BAREKENG: Journal of Mathematics and Its Applications
December 2024 Volume 18 Issue 4 Page 2471–2478
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss4pp2471-2478

MODELING EMPLOYEE RESIGNATION USING A SEMIPARAMETRIC APPROACH COX PROPORTIONAL HAZARD

Ni Wayan Widya Septia Sari[1], Ardi Kurniawan[2*], Elly Ana[3]

[1,2,3]*Mathematics Department, Faculty of Science and Technology, Universitas Airlangga*
Jl. Dr. Ir. H. Soekarno, Surabaya, 60116, Indonesia

*Corresponding author's e-mail: * **ardi-k@fst.unair.ac.id***

Article History:

Received: 4th, April 2024
Revised: 4th, June 2024
Accepted: 31st, July 2024
Published: 14th, October 2024

Keywords:

Cox Proportional Hazard;
Length of Work;
Semiparametric Survival Analysis;
Type III Censored.

ABSTRACT

Survival analysis is a research method that studies the duration individuals or experimental units endure against events like death, disease, recovery, or other experiences. This study employs a semi-parametric survival analysis model using the Cox proportional hazards regression method to identify factors such as age, gender, marital status, and education influencing how long employees stay with a company before resigning. The aim is to describe and interpret significant factors affecting employee resignation using the Cox Regression method. The results indicate that age significantly influences employee tenure. The average tenure is eight years. The probability of an employee still working at age 32 for up to eight years is 0.0057, while the likelihood for an employee who has worked more than eight years at age 32 is 0.9943. The study uses secondary data on the tenure of 521 employees, analyzed with the Cox proportional hazards regression method. The data, however, has limitations due to type III censoring, where some subjects leave observation, resulting in incomplete data. The study concludes that age significantly impacts employee tenure. Younger employees tend to explore career opportunities, while older employees seek stability, pension benefits, and a comfortable work environment.

How to cite this article:

N. W. W. S. Sari, A. Kurniawan and E. Ana., "MODELING EMPLOYEE RESIGNATION USING A SEMIPARAMETRIC APPROACH COX PROPORTIONAL HAZARD," *BAREKENG: J. Math. & App.*, vol. 18, iss. 4, pp. 2471-2478, December, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Employee resignation, a perennial challenge in the corporate landscape, profoundly impacts a company's productivity and incurs substantial costs associated with recruiting replacements [1]. As businesses navigate the multifaceted terrain of the 5.0 Industrial Revolution, implementing robust employee retention strategies becomes increasingly salient to foster operational stability and actualize organizational objectives [2]. The repercussions of elevated employee turnover reverberate far beyond mere financial implications, precipitating disruptions in company productivity and casting shadows upon the organization's reputation in the eyes of prospective job seekers [3]. Consequently, companies are compelled to redouble their efforts in employee retention to bolster productivity and lessen the harmful effects of frequent resignations [4]. Notably, external factors such as age and tenure assume pivotal roles in shaping employees' propensity to resign, exerting considerable influence on workforce dynamics [16]—for instance, PT. Zhi Sheng Indonesia in Sumedang grappled with alarmingly high turnover rates, necessitating urgent interventions to fortify the company's prospects and ensure its enduring viability. This research sample involved 57 respondents and used a questionnaire sent via the WhatsApp group and work application [5].

Survival analysis, recognized as event time analysis, stands as a robust methodology offering invaluable insights into the intricate interplay of factors shaping the timing of specific events for individuals or experimental units [6]. Distinguished by censored data, where observations of survival times are often limited by time constraints or other related factors, survival analysis presents a series of challenges in managing censored data and requires special statistical methods and opportunities to estimate survival functions, compare survival between groups, and identifying risk factors associated with certain outcomes that are unique to researchers [7]. Building on this foundation, previous inquiries have leveraged survival analysis techniques to delve into the resilience of the workforce, unraveling key determinants and furnishing actionable insights for strategic planning and intervention design [8]. At its core, survival data encapsulates the temporal journey until a predefined event materializes, with survival time serving as a pivotal metric delineating the temporal span until the occurrence of the event of interest [9]. By harnessing the rich insights afforded by survival analysis, researchers can navigate the complexities of time-dependent phenomena, facilitating a nuanced understanding of dynamic processes and informing evidence-based decision-making across diverse domains. Lawless (1982) conceptualizes the survival function as the probability of an individual enduring until time t [10]. This methodology facilitates a nuanced understanding of temporal dynamics, enabling organizations to proactively address challenges and optimize resource allocation for sustainable growth. If T is a random variable representing an individual's lifespan in the interval $[0, \infty)$, then the survival function $S(t)$ can be expressed as $S(t) = P(T \geq t) = \int_t^\infty f(t)dt$ [11]. Therefore, an equation arises that explains the relationship between the survival function and the cumulative distribution function, represented as $S(t) = 1 - F(t)$.

Censoring indeed holds a pivotal position in survival analysis, endowing it with distinctiveness compared to other statistical domains. Censored observations denote instances where the precise survival duration of an individual remains unknown, constrained by temporal limitations or other pertinent factors [12]. In the realm of this research, the overarching objective revolves around delineating employees' tenure utilizing a semiparametric approach via the Cox Proportional Hazard regression methodology. This statistical framework empowers the amalgamation of diverse predictor variables, encompassing dimensions such as age, gender, marital status, and education, among others, to discern their impact on the duration of an employee's tenure before opting to resign [13].

Collet (2003) elucidates the formulation for the Cox Proportional Hazard model of proportional failure, underpinning the methodology with robust theoretical foundations [14]. By harnessing this analytical framework, researchers can unveil nuanced insights into the dynamics governing employees' tenure, elucidating the interplay between various predictor variables and the propensity for resignation. Consequently, this holistic understanding enables organizations to devise targeted interventions aimed at enhancing employee retention and fostering organizational sustainability. Thus, the utilization of the Cox Proportional Hazard regression model emerges as a potent tool for unraveling the intricate tapestry of employee attrition dynamics within organizational settings.

$$h_i(t) = h_0(t) \exp\left(\beta_{1i}X_{1i} + \beta_{2i}X_{2i} + \cdots + \beta_{pi}X_{pi}\right) \quad (1)$$

with,

$h_i(t)$ = failure function of individual i at time t

$h_0(t)$ = basic failure function (basic hazard function)

$\beta_{1i}, \beta_{2i}, \ldots, \beta_{pi}$ = regression parameters

$X_{1i}, X_{2i}, \ldots, X_{pi}$ = value of predictor variable for individual i

Dummy variables serve as indispensable tools in regression analysis, facilitating the interpretation of results and enabling hypothesis testing, particularly for qualitative variables [15]. Their strategic incorporation allows researchers to capture the nuances of categorical predictors, thus enhancing the robustness of the analytical framework.

In alignment with this methodological rigor, the present study adopts a semiparametric approach leveraging the Cox Proportional Hazard regression methodology. By doing so, it endeavors to make a substantive contribution to elucidating the multifaceted determinants shaping employees' decisions to resign. In essence, this analytical framework empowers researchers to disentangle the intricate interplay between various predictor variables—ranging from demographic attributes such as age and gender to organizational factors like job satisfaction and leadership effectiveness—and the temporal dynamics governing employee tenure.

The overarching objective of this study extends beyond mere scientific research; it aspires to inform the design and implementation of targeted retention strategies tailored to the contemporary challenges posed by the increasingly competitive economic and business landscape. By unraveling the underlying drivers of employee attrition, organizations can proactively institute measures aimed at bolstering retention rates and fortifying their workforce stability.

Furthermore, the implications of this research extend beyond organizational boundaries, resonating with broader socio-economic imperatives encapsulated within the Sustainable Development Goals (SDGs). In particular, the findings of this study hold the potential to contribute to the realization of SDG 8, which pertains to fostering sustainable economic growth and promoting decent work for all. By fostering a deeper understanding of the factors influencing employee retention, this research endeavors to catalyze positive change, fostering environments conducive to both organizational prosperity and societal well-being.

2. RESEARCH METHODS

The research data used is data on employee resignations in 2021 obtained via the website https://www.kaggle.com/, totaling 521 which will be analyzed to determine the factors that cause employee resignations, cox regression analysis uses the help of SPSS 16 software. The steps taken to achieve the objectives of this research are:

a. Describe the characteristics of employees' length of service and the factors that influence the decision to resign using a method, namely by carrying out a descriptive statistical analysis which includes calculating the average, median, standard deviation, maximum value, and minimum value for continuous scale variables.
b. Modelling employees based on the Cox Proportional Hazard method with stages:
   (i) Classifying data related to response variables and predictor variables.
   (ii) Estimate the parameters of the Cox Proportional Hazard model of employees and estimate the baseline hazard function.
   (iii) Testing model parameters using the partial likelihood ratio test use statistic test and interpreting them with the predetermined hypothesis, namely:

$$X_{LR}^2 = -2\left[\ln \frac{L(0)}{L(\beta_i)})\right] \quad (2)$$

ln $L(0)$ is the ln partial likelihood of a model without predictor variables, while ln $L(\beta_i)$ or l_p is the ln partial likelihood of a model consisting of p predictor variables.

H_0 : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: there is at least one $\beta_k \neq 0$

(iv) Evaluate the model parameters individually using the Wald test use statistic test and interpret them with the following hypothesis:

$$z^2 = \left(\frac{\beta_j}{SE(\beta_j)}\right)^2 \tag{3}$$

with $SE(\beta_j) = \sqrt{var(\beta_j)}$

$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$

Where $k = 0,1, \ldots,5$

c. Carry out the goodness of fit testing to verify the proportional hazard assumption with the following hypothesis:

H_0: No relationship exists between Schoenfeld residual and survival time rating (proportional hazard assumption is met).

H_1: There is a relationship between Schoenfeld residual and survival time rating (proportional hazard assumption is not met).

With the critical region: reject H_0 if ρ-count $> \rho$-table.

d. Interpret the Cox Proportional Hazard model applied to the case of employees based on the results obtained from the model.

3. RESULTS AND DISCUSSION

3.1 Description of Sample Characteristics and Factors Suspected of Influence

The dataset utilized in this research draws from employee length of service data, comprising a total of 521 observations. **Table 1** provides comprehensive descriptive statistics for variables characterized by a continuous scale, shedding light on key attributes central to the study's analytical framework.

Table 1. Descriptive Statistics for Variables that have a continuous scale

Variable	Average	St.Dev	Median	Max	Min
Length of work	8.121	6.733	7	40	< 1
Age	32.461	7.815	32	58	18

Upon examination, it is evident that the average length of service among employees stands at 8.121 years, with a standard deviation of 6.733 years, underscoring the variability inherent within the dataset. Notably, the median length of service is reported at seven years, indicative of the central tendency within the distribution. The range of observed values spans from less than one year, denoting relatively recent hires, to a maximum tenure of 40 years, reflecting the presence of longstanding employees who have contributed significantly to the organization's journey.

The descriptive statistics unveil compelling insights into the demographic composition of the workforce. The dataset showcases a diverse age spectrum among employees, with an average age of 32.461 years and a standard deviation of 7.815 years. The median age of 32 years underscores the central tendency within the distribution, suggesting a balanced distribution of age groups. Notably, the age range spans from a minimum of 18 years, representing youthful entrants to the workforce, to a maximum of 58 years, reflective of seasoned professionals with extensive experience.

3.2 Modeling Employee Resignation Based on the Cox Proportional Hazard Method

Cox proportional hazard modeling will use the maximum partial likelihood method. Based on the testing, the results obtained for the model parameter estimator values for each predictor variable are as follows:

Table 2. Results of Parameter Estimation for the Cox Proportional Hazard Model

Variable	B	SE	Wald	Exp(B)	Sig	Conclusion
Age	-0.136	0.011	148.225	0.873	0.000	**Sig.**
Department			1.113		0.573	**No sig.**

Variable	B	SE	Wald	Exp(B)	Sig	Conclusion
Department (1)	0.300	0.358	0.703	1.350	0.402	**No sig.**
Department (2)	0.121	0.141	0.741	1.129	0.389	**No sig.**
Education			0.828		9.75	**No sig.**
Education (1)	0.456	0.640	0.507	1577	0.477	**No sig.**
Education (2)	0.002	0.267	0.000	1.002	0.993	**No sig.**
Education (3)	0.022	0.333	0.004	1.022	0.948	**No sig.**
Education (4)	-0.047	0.271	0.030	0.854	0.862	**No sig.**
Education (5)	-0.051	0.318	0.025	0.951	0.873	**No sig.**
Gender	0.077	0.115	0.454	1.081	0.501	**No sig.**
Marital status			1.538		0.464	**No sig.**
Marital status (1)	-0.075	0.157	0.225	0.928	0.635	**No sig.**
Marital status (2)	-0.159	0.128	1.531	0.853	0.216	**No sig.**

From **Table 2.** it can be inferred that the interim Cox Proportional Hazard model in this study is:

$$h_i(t) = h_0(t)\exp(-0.136\,\mathrm{X}_1 + 0.300\,\mathrm{X}_2(1) + 0.121\,\mathrm{X}_2(2) + 0.456\,\mathrm{X}_3(1) - 0.002\,\mathrm{X}_3(2) \\ -0.022\,\mathrm{X}_3(3) - 0.047\,\mathrm{X}_3(4) - 0.051\,\mathrm{X}_3(5) + 0.077\,\mathrm{X}_4 - 0.075\,\mathrm{X}_5(1) - 0.159\,\mathrm{X}_5(2) \quad (4)$$

Next, partial or Wald testing is employed to assess the significance of each explanatory variable in the model and its impact on the response variable. As indicated in **Table 2**, variables such as department (X_2), education (X_3), gender (X_4), and marital status (X_5) exhibit sig values $<$ 0.05, they suggest a statistically significant effect on the length of work. Consequently, these four variables do not significantly influence the duration of employment.

However, given the presence of variables still lacking significant impact, a retest is warranted, prioritizing the critical variable, age (X_1). The results of this retest involving the predictor variable age (X_1) are presented in **Table 3** below, elucidating the nuanced relationship between age and length of work tenure. Through this iterative process, researchers endeavor to refine the model and enhance its predictive capacity, thereby fostering a deeper understanding of the determinants shaping employee tenure dynamics within organizational settings.

Table 3. Significant β Parameter Estimation Results

Variable	B	SE	Wald	Exp(B)	Sig	Conclusion
Age	-0.136	0.011	152.725	0.873	0.000	Sig.

From these results, it can be concluded that the best Cox Proportional Hazard model in this study is:

$$h_i(t) = h_0(t)\exp(-0.136\,X_1) \quad (5)$$

To ensure the accuracy of this model, this model is 4.697 better at explaining work experience data than the temporary model. The next step is to test the Proportional Hazard assumption for the age variable (X_1). This test is critical to verify that the age variable's risk ratio remains constant. With the following hypothesis:

H_0 : No correlation exists between the Schoenfeld residual age variable and the rank survival time.
H_1 : There is a correlation between the Schoenfeld residual age variable and the rank survival time.

Based on the obtained results, the analysis reveals a p-value of 0.684 for the age variable, surpassing the significance threshold of 0.05. Consequently, the null hypothesis (H_0) is accepted, indicating no statistically significant correlation between the Schoenfeld residual age variable and the rank survival time. Moreover, this finding suggests that the proportional hazard assumption is met, underscoring the stability of the hazard ratios over time.

After testing the Proportional Hazard assumption for the age variable (X_1), it was found that there was no correlation between the Schoenfeld residual age variable and the rank survival time, and the Proportional Hazard assumption was met. The next step is to calculate the essential probability survival function and cumulative distribution function to calculate the chances of employees still working or resigning at certain times. Before carrying out these calculations, parameter estimates, hazard estimates, and basic survival are required as follows:

Table 4. Basic Hazard and Survival Estimation

Long working	Hazard basic	Survival basic
<1	1.133	0.986
1	11.098	0.873
2	15.406	0.828
3	19.425	0.788
4	23.277	0.752
5	33.407	0.664
6	46.943	0.562
7	56.318	0.514
8	66.080	0.445
9	73.074	0.408
10	77.160	0.388
11	78.527	0.382
13	80.609	0.372
14	83.143	0.361
15	86.401	0.346
18	92.807	0.320
21	103.294	0.282
36	1690.770	0.000

Thus, the primary hazard and survival estimation results are obtained in Table 4 based on Equation (3). The next step is to calculate the chances of employees who are still working or have resigned at various times.

As an application, if you have data on employees who have worked for more than eight years and are 18 years old, then calculate the probability of employees who are still working ($S(t,X)$) and the probability of employees who have resigned ($F(t,X)$) as follows:

$$S(8|18) = S_0(8)^{\exp\,(-0.136\cdot 18)} = 0.445^{-2.448} = 0.0385$$

$$F(8|18) = 1 - S(8|18) = 1 - 0.0385 = 0.9615$$

The calculation results show that the probability that an employee who is 18 years old is still working for more than eight years is 0.0385 and that an employee who is 18 years old has resigned for up to 8 years is 0.9615.

For the second scenario, wherein an employee has accrued up to 8 years of tenure and is 26 years old, the probability of continued employment can be computed using the survival function $S(t,X)$ while the likelihood of resignation can be estimated using the cumulative distribution function $F(t,X)$. The calculation results facilitate a nuanced understanding of employee retention and resignation probabilities at different time intervals and demographic profiles, thereby informing strategic decision-making to bolster workforce stability and organizational resilience.

$$S(8|26) = S_0(8)^{\exp(-0.136\cdot 26)} = 0.445^{-3.536} = 0.0130$$

$$F(8|26) = 1 - S(8|26) = 1 - 0.0130 = 0.9870$$

The calculation results show that the probability that an employee who is 26 years old is still working for more than eight years is 0.0130 and that an employee who is 26 years old has resigned for up to 8 years is 0.9870.

For the third example, if you have data on employees who have worked for more than eight years and are 32 years old, then calculate the probability of employees who are still working ($S(t,X)$) and the probability of employees who have resigned ($F(t,X)$) as follows:

$$S(8|32) = S_0(8)^{\exp\,(-0.136\cdot 32)} = 0.445^{-4.352} = 0.0057$$
$$F(8|32) = 1 - S(8|32) = 1 - 0.0057 = 0.9943$$

The calculation results show that the probability that an employee aged 32 years will still work for more than eight years is 0.0057, and an employee aged 32 years will have resigned for more than eight years is 0.9943.

Table 5. **Estimated Probabilities of Employees Still Working $S(t, X)$ and Resigning $F(t, X)$ in Various Years**

AGE (X)	$S(8, X)$	$F(8, X)$
18	0.0385	0.9615
19	0.0336	0.9664
20	0.0293	0.9707
21	0.0256	0.9744
22	0.0223	0.9777
23	0.0195	0.9805
24	0.017	0.983
25	0.0149	0.9851
26	0.013	0.987
27	0.0113	0.9887
28	0.0099	0.9901
29	0.0086	0.9914
30	0.0075	0.9925
31	0.0066	0.9934
32	0.0057	0.9943
33	0.005	0.995
34	0.0044	0.9956
35	0.0038	0.9962
36	0.0033	0.9967
37	0.0029	0.9971
38	0.0025	0.9975
39	0.0022	0.9978
40	0.0019	0.9981
41	0.0017	0.9983
42	0.0015	0.9985
43	0.0013	0.9987
44	0.0011	0.9989
45	0.001	0.999
46	0.0009	0.9991
47	0.0007	0.9993
48	0.0007	0.9993
49	0.0006	0.9994
50	0.0005	0.9995
51	0.0004	0.9996
52	0.0004	0.9996
53	0.0003	0.9997
55	0.0003	0.9997
56	0.0002	0.9998
58	0.0002	0.9998

Based on the results of data analysis, it can be shown that being in a productive age does not guarantee that an employee will stay in a particular company because, in a productive age, it can happen to employees who want to explore areas of work and look for higher salaries. In contrast, employees choose to resign from particular companies as they age, it could be due to working conditions such as high-pressure meltdowns every day, even on weekends, as is often the case in the surrounding environment. There are also motivational factors, job satisfaction, and personal factors such as feeling unused and not having opportunities to advance your career.

4. CONCLUSIONS

The data analyzed in this research consists of observations from 522 employees obtained from the Kaggle site, focusing on factors contributing to employee resignations. The majority of employees were aged between 18 and 32 years (49%), worked in the R&D department (65%), had education in the field of life sciences (44%), were male (61%), and were married (44%). The analysis aims to calculate the duration of employment until resignation based on the length of work, providing insights into when employees tend to resign.

The results of temporary model testing show that age significantly influences the survival time of employee resignation with a P-value of 0.000. So, the best model for data on employee length of service up to 8 years using the Cox proportional hazards regression method is:

$$h_i(8) = 66.080\ \exp(-0.136\ X_1)$$

This means that the estimated probability of an employee who is still working for up to 8 years and is 32 years old is 0.0057, while the likelihood of an employee who has worked for more than eight years and is 32 years old is 0.9943. Based on the results of data analysis, it can be shown that being in a productive age does not guarantee that an employee will stay in a particular company because, in a productive age, it can happen to employees who want to explore areas of work and look for higher salaries. In contrast, employees choose to resign from particular companies as they age, it could be due to working conditions such as high-pressure meltdowns every day, even on weekends, as is often the case in the surrounding environment. There are also motivational factors, job satisfaction, and personal factors such as feeling unused and not having opportunities to advance your career.

REFERENCES

[1] G. Syahronica, "The influence of job satisfaction and work stress on turnover intention (Study on employees of the fantasy world department of PT Pembangunan Jaya Ancol, Tbk)," *Journal of Business Administration*, vol. 20, no. 1, 2015.

[2] N. Gunaprasida and A. Wibowo, "The effect of work-family conflict and flexible work arrangements on turnover intention: do female and male employees differ?," *Journal of Business Strategy*, vol. 23, no. 1, pp. 27–36, 2019.

[3] A. Setiawan and R. Pratiwi, "The impact of high employee turnover on organizational productivity and employer branding: A study in the retail industry," *Journal of Management Studies*, vol. 35, no. 2, pp. 67-82, 2022.

[4] B. Susanto and R. Utami, "The role of employee retention strategies in enhancing organizational productivity: A case study in the technology industry," *Journal of Human Resource Management*, vol. 30, no. 3, pp. 78-92, 2021.

[5] M. Ahyar and Y. Maulana, "Factors that Influence PT Turnover Intention. Zhi Sheng Indonesia Sumedang Area," *Journal of Business Disruption*, vol. 5, no. 2, pp. 100-114, 2022.

[6] B. Suryadi and A. Wibisono, "The application of survival analysis in predicting customer churn is a case study in the telecommunications industry," *Journal of Data Science and Analytics*, vol. 15, no. 3, pp. 45-58, 2021.

[7] E. T. Lee, "*Statistical Methods for Survival Data Analysis*," John Wiley & Sons, 2003.

[8] A. Brahmantyo et al., "Survival Analysis with the Cox Regression Method," Ph.D. dissertation, Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Indonesia, 2015.

[9] R. Sari and E. Susanto, "Understanding loan default prediction using survival analysis: A case study in the banking industry," *Journal of Financial Analytics*, vol. 15, no. 3, pp. 67-80, 2021.

[10] J. F. Lawless, "*Statistical Models and Methods for Lifetime Data*," John Wiley & Sons, 1982.

[11] D. Collett, "*Modelling survival data in medical research*," CRC Press, 2023.

[12] D. Maruddani et al., "Censored Data Analysis in Health Research," *Journal of Public Health*, vol. 9, no. 2, pp. 123-135, 2021.

[13] H. Latan, "*Application of Statistical Data Analysis for Social Sciences and Science with IBM SPSS*," Alphabeta, 2014.

[14] D. Collett, "*Modelling Survival Data in Medical Research*," Chapman & Hall, 2003.

[15] D. Prasetyo and E. Santoso, "Analyzing the impact of training programs on workforce resilience: A survival analysis perspective in the service industry," *Journal of Training and Development*, vol. 30, no. 2, pp. 67-8, 2022.

[16] Wahyunur and E. Santoso,2022. [Online]. "*Analysing the impact of workplace environment and leadership style on employee loyalty at the Bone Arasoe Sugar Factory*," Skripsi. Available: http://repository.unhas.ac.id/id/eprint/22940/2/A021181028_skripsi_27-10-2022%201-2.pdf. [Accessed 30 Maret 2024].