

Renewal Prediction For Individual Property Insurance Using Binary Logistic Regression in PT XYZ Insurance

Mutia Karunia Antap Wulan¹, Bambang Permadiyansach², Suududdin³, Jerry Heikal⁴

^{1,2,3,4} Magister Management Program, Universitas Bakrie Jakarta, Indonesia

E-mail: mutia.karunia2@gmail.com¹, n4n4nk@gmail.com², soehood@gmail.com³,
jerry.heikal@bakrie.ac.id⁴

Article History:

Received: 01 Agustus 2025

Revised: 19 Agustus 2025

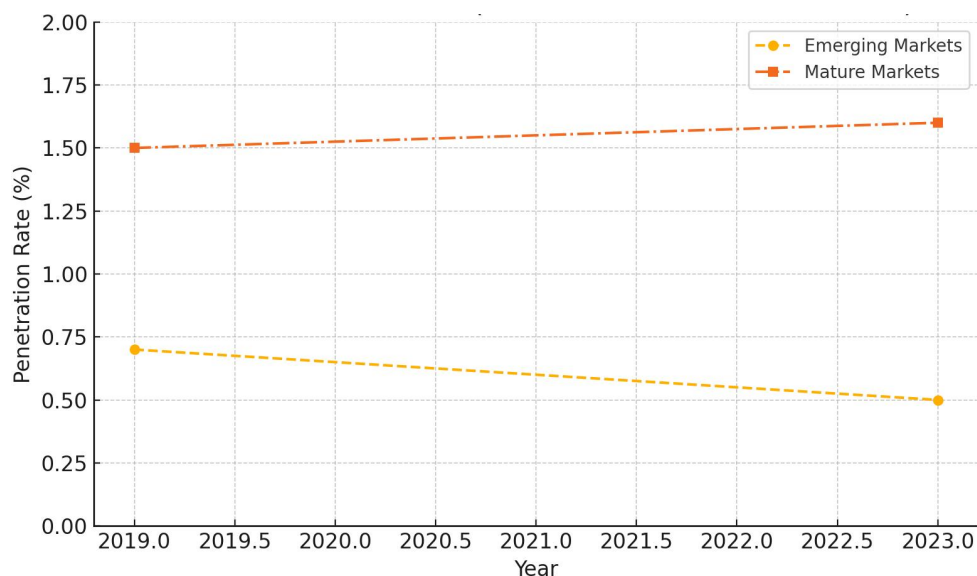
Accepted: 21 Agustus 2025

Keywords: Binary logistic, Renewal, Insurance Products, Customer Retention, Property Insurance

Abstract: Individual property insurance is a crucial product that offers financial protection for personal assets such as residential buildings, commercial properties, and private dwellings against risks such as fire, theft, and natural disasters. However, retaining customers and encouraging the renewal of policies continues to be a critical challenge in the insurance industry. This study aims to predict renewal decisions of individual property insurance policyholders at PT XYZ Insurance using a binary logistic regression model and to identify key factors that influence these decisions. The model was built using 178 policyholder records and evaluated several independent variables, including premium, discount, and admin cost. The analysis indicates that discount and admin cost are statistically significant in influencing renewal behavior. Discount, with a p-value of 0.034, shows a negative relationship suggesting that higher discounts may reduce the likelihood of renewal. On the other hand, admin cost, with a p-value of 0.053, shows a positive association, indicating that higher operational spending may contribute to increased renewal probability. The premium variable was found to be statistically insignificant in this model. These findings underline the importance of pricing strategy and cost structure in influencing customer retention. The study provides practical implications for insurance companies to refine their marketing and pricing strategies, particularly by reassessing discount allocation and expense effectiveness. Future research may incorporate demographic, behavioral, and regional variables to enhance the robustness and accuracy of renewal prediction in the property insurance domain.

INTRODUCTION

Personal property insurance serves as a crucial form of financial protection against risks such as fire, burglary, earthquakes, and other unforeseen events that can cause damage or loss to privately owned assets like houses, apartments, and private buildings. Despite the increasing awareness of risk mitigation, the insurance penetration rate in Indonesia remains relatively low. As reported by McKinsey & Company (2025), “in emerging economies, personal lines market growth in the 2022–2023 period lagged behind GDP by five percentage points, and average 2023 relevance remained below 0.5 percent of GDP.”



Graphic 1. Insurance Penetration (Personal Lines Premium / GDP)

The insurance penetration rate for personal lines (measured as premiums to GDP):

- The dashed yellow line (Emerging Markets) illustrates a decline from 0.7% in 2019 to 0.5% in 2023, reflecting stagnant public participation in personal insurance products in developing countries such as Indonesia.
- The dotted orange line (Mature Markets) shows a slight increase from 1.5% to 1.6%, indicating a more stable and developed insurance sector.

This graph highlights the gap between market potential and actual uptake of personal insurance in emerging economies. This indicates that although the market potential is considerable, the public's participation and uptake of property insurance are still limited.

One of the major challenges faced by general insurance companies, especially in the personal property segment, is the low policy renewal rate. Upon expiration, many policyholders choose not to renew their insurance contracts for reasons such as affordability, dissatisfaction with claims experiences, or the lack of personalized engagement from insurers. As reported by MediaAsuransi.com (2023), “poor claims experience, such as slow and non-transparent processes, has been a major factor contributing to policyholders’ decisions not to renew.”

From a pricing perspective, customers often feel that the annual premium burden is too high compared to the perceived benefits of coverage. A report from Investor.co.id (2024) stated that “many customers decided not to renew their policies because the premiums were deemed expensive and did not meet their protection expectations.” This points to a significant value-perception gap between insurers and policyholders.

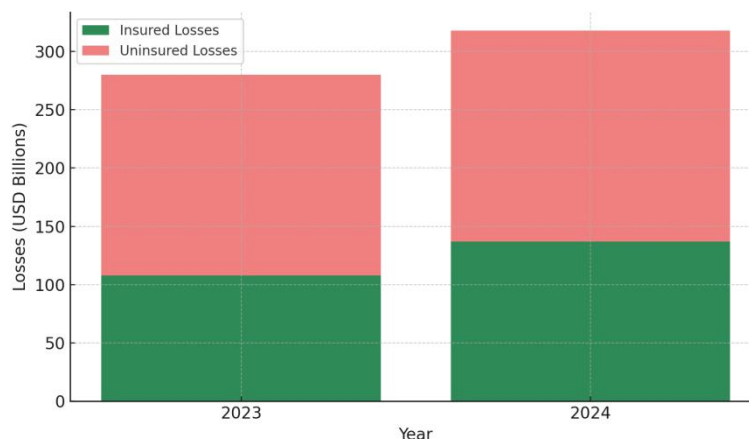
In addition, ineffective communication and education efforts have further worsened the situation. As noted by Bisnis.com (2024), “insurance companies have not maximized their financial literacy programs, especially among younger and urban consumer segments.” Without adequate understanding of insurance’s long-term value, many individuals deprioritize renewing their policies.

Another important factor is customer trust in insurance providers. Wagan and Sidra (2025) emphasize that “trust has a greater impact on customer loyalty than service quality,” and that “insurance companies need to scrutinize customer loyalty to ensure that their customers renew their insurance policies.” In the Indonesian context, a trust gap has been repeatedly cited as a root cause for low policyholder retention, especially when the insurance process lacks transparency or responsiveness.

For insurance companies, customer attrition has direct implications for premium income fluctuations and increasing acquisition costs. While significant marketing investments are required to acquire new customers, failing to retain existing ones undermines long-term profitability. At the same time, regulatory constraints such as tariff and product structures mandated by SEOJK No. 6/SEOJK.05/2017 limit pricing flexibility and innovation.

While technology and digital transformation present opportunities for enhanced underwriting accuracy and marketing personalization, adoption in Indonesia remains uneven. The OECD (2023) noted that “the use of technologies such as artificial intelligence and digital engagement platforms can improve risk assessment and support insurers in offering more targeted products.” However, implementation is hampered by limitations in data access, workforce readiness, and data privacy concerns.

On a broader scale, low personal property policy renewal rates may also affect the country’s financial resilience to natural disasters. Aon (2025) reported that “annual global losses due to natural catastrophes reach hundreds of billions of dollars, yet only around 40% of those losses are insured.” This underscores the critical role of property insurance in building financial protection systems for households and businesses alike.

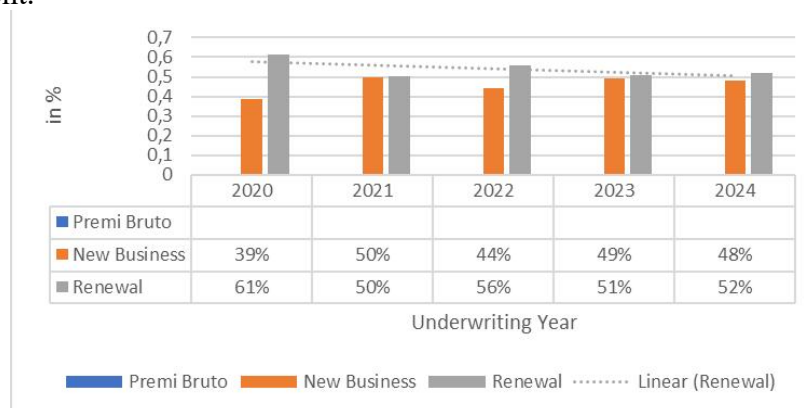


Graphic 2. Global Natural Disaster Losses: Insurance vs Uninsured (in USD Billions)

The chart shows global natural disaster losses in 2023 and 2024, distinguishing between insured and uninsured losses. It highlights the large protection gap, with a significant portion of losses remaining uninsured, reinforcing the critical need for improved insurance penetration worldwide.

Empirical studies highlight that predictive statistical models, especially binary logistic regression are effective in estimating the likelihood of policy renewal based on customer profiles and product attributes. Hendrawan et al. (2024) demonstrated that “binary logistic regression significantly identifies the key variables influencing policyholders’ renewal decisions.” Notably, customer age, premium size, policy tenure, location, and discount levels were identified as significant predictors of renewal behavior.

PT XYZ Insurance, a general insurance provider in Indonesia, has recently emphasized expanding its retail business line, including personal property insurance. As part of its business development efforts, the company seeks to enhance its customer retention strategy by identifying patterns in policy renewal decisions and leveraging predictive analytics. With a growing focus on digital innovation and customer-centric models, XYZ is well-positioned to adopt data-driven approaches to address its renewal rate challenges and strengthen its presence in the retail insurance segment.



Graphic 3. Trend Declining of Renewal Policy at XYZ Insurance

The graphic illustrates a consistent downward trend in the renewal rate of personal property insurance policies at XYZ Insurance from 2020 to 2024. While new business acquisition has remained relatively stable, with percentages fluctuating between 39% and 50%, the renewal rate has shown a clear decline—from 61% in 2020 to just 52% in 2024. This pattern indicates a challenge in sustaining customer retention and highlights an issue in persistency management. Despite ongoing efforts in underwriting and sales, the company appears to struggle with fostering long-term policyholder loyalty, which could impact the stability of its portfolio and future revenue streams.

Another root cause contributing to low policy renewal rates is the mismatch between insurance products and customer life stages. Many personal property insurance offerings in Indonesia are standardized and fail to evolve with policyholders’ changing needs. For example, young professionals purchasing their first property may require flexible, entry-level coverage, whereas families or retirees may prioritize comprehensive protection. The lack of product differentiation often results in customer disengagement once their initial contract expires.

Additionally, the absence of proactive retention efforts such as timely reminders, customized renewal packages, or loyalty rewards means insurers lose critical touchpoints with policyholders. According to a survey by InsuranceTech Asia (2023), over 60% of lapsed customers did not receive any personalized follow-up before their policy expired. This indicates that insurers are often reactive rather than strategic in managing their customer lifecycle.

From a service standpoint, the traditional claims process is still a pain point. Lengthy

documentation, rigid verification requirements, and slow settlement times create negative experiences, discouraging customers from renewing. Moreover, in Indonesia's diverse archipelagic setting, service availability in remote areas remains inconsistent, reinforcing a perception of inaccessibility among rural and semi-urban customers.

Behavioral economics also plays a role. Policyholders often underestimate the likelihood of rare but catastrophic events and overweight immediate expenses such as premium payments. This "present bias" reduces the perceived value of insurance, particularly among younger customers or those with limited financial literacy. Without behavioral nudges or structured renewal incentives, many customers default to non-renewal simply due to inertia.

Internally, many insurance companies operate with fragmented customer databases, limiting their ability to gain a unified view of the policyholder. This hampers efforts to conduct predictive analytics or deliver timely, relevant offers. The integration between marketing, underwriting, and claims departments is often suboptimal, resulting in disjointed communication and missed opportunities to engage customers effectively.

Taken together, these factors indicate that improving personal property policy renewal rates requires a multi-dimensional approach spanning product design, customer engagement, digital infrastructure, and operational efficiency. For companies like PT XYZ Insurance, addressing these root causes could unlock significant growth in the retail segment, enhance customer trust, and contribute to broader financial inclusion efforts in Indonesia.

The objective of this research is to build a data-driven predictive model that enables PT XYZ Insurance to proactively identify policyholders with high renewal probability and those at risk of lapsing, thereby allowing for tailored retention strategies and more efficient allocation of marketing resources.

RESEARCH METHODOLOGY

This study adopts a quantitative research approach utilizing binary logistic regression to analyze the factors that influence policyholder decisions to renew individual property insurance policies. The logistic regression method is appropriate for this analysis, as the dependent variable is binary in nature coded as 1 for renewal and 0 for non-renewal (Renewal Flag).

The dataset comprises 178 policyholders collected from internal insurance records. Before modeling, the data underwent a comprehensive preprocessing phase, which included data cleaning, variable normalization, and encoding to ensure all features were compatible with logistic regression requirements. This step aimed to eliminate data inconsistencies and ensure the accuracy of the model outputs.

The operating variables used in this binary logistic regression model include Premium, Discount, Admin Cost, and Occupation. These independent variables were selected based on their relevance in insurance pricing and policyholder characteristics. Each variable underwent preprocessing to fit the model specification, and their respective impacts on the likelihood of policy renewal were subsequently evaluated.

The binary logistic regression method estimates the probability that a policyholder will renew their policy based on these input variables. The relationship between the independent variables and the renewal likelihood is expressed using the following logistic regression function:

$$Y = \frac{1}{1 + \exp^{-(C + \beta_1 X_1 + \beta_2 X_2)}}$$

Where:

- Y = Probability of policy renewal (Renew_Flag)
- C = Intercept (constant term)
- β_1, β_2 = Coefficients of independent variables X_1, X_2
- X_1, X_2 = Independent variables (e.g., discount, cost)

All stages of data handling and model estimation were performed using Python. Binary logistic regression was used not only for its predictive capability but also for its interpretability in identifying statistically significant factors that affect renewal decisions. The model provides a basis for future customer retention strategies by highlighting which variables most influence the likelihood of renewal.

RESULT

This study employed a binary logistic regression approach to examine the renewal behavior of individual property insurance policyholders. The dependent variable, Renewal, represents a binary outcome indicating whether or not a customer decided to renew their insurance policy. The goal of the analysis was twofold: first, to identify which variables significantly influence the renewal decision, and second, to construct a reliable predictive model based on those significant indicators.

Table 1. Result from logistics binary regression with Python

Optimization terminated successfully.
Current function value: 0.539291
Iterations 14

Logit Regression Results

Dep. Variable:	Renew_Flag	No. Observations:	178
Model:	Logit	Df Residuals:	173
Method:	MLE	Df Model:	4
Date:	Sun, 22 Jun 2025	Pseudo R-squ.:	0.04618
Time:	05:15:17	Log-Likelihood:	-95.994
converged:	True	LL-Null:	-100.64
Covariance Type:	nonrobust	LLR p-value:	0.05414

	coef	std err	z	P> z	[0.025	0.975]
const	-4.5285	57.639	-0.079	0.937	-117.499	108.442
Premi Langsung	-2.784e-07	2.85e-07	-0.977	0.328	-8.37e-07	2.8e-07
Diskon	-1.078e-05	5.61e-06	-1.920	0.055	-2.18e-05	2.26e-07
Biaya	3.565e-05	1.91e-05	1.869	0.062	-1.75e-06	7.3e-05
Occupation	0.0016	0.019	0.082	0.935	-0.036	0.040

Classification Report:

	precision	recall	f1-score	support
0	0.57	0.09	0.15	45
1	0.76	0.98	0.86	133
accuracy			0.75	178
macro avg	0.67	0.53	0.50	178
weighted avg	0.71	0.75	0.68	178

Accuracy Score:
0.7528089887640449

In the first step of the analysis, a full logistic regression model was developed by incorporating all available independent variables: Premium, Discount, Admin Cost, and Occupation. The results of this full model provided an initial understanding of the relationship between these predictors and the likelihood of policy renewal. The model achieved convergence and returned a pseudo R-squared value of 0.0461, which, while modest, suggests some explanatory power. The regression coefficients for Premium and Occupation were not statistically significant, with p-values of 0.328 and 0.935, respectively. This indicates that these variables do not meaningfully contribute to the prediction of renewal behavior.

Interestingly, Discount and Admin Cost showed near-significant results with p-values of 0.055 and 0.062, respectively. Although these values are slightly above the conventional 0.05 threshold for significance, they suggested potential influence and warranted further investigation in a refined model.

Table 2. Result from logistics binary regression with Python for the second time.

```

Optimization terminated successfully.
Current function value: 0.543297
Iterations 5
  
```

Logit Regression Results						
Dep. Variable:	Renew_Flag	No. Observations:	178			
Model:	Logit	Df Residuals:	175			
Method:	MLE	Df Model:	2			
Date:	Sun, 22 Jun 2025	Pseudo R-squ.:	0.03909			
Time:	09:30:41	Log-Likelihood:	-96.707			
converged:	True	LL-Null:	-100.64			
Covariance Type:	nonrobust	LLR p-value:	0.01956			

	coef	std err	z	P> z	[0.025	0.975]
const	0.0844	0.610	0.138	0.890	-1.112	1.281
Diskon	-1.1575	0.545	-2.123	0.034	-2.226	-0.089
Biaya	3.6222	1.871	1.936	0.053	-0.044	7.288


```

Classification Report:
precision    recall  f1-score   support

     0       0.50      0.04      0.08        45
     1       0.75      0.98      0.85       133

 accuracy          0.75        178
 macro avg          0.63        178
 weighted avg          0.69        178
  
```



```

Accuracy Score:
0.7471910112359551
  
```

Based on these preliminary findings, a second logistic regression model was constructed using only the two most promising predictors: Discount and Admin Cost. This focused model aimed to enhance interpretability while retaining variables with meaningful effects. In the results of this second model, Discount emerged as statistically significant ($p = 0.034$), and Admin Cost approached significance ($p = 0.053$). The negative coefficient for Discount implies that higher discount amounts are associated with a lower probability of renewal, which may seem

counterintuitive but could reflect opportunistic behavior among certain policyholders who take advantage of promotions but do not remain loyal. On the other hand, Admin cost showed a positive coefficient, suggesting that customers who incurred higher administrative or policy-related costs were more likely to renew, possibly due to their greater investment or commitment to the policy.

To quantitatively estimate the renewal probability y , the final logistic regression model is expressed as:

$$\hat{y} = \frac{1}{1 + \exp(-(0.0844 - 1.1575 \cdot \text{Discount} + 3.6222 \cdot \text{Admin Cost}))}$$

Where:

- $\hat{y}$ = Predicted probability of policy renewal
- *Discount* = Standardized value of discount offered
- *Admin Cost* = Standardized value of administrative costs
- 0.0844 = Intercept (const)
- -1.1575 and 3.6222 = Estimated coefficients from the regression model

This model was applied to the entire dataset of 178 policyholders. The resulting predicted values y from this equation were interpreted to classify each policyholder's likelihood of renewal. A threshold of 0.5 was used to distinguish between renewal and non-renewal behavior:

- Policyholders with $y \geq 0.5$ were classified as likely to renew
- Those with $y < 0.5$ were classified as likely not to renew

Based on this classification rule and the model outputs, it was found that 133 out of 178 policyholders (74,8%) had predicted probabilities above 0.5, indicating a high likelihood of renewal. Meanwhile, 45 policyholders (25.3%) fell below the 0.5 threshold, suggesting a higher risk of non-renewal.

These results offer practical implications for the company. The predicted probability scores can be operationalized by the marketing and customer experience teams to prioritize retention efforts. Customers identified as high-risk for non-renewal (i.e., $y < 0.5$) may be targeted with personalized offers, enhanced service touchpoints, or post-sales engagement. Meanwhile, customers likely to renew could be engaged with loyalty incentives or cross-sell opportunities. By anchoring such efforts on the model outputs, PT. Asuransi XYZ can make more efficient use of its resources while improving customer lifetime value.

In conclusion, the two-step binary logistic regression analysis revealed that not all traditional pricing components, such as premium amounts, are influential in predicting renewal behavior. Instead, Discount and Admin Cost appear to play a more pivotal role. The refined model provides actionable insights that can be used by insurers to tailor retention strategies, optimize discounting policies, and improve overall customer lifetime value. Future research might include interaction terms or explore nonlinear effects, while also integrating more granular behavioral or demographic variables to enhance model accuracy.

CONCLUSION

This study applied a binary logistic regression model to explore the determinants of individual property insurance policy renewal. The results indicate that while conventional

variables such as premium and occupation do not significantly influence renewal decisions, discount and administrative cost exhibit stronger predictive relevance. Specifically, greater discounts were associated with a reduced likelihood of renewal, potentially reflecting short-term, promotion-sensitive behavior, whereas higher administrative costs were positively linked to renewal, suggesting stronger customer engagement. The refined model achieved satisfactory predictive performance, particularly in identifying policyholders likely to renew, which aligns with insurer priorities on customer retention. These findings underscore the need for more nuanced retention strategies that move beyond price incentives and consider customer investment as a proxy for loyalty. Future research should consider integrating behavioral and demographic factors to further refine predictive accuracy and strategic insight.

REFERENCES

- Alamsyah, S., & Heikal, J. (2024). Application of binary logistic regression to predict COVID-19 vaccination willingness. *Health Analytics Journal*, 7(4), 112–125. idm.or.id
- Aon. (2025). *2025 Climate and Catastrophe Insight Report*. Aon Global Risk Solutions.
- Arda, E., Aulia, A., & Heikal, J. (2024). Analysis of factors influencing the performance of Payakumbuh City public housing and residential area employees using binary logistic regression. *Jurnal Ekonomi dan Bisnis Digital*, 1(3), 379–384.
- Bisnis.com. (2024, March 15). Lack of financial education undermines customer trust in insurance. Retrieved from <https://bisnis.com>
- Dewi, P., Aulia, R. N., Taufiqillah, R., & Heikal, J. (2024). Customer churn prediction for life insurance using binary logistic regression. *Economic Reviews Journal*, 3(3), 101–112. idm.or.id
- Erwan, E., Rica, R., & Heikal, J. (2024). Analysis of factors affecting employee performance at BKPSDM Payakumbuh City using binary logistic regression. *Jurnal MSDA (Manajemen Sumber Daya Aparatur)*, 12(2), 171–178. <https://doi.org/10.33701/jmsda.v12i2.3849>
- Faisal, F., Sari, I. R., Saraswati, I., & Heikal, J. (2024). Analisis pengaruh karakteristik produk terhadap niat beli ulang pelanggan menggunakan metode regresi logistik biner. *AKADEMIK: Jurnal Mahasiswa Humanis*, 4(3), 1182–1190. <https://doi.org/10.37481/jmh.v4i3.1064>
- Hendrawan, E., Zakaria, D., Salwa, E., & Heikal, J. (2024). Customer renewal prediction for motor vehicle insurance using binary logistic regression in PT XYZ Insurance. *Innovative: Journal of Social Science Research*, 4(6), 2311–2320.
- Hidayat, T., & Heikal, J. (2024). Binary logistic regression in predicting online customer satisfaction: A case study in Indonesian e-commerce. *Journal of Digital Innovation Studies*, 6(2), 123–134. idm.or.id
- Investor.co.id. (2024, April 12). Premiums too expensive become major reason for policy non-renewals. Retrieved from <https://investor.co.id>
- McKinsey & Company. (2025). *Global Insurance Report 2025: The pursuit of growth*. Retrieved from <https://www.mckinsey.com>
- MediaAsuransi.com. (2023, October 22). Negative claim experiences drive policyholder attrition. Retrieved from <https://mediaasuransi.com>
- Mohamad, A., Setiyanto, A., Raharjo, B. W., & Heikal, J. (2024). Analisis faktor yang mempengaruhi kepuasan pembeli di Kopi Kenangan menggunakan metode regresi logistik biner. *Syntax Admiration / Jurnal Syntak Admiration*, 5(8), 123–134.
- OECD. (2023). *Leveraging technology in insurance to enhance risk assessment and policyholder*

risk reduction. OECD Publishing.

- Riyani, S., Kristianto, F., Wulandari, R., & Heikal, J. (2024). Penerapan metode regresi logistik biner dengan menggunakan Python untuk menganalisis pengguna media sosial terhadap probabilitas pembukaan rekening pada Bank X. *Scientific Journal of Reflection: Economic, Accounting, Management and Business*, 7(2), 438–449.
- Sani, P. D., Ustriyana, I. N. G., & Wijayanti, P. U., & Heikal, J. (2025). Analysis of factors affecting Indomie instant noodle repurchase decisions using the binary logistic regression method. *Journal of Social and Economics Research*, 7(1), 221–233.
<https://doi.org/10.54783/jser.v7i1.816>
- Wagan, S. M., & Sidra, S. (2025). Insurance companies' activities to maintain customer loyalty in a competitive market environment. *Malaysian Journal of Business, Economics and Management*, 4(1), 31–40.
- Wardani, D. A., & Heikal, J. (2024). Analisis keberlanjutan adopsi energi terbarukan di sektor industri menggunakan pendekatan biner. *Jurnal Energi dan Lingkungan*, 7(1), 45–56.
idm.or.id
- Zulfahmi, M. R. Y., & Heikal, J. (2024). Analisis prediksi financial distress perusahaan industri kimia dasar. *Jurnal Mirai Management*, 9(1), 488–505.