Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 17(2) 2024: 117-127
http://ejournal.undip.ac.id/index.php/media_statistika

EXPLORE THE DETERMINANTS OF CUSTOMERS TIME TO PAY HOUSE OWNERSHIP LOAN ON DATA WITH HIGH MULTICOLLINEARITY WITH PCA-COX REGRESSION

Rangga Ramadhan[1], Adfi Bio Fimba[1], Adji Achmad Rinaldo Fernandes[1], Solimun[1], Fachira Haneinanda Junianto[1], Devi Veda Amanda[1], Rauzan Sumara[2]
[1] Department of Statistics, Brawijaya University, Malang, Indonesia
[2] Faculty of Mathematics and Information Science, University of Technology, Warsaw, Poland

email: *rramadhan160702@student.ub.ac.id*

DOI: 10.14710/medstat.17.2.117-127

Article Info:
Received: 27 November 2023
Accepted: 30 December 2024
Available Online: 31 December 2024

Keywords:
Survival Analysis; Credit Collectability; PCA; Time to Pay Credit

Abstract: One of the models in survival analysis is the Cox proportional hazards model. This method ignores assumptions regarding the distribution of survival times studied. If there are indications of multicollinearity in data handling, one way that can be done is to use PCA (Principal Component Analysis). PCA-Cox regression is a combination of survival analysis and PCA which can be an alternative in analyzing multicollinearity survival data. The large number of cases of bad credit means that customers must be careful in providing credit to prospective customers. Character, capacity, capital and collateral variables are thought to influence the length of time customers pay house ownership loans at the bank. The data used is secondary data (n=100) regarding the assessment of character variables, capacity, capital and collateral, credit collectibility, and time to pay customer house ownership loans at the bank. The results of the analysis using PCA-Cox regression show that the variables character, capacity, capital and collateral have a significant effect on the length of house ownership loan payment time for Bank X customers. The originality of this research is the use of the PCA-Cox regression integration model in bank credit risk analysis.

1. INTRODUCTION

Survival analysis is a robust statistical approach designed for analyzing time-to-event data, where the main variable of interest is the duration until a particular event occurs (Kleinbaum & Klein, 2005). The primary goal is to explore the relationship between survival time and predictor variables that influence the timing or likelihood of the event. Survival time refers to the interval between a defined starting point and the occurrence of an event, such as failure, recovery, relapse, or bankruptcy. This makes survival analysis highly applicable in diverse fields, including healthcare, engineering, and finance, where understanding the time to an event is crucial. There are three primary approaches to survival analysis: parametric, non-parametric, and semi-parametric (Pourhoseingholi et al., 2007). The parametric approach assumes that survival times follow a specific probability distribution, such as Weibull or exponential, making it suitable when the underlying data distribution is well-understood. Non-parametric methods, like Kaplan-Meier and life tables, do not rely on assumptions about the survival time distribution, making them more flexible but less powerful in cases with substantial prior knowledge of the data. Semi-parametric methods, notably Cox proportional hazards regression, strike a balance by not requiring distributional assumptions while allowing for the inclusion of covariates (George et al., 2014; Mardhiah et al., 2022).

Despite its widespread use, Cox regression has limitations, particularly when multicollinearity exists among predictor variables. Multicollinearity arises when predictor variables are highly correlated, leading to biased coefficient estimates and reduced model reliability (Fernandes & Solimun, 2016). To address this, advanced techniques like Partial Least Square-Cox Regression (PLS-Cox) and Principal Component Cox Regression (PCA-Cox) have been developed. PCA-Cox regression combines the principles of Principal Component Regression (PCR) and Cox Regression by first reducing the dimensionality of the data through principal component analysis and then regressing the component scores against survival time. This approach mitigates multicollinearity while preserving the interpretability of survival relationships (Fernandes & Solimun, 2014).

The financial sector, particularly banking, often faces credit risks associated with non-performing loans. One example is house ownership loans, a prominent credit facility provided by banks for purchasing or renovating homes. These loans are critical in addressing Indonesia's substantial housing demand, with a shortage of 11 million habitable units reported in 2021 (Portal Informasi Indonesia, 2022). House ownership loans remain the primary choice for home financing in Indonesia, accounting for 69.54% of consumer financing, according to Bank Indonesia (Rahman, 2022). This highlights the need for banks to carefully evaluate creditworthiness to mitigate potential losses from non-performing loans. Banks assess creditworthiness using the 4C framework: character, capacity, capital, and collateral (Wahyuni, 2017). These factors collectively measure a debtor's reliability, financial ability, resources, and loan security. However, effective assessment becomes challenging when these variables exhibit multicollinearity, which can distort statistical analyses. Survival analysis, particularly when combined with PCA, provides a valuable tool for evaluating the timing of loan repayments and identifying factors influencing repayment behavior (Lin et al., 2006; Maxwell et al., 2019).

Research on survival analysis in credit risk often focuses on specific techniques. For example, the Accelerated Failure Time (AFT) model has been used to describe repayment timing (Kawi & Purwono, 2022). Meanwhile, PCA has proven effective in reducing dimensionality and multicollinearity before applying Cox regression (Lin et al., 2006). Studies comparing methods for addressing multicollinearity, such as PCR, Ridge Regression, and LASSO, have shown PCA-based approaches to be superior in many contexts (Maxwell et al., 2019). Considering these advancements, this study aims to develop a survival analysis model combining PCA and Cox regression. The model will examine the relationship between 4C framework variables and the timing of house ownership loan repayments for Bank X customers. By addressing multicollinearity, the study seeks to improve the accuracy and interpretability of credit risk assessments, providing insights into factors that influence loan performance and helping banks make informed decisions to minimize non-performing loans and financial losses.

2. LITERATURE REVIEW

2.1. Survival Analysis

Survival analysis is a technique in statistical analysis where the response variable studied is the time until a specific event occurs. In this analysis, the main focus is on time. analysis, according to Kleinbaum & Klein (2005), is a statistical technique designed to handle or assess data with a dependent variable representing the duration until an event transpires. In this analytical approach, various factors that influence the survival duration, denoting the period an individual remains event-free, can elucidate the phenomenon.

When observing survival data, incomplete subject data can often be found because the subject did not experience the specific event observed (failure time) until the study period ends. It usually takes a long time and a lot of money if you want to get complete and intact survival data where all individuals experience failure time. In overcoming this problem, survival analysis has special considerations for incomplete data which is commonly known as data censoring (Katzman *et al*., 2018). Data censoring occurs because the survival time of a subject cannot be known exactly, but there is some information that can provide clues about the subject's survival time (Kleinbaum & Klein, 2005).

Survival Function S(t) states the probability that a subject will survive (not experience an incident) up to or beyond time t (Lawless, 1982). The probability density function (PDF), denoted as $f(x)$ represents the likelihood of an individual experiencing a specific event and fundamental in quantifying the instantaneous rate of occurrences, at an exact time t within a continuous time interval. According to the definition above, the probability density function can be formulated as shown in Equation (1).

$$f(x) = \lim_{\Delta t \to 0} \left[\frac{P(t < T < (t + \Delta t)}{\Delta t} \right] = \lim_{\Delta t \to 0} \left[\frac{F(t + \Delta t) - F(t)}{\Delta t} \right] \qquad (1)$$

If T is a non-negative random variable then $F(t)$ is the cumulative distribution function of T. $F(t)$ is defined as the probability of an individual experiencing a failure event at time t written as shown in Equation (2).

$$F(t) = P(T \leq t) = \int_0^t f(x)\, dx, \qquad for\ t > 0 \qquad (2)$$

The PDF $f(x)$ is derived as the limit of the probability that the event occurs within a very small interval $[t, t + \Delta t]$, divided by the interval's length (Δt), as Δt approaches zero. It provides insight into the timing of events and forms the basis for deriving other survival metrics, such as the cumulative distribution function (CDF) $F(t)$ and the survival function $S(t)$. These relationships allow us to model and analyze the probability of survival over time.

$$S(t) = P(T \geq t) = \int_t^\infty f(x)\, dx \qquad (3)$$

Hazard function h(t) or also known as the hazard rate states the instantaneous rate at which a subject experiences the incident at a specific time t (Katzman *et al*., 2018; Kleinbaum & Klein, 2005). The hazard function can be formulated as written in Equation (4) (Lawless, 1982).

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P(t \leq T < (t + \Delta t)|T \geq t}{\Delta t} \right] \qquad (4)$$

2.2. Cox Proportional Hazard Model

Cox proportional hazards, the most popular model in semiparametric survival analysis, can analyze how a predictor variable influences the time until an event occurs (survival time). However, there is a weakness in this modeling, namely that it cannot estimate with certainty what the functional hazard form will be from the survival time studied (Kleinbaum & Klein, 2005; Rabe-Hesketh & Skrondal, 2008). Cox proportional hazards modeling does not require assumptions regarding the distribution of survival times, so this model is flexible to problems involving the distribution of survival times. The Cox proportional hazard model can be formulated in Equation (5).

$$\hat{h}(t,x) = h_0(t)\exp\left(\beta_1 x_1 + \cdots + \beta_p x_p\right) \tag{5}$$

where $h_0(t) =$ basic failure function (baseline hazard); $x_j =$ independent variable value of x_j, $j = 1,2,\ldots,p$; $\beta_j =$ regression parameter, $j = 1,2,\ldots,p$.

The survival time depends on the values $x_1, x_2, \ldots, x_p$combined in the vector x of size $p \times 1$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$. The vector $\boldsymbol{\beta}$ contains regression coefficients resulting from parameter estimation which provides information on how much influence the predictor variables have on survival time.

Hazard Ratio (HR) is a comparison of the hazard of one individual with another individual. Kleinbaum & Klein (2005) explains that the hazard ratio is a measure that can be used to determine the level of risk seen by comparing an individual who has a value of variable X in the success category with another individual who has a value of variable X in the category of failure. Stensrud & Hernán (2020) explains the hazard ratio as the effect or magnitude of the influence of the predictor variable on the observed time. By using the hazard ratio, the Cox regression coefficient obtained can be easily interpreted and understood. Hazard ratio can be calculated with Equation (6).

$$\begin{aligned} HR &= \frac{h(t,x^*)}{h(t,x)} \\ &= \frac{h_0(t)\exp\left(\beta_1 x^*{}_1 + \cdots + \beta_p x^*{}_p\right)}{h_0(t)\exp\left(\beta_1 x_1 + \cdots + \beta_p x_p\right)} \\ &= \exp\left(\beta_1(x^*{}_1 = x_1) + \cdots + \beta_p(x^*{}_p - x_p)\right) \\ &= \exp\left(\sum_{j=1}^{p} \beta_j (x_j^* - x_j)\right) \end{aligned} \tag{6}$$

Where $h(t,x^*) =$ hazard function for an individual with variable value x^*; $h(t,x) =$ hazard function for an individual with variable value x; $h_0(t) =$ basic failure function (baseline hazard); $\beta =$ regression coefficient; $x^* =$ predictor variable value for success/comparison group; $x =$ predictor variable value for base/reference group; $p =$ number of predictor variables.

2.3. PCA-Cox Regression

According to Candès *et al*. (2011), the Principal Component Analysis (PCA) method was first introduced by Harold Hotelling. PCA functions to eliminate multicollinearity from predictor variables. According to Bair *et al*. (2006) PCA can be applied to multicollinearity problems in regression analysis such as Cox regression in survival analysis. By forming the main components of the predictor variables that contain multicollinearity, it can be seen which predictor variables have a significant effect and which predictor variables can be identified which are important. PCA-Cox regression aims to predict survival chances from response variables. The PCA-Cox regression method involves principal components formed from a linear combination of the predictor variables studied. The main components that have been formed from the PCA-Cox regression model are components that are independent of each other or do not have multicollinearity, so that in their application they can be used to overcome non-fulfillment of the non-multicollinearity assumption in modeling (Fernandes & Solimun, 2016).

Component score regression sk_{ij} using the PCA method is carried out by regressing the component scores on the response variable Y. This means using the component scores obtained from the PCA results as predictor variables for the Cox regression on the response variable Y. The following equation for the regression results of component scores sk_{ij} on the response variable Y can be written as Equation (7).

$$h_i(t) = h_0(t)\, exp(c_1 sk_{i1} + c_2 sk_{i2} + \cdots + c_m sk_{np}) \tag{7}$$

where $h_0(t) =$ basic failure function (baseline hazard); $sk_{ij} =$ score of the i-th component of the j-th original variable; $c_j = j$-th Cox regression coefficient, $j = 1,2, \dots, m$

The equation above is a Cox regression model with predictor variable component scores obtained using the PCA method, so in summary the above equation can be called the Principal Component Analysis Cox model Regression or PCA-Cox regression.

Basically, the PCA-Cox method regression aims to determine the form of the connection or association between...the predictor variable (X) and the response variable (Y). Based on this, a transformation is needed from the main component scores to the original variables.

3. MATERIAL AND METHOD

The variables used in this research are character, capacity, capital, collateral, credit collectibility, and the time for house ownership loan payments for each customer. Data were obtained from research conducted at Bank X in 2023. The event studied in this research is the timing of house ownership loan payments. This research uses data from 100 subjects who were customers with house ownership loans at Bank X and were willing to be interviewed. The variables include assessments of character, capacity, capital, collateral, and credit collectibility, while the response variable is the time for house ownership loan payments. The data were analyzed using R Studio software. The data measurement scale employed in this study is presented in Table 1.

Table 1. Data Variables and Scales

Variable	Information	Data Type
Character (X_1)	Age	Continuous
Capacity (X_2)	RPA (Installment Income Ratio)	Continuous
Capital (X_3)	Loan To Value	Continuous
Collateral (X_4)	Length of Residence	Continuous
Credit Payment Time (t)	Time	Continuous

The analysis used is PCA-Cox Regression. The important point in survival analysis is being able to analyze censored data. In this research, the status of customer censored data is seen from credit collectability, so that the analysis still considers data on customers whose credit is bad. Principal Component Analysis aims to accommodate multicollinearity in the data. The main components formed from PCA are a linear combination of the original variables. The resulting component scores were then formed into a Cox Proportional Hazard model. The model formed at this stage is the Cox proportional hazard model between the customer's house ownership loan payment time and the main components formed from PCA. This modeling is a model that is more sensitive to the multicollinearity between variables. The relationship between variables can be explained without bias using this modeling. The final step is to carry out a transformation to the original variables to form a Cox proportional hazard model which can explain the relationship between the customer's house ownership loan payment time and the character, capacity, capital and collateral variables.

4. RESULTS AND DISCUSSION

The foundation of Cox proportional hazard modeling relies on the Cox proportional hazard assumption. This assumption signifies that the ratio between individuals in different categories remains consistent throughout time and is not influenced by time. The outcomes of the Cox proportional hazard assumption assessment via the Global Test are detailed in Table 2.

Table 2. Proportional Hazard Assumption Test Results

Variable	Test Statistics	p-value
Character (X_1)	0.23	0.63
Capacity (X_2)	1.36	0.24
Capital (X_3)	1.41	0.23
Collateral (X_4)	1.42	0.36

Testing the Cox proportional hazard assumption uses the following hypothesis test.

$H_0 : \rho = 0$ (Assuming PH is met)

$H_1 : \rho \neq 0$ (PH assumption not met)

Based on Table 2, it is found that all predictor variables have a p-value > 0.05. So the decision taken is to accept H_0. So, it can be concluded that the hazard ratio for character, capacity, capital and collateral does not depend on time or the proportional hazard assumption is met.

The existence of multicollinearity between predictor variables can cause the estimator to have a variance that tends to be large. A large variance value causes the confidence interval to be wider so that the test results accept more of the null hypothesis. In fact, in the case of perfect multicollinearity between variables, the estimated model coefficients cannot be estimated. This also applies to Cox regression modeling in survival data analysis. Multicollinearity checking can use the VIF value. In this case, the VIF values between variables is presented in Table 3.

Table 3. Predictor Variable VIF Value

	Variable			
	Character	Capacity	Capital	Collateral
VIF value	12.9998	15.95228	10.5694	10.7558

From Table 3, it can be seen that all predictor variables have a value of $VIF > 10$. So, it can be concluded that the data contains multicollinearity between variables or there is a close relationship between one predictor variable and another.

Principal Component Analysis (PCA) is a commonly used technique for addressing data that is affected by multicollinearity. Within this analysis, a primary component is constructed, representing a linear combination of predictor variables affected by multicollinearity. These primary components are assured to be mutually independent.

Therefore, principal component analysis is one way that is commonly used to handle data that contains multicollinearity. Principal component analysis begins with determining the input matrix. In this case, the editorial variables involved have different units, so the input matrix used is the R matrix. Once the input matrix has been determined, we proceed with the formation of the main components. Formation of principal components begins with calculating the eigenvalues and eigenvectors of the input matrix.

Table 4. Eigenvalue, Diversity Proportion, and Cumulative Diversity Proportion of Principal Components

	PC_1	PC_2	PC_3	PC_4
Eigenvalues	1.2386	1.2040	0.8712	0.5071
Diversity Proportion	0.3836	0.3624	0.1898	0.0643
Cumulative Diversity Proportion	0.3836	0.7460	0.9357	1

Cox regression modeling is carried out by regressing the component scores resulting from PC_1, PC_2, PC_3 the customer credit payment time variable (t). The PCA-Cox regression equation that models the component scores $sk_1, sk_2$and sk_3 the time to pay customer credit variables (t) is as follows.

$$h_i(t) = h_0(t)\exp\left(-0.2694\ sk_1 + 0.1902\ sk_2 + 0.6025\ sk_3\right)$$

Cox regression model into the form of original variables to model the relationship between variables X_1, X_2, X_3, X_4 and X_5 the timing of customer credit payments (t). The following is the Cox proportional hazards regression model that is formed:

$$h_i(t) = h_0(t)\ \exp\left(0.0645\ X_{1i} - 0.4397\ X_{2i} + 0.1265\ X_{3i} - 0.3685\ X_{4i}\right)$$

The hazard ratio of the Cox proportional hazard model is described as follows.

a. Character (X_1)

$$HR = \frac{h_A}{h_B} = \exp\beta_1 = \exp(0.0645) = 1.0667$$

The hazard ratio for the Character variable shows a value of, meaning that a one year increase in the customer's age will increase the potential for the customer to have 1.0667a longer house ownership loan payment time.1.0667

b. Capacity (X_2)

$$HR = \frac{h_A}{h_B} = \exp\beta_2 = \exp(-0.4397) = 0.6442$$

The hazard ratio for the Capacity variable shows a value of 0.6442, meaning that increasing the customer's Installment Income Ratio (RPA) by 1 unit will increase the customer's potential to have a $0{,}6442$ faster house ownership loan payment time.

c. Capital (X_3)

$$HR = \frac{h_A}{h_B} = \exp\beta_3 = \exp(0.1265) = 1.1349$$

The hazard ratio for the Capital variable shows a value of 1.1349, meaning that an increase in the Loan to Value value owned by a customer by 1 unit will reduce the potential for the customer to have a 1.1349 longer house ownership loan payment time.

d. Collateral (X_4)

$$HR = \frac{h_A}{h_B} = \exp\beta_4 = \exp(-0.3685) = 0.6917$$

The hazard ratio for the Collateral variable shows a value of 0.6917, meaning that increasing the time of residence for one year will have the potential to increase the customer's house ownership loan payment time 0.6917 faster.

Significance testing using the Jackknife method begins by carrying out Jackknife resampling. The purpose of this testing is to evaluate the stability and reliability of the model's coefficients while reducing bias from a single data sample. At each resampling, the PCA-Cox regression coefficient is then calculated which is a Cox regression of component and variable sk_3 scores $sk_1,\ sk_2$. Then a transformation is carried out to the initial variables to form a Cox proportional hazards regression model containing the original variables X. The results of the transformation of the PCA-Cox model to the Cox proportional hazards model at each resampling can be seen in Table 5.

Table 5. Transformation Result Coefficient for Each Jackknife Resampling

Resampling to-	$\hat{\beta}^*{}_1$	$\hat{\beta}^*{}_2$	$\hat{\beta}^*{}_3$	$\hat{\beta}^*{}_4$
1	0.0628	-0.4368	0.1298	-0.3709
2	0.1160	-0.4150	0.0867	-0.3554
3	0.0788	-0.4536	0.1266	-0.3851
4	0.0512	-0.4505	0.1245	-0.3761
5	0.0479	-0.4558	-0.0076	-0.3503
⋮	⋮	⋮	⋮	⋮
500	0.0887	-0.4456	0.1260	-0.3914

Based on the coefficient estimates produced at each resampling, standard error values are obtained for each coefficient $\hat{\beta}^*$. Then a Z test was carried out for partial testing using a jackknife standard error. The hypothesis used is as follows.

$H_0 :\ \beta_i = 0; i = 1, 2, 3, 4$
$H_1 :\ \beta_i \neq 0; i = 1, 2, 3, 4$

The significancy level used is 5%. The results of the test statistics and p-value calculations are summarized in Table 6.

Table 6. Z test statistics and P-value Partial Test

Variable	$\hat{\beta}_i$	$se(\hat{\beta}_1)$	Z Test Statistics	p-value	Decision
Character(X_1)	0.0645	0.0331	2.3176	0.0102	Reject H_0
Capacity (X_2)	-0.4397	0.1426	-2.7679	0.0028	Reject H_0
Capital (X_3)	0.1265	0.0545	2.1468	0.0159	Reject H_0
Collateral (X_4)	-0.3685	0.1550	-3.1494	0.0008	Reject H_0

From Table 6, we can observe the p-value associated with the Character variable (X_1), Capacity (X_2), Capital (X_3), and Collateral (X_4) have values greater than the significance level 0.05, so the decision taken is to reject H_0 for each variable. It can be concluded that the Character, Capacity, Capital, and Collateral variables influence the speed or slowness of customer credit payment times.

The quality or goodness of a model can be assessed by examining the R^2-value the model has. R^2 explains how much diversity of data the model can explain. The PCA- Cox regression model was formed which has a value $R^2$of 77.55%, meaning that the model can account for 77.55% of the variance in the data, with the remaining 22.45% being attributed to unexplained factors not addressed in the study. So, it can be concluded that the model formed is quite good in explaining the relationship between character, capital, capacity and collateral variables on the house ownership loan payment time for Bank X customers, but the influence of research factors is still quite strong.

There is a high multicollinearity relationship between predictor variables. This is indicated by the VIF for each variable having a value of more than 10. Before modeling the relationship between the character, capital, capacity and collateral variables on the customer house ownership loan payment time variable, it is necessary to handle multicollinearity. The existence of multicollinearity between predictor variables is because the assessment aspects (predictor variables) are interrelated with each other. In this research, multicollinearity is handled by conducting principal component analysis. The resulting main components are a linear combination of predictor variables and it is guaranteed that the main components are independent of each other. Modeling the relationship between the variables character, capital, capacity, and collateral on the time to pay a customer's house ownership loan is modeled using the Cox proportional hazards model. because the modeling is carried out using component scores resulting from principal component analysis, it is called PCA-Cox regression modeling.

After checking the proportional hazard assumption, it was concluded that the assumption was met for all predictor variables, so that we could proceed to PCA-Cox regression modeling. The results of the analysis show that all assessments of character, capital, capacity and collateral variables have a significant effect on the length of house ownership loan payment time for Bank. This is in line with Ardani & Herawati (2021) research, which concluded that character, capital, capacity and collateral influences the effectiveness of providing credit. In the Capacity variable, increasing the customer's Installment Income Ratio (RPA) by 1 unit will increase the customer's potential to have a customer house ownership loan payment time of 0.6442 times faster. In the Capital variable, an increase in the Loan to Value value owned by a customer by 1 unit will reduce the customer's potential to have a customer house ownership loan payment time of 1.1349 times longer. In the Collateral variable, increasing the time to occupy a residence for one year will have the potential to increase the customer's house ownership loan payment time by 0.6917 times faster.

To improve loan repayment performance, Bank X should refine its customer assessment by incorporating additional variables, such as credit history and behavioral scoring. Risk management should focus on customers with favorable Installment Income Ratio (RPA) and Loan to Value (LTV) ratios, while monitoring high-risk clients to reduce defaults. Loan terms, like repayment duration and interest rates, can be adjusted based on customer profiles for better repayment outcomes. Additionally, educating customers on maintaining optimal financial ratios can enhance their repayment capacity. Lastly, similar analyses for other loan types could provide further insights for credit management.

5. CONCLUSION

Based on the research conducted, the relationship between house ownership loan payment times and customer characteristics such as character, capacity, capital, and collateral were analyzed using the PCA-Cox regression method. This method effectively addressed multicollinearity issues within the data, resulting in a more stable and interpretable model. The findings emphasize the importance of thoroughly understanding customer profiles to estimate risks comprehensively. This approach enables financial institutions, such as banks, to anticipate potential risks more effectively related to house ownership loans.

For future research, it is recommended to include additional predictor variables, such as the 7Ps (Personality, Party, Purpose, Prospect, Payment, Profitability, and Protection), which are believed to significantly influence loan payment durations. Furthermore, the use of the

PLS-Cox regression method is suggested as an alternative for modeling survival data with multicollinearity, offering potentially more accurate and in-depth results. These advancements aim to provide financial institutions with enhanced insights into credit risk management and support more efficient decision-making processes.

ACKNOWLEDGMENT

I would like to thank all parties who have provided support, assistance and prayers during the preparation and completion of this article.

REFERENCES

Ardani, N. W. S., & Herawati, N. T. (2021). Pengaruh Penerapan Prinsip 5C Dan Sistem Pengendalian Internal Terhadap Efektivitas Pemberian Kredit Pada Lembaga Pekreditan Desa (LPD) Di Kabupaten Gianyar. *JIMAT (Jurnal Ilmiah Mahasiswa Akuntansi) Undiksha*, *12*(2), 547–556. https://doi.org/10.23887/JIMAT.V12I2.30302

Bair, E., Hastie, T., Paul, D., & Tibshirani, R. (2006). Prediction by Supervised Principal Components. *Journal of the American Statistical Association*, *101*(473), 119–137. https://doi.org/10.1198/016214505000000628

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust Principal Component Analysis. *Journal of the ACM*, *58*(3), 1–37. https://doi.org/10.1145/1970392.1970395

Fernandes, A. A. R., & Solimun. (2014). Comparative Method in PCA and PLS Cox Regression to Solve Multicollinearity. *Global Journal of Pure and Applied Mathematics*, *10*, 581–590.

Fernandes, A. A. R., & Solimun. (2016). *Pemodelan Statistika Pada Analisis Reliabilitas dan Survival*. https://books.google.com/books?hl=id&lr=&id=dHhNDwAAQBAJ&oi=fnd&pg=PR5&dq=fernandes+solimun+pemodelan+statistika+pada+analisis+reliabilitas&ots=9Pqu34LmOm&sig=nKRtQa5arXSfTUdDVWEW6zA451Q

George, B., Seals, S., & Aban, I. (2014). Survival Analysis and Regression Models. *Journal of Nuclear Cardiology*, *21*(4), 686–694. https://doi.org/10.1007/S12350-014-9908-2

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: Personalized Treatment Recommender System using a Cox Proportional Hazards Deep Neural Network. *BMC Medical Research Methodology*, *18*(1), 1–12. https://doi.org/10.1186/S12874-018-0482-1/FIGURES/6

Kawi, Y. N., & Purwono, Y. (2022). Aplikasi Metode Accelerated Failure Time (AFT): Analisis Risiko Prepayment pada Kredit Kendaraan Bermotor. *Journal of Management and Bussines (JOMB)*, *4*(1), 234-252. https://doi.org/10.31539/jomb.v4i1.3677

Kleinbaum, D. G., & Klein, M. (2005). *Survival Analysis*. New York: Springer. https://doi.org/10.1007/0-387-29150-4

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data (Wiley Series in Probability & Mathematical Statistics)*. United States: John Wiley & Sons.

https://books.google.com/books/about/Statistical_Models_and_Methods_for_Lifet.html?hl=id&id=-hrvAAAAMAAJ

Lin, D., Banjevic, D., & Jardine, A. K. S. (2006). Using Principal Components in a Proportional Hazards Model with Applications in Condition-Based Maintenance. *Journal of the Operational Research Society*, *57*(8), 910–919. https://doi.org/10.1057/PALGRAVE.JORS.2602058

Mardhiah, K., Wan-Arfah, N., Naing, N., Hassan, M., & Chan, H. K. (2022). Comparison of Cox Proportional Hazards Model, Cox Proportional Hazards with Time-Varying Coefficients Model, and Lognormal Accelerated Failure Time Model: Application in Time to Event Analysis of Melioidosis Patients. *Asian Pacific Journal of Tropical Medicine*, *15*(3), 128–134. https://doi.org/10.4103/1995-7645.340568

Maxwell, O., Chukwudike, C. N., Chinedu, O. V., Valentine, C. O., & Paul, O. C. (2019). Comparison of Different Parametric Methods in Handling Critical Multicollinearity: Monte Carlo Simulation Study. *Asian Journal of Probability and Statistics*, *3*(2), 1–16. https://doi.org/10.9734/ajpas/2019/v3i230085

OJK. (2019). *Peraturan Otoritas Jasa Keuangan Nomor 40/POJK.03/2019*.

Portal Informasi Indonesia. (2022). *Indonesia.go.id - Rumah Layak agar Warga Sehat dan Produktif*. https://indonesia.go.id/kategori/sosial/3971/rumah-layak-agar-warga-sehat-dan-produktif

Pourhoseingholi, M. A., Hajizadeh, E., Dehkordi, B. M., Safaee, A., Abadi, A., & Zali, M. R. (2007). Comparing Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma. *Asian Pacific Journal of Cancer Prevention*, *8*(3), 412–416. https://journal.waocp.org/article_24627.html

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling using Stata*. https://books.google.com/books?hl=en&lr=&id=woi7AheOWSkC&oi=fnd&pg=PR21&ots=eeOxak0ONz&sig=88HHQVj52JsbDTbNLTodPpVcZNc

Rahman, D. F. (2022). *KPR Masih Jadi Pilihan Favorit Masyarakat Membeli Rumah pada Triwulan I 2022*. https://databoks.katadata.co.id/datapublish/2022/05/23/kpr-masih-jadi-pilihan-favorit-masyarakat-membeli-rumah-pada-triwulan-i-2022

Stensrud, M. J., & Hernán, M. A. (2020). Why Test for Proportional Hazards? *JAMA Guide to Statistics and Methods*, *323*(14), 1401–1402. https://doi.org/10.1001/jama.2020.126

Wahyuni, N. (2017). Penerapan Prinsip 5c dalam Pemberian Kredit sebagai Perlindungan Bank. *Lex Journal: Kajian Hukum &Amp; Keadilan*, 1(1). https://doi.org/10.25139/lex.v1i1.236