A Comparative Analysis of Macroeconomic Indicators in Optimising Credit Risk Prediction

Evelyn Sierra
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025231091@student.its.ac.id

Erick Delenia
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025231057@student.its.ac.id

Eric Saputra Lays
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025231020@student.its.ac.id

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Agus Tri Haryono
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025231020@student.its.ac.id

2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA) | 979-8-3503-5161-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICTIIA61827.2024.10761138

***Abstract*— Credit risk modelling plays a role in financial institutions' evaluation of the creditworthiness of borrowers and in managing lending risks effectively. Feature selection is critical in developing robust and interpretable credit risk models. This paper presents a comparative analysis of various macroeconomic indicators applied to linear regression analysis for credit risk modelling. Specifically, this study compares three feature selection techniques: clustering feature, feature combination and correlation. This comparative study aims to identify the most compelling feature selection strategy regarding model performance, interpretability, and computational efficiency. This study employs a real-world dataset comprising various macroeconomic indicators in Indonesia and a financial institute's default scores to train and evaluate the linear regression models. Experimental results demonstrate that the clustering feature outperforms feature combinations and correlation features in optimizing credit risk modelling by achieving higher predictive accuracy with fewer features and improved interpretability.**

Keywords—Clustering, Credit Risk, Feature Selection, Machine Learning, Macroeconomic

I. INTRODUCTION

Predicting default rates for individuals requires information about the financial sources of prospective borrowers, such as basic salary, marital status, and total assets [1]. However, predicting retail and corporate banking default rates entails different data requirements. According to Pernyataan Standar Akuntansi Keuangan (PSAK) 71 or International Financial Reporting Standards (IFRS) 9 [2] regulations governing default rate predictions in the banking business sectors, both retail and corporate, macroeconomic data serves as independent data alongside historical default rates of customers within a specific banking segment For instance, in banking segments offering credit cards, there are instances where some customers default on instalment payments because credit fraud [3][4], necessitating financial reserves from the bank to ensure compensation for these customer shortfalls [5]. That problem underscores the necessity for banks to predict future default rates. Consequently, the probability of default assumes utmost importance as a critical risk parameter in credit institutions, notably banks, playing a pivotal role in the comprehensive analysis and management of credit risk.

Estimating the probability of default (PD) from a loan applicant often uses classification techniques [6], such as linear regression. One of the independent indicators that can be used to model the probability of default is the macroeconomy. Quagliariello [7] found that a decline in macroeconomic performance can increase the risk of default. According to Caro [6], macroeconomic indicators that can affect the performance of Microfinance Institutions are Gross Domestic Product (GDP) growth, unemployment rate, inflation, and foreign investment [9].

However, these are not the only economic indicators that can be observed; macroeconomic indicators, such as imports, exports, revenue, GNI, and others, are also observed. Referring to the vast array of macroeconomic indicators available, the need for feature selection becomes apparent. Feature selection [10]-[14] is essential to streamline the modelling process by identifying and prioritizing the most influential and relevant indicators. Analysts can build more robust and accurate models that effectively capture the nuances of economic trends and relationships in the data by strategically selecting features based on their predictive power and significance.

This study delves into a comprehensive comparative analysis of different feature selection strategies explicitly applied to linear regression analysis for credit risk modelling. The primary aim of this study is to scrutinize and compare three distinct feature selection techniques: clustering feature, feature combination, and correlation. This study evaluates these strategies to pinpoint the most effective approach to enhance model performance, ensure interpretability, and streamline computational efficiency.

The rest of the paper is organized as follows: section 2 contains a comprehensive review of previous related research studies about the probability of default, feature clustering and combination. After that, in section 3, the proposed methodology will be described from the dataset using Indonesia macroeconomic indicators and probability of default data from financial institutions. In section 4, experimental results are presented, featuring performance metrics and comparisons with existing methods to evaluate the effectiveness of the proposed approach. Finally, the discussion section offers an interpretation of the results, highlighting implications, limitations, and avenues for future research.

II. Literature Survey

A. Probability of Default

Probability of Default (PD) [15] is a concept in finance used to assess the likelihood of a borrower defaulting on their financial obligations within a specific timeframe. It serves as a metric for evaluating credit risk, enabling lenders and investors to make informed decisions regarding loan approvals, pricing, and portfolio management. Calculating the PD involves employing statistical models and data analytics techniques to estimate the likelihood of default events. Several commonly used methods include machine learning and survival analysis. [16] demonstrated the Probability of Default prediction using the Quantitative Input Influence (QII) [17] method and machine learning. These methods are well-suited for analyzing financial models, allowing insight from the model by clustering observations with similar input influences. However, this method may be complex, require a certain level of expertise to interpret, and is limited to only financial context. Another study by Bruche, M. and González-Aguado, C. [18] proposes an econometric model in which an unobserved Markov chain drives this joint time-variation in default rates and recovery rate distributions. These models leverage historical data on borrower characteristics, loan performance, and macroeconomic indicators to accurately quantify the probability of default. Credit rating agencies and financial institutions also develop proprietary models tailored to their specific risk assessment requirements. By incorporating PD as a dependent indicator, linear regression models can assess the impact of factors using macroeconomic indicators. Furthermore, linear regression analysis facilitates the development of predictive models for estimating the PD and enhancing credit scoring systems and risk management strategies.

B. Macroeconomic Indicator Clustering and Combination

In credit risk prediction, feature combination selects and merges multiple input indicators (features) to enhance the model's predictive power. This approach captures complex relationships and interactions between macroeconomic indicators and their impact on credit risk. Clustering [7] techniques can help identify patterns or groupings within the dataset based on the similarities or dissimilarities between observations. In credit risk prediction, clustering can segment borrowers or financial institutions into homogeneous groups based on their macroeconomic characteristics [4].

Kou, G., Peng, Y., and Wang, G. [19] explored the application of feature clustering techniques such as hierarchical clustering and k-means clustering in grouping related financial indicators for credit risk prediction. Their study demonstrated that clustering similar features improved model interpretability, and MCDM methods were used to rank the clusters. Furthermore, that study found that no algorithm can perform best on all measurements for any data set. Utilizing more than one single performance measure to evaluate clustering algorithms is necessary. In similar research, [20] investigated the efficacy of feature combination methods such as principal component analysis (PCA). Their research revealed that PCA-based feature combination facilitated dimensionality reduction while preserving the most informative aspects of the data, thereby improving model generalization and robustness.

Moreover, studies by Chen, P., & Wu, C. [21] on the macroeconomic frailty model include sectoral factors that capture default correlations among firms in a similar business. Their findings suggest that incorporating sectoral frailties into default intensity models improves the estimation of default correlations among firms operating in similar business sectors. The previous study finds that the default intensity model with sectoral frailties fits the data very well by using data from Japanese firms from 1992 to 2010. That indicates the model's generality in explaining default dynamics across different economies and suggests its applicability to diverse datasets across countries.

C. Macroeconomic Indicator for Default Predictions

Macroeconomic indicators are pivotal in optimizing credit risk prediction models, especially in the context of corporate credit and banking institutions. This section conducts a comparative analysis of macroeconomic indicators employed in forecasting the probability of default within these sectors. By examining existing literature, this review aims to elucidate the efficacy of different indicators and their implications for credit risk management in corporate lending and banking.

Research by [22] focused on approaches for feature clustering and combination specifically to credit risk modelling in emerging markets. Their findings highlighted that macroeconomic indicators, particularly interest rates and inflation rates, have a significant long-term impact on third-party funds in Islamic banks in Indonesia. However, economic growth was not related to third-party funds in Islamic banks in the country. Overall, the study highlights the importance of considering macroeconomic factors, especially interest rates and inflation, when analyzing the dynamics of third-party funds in Islamic banks in Indonesia, both in the long and short term. Similarly, Carvalho, Paulo V., José D. Curto, and Rodrigo Primor [23] examined the relationship between macroeconomic indicators and the probability of default for firms in the Eurozone. The study proposed and estimated several models that control for firm-specific information to reveal related macroeconomic factors that impact credit default prediction. Their findings reveal that GDP growth harms the probability of default, indicating that economic growth generally reduces the likelihood of firms defaulting on their credit obligations. However, the researchers also uncover compelling evidence of asymmetries within the Eurozone regarding the beneficial effects of GDP growth on credit risk. The study observe that the reduction in the probability of default due to economic growth is more pronounced in economies that are more exposed to financial stress conditions.

Furthermore, recent studies by Nigmonov, A., Shams, S., and Alam, K. [24] represent a significant contribution to the emerging field of online peer-to-peer (P2P) lending by examining the influence of economic indicators, specifically interest rates and inflation, on the probability of default. Utilizing an extensive dataset spanning 11 years and focusing on delinquencies as a measure of default risk, it stands out as one of the first empirical investigations of P2P lending markets. Findings from the study reveal that higher interest rates and inflation are correlated with an increased likelihood of default in the P2P lending market. The study suggests that economic conditions, including fluctuations in interest rates and inflationary pressures, significantly shape borrower default 2ehavior within online lending platforms.

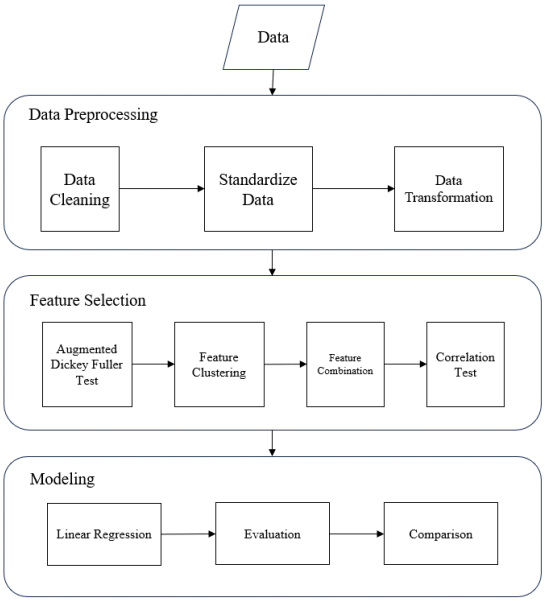

Fig 1. Flowchart Methodology

In summary, the literature presents various macroeconomic indicators utilized to predict default probabilities within corporate credit and banking institutions. However, determining the optimal combination of indicators necessitates careful consideration of sector-specific risk factors and market dynamics, ultimately informing more robust credit risk management practices.

III. Proposed Method

The proposed study involves datasets from banking institutions and macroeconomic data from 1960 to 2022, covering 52 macroeconomic indicators in Indonesia. Table I provides a glimpse of the macroeconomic data for some countries. For example, five key indicators: average lending rate, bank interest rate, unemployment rate, working capital credit rate, and the number of countries included in each month (Number). As shown in Fig. 1, the methodology starts with data collection, followed by data preprocessing, then feature selection and finally, modelling, evaluation and comparison of the resulting models.

A. Preprocessing Data

The data to be used comprises macroeconomic data as independent indicators in a linear regression equation and the percentage of default as the dependent indicator. These datasets will be merged based on dates and undergo a data preprocessing stage. Below are examples of macroeconomic data and default percentage data.

Credit default data are obtained from one of the financial institutions in Indonesia and are available privately. Every financial institution has its business segment, and each segment reports how many defaults occur. Hence, the credit default data is always at the end of the month.

During the data preprocessing stage, the macroeconomic data will be cleaned to prevent missing dates and empty data points. Empty data will be addressed using interpolation. The cleaned macroeconomic data will then undergo sample standardization for data normalization. Subsequently, the macroeconomic data will undergo data transformation to obtain more features. The transformations employed include 12-month lagging transformation, logarithmic transformation, and 12-month differencing. This stage focuses on obtaining cleaned macroeconomic data with various transformations applied. Following this, the default percentage data will be processed. The cleaning process for the default percentage data involves checking for missing dates and applying a logit transformation. Subsequently, the two cleaned datasets will be merged based on dates. The output will yield a dataset with macroeconomic data as independent indicators and default percentage data as the dependent indicator.

B. Feature Selection

The feature selection process commences with macroeconomic data and transformed default percentage data. Initially, the Augmented Dickey-Fuller (ADF) Test is conducted to examine the stationarity of the macroeconomic features. The macroeconomic features undergo three types of ADF methods: ADF for no constant and no trend, constant and no trend, and constant and trend. Any macroeconomic features that are stationary from any ADF Test methods are retained for the subsequent stage. Conversely, if not, those features are eliminated.

Subsequently, the process transitions to the stage of feature clustering and combination. In this study experimentation, the number of independent indicators is restricted to three. Hence, indicator grouping is essential to mitigate high multicollinearity levels among the three independent indicators within one linear regression equation. Indicator grouping is executed based on the covariances of the macroeconomic data, and from each clustering group, one indicator is selected. After obtaining several linear regression equations, results are evaluated and compared to ascertain the best model. This evaluation may encompass R-squared, Mean Absolute Error (MAE), or Root Mean Square Error (RMSE) to assess the model's performance in predicting default percentages.

The following feature selection is clustering, which will begin by performing PCA on the indicators at this stage. Principal Component Analysis (PCA) is a dimension-

TABLE I. Macroeconomc data Statistics

	AVG_LEND_RATE	BI_RATE	….	UNEMP_EAP	WC_LOAN_RATE
Count	139	140	….	27	139
Mean	11.61	5.59	….	0.04	111.08
Std	1.31	1.38	….	0.02	1.32
Min	9.01	3.50	….	0.03	8.40
25%	10.71	4.43	….	0.04	10.31
50%	12.06	5.75	….	0.04	11.44
75%	12.76	6.75	….	0.04	12.26
max	14.15	7.75	….	0.04	12.84

reduction technique that converts complex data into fewer dimensions while retaining relevant information. PCA is used to reduce data complexity before applying the K-means clustering algorithm. K-means clustering is a technique used to group data into k groups based on patterns in the data. After K-means clustering, cluster evaluation is performed using the Silhouette value to determine the optimal number of clusters, where a higher Silhouette value indicates optimum clusters. After all indicators have been clustered, modelling is carried out under the condition that all selected independent indicators are in different clusters.

C. Modelling

In the modelling stage using OLS Regression, the resulting equation will be tested using the Shapiro Wilk, Durbin Watson, and Multicollinearity tests. The Shapiro-Wilk test assesses the normality of residuals or model errors. The tests are important because the OLS model's success relies on the assumption that the errors follow a normal distribution. By testing the normality of residuals, this study can ensure that this assumption is met so that the model results can be interpreted accurately. Next, the Durbin-Watson test is used to detect the presence of autocorrelation in the model residuals. Autocorrelation occurs when there are patterns in the model residuals, which can disrupt the validity of regression results and lead to biased estimates. This test is essential to ensure no autocorrelation in the model residuals, thus making the regression results reliable. Lastly, the Multicollinearity test is conducted to evaluate the level of correlation between independent indicators in the model. Multicollinearity can cause interpretation problems in regression models, making independent indicators appear insignificant or have inconsistent coefficients. This study can identify and address this issue by testing for multicollinearity, making the regression results more reliable and interpretable.

By conducting these three tests, this study can ensure that the resulting regression model is valid, consistent, and interpretable, thus enabling accurate predictions and valuable insights in the context of analysis.

IV. Results and Discussion

The feature selection process begins with selecting macroeconomic features that pass the Augmented Dickey-Fuller (ADF) Test. The results of the augmented Dickey-Fuller test eliminated 360 out of a total of 360 macroeconomic transformations, leaving only 283 types of transformations. Subsequently, these features undergo preprocessing to obtain macroeconomic features that exhibit a correlation above 0.5 with the percentage of default as the target data.

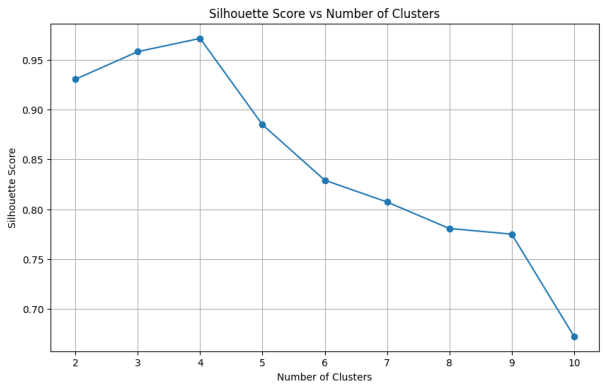

Fig. 2 Silhouette Score vs Number of Clusters

A. Feature Clustering

The 283 features or macroeconomic indicators that have passed the correlation and ADF tests (including feature transformation) are then grouped into 10 clusters and evaluated using the silhouette value. The highest silhouette value results from a cluster with four clusters, as shown in Figure 2. The cluster results are shown in Figure 3. From the plot results, it can be seen that the distance between clusters is separated from each other, with the distance of points in the cluster being close to each other, which indicates that the clustering results are good so that they can proceed to the modelling stage.

After all indicators have been clustered, modelling is done by combining indicators with different clusters. The combination resulted in 1353 models with different independent variable structures. Table III. Shows the five models with the highest r-squared score. It can be seen that the features or indicators that consistently produce the best models are 'PRIVATE_CONS', 'IP_Lag_1' and 'PRIVATE_CONS'.

From Table III. There are five models with the highest r-squared value followed by the lower MSE and RMSE, indicating that the regression model has good performance, provides accurate predictions and explains most of the variance in the data. However, this is not necessarily the main consideration in model selection, and there needs to be other evaluation parameters, such as statistical tests. If 1 of the 5 sample models in Table III is taken, namely the model with the highest r-squared value of 0.0986, then it shows a relatively low level of prediction error. However, judging from the statistical test results of the Shapiro-Wilk test, a p-value of 0.0559 is obtained, which indicates that there is doubt about the normality of the residual data. In addition, high VIF values are seen in the PRIVATE_CONS indicator (VIF = 1054.987002), DPK_Lag_1 (VIF = 140.3), GOV_EXT_DEBT_Lag_1 (VIF = 448), and IP_Lag_1 (VIF = 426.1), indicating significant multicollinearity. The low Durbin-Watson statistic of 0.406 also indicates the presence of positive autocorrelation in the residuals, which may reduce the reliability of the regression model's predictive results. Therefore, it is important to review the model not only from the r-squared matrix but also through statistical tests to stabilize the resulting model.

B. Feature Combination

The selected macroeconomic features will undergo feature combination based on the smallest VIF (Variance Inflation Factor), with an equation only comprising three types of macroeconomic features. The analysis reveals several combinations of lagged macroeconomic indicators that exhibit characteristics for predicting the default percentage. These combinations demonstrate notably high R-squared values, indicating their ability to explain a substantial portion of the variance in the target indicator. The selected features capture essential aspects of the economic landscape that influence default rates. Furthermore, the low VIF values associated with these combinations suggest minimal

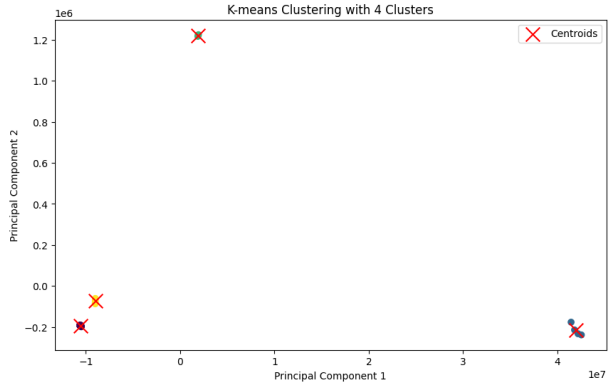

Fig. 3 K Means Clustering Result

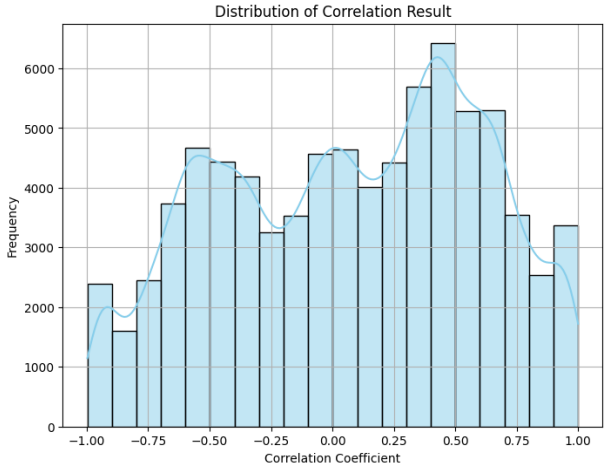

Fig. 4 Distribution of Correlation Result

multicollinearity among the predictors, enhancing the reliability of the regression models. That is crucial for ensuring the estimated coefficients retain their interpretability and statistical significance. However, it is essential to note the normality of residuals, as indicated by Shapiro-Wilk p-values.

From Table IV, the examined models exhibit high R-squared values ranging from 0.9164 to 0.9197. That signifies a robust positive association between the independent and dependent indicators. Furthermore, the exceptionally low Mean Squared Error (MSE) values (between 0.000003 and 0.000004) and correspondingly low Root Mean Squared Error (RMSE) values indicate that the models fit the data remarkably well, with minimal prediction errors on average. While individual feature importance cannot be definitively determined from this Table, the VIF column allows us to assess multicollinearity. VIF values exceeding 5 for all combinations suggest a potential issue with multicollinearity. That implies that some features might be highly correlated, leading to inflated standard errors and making it challenging to isolate the actual effect of each feature on the target indicator. Among the 52 macroeconomic indicators, the most frequently used to obtain the highest R-squared result is the Government External Debt (code : GOV_EXT_DEBT) macroeconomic indicator, with the transformation of lagging one and two months, resulting in the highest accuracy.

While a detailed assessment of these p-values is necessary to determine the adequacy of the regression models, the overall findings suggest promising avenues for further exploration and validation. Further analysis, including robustness checks and out-of-sample testing, is warranted to ascertain the predictive power of these models in real-world scenarios and to inform more informed decision-making processes in financial contexts.

V. Conclusion

The data used consists of macroeconomic data in Indonesia from 1960 to 2022, with 52 types of

TABLE III. CLUSTERING RESULTS

Features				R-Squared	MSE	RMSE	Shapiro-Wilk	VIF				Durbin-Watson
1	2	3	4					1	2	3	4	
'PRIVATE_CONS'	'DPK_Lag_1'	'GOV_EXT_DEBT_Lag_1'	'IP_Lag_1'	0.9086	0.01057	0.10284	0.0559	1054.98	140.31	448.05	426.15	0.4056
'PRIVATE_CONS'	'DPK_Lag_2'	'GOV_EXT_DEBT_Lag_1'	'IP_Lag_1'	0.9085	0.01058	0.10288	0.0531	1056.05	140.97	447.82	426.60	0.4117
'PRIVATE_CONS'	'DPK_Lag_3'	'GOV_EXT_DEBT_Lag_1'	'IP_Lag_1'	0.9084	0.01059	0.10293	0.0514	1065.93	145.64	448.73	433.31	0.411
'DPK_Lag_1'	'GOV_EXT_DEBT_Lag_1'	'IP_Lag_1'	'PRIVATE_CONS_Lag_1'	0.9083	0.01060	0.10299	0.0811	136.91	449.62	320.81	815.11	0.417
'DPK_Lag_2'	'GOV_EXT_DEBT_Lag_1'	'IP_Lag_1'	'PRIVATE_CONS_Lag_1'	0.9082	0.01061	0.10304	0.0755	137.56	449.06	321.23	815.99	0.4235

TABLE IV. COMBINATION RESULTS

Features			R-Squared	MSE	RMSE	Shapiro-Wilk	VIF			Durbin-Watson
1	2	3					1	2	3	
PRIVATE_CONS_Lag2	PRIVATE_CONS_Lag1M	GOV_EXT_DEBT_Lag2M	0.9197	0.000003	0.001837	0.5601	569.54	714.76	1251.41	0.486
PAT_BLEND_Lag1M	PRIVATE_CONS_Lag1M	GOV_EXT_DEBT_Lag2M	0.9178	0.000003	0.001859	0.4836	608.30	714.76	1251.41	0.374
PRIVATE_CONS_Lag1M	INTL_RESERVE_MTH_Lag3M	GOV_EXT_DEBT_Lag2M	0.9166	0.000004	0.001872	0.2448	714.76	1046.44	1251.41	0.244
PRIVATE_CONS_Lag1M	PAT_BLEND_Lag2M	GOV_EXT_DEBT_Lag2M	0.9165	0.000004	0.001874	0.3575	714.76	857.09	1251.41	0.353
MIN_WAGE_Lag2M	PRIVATE_CONS_Lag1M	GOV_EXT_DEBT_Lag2M	0.9164	0.000004	0.001874	0.2284	552.30	714.76	1251.41	0.333

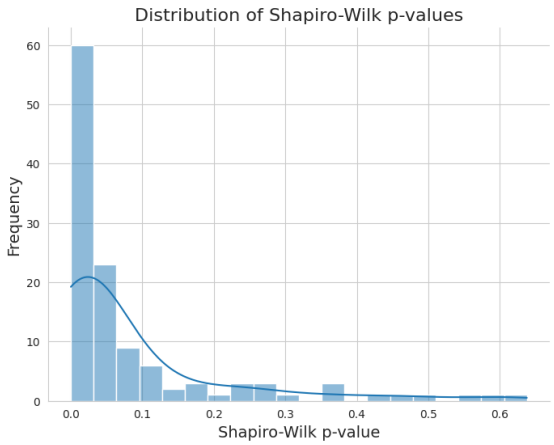

Fig. 5 Distribution of Shapiro Wilk p-values

macroeconomic indicators as independent indicators. Subsequently, the default percentage data from the Indonesian banking industry was collected from 2012 to 2022. The methodology begins with data collection and preprocessing, including data transformation and the Augmented Dickey-Fuller (ADF) Test to assess the stationarity of the macroeconomic data and the default percentage data. The sorted data undergoes feature combination and clustering to derive linear regression equations. The results indicate that feature combinations yield higher accuracy. Although high R-squared suggests a good fit, multicollinearity can impact the reliability of coefficient estimates, necessitating further investigation. Hence, it is needed for future works in obtaining good multicollinearity before modeling.

ACKNOWLEDGMENT

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Dana Departemen Program, Penelitian Keilmuan, Penelitian Flagship, Penelitian Kolaborasi Pusat and under the Project Scheme of the Publication Writing and Intellectual Property Rights (IPR) Incentive (Penulisan Publikasi dan Hak Kekayaan Intelektual/PPHKI) Program 2024.

REFERENCES

[1] X. Zhu, Q. Chu, X. Song, P. Hu and L. Peng, "Explainable prediction of loan default based on machine learning models," *Data Science and Management,* vol. 6, no. 3, pp. 123-133, 2023.

[2] Schutte, W. Daniel, T. Verster, D. Doody, H. Raubenheimer, P. J. Coetzee and D. McMillan, "A proposed benchmark model using a modularised approach to calculate IFRS 9 expected credit loss," *Cogent Economics and Finance,* vol. 8, p. 1735681, 2020.

[3] S. Huda, T. Ahmad, R. Sarno and H.A. Santoso, " Identification of process-based fraud patterns in credit application," in *2014 2nd International Conference on Information and Communication Technology (ICoICT)*, Indonesia, 2014.

[4] S. Huda, S. Sarno and T. Ahmad, " Fuzzy MADM approach for rating of process-based fraud," *Journal of ICT Research and Applications,* vol. 9, no. 2, pp. 111-128, 2015.

[5] S. Kerr and J. J. Canals-Cerda, "Forecasting Credit Card Portfolio Losses in the Great Recession: A Study in Model Risk," *Journal of Credit Risk,* pp. 14-10, 2015.

[6] A. A. Peresetsky, A. A. Karminsky and S. V. Golovan, "Probability of default models of Russian banks," *Economic Change and Restructuring,* vol. 44, pp. 297-334, 2011.

[7] M. Quagliariello, "Banks' riskiness over the business cycle: a panel analysis on Italian intermediaries," *Applied Financial Economics,* vol. 17, no. 2, pp. 119-1138, 2006.

[8] E. J. Caro, "Effects of Macroeconomic Factors in the Performance of Micro Finance Institutions in Ecuador," *International Journal of Economics and Financial Issues,* vol. 7, no. 5, pp. 547-551, 2017.

[9] K. Xing, D. Luo and L. Liu, "Macroeconomic conditions, corporate default, and default clustering," *Economic Modelling,* p. 106079, 2023.

[10] E. Tobback, D. Martens, T. V. Gestel and B. Baesens, "Forecasting Loss Given Default models: impact of account characteristics and the macroeconomic state," *Journal of the Operational Research Society* , vol. 65, pp. 376-392, 2014.

[11] S. J. Koopman, A. Lucas and B. Schwaab, "Modeling frailty-correlated defaults using many macroeconomic covariates," *Journal of Econometrics,* vol. 162, no. 2, pp. 312-325, 2011.

[12] S.I. Sabilla, S. Sarno and K. Triyana, "Optimizing threshold using pearson correlation for selecting features of electronic nose signals," *Int. J. Intell. Eng. Syst,* vol. 12, no. 6, pp. 81-90, 2019.

[13] D.R. Wijaya, S. Sarno and E. Zulaika, "Sensor array optimization for mobile electronic nose: Wavelet transform and filter based feature selection approach," *International Review on Computers and Software,* vol. 11, no. 8, pp. 659-671, 2016.

[14] F.A. Haq, R. Sarno, R.Abdillah, T.C. Amri, A.F. Septiyanto and K.R. Sungkono, "Feature Selection of Photoplethysmograph Data in Machine Learning," in *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Indonesia, 2023.

[15] F. Fiordelisi and D. S. Mare, "Probability of default and efficiency in cooperative banking," *Journal of International Financial Markets, Institutions and Money,* vol. 26, pp. 30-45, 22013.

[16] P. Bracke, A. Datta, C. Jung and S. Sen, "Machine Learning Explainability in Finance: An Application to Default Risk Analysis," Bank of England, England, 2019.

[17] A. Datta, S. Sen and Y. Zick, "Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems," in *2016 IEEE Symposium on Security and Privacy (SP)*, San Jose, 2016.

[18] M. Bruche and C. González-Aguado, "Recovery rates, default probabilities, and the credit cycle," *Journal of Banking & Finance,* vol. 34, no. 4, pp. 754-764, 2010.

[19] G. Kou, Y. Peng and G. Wang, "Evaluation of clustering algorithms for financial risk analysis using MCDM methods," *Information Sciences,* vol. 275, pp. 1-12, 2014.

[20] M.-J. Yang, H.-R. Zheng, H.-Y. Wang, S. Mcclean and N. Harris, "Combining feature ranking with PCA: An application to gait analysis," in *2010 International Conference on Machine Learning and Cybernetics*, Qingdao, 2010.

[21] P. Chen and C. Wu, "Default prediction with dynamic sectoral and macroeconomic frailties," *Journal Banking and Finance,* vol. 40, pp. 211-226, 2014.

[22] L. Y. H. a. T. L. S. Amaliawiati, "The impact of macroeconomics towards Islamic banking third party funds in Indonesia," *International Journal of Innovation, Creativity and Change,* vol. 6, no. 6, pp. 291-303, 2019.

[23] E. S. K. F. –. M. Maria Misankova, "Determination of Default Probability by Loss Given Default," *Procedia Economics and finance,* pp. 411-417, 2015.

[24] I.-C. Yeh and C.-h. Lien, "The comparisons of data mining techniques for the predictive," *Expert Systems with Applications,* p. 123607, 2017.