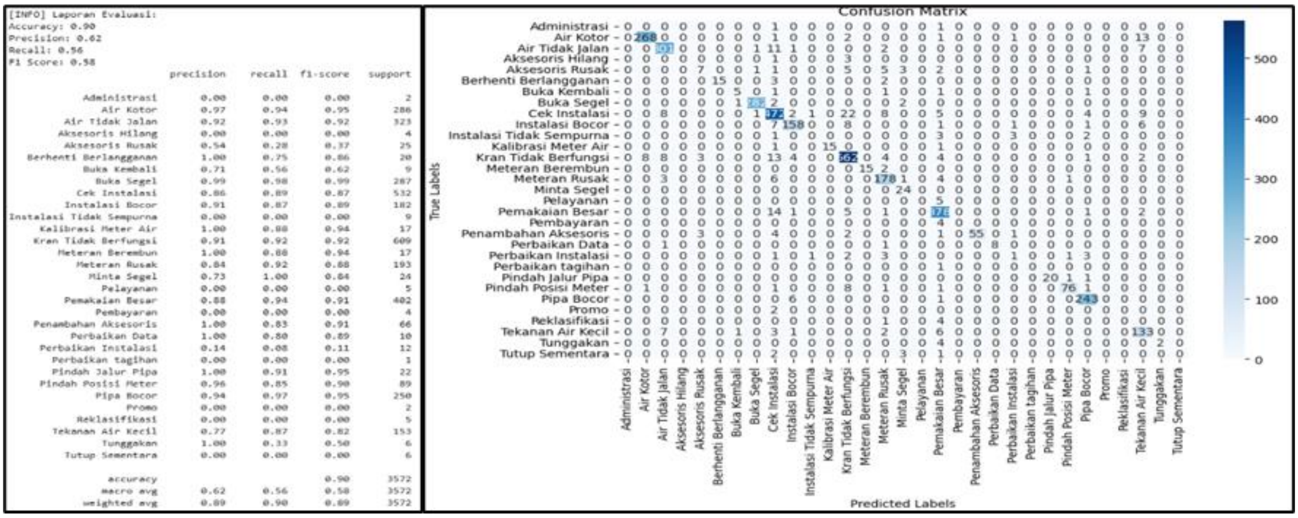


Figure 6. Accuracy and Confusion Matrix

3.4 *Discussion*

The three algorithms evaluated in this study—Logistic Regression, LightGBM, and Random Forest—demonstrated varying levels of performance, each with distinct strengths and limitations.

a) Logistic Regression

The Logistic Regression model achieved an average precision of 67.22% and recall of 61.89%. Despite its simplicity, the algorithm is effective for datasets with a linear relationship between features and the target variable. Its computational efficiency and ease of interpretation make it an ideal choice for applications requiring transparent analysis of the factors influencing classification results. However, its performance declines on datasets with nonlinear or complex patterns, limiting its effectiveness when handling highly variable data.

b) LightGBM

LightGBM excelled in terms of processing speed, recording an average precision of 65.33% and recall of 62.78%. Among the three models, it proved most effective in handling imbalanced data. This advantage was enhanced by the application of the SMOTE method, which successfully improved class distribution, enabling LightGBM to better identify patterns in minority classes. Nonetheless, while its speed is a major strength, LightGBM requires complex hyperparameter tuning to achieve optimal performance.

c) Random Forest

Random Forest achieved the highest accuracy during testing at 90.09%, demonstrating its reliability in handling complex data patterns. The model showed high stability even with datasets containing significant variation. Its ensemble mechanism effectively prevents overfitting, making it a robust choice for text classification. However, while its precision and recall were only slightly lower than those of LightGBM, this suggests that Random Forest may be less effective in detecting minority classes in imbalanced datasets.

Overall, the findings indicate that LightGBM is the most suitable algorithm for applications where processing speed and the ability to handle imbalanced data are critical. Its strong precision and recall highlight its capability to deliver better results for minority classes, which often present challenges in text classification tasks. Random Forest, with its highest overall accuracy, offers exceptional stability in recognizing complex data patterns and is best suited for scenarios where accuracy takes precedence over speed. Logistic Regression remains a viable option for applications requiring a straightforward, efficient, and interpretable model.

4. CONCLUSION

This study demonstrates that machine learning algorithms can be effectively utilized to classify customer complaints at PERUMDAM, providing a technology-based solution capable of overcoming classification challenges and improving the efficiency of data processing. The evaluation of Logistic Regression, LightGBM, and Random Forest algorithms reveals that each has distinct strengths and limitations, making them adaptable to specific operational requirements. LightGBM emerged as the most suitable algorithm for handling imbalanced datasets due to its high processing speed and accurate predictions, making it the primary choice for implementing technology-driven complaint management systems. Random Forest offers advantages in terms of stability and overall accuracy, making it ideal for applications that prioritize the identification of complex data patterns. Meanwhile, Logistic Regression serves as a simple yet efficient alternative, particularly for applications requiring clear model interpretation.

The findings provide practical value for the development of more responsive and efficient customer complaint management systems, while also contributing academically by broadening the understanding of machine learning applications in text data classification. Furthermore, the results hold significant social relevance by supporting improvements in customer satisfaction and public trust in service delivery. In conclusion, the choice of the most appropriate algorithm depends on the specific needs of the application, whether in terms of speed, accuracy, or interpretability. This research is expected to serve as a strong foundation for the implementation of technology-based solutions in the public service sector and to inspire further studies in the field of text data classification.

REFERENCES

Alsubayhin, A., Ramzan, M. S., & Alzahrani, B. (2024). Crime prediction model using three classification techniques: Random Forest, Logistic Regression, and LightGBM. International Journal of Advanced Computer Science and Applications, 15(1), 240–251.

Baihaqi, W. M., Yunita, I. R., Damayanti, A. S. T., & Akhaerunnisa, L. (2023). Analysis of classification algorithm performance on user review sentiment of the Muamalat DIN application. CogITo Smart Journal, 9(2), 241–251.

Empiris, S., Generasi, P., Di, Y. D. A. N. Z., & Jabodetabek, W. (2023). Jurnal Kewidyaiswaraan, 8(62), 45–57.

Hutauruk, W. I. P., Nabila, A. N., & Furqan, M. (2023). Evaluasi efektivitas sistem informasi pengaduan pelanggan PDAM Tirtanadi Cabang Medan Kota berbasis web. Jurnal Informatics Business, 1(3), 124–131.

Istiqamah, N., & Rijal, M. (2024). Klasifikasi ulasan konsumen menggunakan Random Forest dan SMOTE. Journal of System and Computer Engineering, 5(1), 66–77.

Lestari, N., & Fathan, M. R. S. (2024). Penerapan algoritma klasifikasi untuk menentukan segmentasi pelanggan berbasis machine learning. Jurnal Teknologi dan Sistem Informasi, 4.

Muthohhar, J. D., & Prihanto, A. (2023). Analisis perbandingan algoritma klasifikasi untuk penyakit jantung. Jurnal Informatics and Computer Science, 4, 298–304.

Pulungan, M. P., Purnomo, A., & Kurniasih, A. (2023). Penerapan SMOTE untuk mengatasi imbalance class dalam klasifikasi kepribadian MBTI menggunakan Naive Bayes classifier. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(7), 1493–1502.

Saepudin, A., Faqih, A., & Dwilestari, G. (2024). Perbandingan algoritma klasifikasi Support Vector Machine, Random Forest dan Logistic Regression pada ulasan Shopee. Jurnal Tekno Kompak, 18(1), 178.

Sari, L., Romadloni, A., Lityaningrum, R., & Hastuti, H. D. (2023). Implementation of LightGBM and Random Forest in potential customer classification. TIERS Information Technology Journal, 4(1), 43–55.