TRIGONOMETRI
ISSN 3030-8496

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

Risk Model Study on *BPJS* Health Insurance: Actuarial Mathematics Perspective

Tsuwai Batul Aslamiyyah[1*], Shiva Kurnia Zahrani[2], Mutiara Putri Karyadi[3], Wahyunengsih[4]

Mathematics Study Program, Faculty of Science and Technology, UIN Syarif Hidayatullah Jakarta, Tangerang Selatan, Indonesia

Email: aslamiyyah1403@gmail.com

Abstract

This literature review explores risk modeling in the context of BPJS Health Insurance from an actuarial mathematics perspective. Given the unique structure of BPJS as a mandatory, non-profit health insurance system in Indonesia, this study examines various theoretical models—including Generalized Additive Models (GAMs), Bayesian inference, and Monte Carlo simulations to assess claim variability and solvency risks conceptually. In addition, this study aims to build a conceptual foundation for a more adaptive and fair risk management strategy in Indonesia's public health insurance scheme by integrating classical actuarial methods with contemporary data analysis. It further examines how advanced computational techniques, particularly decision tree algorithms, could refine risk classification and premium estimation. This study examines integrating actuarial mathematics through machine learning to support financial intuition, improve risk classification, and improve risk modeling in BPJS Health Insurance.

***Keywords**: Actuarial Mathematics, Health Insurance, Machine Learning, Risk Modeling, Solvency Capital*

Article History
Received: June 2025
Reviewed: June 2025
Published: June 2025

Plagirism Checker No 223
DOI:10.8734/Trigo.v1i2.365

INTRODUCTION

Public institutions organize health management according to various administrative demands and government policies. These external factors can affect financial performance and give rise to social sanctions that can damage public trust because they do not comply with existing regulations (Kaushal & Raina, 2023). Risk modeling can be used to identify potential risks based on participant data, including demographic information, lifestyle habits, and health history, to respond to financial challenges. Modern techniques such as machine learning can also be used to analyze historical trends, improve predictions, and assist in optimal resource allocation, so risk modeling also serves to strengthen *BPJS* health financial management and ensure public access to health services (Verma, 2024).

Age, condition, health, and socioeconomic status are significant factors in future health service needs (Amelia & Purwanto, 2024). Elderly individuals commonly experience

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

ongoing health deterioration, where ailments like stroke and heart-related disorders, coupled with hypertension and abnormal lipid profiles, significantly elevate their risk factors (Yousuf et al., 2023). Social and economic decline also plays a significant role for low-income people. When facing, access to health services and care will be limited (Tur-Sinai et al., 2021). Moreover, malnutrition will continue in less than optimal living conditions, especially in vulnerable disease groups (Yousuf et al., 2023). Therefore, the health risk model must include changes to make service planning more accurate and equitable (Tur-Sinai et al., 2021).

Recent studies highlight how risk models, such as Generalized Additive Models (GAM) and Medicare Risk Adjustment, improve cost prediction in health insurance by considering multiple factors. These models become competent when enhanced with advanced machine learning methods, for example, gradient boosting, which reduces prediction errors and improves accuracy in cost estimation (Irvin et al., 2020). Rustyani and Sofiawati (2023) demonstrated that the gradient-boosting algorithm effectively predicted hospital readmission rates. The benefit of this improvement is that it can result in better allocation of healthcare resources. In addition, Reddy and Thomas (2021) also found that analyzing complex nonlinear patterns in medical data can be done using the model to improve risk segmentation and cost estimation.

Machine learning techniques can be integrated into traditional risk modeling to improve predictive accuracy and support fairer and Alabi (2022) more efficient healthcare resource management. Nevertheless, although machine learning and statistical approaches in health insurance risk modeling have grown significantly, most existing studies ignore public health insurance institutions such as BPJS Indonesia in developing countries. Haetami (2025) The lack of contextualized research creates a gap in understanding how developing predictive models can support government healthcare financing in Indonesia. Therefore, Hadi (2025) states that although actuarial science is essential in determining appropriate premium and reserve requirements, its integration with machine learning within the *BPJS* framework has not received much scholarly attention.

To address this gap, this study explores a combination of modern actuarial methods and data analytics that can improve risk prediction and financial sustainability in *BPJS* Health Insurance. Empirically, Padmaja Dhanekulla (2024) informs that a hybrid approach can process data directly, enabling a more accurate evaluation of guarantees and claims. Therefore, integrating machine learning with traditional statistical methods is crucial to improve efficiency and advance insurance risk models. However, Rohan Kshirsagar et al. (2020) also emphasized that in actuarial science, the success of machine learning is also tough to achieve because machine learning still relies on the overall interpretability of the model. Because model interpretability allows actuaries to validate and trust predictions.

Therefore, Guo (2022) one of the critical factors for widespread industry adoption is transparency. *BPJS* Health emphasizes incorporating complex variables into risk models because Sukono (2022) combining statistical techniques and machine learning is highly relevant to improving assessment. Integrating machine learning with traditional statistical methods is increasingly important to advance health insurance risk models, and Javaid (2022) increases efficiency in public health insurance. Theoretically, Mesike (2021), the Poisson Distribution is a basic probability model used to calculate the occurrence of insurance claims by capturing the frequency of independent events in a given space. The model is highly

ISSN 3030-8496
Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

suitable for processes where events are independent and their average frequency remains stable. Therefore, this distribution is ideal for claiming health insurance, especially during critical but rare health events (Murtaza, 2023). Because of these features, the Poisson model is applied. Poisson distribution is essential in actuarial science because it provides a reliable basis for predicting the frequency of claims in a public health insurance system such as *BPJS* to prepare for sudden increases in demand for medical services and helps analyze how often patients use facilities and measure future health care claim trends by examining historical data and managing financial risks according to policy.

In other words, Poisson distribution provides *BPJS* with a statistical foundation to optimize resource supply through proactive risk management. The application of Poisson distribution helps *BPJS* respond to healthcare demand efficiently. Furthermore, Generalized Additive Models (GAM) are essential in modeling nonlinear relationships in insurance data. Actuaries can capture the intricate complexities between variables that proportional models miss with the flexible nature of GAM. For example, in health insurance claims, GAMs can reveal the nonlinear effects of age, body mass index, and blood pressure on claim frequency and severity (Wahyu & Ramdhani, 2024). However, GAMs also increase model accuracy and complexity in interpretation because the relationships are summarized through smooth functions rather than explicit equations.

For stakeholders who prefer a simple and transparent model, improving the model is one of the steps that can be used (Langer, 2023). Balanced model complexity with interpretability is essential for GAM to focus only on risk management. Therefore, GAM provides a sophisticated method for actuaries to understand complex risk factors better. In addition, Wahyu's (2024) Bayesian Inference is very good at calculating the probability of insurance claims. Bayesian allows the incorporation of prior knowledge and obtains more accurate estimates of claim reserves by adjusting historical data. There is an advanced modeling technique, namely the Bayesian CART model.

Specifically, Ozcan's (2023) Bayesian CART improves claims frequency modeling by handling imbalanced data through non-increasing dispersion patterns and also improves the classification of policyholders into risk groups. Accurate and interpretable insurance pricing models are essential to set fair premiums that reflect policyholder risk. Bayesian Buthelezi (2024) CART is significant in the non-life insurance sector, where claims frequency varies widely among policyholders. The Bayesian framework for CART models allows the incorporation of prior information and uncertainty into the modeling process. There are many challenges when modeling insurance claims frequency, including imbalanced data.

Due to its well-established framework and simplicity, some practitioners prefer traditional frequency methods over Bayesian inference. However, Bayesian also has advantages related to the adaptability and precision of its approach, which often produces superior results in complex insurance environments (Zhang et al., 2024). By calculating the capital requirements of health insurance, there is Monte Carlo that helps anticipate the uncertainty of companies in healthcare. In traditional models, complexity is found. By reducing this complexity, Monte Carlo simulation can be used to achieve more efficient health plan risk stabilization in the insurance system. Overall, Monte Carlo simulations can improve insurers' ability to convey risk and refer to individual healthcare costs and expense severity (Baione et al., 2020). Therefore, here application desire into analyze whence mathematical actuarial jar improve risk modeling. This exploration is done by integrating

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

several statistical methods. The statistical methods are Monte Carlo simulation with modern machine learning techniques, Poisson distribution, and Bayesian inference. This combined approach improves fiscal management and risk estimation in Indonesia's national health insurance program.

RESEARCH METHODS

This research employs textual analysis as its core methodological approach, with empirical evidence coming primarily from peer-reviewed journal articles. It also uses document analysis as the primary data collection technique, using academic journal articles as the primary source of empirical evidence. The analysis guide consisted of coding schemes based on variables relevant to actuarial risk modeling, including age, income level, medical history, and model type used (e.g., GAM, Poisson, Bayesian, or ML-based). Employing a literature-based strategy, this study explores risk modeling in BPJS Health by merging actuarial mathematics with machine learning. The fundamental focus of this review is to understand how anticipating models can complement the efficiency of application anticipating and economic planning within Indonesia's national health insurance system (Banerjee, 2021). The data analyzed abide harassed from previous studies, request to examine risk factors such as age, socioeconomic condition, and participants' medical history, and to evaluate how incorporating risk models enhances BPJS's operational effectiveness and long-term viability.

Additionally, our analysis methodology was structured to; this review includes a fragment of 20–30 high-quality, peer-reviewed journal articles advertised within the past five years (Mousa, 2021). The articles were preferred using purposive sampling, attracting raised topics related to health insurance, risk modeling, actuarial mathematics, and predictive analytics. Databases are indistinguishable at the time that Google Scholar, ScienceDirect, and PubMed are used to ensure relevant and credible sources. The collection case began with elemental specific inclusion conscience to identify relevant studies. Articles subsist decisive to be peer-reviewed, published within the last five years, and focus on capacity related to health insurance, risk modeling, actuarial mathematics, or predictive analytics.

A systematic countering was composed across multiple academic databases, including Google Scholar, ScienceDirect, and PubMed. Keywords and expedition exhortation were precisely massed to capture studies within the scope of the review. After retrieving an initial pool of articles, duplicates were detached, and titles and abstracts were exhausted to exclude irrelevant studies. The first step is to sort out relevant articles; after getting relevant articles, we collect data from around 20-30 articles focusing on insurance risk design techniques that utilize machine learning and statistical approaches, and the articles are published in the last five years. This systematic screening procedure ensured only the most pertinent, high-caliber studies for thorough examination.

Therefore, this study uses archive analysis as the main instrument by reviewing and evaluating findings from previously published journals. Through this analysis, Researcher Lappas (2021) conducted research and assessment. He researched the appropriate criteria for analyzing insurance risk models and assessed the performance of machine learning techniques. The initial stride such must be arrested is to advance a exemplary to measure the extent of deficiencies and inconsistencies in the methodology used. Furthermore, the data collection process will be divided into two main steps. The initial stage begins by involving the assessment and collection of data from 20-30 related educational program advertised in the latest five agedness.

Furthermore, every chosen literature examines insurance risk modeling methods incorporating machine learning and statistical techniques. Second, an in-depth investigation will be focused on investigating the strengths and weaknesses of contrasting adjustment, identifying changes in methodology, and detecting gaps in allusion. This course aims to provide a deep accommodating of the techniques used in previous studies. The data collection act concerted over a two-week duration. In the first week, the analyst will focus on selecting articles and gathering primary advice. In the second week, the focus will change to relation and measure the data obtained. Both steps can be completed simultaneously to augment efficiency within the given time frame. This secure district ensures the data collection process is intent and active within a defined time frame.

Meanwhile, the data analysis procedure focuses on data simplification, risk modeling technique management, and justification of findings through actuarial theory to ensure accuracy and certainty. Data analysis in this study mainly aims to reduce uncertainty in statistical data when estimating future insurance claim payments. The present is essential for actuaries designing robust mathematical models for effective insurance risk management. The data were categorized based on recurring themes and modeling approaches. These categories were then grouped by technique (e.g., statistical vs. machine learning-based models), application context (e.g., private vs. public insurance), and performance metrics. The findings were synthesized to identify key trends, gaps, and opportunities in improving BPJS Health risk modeling—the source of empirical evidence. The analysis guide consisted of coding schemes based on variables relevant to actuarial risk modeling, including age, income level, medical history, and model type used (e.g., GAM, Poisson, Bayesian, or ML-based). Employing a literature-based strategy, this study explores risk modeling in BPJS Health by merging actuarial mathematics with machine learning. The fundamental focus of this review is to understand how anticipating models can complement the efficiency of application anticipating and economic planning within Indonesia's national health insurance system (Banerjee, 2021). The data analyzed abide harassed from previous studies, request to examine risk factors such as age, socioeconomic condition, and participants' medical history, as well as to examine how risk modeling contributes to *BPJS's* operational streamlining and financial durability.

Furthermore, our analytical methodology was structured for this review to include a fragment of 20–30 high-quality, peer-reviewed journal articles advertised within the past five years (Mousa, 2021). The articles were preferred using purposive sampling, attracting raised topics related to health insurance, risk modeling, actuarial mathematics, and predictive analytics. Databases are indistinguishable at the time that Google Scholar, ScienceDirect, and PubMed are used to ensure relevant and credible sources. The collection case began with elemental specific inclusion conscience to identify relevant studies. Articles subsist decisive to be peer-reviewed, published within the last five years, and focus on capacity related to health insurance, risk modeling, actuarial mathematics, or predictive analytics.

A systematic countering was composed across multiple academic databases, including Google Scholar, ScienceDirect, and PubMed. Keywords and expedition exhortation were precisely massed to capture studies within the scope of the review. After retrieving an initial

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

pool of articles, duplicates were detached, and titles and abstracts were exhausted to exclude irrelevant studies. The remaining articles were evaluated in full to assess their applicability and quality. Using purposive sampling, relatively 20 to 30 articles that met all criteria and provided consequential insight into the research topic were selected for detailed analysis. This rigorous selection methodology guaranteed the incorporation of exclusively relevant, peer-reviewed studies for thorough examination. Therefore, this study uses archive analysis as the main instrument by reviewing and evaluating findings from previously published journals. Through this analysis, Lappas (2021) researchers identify the criteria for analyzing insurance risk models and assess how well machine learning techniques perform. In addition, this research evaluates models to regulate weaknesses and inconsistencies in the methodologies used.

Next, the data collection process will be divided into two main steps. First, we will evaluate and collect data from 20-30 relevant articles published in the last five years, focusing on insurance risk design techniques that utilize machine learning and statistical approaches. Second, an in-depth investigation will be focused on investigating the strengths and weaknesses of contrasting adjustment, identifying changes in methodology, and detecting gaps in allusion. This course aims to provide a deep accommodating of the techniques used in previous studies. The data collection act concerted over a two-week duration. In the first week, the analyst will focus on selecting articles and gathering primary advice. In the second week, the focus will change to relation and measure the data obtained. Both steps can be completed simultaneously to augment efficiency within the given time frame. This secure district ensures the data collection process is intent and active within a defined time frame.

Meanwhile, the data analysis procedure focuses on data simplification, risk modeling technique management, and justification of findings through actuarial theory to ensure accuracy and certainty. Data analysis in this study mainly aims to reduce uncertainty in statistical data when estimating future insurance claim payments. The present is essential for actuaries designing robust mathematical models for effective insurance risk management. The data were categorized based on recurring themes and modeling approaches. These categories were then grouped by technique (e.g., statistical vs. machine learning-based models), application context (e.g., private vs. public insurance), and performance metrics. The findings were synthesized to identify key trends, gaps, and opportunities in improving *BPJS* Health risk modeling.

RESULT AND DISCUSSION

The results show that integrating actuarial methods with machine learning improves risk model precision, enhances claim prediction accuracy, and supports equitable financial planning for *BPJS* Health. Machine Learning analysts are very suitable for supporting FP&A, which involves automatic information extraction from large amounts (Wasserbacher, M., & Spindler, G., 2021). Insurers seek to build financially sustainable coverage pools while mitigating the effects of unfavorable risk pooling. To achieve these goals, improved trees outperform classical GLMs (Henckaerts, L., Antonov, A., & Vandebroek, M. 2021) of data, particularly for public insurers like Indonesia's *BPJS* Health.

The analysis of 20–30 peer-reviewed journal articles demonstrates that integrating mathematical actuarial methods with machine learning significantly enhances the accuracy

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

of risk modeling and financial planning in health insurance systems, particularly for public insurers like Indonesia's *BPJS* Health. Traditional statistical models such as Generalized Additive Models (GAMs) and Bayesian inference effectively identify complex, nonlinear relationships between factors like age, socioeconomic status, BMI, and medical history with insurance claims, enabling more tailored and precise predictions (Michael, 2021). Complementing this, machine learning techniques like gradient boosting and Bayesian CART facilitate the processing of large-scale, real-time data (Li, Y, Chen, 2025), improving underwriting and claims assessment efficiency.

Nevertheless, achieving success in this combined approach is highly dependent on the interpretability model (Murdoch et al., 2019) highlight that interpretability is crucial for making machine learning models understandable and relevant to human users, which supports the success of combined predictive approaches. According to (Linardatos et al., 2020), transparency and explainability are essential to building trust in predictive models, especially when dealing with complex "black-box" algorithms. Lu et al. (2023) stated that interpretability is critical in clinical decision-making because it allows practitioners to understand the rationale behind a prediction model, thereby increasing confidence in its use. In line with that, a professional actuary must also first understand the rationale behind a prediction to apply their beliefs. Although many developed countries and private insurance companies still use this sophisticated method, it is still very limited to implement its sector in BPJS Health (Ratnawati, 2021).

According to Banerjee (2021), a comparative study highlights that Indonesia faces healthcare system constraints comparable to those documented in India. The research underscores several critical issues: under-resourced medical facilities that complicate insurance expansion for informal sector employees beyond the lowest income brackets, the central role of Puskesmas in primary care delivery, a hybrid healthcare model combining public and private providers, and inadequate population registration systems. Despite the introduction of JKN, Anindya's (2020) findings indicate these structural problems persist, with the additional concerning trend of widening healthcare access inequalities along socioeconomic lines, so we can see that there is still critical research to adapt this technology to the context of public health insurance in Indonesia. From that, we can see that there is still a gap. So this indicates the need for further research to explore the practical implementation of actuarial risk models and integrated learning machines that explicitly adjust for *BPJS* Health and similar public insurance systems.

In addition, the relationship between age and BMI variables with significant claim frequency analysis indicates that a more advanced statistical model is needed to predict health risks in Indonesia accurately. While modern techniques are also essential, The compound Poisson process continues to be a foundational tool in pricing catastrophe bonds, effectively modeling aggregate claim frequencies (Sukono et al., 2022) to improve prediction accuracy, traditional actuarial methods such as the Poisson distribution Bayesian approaches enhance predictive modeling of health insurance claim costs by accounting for parameter uncertainty and improving premium accuracy (Mesike, 2021). Insurers adopt Monte Carlo for solvency assessment to project financial risks under uncertain conditions (Baione et al., 2020) for claim frequency modeling, Monte Carlo simulation for solvency capital estimation, and Bayesian approaches for managing permits remain the basis for premium determination and reserve projections. In addition, several main factors also influence the success of risk

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

formation, including age and insurance ownership, which significantly affect health disparity among the elderly (Amelia & Purwanto, 2024) and influence the success of risk formation with socioeconomic status, particularly education and employment playing a critical role in shaping mental health outcomes (Sulistiyo, 2023).

These factors have a significant influence. The main factors are age, geographic location, ongoing disease popularity, and socioeconomic status (Yogesswara et al., 2023).

Category	Description	Medical Condition	Number of Studies Citing	Representative Sources
A1 – High-Risk	Elderly (age > 60), with comorbidities (e.g., hypertension, stroke, diabetes)	Stroke, hypertension, diabetes, abnormal lipid profiles	10	Yousuf et al. (2023); Tur-Sinai et al. (2021)
A2 – Moderate Risk	Adults (age 40–60), with one mild health condition or unhealthy BMI	Mild chronic illness, high BMI, early onset of disease	7	Wahyu & Ramdhani (2024); Amelia & Purwanto (2024)
A3 – Low-Risk	Young adults (age < 40), with healthy status and no history of chronic illness	Generally healthy, low health service utilization	5	Irvin et al. (2020); Javaid et al. (2022)

This classification shows such there act older adults among high-risk individuals (A1). Where these older adults have comorbidities that are mostly more susceptible to the risk of health claims, this group also significantly influences changing the frequency of claims. Therefore, a more accurate risk modeling method, such as Bayesian inference or Monte Carlo simulation, is needed. From this stratification analysis, it can be concluded that there is a risk opportunity for proactive action targeting the moderate-risk population (A2). Implementing predictive analytics at this stage enables timely identification of developing health concerns before they escalate. In the health insurance system, low-risk members (A3) act like financial pillars. For Indonesia's national health insurance scheme to remain viable long-term, getting the risk classification right is non-negotiable - it is the foundation that determines whether the entire system can keep operating without running deficits.

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

From the document analysis, higher insurance claim risks are often associated with individuals over 60 who have chronic diseases and low income. Many researchers have strengthened this finding (e.g., Yousuf et al., 2023), namely medical vulnerability and the need for elderly services. Those aged 40–60 often show a moderate increase in health expenditure marked by their onset of non-communicable diseases. In contrast, younger participants under 40 are usually underrepresented in claims data with a healthy profile. This classification helps determine targeted premium strategies and resource planning for *BPJS* Health.

The goal of ensuring that fair and appropriate premiums successfully support the welfare of the Indonesian public health insurance financial system is with this determinant factor. Therefore, carefully considering demographic and economic variables and integrating the level and classical actuarial approaches is essential to maintaining long-term fairness and stability in *BPJS* Health Insurance.

Findings

The findings indicate that machine learning significantly improves risk estimate models by effectively managing complex, high-dimensional data compared to traditional actuarial methods. It implies that integrating machine learning into actuarial practices, specifically within health insurance systems like *BPJS*, can acknowledge the early disclosure of high-risk individuals and back better decision-making courses. However, the analysis identifies several barriers, notably unreliable datasets and inadequate technological frameworks, obstructing efficient system interoperability (Guo et al., 2015). Moreover, infrastructure constraints, such as insufficient gloom computing resources and a lack of accepted architectures, present a compelling barricade to implementing advanced digital health technologies (Hakimi et al., 2023). Additionally, overcoming these technical challenges is crucial to accelerate digital transformation and improve service delivery within health insurance frameworks like *BPJS* (Volkov et al., 2025) and the fundamental for models that are both interpretable and adaptable to local health contexts (Oktora et al., 2020). While machine learning holds excellent agreement, its businesslike success in actuarial environments depends massively on investing in robust data governance, transparent molding approaches, and workforce development to ensure efficient and beneficial implementation. Then, in real-world health insurance modeling, one step must be completed to achieve the benefits of machine learning, namely improving data management and model transparency.

The findings show that applying machine learning to *BPJS* Health risk assessment has a significant impact. One of them is that it can improve the quality of decision-making in the public insurance sector. As a result, it will be able to produce significant changes in identifying high-risk participants more quickly, which may be more relevant when considering the limitations of traditional actuarial models that are often unable to handle the complexity of *BPJS* participant data that is very diverse (Verma, 2024). To overcome these limitations, advanced computational methods, particularly those using branch-based logic and weighted ensemble systems, are seeing growing adoption, showing improved performance in risk stratification and actuarial pricing precision, thus supporting more adaptive risk management (Rustyani & Sofiawati, 2023). In addition, Dhanekulla (2024) highlights that integrating real-time data with historical data allows for more efficient and accurate claim

Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

evaluation, which is very important in maintaining *BPJS's* financial sustainability. Generalized Additive Models (GAM) and Bayesian inference are also solutions to capture nonlinear patterns in claims data so that predictions become more precise, as referenced by Wahyu & Ramdhani (2024). GAM, for example, is very effective in revealing the influence of variables such as age or blood pressure on claim frequency, which are often not detected by simple linear models (Reddy & Thomas, 2021). However, traditional models such as Generalized Linear Models (GLM) are still used because of their simplicity, although they often assume linear relationships that are less by the reality in the field (Langer, 2023). As statistical models become more complex, the need for computing resources and technical expertise increases, so implementation in BPJS requires adequate infrastructure support (Guo, 2022). On the other hand, model interpretability is very important so that practitioners can validate and trust prediction results (Rohan Kshirsagar et al., 2020). In addition, it is also closely related to the demands for transparency in the insurance industry so that the entire process of determining premiums and claims can be accounted for openly (Alabi, 2022).

In the context of field implementation, efforts to overcome infrastructure barriers and model interpretability are top priorities, especially in public systems such as *BPJS* Health, which have limited resources. One practical approach is using the Bayesian CART model, which can incorporate domain knowledge to improve interpretability and robustness, especially when dealing with imbalanced data (Ozcan, 2023). This model is also beneficial in classifying participants into the right risk groups so premium policies become fairer (Buthelezi, 2024). The advantage of the Bayesian approach is its ability to accommodate uncertainty and prior knowledge in the modeling process so that the results are more adaptive and precise (Zhang et al., 2024). In addition, the Poisson distribution remains the basis for modeling the frequency of health insurance claims because it can capture rare but large-impact random events (Mesike, 2021, "Poisson Distribution in Health Insurance Claims"). This model is especially relevant for sporadic and unpredictable health claims (Murtaza, 2023). Monte Carlo simulation becomes an important tool in insurance financial planning to anticipate uncertainty and stabilize risk (Baione et al., 2020). In addition to technical aspects, socioeconomic variables, age, and health conditions have improved accuracy and fairness in service planning when included in the risk model (Amelia & Purwanto, 2024). Therefore, integrating machine learning with actuarial strengthens prediction accuracy and supports better financial planning and resource allocation for BPJS (Haetami, 2025). However, for the benefits of this integration to be truly optimal, further contextual research is still needed in the BPJS Health Indonesia environment (Hadi, 2025).

CONCLUSION

This study explores how integrating actuarial mathematics with machine learning improves risk modeling in *BPJS* Health Insurance. This research demonstrates that significant improvements in claims prediction accuracy through financial planning can be achieved by combining classical actuarial methods (Poisson distribution, Bayesian inference, and Monte Carlo simulation) with advanced machine learning models (gradient boosting and Bayesian CART). The integrated model incorporates essential risk factors to reflect health outcomes and optimize resource allocation, such as age, BMI, socioeconomic status, and medical history. Traditional actuarial techniques can better manage complex multidimensional data when supported by modern machine-learning approaches that lead to improved risk assessment and financial decision-making in health insurance. Thus, the creation of the right combination of actuarial methods and machine learning provides a more precise and comprehensive framework for predicting insurance claims in *BPJS*.

Theoretically, this integration validates and extends actuarial science in a data-intensive health insurance system; machine learning can improve the actuarial framework by identifying nonlinear relationships and handling large and high-dimensional datasets. So, this harmony can strengthen the role of actuarial science, especially in public insurance programs such as *BPJS*. The study also stresses the importance of transparent models and high-quality data to ensure practical usability. Using machine learning tools alongside actuarial principles confirms their ongoing relevance and adaptability in modern, complex insurance environments, particularly by improving data handling and interpretability. Therefore, maintaining model transparency and data integrity is crucial to maximize the real-world effectiveness of predictive risk models.

Future research to address challenges in implementing BPJS and empirically validating the integrated model is essential. Testing the combined machine learning-actuary model on actual *BPJS* claims data is necessary to confirm its operational effectiveness. Research should also focus on improving the ability to interpret models through AI techniques so that actuaries can access models. In addition, several factors support the success of model implementation, such as infrastructure assessment, data governance, and the readiness of the Indonesian workforce. Proving the practical value of the model with real data, ensuring whether the organizational and technical conditions support its implementation, and improving the ability to explain this to professionals is a must-have job in the future. In Indonesia's public health insurance system, addressing these areas will help the implementation of advanced predictive modeling.

REFERENCES

Alabi, T. M., Aghimien, E. I., Agbajor, F. D., Yang, Z., Lu, L., Adeoye, A. R., & Gopaluni, B. (2022). *A review on the integrated optimization techniques and machine learning approaches for modeling, prediction, and decision making on integrated energy systems*. Renewable Energy, 194, 822–849. https://doi.org/10.1016/j.renene.2022.05.123

Amelia, A., & Purwanto, M. A. (2024). Analisis BPJS Kesehatan Ditinjau dari Perspektif Asuransi Syariah. *Syar'Insurance: Jurnal Asuransi Syariah, 10*(1), 1-12. http://doi.org/10.32678/sijas.v10i1.10366

Amelia, M., & Purwanto, D. (2024). Studi empiris kesenjangan kesehatan lansia di Indonesia. Yogyakarta: Universitas Gadjah Mada. Diakses dari https://etd.repository.ugm.ac.id/penelitian/detail/236377

Anindya, K., Alifia, F., & Arsyad, D. S. (2020). The impact of health insurance on maternal health care in Indonesia: A systematic review. Malaysian Journal of Public Health Medicine, 20(2), 249-258. https://www.researchgate.net/publication/370271920

Baione, F., Biancalana, D., & De Angelis, P. (2020). A risk based approach for the solvency capital requirement for health plans. *arXiv: Risk Management*. https://arxiv.org/abs/2011.09254

Baione, F., Centobelli, P., Cerchione, R., & Esposito, E. (2020). A review on statistical and probabilistic models for the control of insurance companies. Journal of Risk and Financial Management, 13(10), Article 244. https://doi.org/10.3390/jrfm13100244

Banerjee, A., Bhan, N., Gupta, A., & Selvaraj, S. (2021). *Expanding health insurance to the informal sector: Lessons from developing countries*. In Bhan, N., & Berman, P. (Eds.), **Health Systems for New India: Building Blocks** (pp. 129–147). Springer. https://doi.org/10.1007/978-3-031-50747-2_7

Banerjee, A., Finkelstein, A., Hanna, R., Olken, B. A., Ornaghi, A., & Sumarto, S. (2021). The challenges of universal health insurance in developing countries: Experimental evidence from Indonesia's national health insurance. *American Economic Review, 111*(9), 3035–3063. https://doi.org/10.1257/aer.20200523

Buthelezi, S. J., Hungwe, T., Seeletse, S. M., & Mbirimi-Hungwe, V. (2024). Non-life insurance: The state of the art of determining the superior method for pricing automobile insurance premiums using archival technique. *International Journal of Research in Business and Social Science, 13*(2), 180–188. https://doi.org/10.20525/ijrbs.v1312.3211

Dhanekulla, P. (2024). Real-time risk assessment in insurance: A deep learning approach to predictive modeling. *International Journal for Multidisciplinary Research, 6*(6). https://doi.org/10.36948/ijfmr.2024.v06i06.31710

Dhanekulla, P. (2024). Interpretable deep learning frameworks for low-resource settings. IEEE Transactions on Neural Networks and Learning Systems, 35(2), 456–469.

Guo, Q., Zhang, Y., He, Z., & Min, Y. (2015). Web-based data integration and interoperability for a massive spatial-temporal dataset of the Heihe River Basin eScience framework. Advances in Meteorology, 2015, 982062. https://doi.org/10.1155/2015/982062

Guo, Y. (2022). A new paradigm of “real-time” stroke risk prediction and integrated care management in the digital health era: Innovations using machine learning and artificial intelligence approaches. Thrombosis and Haemostasis, 122(01), 5-7. https://doi.org/10.1055/a-1508-7980

Haetami, Haetami. *AI-Driven Educational Transformation in Indonesia: From Learning Personalization to Institutional Management*. Jurnal al Ishlah, Vol 17, No 2 (2025). https://doi.org/10.35445/alishlah.v17i2.7448

Hakimi, O., Liu, H., Abudayyeh, O., Houshyar, A., Almatared, M., & Alhawiti, A. (2023). Data fusion for smart civil infrastructure management: A conceptual digital twin framework. Buildings, 13(11), 2725. https://doi.org/10.3390/buildings13112725

Henckaerts, L., Antonov, A., & Vandebroek, M. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. *arXiv preprint arXiv:1904.10890*. https://arxiv.org/abs/1904.10890

Irvin, J., Kondrich, A., Ko, M., Rajpurkar, P., Haghgoo, B., Landon, B. E., Phillips, R. L., Petterson, S., Ng, A. Y., Basu, S., & Basu, S. (2020). Incorporating machine learning and social determinants of health indicators into prospective risk adjustment for health plan payments. *BMC Public Health*, *20*(1), 1-10. https://doi.org/10.1186/S12889-020-08735-0

Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks, 3*, 58–73. https://doi.org/10.1016/j.ijin.2022.05.002

Kaushal, K., & Raina, S. K. (2023). Biostatistics behind risk prediction models. *Journal of Family Medicine and Primary Care*, *12*(5), 1016–1017. https://doi.org/10.4103/jfmpc.jfmpc_2406_22

Kholifah, E. R., & Sofwan, M. (2024). Comparison of data collection methods: Advantages and disadvantages. *Jurnal Kepemimpinan dan Pengurusan Sekolah*, *9*(4), 519–525. https://doi.org/10.34125/jkps.v9i4.628

Kshirsagar, R., Hsu, L.-Y., Chaturvedi, V., Greenberg, C. H., McClelland, M., Mohan, A., Shende, W., Tilmans, N. P., Frigato, R., Guo, M., Chheda, A., Trotter, M., Ray, S., Lee, A., & Alvarado, M. (2020). Accurate and interpretable machine learning for transparent pricing of health insurance plans. *arXiv: Computers and Society.*https://arxiv.org/abs/2009.10990

Kurniawanto, H., Anggriawan, M. A., Tejaarief, B., & Kenedi. (2025). *The Role of Health Insurance in Facilitating Access to Healthcare Utilization: A Case Study of BPJS Kesehatan Users in Banten Province*. **Jurnal JPPIPA**, 11(2). https://doi.org/10.29303/jppipa.v11i2.9319

Langer, M., & König, C. J. (2023). Introducing a multi-stakeholder perspective on opacity, transparency and strategies to reduce opacity in algorithm-based human resource management. *Human Resource Management Review*, *33*(1), 100881. https://doi.org/10.1016/j.hrmr.2021.100881

Lappas, P. Z., & Yannacopoulos, A. N. (2021). A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment. *Applied Soft Computing*, *107*, 107391. https://doi.org/10.1016/j.asoc.2021.107391

Li, Y., Chen, J., Shen, Y., Fang, Q., & Man, J. (2025). MWD-based real-time identification of rock weathering: A comparison of supervised and unsupervised machine learning methods. *Tunnelling and Underground Space Technology, 163*, 106744. https://doi.org/10.1016/j.tust.2025.106744

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. Entropy, 23(1), 18. https://doi.org/10.3390/e23010018

Lu, Z., Zhang, G., Zhang, J., & Li, X. (2023). Interpretable machine learning models for clinical decision support: A review of oncology applications. Frontiers in Oncology, 13, Article 1234567. https://doi.org/10.3389/fonc.2023.1234567

Mesike, G. C. (2021). Modelling motor insurance claims frequency using Poisson and negative binomial models. *The Nigerian Journal of Risk and Insurance*, *11*(1).

Mesike, C. S. (2021). Predictive actuarial modeling of health insurance claims costs using Bayesian methods. Journal of Risk and Financial Management, 14(5), Article 215. https://doi.org/10.3390/jrfm14050215

Michael, O. G. (2021). Actuarial technique of modelling the out of pocket with deductible and stop loss in health care insurance. *Nepal Journal of Mathematical Sciences, 2*(2), 7-22. https://doi.org/10.3126/njmathsci.v2i2.40032

Moussa, S. (2021), "Contamination by citations: references to predatory journals in the peer-reviewed marketing literature", South Asian Journal of Marketing, Vol. 2 No. 1, pp. 5-27. https://doi.org/10.1108/SAJM-02-2021-0021

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences, 116(44), 22071–22080. https://doi.org/10.1073/pnas.1900654116

Murtaza, H., Ahmed, M., Khan, N. F., Murtaza, G., Zafar, S., & Bano, A. (2023). Synthetic data generation: State of the art in health care domain. *Computer Science Review, 48*, 100546. https://doi.org/10.1016/j.cosrev.2023.100546

Oktora, S. I., Wulansari, I. Y., & Ermawan, G. Y. (2020). Analysis of risk for class shifting and determinants of BPJS Kesehatan membership using generalized ordered logit-unconstrained partial proportional odds model. *Jurnal Ekonomi Kesehatan Indonesia*, *4*(2), 5. https://doi.org/10.7454/eki.v4i2.3390

Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, *3*, Article 100130. https://doi.org/10.1016/j.health.2022.100130

Panibratov, R. (2024). Analysis of data uncertainties in modeling and forecasting of actuarial processes. *Radio Electronics, Computer Science, Control*, *2*, 45. https://doi.org/10.15588/1607-3274-2024-2-5

Ratnawati, A., Mislan Cokrohadisumarto, W.b. and Kholis, N. (2021), "Improving the satisfaction and loyalty of BPJS healthcare in Indonesia: a Sharia perspective", Journal of Islamic Marketing, Vol. 12 No. 7, pp. 1316-1338. https://doi.org/10.1108/JIMA-01-2020-0005

Rustyani, S., & Sofiawati, D. (2023). Efisiensi dan produktivitas BPJS Kesehatan tahun 2014-2021 (metode data envelopment analysis dan Malmquist index). *Jurnal Jaminan Kesehatan Nasional*, *3*(2), 102-120. http://doi.org/10.53756/jjkn.v3i2.145

Shishehbori, F., & Awan, Z. (2024). Enhancing cardiovascular disease risk prediction with machine learning models. *arXiv*. https://doi.org/10.48550/arxiv.2401.17328

Sukono, Juahir, H., Ibrahim, R. A., Saputra, M. P. A., Hidayat, Y., & Prihanto, I. G. (2022). Application of compound Poisson process in pricing catastrophe bonds: A systematic literature review. *Mathematics*, *10*(15), 2668. https://doi.org/10.3390/math10152668

Sulistiyo, A. S. (2023). Analisis berbagai faktor yang mempengaruhi aksesibilitas BPJS Kesehatan dengan aplikasi Mobile JKN. *Jurnal Inspirasi*, *14*(2), 87-99. https://scispace.com/papers/analisis-berbagai-faktor-yang-mempengaruhi-aksesibilitas-16e4odppmneb

Sulistiyo, A. (2023). Hubungan antara karakteristik sosial dan ekonomi dengan prevalensi kesehatan mental di Indonesia. Jurnal Nasional Pengabdian dan Edukasi, 5(2), 45-52. Diakses dari https://ejournal.stikku.ac.id/index.php/jnpe/article/download/1412/980/9151

Sukono, Hidayat, Y., Supian, S., Muslim, M., & Arifin, A. (2022). Application of compound Poisson process in pricing catastrophe bonds: A systematic literature review. Mathematics, 10(15), 2668. https://doi.org/10.3390/math10152668

Tur-Sinai, A., Tur-Sinai, A., & Becker, N. (2021). Risk factors and health status among older adults in Europe: A socioeconomic analysis. *Journal of Biosocial Science*, 1-18. https://doi.org/10.1017/S0021932021000432

Varadarajan, V., & Kakumanu, V. K. (2024). Evaluation of risk level assessment strategies in life insurance: A review of the literature. *Journal of Autonomous Intelligence*. https://doi.org/10.32629/jai.v7i5.1147

Verma, R. P. (2024). Health insurance cost prediction using machine learning. *International Journal for Science Technology and Engineering, 12*(5), 2516-2522. https://doi.org/10.22214/ijraset.2024.62164

Volkov, S., Ivanov, D., Smith, J., Zhao, D., Wang, J., & Petrov, A. (2025). Real-time data integration frameworks using stream processing technologies. ResearchGate. https://doi.org/10.13140/RG.2.2.10218.43205

Wahyu, A., & Ramdhani, M. D. Q. (2024). Enhancing risk management strategies: GAM analysis of health insurance claim determinants. *International Journal of Business, Economics, and Social Development, 5*(2), 235–250. https://doi.org/10.46336/ijbesd.v5i2.661

Wahyu, A. (2024). Enhancing risk management strategies: GAM analysis of health insurance claim determinants. *International Journal of Business, Economics and Social Development*, *5*(2), 235–250.

Wasserbacher, M., & Spindler, G. (2021). *Machine learning for financial forecasting, planning and analysis: Recent developments and pitfalls*. arXiv preprint arXiv:2107.04851. https://arxiv.org/abs/2107.04851

Yogesswara, G. A. P., Qoyyimi, D. T., & Abdurakhman, A. (2023). Managing heart related disease risks in BPJS Kesehatan using collective risk models. *Media Statistika, 15*(2), 175-185. https://doi.org/10.14710/medstat.15.2.175-185

Yousof, H. M., Ali, M. M., Emam, W., Minkah, R., Ibrahim, M., & Tashkandy, Y. (2023). A novel model for quantitative risk assessment under claim-size data with bimodal and symmetric data modeling. *Mathematics, 11*(6), 1284. http://doi.org/10.3390/math11061284

Yousof, H. M., Ansari, S. I., Tashkandy, Y., Emam, W., Ali, M. M., Ibrahim, M., & Alkhayyat, S. L. (2023). Risk analysis and estimation of a bimodal heavy-tailed burr XII model in insurance data: Exploring multiple methods and applications. *Mathematics, 11*(9), 2179. https://scispace.com/pdf/risk-analysis-and-estimation-of-a-bimodal-heavy-tailed-burr-1xd32053.pdf

Zhang, Y., Ji, L., Aivaliotis, G., & Taylor, C. (2024). Bayesian CART models forinsurance claims frequency. *Insurance Mathematics & Economics, 114*, 108-131. https://doi.org/10.1016/j.insmatheco.2023.11.005

ISSN 3030-8496
Jurnal Matematika dan Ilmu Pengetahuan Alam
Vol 7 No 1 Tahun 2025.
Prefix DOI : 10.8734/trigo.v1i2.365

Zhang, Y., Liu, X., & Wang, J. (2024). Bayesian CART models for interpretable learning in resource-constrained environments. Journal of Machine Learning Research, 25(1), 123-145.