Desimal: Jurnal Matematika Vol 5 No 3 (2022) 315-328

Contents lists available at DJM

DESIMAL: JURNAL MATEMATIKA

p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm

http://ejournal.radenintan.ac.id/index.php/desimal/index

Machine learning survival analysis on couple time-to-divorce data

Muhammad Luthfi Setiarno Putera[1,*], Setiarno[2]

[1] Institut Agama Islam Negeri Palangka Raya, Indonesia
[2] Universitas Palangka Raya, Indonesia

ARTICLE INFO

Article History

Received : 11-09-2022
Revised : 18-11-2022
Accepted : 07-12-2022
Published : 20-12-2022

Keywords:
Backward Elimination; Concordance Index; Cox Proportional Hazard; Divorce; Survival Support Vector Machine.

*Correspondence: E-mail: m.luthfi@iain-palangkaraya.ac.id

Doi: 10.24042/djm.v5i3.13742

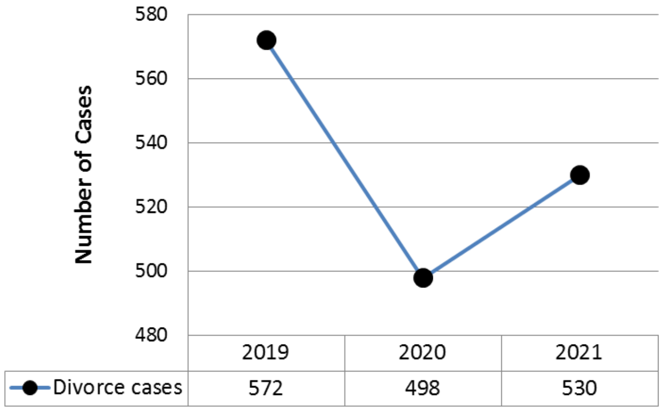
ABSTRACT

Marriage life does not always last harmoniously and occasionally can lead to divorce. The trend for the last three years since 2019 shows that divorce cases in Palangka Raya occur with a fluctuating trend that has recently been increasing. This research used a machine learning method called Survival Support Vector Machine on the divorce dataset in Palangka Raya. This research developed a feature selection technique using backward elimination to determine the factors influencing the couple's decision to have their divorce registered in the religious court. The backward elimination method yielded the variables contributing to divorce: the number of children, the defendant's occupation, the plaintiff's age at marriage, the cause of divorce, and the defendant's education. Based on the comparison of the survival model performance between the Cox proportional hazard and the Survival Support Vector Machine, it was found that the latter was better since it had a higher concordance index and hazard ratio, which were 61.24 and 0.54, respectively. Thus, 61.24% of divorce cases were classified precisely by SUR-SVM in terms of the time sequence of events. Moreover, the hazard ratio of 0.54 indicated that the divorce rate of couples with censored status was 0.54 times than that of couples with failed/endpoint status.

http://ejournal.radenintan.ac.id/index.php/desimal/index

INTRODUCTION

Harmonious marriage life has always been a dream of every couple. The problem in marriage life is unavoidable (Samsuri, 2018). Few problems can be handled well by many couples, while others may force a divorce. Based on the annual report of the Palangka Raya Religious Court, the number of received divorce cases, as shown in Figure 1, has always been in the range of 500 cases in the last 3 years. The figure reached 572 cases in 2019 and 498 cases in 2020. Though, divorce cases in 2021 increased to 530 cases.



(*Source* : Pengadilan Agama Palangka Raya, 2021)

Figure 1. Number of Divorce Cases in 2019-2021

Figure 1 indicates a fluctuation in divorce cases in Palangka Raya for the past 3 years, where the trend in the last year has shown an increase.

Referring to Kleinbaum & Klein (2012), survival analysis could be used to identify factors related to the time an event occurred, such as divorce. However, such cases filed in religious courts are not always classified as legally binding divorces, because the case may have been successfully mediated or the divorce has not yet been sentenced at the end of the research. The state of such censored data makes it impossible to observe the time for a divorce.

Survival analysis could be carried out using Cox proportional hazard regression (Kleinbaum & Klein, 2012). Suryaningrum (2019) applied such a method to analyze the survival of marriage time in East Jakarta and found that couples without any children and the existence of disputes contributed to divorce. However, this method has flaws, particularly when the proportional over-time assumption for the hazard of two individuals is not met, the occurrence of covariate dependencies, and so on (Riyadi et al., 2018). To address this, there are several machine learning methods employing a non-parametric approach, including the Survival Support Vector Machine (SUR-SVM).

Unlike Cox proportional hazard (Cox PH), SUR-SVM does not require a proportional hazard assumption (Spooner et al., 2020). Besides, based on Prastyo et al. (2020), SUR-SVM could also be applied to data whose predictors exceed ten variables and involve numerous observations, as found in divorce cases in many regions, which relate to various socio-economic variables and happen to many couples within a certain time.

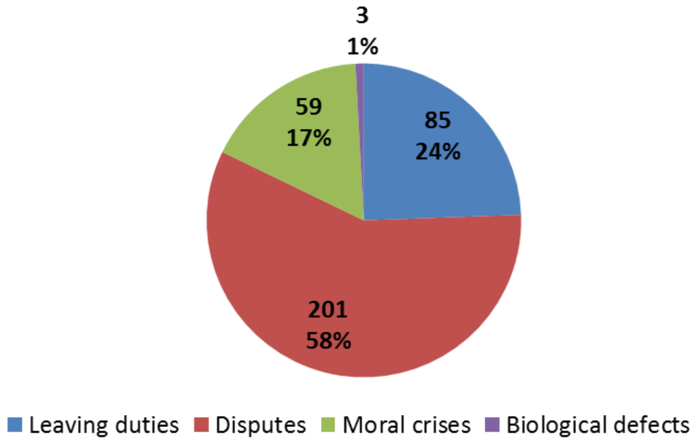
Research using SUR-SVM were carried out by Fouodo et al. (2018) and Mihaylov et al. (2019). Fouodo et al. (2018) examining several different datasets, indicated that SUR-SVM was slightly superior to Cox PH and could be an appropriate alternative to non-parametric models. Besides, Mihaylov et al. (2019) studying breast cancer, found that SUR-SVM had a better predictive performance than Cox PH. It could be seen from the higher SUR-SVM concordance index (c-index).

The increasing number of couples filing for divorce in Palangka Raya encourages this research to identify the factors influencing the time-to-divorce. However, unlike linear regression, the SUR-SVM model cannot directly determine the significant factors. Therefore, this research offers a feature selection technique using backward elimination to identify factors influencing the time to divorce based on the c-index indicator. This research also compares the performances of the Cox PH and SUR-SVM models to look for a better survival model.

METHOD

This research is quantitative and employs a machine learning method, namely Support Vector Machine (SVM). Survival analysis combined with SVM resulted in Survival SVM (SUR-SVM). This research uses R as the main software to process and analyze data.

The research data is accessed from the Supreme Court website, which contains the directory of divorce decisions of the Palangka Raya Religious Court. The observed study period is one year, starting from April 2019 to March 2020. Considering that not all divorce cases are freely accessible, the sampling is carried out using a conventional approach by downloading decisions that could be accessed in general. There were 348 divorce decisions between April 2019 and March 2020 that are publicly accessible. Figure 2 shows the number of divorce cases with respect to the cause of divorce.



(*Source* : Mahkamah Agung Republik Indonesia, 2020)

Figure 2. The Number of Divorce Cases with Respect to the Cause of Divorce

SUR-SVM analysis for regression purposes involves several variables, both response variables and predictors. The responses consist of the time to divorce (marital duration) (t) and censored status (θ). Table 1 contains an explanation of the response and predictors.

Table 1. Variable Description

Variable	Description	Scale
Survival time (t)	Duration of marriage to divorce (time to divorce), expressed in years	Ratio
Status (θ)	The status of event 1: sentenced to divorce 0: not sentenced to divorce(censored)	Nominal
x_1	Plaintiff's age at marriage	Ratio
x_2	Plaintiff's education	Ordinal
x_3	Plaintiff's occupation	Nominal
x_4	Defendant's age at marriage	Ratio
x_5	Defendant's education	Ordinal
x_6	Defendant's occupation	Nominal
x_7	Number of children	Ratio
x_8	Cause of divorce	Nominal

In Table 1, variables x_2 and x_5 have the same four categories: elementary school or equivalent, junior high school or equivalent, senior high school or equivalent, and college. Variables x_3 and x_6 include the same four categories: untrained workers, skilled workers, semi-professionals, and professionals. Variable x_8 has four categories: leaving duties, disputes, moral crises, and biological defects. Before being regressed, all categorical variables are converted to dummies, calculated from the number of categories (n_k) – 1. The steps for conducting this research are shown in Figure 3.

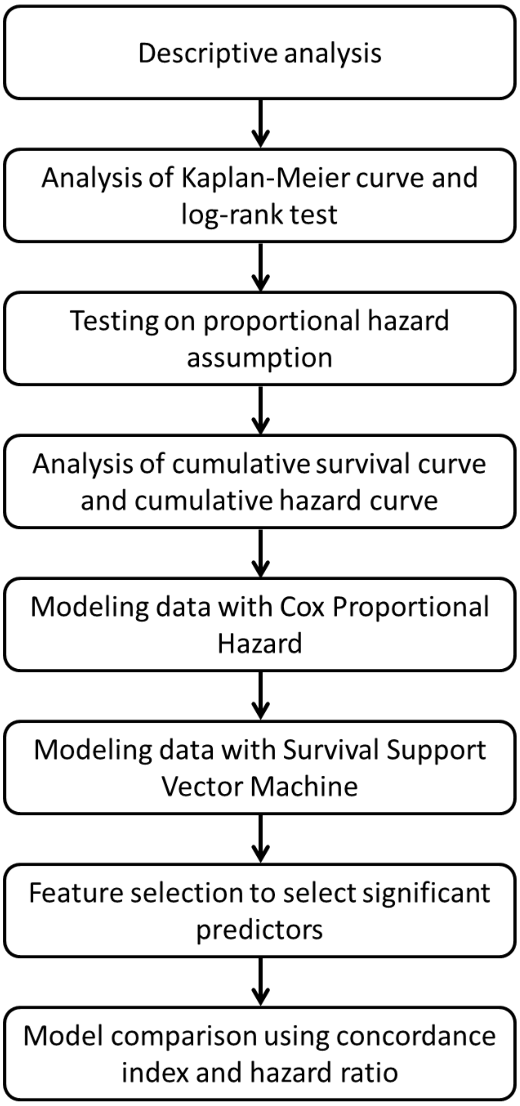


Figure 3. Research Flowchart

Kaplan-Meier curve and log-rank test

The Kaplan-Meier curve is used to determine the survival characteristics of divorce cases. Figure 4 shows the Kaplan-Meier curve of divorce cases.

Figure 4. Kaplan-Meier Curve of Divorce Cases

Figure 4 shows that divorce cases have particular characteristics in which couples with a longer marital duration have a higher probability of divorce.

The log-rank test is to compare whether there are differences in survival curves among categories in a variable (Barman et al., 2020). The hypothesis is

H_0 : No significant difference among categories of the survival curve

H_1 : Any significant difference among categories of the survival curve

with test statistics formulated as

$$\chi^2 = \sum_{p=1}^{P} \frac{\left(O_p - E_p\right)^2}{E_p} \tag{1}$$

where

$$O_p - E_p = \sum_{k=1}^{n} \left(m_{pk} - e_{pk}\right), \text{ and}$$

$$e_{pk} = \left(\frac{n_{pk}}{\sum_{p=1}^{P}\sum_{k=1}^{n} n_{pk}}\right)\left(\sum_{p=1}^{P}\sum_{k=1}^{n} m_{pk}\right).$$

Description :

O_p : number of cases in the p-th category

E_p : expected number of cases in the p-th category

M_{pk} : number of cases in the p-th category experiencing the event at time t_k

N_{pk} : number of at-risk cases experiencing an instantaneous event in the p-th category before time t_k

E_{pk} : expected value in the p-th category at time t_k

P : number of categories in a variable.

The decision is to reject H_0 if $\chi^2 > \chi^2_{\alpha,(P-1)}$ that there is at least one difference in the survival curve in a variable (Mihaylov et al., 2019).

Proportional hazard test

The proportional hazard assumption should be fulfilled by the Cox PH model, meaning that the hazard ratio is time-independent (Faruk, 2018). This test relies on the Schoenfeld error calculated by

$$SR_{mk} = x_{mk} - E\left(x_{mk} \mid R\left(t_{(mk)}\right)\right) \tag{2}$$

and the conditional probability component is obtained from

$$E\left(x_{mk} \mid R\left(t_{(mk)}\right)\right) = \frac{\sum_{l \in R\left(t_{(mk)}\right)} x_{mk} \exp\left(\beta x_l\right)}{\sum_{l \in R\left(t_{(mk)}\right)} \exp\left(\beta x_l\right)} \tag{3}$$

with

SR_{mk} : Schoenfeld error of the m-th predictor for cases experiencing an event at time $t_{(k)}$

x_{mk} : the value of the m-th predictor for cases experiencing an event at time $t_{(k)}$

Next is to create a rank variable v_r based on survival time. The case experiencing the event for the first time is given a value of 1, and so on.

Then, test the correlation between the Schoenfeld error and the ranking variable with the hypothesis,

H_0 : $\rho = 0$

H_1 : $\rho \neq 0$

and below is the statistics test

$$t_{\text{test}} = \frac{r_{v_r,SR_{mk}}\sqrt{n-2}}{\sqrt{1-\left(r_{v_r,SR_{mk}}\right)^2}} \tag{4}$$

where

$$r_{v_r,SR_{mk}} = \frac{\operatorname{cov}\left(v_r, SR_{mk}\right)}{\sqrt{\operatorname{var}\left(v_r\right)\operatorname{var}\left(SR_{mk}\right)}}.$$

The decision is to reject H_0 if $\left|t_{\text{test}}\right| > t_{(\alpha/2,n-2)}$. In any words, the PH assumption does not hold since there is a significant correlation between the Schoenfeld error and the survival time rank variable (Faruk, 2018).

Cumulative survival curve and cumulative hazard curve

The survival function $S(t)$ is the probability that an object does not experience an event until a certain time. The function $S(t)$ is expressed as

$$S(t) = P(T > t) = \int_t^\infty f(u)\,du \tag{5}$$

with T is the time until an event occurs. The hazard function $h(t)$ is the rate at which the event occurs. The frequency of occurrence of events will increase over time (Kleinbaum & Klein, 2012). The cumulative hazard function is written as

$$H(t) = \int_0^t h(u)\,du \text{ or } -H(t) = \ln S(t). \tag{6}$$

Cox Proportional Hazard model

Generally, Kleinbaum & Klein (2012) states that Cox regression can be formulated in the following equation

$$h(t,\mathbf{x}) = h_0(t)\exp\left(\boldsymbol{\beta}^T\mathbf{x}\right) \tag{7}$$

with

$h(t,\mathbf{x})$: hazard function,

$\mathbf{x}$: vector of predictor,

$h_0(t)$: baseline hazard function.

Survival Support Vector Machine

Survival Support Vector Machine (SUR-SVM) is one of the novel machine learning models. Instead of using the hazard function, SUR-SVM uses a prognostic index. The prognostic can be interpreted as an estimate of the couple's reconciliation as a result of mediation or other factors.

In Prastyo et al. (2020), the utility function of the SUR-SVM model is calculated by

$$\mathbf{u}(\mathbf{x}) = \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}) \tag{8}$$

where **w** is the vector of the parameter and $\boldsymbol{\varphi}(\mathbf{x})$ is the transformation of predictor x. The objective function of SUR-SVM is

$$\min_{\mathbf{w},\xi} \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{\gamma}{2}\sum_i \sum_{j,i<j} v_{ij}\xi_{ij};\gamma \geq 0 \tag{9}$$

and the constraint function is

$$\begin{aligned}&\mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_j) - \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i) \geq 1-\xi_{ij};\forall i<j\\ &\xi_{ij} \geq 0;\forall i<j.\end{aligned} \tag{10}$$

Indicator v_{ij} is a comparison between i-th case and j-th case that meet

$$v_{ij} = \begin{cases} 1, \left(t_i < t_j, \delta_i = 1\right) \\ 0, \left(t_i < t_j, \delta_i = 0\right) \end{cases} \quad (11)$$

and ξ_{ij} is the value for violations due to an error in ranking the time of the event (Prastyo et al., 2020).

Feature selection

Feature selection is a method to select features representing a significant set of predictors so as to reduce the possibility of obtaining an unfavorable model. One of the methods used is backward elimination.

The technique is to model all predictors with Cox PH and SUR-SVM, then eliminate one of the predictors and regress with both models (Maulidina et al., 2021). Then, repeat the elimination process again until each predictor gets a turn to be eliminated. The less contributing predictor can be determined from the higher c-index (Prastyo et al., 2020). The Cox PH and SUR-SVM models are then remodeled based only on the set of predictors that previously had the highest c-index.

Concordance index and hazard ratio

The c-index, or concordance index, is a measure of order between the prognostic function and the observed survival time, for both censored and uncensored data. A high c-index value represents a good survival model (Prastyo et al., 2020). Another criterion to determine the goodness of the survival model is the hazard ratio. The higher the hazard ratio, the better the survival model (Bramantoro & Virdyna, 2022).

RESULTS AND DISCUSSION

The number of divorce cases from March 2019 to April 2020 was 348. Those 283 cases were uncensored and the remaining 65 cases were censored. The average time to divorce (t) is 10.43 years, with a median of 8 years. The difference between the average and median values indicates that the distribution of the variable time to divorce, also known as the duration of marriage to divorce, is asymmetric.

It is also shown that the shortest duration of marriage to divorce is 0 years, and the longest one is 36 years. The pattern of the divorce case survival curve is shown in Figure 5.

Figure 5. Kaplan-Meier Curve on Time to Divorce (Marriage Duration to Divorce)

Based on Figure 5, the time-to-divorce (duration of marriage) for couples who file for divorce tends to decrease, with a drastic decrease can be seen at the early marriage intervals of 0 – 5 years and 5 – 10 years. The stability of a couple's chances of survival tends to be stable after 25 years of marriage. This is relevant to the research of Samsuri (2018) that household integrity will be relatively more maintained by couples with longer marriage ages. Table 2 shows the results of the log-rank test for each predictor calculated using Equation (1), with the predictor that is statistically significant at α = 5% in bold.

Table 2. Log-rank Test

Variable	Log-rank Value	d.f	p-value
$\boldsymbol{x_1}$	**18.3**	**1**	**0.00**
x_2	3.6	3	0.30
x_3	3.9	3	0.30
$\boldsymbol{x_4}$	**11.1**	**1**	**0.00**
x_5	2.5	3	0.50
x_6	4.3	3	0.20
$\boldsymbol{x_7}$	**158**	**1**	**0.00**
$\boldsymbol{x_8}$	**10.3**	**3**	**0.02**

Based on Table 2, there are four predictors (in bold) that reject H0 of Equation (1) since the p-value of those variables is less than 0.05, such as the plaintiff's age at marriage (x_1), the defendant's age at marriage (x_4), the number of children (x_7), and the cause of divorce (x_8). It is since the p-value of the four variables is lower than 0.05. It can be said that four variables have different survival curves between groups (categories). Thus, the plaintiff's age at marriage and the defendant's age at marriage can cause significant differences in the chances of marriage survival.

In addition, the number of children and the cause of divorce also provide a significant difference in the chances of marriage survival. The cause of divorce, which can contribute to the chances of marriage survival, is also relevant to previous research by Suryaningrum (2019). As a result, each of the four categories of divorce causes in this research has a different curve.

There are 8 variables to determine which predictors significantly influence the survival time of divorced couples. Table 3 shows the result of the proportional hazard assumption test for each variable calculated based on Equation (2), Equation (3), and Equation (4). The predictor with a statistical significance of α = 5% is highlighted.

Table 3. Proportional Hazard Assumption Test

Variable	Correlation ($\boldsymbol{\rho}$)	Chi-square	p-value
$\boldsymbol{x_1}$	0.14	5.56	**0.02**
x_2	0.05	0.63	0.43
x_3	-0.02	0.12	0.73
$\boldsymbol{x_4}$	-0.01	0.02	0.90
x_5	-0.05	0.69	0.41
x_6	-0.02	0.11	0.74
$\boldsymbol{x_7}$	0.27	18.46	**0.00**
$\boldsymbol{x_8}$	-0.01	0.04	0.83

Table 3 shows the plaintiff's age at marriage (x_1) and the number of children (x_7), highlighted in bold, whose test results reject H0 of Equation (4). This indicates a correlation between Schoenfeld error and the rank of survival time such that the proportional assumption is not satisfied for these two variables. Due to the importance of these variables in modeling the time-to-divorce, another model that does not require the proportional hazard assumption is needed, which is SUR-SVM.

Before modeling the predictors on survival time, Figure 6 presents the cumulative survival function and the cumulative hazard function calculated using Equation (5) and Equation (6).

Figure 6. Cumulative Survival and Cumulative Hazard of Marriage Duration

Based on the left-side picture of Figure 6, couples with 0 – 10 years of marital duration are more likely to file for divorce than couples with longer marital duration. This is indicated by the pattern, which decreases drastically at the interval of 0 – 10 years for divorced couples; hence, it can be said to be more stable.

As the right-side picture of Figure 6 represents the cumulative hazard, the observed pattern is increasing from the bottom left to the top right resembling a ladder. Such a rise indicates that the longer the marital duration of the divorced couple, the higher the rate of couples experiencing divorce events. This finding is similar to research of Suryaningrum (2019), which revealed couples with a long marital duration in which the marriage is filled with many disputes will lead to an increased divorce rate.

The Cox proportional hazard model on survival data from marriage to divorce produces parameter estimates as shown in Table 4. Categorical variables are followed by a number after the dot (.) sign, e.g., x2.1, x2.2, and so on.

Table 4. Coefficient and Hazard Ratio of Cox PH Predictors

Variable	Coefficient (β)	Hazard Ratio	p-value	Variable	Coefficient (β)	Hazard Ratio	p-value
x_1	0.01	1.01	0.27	$x_{5.2}$ (senior high school)	-0.05	0.95	0.85
$x_{2.1}$ (junior high school)	**0.91**	**2.47**	**0.00**	$x_{5.3}$ (college)	0.03	1.04	0.87
$x_{2.2}$ (senior high school)	**0.77**	**2.16**	**0.00**	**$x_{6.1}$ (untrained workers)**	**0.58**	**1.79**	**0.01**
$x_{2.3}$ (college)	**0.87**	**2.38**	**0.00**	$x_{6.2}$ (skilled workers)	0.25	1.29	0.26
$x_{3.1}$ (untrained workers)	-0.10	0.90	0.66	$x_{6.3}$ (semi-professionals)	-0.00	1.00	0.98
$x_{3.2}$ (skilled workers)	-0.05	0.96	0.84	**x_7**	**-0.87**	**0.42***	**0.00**
$x_{3.3}$ (semi-professionals)	-0.21	0.81	0.38	$x_{8.1}$ (leaving duties)	0.69	1.99	0.28
x_4	-0.00	1.00	0.79	$x_{8.2}$ (disputes)	0.64	1.90	0.30
$x_{5.1}$ (junior high school)	-0.27	0.76	0.33	$x_{8.3}$ (moral crises)	0.55	1.73	0.39
Likelihood ratio test		202.7		d.f = 18	p-value = 0.001		

The likelihood ratio test to see the significance of the parameters simultaneously, as seen in Table 4, produces a test statistic of 202.7 with a p-value of 0.001. Such a small figure leads to the decision to reject H_0 where there is at least one variable that has a significant effect on the divorce rate.

Referring to Table 4, the test on parameters partially showed that five variables are significant to the divorce rate, as shown in bold. These variables are the plaintiff's education (junior high school, senior high school, and college), the defendant's occupation (untrained workers), and the number of children. Based on Table 4, the best model of Cox Proportional Hazard referring to Equation (7) is

$$h(t,\mathbf{x}) = h_0(t)\exp(0.91x_{2.1} + 0.77x_{2.2} + 0.87x_{2.3} + 0.58x_{6.1} + -0.87x_7)$$

The negative sign of the number of children (x7) in this model indicates that more children could lead to a lower hazard ratio.

The hazard ratio in Table 4, particularly that of significant variables, can be interpreted as the size of the influence of these variables on the divorce rate. For instance, the hazard ratio of the number of children, shown by the * sign, is 0.42. It means that increasing the number of children by one will reduce the divorce rate by 0.42 times. Thus, couples with more children might have had a longer marital duration.

Meanwhile, the hazard ratio for categorical variables, e.g., that of the defendant's occupation as an untrained worker, is 1.79. This figure shows that defendants who work as untrained workers have a divorce rate that is 1.79 times higher than that of those who work as professionals. Thus, the defendant's occupation could encourage the plaintiff to file for divorce. Table 5 summarizes the performance of the Cox PH model's overall predictors.

Table 5. Cox PH Model Performance

Indices	Value
C-index	21.70
Hazard ratio (HR)	0.15

Table 5 presents the hazard ratio, whose calculations are based on the group of divorce cases corresponding to their prognostic value. Based on the prognostic value, cases are grouped into high-risk and low-risk categories. As could be seen in the HR value, there is a difference between high-risk and low-risk divorce cases because the HR value is far from 1. The model of Cox PH produces a c-index of 21.70%, which means that the Cox PH model produces an appropriate prognostic sequence and survival time of 21.70%.

In addition to applying the semi-parametric method using Cox PH, divorce case data is also analyzed through SUR-SVM as a developed machine-learning approach. Equations (8), (9), (10), and (11) are simulated with real data, which have parameters γ and σ^2. Parameter γ which represents SUR-SVM is 1, while σ^2 which represents the Kernel parameter, is 0.05. Table 6 summarizes the performance of the SUR-SVM model's overall predictors.

Table 6. SUR-SVM Model Performance

Indices	Value
C-index	60.67
Hazard ratio	0.64

The hazard ratio (HR) and c-index value in Table 6 are 0.64 and 60.67, respectively. The HR figure of 0.64 indicates that the SUR-SVM model is better than the Cox PH in terms of classifying cases into risk groups. The c-index value of 60.67% indicates that the SUR-SVM model produces an appropriate prognostic sequence and survival time of 60.67%.

Given that the predictors did not contribute significantly as a whole to the duration of marriage-to-divorce, the c-index values of the Cox PH and SUR-SVM models after carrying out the feature selection using backward elimination are shown in Table 7.

Table 7. Contribution of Eliminated Predictor to the C-index of All Models

Eliminated Variable	C-index of Cox PH	C-index of SUR-SVM
x_1	21.73	60.19
x_2	21.41	60.86
x_3	21.69	60.66
x_4	21.70	61.36
x_5	21.83	60.38
x_6	21.31	60.16
x_7	17.02	59.86
x_8	21.70	60.38

Based on Table 7, it can be seen that the variable stirring the c-index to decrease the most if the variable is removed from the SUR-SVM model is the number of children (x_7). The difference is 0.81, yielded from 60.67 (in Table 6) – 59.86 (in Table 7). The number of children acts as the biggest factor that affects the sequence between marriage survival time and the prognostic index. Consecutively, the predictors contributing to the c-index of the SUR-SVM model are the defendant's occupation (x_6), the plaintiff's age at marriage (x_1), the cause of divorce (x_8), and the defendant's education (x_5). The elimination of other predictors that cause the c-index to increase indicates that corresponding predictors are not significant to the time to divorce, such as the plaintiff's education (x_2) and the defendant's age at marriage (x_4).

Using the Cox PH model as a comparison, Table 7 shows that the variable stirring the c-index to decrease the most if it is eliminated from the Cox PH model is the number of children. Besides the number of children, the predictors contributing to the c-index of the Cox PH model are the defendant's occupation and the plaintiff's education. Meanwhile, other predictors could be said to have no significant contribution to the c-index.

Referring to Table 7, the selected features for the SUR-SVM model are the number of children, the defendant's occupation, the plaintiff's age at marriage, the cause of divorce, and the defendant's education. The selected features for the Cox PH model are the number of children, the defendant's occupation, and the plaintiff's education. After both survival models use the selected features, their performance is summarized in Table 8.

Table 8. Summary of Survival Model

Predictors Selected	Indices	Cox-PH	SUR-SVM
All	C-index	21.70	60.67
	Hazard ratio (HR)	0.15	0.64
Feature selection	C-index	22.33	61.24
	Hazard ratio (HR)	0.19	0.54

Table 8 shows that the c-index of the SUR-SVM model after feature selection is 0.63 higher than that of the SUR-SVM overall predictors, yielding from 22.33 to 21.70. The c-index of Cox PH after feature selection is 0.57 higher than that of Cox PH overall predictors.

Based on Table 8, the SUR-SVM model is better than the Cox PH model. It could be seen from the c-index and hazard ratio of the SUR-SVM, which are higher than those of the Cox PH model. Thus, the post-feature selection SUR-SVM model can produce an appropriate prognostic sequence and time to divorce of 61.24%. It means that 61.24% of divorce cases are classified precisely by SUR-SVM in terms of the time sequence of events.

CONCLUSIONS AND SUGGESTIONS

Research about the application of the Cox PH model and SUR-SVM on the couples' time to divorce duration obtained results showing that the SUR-SVM had better performance than the semi-parametric Cox PH model. The application of feature selection in both models found that the predictors contributing to the c-index of SUR-SVM were the defendant's occupation, the plaintiff's age at marriage, the cause of divorce, and the defendant's education.

Meanwhile, the predictors contributing to the c-index of the Cox PH model were the defendant's occupation and the plaintiff's education. Social and economic factors in this case can affect the marital duration of the divorced couple in Palangka Raya. Further research being considered can be the comparison of marriage survival times prior to the COVID-19 pandemic and post-pandemic as modeled with several machine learning algorithms.

REFERENCES

Barman, M. P., Rahman, T., Bora, K., & Borgohain, C. (2020). Covid-19 pandemic and its recovery time of patients in india: A pilot study. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *14*(5). https://doi.org/10.1016/j.dsx.2020.07.004

Bramantoro, A., & Virdyna, I. (2022). Classification of divorce causes during the covid-19 pandemic using convolutional neural networks. *PeerJ Computer Science*, *8*. https://doi.org/https://doi.org/10.7717/peerj-cs.998

Faruk, A. (2018). The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data. *Journal of Physics: Conference Series*, *974*(1). https://doi.org/10.1088/1742-6596/974/1/012008

Fouodo, C. J. K., König, I. R., Weihs, C., Ziegler, A., & Wright, M. N. (2018). Support vector machines for survival analysis with R. *R Journal*, *10*(1). https://doi.org/10.32614/rj-2018-005

Kleinbaum, D. G., & Klein, M. (2012). *Introduction to survival analysis*. https://doi.org/10.1007/978-1-4419-6646-9_1

Mahkamah Agung Republik Indonesia. (2020). *Direktori putusan pengadilan agama palangkaraya*. https://putusan3.mahkamahagung.go.id/pengadilan/profil/pengadilan/pa-palangkaraya.html

Maulidina, F., Rustam, Z., Hartini, S., Wibowo, V. V. P., Wirasati, I., & Sadewo, W. (2021). Feature optimization using backward elimination and support vector machines (svm) algorithm for diabetes classification. *Journal of Physics: Conference Series*, *1821*(1). https://doi.org/10.1088/1742-6596/1821/1/012006

Mihaylov, I., Nisheva, M., & Vassilev, D. (2019). Application of machine learning models for survival prognosis in breast cancer studies. *Information (Switzerland)*, *10*(3). https://doi.org/10.3390/info10030093

Pengadilan Agama Palangka Raya. (2021). *Laporan tahunan*. https://pa-palangkaraya.go.id/laporan-tahunan/

Prastyo, D. D., Khoiri, H. A., Purnami, S. W., Suhartono, Fam, S.-F., & Suhermi, N. (2020). *Survival support vector machines: A simulation study and its health-related application*. https://doi.org/10.1007/978-3-030-22475-2_5

Riyadi, M. A. A., Prastyo, D. D., & Purnami, S. W. (2018). How svm can compensate logit based response label with various characteristics in predictor? A simulation study. *Proceedings of 2018 10th International Conference on Information Technology and Electrical Engineering: Smart Technology for Better Society, ICITEE 2018*. https://doi.org/10.1109/ICITEED.2018.8534762

Samsuri, S. (2018). Relevansi kedewasaan dalam pernikahan dengan upaya pencapaian tujuan hidup berkeluarga. *Hikmah: Journal of Islamic Studies*, *14*(1). https://doi.org/10.47466/hikmah.v14i1.101

Spooner, A., Chen, E., Sowmya, A., Sachdev, P., Kochan, N. A., Trollor, J., & Brodaty, H. (2020). A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-77220-w

Suryaningrum, N. (2019). Determinan perceraian di jakarta timur tahun 2014 (studi data pengadilan agama dan pengadilan negeri). *Forum Ilmu Sosial*, *46*(2).