Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 130
DOI: 10.11159/icsta24.130

Factors Influencing Business Survival: A Survival Analysis in the Albanian Business Landscape

Lule Basha[1], Eralda Gjika[2], Ornela Gordani[3]
[1,3]Department of Applied Mathematics, Faculty of Natural Science, University of Tirana, Tirana, Albania
lule.hallaci@fshn.edu.al; ornela.gordani@fshn.edu.al
[2]Royal College of Physician and Surgeons of Canada, Ottawa, Canada
egjika@royalcollege.ca

Abstract **-** Investigating the resilience of businesses can significantly enhance the robustness of an economy, while also offering insights into the factors that impact growth and stability. Understanding of business survival supports policymakers and stakeholders in crafting initiatives that promote entrepreneurship, drive innovation, and enhance long-term economic sustainability. Small country economies pose unique challenges, and there is limited prior research addressing these specific contexts. In our study, we utilize data from 1560 Albanian firms established between the first day of July 1992 and the last day of December 2023. The objective of this study is to analyse the monthly survival duration of these firms by evaluating two different approaches. First, a traditional strategy that estimates their survival using the Kaplan-Meier method and verifies the statistical significance of the variables using the Cox regression model. Kaplan-Meier survival curves show a differential survival pattern among several types of legal forms for firms, with an average duration of survival of 143 months for these recently established Albanian firms. Furthermore, the second approach employed in this study involves the utilization of survival support vector machines, decision trees, and random survival forests with survival data. The support vector regression model with the radial basis function, has the highest AUC of 0.93 and the highest predictive accuracy when it comes to ranking survival periods. Its C-index value is 0.72, which is also the same as the Cox proportional hazards model, which is used as the reference model. This study contributes to the existing literature by illuminating critical factors influencing firm survival and methodologies used to achieve the results. The insights gained from this study can benefit policymakers, entrepreneurs, and stakeholders involved in fostering business resilience and sustainable economic development.

Keywords**:** firm survival; Kaplan-Meier; Cox Proportional Hazards model; machine learning; random survival forest; Support Vector Machine

1. Introduction

Given their significant impact on national economies and several participants, the survival of firms remains a primary objective for most companies [1]. However, achieving durability in the ever-changing business setting of today presents significant challenges. McKinsey's research provides a sharp decline in the average lifespan of companies listed in the Standard and Poor's 500 index: from 61 years in 1958 to less than 18 years in 2016. Concerning, McKinsey's projections suggest that a 75% of the standard current and poor's 500 companies cannot survive more than 20 years, [2]. In 2008, the Bank of Korea, [3] conducted a thorough investigation covering 41 countries across Asia, Africa, Europe, North America, South America and Oceania. Their findings revealed remarkable durability among some companies, with 5,586 firms lasting more than two centuries. Among these legacy companies, Japan boasted the highest number, with 3,146 entities, accounting for 56.3% of the total. Germany followed with 837 companies (15.0%), the Netherlands with 222 (4.0%) and France with 196 (3.5%). Numerous studies have investigated the factors affecting the longevity of manufacturing firms, which have identified various factors related to firm survival, includes, among other things, loans, assistance, revenue, and creativity [4-8]. Also, the size and age of the firm are main factors related to the survival of the firm [9-11]. A firm's longevity plays a modest but significant role in reducing the likelihood of failure. Clearly, further research on firm survival is crucial.

Survival analysis is a basic statistical method widely used in various fields of research, which deal with the time until the occurrence of a specific event. Because of their wide use, these methods are commonly used to analyze firm survival over time. The Cox proportional hazards model, one of the most important parts of survival analysis models, allows the identification of factors and the impact they have on the firm's survival time [10]. In the study [12], the authors conducted an empirical study using survival analysis models, specifically the Cox proportional hazard model, focusing on Chilean firms from 2011 to 2015. First, they examined the factors affecting the survival of firms for different groups using a probit model, which includes all firms and not only those created in 2010. Some other works that use survival analysis models to estimate the longevity of firms and the factors that affect this lifespan was carried out by the authors [13-15]. Authors in [16], studies the survival time of non-cultural firms compared to cultural firms, and how factors such as profitability, solvency and debt can affect their longevity. Their analysis included the application of the Kaplan-Meier method to estimate survival time, the use of the Harrington-Fleming test and the Cox regression model to determine the statistical significance of these factors included in the study.

Machine learning methods, including decision trees, random forests and neural networks, are increasingly used to predict the survival time of firms, in recent years [17-20]. Authors in [21], studies the predictive accuracies of different machine learning models for predicting the bankruptcy of firms one year before their closure. They benchmark their performance against results obtained from discriminant analysis using a data set spanning 1985 to 2013 and comprising over 10,000 observations per year from North American firms. An extensive review of these applications in finance through the use of bibliometric techniques is provided in [22]. Authors in [23], analyses how machine learning methods, such as Bayesian additive regression trees (BART), spanning trees, bootstrap forests, and regression tree algorithms, combined with economic theory, can be used to predict performance financial of manufacturing firms in developing countries. Reference [24] compares two distinct methodologies for assessing credit risk in small and medium-sized businesses, in a selection of 464 Italian small and medium-sized businesses spanning the period from 2015 to 2017. The models studied are a model ordered probit and a machine learning random forest model, where the results show that the random forest approach outperforms the traditional ordered probit model. Machine learning methods are often used to solve complex problems encountered in survival analysis. By utilizing machine learning techniques, researchers can effectively model and predict survival analysis results [25]. Recent studies pointed out that machine learning algorithms are an excellent solution to calculate survival time in classical models of survival analysis [26-28], constructing the most recent solutions in this domain.

This paper aims to study the survival time per month for 1560 firms in Albania. To accomplish this goal, statistical methods and survival analysis, along with the development of machine learning classifiers for survival methods, were applied to these data, in order to identify factors that contribute to firm longevity. Section 2 provides an overview of survival analysis, Kaplan-Meier and Cox proportional hazards model. Along with an overview of machine learning algorithms for survival analysis methods, such as decision tree, random survival forest, and survival Support Vector Machine. Section 3 applies various statistical models as well as machine learning methods to the firm dataset and shows the comparison between these models. The sample obtained from the National Business Center (NBC) database in Albania consists of 1560 firms established between the first day of July 1992 and the last day of December 2023. Lastly, section 4 presents the conclusions derived from result analysis along with additional recommendations. It's important to note that the study has certain limitations that should be considered when interpreting the findings.

2. Data Analysis Methods

2.1. Survival Data Analysis

Survival analysis studies the time until an event of interest, Y, occurs. Survival data often include censored observations, where the time taken in the study is not shown in full, the censoring time, C. This issue represents a key characteristic of this type of analysis. Due to random right-censoring, we may not always observe the Y variable. Therefore, for each individual, we have the pair (T, δ), where $T=min(Y,C)$ is the time until an individual either experiences the event of interest or their follow-up period ends, and $\delta = I(Y \leq C)$ is the censoring indicator. A sample is a set of observations represented by $\{(T_i, \delta_i), i = 1, ..., n\}$ independent and identically distributed variables of the random vector (T, δ). A fundamental goal in survival analysis is to estimate the density and survival function for the variable Y. To this end, Kaplan [29], introduced the nonparametric maximum likelihood estimator of the survival function, known as the product-limit estimator.

$$\hat{S}(t) = \prod_{i:T_{(i)} \leq t}^{n} \left(1 - \frac{\delta_{[i]}}{n - i + 1}\right) \tag{1}$$

where $\delta_{[i]}$ is the corresponding censoring indicator concomitant of $T_{(i)}$.

The Cox proportional hazards model [30] is one of the most widely used statistical approaches in survival analysis, to evaluate the impact that the factors under study have on survival time. Its basic assumption is that the risks remain proportional over time for all individuals in the studied population and is represented by the formula:

$$h_{(i)}(t) = \exp\left(\sum_{j=1}^{p} X_{ij}\beta_j\right)h_0(t) \tag{2}$$

where $h_{(i)}(t)$ is the hazard function for the ith individual, X_{ij} is the explanatory variable j for individual i, β_j is the slope term for the j^{th} explanatory variable, and $h_0(t)$ refers to a hazard function for an individual with zeros for all features.

2.2. Machine Learning Techniques Applied to Survival Analysis

In recent years, machine learning has advanced significantly in its ability to solve standard regression and classification problems. This development has generated interest within the machine learning community to extend these methodologies to various fields, including survival analysis. However, a major challenge facing researchers in this field is the existence of censored data, a less common characteristic in traditional machine learning problems. Generally, we have been employed the following models for survival analysis:

Decision Trees and Random Survival Forests: tree-based methods were among the earliest machine learning models successfully applied to censored data [31, 32]. Among these methods, random survival forests (introduced by Ishwaran in 2008 [33]) are the most advanced tree-based survival model to date, where the ensemble is created by stacking Nelson-Aalen tree-based estimators $\hat{H}_b(t \mid y)$. The survival prediction from the random survival forest, for B bootstrap samples is:

$$\hat{S}^{rsf}(t \mid y) = \exp\left(-\frac{1}{B}\sum_{b=1}^{B}\hat{H}_b(t \mid y)\right) \tag{3}$$

Support Vector Machines (SVMs): Shivaswamy introduced the support vector censored regression approach to create a model competent of using all available information from survival data including censored data [34]. Support vector machines essentially divide the dependent space into two subspaces using a hyperplane that maximizes the boundaries between the hyperplane and the points from the two subspaces closest to it. The optimization of the convex objective function for this model, considering the censored data, with α_i Lagrange multipliers, can be expressed as follows:

$$\max_{\alpha,\alpha^*}\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)y_iy_j-\sum_{i\in L}t_i\alpha_i+\sum_{i\in U}t_i\alpha_i^*\right)$$
$$\text{subject to}\begin{cases}\sum_{i\in L}\alpha_i-\sum_{i\in U}\alpha_i^*=0\\ 0<\alpha_i,\alpha_i^*\le C\end{cases} \tag{4}$$

An exhaustive grid search with *k*-fold cross-validation has been used to assess the models during the hyper parameter exploration. The hold-out test set has been used to evaluate the final models. The research demonstrates that this technique is also utilised to increase classification accuracy. The *k*-fold cross validation has been integrated into the grid search and offers the benefit of precisely estimating the model's accuracy and utilising additional data for validation.

To evaluate models during hyper parameter exploration we use an exhaustive network search with *k*-fold cross-validation. This method offers the benefit of accurately estimating the accuracy of the model and using additional data for validation. The retention test set is used to evaluate the final models and also to increase the classification accuracy, as researches have shown.

3. Results

3.1. Data Pre-Processing and Descriptive Statistics

This paper examines the duration of time of newly established firms by analysing a sample obtained from the National Business Center (NBC) database in Albania, [35]. The sample consists of 1560 firms established between July 1992 and

December 2023. The data were downloaded from the website of the NBC, as PDF format. Then, through an algorithm created by the authors in Python, the necessary information was extracted from PDF, in Excel format. Also, the data have been subjected to their pre-processing phase, dealing with missing data and outliers, where 1700 pdf files were extracted from the web, but because they had a lot of missing values, only 1560 of them were used for analysis. The registration time in the NBC marks a firm's birth, while deregistration signifies its exit. The study faces the challenge of censored data, referring to firms still in operation (or 'alive') as of last day of February 2024, the end of the sample period. Of the 1560 firms, 799 (51%) were still in business ('alive') in the last day of February 2024 (the end of the sample period), while 763 (49%) were closed ('dead'). We also conduct a comparison among three categories of firms by grouping them into classes.: Class I firms, Natural Person (NP), constituting 33.1% of the firms; Class II firms, Limited Liability Company (LLC), constituting 64.5% of the firms and Class III firms, Others (Branch of the foreign company; Joint stock company; representative office; collective society etc.). The survival time we study is in months. Then to analyze the survival time for these firms, we used survival analysis methods and the aforementioned machine learning methods for survival data, by using the R packages (*survival; randomForestSRC; survivalsvm; survivalROC*).

3.2. Survival Analysis

In this study, the Kaplan–Meier method has been used to estimate the survival time, by plotting the survival probabilities in relation to the survival months for firms, Figure 1.

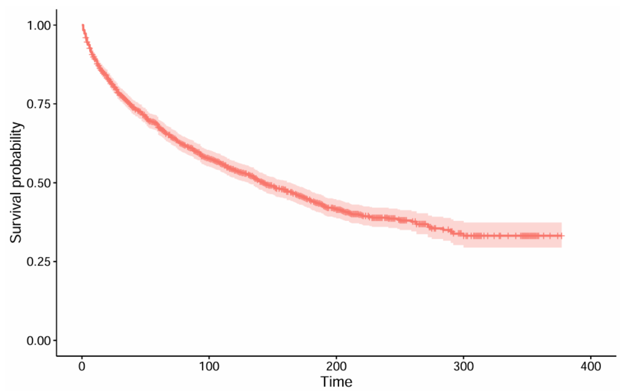

Fig. 1: Survival probability over time.

For these firms, the average expected survival period is 143 months, so almost half of the businesses close 143 months after they are established and 75% of them close in 37 months. The range from 131 to 164 months of the 95% confidence interval for the average survival time shows a considerable heterogeneity, where many factors can be attributed to these fluctuations, including economic conditions, financial markets, the dynamics of the industry, etc. We have also examined the survival time of these firms, categorizing them based on their legal structure and geographic location. After constructing Kaplan-Meier survival curves, statistically distinct survival patterns emerged between different groups of firms ($p < 0.0001$). Limited Liability Companies (LLCs) make up a significant percentage of the data set and have an average survival time of 181 months compared to companies classified as Natural Person (NP), which have a much lower average survival period of 93 months. Authors in [36] estimate Albanian firms' hazards of failure and density function using kernel smoothing methods for survival analysis, while [37], develop a guided kernel estimator based on a gamma kernel, and provides an application of this model to the Albanian firm survival.

3.3. Decision Trees and Random Forests

One of the problems encountered with decision trees is their predisposition to over fit the training data set, which leads to poorer performance on the test set compared to the training set. The reason this situation occurs is because the models learn very specific rules that may not be statistically significant. Figure 2 illustrates the receiver operating characteristic (ROC) curves for decision tree, for our data.

For each of the models we used a 5-out-of-10-fold nested cross-validation was used for tuning, so the data sets were divided into 5 almost equal sub-samples randomly and each iteration used a test set of data from one of 5 groups. The constructed models were trained on the remaining sets after a 10-fold cross-validation was used to fine-tune the model parameters. After obtaining the results the decision tree model gave an AUC of 0.65, with a sensitivity of 0.68 and specificity specificity of 0.73. with a 95% confidence interval for AUC, estimated through k-fold cross-validation on the training set, [0.73, 0.79].

The Random Survival Forest (RSF), pruned to the survival data, was constructed with 1000 trees, each with an average terminal node size of 5 observations, where the average number of terminal nodes in all trees was approximately 27,822. The censoring data management model used in this model is the same as the main survival model. Trees were developed using ranked sampling without replacement, and the log-rank split rule.

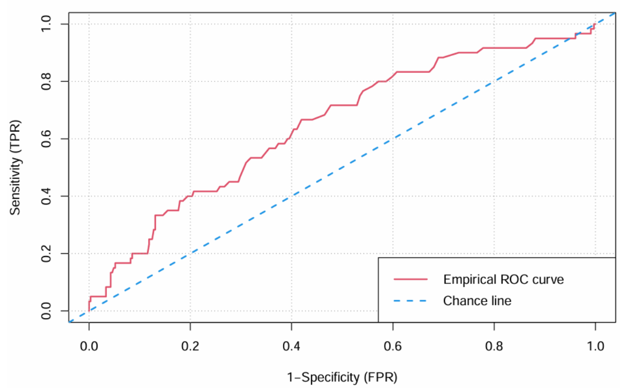

Fig. 2: Receiver Operating Curve (ROC) for the Decision Tree model

Through this model, the importance that the factors taken in the study have on survival time has been studied, where among the predictor variables considered in the model, the most influential factor is “legal form”, with a significance point of 0.1201. Geographical location with a significance score of 0.0333 and a relative significance of 0.2774 compared to “legal form”, it also has a level of significance, albeit to a lesser extent.

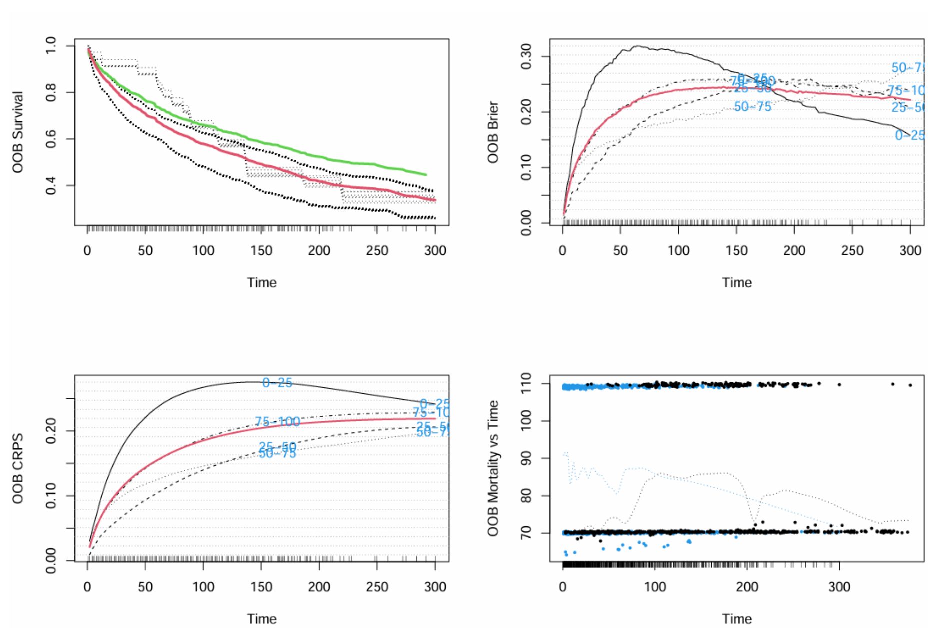

Fig. 3: Survival curves and OOB estimated from a Random Survival Forest model

Figure 3 presents the survival curves estimated from a random survival forest model, stratified by levels of legal form, together with OOB. The out-of-bag (OOB) Brier score evaluates the predictive performance of the model, that is, it tells us how well the model’s predicted probabilities match the observed outcomes. A lower OOB Brier score, in our case less than

0.3, serves to help assess the reliability and accuracy of models in predicting survival probabilities over time. For RSF, the package *randomForestSRC*, [33] was used.

3.4. Support Vector Machines

To evaluate the survival Support Vector Machines (SSVM) models for our study, we fitted four models, the survival support vector regression models [34] and the ranking approach, as proposed by [38] (*vanbelle*), using linear and radial basis function (RBF) kernels. For this we have used the *survivalsvm* package in R. Five 10-fold nested cross validation was utilised for tuning: data sets were split into five nearly equal-sized subsamples at random, and one of the five groups was used as the test data set for each iteration. After fine-tuning the model parameters using a 10-fold cross validation, the models were trained on the remaining groups. The C-index was used to choose the top models. For these models, the 95% confidence interval for the AUC, which was calculated during training using the k-fold, is [0.83, 0.94]. As a result, that interval contains the AUC for the test set, and the k-fold sample mean and the test AUC do not differ significantly (p-value = 0.50 one sample t-test). We can also say that these models have a high degree of reliability.

3.5. Models' Comparison

Table 1, summarize the results for each off the methods given above, together with the Cox PH model, which served as reference model. Among the models, the Support Vector Regression with Radial Basis Function model has the highest AUC of 0.93, followed by the Random Survival Forest model with an AUC of 0.81. Also, this model has the highest sensitivity of 0.91, indicating that it correctly identifies 91% of positive cases.

Table 1. Performance of the models.

Model	AUC	Sensitivity	Specificity	C-Index
Decision Tree	0.65	0.68	0.73	0.595
Random Forest	0.81	0.78	0.79	0.667
SSVR linear	0.89	0.87	0.82	0.698
SSVR RBF	0.93	0.91	0.91	0.720
Vanbelle linear	0.86	0.79	0.82	0.686
Vanbelle RBF	0.91	0.89	0.90	0.698
Cox PH model	--	--	--	0.722

The SSVR linear model and Vanbelle RBF model have the highest specificity of 0.82 and 0.90, respectively, indicating that they correctly identify 82% and 90% of negative cases, respectively. Figure 4 present the performance estimates of the compared models, based on the C-index.

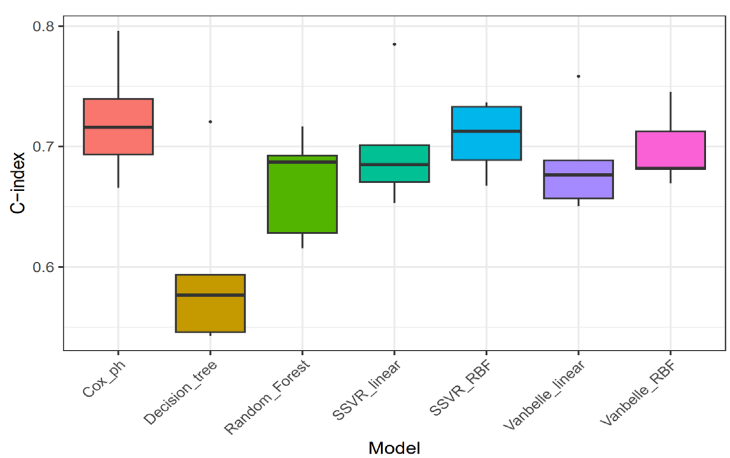

Fig. 4: Prediction performance for the models.

From the comparative analysis of C-index values, which serves as a key metric for evaluating the predictive performance of survival analysis models, it is evident that both support vector regression with the radial basis function kernel; and the Cox Proportional Hazards (PH) model, which is used as a reference model, exhibits the highest predictive accuracy in ranking survival times, with a C-index of 0.72. Where a higher C-index value means better discriminating power in ranking the survival time. These models are better than other methods in terms of their ability to distinguish the survival durations of firms in Albania, followed by Support Vector Regression with a linear kernel.

4. Conclusion

This study investigates the survival time of firms in the Albanian economy, using a combined analysis of traditional statistical methods with advanced machine learning techniques, applied on 1560 firms created between July 1992 and December 2023. The main factors that affect the survival outcomes of these firms were identified, and the predictive efficiency of the studied models was evaluated through rigorous analysis and testing. Our results highlight statistically different survival patterns between different legal forms of enterprises, with a p-value <0.0001. In these results, limited liability companies (LLCs) demonstrate a significantly longer average survival time compared to firms with natural persons. Following legal structure, the geographic location of the firms also appears to have a moderate impact on their survival outcomes, with an importance score of 0.0333 and a relative importance of 0.2774.

In conjunction with survival analysis models, the development of machine learning classifiers for survival methods, were carried out on these data, such as decision-tree, random survival forest, and survival Support Vector Machine. A 5-by-10-fold nested cross-validation was employed for tuning, for each of the model. Decision tree model yielded an AUC of 0.65, with a sensitivity of 0.68 and specificity of 0.73. Random Forest model yielded an AUC of 0.81, with a sensitivity of 0.78 and specificity of 0.79. From the group of Support Vector Machines models for our study, we fitted four models, the survival support vector regression models and the ranking approach, using linear and radial basis function (RBF) kernels. The Support Vector Regression with Radial Basis Function model emerges as a standout performer, exhibiting superior predictive accuracy with an AUC of 0.93 and a C-index of 0.72, aligning closely with the Cox Proportional Hazards model.

The study's conclusions provide valuable new insights into the factors influencing business survival not only in the Albanian economy but also in other small and medium countries facing similar economic conditions. The sheer volume of observations and factors within the dataset may pose a challenge, potentially limiting the depth of analysis that can be achieved. Additionally, the study's timeframe may not have captured long-term trends or macroeconomic shocks that could significantly impact a firm's survival prospects.

References

[1] P. F. Drucker, *The Practice of Management*, Harper & Row, New York, NY, USA, 1954.

[2] S. Garelli, "Why you will probably live longer than most big companies," IMD, 2016. https://www.imd.org/research-knowledge/articles/why-you-will-probably-live-longer-than-most-big- companies/

[3] The Bank of Korea. The Factors and Implications of Japanese Corporate Longevity. 2008. Available online: https://www.bok.or.kr/viewer/skin/doc.html?fn=FILE_201803300723135521.pdf&rs=/webview/result/P0000537/200805

[4] A. Segarra and M. Callejón, "New Firms' Survival and Market Turbulence: New Evidence from Spain," *Rev. Ind. Organ.*, vol. 20, pp. 1–14, 2002. https://doi.org/10.1023/A:1013309928700

[5] S. Mateut, S. Bougheas, and P. Mizen, "Trade credit, bank lending and monetary policy transmission," *Eur. Econ. Rev.*, vol. 50, pp. 603–629, 2006. https://doi.org/10.1016/j.euroecorev.2005.01.002

[6] F. Delmar, A. McKelvie, and K. Wennberg, "Untangling the relationships among growth, profitability and survival in new firms," **Technovation**, vol. 33, pp. 276–291, 2013. https://doi.org/10.1016/j.technovation.2013.02.003

[7] G. Pellegrini and T. Muccigrosso, "Do subsidized new firms survive longer? Evidence from a counterfactual approach," *Reg. Stud.*, vol. 51, pp. 1483-1493, 2017. https://doi.org/10.1080/00343404.2016.1190814

[8] K. Shin, G. Park, J. Choi, and M. Choy, "Factors affecting the survival of SMEs: A study of biotechnology firms in South Korea," *Sustainability*, vol. 9, p. 108, 2017. https://doi.org/10.3390/su9010108

[9] S. Kaniovski and M. Peneder, "Determinants of firm survival: A duration analysis using the generalized gamma distribution," *Empirica*, vol. 35, pp. 41–58, 2008. https://doi.org/10.1007/s10663-007-9050-3

[10] M. C. Manjón-Antolín and J. M. Arauzo-Carod, "Firm survival: Methods and evidence," *Empirica*, vol. 35, pp. 1–24, 2008. https://doi.org/10.1007/s10663-007-9048-x

[11] S. Esteve-Pérez, A. Sanchis Llopis, and J. A. Sanchis Llopis, "The determinants of survival of Spanish manufacturing firms," *Rev. Ind. Organ.*, vol. 25, no. 3, pp. 251-273, 2004. DOI:10.1007/ s11151-004-1972-3
[12] D. López, D. O'Farrill, J. Pérez, and B. Velasquez, "Determinants on firm survival in Chile: Evidence from cohort 2010 for the period 2011-2015", *IFC-Bank Indonesia Satellite Seminar on "Big Data" at the ISI Regional Statistics Conference* 2017, Bali, Indonesia.
[13] E. Taymaz and S. Özler, "Foreign ownership, competition, and survival dynamics," *Rev. Ind. Organ.*, vol. 31, no. 1, pp. 23-42, 2007. DOI:10.1007/s11151-007-9144-x
[14] Y. Che, Y. Lu, and Z. Tao, "Institutional quality and new firm survival," *Economics of Transition*, vol. 25, no. 3, pp. 495-525, 2017. DOI: 10.1111/ecot.12119.
[15] S. Esteve-Pérez, F. Pieri, and D. Rodriguez, "Age and productivity as determinants of firm survival over the industry life cycle," *Ind. Innov.*, vol. 25, pp. 167–198, 2018. https://doi.org/10.1080/13662716.2017.1291329
[16] M. Muñoz Dueñas, A. Vaamonde Liste, and M. d. R. Cabrita, "The survival of cultural firms: A study of multiple accounting parameters in Spain," *Sustainability*, vol. 12, p. 1159, 2020. https://doi.org/10.3390/su12031159
[17] D. Delen, C. Kuzey, and A. Uyar, "Measuring firm performance using financial ratios: A decision-tree approach," *Expert Systems with Applications*, vol. 40, no. 10, pp. 3970-3983, 2013. http://dx.doi.org/10.1016/j.eswa.2013.01.012.
[18] M. Bohanec, M. K. Borštnar, and M. Robnik-Šikonja, "Explaining machine learning models in sales predictions," *Expert Systems with Applications*, vol. 71, pp. 416-428, 2017. https://doi.org/10.1016/j.eswa.2016.11.010.
[19] F. J. Bargagli-Stoffi, J. Niederreiter, and M. Riccaboni, "Supervised learning for the prediction of firm dynamics," in *Data science for economics and finance*, pp. 19-41, Springer, Cham, 2021. http://dx.doi.org/10.1007/978-3-030-66891-4_2.
[20] D. Bragoli, C. Ferretti, P. Ganugi, G. Marseguerra, D. Mezzogori, and F. Zammori, "Machine-learning models for bankruptcy prediction: Do industrial variables matter?" *Spatial Economic Analysis*, pp. 1-22, 2021. https://doi.org/10.1080/17421772.2021.1977377
[21] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, vol. 83, pp. 405–417, 2017. https://doi.org/10.1016/j.eswa.2017.04.006
[22] J. W. Goodell, S. Kumar, W. M. Lim, and D. Pattnaik, "Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis," *Finance Research Letters*, vol. 32, Article 100577, 2021.
[23] D. Saha, T. M. Young, and J. Thacker, "Predicting firm performance and size using machine learning with a Bayesian perspective," *Machine Learning with Applications*, no. 100453, p. 100453, 2023. https://doi.org/10.1016/j.mlwa.2023.100453
[24] A. Bitetto, P. Cerchiello, S. Filomeni, A. Tanda, and B. Tarantino, "Machine learning and credit risk: Empirical evidence from small- and mid-sized businesses," *Socio-Economic Planning Sciences*, vol. 90, no. 101746, p. 101746, 2023. https://doi.org/10.1016/j.seps.2023.101746
[25] S. Huang, J. Yang, S. Fong and Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges," *Cancer Lett.*, vol. 471, pp. 61–71, 2020.
[26] A. J. Steele, S. C. Denaxas, A. D. Shah, H. Hemingway, and N. M. Luscombe, "Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease," *PLoS ONE*, vol. 13, p. e0202344, 2018. https://doi.org/10.1371/journal.pone.0202344
[27] H. Kvamme, O. Borgan, and I. Scheel, "Time-to-event prediction with neural networks and Cox regression," *J. Mach. Learn. Res.*, vol. 20, pp. 1-30, 2019.
[28] C. Nagpal, X. Li, and A. Dubrawski, "Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks," *IEEE J. Biomed Health Inform.*, vol. 25, pp. 3163-3175, 2021. DOI: 10.1109/JBHI.2021.3052441
[29] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Am. Stat. Assoc.*, vol. 53, pp. 457-481, 1958. https://doi.org/10.1080/01621459.1958.10501452
[30] D. R. Cox, "Regression models and life-tables," *J. R. Stat. Soc.*, vol. 34, pp. 187–220, 1972. https://doi.org/10.1111/j.2517- 6161.1972.tb00899.x
[31] F. Dannegger, "Tree stability diagnostics and some remedies for instability," *Statistics in Medicine*, vol. 19, pp. 475–491, 2000.

[32] A. Benner, “Application of ‘Aggregated Classifiers’ in Survival Time Studies,” in *COMPSTAT 2002 - Proceedings in Computational Statistics: 15th Symposium Held in Berlin, Germany, 2002*, MR1973489, 2002.
[33] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, “Random survival forests,” *Ann. Appl. Stat.*, vol. 2, pp. 841–860, 2008. DOI: 10.1214/08-AOAS169
[34] P. K. Shivaswamy, W. Chu, and M. Jansche, “A support vector approach to censored targets,” in *Proceedings of the Seventh IEEE International Conference on Data Mining (ICDM 2007)*, Omaha, NE, USA, 2007, pp. 655–660.
[35] National Business Center (NBC) https://qkb.gov.al/
[36] L. Basha and F. Hoxha, “Guided kernel estimator based on gamma kernel,” *International Journal of Trend in Research and Development*, vol. 5, no. 6, Nov-Dec 2018, ISSN 2394-9333. http://www.ijtrd.com/papers/IJTRD19140.pdf
[37] L. Basha and M. Muça, “Kernel smoothing method for hazard rate estimation: An application to Albanian firm survival,”, *the 2nd International Conference of Natural Sciences and Mathematics*, 2018, Macedonia, ISSN 2545-4072.
[38] V. Van Belle, K. Pelckmans, J. Suykens, and S. Van Huffel, “Support vector machines for survival analysis,” in *Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007)*, 2007, pp. 1–8. [p. 413–415]