

## An advanced logistic regression model for forecasting payer revenue in private hospitals: a case study in manado

Hartiny Pop Koapaha<sup>1</sup>, Wilsen Grivin Mokodaser<sup>2</sup>

<sup>1</sup>Faculty of Economics and Business, Universitas Klabat, Indonesia

<sup>2</sup>Faculty of Computer Science, Universitas Klabat, Indonesia

### Article Info

#### Article history:

Received February 21, 2025

Revised March 4, 2025

Accepted March 6, 2025

#### Keywords:

Hospitals

Forecasting

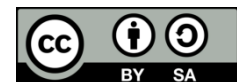
Logistic regression

Confusion Matrix

### ABSTRACT

Manado, the provincial capital, stands as a vital center for healthcare services, where private hospitals compete intensively to attract patients from various economic and social backgrounds. Accurate revenue forecasting for partnered payers is essential for effective management strategies. This study employs a logistic regression model, achieving a notable accuracy of 79.55% in predicting hospital revenue based on payer partnerships. The confusion matrix reveals 21 true negatives (TN), confirming the model accurately identified low-revenue customers, with zero false positives (FP), indicating no misclassification of these individuals. However, 9 false negatives (FN) highlight a critical risk, as high-revenue customers were miscategorized as low revenue, even though 14 true positives (TP) were precisely identified. Based on these insights, hospitals can strategically target 61 payers projected to exceed median revenue, presenting a significant opportunity for income growth. Conversely, the 159 payers identified as below median revenue warrant urgent attention. To enhance engagement and increase revenue from these lower-revenue groups, targeted business strategies such as intensified marketing, personalized service offerings, and promotional discounts are recommended. This research contributes a novel approach to leveraging predictive analytics in healthcare, underscoring the pressing need for hospitals to innovate their operational strategies to optimize revenue in a competitive landscape.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Wilsen Grivin Mokodaser,

Fakultas Ilmu Komputer,

Universitas Klabat,

Jl Arnold Mononutu, Airmadidi Bawah, Kec. Airmadidi, Kabupaten Minahasa Utara, Sulawesi Utara 95371

Email: [wilsenm@unklab.ac.id](mailto:wilsenm@unklab.ac.id)

<https://doi.org/10.52465/joscecx.v6i1.555>

## 1. INTRODUCTION

The healthcare industry, particularly hospitals, has been experiencing rapid growth in line with the increasing public demand for quality healthcare services. Competition among hospitals, both private and public, has become more intense due to factors such as advancements in medical technology, the use of artificial intelligence in the field of hospital services which can influence satisfaction with the services provided [1], healthcare policies, changing disease patterns, and heightened public awareness of the importance of health [2].

Private hospitals, in particular, face challenges in remaining competitive amid an increasingly dynamic market [3]. The key factors influencing the competitiveness of private hospitals include service quality, service fees, ease of access, and patient satisfaction [4]. Additionally, hospitals must effectively manage human resources, improve operational efficiency, and adopt the latest technologies to provide better services [5]. Another equally important factor is marketing and branding strategies, which can attract more patients and build their loyalty to the hospital [6]. In a business context, private hospitals often rely on various types of payers, such as health insurance providers, national health coverage programs, and self-paying patients [7]. With a population of approximately 2.66 million, the province of North Sulawesi requires healthcare facilities capable of meeting the needs of all its residents [8].

With increasing competition in the healthcare sector, private hospitals in Manado also face similar pressures to enhance their efficiency and profitability. North Sulawesi has several private hospitals that play a crucial role in providing healthcare services to the community [9]. Manado, as the provincial capital, serves as the center of healthcare services, with various private hospitals offering high-quality medical facilities and services. Private hospitals in this region compete in terms of service quality, medical technology, and business strategies to attract patients from diverse economic and social backgrounds. Some of the well-known private hospitals in North Sulawesi, such as Siloam Hospital Manado, Advent Hospital Manado, and Hermina Hospital Manado, have adopted modern technology and more efficient service systems. They also collaborate with various health insurance providers to facilitate patient access [10].

Several previous studies have explored how algorithms and research methods can predict the revenue of various organizations and companies. One such study focused on predicting potential donors using a logistic regression model, achieving a model accuracy of 0.6129, which indicates good predictive performance [11]. Another study discussed the application of machine learning using logistic regression algorithms for diabetes prediction, which achieved an accuracy rate of 82% [12]. Another study utilized decision tree and logistic regression algorithms to predict employee promotions at Prasama Bhakti Foundation, finding that the decision tree model outperformed logistic regression in accuracy [13]. Additionally, a study conducted sentiment analysis on Twitter users using a logistic regression model, achieving an accuracy rate of 78.57% [14]. Furthermore, a study on web-based SMS classification using a logistic regression algorithm was conducted on 1,140 labeled messages, achieving an accuracy rate of 97% [15].

Considering these related studies, hospitals need to implement the right strategies to predict revenue from various payer sources to better manage their finances. Forecasting payer revenue is a crucial aspect of financial planning and strategic decision-making, enabling hospitals to ensure operational sustainability and enhance the quality of their services.

## **2. METHOD**

This research was developed by considering the stages that will be undertaken to achieve optimal results. Below is an overview of the research process used in the research workflow.

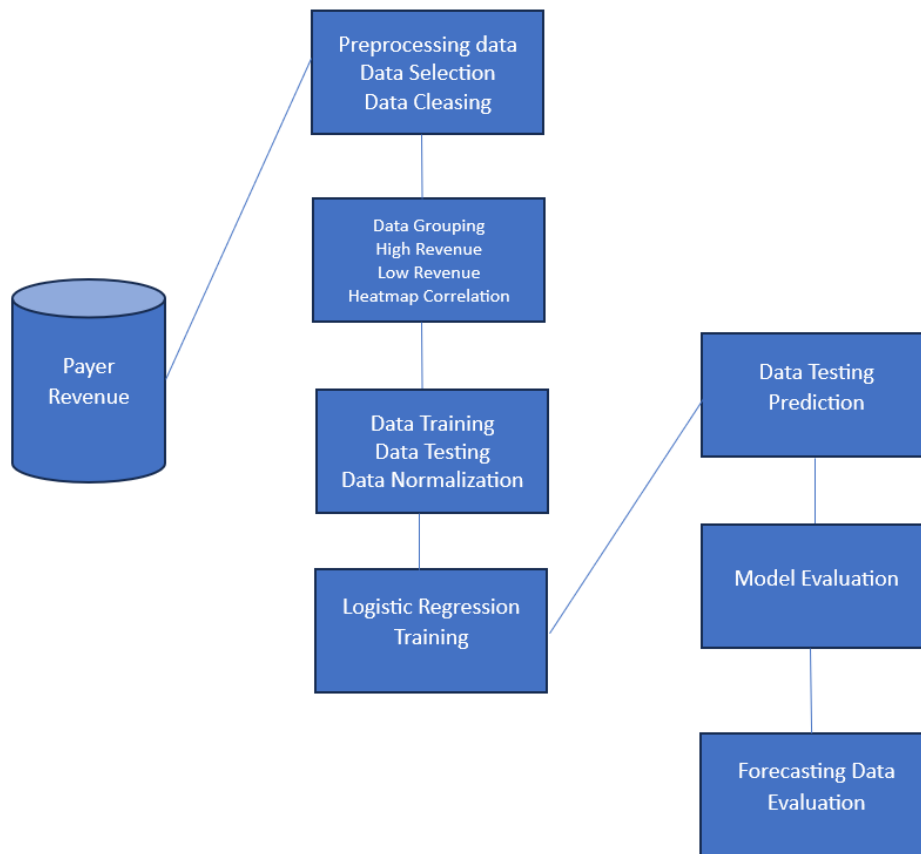


Figure 1. Research flow

Figure 1 illustrates the research flow as follows:

#### **Revenue Data**

At this stage, a data tabulation process is conducted, consisting of revenue data from each payer collaborating with the hospital. This data includes payer information and the total revenue generated throughout the partnership.

#### **Data Preprocessing**

In this stage, the collected data is reviewed to identify any data that requires cleaning. If there are missing values or unnecessary data, appropriate actions will be taken according to the research needs [16].

#### **Data Grouping**

Once the data has undergone preprocessing, the payers are grouped into categories based on revenue—high revenue and low revenue—before applying the predictive algorithm [17].

#### **Data Training and Normalization**

At this stage, the grouped data is split into training and testing datasets, with an 80% allocation for training and 20% for testing. The 80/20 training and testing data split is commonly used in machine learning because it provides a balanced approach to training and evaluation. 80% training data allows the model to learn patterns effectively, while 20% testing data ensures sufficient evaluation without sacrificing too much training data. This ratio helps prevent overfitting, as using too much training data (e.g., 90-95%) may cause the model to perform well on training data but poorly on new data, while using too much testing data (e.g., 40-50%) may limit the model's learning ability. The 80/20 split is also a standard in industry and research, offering a good balance in most cases, though variations like 70/30 or 90/10 may be used depending on dataset size and specific needs. Additionally, this split maintains computational efficiency, allowing the model to learn effectively without excessive resource consumption. Overall, the 80/20 ratio is an effective compromise between optimal training and accurate evaluation. After splitting, data normalization is performed [18].

#### **Algorithm Implementation**

At this stage, the predictive algorithm—logistic regression—is applied to predict the revenue generated by each payer. Using the following equation [19].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (1)$$

Where:

$Y$  = Response Variable (the affected variable)

$X_i$  = Predictor Variable (the influencing variable)

$\beta_0$  = Intercept

$\beta_1$  = Regression Coefficient of the Predictor Variable

### Data Testing Prediction

After implementing the logistic regression model, predictions are made using the predefined test data. The prediction process for testing data using Logistic Regression begins with data preprocessing, where numerical features are processed by removing missing values using SimpleImputer and normalized using StandardScaler to ensure a better distribution. Next, the dataset is split into 80% Training Set and 20% Testing Set, allowing the model to learn from the training data and be tested on new data. The model is then trained using  $X_{train\_scaled}$  and  $y_{train}$ , where Logistic Regression optimizes the weights for each feature to predict the probability of a class (0 = Low Revenue, 1 = High Revenue). Subsequently, the model performs predictions on test data by calculating the probability of each sample belonging to the "High Revenue" class; if the probability is  $\geq 0.5$ , the class is predicted as 1 (High Revenue), and if  $< 0.5$ , the class is predicted as 0 (Low Revenue) [20].

### Model Evaluation

In this stage, the generated model is evaluated using a confusion matrix to determine the accuracy of the applied predictive model [21].

### Algorithm Implementation

At this stage, the predictive algorithm—logistic regression—is applied to predict the revenue generated by each payer. Using the following equation [19].

	Prediction:		
	No	Yes	
Actual: No	TN	FP	$Y1 = TN + FP$
Actual: Yes	FN	TP	$Y2 = FN + TP$
	$X1 = TN + FN$	$X2 = FP + TP$	

TP = True Positive

TN = True Negatif

FP = False Positive

FN = False Negative

### Data Evaluation

Based on the prediction results, each payer will be analyzed based on the predicted values. The payers will then be regrouped into high-revenue and low-revenue categories. From these results, it is expected that the hospital can use them to develop marketing or service strategies for each payer with a potential decline in revenue and those likely to experience revenue growth.

## 3. RESULTS AND DISCUSSIONS

In this study, the implementation process was carried out using the Python programming language and Google Colab for data analysis and model application, resulting in the following:

### Data Tabulation

At this stage, a total of 220 data entries will be processed, consisting of payers collaborating with the hospital and the revenue generated by each payer.

Tabel 2. Payer data

Payer	IGD	MCU	Rawat Inap	Rawat jalan
BIOMEDILAB KLINIK, PT	2.766.600	2.650.000	2875.250	2.806.350
PLN (PERSERO) DISTRIBUSI JAWA TIMUR, PT	4.940.846	4.891.373	4.945.633	4.787.641
PLN UIP MALUKU, PT	4.094.383	3.944.746	6.614.930	1.124.925
ARTA BOGA CEMERLANG, PT	2.274.750	4.431.512	2.297.498	22.511.245
MATAHARI PUTRA PRIMA TBK, PT	1.087.785	1.035.000	1.070.190	1.042.245
PT - REDPATH INDONESIA	17.923.905	16.945.348	16.602.853	16.309.286
AJ INHEALTH INDONESIA, PT	2.699.667	21.089.187	39.352.552	21.215.344
ADM-ASKRIDA	5.034.309	69.447.886	156.274.891	47.034.459
AIA FINANCIAL INDONESIA	6.011.261	126.863.143	327.169.250	47.408.918
Dst...				

When exporting data to Google Colab, an initial data exploration is conducted, as shown in Figure 2 below:

```

♦ Data Awal:
      Payer      Emergency ...      Inpatient      Outpatient
0      BIOMEDILAB KLINIK, PT 2766600.000 ... 2.875250e+06 2806350.0
1      PLN (PERSERO) DISTRIBUSI JAWA TIMUR, PT 4940845.512 ... 4.945633e+06 4787641.0
2      PLN UIP MALUKU, PT 4094383.295 ... 6.614930e+06 1124925.0
3      ARTA BOGA CEMERLANG, PT 2274750.000 ... 2.297498e+06 2297497.5
4      LIPPO GENERAL INSURANCE, PT - MEARES SOPUTAN ... 9056901.000 ... 3.359739e+08 214292169.0

[5 rows x 5 columns]

♦ Info Dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 219 entries, 0 to 218
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ---          -
0   Payer           219 non-null   object
1   Emergency       219 non-null   float64
2   Health Checkup 219 non-null   float64
3   Inpatient       219 non-null   float64
4   Outpatient      218 non-null   float64
dtypes: float64(4), object(1)

```

Figure 2. Data Exploration in colab

### Data Preprocessing

During the data preprocessing stage, it was found that some columns had missing values. These missing values were handled by filling NaN values with the median. Additionally, data normalization was performed using a scaler in Python, as shown in Figure 3 below.

```
imputer = SimpleImputer(strategy="median")
X_train_imputed = imputer.fit_transform(X_train)
X_test_imputed = imputer.transform(X_test)

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train_imputed)
X_test_scaled = scaler.transform(X_test_imputed)
```

Figure 3. Data preprocessing

### Data Grouping

At this stage, grouping is performed to analyze the composition of the data to be processed and to observe the distribution based on high and low revenue. As shown in Figure 4, the plot of payer distribution based on revenue indicates that before prediction, the number of payers categorized as Low Revenue was 110, while the number of payers categorized as High Revenue was 109.



Figure 4. Grouping revenue

After visualizing the data based on revenue, a correlation analysis between variables was conducted using a correlation heatmap, as shown in Figure 5 below. This heatmap illustrates the relationship between different variables, where a value of 1 indicates a strong positive correlation—meaning that as one variable increases, the other also increases. Conversely, a value of -1 represents a strong negative correlation, meaning that as one variable increases, the other decreases. Through this correlation heatmap, we can also observe that the revenue class has a high correlation with certain features, indicating that these features significantly influence the classification of revenue.

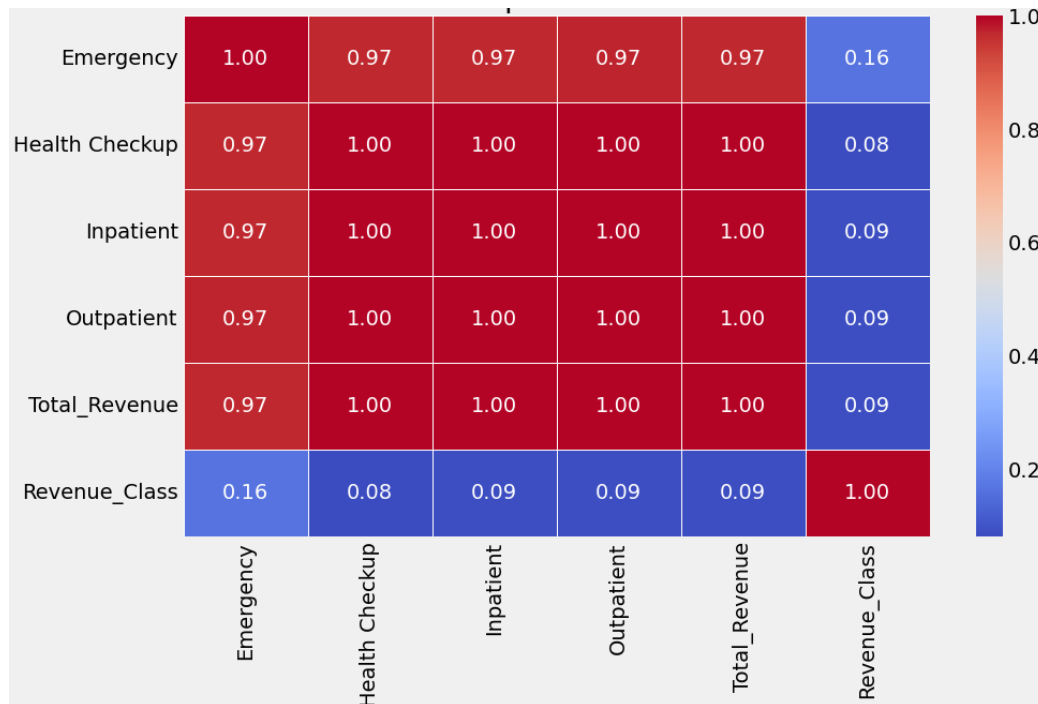


Figure 5. Heatmap correlation

### Data Training and Normalization

After analyzing the data to be processed, the dataset is split into 80% for training and 20% for testing, using a random state value of 42. Then, data normalization is performed using a scaler in Python.

### Algorithm Implementation

Once the training and testing data are determined and normalization is completed, the next step is training the model using the logistic regression algorithm. Using Python, the model training process is successfully conducted, as shown in Figure 6 below.

```
#Latih Model Logistic Regression
model = LogisticRegression()
model.fit(X_train_scaled, y_train)
```

Figure 6. logistic regression training

### Data Testing Prediction and Model Evaluation

After training the model, the next step is to make predictions on the test data by defining  $y\_pred = model.predict(X\_test\_scaled)$ . This is followed by evaluating the generated model, which achieved an accuracy of 79.55%, with the resulting confusion matrix shown in Figure 7 below.

```

✅ Akurasi Model: 79.55%

♦ Confusion Matrix:
[[21  0]
 [ 9 14]]

♦ Classification Report:
              precision    recall  f1-score   support

     0         0.70      1.00      0.82         21
     1         1.00      0.61      0.76         23

 accuracy          0.80         44
 macro avg          0.85         44
 weighted avg       0.86         44

```

Figure 7. Confusion matrix, precision, recall F1 score and support

In Figure 7, the classification report results show that the Logistic Regression model has an accuracy of 79.55%, meaning that 79.55% of all predictions are correct. From the confusion matrix, the model successfully classified 21 Low Revenue samples correctly and 14 High Revenue samples correctly, but there were 9 misclassifications, where *High Revenue* was classified as *Low Revenue*. In terms of precision, the *Low Revenue* class has a value of 0.70, while *High Revenue* reaches 1.00, meaning that all instances predicted as *High Revenue* are actually correct. However, the recall for *High Revenue* is only 0.61, indicating that the model correctly identifies only 61% of the total *High Revenue* samples, while the rest are misclassified as *Low Revenue*. The F1-score for *Low Revenue* is 0.82, while for *High Revenue* it is 0.76, reflecting a balance between precision and recall. The macro average for precision, recall, and F1-score is 0.85, 0.80, and 0.79, respectively, while the weighted average has values of 0.86, 0.80, and 0.79. Overall, the model performs fairly well in classifying *Low Revenue* but still struggles to correctly detect *High Revenue*, which could impact business decision-making.

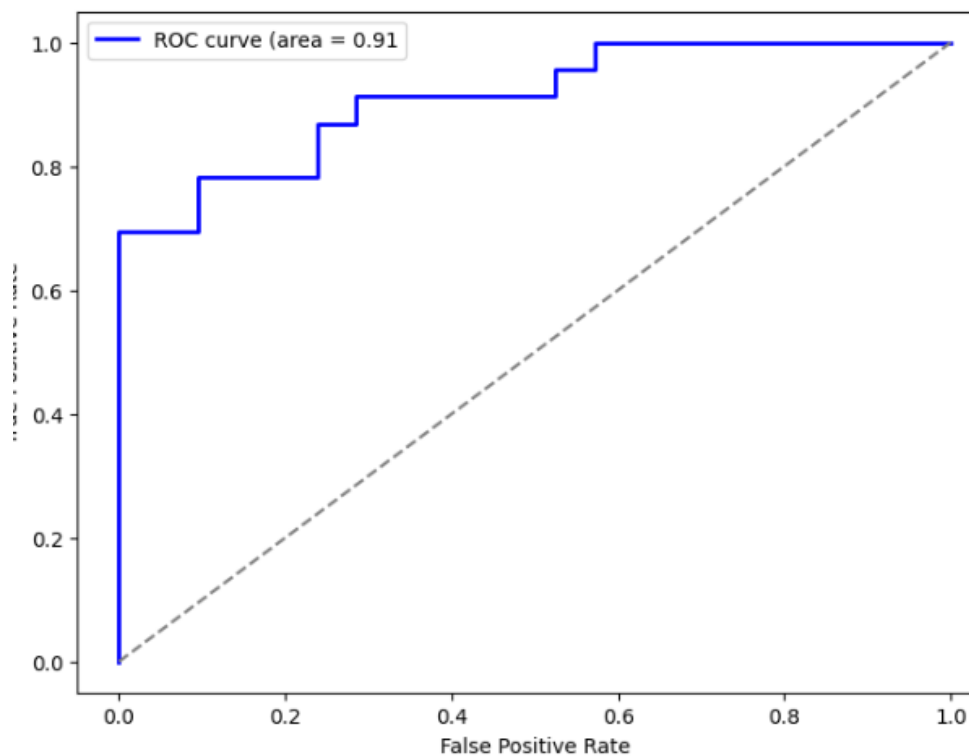


Figure 8. ROC curve

The results of the ROC (Receiver Operating Characteristic) curve above show the performance of the Logistic Regression model in distinguishing between the High Revenue and Low Revenue classes. The X-axis (False Positive Rate - FPR) represents the error rate where the model incorrectly classifies Low Revenue as High Revenue, while the Y-axis (True Positive Rate - TPR) indicates how well the model correctly identifies High Revenue. If the ROC curve approaches the upper left corner of the graph (low FPR, high TPR), the model has good performance, whereas if the curve is close to the diagonal line, the model is only as good as random guessing. The Area Under Curve (AUC) measures how well the model separates the two classes, where an AUC value close to 1 indicates excellent prediction, while an AUC around 0.5 suggests that the model is no better than random guessing. If the AUC value is high (e.g., above 0.8), the model has good predictive capability; however, if it is low, the model can be improved by adjusting features or using a more complex model.

#### 4. CONCLUSION

From the application of the logistic regression algorithm, it can be concluded that the model's accuracy of 79.55% is a fairly good result in predicting hospital revenue based on partnered payers. The model effectively identifies low-revenue payers, as there are no false positives—meaning no errors in predicting low-revenue customers as high-revenue ones. However, while the model performs well in avoiding false positives, it has a relatively high number of false negatives. The number of payers with the potential for increased revenue is 61 out of 109 payers who were initially classified as high-revenue payers. Based on these predictions, the hospital can use these insights to focus more on these 61 payers, who are expected to contribute additional revenue in the following year. The 61 potential payers who present an opportunity to increase revenue were identified after applying regression, as their revenue was predicted to be above the median value. Meanwhile, 158 payers, after applying logistic regression, had their revenue predicted to be below the median value. Additionally, the hospital can reassess its service strategies to ensure that the remaining 158 payers, who are likely to fall into the low-revenue category, continue their partnerships and potentially increase their revenue contributions. Based on the obtained results, it is recommended that decision-makers implement more precise business strategies to maintain revenue opportunities from the high-revenue category. Additionally, for payers who are likely to experience a decline in revenue, the business can introduce specific policies, such as more intensive visits from the marketing team to offer services like discounts or price reductions for each visit.

#### REFERENCES

- [1] L. Irgang, A. Sestino, H. Barth, and M. Holmén, "Healthcare workers' adoption of and satisfaction with artificial intelligence: The counterintuitive role of paradoxical tensions and paradoxical mindset," *Techmol. Forecast. Soc. Change*, vol. 212, p. 123967, Mar. 2025, doi: 10.1016/j.techfore.2024.123967.
- [2] B. C. Pratama, F. Bagis, M. Retnaningrum, and M. N. Innayah, "Peningkatan Kompetensi Usaha Mikro Kecil dan Menengah (UMKM) Berbasis Analisa Studi Kelayakan Bisnis," *BERDIKARI J. Inov. dan Penerapan Ipteks*, vol. 7, no. 2, 2019, doi: 10.18196/bdr.7262.
- [3] M. A. Suawah, "ANALISIS SISTEM INFORMASI AKUNTANSI PENERIMAAN KAS DALAM MENINGKATKAN PENGENDALIAN INTERNAL PADA RUMAH SAKIT GMIM SILOAM SONDER," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 9, no. 3, 2021, doi: <https://doi.org/10.35794/emba.v9i3.35785>.
- [4] U. K. Nisak, *Buku Ajar Statistik Di Fasilitas Pelayanan Kesehatan*. Umsida Press, 2020. doi: 10.21070/2020/978-623-6833-94-0.
- [5] N. K. C. PRATIWI, N. IBRAHIM, and S. SAIDAH, "Prediksi Kanker Paru menggunakan Grid search untuk Optimasi Hyperparameter pada Algoritma MLP dan Logistic Regression," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 12, no. 3, p. 556, Jul. 2024, doi: 10.26760/elkomika.v12i3.556.
- [6] R. Fauzan, R. Supryanita, and Rahmatika, "Analisis Strategi Pemasaran untuk Peningkatan Daya Saing Pada Bisnis Kafe di Kota Bukittinggi (Studi Kasus Kafe Teras Kota), 2021," *J. Manaj. Bisnis Syariah*, vol. 1, no. 1, pp. 1–14, 2021.
- [7] Monika Shanty Ista Purta, Nur Fitriyah, and Adhitya Bayu Suryantara, "ANALISIS PENGAKUAN, PENGUKURAN DAN PENGUNGKAPAN PENDAPATAN PADA RUMAH SAKIT UMUM DAERAH KOTA MATARAM," *J. Ris. Mhs. Akunt.*, vol. 2, no. 1, pp. 29–42, Jun. 2022, doi: 10.29303/risma.v2i1.183.
- [8] B. P. S. S. Utara, "Jumlah Penduduk Menurut Kabupaten/Kota (Ribu Jiwa)," *sulut.bps.go.id*, 2024. <https://sulut.bps.go.id/id/statistics-table/2/NDUjMg==/jumlah-penduduk-menurut-kabupaten-kota.html> (accessed Jan. 24, 2025).
- [9] B. P. Statistik, "Nama Rumah Sakit dan Kapasitas Tempat Tidur di Kota Manado, 2019-2020," *manadokota.bps.go.id*, 2021. <https://manadokota.bps.go.id/id/statistics-table/2/MTI5IzI=/nama-rumah-sakit-dan-kapasitas-tempat-tidur-di-kota-manado.html> (accessed Jan. 25, 2025).
- [10] D. Dahlia, E. M. Putri, and F. Edri, "Analisis Penerapan Akuntansi Lingkungan Pada Rumah Sakit Islam Ibnu Sina Bukittinggi," *J. BANSI - J. Bisnis Manaj. Akuntansi*, vol. 1, no. 1, pp. 28–38, Dec. 2021, doi: 10.58794/bns.v1i1.34.
- [11] S. rahmah Jabir, "Prediksi Potensi Donatur Menggunakan Model Logistic Regression," *Indones. J. Data Sci.*, vol. 4, no. 1, pp. 31–37, Mar. 2023, doi: 10.56705/ijodas.v4i1.64.
- [12] Aditya Pratama, A. C. Nurcahyo, and L. Firgia, "Penerapan Machine Learning dengan Algoritma Logistik Regresi untuk Memprediksi Diabetes," *Pros. CORISINDO 2023*, no. SE-Article, Aug. 2023, [Online]. Available: <https://ojs.stmikpontianak.ac.id/corisindo/article/view/30>
- [13] E. D. Anggara, A. Widjaja, and B. R. Suteja, "Prediksi Kinerja Pegawai sebagai Rekomendasi Kenaikan Golongan dengan Metode Decision Tree dan Regresi Logistik," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, Apr. 2022, doi:

- 10.28932/jutisi.v8i1.4479.
- [14] S. A. Assaidi and F. Amin, "Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Persen pada Pengguna Twitter menggunakan Metode Logistic Regression," *J. Pendidik. Tambusai*, vol. 6, no. 2 SE-Articles of Research, pp. 13217–13227, Jun. 2022, doi: 10.31004/jptam.v6i2.4543.
- [15] F. D. Pramakrisna, F. D. Adhinata, and N. A. F. Tanjung, "Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression," *Teknika*, vol. 11, no. 2, pp. 90–97, Jun. 2022, doi: 10.34148/teknika.v11i2.466.
- [16] A. Mutiarachim and Jaluanto Sunu Punjul Tyoso, "Optimasi Prediksi Pemasaran Nasabah Deposito Bank dengan Metode Klasifikasi Logistic Regression," *J. Cakrawala Inf.*, vol. 4, no. 1, pp. 20–28, Jun. 2024, doi: 10.54066/jci.v4i1.390.
- [17] F. Handayani, "Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. 329, Dec. 2021, doi: 10.26418/jp.v7i3.48053.
- [18] Brury Barth Tangkere, "Analisis Performa Logistic Regression dan Support Vector Classification untuk Klasifikasi Email Phising," *J. Ekon. Manaj. Sist. Inf.*, vol. 5, no. 4, pp. 442–450, Apr. 2024, doi: 10.31933/jemsi.v5i4.1916.
- [19] M. S. Ayu Hapsary, S. Subiyanto, and H. S. Firdaus, "ANALISIS PREDIKSI PERUBAHAN PENGGUNAAN LAHAN DENGAN PENDEKATAN ARTIFICIAL NEURAL NETWORK DAN REGRESI LOGISTIK DI KOTA BALIKPAPAN," *J. Geod. Undip*, vol. 10, no. 2, Apr. 2021, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/geodesi/article/view/30637>
- [20] A. Bimantara and T. A. Dina, "Klasifikasi Web Berbahaya Menggunakan Metode Logistic Regression," in *Annual Research Seminar (ARS)*, Universitas Sriwijaya, Fakultas Ilmu Komputer, 2019. [Online]. Available: <https://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1932>
- [21] D. Y. Utami, E. Nurlelah, and F. N. Hasan, "Comparison of Neural Network Algorithms, Naive Bayes and Logistic Regression to predict diabetes," *J. INFORMATICS Telecommun. Eng.*, vol. 5, no. 1, pp. 53–64, Jul. 2021, doi: 10.31289/jite.v5i1.5201.