Journal of Economic Statistics 2023 1 (3) 1-15

Journal of Economic Statistics

Homepage: https://www.anserpress.org/journal/jes

Machine Learning Survival Models restrictions: the case of startups time to failed with collinearity-related issues

Diego Vallarino [a,*]

[a] *Independent Researcher, Spain*

ABSTRACT

This research evaluates the efficacy of survival models in forecasting startup failures and investigates their economic implications. Several machine learning survival models, including Kernel SVM, DeepSurv, Survival Random Forest, and MTLR, are assessed using the concordance index (C-index) as a measure of prediction accuracy. The findings reveal that more sophisticated models, such as Multi-Task Logical Regression (MTLR) and Random Forest, outperform the standard Cox and Kaplan Meier (K-M) models in terms of predicted accuracy.

KEYWORDS

Startups Failure; Survival Analysis; Survival Machine Learning Models; Concordance Index

* Corresponding author: Diego Vallarino
E-mail address: diego.vallarino@gmail.com

ISSN 2972-3728
doi: 10.58567/jes01030001

Received 18 August 2023; Accepted 10 November 2023; Available online 1 December 2023

1. Introduction

Forecasting corporate failure is crucial from both an economic and societal standpoint. bankruptcies affect the stability of the corporate environment, making assessing the sustainability of partners, consumers, and other stakeholders a particularly difficult and crucial problem for business players.

There are currently a huge number of bankruptcy prediction models (Adnan Aziz, M and Dar, Humayon A, 2006), however almost all of them are classification based, which means they can estimate the posterior probability that a certain firm would fail based on its financial parameters. The expected time to failure is not explicitly considered. For example, if a classification model is based on data gathered one year prior to failure, the model's output is the posterior probability that a certain firm would fail within one year. Decisions based on this likelihood may not be done in time to prevent a failure that happens in less than a year.

A survival analysis, on the other hand, is concerned with the time of occurrence of the event of interest. Despite its prominence in medical and technical areas, survival analysis is seldom used to anticipate financial failure. Adnan et al. (2006) considered 12 types of classification models in their study of bankruptcy prediction models (ranging from discriminant analysis and logit to case-based reasoning, neural networks, and rough sets), but did not address survival analysis. According to this publication, the most often used methodologies are multiple discriminant analysis and logistic regression; these two models account for more than half of the publications evaluated. Alaka et al. (2018) identified eight prevalent technologies, including two statistical techniques (multiple discriminant analysis and logistic regression) and six machine learning models.

As a consequence, we may conclude that survival analysis is not a key focus of financial failure prediction specialists. The purpose of this study is to determine the utility of survival analysis (SA) in predicting bankruptcy. SA models and classification procedures are divided into two types: statistical and machine learning based. Statistical SA models first appeared in the early 1970s, but machine learning SA models are the result of more recent research. A vast amount of research demonstrates that machine learning models outperform statistical models in classification and regression tasks, notably in classification-based bankruptcy prediction (Barboza, Flavio and Kimura, Herbert and Altman, Edward, 2017). Several studies provide comparable conclusions on the superiority of machine learning technology in various areas of survival analysis.

Despite these results, most authors of bankruptcy prediction systems, particularly when utilizing SA, use the most basic statistical models (Cox, Raymond AK and Kimmel, Randall K and Wang, Grace WY, 2017)

In this work, we examine the outcomes of our model comparison and their economic interpretation. Our investigation focuses on the efficacy of several models in forecasting startup failures using a collection of important factors. We examine the prediction power of multiple machine learning survival models, including the Kernel SVM, DeepSurv, Survival Random Forest, and MTLR models. We utilize the concordance index to compare various machine learning techniques (C-index)

Our objective is to determine which model delivers the most accurate and informative forecasts of startup failures, as well as to understand the economic importance of the model's findings. To do so, we analyze the relevance and size of the estimated coefficients for each variable in the model and compare the findings to economic theory and intuition.

By studying the outcomes of our model comparison and the economic interpretation of these results, we seek to gain insights into the variables that lead to startup failures and a better understanding of how alternative models might be used to anticipate these failures.

The rest of the paper is organized as follows. Following a theoretical viewpoint on the survival analysis models used in our work, we perform a brief evaluation of articles that utilize survival analysis to answer the financial collapse problem. Following that, we discuss the empirical analysis part, which contains the models utilized, the data source, and the assessment measures.

The analytical results are then presented, including a comparison of the various models, and the economic implications of our findings are discussed. Finally, we end the study by summarizing the important results and their implications for future research and policymaking.

Overall, our work adds to the literature on the use of survival analysis in finance and gives insights into the characteristics that cause financial crashes, which may help policymakers build more effective policies to avoid similar disasters in the future.

2. Theorical perspective

Signals representing a company's operational health may reveal signs of financial distress, which may then be integrated into prediction models. (Beaver, William H, 1966) was the first to foresee bankruptcy using financial ratios, and financial ratios have been the most essential piece of information in predicting financial difficulty for decades (Ohlson, James A., 1980).

Market-based information may offer us with a timely prediction; that is, under the assumption of efficient markets, the market price includes all future perspectives (Bharath, Sreedhar T and Shumway, Tyler, 2008). Corporate governance and corporate efficiency Li et al. (2021), external resource concerns (Hu, Dan and Zheng, Haiyan, 2015), and macroeconomic difficulties are all important factors to address (Tinoco, Mario Hernandez and Wilson, Nick, 2013).

Furthermore, unstructured data has attracted a lot of interest in business research in recent years. In some works (Mai, Feng and Tian, Shaonan and Lee, Chihoon and Ma, Ling, 2019) the authors used textual data to predict corporate insolvency, while others (Hosaka, Tadaaki, 2019) used image data generated from financial papers to forecast firm bankruptcy utilizing convolutional neural networks in information extraction. Statistical analysis and data mining methodologies have been applied in bankruptcy and financial distress prediction studies to enhance decision-making tools (Yang, Zijiang and You, Wenjie and Ji, Guoli, 2011). Altman (1968) pioneered the use of multiple discriminant analysis (MDA), which was further developed upon by Deakin (1972), and others.

Later, logistic regression (or Logit) replaced the Z-score as a Basel II criteria since it may yield probabilistic results (Ohlson, James A., 1980). Since the late twentieth century, machine learning algorithms have appeared in the literature. Tam et al. (1992) and Jabeur (2023) used neural networks to classify both bankruptcy and non-bankruptcy publicly traded enterprises.

Other novel algorithms include genetic algorithms, rough sets, decision trees, support vector machines (Lin, Wei-Yang and Hu, Ya-Han and Tsai, Chih-Fong, 2011). Mathematical programming is another kind of algorithm. Data envelopment analysis (DEA) is a nonparametric technique for evaluating businesses and determining relative efficiency based on the distance to the ideal frontier. DEA has been used to anticipate insolvency and financial problems assessed model discussions Emrouznejad at al. (2018) and Henriques et al. (2020).

While the previous methodologies treat financial difficulty as a categorization problem, a survival analysis approach is concerned with both the time and the occurrence of the event. Survival analysis, like static classification approaches, may benefit from time-varying variables and censoring in models. Luoma and Laitinen (1991) used Cox proportional hazard models to anticipate the failure of Finnish industrial and retail firms, although they were proven to be considerably inferior to both discriminant and logit analysis.

Shumway (2021) develops a discrete-time bankruptcy hazard model using accounting and market data. Because of the advantages in parameter computation and the kind of variables reported on a regular basis for businesses, De Leonardis and Rocci (2014) employed the discrete hazard model.

In terms of prediction accuracy, Gepp and Kumar (2008) observed that the Cox model was comparable to discriminant analysis and logistic regression at equal misclassification costs but poorer when compensating for higher Type I error costs. Kristanti and Herwany (2017) found promising results utilizing survival analysis on struggling Indonesian businesses. Recurring event data are often employed in medical research, particularly in the study of epilepsy, asthma, heart attacks, and hospital admissions Alhurani et al. (2022). Within-subject correlation is a critical aspect of recurrent event data, in which one event increases or decreases the likelihood of subsequent occurrences (Box-Steffensmeier and Janet M and De Boef, Suzanna, 2006).

Traditional statistical approaches, such as logistic regression and Cox proportional hazards regression, either neglect or fail to account for within-subject correlation, resulting in an inaccurate calculation of standard errors and a divergence from the original research topic (Twisk, Jos WR and Smidt, Nynke and de Vente, Wieke, 2005). Many methods for assessing recurrent events that take into account all available information and within-subject correlations have been proposed. Based on various definitions of risk sets, marginal intensity approaches allow all cases to be at risk for each repeated event (Wei, Lee-Jen and Lin, Danyu Y and Weissfeld, Lisa, 1989), whereas conditional intensity models are estimated in elapsed time or gap time, and cases are designated at risk for the kth repeated event only after experiencing the (k-1)th event (Chang at al. 1999; Zhou et al. 2022).

In the Andersen-Gill (AG) model (Andersen, Per Kragh and Gill, Richard, 1982), repeating events are believed to be ordered yet have an equal probability of occurring. According to the Prentice, Williams, and Peterson (PWP) model (Prentice, Ross L and Williams, Benjamin J and Peterson, Arthur V, 1981), a person is not at risk for a future event until the preceding event happens. Despite the fact that there is a substantial body of literature on modeling recurrent events using the PWP model in the fields of medicine (Ejoku, Jonatha, 2020), consumer behavior (Bijwaard, Govert E and Franses, Philip Hans and Paap, Richard, 2006), and (Zhou, Fanyin and Fu, Lijun and Li, Zhiyong and Xu, Jiawei, 2022).

In their case, Zhou et al. (2022) have used the Cox analysis to study financial distress. In this work they have developed 3 different models where each one contains different variables and he seeks to be able to understand, through a single survival model, which are the factors that most explain financial distress. There is no comparison between these Cox models and other survival machine learning models.

Corporate finance studies are few and far between. Parker et al. (Parker, Susan and Peters, Gary F and Turetsky, Howard F, 2005), for example, used the Cox and PWP models to investigate the impact of corporate governance features on auditors repeated going-concern ratings of failed firms. (Wang, Yuling and Carson, James M, 2010) examined insurers' recurring rating changes using the PWP model. (Godlewski, Christophe J., 2015) used the PWP model to explore the variables affecting debt contract renegotiations between startups and European firms in the context of corporate loans.

3. Empirical Analysis

While conducting our research, we gathered information from a variety of data sources to get insights into the characteristics and dynamics of small and medium-sized businesses (SMEs) in the United States. In our investigation, we utilized data from a variety of publicly available sources, including but not limited to the U.S. Census Bureau, Bureau of Labor Statistics, Small Business Administration, and Federal Reserve. The study of startup failures is important in finance and economics because it has substantial consequences for financial stability and the economy as a whole. This section describes the models utilized in our investigation, as well as the data source and assessment measures.

3.1. Models

3.1.1. Cox Proportional Hazards Model (coxph)

The Cox proportional hazards model is a widely used semi-parametric model in survival analysis. It assumes that the hazard function can be represented as the product of a time-independent baseline hazard function and a time-varying covariate function. Mathematically, the model can be represented as:

$$h(t|x) = h_0(t)\exp(\beta^T x)$$

where $h(t|x)$ is the hazard function for a given time t and covariate values x, $h0(t)$ is the baseline hazard function, β is a vector of regression coefficients, and $exp(\beta X)$ is the hazard ratio, which represents the change in hazard associated with a unit change in the covariate.

3.1.2. Multi-Task Logistic Regression (MTLR)

Multi-task logistic regression is a machine learning method that can be used for survival analysis. It is a multi-output learning algorithm that can predict the probability of an event occurring at different time points. Mathematically, the model can be represented as:

$$h(t|x) = exp\left(\Sigma_{k=1}^{K}\Sigma_{j=1}^{p}\beta_{kj}x_{kj}\right)$$

Where $h(t|x)$ is the hazard rate for an individual with covariates x , β_{kj} are the regression coefficients for the kth characteristic of the jth group, and x_{kj} is the kth feature of the jth group.

3.1.3. Kernel Support Vector Machine (Kernel SVM)

Kernel support vector machines are a popular machine learning method for survival analysis. They can handle non-linear relationships between covariates and outcomes by projecting the data into a higher-dimensional space using a kernel function. The model can be represented as:

$$f(x) = sign(\Sigma_{i=1}^{n}\alpha_i y_i K(x_i, x) + b)$$

Where $K(x_i, x)$ is a kernel function that measures the similarity between the feature vectors x_i and x, y_i is the class label of the i-th instance, α_i are the weights of the support vectors and b is the bias.

3.1.4. Random Survival Forest

Random survival forests are an extension of random forests for survival analysis. They use an ensemble of decision trees to predict the survival function. The model can be represented as:

$$h(t|x) = (1/B)\Sigma_{b=1}^{B}h_b(t|x)$$

Where $h_b(t|x)$ is the hazard rate for an individual with covariates x in the bth decision tree and B is the number of trees in the random forest.

3.1.5. DeepSurv

DeepSurv is a deep learning model for survival analysis. It uses a neural network with a flexible architecture to predict the survival function. The model can be represented as:

$$h(t|x) = exp\left(\Sigma_{i=1}^{p}\beta_i f_i(x) + g\big(h_\theta(x)\big)\right)$$

Where $h(t|x)$ is the hazard rate for an individual with covariates x and β_i are the regression coefficients for the input features $f_i(x), g(\cdot)$ is a non-linear function that transforms the output features and $h_\theta(x)$ is a neural network with θ parameters.

3.2. Data

We collected information from numerous data sources to get insights into the characteristics and dynamics of small and medium-sized firms (SMEs) in the United States while conducting our study. Several publicly accessible data sources, including but not limited to the U.S. Census Bureau, Bureau of Labor Statistics, Small Business Administration, and the Federal Reserve, were used in our study. We have gotten a complete insight of the major qualities and trends within the SME sector by using the strength of these data sources, research papers, publications, working papers, among various other documents.

Furthermore, we chose and emphasized on 10 critical features to replicate a representative database based on an intensive review of relevant literature. We employed a data simulation model based on an algorithm built specifically for this study, and these features, together with a *state* variable and a *time* variable, were used to represent the typical behavior of startups in the United States. It is vital to note that censored data was not taken into account in the settings used to simulate the database of 1000 signatures and 10 variables, in a period of 10 years.

The 10 main variables are the following: *Sales, Investments, Employees, Expenses, Innovation, Competition, Marketing, Profitability, FounderAge, and ProductQuality.* We can describe them as follows:
Sales: Sales refers to the revenue generated by a startup through the sale of its products or services. It represents the financial success and market demand for the startup's offerings. Investments: *Investments* refer to the capital or funds that are injected into a startup by external investors or stakeholders. These funds are typically used to fuel the growth and expansion of the startup.

Employees: Employees represent the workforce or personnel working for the startup. The number and quality of employees can greatly impact the startup's productivity, efficiency, and ability to innovate. *Expenses*: Expenses refer to the costs incurred by a startup in running its operations, including salaries, rent, utilities, marketing costs, and other overhead expenses. Monitoring and managing expenses is crucial for maintaining financial stability. *Innovation*: Innovation represents the ability of a startup to develop and introduce new ideas, products, or processes. Emphasizing innovation is important for startups to stay ahead of the competition and drive growth.

Competition: Competition refers to other companies or startups operating in the same industry or offering similar products or services. Understanding the competitive landscape helps a startup identify market opportunities, differentiate itself, and develop effective marketing and sales strategies. *Marketing*: Marketing encompasses the activities undertaken by a startup to promote its products or services, attract customers, and build brand awareness. Effective marketing strategies can drive sales and help a startup gain a competitive edge.

Profitability: Profitability measures the financial success of a startup by comparing its revenues to expenses. A startup needs to achieve and maintain profitability to sustain its operations, attract investors, and support future growth. *Founder Age*: Founder age refers to the age of the individuals who started the startup. The age and experience of the founders can influence the decision-making process, leadership style, and industry connections, which in turn can impact the startup's success. *Product Quality*: Product quality reflects the level of excellence or value provided by a startup's products or services. Delivering high-quality offerings is crucial for attracting and retaining customers, building a strong reputation, and gaining a competitive advantage.

Time: Time represents the duration of a startup's existence or the stage of its development. The time factor influences various aspects such as market conditions, customer preferences, and industry trends, which can impact the startup's strategies and decisions. *Status*: Status refers to the current situation of the startup, if the startup is alive or not at the time of analysis. It is a variable that only takes the values 0 or 1, depending on whether the company stopped to exist in the market, or if the company continues to exist, developing its common functions.

3.3. Metrics

3.3.1. C-Index

The C-index (also known as the concordance index or the area under the receiver operating characteristic curve) is a widely used metric in survival analysis and medical research to assess the performance of predictive models that estimate the likelihood of an event occurring over a given time period.

The C-index is generated using the rankings of anticipated event occurrence probability for each participant in a dataset. It calculates the percentage of pairings of people in whom the person with the higher anticipated probability experienced the event before the person with the lower projected probability. In other words, it assesses a predictive model's capacity to rank people in order of their likelihood of experiencing the event of interest.
The C-index scales from 0 to 1, with 0.5 representing random prediction and 1 indicating perfect prediction. In medical research, a C-index value of 0.7 or above is considered satisfactory performance for a prediction model. Here is the formula of non-censored data C-Index.

$$C - index = \frac{\Sigma_{\mathrm{ij}}\, 1_{T_j<T_i} \cdot 1_{\eta_j>\eta_i} \cdot \delta_j}{\Sigma_{\mathrm{ij}}\, 1_{T_j<T_i} \cdot \delta_j}$$

$$\eta_i, \, the \, risk \, score \, of \, a \, unit \, i$$

$$1_{T_j<T_i} = 0 \quad if \;\; T_j < T_i \;\; else \; 0$$

$$1_{\eta_j<\eta_i} = 0 \quad if \;\; \eta_j < \eta_i \;\; else \; 0$$

$$\delta_j, \, represents \, whether \, the \, value \, is \, censored \, or \, not$$

4. Results

The Kaplan-Meier curve displays the survival probability over time for a group of startups. The x-axis shows the time, and the y-axis displays the survival probability. At the start of the observation period, all startups are assumed to be "alive," represented by the value of 1. Over time, some startups may "die," meaning they fail, and their survival probability decreases.

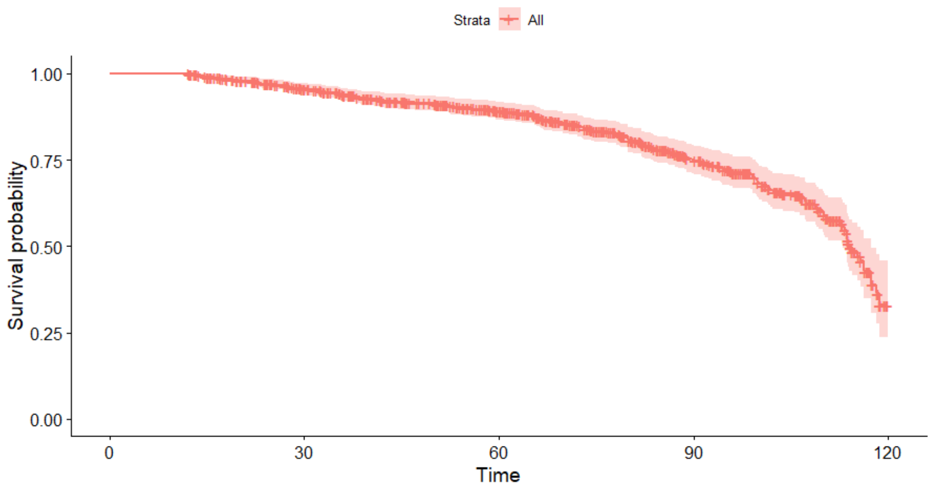

Figure 1. Kaplan-Meier survival curve.

Source: own elaboration.

The table below shows the evolution of the risk of startup failures over time. The analysis reveals intriguing patterns in startup survival. At the beginning of the observation period (time = 12 months), all 700 startups in the sample (*data.train*) were at risk, and none had failed. The estimated survival probability was 1.000, indicating a 100% likelihood of survival.

As time progressed, the number of startups at risk decreased, accompanied by an increase in the number of failures. This resulted in a gradual decline in survival probabilities. At time 24 months, the survival probability stood at 0.970, indicating that approximately 97% of the startups were expected to have survived up to that point.

Table 1. Kaplan-Meier survival probabilities (survival) at different time points.

Call: surfit (formula = Surv (time, status) ~ 1, data = data. train						
time	n. risk	n. event	survival	std. error	lower 95% CI	upper 95% CI
12	700	0	1.000	0.00000	1.000	1.000
24	627	20	0.970	0.00657	0.857	0.983
36	557	20	0.938	0.00956	0.919	0.957
48	493	13	0.915	0.01123	0.893	0.937
60	430	12	0.891	0.01283	0.867	0.917
80	287	34	0.808	0.01801	0.773	0.844
100	150	34	0.698	0.02481	0.640	0.737
110	85	16	0.597	0.03026	0.541	0.659

Source: own elaboration.

Further analysis demonstrated a continuous decrease in survival probabilities. At time 36 months, the survival probability dropped to 0.938, suggesting a decline in the likelihood of survival to 93.8%. The downward trend continued, and at time 48 months, the survival probability was 0.915, indicating a further reduction to 91.5%. The standard errors associated with the estimates were relatively small, suggesting precise estimates of the survival probabilities. Confidence intervals provided additional insights, indicating the likely range within which the true survival probabilities fell. The intervals tended to narrow as time progressed, indicating increased precision in the estimates.

The results highlight the changing risk landscape for startups over time. The decreasing survival probabilities suggest a heightened risk of failure as startups mature. This underlines the challenges faced by entrepreneurs in sustaining their ventures and the importance of strategic decision-making.

Understanding the time-to-failure patterns and associated risks can aid stakeholders in evaluating investment opportunities, designing support mechanisms, and formulating policies to foster startup resilience. Moreover, the precise estimates obtained from the analysis offer valuable insights for entrepreneurs seeking to optimize their strategies and mitigate potential pitfalls.

4.1. Model comparison

The paper analyzed the performance of different machine learning survival models in predicting startup failures using a set of relevant variables. This procedure divided the dataset into a training set and a testing set for machine learning design. The code randomly selects 70% of the rows from the data frame *df* and assigns them to *data.train*. The *train_index* variable stores the numeric row indices of *data.train*. The remaining rows, which constitute 30% of the original data, are assigned to *data.test*. This separation allows for training a model on the training set and evaluating its performance on the testing set to assess its effectiveness and generalization capabilities. The concordance index (C-index) was used to compare the predictive power of different models.

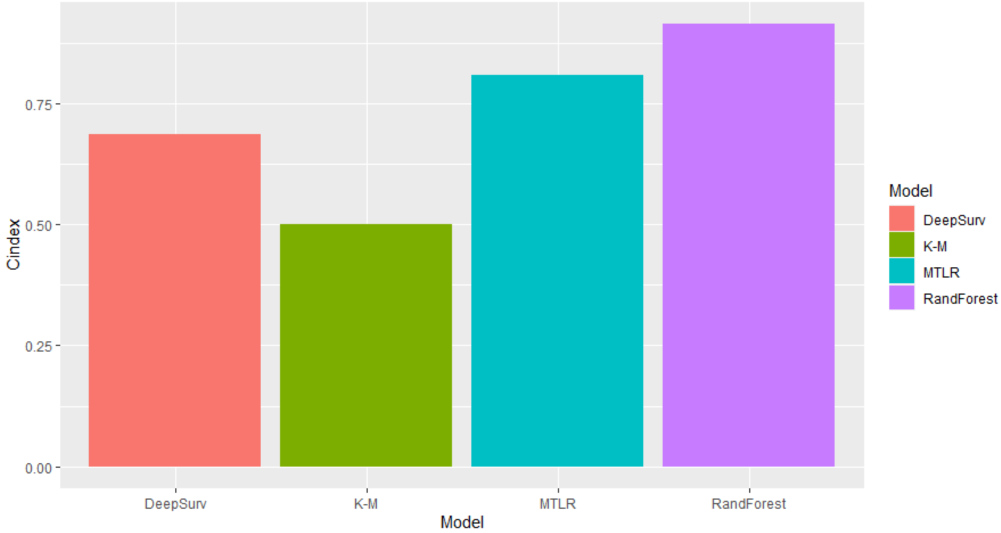

Figure 2. results from different machine learning models.

Source: own elaboration

The comparison research produced intriguing results about the efficacy of the survival models for predicting startup failure. The C-index findings showed that the models had various degrees of prediction accuracy. The C-index for the Kaplan-Meier (K-M) model was 0.500, suggesting random or poor prediction ability. The model's simplicity and assumption of predictor independence may restrict its capacity to reflect the intricacies of startup failure.

The Multi-Task Logistic Regression (MTLR) model performed better in terms of prediction, with a C-index of 0.810527. This model contains numerous variables and takes into account their interdependence, making it more suited to capture the multidimensional character of startup failure. With a C-index of 0.915297, the Random Forest model beat the prior models. This model displayed a greater predictive accuracy by employing an ensemble of decision trees and including feature significance, making it a potential tool for startup failure prediction.

Deep neural network-based DeepSurv model got a C-index of 0.687809. While this model performed well, it did not outperform the Random Forest model in terms of predicted accuracy. The DeepSurv model's capacity to capture complicated nonlinear interactions, as well as its potential for improvement with bigger datasets, making it a promising field for future study.

It is crucial to highlight the inability to use the Cox model in this study, which has implications for the analysis of survival data. When attempting to compute the C-index, the result was 'NA,' indicating missing or unavailable values. There can be several plausible explanations for 'NA' values in predictions. In this particular study, one potential explanation presented is the presence of a significant correlation between the variables used to predict the 'status' of individuals. This correlation may introduce a collinearity issue, where the predictor variables are highly correlated with each other. Collinearity can lead to unstable or unidentifiable coefficients in the model, resulting in 'NA' values in the forecasts.

In the context of this study, the correlations between the 'status' and the variables 'innovation' (0.223), 'profitability' (0.293), and 'investments' (0.426) are observed to be positive and moderately substantial. These correlations indicate that these variables are associated with the 'status' outcome variable in a meaningful way within the given dataset.

In this way, it is important to remark that the comparison research indicated significant disparities in the predictive ability of the survival models for predicting startup failure. The findings imply that more advanced models, such as Multi-Task Logistic Regression (MTLR) and Random Forest, outperform the standard Cox and Kaplan-Meier (K-M) modelx in terms of predicted accuracy.

These results have ramifications for startup ecosystem decision-making. These models' insights may help investors and entrepreneurs analyze the risk of startup failure and make educated investment choices. These models may be used by policymakers to establish targeted interventions and support mechanisms for at-risk businesses.

4.2. Economic perspective

4.2.1. Matrix Analysis

The relative weights matrix, shown in the picture below, may be helpful in understanding how regulators and analysts estimate the risk of failure of a business and which elements they deem most essential at various periods. It is crucial to note, however, that these weights may alter over time as markets and the economy develop, and that various authorities and experts may use somewhat different methods to measuring bankruptcy risk.

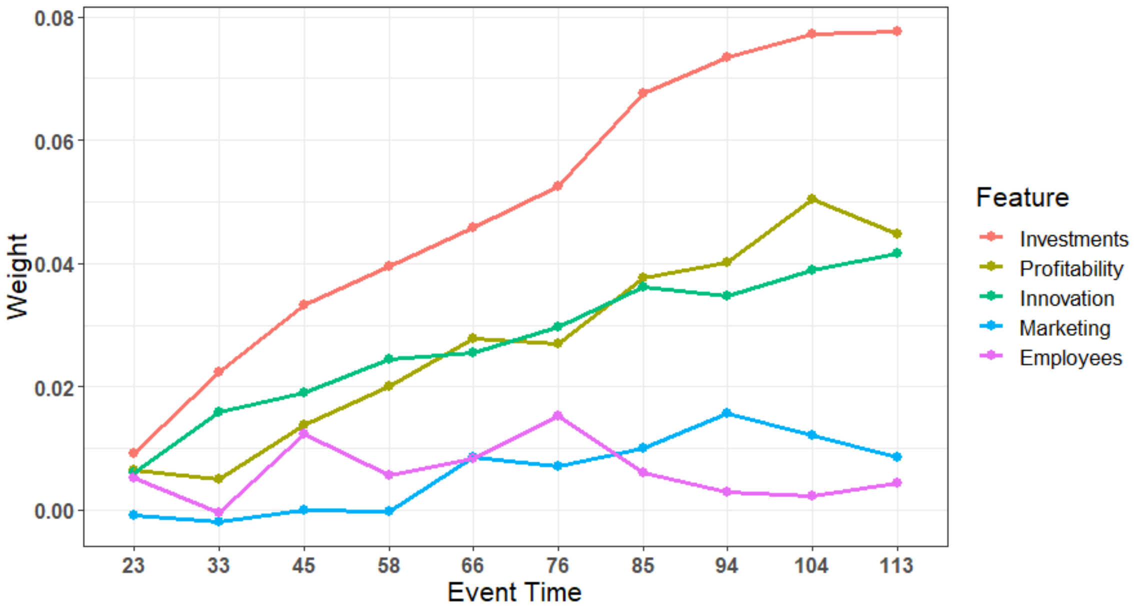

Figure 3. Relative weight of each variable based on MTLR model.

Source: own elaboration.

We employed the MTLR (Multi-Task Logistic Regression) model using time-to-event data to explore the impact of several variables on startup survival in this research. We may estimate the weights associated with each predictor at various time periods using the MTLR model. We looked at starting survival rates at 10 different time intervals ranging from 22.5 to 112.6 units.

The weight matrix sheds light on the impact of various factors on startup survival at each time interval. The weights show the size and direction of each predictor's influence on the log-odds of survival. Positive weights imply a greater chance of survival, whereas negative weights indicate a greater danger of failure.

We can see from the weight matrix that the predictors have various impacts on startup survival at different time periods. Investments, workers, and profitability, for example, seem to have positive weights, indicating a good influence on survival, but costs and competition appear to have negative weights, indicating a negative impact. Weights related with sales, invention, marketing, and founder age change with time, showing time-varying impacts.

It is vital to note that predictor weights indicate their relative relevance inside the MTLR model and may give insights into the elements that impact startup survival. However, interpretation should be done with caution, taking into account the limits and assumptions.

Because of the inherent limitations and assumptions of the MTLR model, as well as the observational character of the research, care should be applied when interpreting the weight matrix findings and making inferences regarding causation.

To begin, the MTLR model implies that the predictors in the study are independent of one another. In practice, however, there may be interdependencies and confounding variables that impact both the predictors and the result variable. Failure to appropriately account for these variables might result in skewed estimates and erroneous interpretations.

Second, using observational data, the MTLR model analyzes the relationships between predictors and startup survival. This implies that causation cannot be determined directly from model findings. While the weight matrix gives information about the relative relevance of predictors, it does not demonstrate a link between those factors and startup survival. Unobserved variables or hidden factors may contribute to both the predictors and the outcome, resulting in erroneous relationships.

Further stringent research designs, such as randomized controlled trials or natural experiments, would be required to demonstrate causal links. These approaches provide for more control over confounding factors and a more solid foundation for inferring causation.

Furthermore, the limitations of the dataset employed in the research should be considered. The results' generalizability may be restricted to the unique environment and population under consideration. Extrapolating the findings to other startup ecosystems or historical periods should be done with caution, given the mechanics of company survival vary among sectors, geographical locations, and economic situations.

5. Conclusion

Our comparative research of survival models for startup failure prediction showed some intriguing results in terms of performance. The findings, as measured by the C-index, show that the models have varied degrees of predictive accuracy.

With a C-index of 0.500, the Kaplan-Meier (K-M) model demonstrated random or poor predictive ability. This is due to its simplicity and the assumption of predictor independence, which restrict its capacity to capture the complexity of startup failure.

The Multi-Task Logistic Regression (MTLR) model, on the other hand, demonstrated enhanced predictive accuracy, with a C-index of 0.810527. The MTLR model showed to be better effective for capturing the multidimensional character of startup failure by including many indicators and taking their interdependence into account.

With a C-index of 0.915297, the Random Forest model beat the prior models. This model displayed improved predictive accuracy by using an ensemble of decision trees and adding feature significance, making it a potential tool for startup failure prediction.

Deep neural network-based DeepSurv model got a C-index of 0.687809. While it performed well, it did not outperform the Random Forest model in terms of predicted accuracy. Nonetheless, the DeepSurv model's capacity to capture complicated nonlinear interactions, as well as its potential for development with bigger datasets, suggest that it is a promising topic for further study.

It is crucial to highlight that our efforts to apply the Cox model were futile, since the C-index computation yielded "NA" results. This shortcoming may be ascribed to the factors used to predict the "status" having a strong correlation. Because of the collinearity generated by this association, the estimated coefficients may become unstable or unidentifiable, resulting in "NA" values in the forecasts. Notably, we found positive and statistically significant correlations between the variable's "status" and "innovation" (0.223), "profitability" (0.293), and "investments" (0.426).

The weight matrix sheds light on the impact of various factors on startup survival at each time interval. The weights show the size and direction of each predictor's influence on the log-odds of survival. At various time periods, the predictors have differing influence on startup survival. Investments, workers, and profitability, for example, seem to have positive weights, indicating a good influence on survival, but costs and competition appear to have negative weights, indicating a negative impact. Weights related with sales, invention, marketing, and founder age change with time, showing time-varying impacts.

Further study and inquiry may be conducted in the future to improve our grasp of startup failure prediction. Here are some possible next steps:

Model Refinement: Continuously refining current machine learning survival models by fine-tuning their parameters, testing with alternative feature engineering strategies, and experimenting with ensemble methods may result in improved forecast accuracy. Incorporating domain-specific information and adding additional predictors related to startup failure may also help to enhance the models.

Generalization and Validation: Validating the models using separate datasets from other startup environments and historical periods might help determine their generalizability. This will allow them to assess their performance in a variety of settings and discover any possible limits or biases.

Model Interpretability: Improving machine learning models' interpretability may give deeper insights into the causes behind startup failure. Techniques such as feature significance analysis, partial dependency plots, and SHAP (Shapley Additive exPlanations) values may aid in elucidating the underlying processes and linkages in the models.

Researchers, policymakers, and stakeholders may increase their knowledge of startup failure prediction, contribute to the creation of successful initiatives, and foster a vibrant entrepreneurial environment by taking these next actions.

Funding Statement

This research received no external funding.

Acknowledgments

Acknowledgments to anonymous referees' comments and editor's effort.

Conflict of interest

The author claims that the manuscript is completely original. The author also declares no conflict of interest.

References

Aalen, Odd O. (1989). A linear regression model for the analysis of life times. *Statistics in medicine*, 907-925. https://doi.org/10.1002/sim.4780080803

Adnan Aziz, M and Dar, Humayon A. (2006). Predicting corporate bankruptcy: where we stand?. *Corporate Governance*: The international journal of business in society, 18-33. https://doi.org/10.1108/14720700610649436

Alaka, Hafiz A and Oyedele, Lukumon O and Owolabi, Hakeem A and Kumar, Vikas and Ajayi, Saheed O and Akinade, Olugbenga O and Bilal, Muhammad. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 164-184. https://doi.org/10.1016/j.eswa.2017.10.040

Alhurani, Abdullah S and Hamdan-Mansour, Ayman M and Ahmad, Muayyad M and McKee, Gabrielle and O'Donnell, Sharon and O'Brien, Frances and Mooney, Mary and Saleh, Zyad T and Moser, Debra K. (2022). The Association of Persistent Symptoms of Depression and Anxiety with Recurrent Acute Coronary Syndrome Events: A Prospective Observational Study. *Healthcare*, 383. https://doi.org/10.3390/healthcare10020383

Altman, Edward I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 589-609. https://doi.org/10.2307/2978933

Andersen, Per Kragh and Gill, Richard. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 1100-1120. https://doi.org/10.1214/aos/1176345976

Barboza, Flavio and Kimura, Herbert and Altman, Edward. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 405-417. https://doi.org/10.1016/j.eswa.2017.04.006

Bauer, Julian and Agarwal, Vineet. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 432-442. https://doi.org/10.1016/j.jbankfin.2013.12.013

Beaver, William H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 71-111. https://doi.org/10.2307/2490171

Beretta, Alessandro and Heuchenne, Cedric. (2019). Variable selection in proportional hazards cure model with time-varying covariates, application to US bank failures. *Journal of Applied Statistics*, 1529-1549. https://doi.org/10.1080/02664763.2018.1554627

Bharath, Sreedhar T and Shumway, Tyler. (2008). Forecasting default with the Merton distance to default model. *The Review of Financial Studies*, 1339-1369. https://doi.org/10.1093/rfs/hhn044

Bijwaard, Govert E and Franses, Philip Hans and Paap, Richard. (2006). Modeling purchases as repeated events. *Journal of Business & Economic Statistics*, 487-502. https://doi.org/10.1198/073500106000000242

Box-Steffensmeier and Janet M and De Boef, Suzanna. (2006). Repeated events survival models: the conditional frailty model. *Statistics in medicine*, 3518-3533. https://doi.org/10.1002/sim.2434

Chang, Shu-Hui and Wang, Mei-Cheng. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association*, 1221-1230. https://doi.org/10.1080/01621459.1999.10473875

Clayton, David. (1994). Some approaches to the analysis of recurrent event data. *Statistical methods in medical research*, 244-262. https://doi.org/10.1177/096228029400300304

Cox, David R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Cox, Raymond AK and Kimmel, Randall K and Wang, Grace WY. (2017). Proportional hazards model of bank failure: Evidence from USA. *Journal of Economic & Financial Studies*, 35-45. https://doi.org/10.18533/jefs.v5i3.290

De Leonardis, Daniele and Rocci, Roberto. (2014). Default risk analysis via a discrete-time cure rate model. *Applied Stochastic Models in Business and Industry*, 529-543. https://doi.org/10.1002/asmb.1998

Deakin, Edward B. (1972). A discriminant analysis of predictors of business failure. *Journal of accounting research*, 167-179. https://doi.org/10.2307/2490225

Du Jardin, Philippe. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, 286-303. https://doi.org/10.1016/j.ejor.2014.09.059

Duffie, Darrell and Saita, Leandro and Wang, Ke. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of financial economics*, 635-665. https://doi.org/10.1016/j.jfineco.2005.10.011

Ejoku, Jonatha. (2020). Analysis of recurrent events with associated informative censoring: Application to HIV data. *International Journal of Statistics in Medical Research*. https://doi.org/10.6000/1929-6029.2020.09.03

Emrouznejad, Ali and Yang, Guo-liang. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978-2016. *Socio-economic planning sciences*, 4-8. https://doi.org/10.1016/j.seps.2017.01.008

Fotso, Stephane. (2018). Deep neural networks for survival analysis based on a multi-task framework. arXiv preprint arXiv:1801.05512.

Gepp, Adrian and Kumar, Kuldeep. (2008). The role of survival analysis in financial distress prediction. *International research journal of finance and economics*, 13-34.

Godlewski, Christophe J. (2015). The dynamics of bank debt renegotiation in Europe: A survival analysis approach. *Economic Modelling*, 19-31. https://doi.org/10.1016/j.econmod.2015.03.017

Henriques, Iago Cotrim and Sobreiro, Vinicius Amorim and Kimura, Herbert and Mariano, Enzo Barberio. (2020). Two-stage DEA in banks: Terminological controversies and future directions. *Expert Systems with Applications*, 113632. https://doi.org/10.1016/j.eswa.2020.113632

Hosaka, Tadaaki. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert systems with applications*, 287-299. https://doi.org/10.1016/j.eswa.2018.09.039

Hu, Dan and Zheng, Haiyan. (2015). Does ownership structure affect the degree of corporate financial distress in China? *Journal of Accounting in Emerging Economies*. https://doi.org/10.1108/JAEE-09-2011-0037

Jabeur, Sami Ben and Serret, Vanessa. (2023). Bankruptcy prediction using fuzzy convolutional neural networks. *Research in International Business and Finance*, 101844. https://doi.org/10.1016/j.ribaf.2022.101844

Kristanti, Farida Titik and Herwany, Aldrin. (2017). Corporate governance, financial ratios, political risk and financial distress: A survival analysis. *Accounting and Finance Review*, 26-34. http://dx.doi.org/10.35609/afr.2017.2.2(4)

Lane, William R and Looney, Stephen W and Wansley, James W. (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance*, 511-531. https://doi.org/10.1016/S0378-4266(86)80003-6

LeBlanc, Michael and Crowley, John. (1992). Relative risk trees for censored survival data. *Biometrics*, 411-425. https://doi.org/10.2307/2532300

Lee, Suk Hun and Urrutia, Jorge L. (1996). Analysis and prediction of insolvency in the property-liability insurance industry: A comparison of logit and hazard models. *Journal of Risk and insurance*, 121-130. https://doi.org/10.2307/253520

Li, Zhiyong and Crook, Jonathan and Andreeva, Galina and Tang, Ying. (2021). Predicting the risk of financial distress using corporate governance measures. *Pacific-Basin Finance Journal*, 101334. https://doi.org/10.1016/j.pacfin.2020.101334

Lin, Wei-Yang and Hu, Ya-Han and Tsai, Chih-Fong. (2011). Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C Applications and Reviews*, 421-436. https://doi.org/10.1109/TSMCC.2011.2170420

Luoma, Martti and Laitinen, Erkki K. (1991). Survival analysis as a tool for company failure prediction. *Omega*, 673-678. https://doi.org/10.1016/0305-0483(91)90015-L

Mai, Feng and Tian, Shaonan and Lee, Chihoon and Ma, Ling. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European journal of operational research*, 743-758. https://doi.org/10.1016/j.ejor.2018.10.024

Ohlson, James A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109-131. https://doi.org/10.2307/2490395

Parker, Susan and Peters, Gary F and Turetsky, Howard F. (2005). Corporate governance factors and auditor going concern assessments. *Review of Accounting and Finance*. https://doi.org/10.1108/eb043428

Pölsterl, Sebastian, Nassir Navab, and Amin Katouzian. (2015). Fast training of support vector machines for survival analysis. *Springer*. https://doi.org/10.1007/978-3-319-23525-7_15

Prentice, Ross L and Williams, Benjamin J and Peterson, Arthur V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 373-379. https://doi.org/10.1093/biomet/68.2.373

Shumway, Tyler. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The journal of business*, 101-124. https://doi.org/10.1086/209665

Tam, Kar Yan and Kiang, Melody Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management science*, 926-947. https://doi.org/10.1287/mnsc.38.7.926

Tian, Shaonan and Yu, Yan. (2017). Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance*, 510-526. https://doi.org/10.1016/j.iref.2017.07.025

Tinoco, Mario Hernandez and Wilson, Nick. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International review of financial analysis*, 394-419. https://doi.org/10.1016/j.irfa.2013.02.013

Twisk, Jos WR and Smidt, Nynke and de Vente, Wieke. (2005). Applied analysis of recurrent events: a practical overview. *Journal of Epidemiology & Community Health*, 706-710. http://dx.doi.org/10.1136/jech.2004.030759

Uno, Hajime and Cai, Tianxi and Pencina, Michael J and D'Agostino, Ralph B and Wei, Lee-Jen. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 1105-1117. https://doi.org/10.1002/sim.4154

Van Belle, Vanya and Pelckmans, Kristiaan and Suykens, Johan AK and Van Huffel, Sabine. (2007). Support vector machines for survival analysis. Proceedings of the third international conference on computational intelligence in medicine and healthcare (cimed2007), 1-8.

Wang, Ping and Li, Yan and Reddy, Chandan K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 1-36. https://doi.org/10.1145/3214306

Wang, Yuling and Carson, James M. (2010). Macroeconomic factors and insurer rating transitions. Available at SSRN 1558456. http://dx.doi.org/10.2139/ssrn.1558456

Wang, Zongjun and Li, Hongxia. (2007). Financial distress prediction of Chinese listed companies: a rough set methodology. *Chinese Management Studies*, 93-110. https://doi.org/10.1108/17506140710758008

Wei, Lee-Jen and Lin, Danyu Y and Weissfeld, Lisa. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association*, 1065-1073. https://doi.org/10.1080/01621459.1989.10478873

Yang, Zijiang and You, Wenjie and Ji, Guoli. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 8336-8342. https://doi.org/10.1016/j.eswa.2011.01.021

Yu, Chun-Nam and Greiner, Russell and Lin, Hsiu-Chin and Baracos, Vickie. (2011). Learning patient-specific cancer survival distributions as a sequence of dependent regressors. *Advances in neural information processing systems*

Zelenkov, Yuri and Fedorova, Elena and Chekrizov, Dmitry. (2017). Two-step classification method based on genetic algorithm for bankruptcy forecasting. *Expert Systems with Applications*, 393-401. https://doi.org/10.1016/j.eswa.2017.07.025

Zhou, Fanyin and Fu, Lijun and Li, Zhiyong and Xu, Jiawei. (2022). The recurrence of financial distress: A survival analysis. *International Journal of Forecasting*, 1100-1115. https://doi.org/10.1016/j.ijforecast.2021.12.005