Review of Economic Assessment 2024 3 (1) 1-19

Review of Economic Assessment

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

# Temporal Dynamics of Countries' Journey to Cluster-Specific GDP per Capita: A Comprehensive Survival Study

Diego Vallarino [a, *]

[a] *Independent Researcher, Madrid, Spain*

## ABSTRACT

This research delves into the temporal dynamics of a nation's pursuit of a targeted GDP per capita level, employing five different survival machine learning models, remarkably Deep Learning algorithm (DeepSurv) and Survival Random Forest. This nuanced perspective moves beyond static evaluations, providing a comprehensive understanding of the developmental processes shaping economic trajectories over time. The economic implications underscore the intricate balance required between calculated risk-taking and strategic vulnerability mitigation. These findings guide policymakers in formulating resilient economic strategies for sustained development and growth amid the complexities inherent in contemporary economic landscapes.

## KEYWORDS

Temporal dynamics; Survival analysis; Machine learning models; Economic development; Calculated risk-taking; Vulnerability mitigation; Developmental processes; Resilient economic strategies

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

ISSN 2972-3515
doi: 10.58567/rea03010001

Received 2 January 2024, Accepted 6 February 2024, Available online 19 February 2024

## 1. Introduction

In contemporary economic research, there exists a critical imperative to delve into the temporal dynamics that underlie a nation's progression toward achieving specific GDP per capita levels. The examination of economic development trajectories, particularly in terms of GDP per capita, extends beyond a mere quantitative evaluation of a country's prosperity. It necessitates a nuanced understanding of the intricate interplay among various socio-economic factors over time (Yunita et al., 2023). Our scholarly pursuit aims to contribute to this discourse by emphasizing the temporal dimension, transcending traditional static analyses, and providing a more comprehensive comprehension of the developmental processes at play.

The motivation for this study lies in the significance of understanding the time required for countries to reach specific GDP per capita thresholds, a pursuit with far-reaching implications for policymakers, economists, and stakeholders. This understanding facilitates the deciphering of the efficacy of different developmental strategies, the discernment of the impact of external shocks on growth trajectories, and the identification of potential bottlenecks hindering economic advancement (Dutta & Mishra, 2023). By focusing on the temporal aspect, our research seeks to unravel intricate patterns and relationships characterizing the developmental journey, offering insights that extend beyond traditional static analyses.

As we embark on this scholarly journey into the temporal intricacies of GDP per capita dynamics, we emphasize the importance of incorporating rigorous methodologies, empirical analyses, and data-driven insights. Adopting a time-centric lens allows researchers not only to contribute to theoretical advancements in economic literature but also to formulate more effective policies aimed at expediting and sustaining economic growth (P. Wang et al., 2017). This exploration into the temporal dimension paves the way for a more informed and nuanced discourse, fostering a deeper understanding of the multifaceted processes shaping the economic destinies of nations.

Commonly used models to study GDP, such as the neoclassical growth model, Keynesian model, Solow growth model, and Harrod-Domar model, provide different perspectives on factors influencing GDP growth. Each model, whether emphasizing technology, capital accumulation, aggregate demand, government policies, labor, capital, technological progress, or investment, contributes valuable insights into the dynamics of GDP and aids economists and policymakers in understanding and managing economic performance (Le-Van & Tran-Nam, 2023).

However, despite the richness of existing models, our investigation reveals a gap in the focus on survival analysis within the realm of GDP per capita prediction. In this paper, we present the results of our survival model comparison and explore their economic implications. Our research assesses the effectiveness of various machine learning survival models, including the Cox, Kernel SVM, DeepSurv, Survival Random Forest, and MTLR models, in projecting the time until a country reaches a GDP per capita level equal to the median of its respective cluster. The concordance index (C-index) serves as a benchmark for comparing the performance of these machine learning algorithms.

Our primary goal is to identify the model that offers the most accurate and informative prediction of the time until a country achieves a GDP per capita level equal to the median of its cluster. Additionally, we aim to comprehend the economic significance of the model's conclusions. To achieve this, we scrutinize the significance and magnitude of vulnerabilities and risks associated with the Multidimensional Vulnerability Index (MVI) of the United Nations Development Programme (Assa, J., et al., 2023; OAS, 2023) and compare the results to economic theory and intuition.

Our research seeks insights into the factors influencing GDP per capita levels and a deeper understanding of how various survival models may contribute to the establishment of public policies. We explore the outcomes of our model comparison and provide an economic interpretation of these findings.

The subsequent sections of the paper are structured as follows. First, we offer a theoretical perspective on the survival analysis models employed in our study, followed by a brief examination of works that employ survival analysis to address the subject of financial collapse. Subsequently, we delve into the empirical analysis, covering the models used, the data source, and the assessment measures.

The analytical results, including a comparison of different models, are then presented, and we explore the economic ramifications of our findings. Finally, we summarize the key findings and discuss their implications for future research and policymaking. In essence, our research contributes to the literature on the use of survival analysis in finance and provides insights into the features that lead to financial catastrophes, aiding policymakers in developing more effective regulations to prevent such disasters in the future.

## 2. Theorical perspective

Survival analysis emerges as a critical tool within the realm of economic analysis, offering a nuanced lens through which to examine dynamic processes such as business longevity, economic policy impacts, and the persistence of economic phenomena over time. The significance of survival analysis in the economic context lies in its ability to accommodate time-to-event data, providing a robust framework for assessing durations until specific economic events occur (Jin et al., 2021; P. Wang et al., 2017; Zelenkov, 2020; Zhou et al., 2022). However, the crux of its effectiveness hinges on a meticulous understanding of the nature and quality of the available data.

Recognizing the type of data at hand is pivotal, as inaccuracies or misinterpretations can substantially impact the reliability of survival analyses. As an academic researcher navigating the intersection of survival analysis and economic inquiry, it becomes imperative to critically evaluate and discern the intricacies of the data landscape, ensuring that the analyses conducted contribute meaningfully to our understanding of temporal economic phenomena and inform more informed decision-making processes (Finch, 2005; Gorfine & Zucker, 2022; Maharana et al., 2022; Mumuni & Mumuni, 2022).

There are several types of data that may affect the use of these methods. For instance, basic patient data consisting of demographic and clinical information such as sickness stage, comorbidities, and treatment (Hair & Fávero, 2019; Maharana et al., 2022).

Using datasets containing censored data, competitive risk data, or even data that displays longitudinal patient information might potentially be problematic (Barrett et al., 2011). Survival algorithms are applicable despite some restrictions (Cuperlovic-Culf, 2018; Jin et al., 2021).

In the next paragraphs, we will discuss the implications of this kind of data for survival analysis.

### *2.1. Baseline Agent Data*

Essential to the development of a survival strategy for healthcare practitioners is basic patient information. Along with clinical data such as illness stage, comorbidities, and treatment history, demographic information such as age, gender, race, and ethnicity may have a substantial influence on a patient's survival rate (Hair & Fávero, 2019; Haradal et al., 2018; Mumuni & Mumuni, 2022). Although we have evidenced the characteristics of the baseline data with the most representative example in the literature on survival analysis, the most characteristics apply to the analysis of customers in the purchasing process, employees in the work process in the company, or the company in its life process over the years.

Developing a survival strategy requires in-depth understanding and analysis of several variables that might influence a patient's prognosis (Cuperlovic-Culf, 2018; Jin et al., 2021). Various algorithms, including trees, forests, neural networks, deep learning, multitasking, boosting, and "others," may be used by healthcare providers to create survival forecasts (Thenmozhi et al., 2019; Zhao et al., 2022). To minimize mistakes and biases, it is vital to consider the limits and restrictions of these algorithms while generating predictions (Azodi et al., 2020). Therefore, fundamental patient data is crucial for constructing accurate survival algorithms for successful patient care, but it is often insufficient for achieving a satisfactory performance in machine learning models.

### 2.2. Censored Data

The idea of suppressing data is characteristic of survival data. If the event of interest is death or bankruptcy of a company, the event time is censored for participants who are still alive at the conclusion of the research. This implies that the statistical analysis must continue without knowledge of the subject's date of death (Basak et al., 2022a; Jiang, 2022; Vinzamuri et al., n.d.).

The only information available on his death is that it occurred after the conclusion of the research. In general, people who drop out of follow-up research are censored since they are often lost to follow-up and the timing of their occurrence is unclear (Raghunathan, 2004). The date of the occurrence is unobserved, but it is not a missing data point either, since these two categories of unobserved data have distinct properties and empirical interpretations (Yuan et al., 2022).

For right-censored topics, the sole known fact is that their incident happened after the censorship period. If the research had been continued (or if the volunteers had not dropped out), ultimately the result of interest would have been seen for all participants (Basak et al., 2022a, 2022b; Jiang, 2022). Conventional statistical approaches for analyzing survival data assume censoring to be independent or non-informative (Khan & Zubek, 2008).

This implies that, at a given point in time, the subjects who remain in follow-up have the same future risk for the occurrence of the event as those who are no longer followed (either due to censorship or abandonment of the study), as if the losses to follow-up were random and therefore not informative (Basak et al., 2022b).

Current research clearly demonstrates that the handling of censored data is essential in order to have an accurate view of the survival analysis experiment to be conducted (Jiang, 2022). Therefore, the focus of this study will be to determine the optimal ways for integrating censored data, both from the right (the most prevalent in analytic models) and from the left. The latter have not been widely analyzed in the literature, although time-to-event statistical analysis may give a lot of hints (Cui et al., 2020; Yuan et al., 2022).

When dealing with survival data, it is typical to encounter censored data, which happens when the precise timing of an event is unknown, but it is known that the event did not occur before or after a certain period. There are three forms of censored data: right-censored data, interval censored data, and left-censored data. There are a number of excellent algorithms available for managing massive quantities of filtered data (Cui et al., 2020; Yuan et al., 2022).

Survival Random Forest is one method that can effectively manage restricted data. It is an approach for machine learning that builds numerous decision trees and combines their predictions (Jin et al., 2021; Jin Ziweiand Shang, 2020; Zhao et al., 2022). Multi-Tasking Linear Regression (MTLR) is an additional approach that can effectively manage censored data. It employs a Bayesian technique to estimate the survival time distribution and is beneficial when dealing with many outcomes (L. Wang et al., 2017). XGboost is another a well-known algorithm that can deal with enormous amounts of censored data with both continuous and categorical variables (Barnwal et al., 2022).

In the forthcoming sections of this research, we will delve into an examination of how various machine learning models for survival analysis perform in predicting the time until diverse countries achieve the GDP per capita of their respective clusters. This investigation aims to scrutinize the efficacy and comparative strengths of these models in capturing the temporal dynamics inherent in economic development across distinct clusters. By evaluating the predictive capabilities of machine learning algorithms in this context, we seek to contribute insights that can refine our understanding of the temporal dimensions associated with the attainment of specific GDP per capita levels, offering valuable implications for economic forecasting and policy formulation.

## 3. Research Design

This empirical investigation delves into the dynamics of economic growth across 160 countries. Employing a nuanced approach, the dataset comprises countries with left-censored (64), right-censored (59), and those reaching the specified event within 144 months (33). This comprehensive sampling strategy ensures a thorough representation of diverse economic scenarios. Furthermore, the research integrates information from authoritative sources such as the World Bank, IMF, and the United Nations, harnessing data from 160 countries to capture multifaceted vulnerabilities (Assa et al., 2021)

In addressing the central query, the research seeks to unravel the intricate web of variables and risks influencing the time required for a country to achieve the average GDP per capita. By leveraging statistical techniques, particularly survival analysis, the study aims to provide a robust understanding of the temporal aspects of economic development. 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, $\beta$ 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 , $\beta_{kj}$ 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, $\alpha_i$ 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:

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

Where $h(t|x)$ is the hazard rate for an individual with covariates $x$ and $\beta_i$ 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 $\theta$ parameters.

## *3.2. Data*

This empirical study looks at the patterns of economic development in 160 nations. Using a sophisticated methodology, the dataset includes nations that have been left-censored (64), right-censored (59), and those that have reached the desired event within 144 months (33). This extensive sampling technique guarantees a complete representation of many economic situations.

For the analysis, a *status* of 1 was assigned if the country reached the per capita GDP of the median of its cluster between 2009 and 2021. In other words, if a country did not reach that per capita GDP level in those 12 years (144 months), the data is considered right-censored. If the country has already reached the mentioned level, it is identified as left-censored data (*status* of 1 in 2009 or before). Furthermore, the study incorporates information from authorized sources such as the World Bank, IMF, and UN, as well as data from 160 nations to identify multiple risks (see Annex).

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

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.

Vul_Hogares: 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_Empresas: Assessing business-related challenges, Vul_Empresas 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 2009 to 2021, measuring the lifespan of a country within its respective GDP per capita cluster. This timeframe serves as a crucial variable influencing several aspects, including market conditions, customer preferences, and industry trends. These temporal dynamics play a pivotal role in shaping the strategies and decisions of countries in their economic development.

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 median GDP per capita of its cluster. A value of 0 indicates that the country has not achieved the cluster's median income, while a value of 1 signifies that the country has attained or surpassed this threshold. This status variable provides a snapshot of the country's economic performance and its alignment with the cluster's income level, distinguishing between countries that persist in their economic development and those that may face challenges or have ceased to exist in the market.

The countries have been categorized into 7 clusters, and the mean and median of GDP per capita have been calculated for each cluster, as outlined in the accompanying table (Annex).

**Table 1.** Comparative Analysis of Median and Average GDP per Capita by Cluster (2009-2021) in current US$.

| | Median GDP per capita/Cluster (2009-2021) | Average GDP per capita/Cluster (2009-2021) |
|---|---|---|
| Cluster 1 | 51.454 | 55.556 |
| Cluster 2 | 30.104 | 31.080 |
| Cluster 3 | 17.319 | 23.515 |
| Cluster 4 | 6.149 | 8.738 |
| Cluster 5 | 2.654 | 3.249 |
| Cluster 6 | 1.639 | 2.925 |
| Cluster 7 | 1.302 | 1.393 |

*Source: own elaboration*

## *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 censored data C-Index.

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

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

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

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

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

## 4. Results

The regression analysis presented examines the Multidimensional Vulnerability Index (MVI) as the dependent variable, with several independent variables representing different aspects of risk and vulnerability. The interpretation of the economically significant results is detailed below.

The intercept (0.064240) signifies the estimated MVI when all other independent variables are zero, implying the inherent vulnerability in the absence of specific risks or factors. Commercial risk, financial risk, endogenous risk, and other variables represent the change in MVI associated with a one-unit increase in the respective independent variable, holding all others constant.

Commercial risk exhibits a statistically significant positive impact on the MVI, implying that an increase in commercial risk corresponds to an elevated level of multidimensional vulnerability. Similarly, financial risk also displays a statistically significant positive association with the MVI, indicating that heightened financial risk is linked to increased vulnerability. Notably, endogenous risk demonstrates the most substantial positive impact on the MVI, suggesting a significant influence of internal risks on vulnerability.

Certain variables, such as inherent vulnerability, fragility related to domestic resources, and state capabilities, do not exhibit statistical significance in affecting the MVI at conventional levels. Conversely, variables like democracy fragility, vulnerabilities in companies and homes, and social cohesion in terms of capabilities, all show a statistically significant positive impact on the MVI. These findings imply that factors such as democratic fragility, vulnerabilities in businesses and households, and social cohesion contribute to a higher level of multidimensional vulnerability.

Model statistics reveal a high multiple coefficient of determination (R-squared) of 0.911, indicating that approximately 91.1% of the variability in the MVI is explained by the included independent variables. The low p-value associated with the F-statistic (< 2.2e-16) signifies the overall significance of the model.

When we go deeper in the survival analysis, the Kaplan-Meier curve visually illustrates the survival probabilities of a cohort of 160 countries over time. The x-axis represents the duration (in months) until reaching the median GDP per capita of their respective clusters, while the y-axis illustrates the survival probability. Initially, all countries are assumed to be "alive" with a survival probability of 1. As time progresses, some countries may experience failure, leading to a decline in survival probability.

```
Call:
lm(formula = MVI ~ ., data = df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.067249 -0.015323  0.001079  0.016923  0.052715

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                      0.064240   0.024677   2.603 0.010371 *
Commercial_risk                  0.073897   0.011173   6.614 1.03e-09 ***
Financial_risk                   0.065800   0.017615   3.735 0.000285 ***
Endogenous_risk                  0.274280   0.024983  10.979  < 2e-16 ***
Vul_Inherent                     0.010698   0.021456   0.499 0.618956
Vul_Fragility_Democracy          0.096715   0.013583   7.121 7.88e-11 ***
Vul_Fragility_Domestic_Resources 0.002734   0.015751   0.174 0.862486
Vul_Companies                    0.093510   0.022200   4.212 4.85e-05 ***
Vul_Homes                        0.107205   0.025640   4.181 5.46e-05 ***
Capabilities_State               0.036997   0.031654   1.169 0.244748
Social_Cohesion_Capabilities     0.162935   0.024478   6.656 8.28e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

s: 0.02501 on 123 degrees of freedom
  (26 observations deleted due to missingness)
Multiple R-squared: 0.911,
Adjusted R-squared: 0.9038
F-statistic: 125.9 on 10 and 123 DF,  p-value: < 2.2e-16
```

**Figure 1.** MVI and independent variables.

*Source: own elaboration*

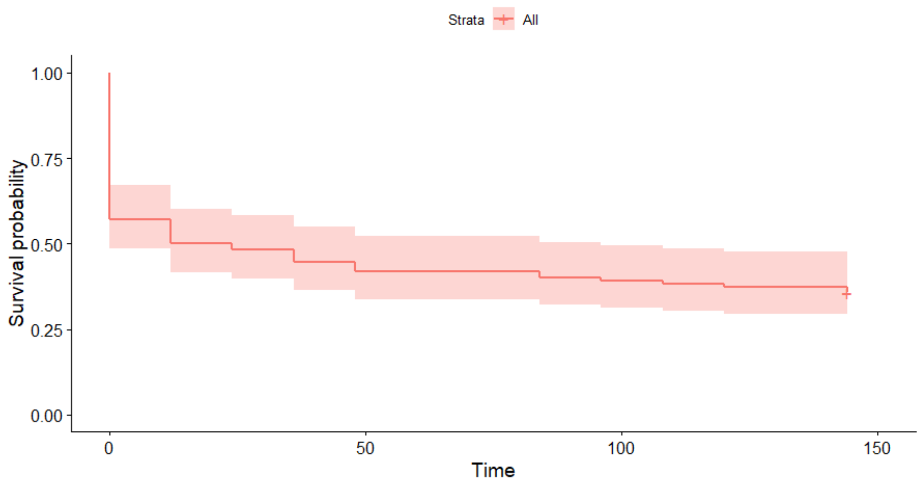



**Figure 2.** Kaplan-Meier survival curve.

*Source: own elaboration*

The table provides a detailed breakdown of the evolution of country failure risk at specific time points. For instance, at 10 months, 64 countries were at risk, with 48 events (failures), resulting in a survival probability of 0.871. This indicates that approximately 87.1% of countries were expected not to reach the median GDP per capita of their clusters up to that point.

As time advances, the number of countries at risk diminishes, accompanied by an increase in failures, contributing to a gradual reduction in survival probabilities. At 50 months, 47 countries were at risk, with 17 events, resulting in a survival probability of 0.420. This suggests that around 42.0% of countries were expected not to reach the median GDP per capita of their clusters up to that point.

**Table 2.** 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 | 64 | 48 | 0.871 | 0.0468 | 0.487 | 0.671 |
| 50 | 47 | 17 | 0.420 | 0.0466 | 0.338 | 0.522 |
| 80 | 47 | 0 | 0.420 | 0.0466 | 0.338 | 0.522 |
| 105 | 44 | 3 | 0.393 | 0.0461 | 0.312 | 0.495 |
| 108 | 44 | 1 | 0.384 | 0.0460 | 0.304 | 0.485 |
| 111 | 43 | 0 | 0.384 | 0.0460 | 0.304 | 0.485 |
| 130 | 42 | 1 | 0.375 | 0.0457 | 0.295 | 0.476 |
| 144 | 42 | 2 | 0.375 | 0.0451 | 0.279 | 0.458 |

*Source: own elaboration*

Continued analysis reveals a consistent decrease in survival probabilities. For example, at 80 months, 47 countries were at risk, with no events, resulting in a survival probability of 0.420. This indicates that approximately 42.0% of countries were expected not to reach the median GDP per capita of their clusters up to that point.

These findings underscore the changing risk landscape for countries, with decreasing survival probabilities suggesting heightened failure risks as they strive to reach the median GDP per capita of their respective clusters. This emphasizes the challenges faced by nations in sustaining economic development and the crucial role of strategic decision-making.

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*

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.

The presented table offers a comparative overview of various survival models applied to a dataset characterized by a substantial number of censoring events both on the right and left sides. This analysis is pivotal as, in scenarios featuring prevalent censoring at both ends, advanced machine learning models tend to outperform conventional Cox proportional hazards models.

The Concordance Index (C-index) serves as a metric for evaluating the discriminatory power of these models. Notably, the DeepSurv model stands out with a remarkable C-index of 0.888889. This exceptional performance underscores DeepSurv's ability to effectively capture intricate, non-linear relationships inherent in the dataset, making it a robust choice for addressing bidirectional censoring.

Furthermore, the Random Forest model demonstrates noteworthy performance, boasting a C-index of 0.715702. The strength of Random Forest lies in its capacity to handle non-linear and complex relationships, coupled with its resilience to overfitting. This makes it particularly well-suited for datasets exhibiting high levels of censoring and inherent complexities.

The figure 3 depicting the relationship between predictor variables and the cumulative hazard, often referred to as "mortality," over time in the context of a random survival forest model.

In this plot, the y-axis labeled as "mortality" signifies the cumulative hazard, representing the cumulative risk or probability of the event of interest (e.g., failure, mortality) occurring up to a given time point. The x-axis typically represents time or some other relevant scale.

Each variable in the plot is represented by a bar, and the height of the bars indicates the variable's importance in predicting the cumulative hazard. Taller bars imply that the corresponding variable has a more substantial impact on the cumulative hazard, reflecting its significance in influencing the risk of the event over time.

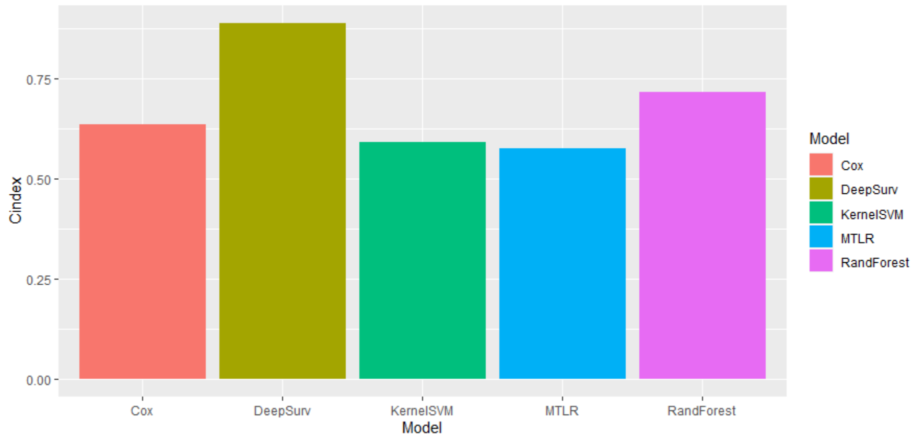



**Figure 3.** Results from different machine learning models.

*Source: own elaboration*

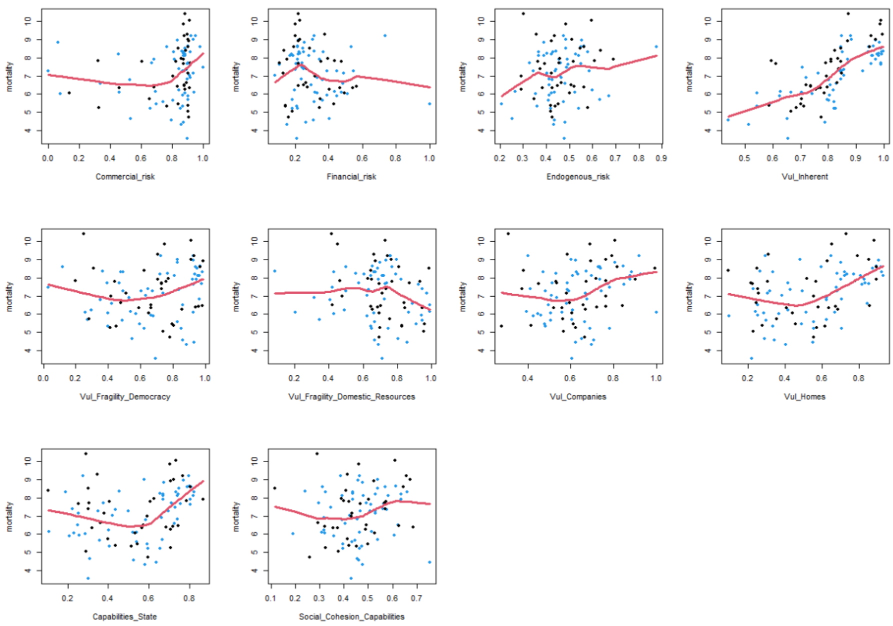



**Figure 4.** Random Survival Forest Model variables relationship.

*Source: own elaboration*

*Note: the importance of variables is represented by the length or height of the bars. The taller the bar, the more important the variable is in predicting survival outcomes.*

*4.2 Economic perspective*

Let's delve into a more detailed and descriptive analysis of the weight matrix, providing a deeper understanding of the economic implications of each variable:

*Commercial Risk, Financial Risk, and Endogenous Risk*: Positive weights on these risk factors suggest that, to some extent, embracing certain levels of commercial, financial, and endogenous risks may be associated with achieving the GDPpc target. This could be indicative of an entrepreneurial environment where taking calculated risks could lead to economic growth. However, an excessive reliance on risky strategies might also amplify the potential for economic downturns or financial crises.

Commercial_risk: Despite the consistently negative coefficient (-0.013113018 at time 50 and -0.001570893 at time 111), the magnitude may vary. In general, higher commercial risk consistently decreases the probability of reaching the event.

Financial_risk: The consistently positive coefficient (0.04194862 at time 50, to 0.01840382 at time 140) indicates that financial risk consistently contributes to an increase in the probability of reaching the event. The model suggests that, within the defined scope, higher financial risk is consistently linked to a higher likelihood of achieving the specified economic event over time. Possible explanations could involve economic mechanisms, policy implications, or other factors that align with the observed patterns in the data. Possible interpretations imply: (a) Investor Perception: Higher financial risk, as perceived by investors in emerging markets, might be associated with economic conditions that are conducive to achieving higher GDP per capita. (b) Risk-Return Dynamics: The positive relationship could reflect a risk-return tradeoff, where higher financial risk is linked to the potential for higher economic returns or growth. And (c) Policy Implications: It could also suggest that in certain contexts, policies or conditions associated with higher financial risk are conducive to achieving the specified economic event. It's important to note that correlation does not imply causation, and further analysis and domain-specific expertise are needed to fully understand the underlying reasons behind these relationships.

Endogenous_risk: The consistently negative coefficient (-0.005939786 to -0.008532349) suggests that higher endogenous risk consistently decreases the probability of reaching the event.

*Vulnerabilities (Inherent, Fragility - Democracy*, *Domestic Resources, Companies, Homes*): Positive weights on vulnerability variables imply that countries facing higher inherent vulnerabilities, fragility in democratic institutions, and vulnerabilities in domestic resources, companies, and homes might find themselves better positioned to attain the GDPpc target. This raises questions about the potential benefits of targeted economic development strategies in vulnerable areas. However, these vulnerabilities could also signify challenges that need to be addressed for sustained economic progress, as excessive fragility may lead to economic instability.

Vul_Inherent: The consistently positive coefficient (0.07718029 to 0.01124327) indicates that inherent vulnerability consistently contributes to an increase in the probability of reaching the event.

Vul_Fragility_Democracy: The coefficient for this variable shows a mixed pattern over time. Initially, there are two positive coefficients (0.01531536 and -0.01267351), indicating that fragility in democracy increases the probability of reaching the event. However, towards the end of the data, there are four consecutive negative coefficients (-0.02448684, -0.01504001, -0.003224489), suggesting that fragility in democracy later decreases the probability of reaching the event.

Vul_Fragility_Domestic_Resources: Although consistently negative (-0.009576292 to -0.026474982), the magnitude may vary. Overall, higher fragility of domestic resources consistently decreases the probability of reaching the event.

Vul_Companies: The consistently negative coefficient (-0.001671509 to -0.016664838) indicates that vulnerability of companies consistently decreases the probability of reaching the event.

Vul_Homes: With a consistently positive coefficient (0.028418399), vulnerability of homes consistently contributes to an increase in the probability of reaching the event. In practical terms, if household-level vulnerabilities represented by "Vul_Homes" worsen or increase, the model suggests that the likelihood of reaching the event (specific GDP per capita target) also increases over the analyzed time period.

*Capabilities (State, Social Cohesion):* Positive weights on state capabilities and social cohesion highlight their critical role in achieving the GDPpc target. A well-functioning state, with effective institutions and governance structures, is likely to create an environment conducive to economic development. Social cohesion, indicating stability and unity within a society, contributes positively as it fosters collaboration and shared economic goals.

Capabilities_State: This variable represents the capabilities of the state and has a consistently positive coefficient (0.009647651). In practical terms, as the capabilities of the state increase, the probability of reaching the specified event (achieving the GDP per capita of the median of the country's cluster) consistently increases over time. The term "capabilities" could encompass a range of factors, including the effectiveness of governance, public policies, and the state's capacity to facilitate economic development.

Social_Cohesion_Capabilities: This variable captures the combined impact of social cohesion and capabilities and has a consistently positive coefficient (0.018017232). This suggests that higher levels of social cohesion coupled with broader capabilities are associated with a consistently increased probability of reaching the specified event over time. Social cohesion could involve factors such as community resilience, inclusiveness, and social stability, while capabilities may encompass the broader capacity of communities and institutions.

In both cases, the positive coefficients indicate that an increase in the respective capabilities (state or social cohesion and capabilities) contributes positively to the likelihood of achieving the specified economic event. These interpretations are based on the assumption that other variables remain constant. Further analysis and domain-specific expertise would be necessary to delve into the specific factors contributing to these relationships.

*Bias*: The positive bias suggests an underlying positive force that influences reaching the GDPpc target, capturing unobserved factors. This could include favorable global economic conditions, geopolitical stability, or other externalities that positively impact economic performance.

On the basis of the stated findings, what public policies may be suggested as possible?

*Balancing Risks and Rewards*: Countries aiming for the GDPpc target should carefully balance the pursuit of commercial and financial opportunities with the inherent risks. Successful economic development often involves calculated risk-taking, but excessive risk exposure could lead to undesirable consequences.

*Addressing Vulnerabilities Strategically*: Understanding the specific vulnerabilities contributing positively to the target is crucial. Policymakers may need to strategically address vulnerabilities, such as enhancing democratic institutions and ensuring sustainable management of domestic resources.

*Investing in Capabilities*: The positive impact of state capabilities and social cohesion underscores the importance of investing in institutions and social structures. Governments should prioritize policies that strengthen governance, reduce corruption, and foster social harmony to support long-term economic development.

*Contextual Decision-Making*: Recognizing the context-specific nature of these weights is paramount. The economic landscape varies across regions and countries, necessitating tailored strategies that consider the unique challenges and opportunities present.

While the weight matrix offers valuable insights, strategic decision-making requires a nuanced understanding of the interplay between risks, vulnerabilities, capabilities, and the broader economic context. Countries aspiring to reach the GDPpc target 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 academic work has delved deeply into the intricacies of the Kaplan-Meier and Cox survival analysis and an extensive comparison of five machine learning models, providing a profound understanding of the temporal dynamics and predictive capacities germane to countries aspiring to attain the median GDP per capita within their respective clusters.

The Cox survival curve, serving as a temporal lens, has unveiled a nuanced narrative of the evolving risk landscape. As countries traverse the trajectory toward economic benchmarks, the diminishing survival probabilities encapsulate the complex interplay of factors and the formidable challenges entwined with sustained economic development.

The model comparison has been a pivotal exploration, shedding light on the nuanced efficiency of diverse survival models, particularly in the context of bidirectional censoring scenarios. Advanced machine learning models, prominently DeepSurv and Random Forest, have showcased exceptional prowess, eclipsing traditional models in navigating the intricate tapestry of complexities arising from censoring at both extremities. The judicious use of the Concordance Index has not only provided a metric for comparison but has underscored the unparalleled ability of DeepSurv to discern intricate non-linear relationships inherent in the dataset.

From an economic standpoint, the exhaustive analysis of the weight matrix has unraveled multifaceted implications associated with commercial risk, financial risk, endogenous risk, vulnerabilities, capabilities, and the discerned positive bias. This analysis transcends a mere surface-level examination, delving into the microeconomic and macroeconomic implications of each variable. It reveals a nuanced tapestry where embracing certain levels of commercial, financial, and endogenous risks may indeed be associated with achieving the GDP per capita target. This nuanced understanding suggests that fostering an entrepreneurial environment, where calculated risks are embraced, may catalyze economic growth. However, a cautious note is sounded, emphasizing the potential pitfalls of an excessive reliance on risky strategies, which might amplify the vulnerability to economic downturns or financial crises.

The positive weights attributed to vulnerability variables raise profound questions about the potential benefits of targeted economic development strategies in areas facing inherent vulnerabilities, democratic fragility, and challenges in domestic resources, companies, and homes. It prompts policymakers to consider whether addressing these vulnerabilities strategically could be a key to unlocking economic potential. However, the cautionary note is clear—excessive fragility may pose a threat to economic stability, demanding a delicate balancing act in policy formulation.

The positive weights on state capabilities and social cohesion underscore the critical role of governance structures and societal unity in achieving the GDP per capita target. This goes beyond a mere economic perspective; it underscores the societal and institutional underpinnings that form the bedrock of economic development. It beckons governments to prioritize policies that strengthen governance, reduce corruption, and foster social harmony for sustained economic progress.

The positive bias in the weight matrix introduces a dimension of unobserved factors that exert a positive influence on reaching the GDP per capita target. While not explicitly identified, these factors could include favorable global economic conditions, geopolitical stability, or other externalities that positively impact economic performance. Acknowledging this positive bias adds a layer of complexity to the economic narrative, prompting policymakers to consider not only the observable variables but also the hidden forces shaping economic outcomes.

The policy implications arising from the nuanced economic analysis presented herein are intricate and underscore the imperative for an all-encompassing and contextually sensitive approach to policymaking. The detailed examination of risk factors reveals the necessity for nations to engage in a meticulous equilibrium between the pursuit of commercial and financial opportunities and the attenuation of inherent risks. While economic development necessitates calculated risk-taking, prudence is advised against undue risk exposure, as it may precipitate undesirable consequences.

Strategic vulnerability mitigation emerges as a pivotal facet of the policy recommendations, as indicated by the positive weights attributed to vulnerability variables. In light of these findings, policymakers are enjoined to devise and implement strategic initiatives tailored to address specific vulnerabilities that contribute positively to the attainment of the GDP per capita target. The call for targeted interventions encompasses areas such as the fortification of democratic institutions and the sustainable management of domestic resources.

The affirmative impact of state capabilities and social cohesion on economic development emphasizes the pivotal role played by governance structures and societal unity. Governments are earnestly encouraged to prioritize policy initiatives directed at fortifying governance, curbing corruption, and fostering social harmony. Such endeavors are deemed essential in cultivating an environment conducive to sustained long-term economic progress.

The imperative of contextual decision-making permeates the policy discourse. Acknowledging the context-specific nature of the weight matrix variables is deemed paramount in informing policy formulation. Tailored strategies, accounting for the unique challenges and opportunities inherent in each country or region, are deemed requisite. This necessitates an acute awareness of the diversity characterizing economic landscapes and underscores the importance of bespoke approaches to policymaking.

The policy implications arising from this nuanced economic analysis underscore the imperative for an all-encompassing and contextually sensitive approach to policymaking. Nations are advised to find a meticulous equilibrium between pursuing opportunities and attenuating inherent risks. Strategic vulnerability mitigation, targeted interventions, and the fortification of democratic institutions are emphasized, along with the importance of governance structures and societal unity in achieving sustained economic progress.

In considering future research directions, a section on the research shortcomings and potential avenues for further investigation is essential. While this study provides valuable insights, it is not without limitations. Future research could explore the robustness of the findings across different geographical regions or consider additional variables that might influence the GDP per capita target. Additionally, a more in-depth analysis of the social and cultural aspects impacting economic development could enrich the understanding of the intricate dynamics involved. Furthermore, exploring the applicability of the models in different economic contexts or under varying external conditions could contribute to a more comprehensive understanding of their predictive capabilities.

In essence, this research extends beyond traditional economic boundaries, offering a holistic view that intertwines microeconomic and macroeconomic considerations. The integration of machine learning models, survival analysis, and economic interpretation is not merely a technical exercise but a holistic endeavor that contributes to the broader discourse on the multifaceted nature of economic development. It is our hope that this academic inquiry serves as a guiding beacon for policymakers, urging them toward nuanced and contextually informed decision-making in the pursuit of sustained economic prosperity.

## Funding Statement

This research received no external funding.

## Acknowledgments

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

## Conflict of interest

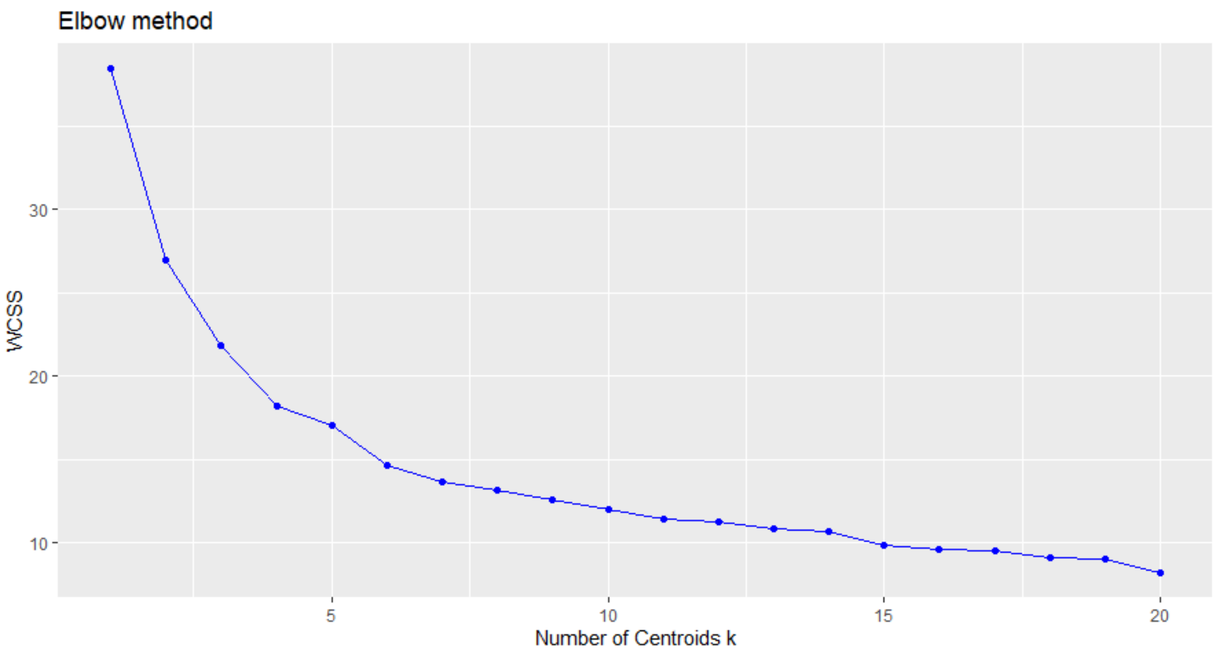
The author claims that the manuscript is completely original. The author also declares no conflict of interest.

# Appendix

**A1.** Country Cluster Mapping**.**

| Country | Cluster | Country | Cluster | Country | Cluster | Country | Cluster |
|---|---|---|---|---|---|---|---|
| 1 Sweden | 3 | 46 Guyana | 1 | 85 Guinea-Bissau | 4 | 130 Ukraine | 2 |
| 2 Denmark | 3 | 47 Greece | 1 | 86 Paraguay | 4 | 131 Congo, Rep. | 2 |
| 3 Australia | 3 | 48 Peru | 1 | 87 Burundi | 4 | 132 Belarus | 2 |
| 4 Finland | 3 | 49 Latvia | 1 | 88 Bolivia | 4 | 133 Nigeria | 2 |
| 5 Germany | 3 | 50 Botswana | 1 | 89 El Salvador | 4 | 134 Nepal | 2 |
| 6 Austria | 3 | 51 Colombia | 1 | 90 Romania | 4 | 135 Equatorial Guinea | 2 |
| 7 Switzerland | 3 | 52 North Macedonia | 1 | 91 Togo | 4 | 136 Myanmar | 2 |
| 8 Norway | 3 | 53 China | 1 | 92 Morocco | 4 | 137 Syrian Arab Republic | 2 |
| 9 Netherlands | 3 | 54 United Arab Emirates | 1 | 93 Benin | 4 | 138 Lebanon | 2 |
| 10 Canada | 3 | 55 South Africa | 1 | 94 Timor-Leste | 4 | 139 Kyrgyz Republic | 2 |
| 11 United Kingdom | 3 | 56 Croatia | 1 | 95 Mongolia | 4 | 140 Senegal | 2 |
| 12 France | 3 | 57 Malaysia | 1 | 96 Mauritania | 4 | 141 Central African Republic | 2 |
| 13 Singapore | 5 | 58 Rwanda | 1 | 97 Albania | 4 | 142 Chad | 2 |
| 14 Belgium | 5 | 59 Philippines | 1 | 98 Guatemala | 4 | 143 Eswatini | 2 |
| 15 Japan | 5 | 60 Suriname | 1 | 99 Jordan | 4 | 144 Zimbabwe | 2 |
| 16 Israel | 5 | 61 Dominican Republic | 1 | 100 Kazakhstan | 4 | 145 Eritrea | 2 |
| 17 United States | 5 | 62 Nicaragua | 1 | 101 Jamaica | 4 | 146 Guinea | 2 |
| 18 Spain | 5 | 63 Papua New Guinea | 1 | 102 Malawi | 4 | 147 Moldova | 2 |
| 19 New Zealand | 5 | 64 Oman | 1 | 103 Bosnia and Herzegovina | 4 | 148 Uzbekistan | 2 |
| 20 Cyprus | 5 | 65 Panama | 1 | 104 Serbia | 4 | 149 Uganda | 2 |
| 21 Estonia | 5 | 66 Argentina | 1 | 105 Gabon | 4 | 150 Libya | 2 |
| 22 Hong Kong SAR, China | 5 | 67 Saudi Arabia | 1 | 106 Tanzania | 4 | 151 Bangladesh | 2 |
| 23 Italy | 5 | 68 Armenia | 1 | 107 Ecuador | 4 | 152 Ethiopia | 2 |
| 24 Chile | 5 | 69 India | 1 | 108 Vietnam | 4 | 153 Burkina Faso | 2 |
| 25 Czechia | 5 | 70 Cuba | 1 | 109 Lesotho | 4 | 154 Venezuela, RB | 2 |
| 26 Mauritius | 5 | 71 Georgia | 1 | 110 Comoros | 4 | | |
| 27 Slovenia | 5 | 72 Indonesia | 1 | 111 Djibouti | 4 | | |
| 28 Poland | 5 | 73 Bahrain | 1 | 112 Sri Lanka | 4 | | |
| 29 Brazil | 5 | 74 Belize | 1 | 113 Liberia | 4 | | |
| 30 Ireland | 0 | 75 Turkiye | 1 | 114 Zambia | 4 | | |
| 31 Uruguay | 0 | 76 Haiti | 1 | 115 Egypt, Arab Rep. | 4 | | |
| 32 Costa Rica | 0 | 77 Mexico | 1 | 116 Sierra Leone | 4 | | |
| 33 Portugal | 0 | 78 Russian Federation | 1 | 117 Lao PDR | 4 | | |
| 34 Qatar | 0 | 79 Algeria | 1 | 118 Afghanistan | 4 | | |
| 35 St. Lucia | 0 | 80 Mali | 1 | 119 Iran, Islamic Rep. | 4 | | |
| 36 St. Kitts and Nevis | 0 | 81 Namibia | 1 | 120 Tajikistan | 4 | | |
| 37 Barbados | 0 | 82 Congo, Dem. Rep. | 1 | 121 Cote d'Ivoire | 2 | | |
| 38 Hungary | 0 | 83 Ghana | 1 | 122 Kenya | 2 | | |
| 39 Bahamas, The | 0 | 84 Tunisia | 1 | 123 Mozambique | 2 | | |
| 40 South Korea | 0 | 155 Cambodia | 6 | 124 Sudan | 2 | | |
| 41 Trinidad and Tobago | 0 | 156 Cameroon | 6 | 125 Iraq | 2 | | |
| 42 Kuwait | 0 | 157 Somalia | 6 | 126 Gambia, The | 2 | | |
| 43 Thailand | 0 | 158 Angola | 6 | 127 Azerbaijan | 2 | | |
| 44 Bhutan | 0 | 159 Madagascar | 6 | 128 Honduras | 2 | | |
| 45 Bulgaria | 0 | 160 Pakistan | 6 | 129 Yemen, Rep. | 2 | | |

**A2.** Number of Clusters Centroids.



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

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

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

Barnwal, A., Cho, H., & 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., & 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., & Lipsitz, S. (2022). 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., & 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., & 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. *Journal of Data Science* 3. http://dx.doi.org/10.6339/JDS.2005.03(1).192

Gorfine, M., & Zucker, D. M. (2022). *Shared Frailty Methods for Complex Survival Data: A Review of Recent Advances*.

https://doi.org/10.48550/arXiv.2205.05322
Hair, J. F., & 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., & 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., & 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, Z., 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. *Web Information Systems Engineering – WISE 2020*, 503–515. https://doi.org/10.1007/978-3-030-62008-0_35
Khan, F. M., & 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., & 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., & 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., & 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., & 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., & 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
Vinzamuri, B., Li, Y., & Reddy, C. K. (2017). Pre-Processing Censored Survival Data using Inverse Covariance Matrix based Calibration. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, *29*(10), 2111-2124. https://doi.org/10.1109/TKDE.2017.2719028
Wang, L., Li, Y., Zhou, J., Zhu, D., & 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., & Reddy, C. K. (2017). *Machine Learning for Survival Analysis: A Survey*. https://doi.org/10.48550/arXiv.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., & 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., & 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., & 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., & 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