Journal of Regional Economics 2025 4(1) 1-18

ANSER PRESS

Journal of Regional Economics

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

Decoding the Puzzle of Joblessness: Machine Learning Predicts Unemployment Trends in the Americas

Diego Vallarino [a, *]

[a] *Independent Researcher, Atlanta, USA*

ABSTRACT

This study investigates the efficacy of diverse machine learning survival models, including Cox, Kernel SVM, DeepSurv, Survival Random Forest, and MTLR models, employing the concordance index to assess their predictive abilities. The primary objective of this research is to identify the most accurate model for forecasting the time it takes for a country to witness a 10% surge in unemployment within a 120-month timeframe (2013-2022), utilizing variables from the MVI dataset of 28 American countries. Through the comparative evaluation of complex survival models, we discovered that DeepSurv, a sophisticated machine learning algorithm, excels in capturing intricate nonlinear relationships, while conventional models exhibit comparable performance under specific circumstances. The weight matrix, a pivotal element of our analysis, meticulously assesses the economic repercussions of various risk factors, vulnerabilities, and capabilities.

KEYWORDS

Probability Forecasting; Regional Forecasting; Survival Forecasting; Machine Learning; Unemployment Forecasting

---

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

ISSN 2972-3434
doi: 10.58567/jre04010001

Received 1 March 2024; Accepted 27 June 2024; Available online 2 July 2025; Version of Record 15 March 2025

1. Introduction

In contemporary economic research, the exploration of temporal dynamics in a nation's response to a significant increase in unemployment gains paramount importance. This empirical investigation, conducted across 28 countries, incorporates a comprehensive dataset that includes countries with right-censored data (9 countries) and those experiencing a 10% increase in unemployment within 120 months (from 2013 to 2022).

In addressing the central query, this research aims to unravel the intricate relationship of variables and risks influencing the time required for a country to undergo the specified 10% increase in unemployment. Leveraging advanced statistical techniques, particularly survival analysis, the study incorporates key variables such as *Vul_Inherent, Vul_Fragility_Democracy*, and *Vul_Human Rights*, offering a robust understanding of multifaceted vulnerabilities.

This academic research emphasizes rigorous methodologies, empirical analyses, and data-driven insights. With a focus on a time-centric lens, the study contributes not only to theoretical advancements in economic literature but also to the formulation of more effective policies aimed at mitigating and managing rising unemployment. The introduction of models commonly used to study economic phenomena, such as the neoclassical growth model, the Keynesian model, the Solow growth model, and the Harrod-Domar model, lays the foundation for a comprehensive exploration (Dutta & Mishra, 2023; Nolan et al., 2019; Yunita et al., 2023).

Moving forward, the research addresses the apparent gap in survival analysis within unemployment prediction specialists (Gabrikova, *at al.* 2023). This work evaluates the effectiveness of various machine learning survival models, including Cox, Kernel SVM, DeepSurv, Survival Random Forest, and MTLR models, using the concordance index to compare their predictive capabilities. The research's primary goal is to determine the most accurate model for projecting the time until a country experiences a 10% increase in unemployment within 120 months.

To achieve this goal, the study scrutinizes the significance and magnitude of vulnerabilities and risks associated with the Multidimensional Vulnerability Index (MVI) of the United Nations Development Programme, comparing the results to economic theory and intuition (Assa, J., et al., 2023; OAS, 2023). The research seeks insights into the factors influencing unemployment levels and aims to provide a better understanding of how survival models can inform public policies.

The paper's structure includes a theoretical perspective on survival analysis models, a brief examination of works using survival analysis to address economic challenges, an empirical analysis encompassing models, data sources, and assessment measures, analytical results, and a conclusion summarizing key findings and implications for future research and policymaking. Overall, this research contributes to the literature on survival analysis in economics, offering insights into features leading to economic challenges and aiding policymakers in developing effective strategies to address such issues in the future.

2. Theorical Perspective

Survival analysis is an important tool in economic research because it provides a nuanced view on dynamic processes such as the longevity of enterprises, the consequences of economic policies, and the persistence of economic events. Survival analysis has significant relevance within the area of economics because to its potential to accommodate time-to-event data, therefore providing a robust framework for analyzing the durations till specific economic events occur (Jin et al., 2021; P. Wang et al., 2017; Zelenkov, 2020; Zhou et al., 2022). Nevertheless, its success is dependent on a detailed understanding of the features and quality of the available data.

It is crucial to determine the nature of the data being evaluated, since mistakes or misconceptions may greatly undermine the reliability of survival forecasts. When dealing with the confluence of survival analysis and economic inquiry in academic research, it is essential to conduct a thorough examination and discernment of the difficulties inherent in the data environment. This means ensuring that the studies undertaken not only enrich our understanding of temporal economic events but also serve as the basis for better informed decision-making processes (Finch, 2005; Gorfine & Zucker, 2022; Maharana et al., 2022; Mumuni & Mumuni, 2022).

Several different types of data may influence the adoption of these tactics. For example, in the biostatistics field basic patient data may contain demographic and clinical details such as the stage of sickness, concomitant conditions, and course of medication (Hair & Fávero, 2019; Maharana et al., 2022).

The usage of datasets, especially censored data, competitive risk data, or longitudinal patient information, might perhaps lead to difficulties (Barrett et al., 2011). Survival algorithms are still useful, despite certain drawbacks (Cuperlovic-Culf, 2018; Jin *et al*., 2021). These problems are even more complex in economic analysis. In the next paragraphs, we will discuss how important this kind of data is for survival analysis.

2.1. Baseline Agent Data

Fundamental patient information is essential for healthcare practitioners to develop a survival strategy. Demographic information, such as race, ethnicity, age, and gender, may have a major influence on a patient's survival probability when combined with clinical data such as illness stage, comorbidities, and treatment history. While we have illustrated the baseline data attributes using the most representative example from the survival analysis literature, the majority of these attributes apply to the examination of customers during the purchasing phase, employees throughout the work process within the organization, or the company throughout its lifespan.

Creating a survival strategy requires a thorough understanding and assessment of several factors that might affect a patient's outlook (Cuperlovic-Culf, 2018; Jin *et al*., 2021). Healthcare practitioners may produce survival predictions using a number of methodologies, including forests, trees, neural networks, deep learning, multitasking, boosting, and "extras" (Thenmozhi *et al*., 2019; Zhao *et al*., 2022). To reduce mistakes and biases, it is vital to consider the limits and restrictions of these algorithms while creating predictions (Azodi *et al*., 2020). As a result, although core patient data is necessary for constructing accurate survival algorithms that assure successful patient care, it is often insufficient to achieve a satisfactory level of performance in machine learning models.

2.2. Censored Data

The notion of data suppression allows us to identify survival data. When the event under studied involves a corporation's collapse or bankruptcy, the remaining participants' event time is suppressed at the end of the research. This involves continuing the statistical study in the absence of the subject's death date (Basak *et al*., 2022; Jiang, 2022; Vinzamuri *et al*., n.d.).

His death is only known to have occurred after the inquiry was completed. People who abandon follow-up research are typically prone to censorship, since their occurrence is commonly unnoticed and their timing is unknown (Raghunathan, 2004). The unobserved date of the event does not qualify as a missing data point, since these two classes of unobserved data have distinct features and are vulnerable to various empirical interpretations (Yuan *et al*., 2022).

The only available information concerning right-censored issues is that the incident happened after the censorship period. If the research had been continued (or if the volunteers had stayed), the final conclusion of interest to all participants would have been observed (Basak *et al*., 2022; Jiang, 2022). Conventional statistical approaches for assessing survival data work under the premise that filtering is non-informative and independent (Khan & Zubek, 2008).

This indicates that at a certain point, the individuals who are still under follow-up have an equal future risk of experiencing the event as those who are no longer under follow-up (either because of censorship or study abandonment). This would be the case if the follow-up losses were arbitrary and so lacked informational value (Basak *et al*., 2022).

The current research clearly indicates that careful handling of censored data is crucial for achieving an accurate picture of the impending survival analysis experiment (Jiang, 2022). As a result, this study will try to determine the most effective techniques for integrating censored data, spanning both the left and the right (which is the predominant approach in analytic models). While the literature has not thoroughly investigated the latter, time-to-event statistical analysis may give significant insights (Cui *et al*., 2020; Yuan *et al*., 2022).

Censored data is prevalent when dealing with survival data; it happens when the precise date of an event is unknown, but it is obvious that the event did not occur before to or after a certain time period. There are three forms of censored data: left-censored, interval-censored, and right-censored. A plethora of special ways are now available to deal with massive amounts of filtered data (Cui *et al*., 2020; Yuan *et al*., 2022).

Survival Random Forest is one technique that performs well with confined data. It is a machine learning approach that combines the predictions of many decision trees (Jin et al., 2021; Jin Ziweiand Shang, 2020; Zhao *et al*., 2022). Multi-Tasking Linear Regression (MTLR) is an additional approach for processing censored data efficiently. It predicts the survival time distribution using a Bayesian technique, which is beneficial when dealing with several outcomes (L. Wang *et al*., 2017). XGboost, another well-known algorithm, can analyze large amounts of censored data, including categorical and continuous variables (Barnwal *et al*., 2022).

This empirical investigation delves into the intricate dynamics of a 10% increase in unemployment across 28 countries, employing a nuanced approach to construct a comprehensive dataset. The dataset comprises countries with right-censored data (9) and those experiencing a 10% increase in unemployment within 120 months, enabling a thorough examination of the temporal features associated with this economic phenomenon.

The subsequent sections of this research will focus on evaluating the prediction performance of various machine learning models for survival analysis. By objectively assessing the efficacy and relative advantages of these models, we aim to contribute to a deeper understanding of the temporal complexities involved in economic development, particularly in the context of unemployment fluctuations. Such insights hold significant implications for policy formulation and economic forecasting.

3. Empirical Analysis

This empirical investigation intricately examines the dynamics of the impact of a 10% increase in unemployment across 28 countries, leveraging a nuanced approach to create a comprehensive dataset. The dataset encompasses countries with right-censored data (9) and those experiencing a 10% increase in unemployment within 120 months. This meticulous sampling strategy ensures a thorough representation of diverse economic scenarios, incorporating information from authoritative sources such as the World Bank, IMF, and the United Nations (Assa et al., 2023; OAS, 2023).

With the intention of answering the primary inquiry, this study endeavors to decipher the complex interconnections among factors and hazards that impact the duration needed for a nation to undergo the designated 10% augmentation in unemployment. By using sophisticated statistical methods, including survival analysis, this research aims to provide a comprehensive comprehension of the temporal dimensions of the economic impact of rising unemployment.

The analysis incorporates key variables, including *Vul_Inherent, Vul_Fragility_Democracy*, and *Vul_Human Rights*, which have been meticulously curated to capture multifaceted vulnerabilities. The rigorous examination of these variables contributes to a deeper understanding of the factors influencing the duration of increased unemployment and provides valuable insights for policymakers seeking effective strategies to address and mitigate the challenges associated with elevated unemployment levels.

3.1. Models

3.1.1. Cox Proportional Hazards Model (coxph)

The Cox proportional hazards model is a popular semi-parametric model in survival analysis. It is assumed that the hazard function may be expressed as the product of a time-independent baseline hazard function and a time-varying covariance function. The mathematical representation of the model is as follows.

Assumption 1: The effect of the predictor variables on the conditional risk function is additive.

Definition 1: The conditional risk function is the function that describes the probability that an event will occur at a given time, given that it has not occurred before.

Definition 2: The baseline risk function is the conditional risk function in the absence of predictor variables.

Suppose we have a dataset of observations t_i, x_i, where t_i is the time at which the event occurs for observation *i* and x_i is a vector of predictor variables for observation *i*.

For each observation *i*, we can define the following conditional risk function:

$$h(t_i|x_i) = P(T_i \leq t_i|x_i)$$

where T_i is the time until the event occurs for observation *i*.

Assumption 1 tells us that the effect of the predictor variables on the conditional risk function is additive. This means that we can write the conditional risk function as the following sum:

$$h(t_i \mid x_i) = h_0(t_i) + f(x_i)$$

where $h_0(t_i)$ is the baseline risk function and $f(x_i)$ is a function that describes the effect of the predictor variables on the conditional risk function.

The baseline risk function is the conditional risk function in the absence of predictor variables. We can estimate the baseline risk function using the data from observations in which the event does not occur.

The function $f(x_i)$ can be any function that describes the effect of the predictor variables on the conditional risk function. In the case of the Cox model, $f(x_i)$ is an exponential function:

$$f(x_i) = exp(B'x_i)$$

where *B* is a vector of coefficients that represents the effect of the predictor variables on the conditional risk function.

Therefore, the Cox model can be written as follows:

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

where $h(t|x)$ is the hazard function for a given time t and covariate values x, $h_0(t)$ 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:

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

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

This empirical study examines the patterns of economic impact in 24 nations, leveraging a sophisticated methodology to create a comprehensive dataset. The dataset includes nations that have been right-censored (9) and those experiencing a 10% increase in unemployment within 120 months. This extensive sampling technique guarantees a complete representation of diverse economic situations.

For the analysis, the experimental setup now considers the event as the country achieving a 10% increase in unemployment. The observation spans from 2013 to 2022. If the 10% threshold has not been reached within these 120 months, the status is designated as 0. Conversely, if the country has experienced the specified increase in unemployment, the status is set to 1. Notably, adjustments have been made to mitigate the impact of changes in 2020 due to the COVID-19 pandemic. Furthermore, the study incorporates information from authorized sources such as the World Bank, IMF, and UN, as well as data from 120 nations to identify multiple risks (see Annex).

The dataset is a comprehensive compilation of various variables, each offering a unique perspective on economic impact scenarios (Balica et al., 2023; Dutta & Mishra, 2023). The primary variables include:

Natural_Risk: EM-DAT characterizes disasters as occurrences or situations that surpass local capacity, compelling the need for external assistance at the national or international level. These events are typically unforeseen and abrupt, resulting in substantial damage, destruction, and human suffering.

Commercial_risk: This variable captures the commercial risks associated with countries and is instrumental in understanding the economic challenges related to trade and market conditions.

Financial_risk: Focused on economic stability, Financial_risk encompasses indicators like the Emerging Markets Bond Index (EMBI) and Risk Rating S&P, providing insights into financial vulnerabilities.

Endogenous_risk: Examining internal economic factors, Endogenous_risk incorporates annual GDP growth, current account balance, inflation, primary balance, public debt, and external debt, offering a holistic view of a country's economic health.

Vul_Inherent: A composite variable, Vul_Inherent encapsulates critical aspects such as proximity to global markets, landlocked status, coastal population proportion, inhabitants in arid lands, total economic loss, fatalities, and affected individuals, providing a comprehensive measure of inherent vulnerabilities.

Vul_Fragility_Democracy: Focusing on democratic institutions, this variable includes indicators like expanded freedom, freedom of association, clean elections, suffrage share, and an elected official's index, offering insights into the fragility of democratic structures. Expanded freedom, freedom of association, clean elections, share of population with suffrage, elected officials index.

Vul_Human Rights: This variable focuses on equal treatment and the absence of discrimination. It assesses the effective guarantee of the right to life and security of the person, due process of the law, rights of the accused, freedom of opinion and expression, freedom of belief and religion, freedom from arbitrary interference with privacy, freedom of assembly and association, and fundamental labor rights.

Vul_Homes: Capturing household-level vulnerabilities, this variable includes the Human Development Index, multidimensional poverty, gender inequality incidence, Gini coefficient, and personal remittances, reflecting the socio-economic conditions at the household level.

Vul_Companies: Assessing business-related challenges, Vul_Companies includes indicators like ease of doing business, permits for construction, property registration, credit accessibility, international trade, and contract enforcement, offering a nuanced perspective on the business environment.

Capabilities_State: Focused on governance and infrastructure, this variable incorporates indices such as the corruption perception index, government effectiveness, Hyogo framework, access to electricity, internet users, adult literacy rate, cell phone subscriptions, road length, basic water services, basic sanitation, doctor density, MCV2 vaccine coverage, DTP3 vaccine coverage, PCV23 vaccine coverage, national health expenditure per capita, and maternal mortality, providing insights into the state's capacity and performance.

Social_Cohesion_Capabilities: Exploring societal dynamics, this variable encompasses power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, indulgence, civil society participation index, direct popular vote index, local government index, and regional government index, shedding light on social cohesion and governance.

Time: The temporal dimension signifies the duration from 2013 to 2022, capturing the span of a country to undergo a 10% increase in unemployment. The temporal dynamics during this period play a pivotal role in shaping the strategies and decisions of countries in addressing the economic impact of rising unemployment.

Status: The status variable reflects the present condition of a country within the analysis period. It is binary, taking values of 0 or 1, signifying whether the country has reached or surpassed the 10% increase in unemployment. A value of 0 indicates that the country has not achieved the 10% threshold, while a value of 1 signifies that the country has attained or surpassed this increment. This status variable provides a snapshot of the country's economic performance and its alignment with the specified increase in unemployment, distinguishing between countries that persist in managing their economic challenges and those that may face difficulties or have undergone significant economic shifts.

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 commonly used metric in survival analysis and medical research to evaluate the performance of predictive models that estimate the probability of an event occurring over a given time period.

The C-index is calculated by rating the predicted event occurrence probability for each participant in a dataset. It computes the proportion of pairs of persons in which the person with the higher expected likelihood saw the occurrence before the person with the lower projected probability. In other words, it evaluates a prediction model's ability to rank persons in order of their chance of attending the event of interest.

The C-index ranges from 0 to 1, with 0.5 reflecting random prediction and 1 showing perfect prediction. In medical research, a C-index value of 0.7 or above indicates excellent performance for a prediction model. The formula for censored data C-Index is as follows.

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

η_i, $the\ risk\ score\ of\ a\ unit\ i$
$1_{T_j<T_i} = 0 \quad if\ \ T_j < T_i\ \ else\ 0$
$1_{\eta_j<\eta_i} = 0 \quad if\ \ \eta_j < \eta_i\ \ else\ 0$
δ_j, $represents\ whether\ the\ value\ is\ censored\ or\ not$

4. Results

The regression study looks at the Multidimensional Vulnerability Index (MVI) as the dependent variable, with many independent variables indicating different dimensions of risk and vulnerability. The explanation of economically important outcomes is provided below.

The intercept (0.169364) shows the estimated MVI when all other independent variables are zero, suggesting inherent vulnerability in the absence of particular hazards or circumstances. Each coefficient for *commercial risk, financial risk, endogenous risk*, and other factors represents the change in MVI associated with a one-unit rise in the relevant independent variable, while keeping all others constant (see Annex).

Following a thorough inspection of the coefficients, *Natural risk*: The coefficient of 0.126599 is statistically significant (p-value = 6.65e-07), indicating that a rise in natural risk is associated with a higher degree of multidimensional vulnerability. *Commercial risk*: The coefficient of -0.047992 is not statistically significant (p-value = 0.325703), suggesting that there is insufficient evidence to establish that commercial risk has a substantial influence on the MVI.

Financial risk: The coefficient of 0.070071 is statistically significant (p-value = 0.004069), showing that higher financial risk is associated with greater vulnerability. Endogenous risk: The coefficient of 0.047028 is not statistically significant (p-value = 0.358206), indicating a non-significant positive influence on the MVI. Vul Inherent: The coefficient of 0.000374 is not statistically significant (p-value = 0.994192), indicating that inherent vulnerabilities do not have a substantial influence on MVI.

Vul Companies: The coefficient of 0.128304 is statistically significant (p-value = 0.003753), indicating that vulnerabilities in businesses correlate to greater levels of multidimensional vulnerability. *Vul Home*s: The coefficient of 0.157584 is statistically significant (p-value = 0.000229), indicating that home vulnerabilities contribute considerably to the MVI. *Capabilities State*: The coefficient 0.004730 is not statistically significant (p-value = 0.932943), showing that state capabilities have no meaningful influence on the MVI.

Social Cohesion Capabilities: The coefficient of 0.167742 is statistically significant (p-value = 6.33e-05), indicating that social cohesiveness in terms of capabilities contributes considerably to increased multidimensional vulnerability.

Model results show a high multiple coefficient of determination (R-squared) of 0.9617, suggesting that the included independent variables account for about 96.2 percent of the variability in the MVI (see Annex).

We use Kaplan-Meier survival analysis to examine the time it takes for a cohort of 24 American countries to experience a 10% increase in unemployment. The Kaplan-Meier curve illustrates the survival probability of each country over the 120-month period, with the x-axis representing time and the y-axis representing the survival probability. Initially, all countries have a survival probability of 1, indicating they are expected to reach the target unemployment rate. As time progresses, some countries may fail to achieve the target, resulting in a gradual decline in survival probability.

Table 1 presents a detailed breakdown of the country failure risk at various time points. At 10 months, there were no countries that failed to reach the target, resulting in a survival probability of 1.000. This indicates that all countries were expected to achieve the target unemployment rate up to that point.

As time goes on, the number of countries at risk decreases, and the number of events (failures) increases, leading to a consistent reduction in survival probabilities. At 50 months, six countries had failed to reach the target, resulting in a survival probability of 0.65. This suggests that around 35% of countries were expected not to achieve the target unemployment rate up to that point.

This trend continues, with a survival probability of 0.55 at 60 months, indicating that approximately 45% of countries were expected not to reach the target unemployment rate up to that point. At 78 months, four countries had failed to achieve the target, resulting in a survival probability of 0.417. This implies that approximately 58.3% of countries were expected not to achieve the target unemployment rate up to that point.

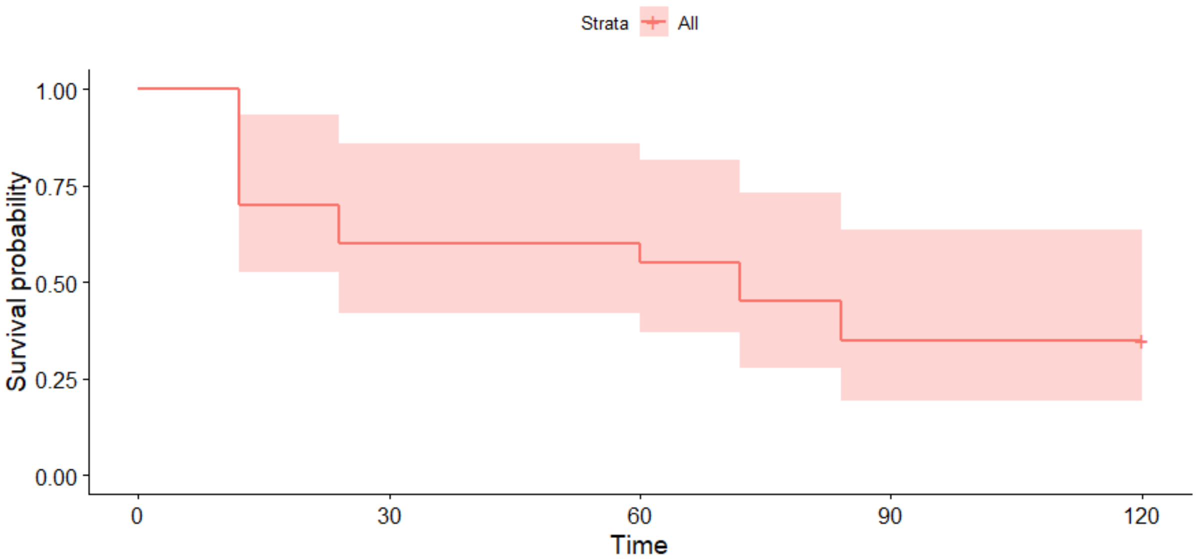

Figure 1. Kaplan-Meier survival curve.

Source: own elaboration.

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
10	20	0	1.000	0.0000	1.000	1.000
12	20	6	0.70	0.102	0.525	0.933
24	14	2	0.60	0.110	0.420	0.858
36	12	0	0.60	0.110	0.420	0.858
60	12	1	0.55	0.111	0.370	0.818
90	7	4	0.35	0.107	0.193	0.636
111	7	0	0.35	0.107	0.193	0.636

Source: own elaboration.

The subsequent time points, at 90 and 120 months, demonstrate a consistent survival probability of 0.35, indicating that approximately 35% of countries were expected not to reach the target unemployment rate during these periods.

These findings highlight the dynamic nature of the risk landscape for countries pursuing economic development goals, with decreasing survival probabilities suggesting heightened failure risks as they strive to reach specific unemployment targets. This emphasizes the importance of strategic decision-making and the potential benefits of proactive risk mitigation strategies.

Understanding time-to-achievement patterns and associated risks can assist stakeholders in evaluating economic development opportunities, designing support mechanisms, and formulating policies to enhance national resilience. The precise estimates obtained from the analysis provide valuable insights for countries seeking to optimize strategies and mitigate potential challenges on the path to economic prosperity.

4.1. Model comparison

Utilizing a collection of pertinent characteristics, the article evaluated the efficacy of several machine learning survival models in forecasting startup failures. For machine learning design, this approach partitioned the dataset into two subsets: a training set and a testing set. 70 percent of the rows from the data frame df are selected at random and assigned to the data.train variable. The train index variable is responsible for storing the row indices in numeric format for data.train. 30 percent of the initial data set is comprised of the remaining rows, which are designated as data.test.

This division enables the training of a model on the designated training set and the subsequent evaluation of its performance on the testing set in order to gauge its efficacy and capacity for generalization. Comparing the prediction capability of various models required the concordance index (C-index).

To ensure the robustness of the findings, a cross-validation process was employed. The dataset was randomly split into k folds, with each fold serving as a testing set once while the remaining folds combined formed the training set. This process was repeated k times, each time using a different fold as the testing set, allowing for a more comprehensive assessment of the model's performance.

Figure 2 presents a comprehensive comparison of five survival models applied to a dataset focused on examining the time it takes for each country to experience a 10% increase in unemployment within a 120-month timeframe. This analysis is particularly relevant in scenarios characterized by a significant number of censoring events on both the right and left sides, where advanced machine learning models often excel over traditional Cox proportional hazards models.

The Concordance Index (C-index) serves as a key metric for evaluating the discriminatory power of these models. Notably, the DeepSurv model emerges with a remarkable C-index of 0.987, highlighting its ability to capture intricate, non-linear relationships inherent in the dataset. This underscores its robust suitability for addressing bidirectional censoring.

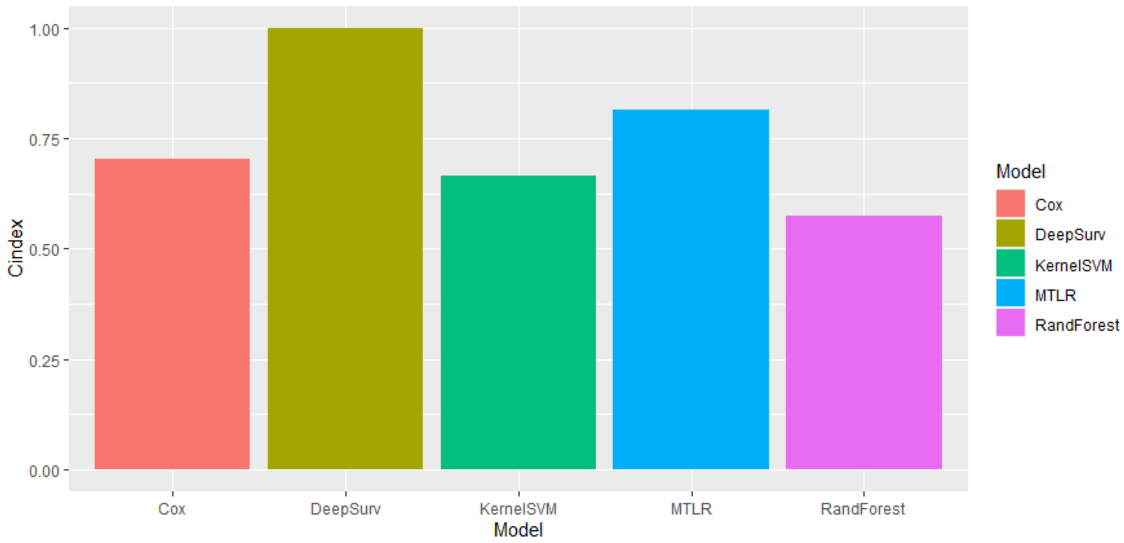

Figure 2. results from different machine learning models.

Source: own elaboration.

In contrast, the Random Forest model demonstrates a C-index of 0.574, suggesting limitations in effectively capturing the complexities of a small sample of censored data. Despite its reputation for handling non-linear and complex relationships, the lower C-index suggests challenges in accurately predicting the time until a country experiences a 10% unemployment increase.

Complementing these findings, both the Cox and MTLR models exhibit identical C-index values of 0.703, showcasing their comparable performance in this analysis. Meanwhile, the KernelSVM model presents a C-index of 0.667, indicating moderate discriminatory power compared to the other models.

This comparative analysis highlights that in scenarios with a limited number of observations and a low percentage of censored data, traditional models and more complex machine learning models may exhibit similar performance. While advanced machine learning models, especially DeepSurv, demonstrate superior performance in efficiently managing datasets with censoring at both extremes under certain conditions, it is important to note that conventional approaches such as Cox proportional hazards may perform poorly in addressing the specific challenges posed by bidirectional censoring, especially with a small sample size and a low proportion of censored data. More complex models, such as MTLR, may be more effective in these cases.

4.2. Economic perspective

The weights in the MTLR weight matrix indicate the influence of each variable on the time until unemployment in American countries increases by 10%. In general, variables with larger weights have a more significant influence on the time to reach the unemployment increase.

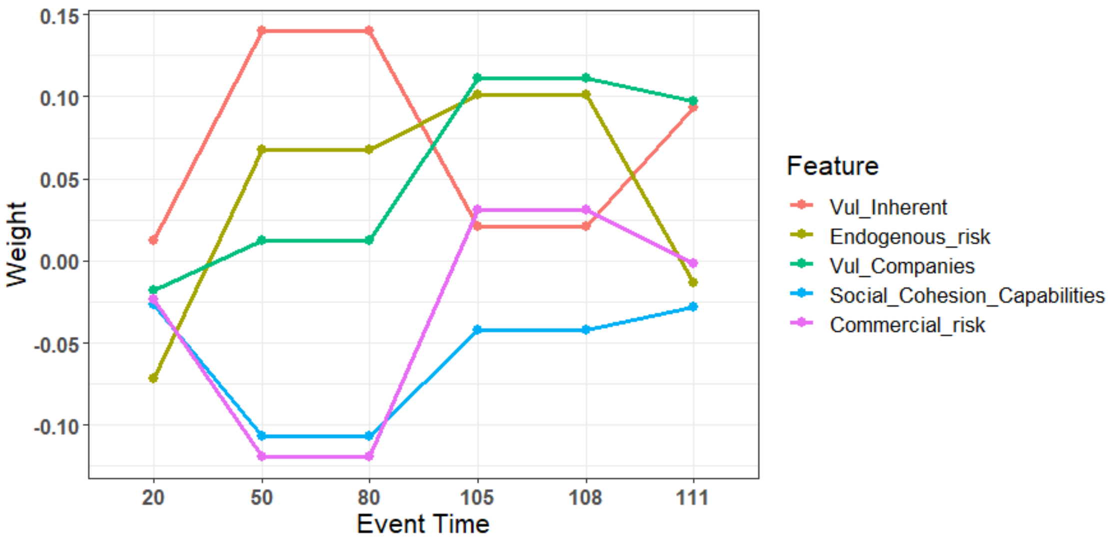

Figure 3. Relative weight of each variable based on MTLR model.

Source: own elaboration.

4.2.1. Variables with positive weights

Natural_risk: Natural risk, such as natural disasters, has a positive weight, indicating that it increases the time to reach the unemployment increase. This is because natural disasters can cause damage to infrastructure and the economy, which can lead to an increase in unemployment. For example, a natural disaster that destroys a factory can lead to layoffs of the factory's workers.

Financial_risk: Financial risk, such as financial crises, also has a positive weight. This is because financial crises can lead to a recession, which can increase unemployment. For example, a financial crisis that causes companies to go bankrupt can lead to layoffs of the workers of those companies.

Vul_Inherent: Inherent vulnerability, such as economic inequality, also has a positive weight. This is because economic inequality can make it difficult to recover from a recession, which can increase unemployment. For example, significant economic inequality can make it difficult for people to find work after a recession.

Vul_Homes: Household vulnerability, such as poverty, also has a positive weight. This is because poor households are more vulnerable to the effects of economic crises, which can increase unemployment. For example, an economic crisis can cause people to lose their homes, which can make it difficult for them to find work.

4.2.2. Variables with negative weights

Capabilities_State: The state of capabilities, such as the quality of education and health, has a negative weight. This indicates that a better state of capabilities can reduce the time to reach the unemployment increase. This is because a better state of capabilities can help people find work and recover from economic crises. For example, a high-quality education system can help people develop the skills they need to find work.

Social_Cohesion_Capabilities: Social cohesion, such as trust between citizens, also has a negative weight. This indicates that greater social cohesion can reduce the time to reach the unemployment increase. This is because greater social cohesion can help people work together to overcome economic crises. For example, a society with high trust between citizens can be more resilient to economic crises.

4.2.3. Specific conclusions by variable

Natural_risk

The positive weight of the variable Natural_risk indicates that natural risk is an important factor that can increase the time until unemployment increases. This is because natural disasters can cause damage to infrastructure and the economy, which can lead to an increase in unemployment. For example, a natural disaster that destroys a factory can lead to layoffs of the factory's workers.

Financial_risk

The positive weight of the variable Financial_risk indicates that financial risk is also an important factor that can increase the time until unemployment increases. This is because financial crises can lead to a recession, which can increase unemployment. For example, a financial crisis that causes companies to go bankrupt can lead to layoffs of the workers of those companies.

Vul_Inherent

The positive weight of the variable Vul_Inherent indicates that inherent vulnerability is an important factor that can increase the time until unemployment increases. This is because economic inequality can make it difficult to recover from a recession, which can increase unemployment. For example, significant economic inequality can make it difficult for people to find work after a recession.

Vul_Homes

The positive weight of the variable Vul_Homes indicates that household vulnerability is also an important factor that can increase the time until unemployment increases. This is because poor households are more vulnerable to the effects of economic crises, which can increase unemployment. For example, an economic crisis can cause people to lose their homes, which can make it difficult for them to find work.

Capabilities_State

The negative weight of the variable Capabilities_State indicates that the state of capabilities is an important factor that can reduce the time until unemployment increases. This is because a better state of capabilities can help people find work and recover from economic crises. For example, a high-quality education system can help people develop the skills they need to find work.

Social_Cohesion_Capabilities

The negative weight of the variable Social_Cohesion_Capabilities also indicates that social cohesion is an important factor that can reduce the time until unemployment increases. This is because greater social cohesion can help people work together to overcome economic crises. For example, a society with high trust between citizens can be more resilient to economic crises.

4.2.4. Policy Recommendations Based on Survival Analysis

The findings of our survival analysis, which examined the time it takes for 24 American countries to experience a 10% increase in unemployment, suggest several policy recommendations for promoting economic resilience and mitigating the risk of unemployment.

Balancing Risks and Rewards: Governments should adopt a balanced approach to economic development, striking a careful equilibrium between pursuing commercial and financial opportunities and managing inherent risks. Calculated risk-taking is essential for growth, but excessive exposure to risk could have detrimental consequences.

Strategically Addressing Vulnerabilities: Policymakers should prioritize strategies that address the fundamental vulnerabilities that can hinder economic progress. This includes strengthening democratic institutions, ensuring sustainable management of domestic resources, and fostering a supportive environment for entrepreneurship.

Investing in Capabilities: Enhancing state capabilities and promoting social cohesion are crucial for long-term economic development. Investing in education, infrastructure, and social welfare programs can empower individuals and strengthen the overall economy.

Contextual Decision-Making: Recognizing the context-specific nature of policy implementation is essential. Tailored strategies that consider the unique challenges and opportunities of each region or country are essential for effective policy outcomes.

Nuanced Understanding: While the weight matrix provides valuable insights into the relative importance of various factors, strategic decision-making requires a comprehensive understanding of the interplay between risks, vulnerabilities, capabilities, and the broader economic context.

Countries aspiring to achieve sustainable economic growth should approach their policies with a careful balance, addressing vulnerabilities, leveraging strengths, and adapting strategies to their specific economic and social landscape.

5. Conclusion

This study provides a comprehensive analysis of the factors that influence the time it takes for 24 American countries to experience a 10% unemployment increase. The findings reveal a complex interplay between various risk factors, vulnerabilities, and capabilities, highlighting the need for a nuanced approach to economic policymaking.

The comparative analysis of survival models demonstrates the superior performance of advanced machine learning models, particularly DeepSurv, in scenarios with data censoring. This underscores the value of machine learning techniques for accurately predicting time-to-event outcomes, particularly in small sample sizes.

The weight matrix analysis provides valuable insights into the relative importance of various risk factors, vulnerabilities, and capabilities. It suggests that a combination of calculated risk-taking, strategic vulnerability mitigation, and investment in state capabilities are crucial for promoting economic resilience and reducing the risk of unemployment.

However, it is important to recognize that the weight matrix represents an aggregated analysis and may not accurately reflect the specific circumstances of individual countries. Tailored policies that consider the unique context of each country are essential for effective policymaking.

In conclusion, the study highlights the dynamic nature of economic factors and the need for a nuanced understanding of their interplay. Governments should adopt a balanced approach that addresses vulnerabilities, leverages strengths, and adapts strategies to their specific economic and social landscape. By doing so, they can promote economic resilience and reduce the risk of unemployment in the American region.

In addition to the findings discussed above, the study also found that the following factors may influence the time it takes for countries to experience a 10% unemployment increase:

- *Fiscal policy*: Countries with more expansionary fiscal policies may be better able to weather economic shocks and reduce the risk of unemployment.
- *Monetary policy*: Countries with more accommodative monetary policies may also be better able to reduce the risk of unemployment.
- *Demographic trends*: Countries with aging populations may be more vulnerable to unemployment shocks.

Future research should explore these factors in more detail to better understand their impact on unemployment risk.

In addition to the recommendations mentioned above, future research should also consider the following:

- The impact of different types of risk-taking on unemployment risk. For example, some types of risk-taking, such as innovation, may be more beneficial for economic growth and employment than other types of risk-taking, such as speculation.

- The effectiveness of different strategies for mitigating vulnerabilities. For example, some strategies, such as strengthening democratic institutions, may be more effective than others, such as reducing the regulatory burden on businesses.
- The interaction between risk factors, vulnerabilities, and capabilities. For example, the impact of a given risk factor may depend on the presence or absence of certain vulnerabilities or capabilities.

By addressing these questions, future research can help to improve our understanding of the factors that influence unemployment risk and inform the development of more effective policies for promoting economic resilience.

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.

Author contributions

The author was solely responsible for all aspects of the research, including conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft, and writing – review & editing.

Appendix

A1. Table 2. Country List.

Argentina	Guyana
Bahamas	Haiti
Barbados	Honduras
Belize	Jamaica
Boliva	Ivlexi co
Brasil	Nicaragua
Canada	Panama
Chile	Paraguay
Colombia	Peru
Costa Rica	Republica Dominicana
Ecuador	San Vicente ylas Granadinas
EI Salvador	Surinam
Estados Unidos de America	Trinidad y Tobogo
Guatemala	Uruguay

```
Call:
lm(formula = MVI ~ ., data = df[, -c(1:2)])

Residuals:
       Min         1Q     Median         3Q        Max 
-0.0302975 -0.0092157 -0.0004935  0.0108401  0.0286271 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  0.169364   0.063444   2.669 0.015631 *  
Natural_risk                 0.126599   0.016991   7.451 6.65e-07 ***
Commercial_risk             -0.047992   0.047499  -1.010 0.325703    
Financial_risk               0.070071   0.021297   3.290 0.004069 ** 
Endogenous_risk              0.047028   0.049875   0.943 0.358206    
Vul_Inherent                 0.000374   0.050679   0.007 0.994192    
Vul_Companies                0.128304   0.038567   3.327 0.003753 ** 
Vul_Homes                    0.157584   0.034364   4.586 0.000229 ***
Capabilities_State           0.004730   0.055431   0.085 0.932943    
Social_Cohesion_Capabilities  0.167742   0.032396   5.178 6.33e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

s: 0.01837 on 18 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared: 0.9617,
Adjusted R-squared: 0.9425 
F-statistic: 50.21 on 9 and 18 DF,  p-value: 6.368e-11
```

A2. Figure 4. MVI and its dependent variables.

A3. Data Sources.

https://data.imf.org/?sk=388dfa60-1d26-4ade-b505-a05a558d9a42&sId=1479329334655
https://www.statista.com/statistics/1086634/emerging-markets-bond-index-spread-latin-america-country/
https://www.theguardian.com/news/datablog/2010/apr/30/credit-ratings-country-fitch-moodys-standard#data
https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG
https://datos.bancomundial.org/indicator/BN.CAB.XOKA.CD
https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG
http://www.worldbank.org/en/research/brief/fiscal-space
https://www.un.org/development/desa/dpad/least-developed-country-category/ldc-data-retrieval.html
https://worldjusticeproject.org/
https://www.visionofhumanity.org/public-release-data/
https://www.visionofhumanity.org/maps/#/
https://www.v-dem.net/data/the-v-dem-dataset/
https://www.v-dem.net/
https://data.worldbank.org/indicator/GC.TAX.TOTL.GD.ZS
https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS
https://www.worldbank.org/en/topic/poverty
https://drmkc.jrc.ec.europa.eu/inform-index
https://www.worldbank.org/en/programs/business-enabling-environment/doing-business-legacy
https://www.hofstede-insights.com/models/national-culture/

References

Assa, J., and Meddeb, R. (2021). Towards a multidimensional vulnerability index. *United Nations Development Programme. Retrieved April*, *14*, 2023.

Azodi, C. B., Tang, J., and Shiu, S. H. (2020). Opening the Black Box: Interpretable Machine Learning for Geneticists. In *Trends in Genetics*, 36(6), 442–455. Elsevier Ltd. https://doi.org/10.1016/j.tig.2020.03.005

Balica, S. F., Dinh, Q., and Popescu, I. (2023). Chapter 4 - Vulnerability and exposure in developed and developing countries: large-scale assessments. In J. F. Shroder, P. Paron, & G. Di Baldassarre (Eds.). *Hydro-Meteorological Hazards, Risks, and Disasters (Second Edition)* (Second Edition, 5, 103–143). Elsevier. https://doi.org/10.1016/B978-0-12-819101-9.00013-3

Barnwal, A., Cho, H., and Hocking, T. (2022). Survival Regression with Accelerated Failure Time Model in XGBoost. *Journal of Computational and Graphical Statistics*, *31*(4), 1292–1302. https://doi.org/10.1080/10618600.2022.2067548

Barrett, J. K., Siannis, F., and Farewell, V. T. (2011). A semi-competing risks model for data with interval-censoring and informative observation: An application to the MRC cognitive function and ageing study. *Statistics in Medicine*, *30*(1), 1–10. https://doi.org/10.1002/sim.4071

Basak, P., Linero, A., Sinha, D., and Lipsitz, S. (2022a). Semiparametric analysis of clustered interval-censored survival data using soft Bayesian additive regression trees (SBART). *Biometrics*, *78*(3), 880–893. https://doi.org/10.1111/biom.13478

Cui, P., Shen, Z., Li, S., Yao, L., Li, Y., Chu, Z., and Gao, J. (2020). Causal Inference Meets Machine Learning. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 3527–3528. https://doi.org/10.1145/3394486.3406460

Cuperlovic-Culf, M. (2018). Machine Learning Methods for Analysis of Metabolic Data and Metabolic Pathway Modeling. *Metabolites*, *8*(1). https://doi.org/10.3390/metabo8010004

Dutta, I., and Mishra, A. (2023). *Economics Measuring Vulnerability to Poverty: A Unified Framework Measuring Vulnerability to Poverty: A Unified Framework **.

Finch, H. (2005). Comparison of Distance Measures in Cluster Analysis with Dichotomous Data. In *Journal of Data Science*, 3.

Gabrikova, B., Svabova, L., and Kramarova, K. (2023). Machine Learning Ensemble Modelling for Predicting Unemployment Duration. *Applied Sciences*, 13(18), 10146. https://doi.org/10.3390/app131810146

Gorfine, M., and Zucker, D. M. (2022). *Shared Frailty Methods for Complex Survival Data: A Review of Recent Advances*. http://arxiv.org/abs/2205.05322

Hair, J. F., and Fávero, L. P. (2019). Multilevel modeling for longitudinal data: concepts and applications. *RAUSP Management Journal*, *54*(4), 459–489. https://doi.org/10.1108/RAUSP-04-2019-0059

Haradal, S., Hayashi, H., and Uchida, S. (2018). Biosignal Data Augmentation Based on Generative Adversarial Networks. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 368–371. https://doi.org/10.1109/EMBC.2018.8512396

Jiang, R. (2022). A novel parameter estimation method for the Weibull distribution on heavily censored data. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *236*(2), 307–316. https://doi.org/10.1177/1748006X19887648

Jin, P., Haider, H., Greiner, R., Wei, S., and Häubl, G. (2021). Using survival prediction techniques to learn consumer-specific reservation price distributions. *PLoS ONE*, *16*(4 April). https://doi.org/10.1371/journal.pone.0249182

Jin Ziwei and Shang, J. and Z. Q. and L. C. and X. W. and Q. B. (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. In W. and W. H. and Z. R. and Z. Y. Huang Zhisheng and Beek (Ed.), *Web Information Systems Engineering – WISE 2020*, (503–515). Springer International Publishing.

Khan, F. M., and Zubek, V. B. (2008). Support Vector Regression for Censored Data (SVRc): A Novel Tool for Survival Analysis. *2008 Eighth IEEE International Conference on Data Mining*, 863–868. https://doi.org/10.1109/ICDM.2008.50

Le-Van, C., and Tran-Nam, B. (2023). Comparing the Harrod-Domar, Solow and Ramsey growth models and their implications for economic policies. *Fulbright Review of Economics and Policy*. https://doi.org/10.1108/frep-06-2023-0022

Maharana, K., Mondal, S., and Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, *3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Mumuni, A., and Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258

Nolan, B., Roser, M., and Thewissen, S. (2019). GDP Per Capita Versus Median Household Income: What Gives Rise to the Divergence Over Time and how does this Vary Across OECD Countries? *Review of Income and Wealth*, *65*(3), 465–494. https://doi.org/10.1111/roiw.12362

Organization of American States (2023). Strategic Counsel for Organizational Development and Management for Results. Internal Report

Raghunathan, T. E. (2004). What Do We Do with Missing Data? Some Options for Analysis of Incomplete Data. *Annual Review of Public Health*, *25*(1), 99–117. https://doi.org/10.1146/annurev.publhealth.25.102802.124410

Thenmozhi, M., Jeyaseelan, V., Jeyaseelan, L., Isaac, R., and Vedantam, R. (2019). Survival analysis in longitudinal studies for recurrent events: Applications and challenges. *Clinical Epidemiology and Global Health*, *7*(2), 253–260. https://doi.org/10.1016/j.cegh.2019.01.013

Vallarino, D. (2023). *Buy when? Survival machine learning model comparison for purchase timing*. arXiv. https://arxiv.org/abs/2308.14343

Vallarino, D. (2024). *A dynamic approach to stock price prediction: Comparing RNN and mixture of experts models across different volatility profiles*. arXiv. https://arxiv.org/abs/2410.07234

Vallarino, D. (2023). Machine learning survival models restrictions: The case of startups time to fail with collinearity-related issues. *Journal of Economic Statistics*, *1*(3), 1–15. https://doi.org/10.58567/jes01030001

Vinzamuri, B., Li, Y., and Reddy, C. K. (n.d.). Pre-Processing Censored Survival Data using Inverse Covariance Matrix based Calibration. In *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING* (Issue 1).

Wang, L., Li, Y., Zhou, J., Zhu, D., and Ye, J. (2017). Multi-task Survival Analysis. *2017 IEEE International Conference on Data Mining (ICDM)*, 485–494. https://doi.org/10.1109/ICDM.2017.58

Wang, P., Li, Y., and Reddy, C. K. (2017). *Machine Learning for Survival Analysis: A Survey*. http://arxiv.org/abs/1708.04649

Yuan, H., Xie, F., Ong, M. E. H., Ning, Y., Chee, M. L., Saffari, S. E., Abdullah, H. R., Goldstein, B. A., Chakraborty, B., and Liu, N. (2022). AutoScore-Imbalance: An interpretable machine learning tool for development of clinical scores with rare events data. *Journal of Biomedical Informatics*, *129*, 104072. https://doi.org/10.1016/j.jbi.2022.104072

Yunita, R., Gunarto, T., Marselina, M., and Yuliawan, D. (2023). The Influence of GDP per Capita, Income Inequality, and Population on CO2 Emission (Environmental Kuznet Curve Analysis in Indonesia). *International Journal of Social Science, Education, Communication and Economics (SINOMICS JOURNAL)*, *2*(2), 217–230. https://doi.org/10.54443/sj.v2i2.130

Zelenkov, Y. (2020). Bankruptcy prediction using survival analysis technique. *Proceedings - 2020 IEEE 22nd Conference on Business Informatics, CBI 2020*, *2*, 141–149. https://doi.org/10.1109/CBI49978.2020.10071

Zhao, Z. L., Yu, H. J., and Cheng, F. (2022). An Analysis of Factors Affecting Agricultural Tractors' Reliability Using Random Survival Forests Based on Warranty Data. *IEEE Access*, *10*, 50183–50194. https://doi.org/10.1109/ACCESS.2022.3172348

Zhou, F., Fu, L., Li, Z., and Xu, J. (2022). The recurrence of financial distress: A survival analysis. *International Journal of Forecasting*, *38*(3), 1100–1115. https://doi.org/10.1016/j.ijforecast.2021.12.005