Bulletin of Applied Mathematics and Mathematics Education
Volume 5, Number 1, April 2025 pp. 45-54
https://dx.doi.org/10.12928/bamme.v5i1.13114
e-ISSN 2776-1029, p-ISSN 2776-1002

Predicting financial distress in Indonesian life insurance companies with classification methods and synthetic features generation

Dwi Purwanto, Dedy Dwi Prastyo*

Department of Statistics, Institut Teknologi Sepuluh Nopember, Jl. Surabaya, Jawa Timur 60111 Indonesia

*Corresponding e-mail: dd.prastyo@its.ac.id

ARTICLE INFO

Article History
Received 28 April 2025
Revised 28 May 2025
Accepted 28 May 2025

Keywords
Extreme Gradient Boosting
Financial distress
Life insurance
Synthetic Features Generation

How to cite this article
Purwanto, D., & Prastyo, D. D. (2025). Predicting financial distress in Indonesian life insurance companies with classification methods and synthetic features generation. *Bulletin of Applied Mathematics and Mathematics Education*, *5*(1), 45-54.

ABSTRACT

Financial problems in life insurance companies can become serious if not addressed immediately. Companies experiencing financial distress, for instance, are unable to meet their obligations to pay their liabilities. A company can be categorized as experiencing financial distress when it has an RBC ratio of less than 120%—based on regulation by the Indonesian Finance Service Authority—or ROA < 0 (suffering loss). Therefore, financial distress prediction is carried out to assess the company's current financial condition so that it can be handled early. In this study, we aimed to predict financial distress of Indonesian life insurance companies. We utilized the Support Vector Machine (SVM) classification method, Generalized Extreme Value Regression (GEVR), and Extreme Gradient Boosting (XGB) and by incorporating synthetic feature generation in variable selection. The results of financial distress prediction obtained the best model that can predict the financial condition of life insurance companies in Indonesia at each size, where for sizes 0 and 1, the XGB model with variable selection produces accuracy values of 98.00% and 94.10%, respectively, and AUC values of 100% and 87.40%. Then, at size 2, we can use Stepwise Generalized Extreme Value Regression with accuracy and AUC results of 90.20% and 82.60%, respectively. Each addition of size to the time window classification results tends to reduce the model's performance in predicting the financial condition of life insurance companies in Indonesia.

Introduction

A company operates primarily with the objective of generating profit. However, in practice, its economic activities are often subject to various uncertainties. Effective management and the formulation of appropriate strategies are crucial for maintaining resilience in various circumstances and preventing financial failure. As the complexity of risks increases, it becomes imperative to implement comprehensive risk management procedures encompassing the stages of identification, measurement, monitoring, and control. Insurance companies that neglect the application of sound risk management practices may face significant financial challenges, potentially leading to financial distress and eventual bankruptcy.

The insurance sector plays a critical role in supporting Indonesia's economic stability.

According to the Indonesian Financial Services Authority, insurance is a contractual agreement in which an insurer commits to provide compensation in exchange for premium payments (OJK, 2023). Life insurance, a key segment of the industry, provides financial benefits upon the death or survival of the insured party, depending on the policy terms. Insurance companies contribute to national development by mitigating financial risks faced by individuals and institutions while simultaneously mobilizing and managing funds that facilitate investment and economic growth.

Between 2017 and 2021, several life insurance companies in Indonesia encountered severe financial distress. Notably, PT Asuransi Jiwasraya failed to fulfill maturing policy obligations amounting to IDR 802 billion in 2018, a figure that escalated to IDR 12.4 trillion in 2019. This crisis stemmed from the provision of long-term, high-return products, such as the JS Savings Plan, which ultimately led to the company's solvency ratio falling below the minimum threshold set by OJK (Azhar & Hidayat, 2021). Similarly, PT Kresna Life Insurance experienced default on multiple products, with pending claims reaching IDR 6.4 trillion. As a consequence, OJK issued a Business Activity Restriction Sanction (PKU) in August 2020. These events underscore the importance of proactive measures in predicting financial distress using financial statement analysis.

Financial distress represents the critical stage preceding bankruptcy, characterized by a persistent decline in a firm's financial health (Platt & Platt, 2002). This condition can be identified if a company shows deteriorating financial performance over three consecutive years (Kordestani et al., 2021). The factors contributing to financial distress can be categorized into internal and external dimensions (Ariesta, 2013). Internal factors, typically microeconomic, include issues such as negative cash flow, excessive liabilities, and operational losses. In contrast, external factors are generally macroeconomic, including rising interest rates and changes in the broader financial environment. A company is considered financially distressed when its Risk-Based Capital (RBC) ratio falls below 120%. Past cases of financial distress in the insurance sector serve as crucial lessons for avoiding similar occurrences in the future, thereby emphasizing the necessity of financial statement analysis over a multi-year horizon.

Numerous studies have been conducted on predicting financial distress in Indonesian companies. Ramadhani (2022) demonstrated that the Generalized Extreme Value Regression (GEVR) model outperforms Support Vector Machine (SVM) and Logistic Regression across various sample sizes, particularly when variable selection is applied. Calabrese and Osmetti (2011) also highlighted the superior performance of GEVR in modeling rare event outcomes compared to logistic regression. Another study from Zieba et al. (2016) found that an ensemble method employing boosted trees and synthetic feature generation yielded better predictive performance than traditional classification models

Building upon the aforementioned studies, the present research aims to predict financial distress among life insurance companies in Indonesia using classification methods, including SVM, GEVR, and Extreme Gradient Boosting (XGB). This study incorporates variable selection and synthetic feature generation techniques to enhance model performance. Variable selection is utilized to address the issue of multicollinearity, particularly due to the high intercorrelation among financial ratios. For the SVM model, regularization techniques are applied. The GEVR model employs stepwise selection, and for the XGB model, feature importance is assessed through gain values.

Method

This study utilizes secondary data from Ramadhani (2022) to predict financial distress in life insurance companies in Indonesia from 2017 to 2021. This data is obtained from the company's

10.12928/bamme.v5i1.13114

financial statements, OJK, BI, and BPS publications, using financial ratio data of life insurance companies and macroeconomic indicators. The data is divided into training data and testing data, with testing using the last year for RBC or ROA variables. The structure of training and testing data (not employing cross-validation) for each memory time window size is illustrated in Figure 1.

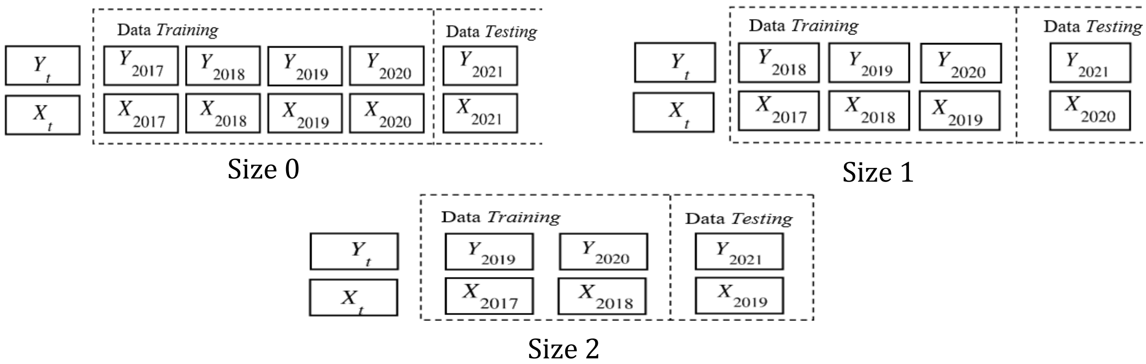

Figure 1. Full-time memory window size

Therefore, the number of samples in each window size is summarized in Table 1. The study variables consist of response variables (RBC and ROA), which categorize the company into financial distress (Y=1) if RBC < 120% or ROA < 0, and 14 predictor variables, including 10 financial ratios such as Underwriting Ratio (UR), Loss Ratio (LR), Commission Expense Ratio (CER), Investment Yield Ratio (IYR), Liability to Liquid Asset Ratio (LLAR), Premium Receivable to Surplus Ratio (PRSR), Investment to Technical Reserve Ratio (ITRR), Net Premium Growth (NPG), Own Retention Ratio (ORR), and Technical Reserve Ratio (TRR), one company size, and three macroeconomic indicators such as inflation, interest rates, and economic growth.

Table 1. Sample size in training and testing data for each window size

Size	Training			Testing		
	Sample size	Distressed (%)	Non-distress (%)	Sample size	Distressed (%)	Non-Distress (%)
0	204	6.37	93.63	51	9.80	90.20
1	153	6.54	93.46	51	9.80	90.20
2	102	7.84	92.16	51	9.80	90.20

This study employs a general equation with one response variable and multivariable predictors, along with the generation of new synthetic features from the three main models used: the Support Vector Machine, Generalized Extreme Value Regression, and Extreme Gradient Boosting models. Each model will be classified to obtain predictions of financial distress in life insurance companies using variable selection models, such as Lasso SVM, Elastic-Net SVM, Stepwise GEVR, and Extreme Gradient Boosting Variable Selection. The variable selection method will eliminate insignificant predictor variables in the model.

The concept of drift is employed in this study to examine how the financial ratios of life insurance companies evolve over time. In addition, the data is divided into three-time window sizes (0, 1, and 2) to evaluate the company's financial distress condition, as shown in Table 2.

To select the model with the best prediction and performance, error rates are calculated to evaluate the classification process's performance (Johnson & Wichern, 2007). Accuracy, recall, and specificity are some evaluation measures used (Hotho et al., 2005). The area under the curve (AUC), which summarizes the model's performance through the ROC curve, is used to evaluate the classification accuracy on imbalanced data.

Table 2. Time window size of research data

Size	Response	Predictor
0		X_t
1	Y_t	X_{t-1}
2		X_{t-2}

Concept of drift

The concept of drift is a grouping over time without loss of generality of batches that are assumed to have the same size, each containing n samples. (Scholz & Klinkenberg, 2007). Concept drift is often referred to as a change in the target concept resulting from changes in the underlying context. Changes can occur directly or indirectly, and can be handled by using a window size on the training data. The solution with a time window size aims to predict events at a specific time by considering the condition at the previous time, based on the time window size. Data grouping is based on the next period, where each batch has the same size.

Synthetic features generation

Synthetic features generation combines econometric measures using arithmetic operations such as addition, subtraction, multiplication, and division (Zięba et al., 2016). Each synthetic feature can be seen as a single regression model. The purpose of using synthetic features is to combine econometric indicators proposed by experts into more complex features. Synthetic features can be viewed as hidden features extracted by neural networks, but their extraction methods differ. The generated synthetic features have values resulting from their arithmetic operations, without any transformation or pre-processing.

Support vector machine

Classification provides more knowledge related to large amounts of data and is then grouped into several categories to facilitate the processing of data analysis funds. There are several classification algorithms, one of which is the SVM method. SVM is a supervised learning algorithm that solves classification problems by finding a hyperplane that can separate two data sets from two different classes (Cortes & Vapnik, 1995). The advantage of using the SVM method is that it determines the distance using a support vector, making the computational process faster. The SVM method consists of two types: linear SVM, which is used to separate data linearly, and non-linear SVM, which is used to separate data non-linearly. Additionally, SVM has a convex objective function, which guarantees that the solution is globally optimal.

SVM in linear classification uses a decision boundary, which will determine the classification of training data, so that a linear model or hyperplane can be formed that is most optimal in classifying the data (Qasthari, 2017). A kernel will be used for non-linear mapping in non-linear SVM, and then the optimal hyperplane is sought using linear SVM (Dewi & Purnami, 2015). Linear SVM does not employ a kernel function, while non-linear SVM employs a kernel function, even if the kernel type is the linear kernel.

Standard Support Vector Machine models do not have the ability to select variables. A specific approach to adding regularization to SVM models is to use a regularization term with sparsity to replace the $L2$-norm in standard SVM. $L1$-norm is a convex function, with Lipschitz continuity,

 10.12928/bamme.v5i1.13114

having better properties than other norms (Li, 2019). Lasso SVM and similar extensions have developed into one of the most important tools for data analysis.

The Lasso SVM method was first introduced by Tibshirani (1996). The Lasso SVM method is one of the methods of Regularized SVM that uses the $L1$-norm penalty to produce rare solutions, such as variable coefficients that shrink to zero. The approach of this method is to eliminate irrelevant variables from the SVM model, thereby obtaining an SVM model with only relevant variables (Hardle et al., 2014).

The Elastic-Net SVM method addresses the shortcomings of the Lasso SVM method. The Elastic-Net SVM method is a combination of $L1$-norm and $L2$-norm penalties, where the $L1$-norm penalty is used to select predictor variables, while the $L2$–norm penalty helps stabilize the coefficient parameter estimator of predictors that exhibit multicollinearity (Hardle et al., 2014). Compared to Lasso, elastic-net has the advantage of selecting highly correlated features in clusters (i.e., clustering effect) while still implementing sparsity in the solution.

Generalized extreme value regression

Generalized Extreme Value Regression (GEVR) modeling is based on the existence of Extreme Value Theory. Extreme value theory is one way to overcome rare events that have a big influence (Putri et al., 2011). In this theory, three extreme value distributions are used, namely the Gumbel Distribution (Type 1), the Frechet Distribution (Type 2), and the Weibull Distribution (Type 3). The difference in the tails of different distributions makes it challenging to determine the distribution pattern of the extreme value. The three distributions can be combined into one type of distribution called the Generalized Extreme Value Distribution, which is highly flexible with a tail shape parameter τ that controls the shape and size of the tails of the three different distribution families included. The combined distribution becomes a single parametric representation, so that the cumulative function of the distribution is obtained as follows (Calabrese & Osmetti, 2013).

Extreme gradient boosting

Extreme Gradient Boosting Model (XGB) is one of the implementations of Gradient Boosting Machines (GBM), which is known as one of the best-performing algorithms used in Supervised Learning methods (Osman et al., 2021). XGB is an ensemble of K Classification and Regression Trees (CART) where xi is the training data given to predict the class label yi (Chen & Guestrin, 2016). Parallel and distributed computing makes the execution process faster, which is favored by data scientists (Mo et al., 2019).

The XGB algorithm can accept continuous and discrete variables as input, but the output variables must be discrete, including binary variables (Mo et al., 2019). When using the XGB algorithm, a Z-statistic is often used to test the significance of each independent variable with a p-value given at the 95% confidence interval (Chen & Guestrin, 2016). Variable selection with the Extreme Gradient Boosting method can be performed by examining the gain value.

Results and discussion

In the results and discussion chapter, this research presents an analysis of financial distress classification in life insurance companies in Indonesia using SVM, XGB, and GEVR methods with variable selection. It involves synthetic features generation (SFG) to improve model accuracy. The involvement of new variables consists of the generation of synthetic features. This study involves combining two new microeconomic indicator variables with the arithmetic method of multiplication and division. However, this method has some weaknesses, namely, the newly generated variables that have a high correlation with the forming variables. Therefore, elimination will be done for variables with a correlation value of greater than 0.5. Then, 10 new variables were obtained from synthetic feature generation that correlated less than 0.5, as presented in Table 3.

Table 3. Synthetic features generation result variables

Variable
UR x IYR; UR x PRSR; UR x ITRR; LR x CER; CER x LLAR; CER x ITRR; IYR x LLAR; IYR x TRR; LLAR x ITRR; PRSR x NPG

This paper does not present descriptive statistics, such as mean, minimum, and maximum, because these have already been reported in Ramadhani (2022). This study presents an analysis of financial distress classification in life insurance companies in Indonesia using SVM, XGB, and GEVR methods with variable selection. The correlation of each generated feature with the factors is presented in Table 4.

Table 4. Correlation of each generated feature with the factors

Variable	Corr with Factor 1	Corr with Factor 2
UR x IYR	0.483	0.180
UR x PRSR	0.019	-0.065
UR x ITRR	0.500	-0.377
LR x CER	0.421	0.463
CER x LLAR	0.389	0.425
CER x ITRR	0.389	0.221
IYR x LLAR	0.398	0.432
IYR x TRR	0.409	0.404
LLAR x ITRR	0.277	0.500
PRSR x NPG	0.031	0.129

Next, the resulting classification of financial distress in life insurance companies, along with an evaluation of classification accuracy, is presented in Table 5.

Table 5. Model comparison with synthetic features generation

Size	Model	Accuracy		AUC	
		Training	Testing	Training	Testing
0	*Elastic-Net SVM*	97.50%	94.10%	*84.30%*	*70.00%*
	Lasso SVM	95.60%	96.10%	*68.90%*	*80.00%*
	XGB	***96.50%***	***98.00%***	***99.20%***	***100.0%***
	Stepwise GEVR	*95.60%*	*98.00%*	*91.50%*	*98.90%*
1	*Elastic-Net SVM*	96.70%	*94.10%*	*75.00%*	*70.00%*
	Lasso SVM	*93.50%*	*90.20%*	*50.00%*	*50.00%*
	XGB	***99.30%***	***94.10%***	***100.0%***	***87.40%***
	Stepwise GEVR	*94.70%*	*90.20%*	*92.60%*	*73.70%*
2	*Elastic-Net SVM*	*92.10%*	*90.20%*	*50.00%*	*50.00%*
	Lasso SVM	*92.10%*	*90.20%*	*50.00%*	*50.00%*
	XGB	*97.10%*	*90.20%*	*100.0%*	*50.00%*
	Stepwise GEVR	***92.10%***	***90.20%***	***89.90%***	***82.60%***

Based on Table 5, the best model obtained is Extreme Gradient Boosting with variable selection, which was achieved by adding new synthetic features generation variables at sizes 0 and 1. Then, at size 2, the best model obtained is the Stepwise Generalized Extreme Value Regression model, as using a higher size tends to reduce the model's performance. The confusion matrices for each method, employed for each size, are presented in Table 6, Table 7, Table 8, and Table 9.

 10.12928/bamme.v5i1.13114

Table 6. Confusion matrix of Lasso SVM

Size	Training			Testing		
	Prediction	Actual		Prediction	Actual	
		Distressed	No-Distressed		Distressed	No-Distressed
0	Distressed	5	8	Distressed	3	2
	No-Distresses	1	190	No-Distressed	0	46
1	Distressed	0	0	Distressed	0	0
	No-Distressed	10	143	No-Distressed	5	46
2	Distressed	0	0	Distressed	0	0
	No-Distressed	8	94	No-Distressed	5	46

Table 7. Confusion matrix of Elastic Net SVM

Size	Training			Testing		
	Prediction	Actual		Prediction	Actual	
		Distressed	No-Distressed		Distressed	No-Distressed
0	Distressed	9	1	1	2	0
	No-Distressed	4	190	0	3	46
1	Distressed	5	0	1	2	0
	No-Distressed	5	143	0	3	46
2	Distressed	0	0	1	0	0
	No-Distressed	8	94	0	5	46

Table 8. Confusion matrix of GEVR

Size	Training			Testing		
	Prediction	Actual		Prediction	Actual	
		Distressed	No-Distressed		Distressed	No-Distressed
0	Distressed	4	9	Distressed	1	4
	No-Distressed	2	189	No-Distressed	0	46
1	Distressed	3	7	Distressed	2	3
	No-Distressed	1	142	No-Distressed	2	44
2	Distressed	1	7	Distressed	1	4
	No-Distressed	1	93	No-Distressed	1	45

Table 9. Confusion matrix of XGB

Size	Training			Testing		
	Prediction	Actual		Prediction	Actual	
		Distressed	No-Distressed		Distressed	No-Distressed
0	Distressed	6	7	Distressed	4	1
	No-Distressed	0	191	No-Distressed	0	46
1	Distressed	9	1	Distressed	2	3
	No-Distressed	0	143	No-Distressed	0	46
2	Distressed	5	3	Distressed	0	5
	No-Distressed	0	94	No-Distressed	0	46

Meanwhile the ROC curve, from which the AUC is calculated, is exhibited in Figures 2, Figure 3, and Figure 4.

In imbalanced data conditions, a suitable model performance measure, such as AUC, is necessary because accuracy results that are more suitable for balanced data can skew the classification results towards the majority class. This can be seen from the results above where the Lasso SVM and Elastic-Net SVM models have good accuracy values at each size.

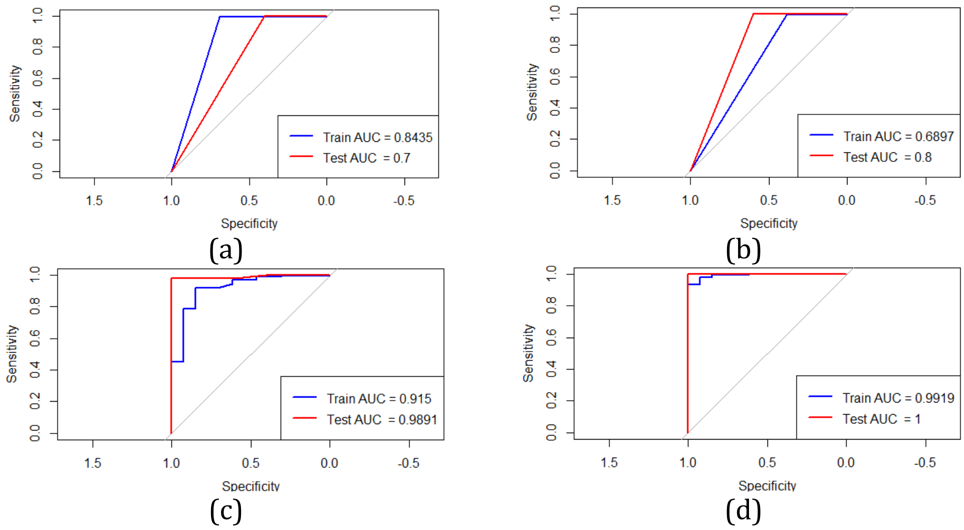

Figure 2. ROC of training and testing data for Size 0 of Lasso SVM (a), E-Net SVM (b), GEVR (c), XGB (d).

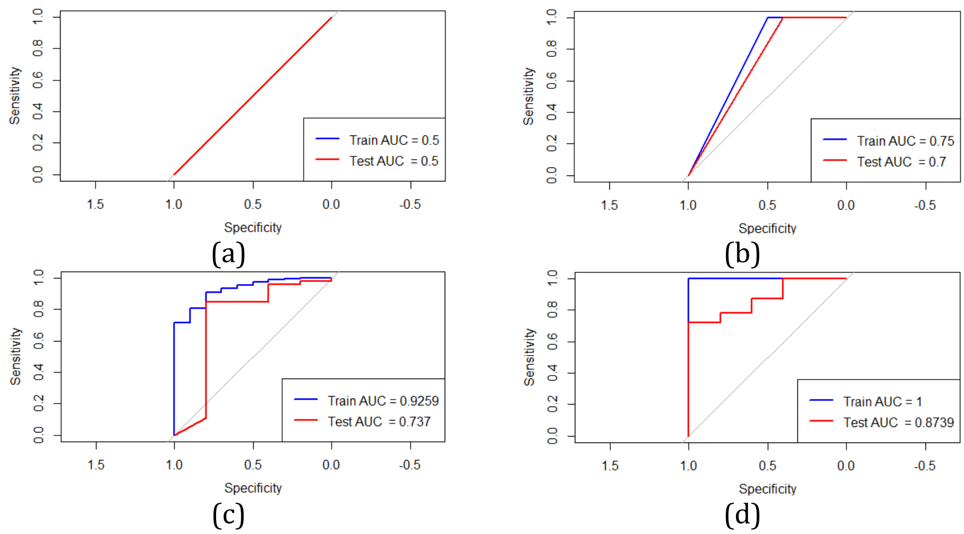

Figure 3. ROC of training and testing data for Size 1 of Lasso SVM (a), E-Net SVM (b), GEVR (c), XGB (d).

However, the AUC value decreases which tends to be low compared to XGB and GEVR when used in training to testing data and tends to experience overfitting and is less suitable for use in financial distress prediction. The XGB and GEVR methods outperform Lasso SVM and Elastic-Net SVM, making them more suitable for predicting financial distress in extreme data.

This study demonstrates that OJK, as the regulator of the life insurance industry, can enhance its supervision system by utilizing XGB and GEVR methods, complemented by synthetic feature generation, thereby improving prediction accuracy. Given the characteristics of unusual cases and imbalanced data with a limited number of samples, this method is highly effective in predicting financial crises.

 10.12928/bamme.v5i1.13114

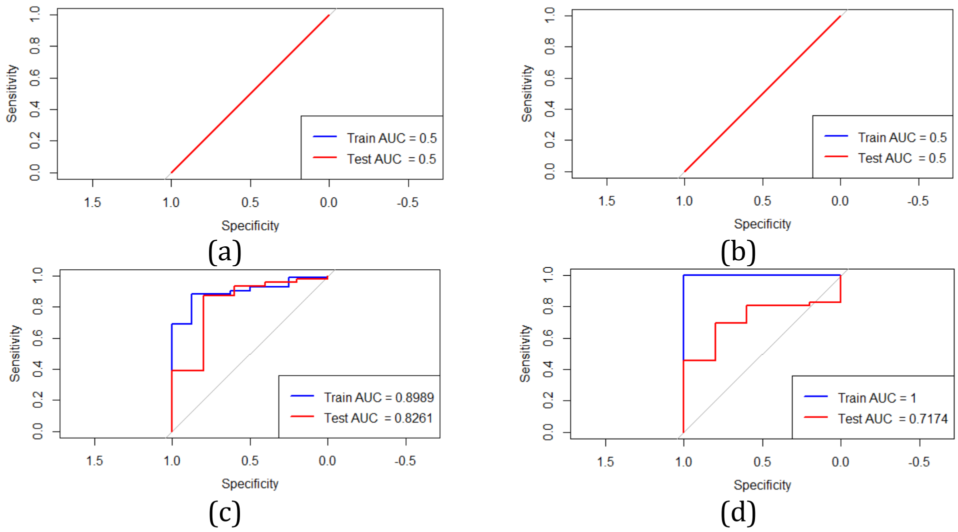

Figure 4. ROC of training and testing data for Size 2 of Lasso SVM (a), E-Net SVM (b), GEVR (c), XGB (d).

By implementing this system, the OJK can establish a more accurate early warning mechanism to identify symptoms of financial crisis in insurance companies in a timely manner. This early identification allows for quick intervention to reduce the risk of bankruptcy by providing policy direction, restructuring advice, or other preventive measures. The results of this analysis not only prevent companies from collapsing but also safeguard the interests of policyholders and maintain the overall financial system's stability.

Conclusion

In this study, the prediction of financial distress in life insurance companies in Indonesia is made using classification methods, including SVM, GEV Regression, and XGB, which involve SFG in both variable selection and prediction. The new variables formed with the generation of the synthetic features are ten variables. Financial distress data has an extreme imbalance proportion because it is not expected to occur. The best model for predicting financial distress in life insurance companies in Indonesia, at sizes 0 and 1, is the XGB model obtained through variable selection, where the accuracy and AUC values between the training and testing data do not differ much, so that no overfitting occurs. The accuracy and AUC of testing data at size 0 are 98.00% and 100.0%, and at size 1, they are 94.10% and 87.40%. Then, at size 2, we can use Stepwise Generalized Extreme Value Regression with accuracy and AUC results of 90.20% and 82.60%, respectively. Each addition of size to the time window classification results tends to reduce the model's performance in predicting the financial condition of life insurance companies in Indonesia. The suggestions for further research are to use other classification methods that are sensitive to imbalance data, use oversampling techniques such as SMOTE to balance classes, increase the amount of data so that the model evaluation results are more optimal, and expand the definition of financial distress with ROA limits less than 120% and ROA less than zero to make the classification categories more diverse.

Acknowledgement

Special thanks to the reviewers for their queries, which have improved the quality of this article.

References

Ariesta, D. R., & Chariri, A. (2013). Analisis pengaruh struktur dewan komisaris, struktur kepemilikan saham dan komite audit terhadap financial distress. *Diponegoro Journal of*

Accounting, 1(1), 1-9.

Azhar, H., & Hidayat, N. (2021). *Penegakan Hukum yang Mengganggu Roda Ekonomi Kasus Jiwasraya dan Dampaknya terhadap Pasar Modal Indonesia*. Lokataru Foundation.

Calabrese, R., & Osmetti, S. A. (2011). Generalized extreme value regression for binary rare events data: an application to credit defaults. *Bulletin of the International Statistical Institute LXII, 58th Session of the International Statistical Institute*, 5631-5634.

Calabrese, R., & Osmetti, S. A. (2013). Modelling small and medium enterprise loan defaults as rare events: the generalized extreme value regression model. *Journal of Applied Statistics*, *40*(6), 1172–1188. https://doi.org/10.1080/02664763.2013.784894

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Dewi, M. R., & Purnami, S. W. (2015). Klasifikasi hasil pap smear test sebagai upaya pencegahan sekunder penyakit kanker serviks di rumah sakit "X" Surabaya menggunakan Piecewise Polynomial Smooth Support Vector Machine (PPSSVM). *Jurnal Sains dan Seni ITS, 4*(1), D61-D66.

Hardle, W., Prastyo, D. D., & Hafner, C. (2014). Support vector machines with evolutionary model selection for default prediction. In J. Racine, L. Su, & A. Ullah, *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press.

Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. J*ournal for Language Technology and Computational Linguistics, 20*(1), 19-62.

Johnson, R., & Wichern, D. (2007). *Applied Multivariate Statistical Analysis Sixth Edition* (6th ed.). Pearson Education, Inc.

Kordestani, M., Saif, M., Orchard, M. E., Razavi-Far, R., & Khorasani, K. (2021). Failure Prognosis and Applications—A Survey of Recent Literature. *IEEE Transactions on Reliability*, *70*(2), 728–748. https://doi.org/10.1109/TR.2019.2930195

Li, Q. (2019). Logistic and SVM Credit Score Models Based on Lasso Variable Selection. *Journal of Applied Mathematics and Physics*, *07*(05), 1131–1148.

Mo, H., Sun, H., Liu, J., & Wei, S. (2019). Developing window behavior models for residential buildings using the XGBoost algorithm. *Energy and Buildings*, *205*, 109564. https://doi.org/10.1016/j.enbuild.2019.109564

OJK. (2023). *Asuransi*. Otoritas Jasa Keuangan.

Osman, A. I. A., Najah Ahmed, A. N., Chow, M. F., Huang, Y. F., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

Platt, H. D., & Platt, M. B. (2002). Predicting corporate financial distress: Reflections on choice-based sample bias. *Journal of Economics and Finance*, *26*(2), 184–199. https://doi.org/10.1007/BF02755985

Putri, R. N., Djuraidah, A., & Rahman, L. O. A. (2011). *Analisis Ukuran Risiko Keuangan Menggunakan Teori Nilai Ekstrim : Studi Kasus Indeks Harga Saham Gabungan Periode 2001-2010*. IPB University. http://repository.ipb.ac.id/handle/123456789/51709

Qasthari, E. T. (2017). *Teknik Pengukuran: Metode Klasifikasi Support Vector Machine (SVM) pada Data Pengukuran*. Universitas Indonesia.

Ramadhani, S. W. (2022). Prediksi financial distress perusahaan asuransi jiwa di indonesia dengan metode support vector machine, generalized extreme value regression, dan logistic regression. *Thesis*. Institut Teknologi Sepuluh November.

Scholz, M., & Klinkenberg, R. (2007). Boosting classifiers for drifting concepts. *Intelligent Data Analysis*, *11*(1), 3–28. https://doi.org/10.3233/IDA-2007-11102

Zięba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, *58*, 93–101. https://doi.org/10.1016/j.eswa.2016.04.001

 10.12928/bamme.v5i1.13114