

# RISK-BASED PREMIUMS OF INSURANCE GUARANTEE SCHEMES: A MACHINE-LEARNING APPROACH

Citra Amanda<sup>1</sup> and Ananta Dian Pradipta<sup>2\*</sup>

<sup>1</sup> Finance Program, School of Accounting, Universitas Bina Nusantara, DKI Jakarta, 11480, Indonesia

<sup>2</sup> Group of Transformation Management, Indonesia Deposit Insurance Corporation, DKI Jakarta, 12190, Indonesia

---

## ABSTRACT

**Introduction/Main Objectives:** This study explores the application of machine-learning techniques to risk-based premium calculations for insurance guarantee schemes within the Indonesian insurance market. This study aims to develop a risk-based premium calculation model using machine-learning techniques in the Indonesian context. **Background Problems:** A gap exists in determining risk-based premiums for both the life and non-life insurance sectors within the Indonesian insurance market. Identifying and understanding the key variables that significantly influence risk-based capital (RBC) is important, and this research addresses this need. **Novelty:** This paper is the first to apply machine learning to calculate risk-based premiums in the context of the Indonesian insurance market. The distinction between the life and non-life insurance sectors in terms of the importance of its variables and its selection of an optimal model further enrich its unique approach. **Research Methods:** We employed gradient-boosted and decision-tree models to identify key factors impacting risk-based capital. Furthermore, we leveraged clustering techniques to categorize companies into distinct risk tiers, aiming to enable more precise risk-based premium rate calculations. **Finding/Results:** The findings reveal significant differences between the life and non-life insurance sectors in terms of key variables that impact their risk-based capital. These insights lead to the categorization of insurance companies into distinct risk tiers which helps to more accurately calculate risk-based premiums. **Conclusion:** Machine learning can serve as a powerful tool in refining insurance risk management practices, ultimately benefiting insurers, policyholders, and regulators alike.

---

## ARTICLE INFO

### *Article information:*

Received: 3 August 2023.  
Received in revised version: 8 November 2023. Received in revised version: 24 January 2024.  
Accepted: 26 January 2024

### *Keywords:*

risk-based capital, insurance guarantee schemes, machine learning, Indonesian insurance market, clustering techniques

### *JEL Code:*

G22, C19

### *ISSN:*

ISSN 2085-8272 (print)  
ISSN 2338-5847 (online)

---

\* Corresponding Author at Transformation Management Division, Indonesia Deposit Insurance Corporation (Lembaga Penjamin Simpanan), Sudirman Central Business District (SCBD) Lot 9 Jalan Jenderal Sudirman Kav. 52-53, Jakarta 12190, Indonesia

E-mail address author #1: citra.amanda@binus.ac.id (Orcid: <https://orcid.org/0000-0002-1982-6182>). author #2 ananta.pradipta@lps.go.id (Orcid: <https://orcid.org/0000-0001-9084-2073>)

## INTRODUCTION

### 1. Background and Motivation

Recognizing the importance of a robust and well-regulated insurance sector, Indonesia has introduced Law No. 4 of 2023 concerning Financial Sector Development and Reinforcement, marking a significant shift in the country's financial system. One of the provisions within the law concerns the role of the Indonesia Deposit Insurance Corporation (LPS) in guaranteeing insurance policies, expanding its responsibilities beyond bank deposit guarantees, and further emphasizing the importance of a stable insurance industry in maintaining financial stability. By implementing stricter regulations, prudent management, and more effective oversight, public confidence in the insurance sector can be increased. The new regulations and the expanding role of the LPS underscore the pressing need for more accurate risk assessments and premium calculations in order to develop fair premiums to maintain financial stability in the Indonesian insurance sector.

The insurance industry is crucial for a country's financial stability due to its essential role in risk management, investment, and social welfare. As Cummins & Weiss (2014) emphasized in their study on systemic risk and the global financial crisis, insurance companies pool and manage risks, encouraging long-term savings and investments that contribute to economic growth. Several countries, including France, Germany, Japan, and Taiwan, have established insurance policy guarantee systems to protect policyholders and maintain the stability of their insurance industries. These countries have adopted the *ex-ante* funding scheme (Ligon & Thistle, 2005).

The case of PT Asuransi Jiwasraya, an Indonesian life insurance company, serves as a

valuable lesson on the consequences of poor management in the insurance industry (Hidajat, 2021). Jiwasraya's financial condition significantly deteriorated in 2018, with a deficit in equity increasing from IDR 3 trillion to over IDR 10 trillion (Solichin, 2021). The COVID-19 pandemic further worsened the situation as the company's obligations surged to over IDR 50 trillion, leading to a negative risk-based capital ratio shown in its financial statement that fell well below the minimum requirement of 120%. To address Jiwasraya's issues, a holding company called IFG was established that implemented various restructuring measures. Interestingly, the company's insolvency was managed using the bridge bank mechanism, a resolution method typically reserved for failing banks. This case underscores the importance of sound management practices and the need for effective regulatory oversight to maintain financial stability and safeguard the interests of policyholders in the insurance sector.

The primary motivation for this study is to investigate comprehensively Indonesia's insurance guarantee scheme (IGS) which adopts the *ex-ante* funding scheme and to develop a machine-learning model that calculates the risk-based premium as an addition to the base premium for insurance companies in Indonesia. The base premium is typically assigned based on the insurer's premium income, the insurer's liabilities, and the discretion of the regulator. By calculating the risk-based premium, insurance guarantee institutions can further adjust the premium payments according to the risk profile of the insurance company. Furthermore, this study will classify the risk-based premium using a multiclassification approach, offering a more tailored and accurate assessment of the insurance guarantee premium for each insurer.

This approach builds upon previous studies, such as those by Cummins (1988), Lee et al.

(1997), Eling & Schmeiser (2010), Duan & Yu (2005), and Nektarios (2010), which have explored insurance guaranty funds and regulatory approaches. However, the calculation of insurance guarantee premiums is predominantly found in the banking industry, with limited literature addressing its application in the insurance sector (Anginer & Demirgüç-Kunt, 2018; Assa et al., 2019; Dugas et al., 2003). The multiclassification model developed in this study will contribute to the existing literature and provide a valuable tool for policymakers and industry practitioners in Indonesia and beyond.

This study utilizes data from various sources, encompassing both structured and unstructured data. The financial data are collected from multiple providers, including the Financial Services Authority of Indonesia (OJK), companies' annual reports, and private data providers such as Bloomberg, Osiris, and Infobank. The methodology employed in this research is machine learning, incorporating techniques such as decision trees and gradient-boosted models. The data used in this study span from 2018 to 2020, providing a relevant and recent overview of the industry's landscape.

The results of this study reveal key variables that significantly impact the risk-based capital (RBC) of life and non-life insurance companies in Indonesia. Using a gradient-boosted decision-tree model, the research identifies important factors such as the investment-to-total-assets ratio, the previous year's annual GDP growth, pre-tax profit-to-total-assets ratio, and internet usage per population, among others. By applying clustering techniques, insurance companies are classified into different risk categories, ranging from lower risk to upper risk. This classification enables stakeholders to better understand the specific risk profiles of insurance companies, develop targeted strategies to manage their risks effectively and calculate risk-based premium

rates. The findings offer valuable insights for insurance companies, policyholders, and regulators into the Indonesian insurance market, highlighting the importance of accurate and efficient insurance guarantee premium rate calculation models.

The paper is organized as follows: Section 1 outlines the purpose, scope, and context of the Indonesian insurance market. Section 2 presents the literature review and discusses insurance guarantee schemes and the application of machine learning to premium rate calculations. Section 3 details the data and methodology employed in the study, while Section 4 explains the model developed for the research. Section 5 presents the results obtained from the analysis, followed by Section 6 which offers a discussion of the findings. Finally, Section 7 concludes the paper with a summary of the conclusions drawn from the research.

## **2. Purpose and Scope of the Paper**

This paper aims to outline the critical considerations that need to be addressed when designing the optimal insurance guarantee premium for insurance companies in Indonesia. Recognizing the importance of striking a balance between maintaining insurance companies' financial stability and safeguarding policyholders' interests, this paper offers an academic perspective on the factors and methodologies that can contribute to determining an effective premium structure using machine learning. While the paper does not provide specific recommendations on how the Indonesia Deposit Insurance Corporation (IDIC) should establish the premium, it aims to serve as a valuable reference for understanding the underlying complexities and challenges in creating a fair and robust insurance guarantee premium framework for the Indonesian insurance market.

### 3. The Indonesian Insurance Market Context

The Indonesian insurance market has seen remarkable growth in recent years, with total gross written premiums increasing from IDR 324.2 trillion (approx. USD 22.8 billion) in 2017 to IDR 400.9 trillion (approx. USD 28.2 billion) in 2020 (General Insurance Association of Indonesia & Indonesian Life Insurance Association, 2021). Despite this growth, insurance penetration in Indonesia remains relatively low at around 3% of GDP in 2020 (World Bank, 2021), indicating significant potential for further expansion. Figure 1 depicts the Indonesian insurance market's contraction of 7.6% in 2020, reaching a total value of USD 20.5 billion. Over the period from 2016 to 2020, the market demonstrated a compound annual growth rate (CAGR) of 4.3%.

However, a factor in the low insurance penetration rate is the limited awareness of the importance of insurance among the Indonesian population. A study by Yamin et al. (2018) found that financial literacy and education play crucial roles in shaping the public's perception and understanding of insurance products. Efforts to improve financial literacy and promote the benefits of insurance are essential to increase insurance penetration in Indonesia.

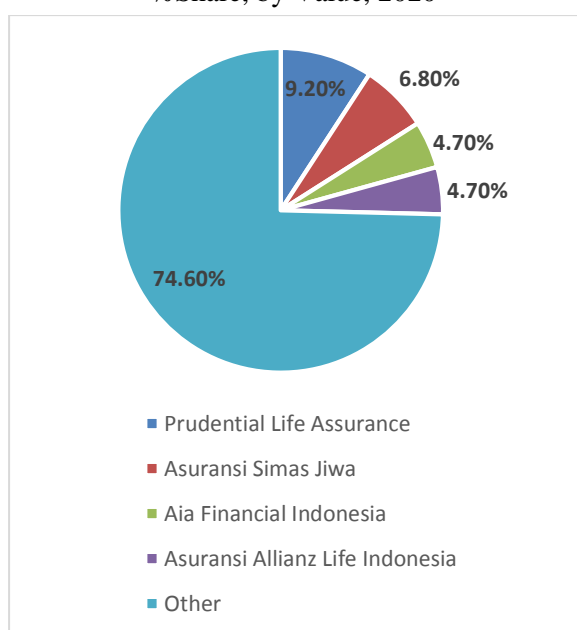
In 2020, the Indonesian insurance market experienced a downturn following three consecutive years of growth, primarily attributed to the repercussions of the COVID-19 pandemic. This scenario has intensified competition among prominent players in the market, necessitating the encroachment of competitors' market shares to achieve growth in a contracting environment. Four leading players dominate the Indonesian insurance market, with life insurance being the most lucrative segment, contributing 72.4% of the market's total value in 2020. Consequently, it is not surprising that companies specializing in life insurance, such as Prudential Life Assurance, Asuransi Simas Jiwa, Asuransi Allianz Life Indonesia, and AIA Financial Indonesia, are the frontrunners in the overall insurance market. Despite the dominance of these players, the market exhibits considerable fragmentation, with other firms holding a 74.6% share of the market, further fueling rivalry. Major players have established a robust market presence through diversified product portfolios and numerous industry accolades. Additionally, the insurance market is characterized by debt offerings, acquisitions, and financial ventures, which are essential elements of companies operating within this sector and crucial approaches for raising capital. Figure 2 presents



the distribution of market share in the Indonesian insurance sector, measured by value, for the year 2020. The visualization highlights the relative proportions of different players within the market, showcasing their contributions to the overall industry value.

Another challenge the Indonesian insurance industry faces is the need for effective risk management, particularly in the context of natural catastrophes. As a country prone to natural disasters, Indonesia requires a strong insurance market to mitigate the economic impact of such events. A study by Sugiarto et al. (2019) emphasizes the importance of improving the insurance sector's risk management capabilities and developing appropriate risk transfer mechanisms.

**Fig 2.** Indonesia Insurance Market Share: %Share, by Value, 2020



Source: Marketline: Authors' own calculation

## LITERATURE REVIEW

### 1. Insurance Guarantee Schemes (IGSs)

Insurance guarantee schemes (IGSs) play a critical role in maintaining confidence in the insurance industry and ensuring the protection of

policyholders when an insurer becomes insolvent. One of the key aspects of IGSs is the funding mechanism. Thomann & Hill (2014) delineated two broad categories: ex-ante and ex-post funding. While the categorization is clear-cut, the choice between them is not. The suitability of either mechanism isn't just a matter of academic debate; it has real-world implications for industry stability. Countries have adopted different funding mechanisms based on their unique market structures and regulatory environments (Ligon & Thistle, 2005).

A number of studies have explored the design and implementation of IGSs in various countries. For example, Cummins & Sommer (1996) analyzed the historical development and current status of IGSs in the United States and found that most states have established life and health insurance guarantee associations to protect policyholders. Nektarios (2010) offered a panoramic view of Europe, advocating for an integrated insurance regulation strategy, weaving in IGSs. Other studies have focused on the optimal design of IGSs, including calculating premiums and managing guarantee funds. Venturing into the more technical realm, studies like those of Eling & Schmeiser (2010) and Duan & Yu (2005) grappled with the nuances of IGS design. While fair premium pricing is a noble goal, the challenge lies in balancing actuarial soundness and market realities.

Besides its benefits, the shadow of 'moral hazard' looms large over IGSs. Grace et al. (2002) showed that the existence of IGSs can lead to moral hazard problems, as insurers may take on excessive risks, due to the protection offered to policyholders in the event of insolvency. One of the foundational studies exploring the moral hazard implications of IGSs was by Cummins and Danzon (1997), who argued that the establishment of guarantee funds reduces the incentives for policyholders to

monitor the financial health of their insurers, leading to an increase in risk-taking behavior by insurance companies. The authors found evidence of this moral hazard effect in the U.S. property-liability insurance market, showing that insurers in states with guarantee funds exhibited riskier investment strategies and higher insolvency rates. Lee et al. (1997) added another layer, pointing out the skewed risk appetites of life insurers with easy access to guarantee funds. This finding is consistent with the notion that the presence of a guarantee fund encourages insurers to engage in riskier activities, as the downside risk is partially absorbed by the guarantee fund. In sum, while IGSs undoubtedly infuse confidence and provide a buffer against industry shocks, they're not without their challenges.

## 2. Premium Rates and Applications of Machine Learning

Some studies have proposed various policy measures to mitigate the moral hazard concerns associated with IGSs. For instance, Epermanis & Harrington (2006) suggested that stronger regulatory oversight and risk-based capital requirements could help reduce the moral hazard effects associated with guarantee funds. Similarly, Han et al. (2016) recommended the implementation of risk-based premium pricing for IGSs to better align the premiums charged to insurers with their risk profiles, thus discouraging excessive risk-taking. Cummins (1988) extended the work on deposit insurance premiums to price single-period premiums for insurance guarantee funds. He proposed a model that incorporates the riskiness of insurers' investment portfolios, the likelihood of insolvency, and the costs of insurer failure. This model allows for the calculation of optimal premiums that reflect the true risk exposure of the guarantee fund to each insurer, thus discouraging excessive risk-taking.

Duan & Yu (2005) developed a multiperiod model to measure the cost of the guarantee fund, incorporating risk-based capital regulations. Their model takes into account the dynamic nature of insurers' risk profiles, the regulatory environment, and the potential losses that guarantee funds may face. By considering these factors, their model aims to determine optimal premium rates that balance the financial stability of the guarantee fund with the need to minimize moral hazard. Eling & Schmeiser (2010) studied the optimal design of insurance guarantee schemes and the pricing of guarantee fund premiums. They emphasized the importance of a fair premium in a competitive market and identified the key factors that should be considered when determining the optimal premium rate, such as the riskiness of the insurer's portfolio, the correlation between insurer risk and market risk, and the costs of insurer failure. Some jurisdictions have insurance guarantee institutions that use a mixed premium assessment based on premium income and liabilities. For example, South Korea uses the arithmetic mean of premium income and liabilities. In contrast, Malaysia uses liabilities as the basis for life insurance premiums and premium income as the basis for general insurance premiums.

Machine learning has emerged as a powerful tool in the insurance industry. The application of machine-learning techniques to the determination of premium rates for insurance guarantee schemes (IGSs) could enhance the accuracy and efficiency of the pricing process. One potential application of machine learning in the context of IGSs is the development of predictive models to estimate insurers' risk profiles more accurately. By leveraging large datasets and advanced algorithms, machine-learning models can identify complex relationships and patterns in the data that may not be apparent using traditional methods (Bose & Chen, 2009).

Another potential application of machine learning in IGS premium pricing is the optimization of risk-based premium rates. Machine-learning algorithms, such as decision trees, support vector machines, and neural networks, can be employed to model the relationship between premium rates and insurers' risk profiles (Gao & Lin, 2015). These models can help determine the optimal premium rates that balance the financial stability of the guarantee fund with the need to minimize moral hazard. In addition to improving the accuracy of premium pricing, machine-learning techniques can also be used to enhance other aspects of IGSs, such as monitoring and supervision of insurance companies. For instance, machine-learning models can be employed to detect early warning signals of potential insolvency or financial distress among insurers, enabling regulatory authorities to take timely action to mitigate risks (Khashman, 2017).

## **METHOD, DATA, AND ANALYSIS**

### **1. Data Sources and Preprocessing**

We collect structured and unstructured data on insurance companies' financial performance, claims history, and other relevant variables to develop the model. The sample consists of 42 life insurance and 59 non-life insurance companies in Indonesia, not only public companies but also private companies. Structured data are obtained from publicly available financial statements from OSIRIS, Bloomberg, OJK, and companies' financial statements. These data include financial ratios, income statements, balance sheets, and cash flows for each insurance company. Some key financial variables considered are total assets, total liabilities, equity, net income, loss ratio, expense ratio, combined ratio, solvency ratio,

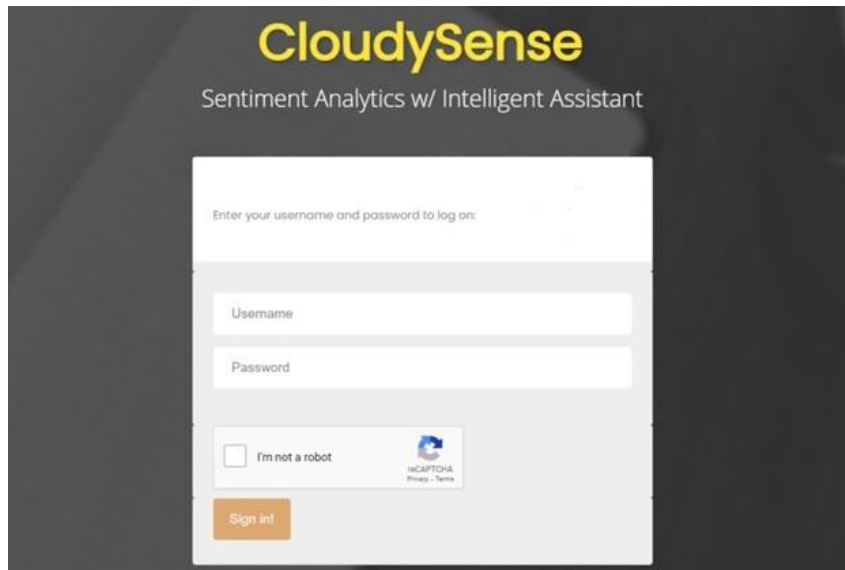
risk-based capital ratio, and other variables gathered from financial statements. In addition, we also use macroeconomic variables such as GDP growth, inflation, employment rate, etc. Our data collection includes the temporal dimension, such as using lag variables. Using lagged variables is pivotal for our analysis, as it allows us to measure the impact of past events on current outcomes.

### **2. Sentiment Variable**

We developed a machine-learning model specifically tailored for sentiment analysis in the insurance industry. We gathered data from Twitter (in Indonesian) for each insurance company and utilized alternative data sources, such as news regarding insurance companies' insolvency, resolution, and fraud. Our model calculates sentiment with a monthly lag to account for irrational aspects of the market. Additionally, we leveraged alternative data from social media to analyze conversations related to individual insurance companies. We designed web-based tools for the Insurance Financial Guarantee (IFG) to extract and evaluate sentiment from social media discussions. Figure 3 shows CloudySense Sentiment Analytic Tools.

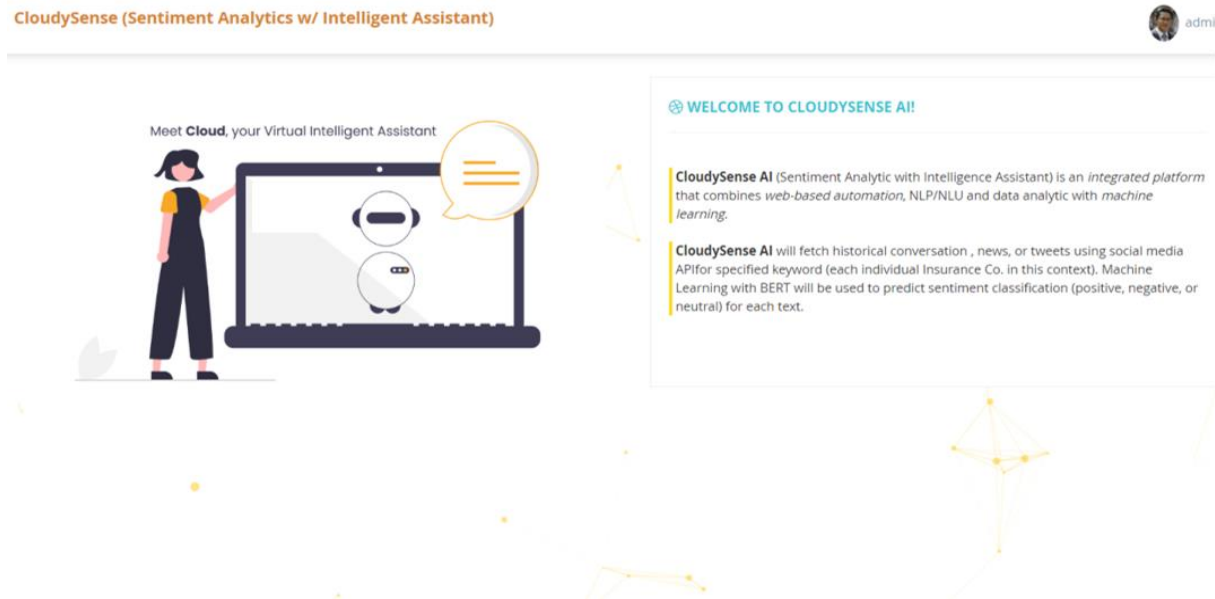
We employed BERT (Bidirectional Encoder Representations from Transformers, available at <https://arxiv.org/abs/1810.04805>) to create a machine-learning model capable of predicting sentiment classifications for each conversation, news article, or tweet related to individual insurance companies. These classifications are categorized as positive, negative, or neutral. To improve the model's understanding of finance-specific terminology and unique words, we pre-trained it with a financial corpus. Figure 4 showcases our approach and the tools used for sentiment analysis in the insurance sector.

**Fig 3.** CloudySense: Sentiment Analytic Tools



Source: Authors' own development

**Fig 4.** Cloudysense AI: Sentiment



Source: Authors' own development

We fetched historical data (news articles or tweets) from 2018 to 2020 for each individual insurance company in our sample. Each piece of data was processed through sentiment analysis, resulting in a sentiment label (positive, negative, or neutral) and a confidence level. The sentiment score was subsequently used as a feature or independent variable in our research. Figure 5

displays the results of the sentiment analysis, where each individual insurance company's historical data (news articles or tweets) were processed, generating sentiment labels (positive, negative, or neutral) and confidence levels, which were later used as features or independent variables in our research.



Fig 5. Sentiment result

CloudySense (Sentiment Analytics w/ Intelligent Assistant)

Source	Year	Sentiment Classification	Confi... Score	OriginalText
Twitter	2018	positive	0.99	"Hadir sekaligus menerima penghargaan Direktur ██████ Insurance, Mr. Hidenobu Ito ██████ selalu hadir memberikan perencanaan masa depan dan perlindungan yang terpercaya dengan segala kemudahannya. #E ██████ #EazyClaim #EazyAccess #EazyConnect #Penghargaan ██████ https://t.co/UJ0dQayPSr"
Twitter	2018	positive	0.99	"Pada hari Kamis, tanggal 27 September 2018 ██████ Insurance memperoleh Penghargaan Best Financial Performance, Life Insurance Company, Category Asset Rp 10 Triliun – Rp 25 Triliun pada acara "Indonesia Insurance Consumer Choice Award 2018" versi Warta Ekonomi. https://t.co/WS0pT1Hsv9"
Twitter	2018	positive	0.99	"Hari Rabu, 26 September 2018 ██████ Insurance melalui unit syariah memperoleh predikat "Sangat Bagus" untuk kategori Perusahaan Asuransi Jiwa Syariah Kontribusi Bruto 100 Miliar ke Atas Sepanjang Tahun 2017" pada acara Rating Infobank Sharia Finance Institution Awards 2018 https://t.co/pb4elwiACQ"
Twitter	2018	negative	0.99	"Mohon perhatian kpd ██████ agar tidak membiarkan data nasabah dapat dengan mudah diperoleh pihak lain. Juga kpd ██████ insurance agar memperhatikan sistem kerja kerjanya dalam mencari nasabah. Cara yg sangat mengganggu kenyamanan."
Twitter	2018	negative	0.99	"sangat terganggu dg telepon dari ██████ insurance yang menepon Magrib & memaksa nasabah untuk menyetujui tawarannya. Kecewa juga karena data-data nasabah ██████ taplus aktif bisa diakses bebas oleh pihak lain, padahal nomor ponsel yg saya gunakan saat ini baru."
Twitter	2019	positive	0.99	"Make great again with the dream team. Thanks for moments. @ ██████ insurance. PT https://t.co/b7Nx7BtznA"
Twitter	2020	positive	0.99	"PT ██████ Insurance mendapatkan apresiasi Top 5 Financial Performance Best Insurance Award 2020 with Assets 10-20T dan Best Good Corporate Governance Report 2020 pada ajang Indonesia Best Insurance Award 2020. https://t.co/sjQy4PE5Z"

CloudySense (Sentiment Analytics w/ Intelligent Assistant)

CloudySense > Sentiment: IFG

**SENTIMENT ANALYTIC IN NEWS AND SOCIAL MEDIA**

Year:

Keyword:

IDBank:  \* optional (use only when assign new keyword to existing bank)

SENTIMENT RESULT

Drag a column header here to group by that column

ID...	Keyword	Source	TotalCount
17	██████ INDONESIA	Twitter	383
21	██████ L INDONESIA	Twitter	378
6	██████	Twitter	304
15	██████ LIFE	Twitter	276
8	██████ SSURANCE	Twitter	257
36	██████ IA	Twitter	208

Source: Authors' own development

Unstructured data are collected through Python and BERT (Bidirectional Encoder Representations from Transformers) for sentiment analysis on customer feedback and social media posts. These data help to better understand customer preferences and adjust

premiums accordingly. Some examples of unstructured data variables are customer satisfaction score, net promoter score, sentiment polarity (positive, negative, neutral), and frequency of specific keywords (e.g., "claims", "service", "price"). Data preprocessing involves

cleaning and transforming the raw data into a format suitable for analysis. This includes handling missing values, outlier detection, and feature scaling.

### 3. Machine-Learning Model Selection and Justification

We inspect several machine-learning approaches, including decision trees, random forests, and gradient-boosted trees, to identify the most appropriate model for our analysis. Neural networks are known for their ability to handle complex relationships and non-linearities in the data (Hornik et al., 1989). Random forests offer robustness against overfitting and can handle large numbers of predictor variables (Friedman, 2001). Gradient-boosted trees are renowned for their high performance in various applications, as they combine weak learners to create a strong learner (Friedman, 2001). We also run an OLS model to provide a comparison for our machine-learning models. The choice of the best model is based on factors such as model performance, interpretability, and computational efficiency.

### 4. Model training, validation, and evaluation

Our machine-learning modeling methodology follows data mining processes, which include: (a) Sample splitting and tuning where the dataset is divided into training, validation, and testing datasets using cross-validation to ensure a robust evaluation of the model's performance (Kohavi, 1995); (b) modeling where we apply various machine-learning algorithms, such as generalized linear models (McCullagh & Nelder, 1989), random forests (Breiman, 2001), boosted-regression trees (Friedman, 2001), and neural networks (Hornik et al., 1989), to develop the premium rate calculation model; (c) performance evaluation where model performance is assessed using metrics such as mean absolute error

(MAE), mean squared error (MSE), and R-squared (Hyndman & Koehler, 2006) to determine the best-fitting model; (d) determining variable importance which means the most influential variables for the selected model are identified to understand better the factors that impact the insurance guarantee's premium. The detailed processes are described in Figure 6.

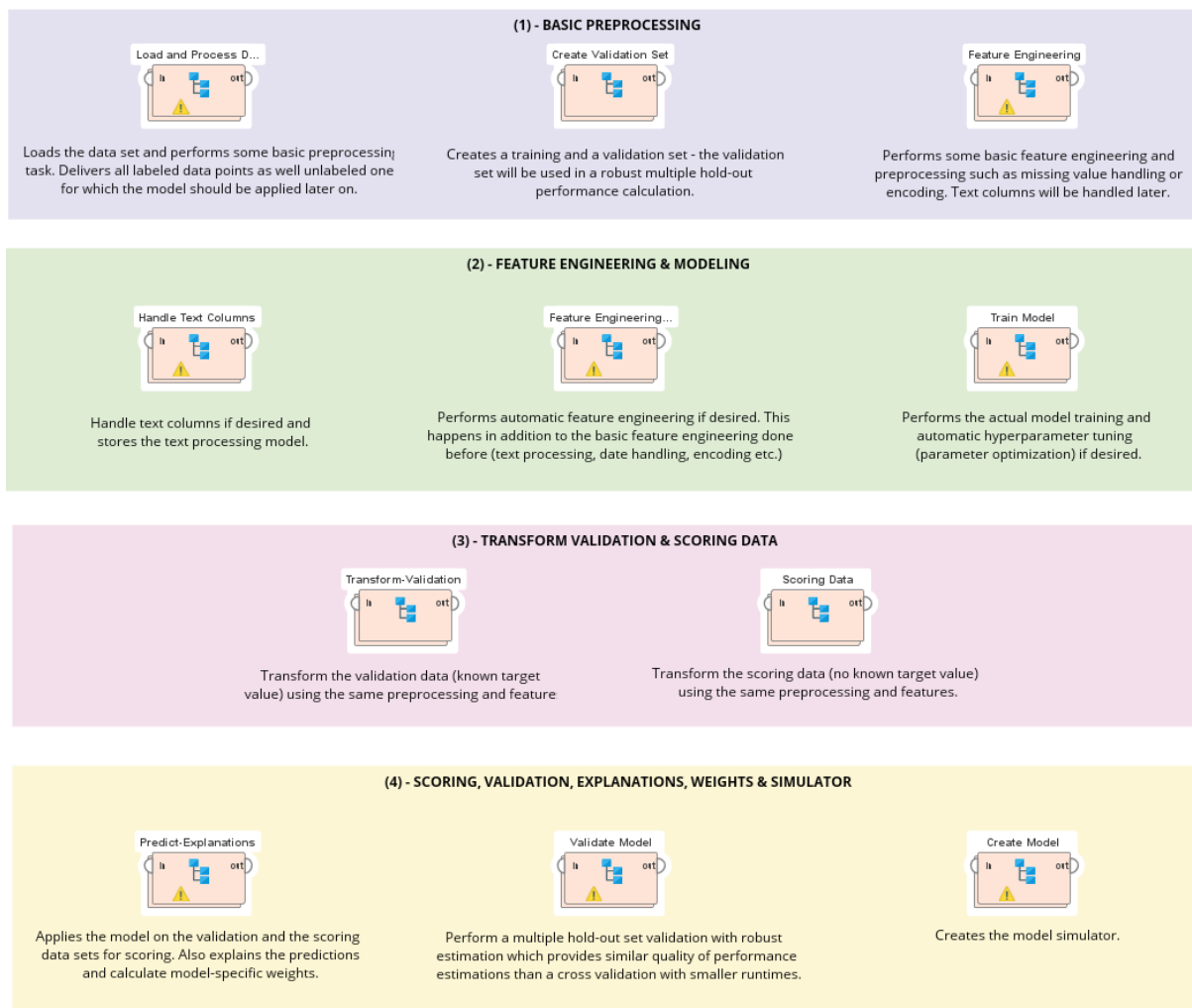
## 5. Model

### 5.1. Levy (insurance premium)

A policy guarantee institution (LPP) requires funding from contributions (premiums) paid by participating insurance companies. This funding is necessary to ensure that the LPP can pay compensation to policyholders and support the continuity of LPP operations. According to the regulations in Indonesia, LPP funding is done on an ex-ante basis, meaning that funds are collected before an insurance company failure occurs. The methods used include premium income, minimum capital requirements, liabilities, and fixed rate. The chosen contribution assessment basis must be relevant to the risk insured by the LPP. Alternatives that can be chosen as the basis for contributions include technical reserves (for life insurance companies) and gross premium income (for general insurance companies).

To determine the contribution percentage for insurance companies, the percentage is utilized to calculate the premium amount payable to the LPP. Factors such as funding targets, policy guarantee program implementation needs, and the insurance industry's capacity are taken into account. We propose a five-tiered premium structure for the LPP, aiming to establish a more comprehensive and equitable approach to determining premiums by considering the varying risk profiles and financial performance among insurance companies.

Fig 6. Machine Learning Process



Source: generated from <https://rapidminer.com/> (2023)

The base premium serves as the foundation for calculating LPP premiums and applies to insurance companies with average risk profiles and financial performance. This balanced approach ensures adequate funding for the LPP while maintaining the continuity of its operations. Recognizing the distinct characteristics and risk factors associated with life insurers and general insurers, the five-tiered structure includes Lower Risk Premium, Moderate-Lower Risk Premium, Moderate-Upper Risk Premium, and Upper Risk Premium for each type of insurance company. Following Cornett et. al (1998), the Federal Deposit Insurance Corporation (FDIC) shifts a flat rate to risk-based

premiums. It can be useful for understanding the potential effects of implementing tiered premium structures in insurance markets, as both contexts involve the use of premiums to manage risk and promote financial stability.

The machine-learning models implemented in this study are specifically tailored for the life insurance and non-life insurance sectors in accordance with the distinct characteristics of each segment as identified in the literature (Cummins & Weiss, 2014; Eling & Pankoke, 2014). By taking these differences into account, the models provide more accurate predictions for risk-based premiums in both life and non-life insurance sectors.

## 5.2. Insurance Company Risk and the Determinants

In this study, to assess insurance company risk, we use risk-based capital (RBC) ratios as a useful proxy variable. RBC ratios, as established by insurance regulators, evaluate the adequacy of an insurer's capital in relation to its risk profile (Koijen & Yogo, 2015). A lower RBC ratio signifies that an insurer is potentially more vulnerable to financial distress, while a higher RBC ratio suggests a more stable financial position and a greater ability to absorb potential losses (Baranoff, 2004). The use of RBC ratios as a proxy for risk has been well-documented in the literature. For instance, Baranoff & Sager (2003) examined the relationship between RBC ratios and the financial health of life insurance companies, finding a strong association between higher RBC levels and lower financial vulnerability. Eling & Schmit (2012) explored the connection between RBC ratios and the solvency of non-life insurance companies, concluding that RBC ratios serve as an effective indicator of insurers' risk exposure.

This study examines the factors affecting the RBC of insurance companies. The factors considered in the model are derived from various aspects, including financial ratios, macro-economic indicators, demographic variables, and market-related data. Some of the key factors include investment to total assets, GDP growth, pre-tax profit to total assets, internet usage, mortality rates, consumer price index, capital adequacy, technical reserves to total assets, and liquidity. Other factors taken into account are related to the insurance market dynamics, such as premium income, market share, and competition. Additionally, the model considers socioeconomic factors such as poverty levels, income inequality, population growth, and urbanization.

## RESULT AND DISCUSSION

The results of the study are presented separately for life insurance and non-life insurance companies to account for the differences in their risk profiles.

### 1. Life Insurance

#### 1.1. The Chosen Model and Variable Importance

For life insurance companies, the most effective model identified in this study is the decision tree, which yields a relative error of 2.7%. This outperforms other machine-learning methods such as the generalized linear model (3.8% relative error), random forest (4.4% relative error), and gradient-boosted trees (3.1% relative error). The decision tree's superior performance can be attributed to its simplicity, interpretability, and ability to handle non-linear relationships between variables (Breiman et al., 1984; Quinlan, 1993). Decision trees are particularly well-suited for the analysis of complex datasets with multiple variables, as they allow for a more intuitive understanding of the relationships between the predictor variables and the response variable (in this case, RBC). Moreover, decision trees can accommodate interactions between variables, which is crucial for capturing the complex nature of the factors influencing the RBC of life insurance companies (Hastie et al., 2009).

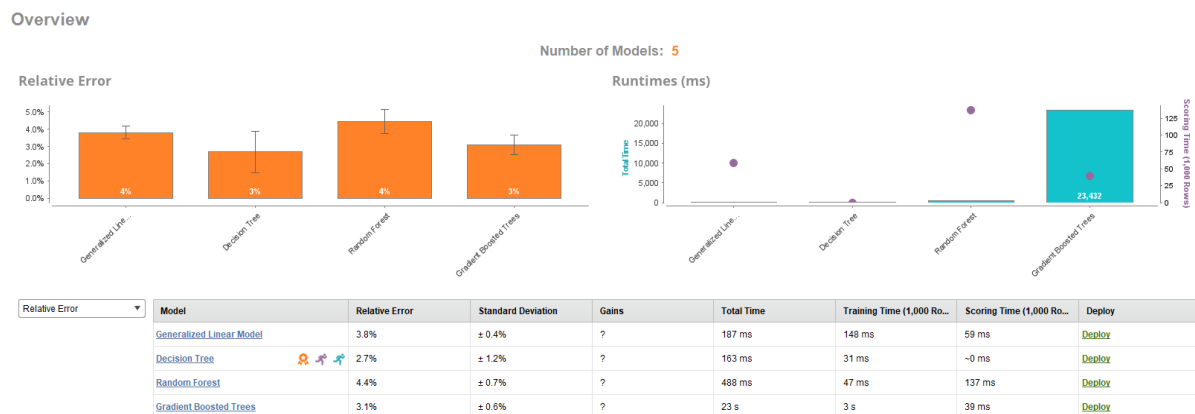
In comparison, generalized linear models are less flexible as they rely on specific assumptions about the distribution of the response variable and the link function relating the predictors to the response (Nelder & Wedderburn, 1972). Although random forests and gradient-boosted trees are powerful ensemble methods that can achieve high prediction accuracy (Breiman, 2001; Friedman, 2001), they tend to be more complex and less interpretable than decision trees, making it harder for practitioners to

understand and apply the results in this study. In light of these findings, the decision tree emerges as the most suitable method for modeling the RBC of life insurance companies. Figure 7 presents the best-performing model for predicting the risk-based capital (RBC) of life insurance companies in Indonesia.

The importance of the variables in this study is significant as it helps us to better understand the factors that influence the risk-based capital (RBC) of life insurance companies in Indonesia. These variables capture different aspects of the

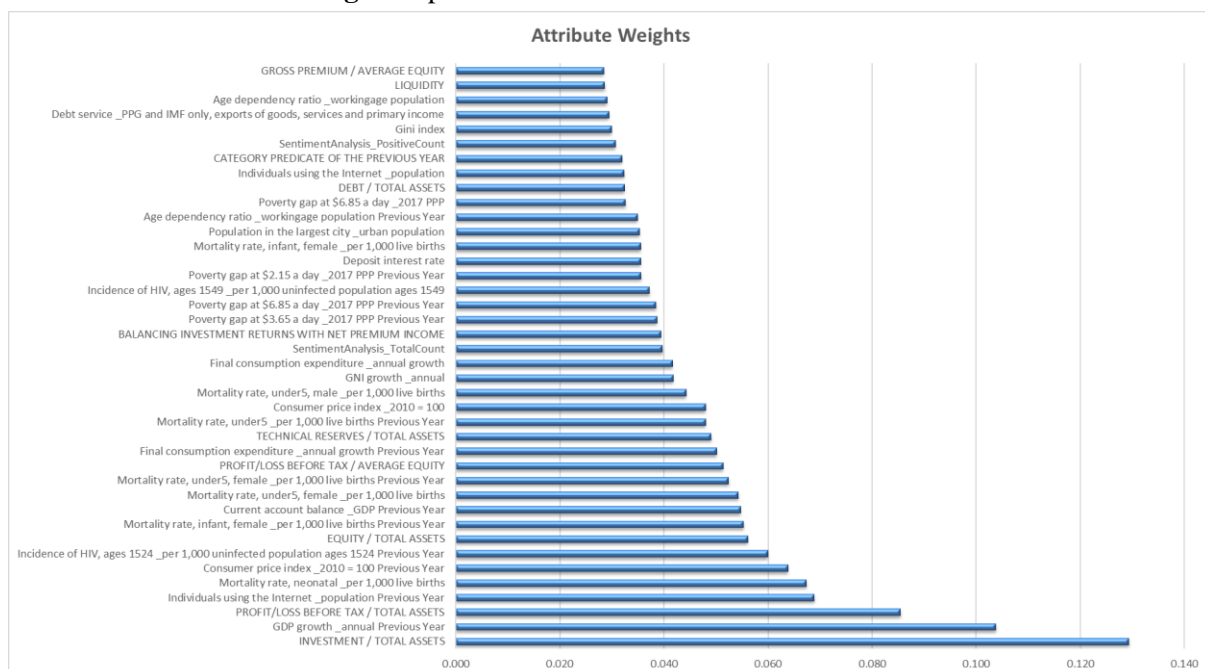
business environment, economic conditions, and socio-demographic factors, which collectively impact the insurance industry. Figure 8 displays the variable importance for the decision-tree model applied to life insurance companies, highlighting the top 10 most important variables with their respective weights. Then, Figure 9 displays the decision-tree model employed for the Indonesian life insurance sector, delineating the hierarchical arrangement of decision nodes and partitions according to the top 10 most influential variables identified in this research.

Fig 7. The Best Model for Life Insurance



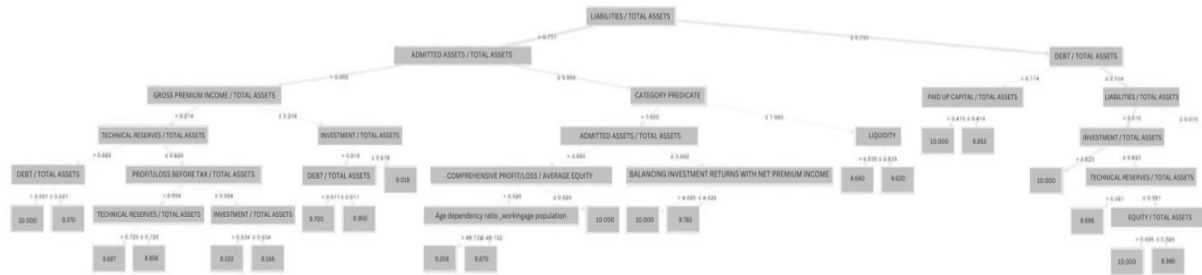
Source: Authors' own calculation

Fig 8. Importance of Variables for Life Insurance



Source: Authors' own calculation

Fig 9. Decision Tree



Source: Authors' own calculation

These variables significantly impact the risk-based capital (RBC) of life insurance companies:

- (1) investment to total assets (0.1293488) measures the proportion of a company's assets invested in various financial instruments, indicating the company's investment strategy and risk exposure;
- (2) previous year's annual GDP growth (0.1037502) reflects the economic growth rate in the previous year, which can influence insurance companies' profitability and their customers' financial stability;
- (3) ration of pre-tax profit to total assets (0.0854684) measures a company's profitability relative to its total assets, providing insights into the efficiency of its asset utilization;
- (4) previous year's internet usage per population (0.0688593) represents the percentage of the population using the internet in the previous year, which may impact the adoption of digital insurance services and customer behavior;
- (5) neonatal mortality rate (0.0673472) captures the number of deaths within the first month of life per 1,000 live births, indicating the level of healthcare quality and potential demand for life insurance products;
- (6) previous year's consumer price index (0.0638234) reflects the inflation rate in the previous year, which may influence the cost of living, disposable income, and insurance product pricing;
- (7) previous year's HIV incidence for ages 15-24 (0.0599372) measures the number of new HIV infections among the population aged 15-24 in the previous year, which could affect the demand for life insurance products and the

risk profile of policyholders;

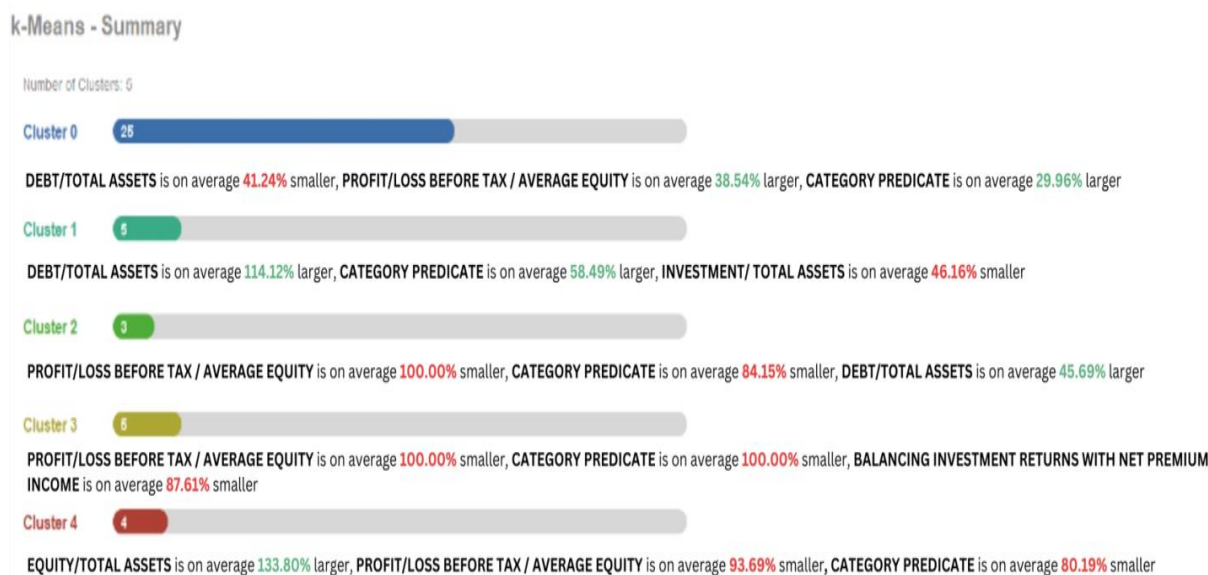
- (8) equity to total assets (0.0560883) indicates the proportion of a company's total assets financed by equity, which can affect its financial stability and risk exposure;
- (9) previous year's female infant mortality rate (0.0552655) measures the number of female infant deaths per 1,000 live births in the previous year, reflecting healthcare quality and potential demand for life insurance products;
- (10) previous year's current account balance to GDP (0.0547058) represents the ratio of a country's current account balance to its GDP in the previous year, indicating the overall economic health and its potential impact on the insurance industry.

The sentiment variable, our variable of interest, is not among the top 10 most important variables. However, it still holds some significance, ranking 21st with a weight of 0.0396780. This indicates that sentiment analysis, although not a primary factor, still plays a role in impacting the risk-based capital of life insurance companies.

### 1.2. Clustering results

The five-tiered structure categorizes insurance companies into Lower Risk Premium (Cluster 0), Moderate-Lower Risk Premium (Cluster 1), Moderate Risk Premium (Cluster 2), Moderate-Upper Risk Premium (Cluster 3), and Upper Risk Premium (Cluster 4), with Cluster 0 representing the lowest risk and Cluster 4 representing the highest risk. Figure 10 shows the clustering.

Fig 10. Clustering



Source: Authors' own calculation

The list of insurance companies and their respective risk clusters presented in a Table 2.

Table 2. Number of Insurance Companies Based on Clusters

Risk Cluster	Insurance Companies
Cluster 0	25
Cluster 1	5
Cluster 2	3
Cluster 3	5
Cluster 4	4

Source: Authors' own calculation

## 2. Non-life Insurance

### 2.1 The Chosen Model and Variable Importance

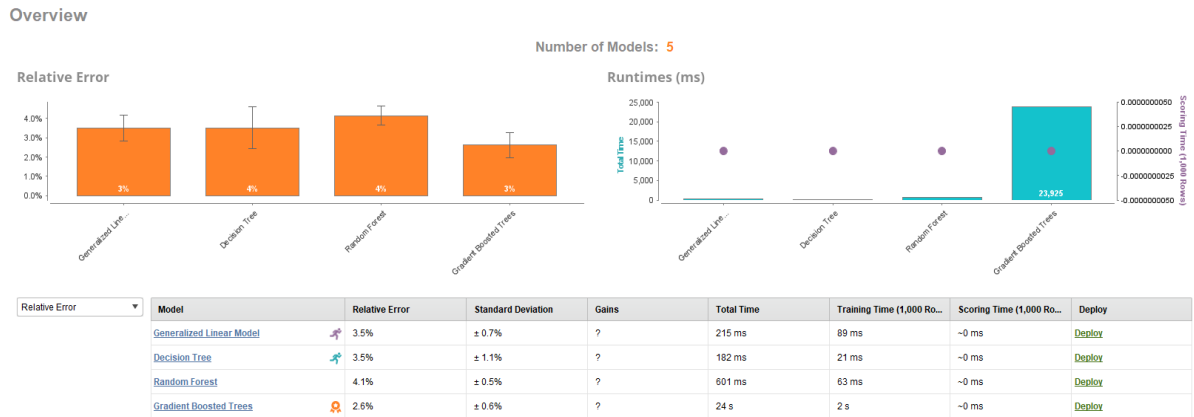
For non-life insurance companies, the gradient-boosted model emerged as the most effective method, yielding a relative error of 2.6%. This outperforms other machine-learning methods such as random forest (4.1% relative error), and decision tree (3.5% relative error). The superiority of gradient-boosted models in this context can be attributed to their ability to create an ensemble of weak learners, which are then combined to create a stronger predictive model

(Friedman, 2001). Gradient boosting is known to be highly effective in reducing bias and variance in regression problems (Natekin & Knoll, 2013), making it a suitable choice for non-life insurance risk prediction. Figure 11 illustrates the best-performing model for predicting the risk-based capital (RBC) of non-life insurance companies in Indonesia.

The importance of variables in this model for non-life insurance also captures different aspects of the business environment, economic conditions, and socio-demographic factors. Figure 12 displays the variables' importance for the decision-tree model applied to non-life insurance companies, highlighting the top 10 most important variables with their respective weights. In addition, Figure 13 displays one of the trees from the gradient-boosted model, showcasing the importance of variables and their interactions in determining the risk clusters for non-life insurance companies. This model allows us to identify the key factors affecting the risk assessment of non-life insurance companies, which can be used to improve risk management strategies and decision-making processes within the industry.



**Fig 11. The Best Model for Non-Life Insurance**

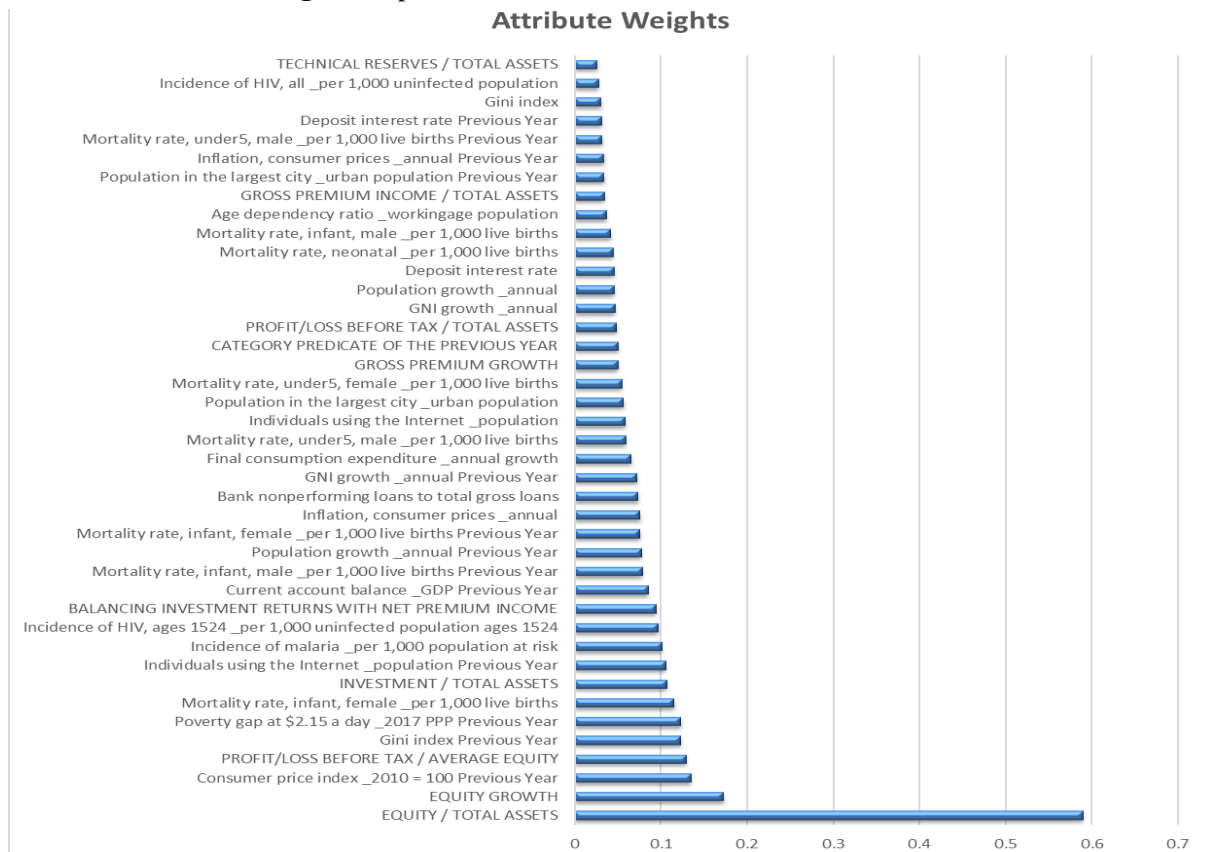


**Gradient-Boosted Model**

Model Metrics Type: Regression  
 Description: N/A  
 model id: rm-h2o-model-production\_model\_po-1211  
 frame id: rm-h2o-frame-production\_model\_po-1211  
 MSE: 0.01166501  
 RMSE: 0.108004674  
 R^2: 0.96778506  
 mean residual deviance: 0.01166501  
 mean absolute error: 0.07899555  
 root mean squared log error: 0.010564373

Source: Authors' own calculation

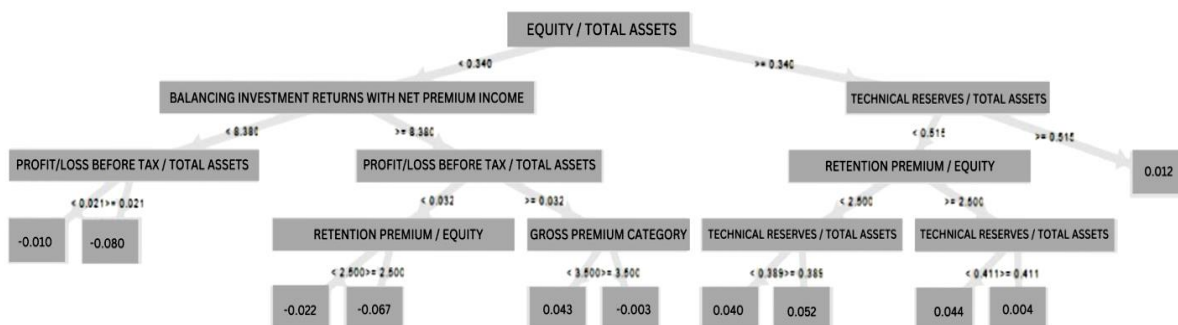
**Fig 12. Importance of Variables for Non-Life Insurance**



Source: Authors' own calculation



Fig 13. Decision Tree



Source: Authors' own calculation

The following variables significantly impact the risk-based capital (RBC) of non-life insurance companies in Indonesia: (1) Own capital to total assets (0.589882094) measures the proportion of a company's assets financed by its own capital, indicating financial stability and risk exposure; (2) growth rate of own capital (0.172249664) reflects the growth rate of a company's own capital, influencing profitability and financial stability; (3) previous year's consumer price index (0.134572027) represents the inflation rate in the previous year, affecting the cost of living, disposable income, and insurance product pricing; (4) pre-tax profit to average own capital (0.129319795) gauges a company's profitability relative to its average own capital, providing insights into financial performance; (5) previous year's Gini index (0.12314214) reflects income inequality in the previous year, impacting insurance companies' customer base and risk profile; (6) previous year's poverty gap at USD 2.15 a day (0.122970814) measures the extent of poverty in the previous year, influencing the demand for insurance products and the risk profile of policyholders; (7) female infant mortality rate (0.114698587) captures the number of female infant deaths per 1,000 live births, indicating the level of healthcare quality and potential demand for insurance products; (8) investment to total

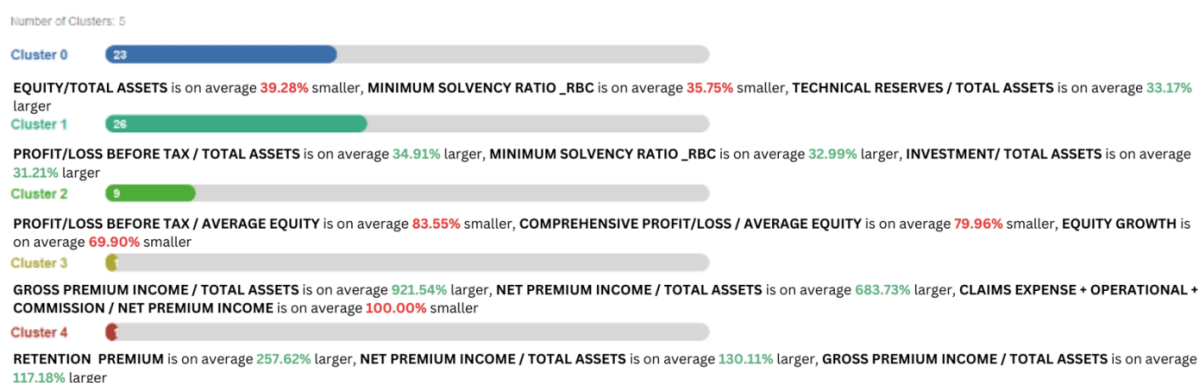
assets (0.106598024) measures the proportion of a company's assets invested in various financial instruments, indicating the company's investment strategy and risk exposure; (9) previous year's internet usage per population (0.105121765) represents the percentage of the population using the internet in the previous year, potentially impacting the adoption of digital insurance services and customer behavior; (10) incidence of malaria per 1,000 population at risk (0.100572084) measures the number of new malaria cases per 1,000 population at risk, affecting the demand for insurance products and the risk profile of policyholders.

## 2.2. Clustering Results

The five-tiered structure classifies non-life insurance companies into distinct risk categories, ranging from Lower Risk Premium (Cluster 0) to Upper Risk Premium (Cluster 4). Cluster 0 represents the lowest risk category, while Cluster 4 signifies the highest risk category. These categories include Moderate-Lower Risk Premium (Cluster 1), Moderate Risk Premium (Cluster 2), and Moderate-Upper Risk Premium (Cluster 3). Figure 14 illustrates the clustering results for non-life insurance companies, providing insights into the risk profiles of different companies within the industry.

Fig 14. Clustering

## k-Means - Summary



Source: Authors' own calculation

The list of insurance companies and their respective risk clusters presented in a Table 3.

**Table 3.** Number of Insurance Companies Based on Clusters

Risk Cluster	Insurance Companies
Cluster 0	22
Cluster 1	26
Cluster 2	9
Cluster 3	1
Cluster 4	1

Source: Authors' own calculation

### 3. Discussion

The analysis of factors that influence risk-based capital (RBC) can help stakeholders make informed decisions and develop strategies to manage the risks in the insurance industry effectively. For life insurance companies, the top variables impacting RBC include investment to total assets, annual GDP growth, pre-tax profit to total assets, internet usage, neonatal mortality rate, consumer price index, HIV incidence, equity to total assets, female infant mortality rate, and current account balance to GDP. These factors encompass a wide range of aspects, from economic conditions to demographic factors, indicating that the RBC of life insurance companies is influenced by both macroeconomic

trends and industry-specific variables. For non-life insurance companies, the most critical variables include capital to total assets capital growth, consumer price index, pre-tax profit to average equity, Gini index, poverty gap, female infant mortality rate, investment to total assets, internet usage, and incidence of malaria. Similar to life insurance companies, these variables represent various aspects of the business environment, economic conditions, and socio-demographic factors, highlighting the complex interplay of factors that shape the RBC of non-life insurance companies.

By applying clustering techniques, we can further classify insurance companies into different risk categories, ranging from lower risk to upper risk. Additionally, introducing risk-based premiums or levies in the insurance industry is essential to account for the varying risk profiles of insurance companies adequately. The clustering results can serve as a basis for determining the base premium, along with the risk-based component, to ensure that the premium structure fairly reflects the risks associated with each insurance company. The risk-based approach will determine the levy rate, acting as a multiplier factor for the total premium, ultimately enabling tailored pricing strategies for both life and non-life insurance

sectors within the Indonesian market. This approach not only encourages better risk management practices among insurance companies but also promotes a more stable and sustainable insurance market in the long run (Grace et al., 2015; (Harrington & Niehaus, 2002).

## **CONCLUSION AND SUGGESTION**

### **1. Summary of Findings**

This study aimed to investigate the factors affecting risk-based capital (RBC) of life and non-life insurance companies in Indonesia and provide a classification of these companies based on their risk profiles. We employed machine-learning techniques, including decision-tree and gradient-boosted models, to identify the most important variables influencing the RBC of insurance companies. The outcome of our research points toward an intricate interplay of economic, financial, and socio-demographic determinants shaping the RBC for both life and non-life insurers. For life insurance companies, the top variables include own capital to total assets, growth rate of own capital, previous year's consumer price index, pre-tax profit to average own capital, previous year's Gini index, poverty gap, female infant mortality rate, investment to total assets, previous year's internet usage per population, and incidence of malaria per 1,000 population at risk. Similarly, for non-life insurance companies, the most important variables are equity to total assets, retained earnings growth, consumer price index, pre-tax profit to average equity, Gini index, poverty gap, female infant mortality rate, investment to total assets, internet usage per population, and incidence of malaria. Although the sentiment variable, which is our variable of interest, does not rank among the top 10 most critical factors, it still bears some importance, standing in 21st position. This suggests that, while sentiment analysis is not a key determi-

nant, it continues to have an influence on the risk-based capital of life insurance companies.

By utilizing clustering techniques, we classified insurance companies into five risk categories, ranging from lower risk to upper risk. This classification can assist stakeholders in understanding the specific risk profile of insurance companies, enabling them to develop targeted strategies to manage risks effectively. The clustering results also provide a foundation for implementing risk-based premiums or levies in the insurance industry, combining base premiums with risk-based components.

From a theoretical point of view, our research adds to what we know about the factors affecting RBC in the insurance sector, particularly in Indonesia's unique context. Managerially, our study provides practical benefits. We've categorized insurance companies by risk levels, helping stakeholders make better business decisions. Investors can better gauge potential risks and rewards, while regulators can use this information to guide their oversight. Additionally, our findings suggest a new way of setting insurance premiums based on a company's risk, leading to fairer pricing for the industry.

### **2. Limitations and Potential Improvement for Future Research**

There are some limitations that need to be acknowledged, as well as suggestions for potential improvements in future research. This study relied on publicly available data, which may be subject to reporting errors or lack of comprehensive coverage. Furthermore, the analysis was limited to the Indonesian market, which may not be generalizable to other contexts. Future research could address these limitations by collaborating with regulatory bodies or insurance companies to access more accurate data and expanding the geographical scope of the study. Additional validation and

external evaluation, as well as the incorporation of more qualitative factors, could also enhance the robustness and generalizability of the findings.

### 3. Contributions to the Field of Insurance Premium Rate Calculation

This research contributes significantly to the field of insurance premium rate calculation. The core problem this research aimed to address was understanding the contributions to the field of insurance premium rate calculation. To this end, this study provides a significant advancement in premium rate calculation methodologies tailored to the unique market dynamics of Indonesia. For insurance companies in Indonesia, this research offers a precise and efficient model for calculating insurance guarantee premium rates. This allows them to price their products accurately, ensuring coverage of claims costs while also securing profitability. For policyholders, the implications of this study instill confidence. By understanding that their insurance premiums are backed by the LPS and are calculated through a rigorous model, they are assured of financial protection against potential insolvency or bankruptcy of the insurer. Lastly, for regulators, this research offers valuable insights into the fiscal behavior of insurance entities and the potency of their pricing methods. Armed with this information, they are better positioned to guarantee that these companies not only operate with financial prudence but also align with the stipulations of Law No.4 of 2023 on Financial Sector Development and Reinforcement. Hence, this study contributes an advanced premium rate calculation method to the stakeholders in the Indonesian insurance sector.

### ACKNOWLEDGEMENT

The authors would like to express their profound gratitude to the Indonesia Financial Group

(IFG). This paper was recognized as the third-best paper in the IFG National Conference in 2023.

### REFERENCES

- Anginer, D., & Demirgüç-Kunt, A. (2018). Bank capital and the structure of the financial system. *Journal of Financial Intermediation*, 36(2), 1–15.
- Assa, H., Geffner, M., & Nektarios, M. (2019). The pricing of deposit insurance: The case of the Hellenic deposit insurance fund. *Journal of Banking Regulation*, 20(2), 152–170.
- Baranoff, E. G. (2004). Risk-Based Capital and Solvency. *In The Geneva Papers on Risk and Insurance - Issues and Practice*, 29(2), 274–285.
- Baranoff, E. G., & Sager, T. W. (2003). The relations among asset risk, product risk, and capital in the life insurance industry. *Journal of Banking & Finance*, 27(10), 2081–2100.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Cummins, J. D. (1988). Risk-based premiums for insurance guaranty funds. *Journal of Finance*, 43(4), 823–839.
- Cummins, J. D., & Sommer, D. W. (1996). Capital and risk in property-liability insurance markets. *Journal of Banking & Finance*, 20(6), 1069–1092.
- Cummins, J. D., & Weiss, M. A. (2014). Systemic risk and the US insurance sector. *Journal of Risk and Insurance*, 81(3), 489–528.
- Duan, J. C., & Yu, M. T. (2005). Fair insurance guaranty premia in the presence of risk-based capital regulations, stochastic interest rate and catastrophe risk. *Journal of Banking & Finance*, 29(10), 2431–2451.
- Dugas, S., Gao, B., & Nektarios, M. (2003). The pricing of deposit insurance considering bankruptcy costs. *The Geneva Papers on Risk and Insurance Theory*, 28(2), 137–150.

- Eling, M., & Pankoke, D. (2014). Basic risk, procyclicality, and systemic risk in the solvency II equity risk module. *J Insur Regul*, 33(2), 1–39.
- Eling, M., & Schmeiser, H. (2010). Legal and economic issues in completing the EU framework for cross-border compensation of insured bank depositors and policyholders. *Journal of Financial Regulation and Compliance*, 18(1), 65–82.
- Eling, M., & Schmit, J. T. (2012). Is there market discipline in the European insurance industry? An analysis of the German insurance market. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 37(3), 460–489.
- Epermanis, K., & Harrington, S. E. (2006). Market discipline in property-liability insurance: evidence from premium growth surrounding changes in financial strength ratings. *Journal of Money, Credit, and Banking*, 38(6), 1515–1544.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 4(2), 1189–1232.
- General Insurance Association of Indonesia, & Indonesian Life Insurance Association. (2021). *Indonesian Insurance Statistics 2020*. Aaui. <https://aaui.or.id/statistik-industri>
- Grace, M. F., Klein, R. W., & Phillips, R. D. (2002). Insurance company failures: why do they cost so much? *The Journal of Risk and Insurance*, 69(3), 301–328.
- Grace, M. F., Leverty, J. T., Phillips, R. D., & Shimpi, P. (2015). The value of investing in enterprise risk management. *Journal of Risk and Insurance*, 82(2), 289–316.
- Han, L., Li, D., Mendoza, J. P., & Ukhov, A. D. (2016). Liquidity creation through efficient M&As: a viable solution for Chinese banks facing regulatory pressure. *Journal of Financial Stability*, 25 (4), 104–114.
- Harrington, S. E., & Niehaus, G. (2002). Capital structure decisions in the insurance industry: Stocks versus mutuals. *Journal of Financial Services Research*, 21(2), 145–163.
- Hidajat, T. (2021). Unprotected Insurance: a Case from Jiwasraya. *Proceedings of the First International Conference of Economics*, 8(2), 1–9.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 2(12), 1137–1143.
- Koijen, R. S., & Yogo, M. (2015). The cost of financial frictions for life insurers. *American Economic Review*, 105(1), 445–475.
- Lee, S. J., Mayers, D., & Smith, C. W. (1997). Guaranty funds and risk-taking: Evidence from the insurance industry. *Journal of Financial Economics*, 44(1), 3–24.
- Ligon, J. A., & Thistle, P. D. (2005). The formation of mutual insurers in markets with adverse selection. *The Journal of Business*, 78(2), 529–556.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. CRC press.
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neuroinformatics*, 7(2), 21–29.
- Nektarios, M. (2010). The future of insurance regulation and supervision in the European Union. *Journal of Financial Regulation and Compliance*, 18(4), 322–336.
- Solichin, R. A. (2021). Legal Protection towards the Beneficiaries of PT Asuransi Jiwasraya Due Payment Defaults of the Jiwasraya Savings Plan: A Critical Review. *Unnes Law Journal*, 7(2), 257–286.
- Sugiarto, K., Suardi, L., & Suryandaru, S. (2019). Catastrophe insurance market

- penetration in Indonesia: Challenges and opportunities. *Jurnal Manajemen*, 23(1), 1–16.
- World Bank. (2021). *Insurance penetration, total (% of GDP) - Indonesia*. Worldbank. <https://data.worldbank.org/indicator/FB.INS>
- .TOTL.GD.ZS?locations=ID
- Yamin, M., Kusumawati, A., & Maulana, T. N. A. (2018). The role of financial literacy and education on the perception of life insurance ownership. *Management Science Letters*, 8(8), 731–742.