Media Statistika
p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 16(1) 2023: 1-12
http://ejournal.undip.ac.id/index.php/media_statistika

CORPORATE FINANCIAL DISTRESS PREDICTION USING STATISTICAL EXTREME VALUE-BASED MODELING AND MACHINE LEARNING

Dedy Dwi Prastyo[1], Rizki Nanda Savera[1], Danny Hermawan Adiwibowo[2]
[1] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2] Bank Indonesia, Indonesia

e-mail: dedy-dp@statistika.its.ac.id

DOI: 10.14710/medstat.16.1.1-12

Article Info:
Received: 9 September 2022
Accepted: 18 April 2023
Available Online: 27 April 2023

Keywords:
Classification; feature selection; financial distress; GEVR; imbalanced data; machine learning.

.

Abstract: The industrial sector plays a leading role in an economy such that the financial stability of companies from this sector be a big concern. Two financial ratios, i.e., the Interest Coverage Ratio (ICR) and the Return on Assets (ROA), are used to determine the corporate financial distress conditions. This work considers two schemes for determining financial distress. First, a company is categorized as distressed if either ICR<1 or ROA<0. The second scheme is for when both ICR<1 and ROA<0 are met. The proportion of distressed and non-distressed companies is imbalanced. Our work views the distressed companies (minority class) as a rare event, causing the proportion to be extremely small, such that the Extreme Value Theory can be employed. The so-called Generalized Extreme Value regression (GEVR), developed from GEV distribution, predicts the distressed labels. The GEVR's performance is compared using machine learning with and without feature selection. The feature selection in GEVR uses backward elimination. The model for prediction employs a drift or windowing concept, i.e., using past-period predictors to predict the current response. The empirical results found that the GEVR, with and without the feature selection, provides the best prediction for financial distress.

1. INTRODUCTION

In 2019, the industrial sector contributed the most to Indonesia's Gross Domestic Product (GDP) at 19.7 percent. National tax revenue from the manufacturing industry until the end of December 2019 was the main contributor, with a grant of 29.23 percent. The global financial crisis that significantly impacts the business area may occur in line with the increasing volatility of the future capital market. This condition encourages the company to strengthen the company fundamentally. If the company cannot deal with global challenges, it will run into a financial distress condition that can lead to default, i.e., failing to service its liability and bankruptcy. In such a case, the shareholders will lose confidence in the company. Subsequently, they will break off the partnership, and the company will run into bankruptcy. According to that problem, financial distress analysis is valuable to be conducted. As an influential sector of the economy, industrial companies have to face this threat and avoid bankruptcy.

An accuracy improvement of financial distress (leading to default) prediction can be converted into significant savings for shareholders. Hence the advancement of financial distress prediction technology attracts much attention from researchers and practitioners. Similar problems in finance, such as fraud of credit cards, defaulted companies, and customer churn, always become a big concern as they impact the economy and investment. A reduced-form technique is commonly known as the primary approach in such kinds of problems. The Logit model obtained much attention during the 1980s, replacing the popularity of univariate and multivariate discriminant analysis. Non-linear classification approaches such as Neural Networks (NN) and Deep Learning, Long Short Term Memory (LSTM) networks (Wang et al., 2019), Deep Learning (Natasha et al., 2019), Support Vector Machines (SVMs) (Haerdle et al., 2014), and other machine learning algorithms like AdaBoost and majority voting (Randhawa et al., 2018), and hybrid approaches are recently popularly applied to solve such financial problems.

The binary class of financial distress and non-distress data, typically like other financial data, is imbalanced, which raises a problem because the classes are not equally represented. The imbalanced data causes an accuracy paradox where a simple model, even random guessing, produces high accuracy but is too crude to be helpful. There are some strategies to work with imbalanced data. First, collect more data to achieve a balance of the class representatives. Unfortunately, it does not work in some instances because the business process that generates the data always produces imbalanced data. Like in finance, the regulator pushes the bad events (fraud, default, financial distress, and others) tend to be zero such that the proportion of classes may never balance.

Second, to change the performance metric by employing a more robust measurement. The area under Receiver Operating Curve (ROC), the so-called AUC (Hanley and McNeil, 1982), is widely used to evaluate the predictive model applied to imbalanced data. The AUC is the probability that the classification model correctly ranks random positive classes (for example, financially distressed companies) higher than random negative ones (non-financially distressed firms). So an AUC close to one is often considered a confirmation of a good model. When applied to balance data, the AUC may perform as well as an accuracy measure, and however, it gives unbiased results compared to the accuracy when applied to imbalanced data. Third, do resampling to make the classes balance. This approach can be made by copying observations from the under-represented class (over-sampling approach) or deleting observations from the over-represented class (under-sampling approach). Instead of just creating copies, generating synthetic samples is the fourth alternative. One of these popular approaches is the Synthetic Minority Over-sampling Technique (SMOTE), which creates synthetic observations from the minor class. One can see the work of Prasetya and Aburakhman (2022), who apply SMOTE on Random Forests and *k*-NN for classifying the imbalanced data. The fifth alternative solution to handle the problem arising from imbalanced data is applying various statistical and machine learning approaches instead of only applying the researcher's favorite methods. Some of them are mentioned in the previous paragraph.

The following alternative solution employs the regularized or penalized models. The penalty function imposes an additional cost on the model for making classification mistakes on the minority class during training. These regularizations or penalties can push the model to pay more attention to the minority class. Penalization also applies a feature section to the underlying model, and the selected features significantly reduce over-fitting prediction accuracy. The financial ratios chosen as features in this research commonly have a multicollinearity problem that causes assumption violation when applied in the parametric approach.

Moreover, the non-linear relationship between features and responses (distress and non-distress) implies that the linear model coefficient no longer represents the feature's relative importance (Haerdle and Prastyo, 2014). Supervised classification methods assign a company to belong to a distressed or non-distressed class. This technique is straightforward to employ all given features, although only subsets are relevant. The embedded feature selection method chooses relevant features along with the estimation process. This technique helps to avoid selecting irrelevant features that lead to overfitting. This procedure also excludes features that cause collinearity. This paper employs an embedded method via the regularization approach, with logistic regression and SVMs as an underlying classification method, to provide an automatic feature selection and parameter estimation. Two regularization forms employed in this work are (i) Least absolute shrinkage and selection operator (Lasso) and (ii) Elastic-net.

This research views the imbalanced data from a different perspective. The minority class (financially distressed companies) is considered a rare event. Thus, the Extreme Value Theory (EVT), commonly used to analyze rare or extreme events, can be adopted to develop a classification model with imbalanced, even severe imbalanced, proportions over the classes. The distribution function of the Generalized Extreme Value (GEV) distribution derived from EVT with the Block Maxima approach is then used to replace the logistic regression model's probability of success. Such a classification model is called the Generalized Extreme Value regression (GEVR) (Calabrese and Osmeti, 2013; Calabrese & Giudici, 2015). This research applies the GEVR to predict the label of distressed and non-distressed companies. The contributing features are financial ratios, including activity, profitability, solvency, and liquidity ratios. Its performance is compared with standard statistical and machine learning algorithms, logistics regression, and SVMs, with and without feature selection, using the Lasso (Tibshirani, 1996) and Elastic-net (Zou and Hastie, 2005). The feature selection in GEVR uses backward elimination. The performance comparison of those methods is evaluated using accuracy and AUC criteria. The definition of financial distress in this research is determined based on two financial ratios, i.e., the Interest Coverage Ratio (ICR) and the Return on Assets (ROA). There are two schemes considered in this work in determining distress. First, a company is categorized as distressed if either ICR<1 or ROA<0, whereas the second scheme is for when both ICR<1 and ROA<0 are met.

Also, a drift concept or windowing, i.e., employing features from past periods to predict response at the current time, is studied here. If the window size is zero, i.e., the response and predictor variables are observed from the same period, the developed model can not be used to forecast the response label in the next (several) years using the current predictors. This case produces a model meaningful for interpretation or prediction but not in the context of forecasting. Likewise, it is useful for interpolation, not extrapolation. Suppose the window size is a positive integer greater than zero. In that case, the developed model established from current period predictors can be used to forecast the response label in the future, depending on the window size. For example, if the window size is one (year), the developed model can forecast one year ahead of the response label using the current period predictor. Likewise, it is helpful for extrapolation. This breakthrough is needed to have broader views about financial distress such that alternative solutions can be offered to solve the problems.

2. LITERATURE REVIEW

2.1. Feature Selection with Regularization

Given training data $\{(x_1^T, y_1), \ldots, (x_n^T, y_n)\}$, each observation consists of a predictor $x_i^T \in R^p, i = 1,2,\ldots,n$, where n is the number of observations, and the associated response $y_i \in Y = \{-1,+1\}$. Each score obtained from each univariate predictor $x_j \in R^n$, for $j = 1,2,\ldots,p$, should reflect the similarity between companies, i.e., a lower score reflects lower default risk—the more similar the two companies, the smaller the difference between their scores. In Support Vector Machine (SVM), a classifier $f: R^p \to \{-1,+1\}$ predicts the label of the response variable for any new observation. A linear classifier $f(x_i) = x_i^T w + b = 0$ divides the space into two regions belonging to each class -1 and $+1$. The SVM aims to maximize the margin between $f(x_i) = -1$ and $f(x_i) = +1$. The non-linear classifier is obtained by applying the kernel trick to the dual problem:

$$\min_{\boldsymbol{\alpha}} \boldsymbol{L_D(\alpha)} \qquad (1)$$
$$\boldsymbol{s.t.} \quad \boldsymbol{0 \le \alpha_i \le C,\ \sum_{i=1}^{n} \alpha_i y_i = 0},$$

where $L_D(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i'=1}^{n} \alpha_i \alpha_{i'}\, y_i y_{i'}\, K(x_i, x_{i'})$, with $K(x_i, x_{i'})$ is a kernel function, and α is a Lagrange multiplier. The constant C in equation (1) is the Lagrange cost parameter controlling the amount of misclassification. See Haerdle et al. (2014) for a more detailed explanation.

Standard L_2-norm SVM can be written in the regularization form as in equation (2):

$$\boldsymbol{\sum_{i=1}^{n}\{1 - y_i f(x_i)\}_+ + \lambda\, \|\mathbf{w}\|_2^2}, \qquad (2)$$

where $\lambda > 0$ is a tuning parameter that controls the trade-off between the loss function and penalty. The hinge-loss function $L\{y_i, f(x_i)\} = \{1 - y_i f(x_i)\}_+ = max\{0, 1 - y_i f(x_i)\}$ is not differentiable at $L\{y_i, f(x_i)\} = 0$ and is a convex upper bound for $\{0 - 1\}$-loss function. The second part of the equation (2) is a penalty function, i.e., $(w) = \|w\|_2^2$. Equation (2) employs all input to construct classifiers such that it can not select relevant predictors. This work uses other penalty terms to select the relevant predictors.

The Lasso technique introduced by Tibshrani (1996) produces sparse feature solutions. Bradley and Mangasarian (1998) and Zhu et al. (2004) applied Lasso to SVM by employing L_1-norm penalty term $P(\mathbf{w})$ instead of L_2-norm as in equation (3).

$$\boldsymbol{\sum_{i=1}^{n}\{1 - y_i f(x_i)\}_+ + \lambda_L \|\mathbf{w}\|_1}, \qquad (3)$$

where large λ_L force some estimates of w_j to be zero.

The elastic-net penalty, a compromise between the ridge and Lasso formulated in (4), can solve Lasso's drawback such that the correlated features are selected or discarded together.

$$\boldsymbol{P(\mathbf{w}) = \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2^2}, \qquad (4)$$

with tuning parameter λ. Fig 1 illustrates the L_1-norm, L_2- norm, and elastic-net penalties (left side) for a single coefficient and its contour (right side) for coefficients corresponding to two features. If the weight λ_1 and λ_2 change, then the shape of $P(\mathbf{w})$ and its contour will change accordingly. The weight λ_1 and λ_2 can be reformulated as γ and $(1-\gamma)$, respectively. The Lasso regularization is employed when $\gamma = 1$. For a very small $\varepsilon > 0$, the elastic-net penalty with $\gamma = 1 - \varepsilon$ performs like the Lasso but removes any degeneracies and wild behavior caused by high correlations between predictors (Friedman et al., 2010). The Lasso and Elastic-net penalties, as illustrated in Fig 1, are then embedded in the logistic regression to enable feature selection works (Herdle et al., 2014).

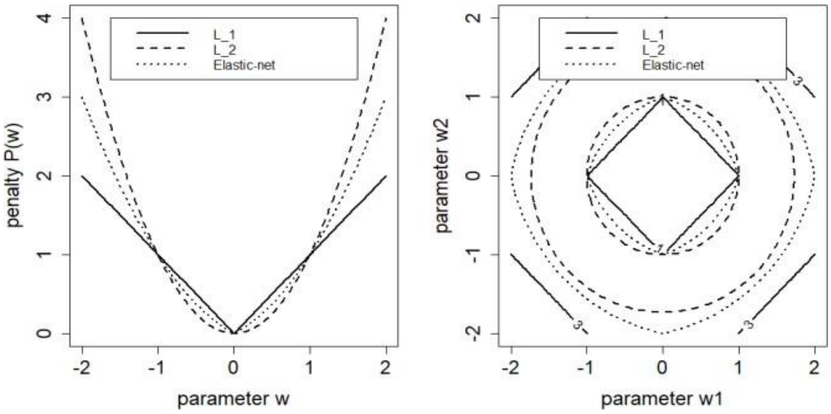


Figure 1. Penalty functions for single w_j (left) and contour for w_1 and w_2 with $\gamma = 0.5$ (right)

2.2. Extreme Value based-Classification

This work views the class of distressed companies as a minority group analogous to rare events. Therefore, the EVT, commonly used to analyze rare or extreme events, can be developed for a classification model with imbalanced data. The GEV distribution function derived from the EVT with Block Maxima approach is employed to replace the probability of success $\pi = P(y = 1)$ or $\pi(\mathbf{x}_i) = P(y_i = 1|\mathbf{x}_i)$ if there are covariates, the GEV distribution combines the following three distributions: Fréchet, Gumbel, and Weibull. The distribution function of GEV is formulated in (5) range between zero and one. It is analogous to a probability of success π.

$$\boldsymbol{F(z)} = \mathbf{exp}\left\{-\left[\mathbf{1} + \boldsymbol{\tau}\left(\frac{\boldsymbol{z-\mu}}{\boldsymbol{\sigma}}\right)\right]^{-\mathbf{1}/\boldsymbol{\tau}}\right\}, \tag{5}$$

with τ is a shape parameter, μ is a location parameter, and σ as a scale parameter. The probability of success π is replaced by $F(z)$. If the binary response variable is affected by the feature's value, then the probability of success, given that the feature is updated, becomes equation (6).

$$\boldsymbol{\pi}(\mathbf{x}_i) = \mathbf{exp}\left\{-[\mathbf{1} + \boldsymbol{\tau}(\boldsymbol{\beta}'\mathbf{x}_i)]^{-\mathbf{1}/\boldsymbol{\tau}}\right\}. \tag{6}$$

Such a model is called Generalized Extreme Value Regression (GEVR) with its link function known as *gevit*. For more details, see Calabrese and Osmeti (2013) and Calabrese & Giudici (2015)—the *gevit* link function as formulated in equation (7) as follows.

$$\frac{\{-\boldsymbol{ln}[\boldsymbol{\pi}(\mathbf{x}_i)]\}^{\boldsymbol{\tau}} - \mathbf{1}}{\boldsymbol{\tau}} = \boldsymbol{\beta}'\mathbf{x}_i \tag{7}$$

$$\boldsymbol{gevit}(\boldsymbol{\pi}(\mathbf{x}_i)) = \frac{\{-\boldsymbol{ln}[\boldsymbol{\pi}(\mathbf{x}_i)]\}^{\boldsymbol{\tau}} - \mathbf{1}}{\boldsymbol{\tau}} = \boldsymbol{\beta_0} + \sum_{j=1}^{p} \boldsymbol{\beta_j x_{ij}}.$$

The Maximum Likelihood Estimation (MLE) does not provide a closed-form solution for the GEVR model such that the estimators are obtained using numerical optimization. The initial value of the estimator plays a vital role as the likelihood function is not convex.

3. METHODOLOGY

3.1. Data and Variables

The financial ratios datasets are calculated from companies' financial reports under Indonesia's industrial sector, spanning from 2005-Q4 until 2018-Q4 (2030 observations). Two financial ratios, the Interest Coverage Ratio (ICR) and Return on Asset (ROA), are used to determine the financial distress condition. Financial distress (FD) condition (Y=1) is assigned to a company when the ICR<1 or the ROA<0. Otherwise, the financial condition is good. The whole analysis is repeated for the ICR<1 and the ROA<0. There are 14 financial ratios as micro-economy covariates and four macro-economy indicators (to capture the indirect effect of when one company defaults to other companies, see Prastyo et al. (2017) and Prastyo et al. (2018). Missing values imputation in financial ratios is calculated using the Nearest Neighbor (k-NN) approach (Kowarik and Templ, 2016). The variables analyzed in this research are summarized in Table 1.

Table 1. Research Variables and Their Descriptions

Response	Variable	Description			
Y	ICR<1 or/and ROA<0	1: The company is financially distressed, 0: The company is in good condition, Scheme one: using "or"; Scheme two: using "and"			
Predictor	**Variable**	**Description**	**Predictor**	**Variable**	**Description**
X_1	EBITA	Earnings Before Income Tax to Asset	X_{10}	CR	Current Ratio
X_2	STA	Sales to Total Asset	X_{11}	QR	Quick Ratio
X_3	ITR	Inventory Turnover Ratio	X_{12}	ETD	Earning to Debt
X_4	DSIR	Day's Sales in Inventory Ratio	X_{13}	WCA	Working Capital to Total Asset
X_5	ROE	Return on Equity	X_{14}	WCS	Working Capital to Sales Ratio
X_6	NPM	Net Profit Margin	X_{15}	RGDPG	Real GDP Growth (%yoy)
X_7	OPM	Operating Profit Margin	X_{16}	BI7DRR	BI 7-Day Repo Rate
X_8	DER	Debt to Equity	X_{17}	USD/IDR	USD/IDR rate
X_9	DAR	Debt to Asset	X_{18}	Inflation	Inflation

3.2. Method

Predictive modeling trained the model from historical data and used that trained model to predict output variables' value or label on the new information of input variables. The mapping function from input to output is approximated by an algorithm or method that expects the prediction to be close to perfect. The relationship between input and output is commonly assumed to be static. This setting is correct for many problems but not for all problems. The relationship between input and output can change over time; in specific cases, it changes sequentially.

The concept of drift is the change over time in the underlying problem's relationship between input and output data in the underlying problem. Specifically, this concept observes the effect between input and output when the data are collected over time. This problem can be resolved by managing the training algorithm's window size. In this paper, it is interesting to know the impact of financial ratios in previous years as covariates of current response (FD). The size equals zero when the response data and covariates are in the same year. The size equal one means using covariates from one year past used to predict the financial distress in the current year. This approach enables using current predictors to predict the response label's year ahead (forecast). The concept drift in this research is illustrated in Fig 2. The size zero is used for interpretation from the model, while the size one, two, and three years are used for prediction.

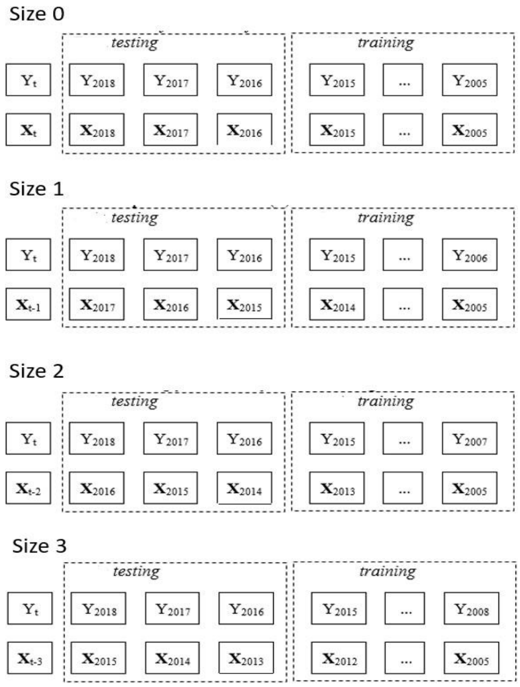


Figure 2. Concept of Drift for Training and Testing Dataset

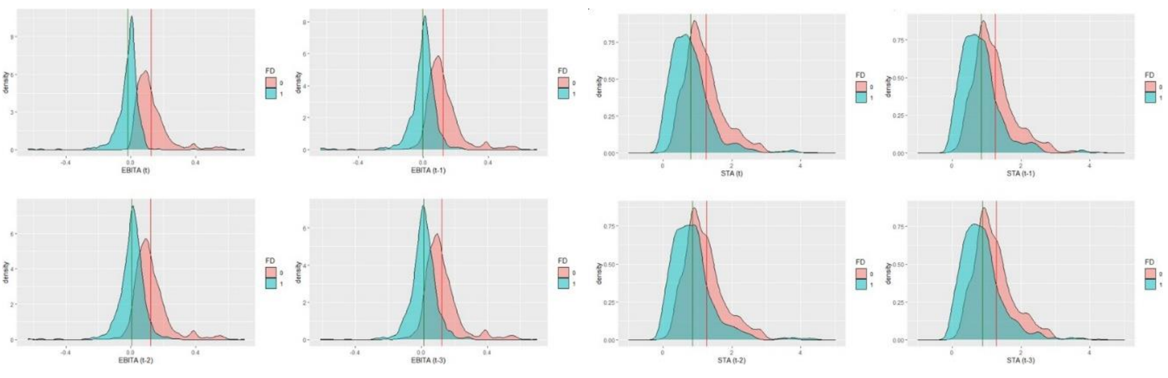
The steps of research in this work can be summarized as follows.

1. Assign the response variable (Y) label based on schemes one and two scenarios. The dataset for schemes one and two will be analyzed separately.
2. Determine the window size (0, 1, 2, and 3 years) and separate them as different datasets.
3. Split the data into training and testing datasets for each window.
4. Apply the proposed and benchmark model, i.e., SVM, logistic regression, and GEVR with and without feature selection approach, for each window size and each scheme.
5. Evaluate the performance of each approach using accuracy and AUC in the training and testing dataset.
6. Make interpretations, explanations, and conclusions.

All the research steps in this work were conducted using open-source *R* software.

4. RESULTS

The distribution of each financial ratio is compared for the class of distressed and non-distressed companies. The distribution is calculated using Kernel Density Estimator (KDE). See Haerdle et al. (2004) for more details. The KDE is a free distribution or nonparametric that enables the data to speak by itself. This data-driven approach is more realistic than the presumed financial ratios following a specific distribution. Fig 3 displays the kernel density of Earnings Before Income Tax to Asset (EBITA) and Sales to Total Asset (STA) financial ratios. The green-filled area is the distribution of each financial ratio for non-distressed companies, whereas the red is for the financially distressed companies. The vertical lines at each area are the mean values for each financial ratio in each group.



(a) (b)

Figure 3. Kernel density of EBITA (a) and STA (b) of distressed and non-distressed companies with window size zero (top-left), one year (top-right), two years (bottom-left), and three years (bottom-right) for first scheme setting

Those three financial ratios displayed in Fig 3 shows that the mean of each ratio of financially healthy companies is always higher than those of distressed companies. Most of each financial ratio from health companies is also higher than those of distressed companies. We made the same graphs for all financial ratios, but they are not shown here because of the limited space. We select these two ratios, among others, since they intuitively explain that the contributing features to the label of output variables should have different patterns across the groups. The opposite condition, i.e., the mean of the ratio and the area are similar, indicates that the financial ratio is not a relevant feature contributing to the learning algorithm for predicting the label (distress/non-distress). Fig 4 of kernel density Working Capital to Sales Ratio (WCS) explains this concept, meaning that the WCS is irrelevant to predicting distress.

The identification through visual exploration using KDE, as in Fig 3 and Fig 4, will be confirmed at the end once the model is thoroughly trained. The significant features within the model have to compromise its KDE pattern. The difference in density over the response categories can easily explain the significance of the contributing features. It is not always as straightforward as this when many features are employed together within the model. Multicollinearity possibly causes the significant individual features to become insignificant when put together into a model with other features.

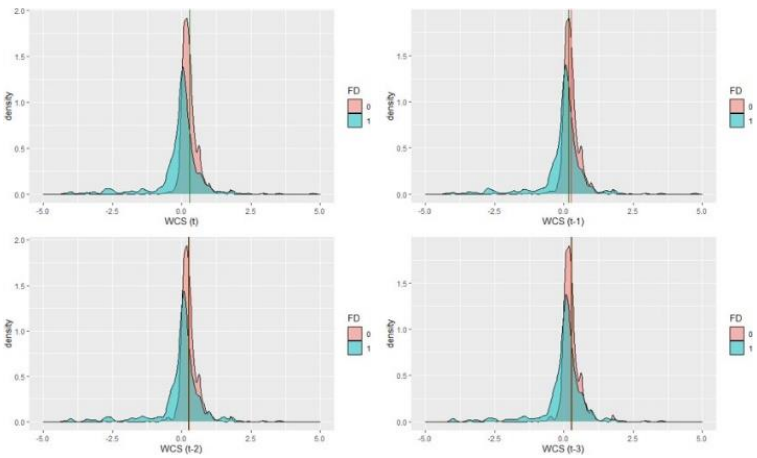


Figure 4. Kernel density of WCS of distressed and non-distressed companies with window size zero (top-left), one year (top-right), two years (bottom-left), and three years (bottom-right) for first scheme setting

Table 2. Performance Evaluation of The Model Without Feature Selection

		AUC		Accuracy	
		Training	Testing	Training	Testing
Scheme one					
SVM	0	0.9412	0.9200	0.9423	0.9195
	1	0.9368	0.8249	0.8483	0.8460
	2	0.8263	0.7855	0.8337	0.8000
	3	0.8347	0.5922	0.8276	0.6230
LOGIT	0	0.8949	0.8985	0.9329	0.9218
	1	0.7417	0.8209	0.8455	0.8621
	2	0.7085	0.7798	0.8360	0.8046
	3	0.6692	0.5545	0.8138	0.7264
GEVR	0	0.976	0.9708	0.9367	0.9241
	1	0.8947	0.928	0.8586	0.8667
	2	0.8721	0.8782	0.8414	0.8207
	3	0.8494	0.8288	0.8233	0.7954
Scheme two					
SVM	0	0.9342	0.8696	0.9436	0.8874
	1	0.8597	0.8106	0.891	0.8299
	2	0.9039	0.9042	0.8858	0.8092
	3	0.9179	0.9042	0.875	0.8092
LOGIT	0	0.8895	0.9241	0.9567	0.9241
	1	0.64	0.8345	0.8855	0.8345
	2	0.6024	0.8815	0.8743	0.8815
	3	0.5752	0.8092	0.8612	0.8092
GEVR	0	0.9587	0.9537	0.9298	0.9218
	1	0.8731	0.8769	0.891	0.8483
	2	0.8492	0.8251	0.8851	0.8276
	3	0.8265	0.7919	0.8698	0.8276

Table 2 summarizes the prediction performance of GEVR and the benchmark models without feature selection. We use AUC as the primary evaluation criterion as it is more appropriate for imbalanced data, and we still provide the accuracy measure as additional information. The larger the window size of the drift, the smaller the AUC values in the testing dataset. The GEVR outperforms the benchmark models using the AUC criterion when the drift's window size is zero. The superiority of GEVR is still happening for one year of concept drift in the testing dataset. On the window size is three and scheme one, the AUC values in the testing dataset produced by SVM and logistic regression drop significantly below 0.6, but the one resulting from GEVR still higher than 0.8. It also seems there is no over or underfitting of AUC yielded from GEVR.

Due to the limited space, the empirical results are not summarized in tables since there will be many tables that make the space run out. The overall comparison focused on the testing dataset is displayed in Fig 5. The solid bullets represent scheme one of the financial distress definition, whereas the light-filled bullets are for scheme two. For all drift's window size, the zero until three years past financial ratios affecting the financial distress condition, the GEVR with or without backward elimination is always in the top two best prediction performance at scheme one of financial distress definition. The GEVR always performs better for scheme one than for scheme two, and this does not still hold for other models. Logistic regression always performs better when applied to scheme two for all drift's window sizes. The SVM behaves in between GEVR and Logistic regression. Both penalizations using Lasso and Elastic-net make the SVM and Logistic regression perform better when applied to scheme one.

The final GEVR model, with or without backward elimination, always results in these significant financial ratios: EBITA, STA, ITR, NPM, OPM, DAR, and ETD for the first scheme of financial distress definition and drift's window size zero. These significant features in the proposed classification model confirm the KDE characteristics presented earlier. One macroeconomy variable, the central bank interest rate, is selected as a significant feature. Another financial ratio, the Quick Ratio (QR), enters the GEVR model as written below. For scheme one, the GEVR model for each drift's window size is written as follows.

Size = 1 (to predict the response label one year ahead using current period predictors)

$$gevit(\pi(\mathbf{x}_i)) = 1.5041 - 24.0229\ \text{EBITA}_i - 0.323\ \text{STA}_i + 0.5857\ \text{OPM}_i + 0.0300\ \text{DER}_i + 1.4987\ \text{DAR}_i + 0.0485\ \text{QR}_i - 0.2820\ \text{RGDPG} + 0.0630\ \text{Inflation}$$

Size = 2 (to predict the response label two years ahead using current period predictors)

$$gevit(\pi(\mathbf{x}_i)) = -6.273 - 4.248\ \text{EBITA}_i - 0.1920\ \text{STA}_i - 0.8804\ \text{NPM}_i - 3.640\ \text{OPM}_i + 0.5861\ \text{DAR}_i + 0.0554\ \text{QR}_i - 0.7346\ \text{WCA}_i - 0.0000787\ \text{USD/IDR}$$

Size = 3 (to predict the response label three years ahead using current period predictors)

$$gevit(\pi(\mathbf{x}_i)) = 0.1168 - 4.8853\ \text{EBITA}_i - 2.2349\ \text{OPM}_i + 0.3381\ \text{DAR}_i - 0.6507\ \text{WCA}_i$$

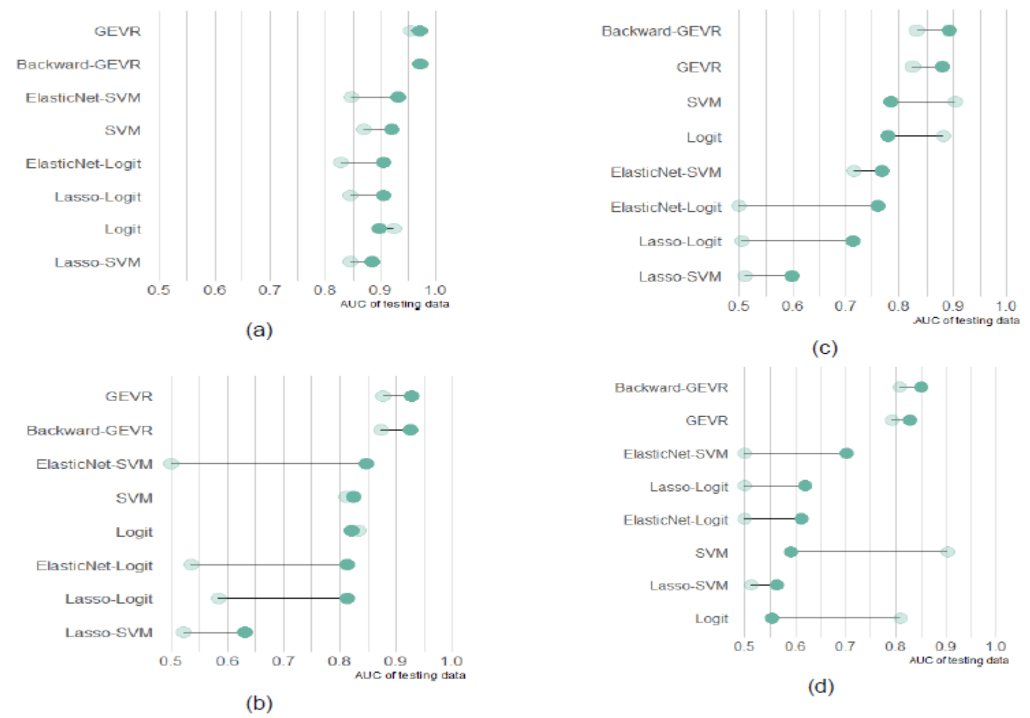


Figure 4. AUC values of the testing dataset for concept drift with window size zero (a), one year (b), two years (c), and three years (d). The dark solid bullet represents the AUC values at scheme one of the financial distress definition.

The predictor variables used in this work are mostly the same as those used in Haerdle and Prastyo (2014), Prastyo et al. (2017), and Prastyo et al. (2018) but with a different approach. Even if they can not be compared directly, they can still compare from a specific point of view, i.e., the sign of coefficient estimator obtained from the fitted model. The proposed approach in this work gives less counterintuitive signs from the financial point of view. The counterintuitive sign is inevitable because the financial ratios as predictor variables suffer the multicollinearity issue. Drop the collinearity predictors may raise other problems as it will be difficult to explain, particularly if the interpretation becomes one of the research's main goals.

5. CONCLUSION

The dataset with a minority class of financially distressed groups considered rare events have been analyzed using the extreme value-based classification model so-called GEVR. The GEVR, with or without feature selection, outperforms the benchmark methods, including SVMs and logistic regression with or without embedded feature selection using the Lasso and Elastic-net. The excellent performance of GEVR applies in all window sizes of concept drift and two schemes of financial distress definition. Also, the significant financial ratios selected in the final GEVR model confirm its characteristics pattern explored using kernel density. The proposed model generally applies to any dataset with imbalanced binary labels. The considerable issues for future research will be, among other possibilities, how to treat the two responses, i.e., the ICR and ROA, as multiresponse modeling. Developing GEVR for multiclass classification is also an exciting research topic, and its' applicability in many areas is no doubt.

ACKNOWLEDGMENT

This research was supported by *Kemendikbudristek* & BRIN scheme Fundamental research, number of contract 008/E5/PG.02.00.PT/2022 & 1506/PKS/ITS/2022.

REFERENCES

Bradley, P.S. and Mangasarian, O.L. (1998). Feature Selection Via Concave Minimization and Support Vector Machines, *Proceeding of the Fifteenth International Conference on Machine Learning* (ICML).

Calabrese, R. and Osmetti, S.A. (2013). Modeling Small and Medium Enterprise Loan Defaults As Rare Even: The Generalized Extreme Value Regression Model, *Journal of Applied Statistics*, Vol. 40, No. 6, pp. 1172-1188.

Calabrese, R and Giudici, P. (2015). Estimating Bank Default with Generalized Extreme Value Regression Models, *Journal of the Operational Research Society*, Vol. 66, pp. 1783-1792.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Path for Generalized Linear Models Via Coordinate Descent, *Journal of Statistical Software*, Vol. 33, No. 1, pp. 1-22.

Haerdle, W.K., Prastyo, D. D., and Hafner, C. M. (2014). Support Vector Machines with Evolutionary Model Selection for Default Prediction," in *The Oxford Handbook of*

Applied Nonparametric and Semiparametric Econometrics, J. S. Racine, L. Su, and A. Ullah (Eds). New York: Oxford University Press.

Haerdle, W.K. and Prastyo, D. D. (2014). "Embedded Predictor Selection for Default Risk Calculation: A Southeast Asian Industry Study," in *Handbook of Asian Finance, Vol. 1: Financial Market and Sovereign Wealth Funds*, D. L. K. Chuen and G. N. Gregoriou (Eds). San Diego. Academic Press, 2014, pp 131-148.

Haerdle, W.K., Werwatz, A., Mueller M., and Sperlich S. (2004). "Nonparametric Density Estimation," In *Nonparametric and Semiparametric Models*. Berlin: Springer. pp.39-83.

Hanley, J.A. and McNeil, B. J. (1982). The Meaning and Use of The Area Under A Receiver Operating Characteristic (ROC) Curve, *Radiology*, Vol. 143, No. 1, pp. 29-36.

Kowarik, A. and Templ, M. (2016). Imputation with the R package VIM, *Journal of Statistical Software*, Vol. 74, No. 7, pp. 1-16.

Natasha, A., Prastyo, D. D., and Suhartono. (2019). Credit Scoring to Classify Consumer Loan using Machine Learning, *AIP Conference Proceedings* 2194, 020070.

Prasetya, J. and Abdurakhman, A. (2022). Comparison of SMOTE Random Forest and SMOTE k-Nearest Neighbors Classification Analysis on Imbalanced Data, Media Statistika, Vol. 15, No. 2, pp. 198-208.

Prastyo, D.D., Miranti, T., and Iriawan, N. (2017). Survival Analysis of Companies' Delisting Time in Indonesian Stock Exchange Using Bayesian Multiple-Period Logit Approach, *Malaysian Journal of Fundamental and Applied Sciences*, Vol. 13, No. 4-1, pp. 425-429.

Prastyo, D.D., Rucy, Y. N., Sigalingging, A. D. C., Suhartono, and Fam, S. (2018). Micro and Macro Determinants of Delisting and Liquidity in Indonesian Stock Market: A Time-dependent Covariate of Survival Cox Approach, *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, Vol. 34, pp. 73-81.

Randhawa, K., Loo, C.K., Seera, M., Lim, C.P., and Nandi, A.K. (2018). Credit Card Fraud Detection Using Adaboost and Majority Voting, *IEEE Access*, Vol. 6, pp. 14277-14284.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via The Lasso, *Journal of the Royal Statistical Society, Series B*, Vol. 58, pp 267-288.

Wang, C., Han, D., Liu, Q., and Luo, S. (2019). A Deep Learning Approach for Credit Scoring of Peer-To-Peer Lending Using Attention Mechanism LSTM, *IEEE Access*, Vol. 7, pp. 2161-2168.

Zou, H. and Hastie, T. (2005). Regularization and Variable Selection Via The Elastic Net, *Journal of the Royal Statistical Society, Series B*, Vol. 67, No. 2. pp 301-320.

Zhu, J., Rosset, S., Hastie, T., and Tibshirani, R. (2004). *1*-norm Support Vector Machine, *Proceeding of Advances in Neural Information Processing System 16*, 2004.