International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration

L. P. Himali
Department of Economics and Statistics, Sabaragamuwa University of Sri Lanka, P.O. Box 02, Belihuloya, Sri Lanka. 12400
Email address: hima099@gmail.com

***Abstract*—** *Unemployment is a major socioeconomic problem worldwide, causing economic vulnerability and mental health problems in people and society alike. This study investigates unemployment patterns in Sri Lanka, especially the duration of unemployment, using standard survival analysis methods combined with modern machine learning algorithms such as XGBoost and Random Forest. Using national labour force survey data from 2017 to 2019, the study uses survival analysis models, which are frequently utilised in medical studies, to determine survival functions and proportional hazards while focusing on their constraints across diverse socioeconomic situations via machine learning techniques' superior non-linear pattern identification capabilities. The study employs survival and machine learning models such as Kaplan-Meier, Cox Proportional Hazards, Weibull, Random Survival Forest, XGBoost, and DeepSurv to assess unemployment duration and associated risk factors. The findings highlight key determinants of unemployment in Sri Lanka, including education level, age, sectoral concentration, and geographic location, with unskilled youth and rural women being particularly vulnerable. Predictive modelling emphasises the need for policy actions addressing gender inequities and geographical inequalities, and aligning education with labour market demands. Furthermore, the research places unemployment in the perspective of Sri Lanka's long-term economic slump, underlining the importance of comprehensive solutions to address both acute and structural labour market difficulties. This study provides useful information for researchers and policymakers by combining survival analysis with machine learning, adding to a data-driven approach to Sri Lanka's labour market improvements.*

***Keywords*—** *Data-driven Analysis, Non-linear Pattern Detection, Policy Modelling, Predictive Modelling, Unemployment Duration.*

I. Introduction

Unemployment is one of the most alarming and shocking socio-economic phenomena which affects both people and societies across the globe. The implications of unemployment levels range from lack of wealth to psychological disorders, as well as the growth of the economy as a whole. Hence, the factors determining the level and the length of unemployment are areas of concern for policymakers and researchers in the quest for policies which may lessen the burden of various forms of unemployment. As far as Sri Lanka is concerned, unemployment is a problem which assumes specific forms and needs specific forms of analysis, since it is a relatively underdeveloped country with a specific socio-economic context.

This paper is more or less subjective in nature on the topic of focus, which is about the patterns of unemployment spell, or rather duration of joblessness, in the case of the people of Sri Lanka, in this case, combining both standard survival analysis and the new age artificial intelligence. Survival analysis is a statistical technique which has become popular in the field of medicine as it seeks to explain time-to-event occurrences, focusing on the effects of certain variables among other issues.

The singular application of survival analysis seems to have gaps in the potential and intuition that may exist when concentrating on or including more familiar factors, such as socioeconomic factors and unemployment, over a duration period. For this, machine learning models like XG Boost and Random Forest seem to have been extremely well suited to deal with such situations due to their predictive power, which stems from the concealed patterns present in the data available.

Few in Sri Lanka. Out of the total number of unemployed people, the unemployed from the groups of single-headed households, natural disasters and localised violence and unemployed children are most likely to defer their search for employment longer than any other groups. Therefore, this paper, in addition to a few Inactive individuals, also provides information which is helpful to a more complete unemployment policy in Sri Lanka.

Sri Lanka's socio-economic situation has changed dramatically over the last decades. Theory relating to the tenets of Voluntary Market Economy took its roots a few decades back and has gone through historical transformations, paving the way for a Global Economy. Such changes have brought several problems, such as unemployment amongst certain segments of the population, due to skill mismatch within its labour market. Is the issue of the span of unemployment, for a length of time, the unemployed individual becomes socially marginalised, ranked lower in the social pedestal than others in good standing, and the quality of life degenerates? Given the non-regulatory nature of unemployment in any country, it is difficult to accurately measure the number of unemployed people and compensation for an injury. With most people still viewing Cox's proportional hazard model, unemployment duration is the focal point of discriminating between various scales of unemployment. Cox's models have widespread approval in the literature due to their ease of use. However, unemployment estimates rely on the proportional hazards and linearity approaches, which for most socioeconomic data are

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

too simplistic. Through all this work, Kuznets has made socio-economic dedications.

These models do not have the constraint of making several stringent assumptions which accompany traditional survival analysis models, and these models have a great capability of non-linear pattern recognition. Two models, XG Boost and Random Forest, are among the best in terms of predictive performance and would be perfect candidates for enhancing the analysis of the unemployment duration, which is of great concern in countries like North America and Europe where these forms of machine learning have also been deployed to predict and model trend of unemployment. However, in the case of Sri Lanka, with its socio-economic context being distinct, the use of such techniques has been very limited. This gap is addressed by the current study, in which one will move beyond survival analysis and also utilise machine learning in analysing unemployment duration in Sri Lanka to provide a more complex view of the forces affecting unemployment in this setting.

II. Problem Statement

Unemployment duration has been a very active field of study in developed nations, but most of the studies conducted in this field historically employed survival analysis-type statistical models, which could not adequately account for the complexity of the Sri Lankan socio-economic environment. In particular, the Sri Lankan labour market presents some unique challenges due to great demographic variations as well as the disability conditions causing unemployment.

This unique challenge is linked to structural inflexibility, cyclical changes, and educational mismatch, which is mostly caused by an imbalance between the formal urban and informal rural sectors. The formal sector, which includes government employment and large commercial firms, pays high benefits and salaries but fails to take in the increasing workforce due to poor economic growth and institutional inefficiencies, whereas the informal sector takes excess labour at cheap wages.

The majority of Sri Lanka's unemployment, unlike other developing countries, results from a mismatch between education and the job market requirements. The educational system generates graduates with academic knowledge but without practical skills, resulting in structural unemployment in industries like as IT and manufacturing. The government's focus on white-collar jobs leads to inefficient labour allocation. Sri Lanka's youth unemployment is also a major issue, with many graduates rejecting job offers owing to misaligned expectations, resulting in extended unemployment due to the 'desire trap'.

Sri Lanka's economy, which depends primarily on tourism and exports, is at risk from external factors like global economic downturns and COVID-19, resulting in cyclical unemployment. The country's lack of economic diversity and gender inequities make employment challenging, with low female labour force participation and hidden unemployment. Addressing these issues is essential for developing employment rates.

The case of Sri Lanka is more complicated because this is a developing country, it has certain socio–economic challenges that cannot be easily solved through the application of classical models. Some of the labour market conditions include the highly informal labour market, occupational uneconomical wide spreads and shortages of occupational training programs, which affect the duration of unemployment among the affected persons. The other factor is the existence of a disability, which makes it more difficult to get employed, thus increasing the duration of unemployment of disabled persons. It is the generation that gets unemployed for such long periods that should be a concern for all because if they do not get appropriate timely aid and support, statistics shoot up, and so does the adverse effect unemployment has on society and the economy.

Such considerations were not addressed in previous studies, and methods relying on machine learning algorithms were not used to analyse the duration of time a person is unemployed. This particular research seeks to bridge this specific gap by merging machine learning with survival techniques to provide a more general framework for the duration of unemployment in the context of Sri Lanka.

III. Objectives

The aims of the research include, but are not limited to, the analytical examination of the survival functions of unemployment in Sri Lanka using advanced economy-specific techniques such as the econometric approach and machine learning. On the two above-mentioned methodologies, the research will produce their strong points into a more complex model capable of predicting the time until a person remains unemployed.

IV. Significance of the Study

Long-lasting unemployment is a concern for almost every policymaker in the world and even more so for those in developing nations like Sri Lanka. Unemployment has far-reaching consequences for individuals, the economy, and society as a whole in the form of lower productivity, higher rates of poverty, and social instability. This makes it all the more important to understand all the primary reasons which govern the increase in the period of occupational unavailability, as this can make this research assist in formulating specific policies aimed at reducing the occurrence of unemployment and improving the functioning of the labour market. The current work attempts to provide new aspects in the examination of unemployment duration spells through the application of machine learning techniques to commonplace survival analysis. As such, unemployment models that employ machine learning techniques can present a better understanding of the phenomenon of prolonged unemployment by improving accuracy in the prediction of variables that have linear and complex relationships with each other.

The combination of both concepts will enable this research to better accommodate all the required information on the unemployment process in Sri Lanka, which is useful for all stakeholders who hold the employment policy there.

V. Literature Review

There is much more to consider aside from the population aspects, for the country's unemployment problem has also to do with the uniform poverty level across the nation. The Sri

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

Lankan labour market includes a broad informal economy which employs a large number of people. Workers in the informal sector usually possess an insecure employment relationship, poorly paid jobs, and limited coverage by social protection provisions, which further increases susceptibility to unemployment (World Bank, 2019). The absence of any rural non-farming jobs leads to an increase in the duration of unemployment of those individuals who could not find a secure, high-salary job (Gunasekara, 2021).

Furthermore, the nature of the human resource market in Sri Lanka has been influenced by the historical perspective during the transitional period involving a pre-transition economy to the change to being more market-oriented economy. As a result, structural unemployment is increased, whereby some groups of the population, such as aged workers and low levels of education, are more susceptible to being unemployed (ILO, 2013). The initiatives of the country as far as combating unemployment through education and vocational studies promotion have not produced satisfactory results, as it has become evident that graduates are produced without the contemporary requirements for the labour market being addressed (Perera & Nandasiri, 2020).

There are many reasons and characteristics which determine how long a person remains unemployed, for example, age, education level, gender, or whether a person has a disability or not. In the context of Sri Lanka, these aspects are aggravated by the specificities of labour supply and demand in the country, in particular, a lack of proper skills, substantial informal employment and inadequate vocational training. Therefore, policies that tackle the most pressing issues for addressing unemployment in Sri Lanka would focus on improving education and vocational training, ensuring gender equality in the labour force, and increasing employment opportunities for persons with disabilities.

Unemployment in modern society is an exacerbating and interrelated socio-economic challenge, affecting both individuals and entire economies on an international scale. The time needed to remain unemployed, or the unemployment duration, greatly varies within certain ethnicities and sociological and economic settings.

The interplay existing between unemployment and socio-economic factors is a subject of considerable inquiry, especially when it comes to the impact of such variables as age, education, gender, marriage, disability, and many others on labour market performance. This section looks at the socioeconomic variables that determine the length of time a person remains unemployed, drawing from different international and regional studies, but with particular emphasis on those conducted in Sri Lanka to present some of the complexities surrounding unemployment in less developed countries.

For older workers, the problem is different. Relating to it, older persons are victims of age discrimination that results in long unemployment periods, as well as having higher wage expectations or not having all the necessary skills. There is evidence that older people, for example, people approaching retirement age, who lose their jobs are often never re-employed (Chan & Stevens, 2001). Older workers are thought to be averse to change or integrating new technologies into the workplace, which sometimes makes it difficult for them to go back to the workforce (Van Dalen & Henkens, 2020). Where the Sri Lankan context is analysed, older workers often seem to reside in the labour constraints. Older workers do have difficulties in securing other forms of employment after losing theirs (Gunasekara, 2021).

Education is another significant aspect that should be looked at regarding unemployment duration. As a general rule, people with higher education levels tend to have shorter periods of unemployment. This is because they have been exposed to factors such as education and training, which give them the right exposure to opportunities in the job market (Psacharopoulos & Patrinos, 2018). The higher the education, the more opportunities and better-paying jobs with stable durations that lower the unemployment level (World Bank Report, 2019). Nonetheless, concerning Sri Lanka, the education–unemployment nexus appears more intricate. On the one hand, advanced education can expand one's opportunities for employment. On the other hand, the education system is blamed for failing to produce a broad range of skills that the economy requires (Perera & Nandasiri, 2020). Many graduates in Sri Lanka acquire degrees in areas of study which are not in demand and which contribute to a high level of graduate unemployment (ILO, 2013). The problem of educated unemployment is most severe in Sri Lanka, as there has been no economic growth in the formal sector to absorb the thousands of graduates who enter the labour market every year (Gunasekara, 2021).

Gender and Unemployment Duration: The variations based on gender concerning the unemployment duration have been outlined in varying labour markets. More often than not, women tend to be unemployed for longer periods than men, especially in patriarchal societies where gender roles and expectations limit female labour force participation (ILO, 2020). The duration of unemployment, on the other hand, is also known to vary by gender, which in itself has some correlates, including discrimination in the labour market, socialisation of women into family roles, and limited education and training opportunities for women (García-Peñalosa, 2020).

The gendered nature of unemployment in Sri Lanka is also reflected by the types of work that women can access. Unemployment among Women is prevalent mostly in the tea and garment sectors due to the volatility associated with these industries (Gunasekara, 2021). These industries, in turn, are dominated by low-skill women who tend to occupy low-paid jobs with no chances of promotion opportunities. Thus, during recession periods, women become more susceptible to losing their jobs and more unemployed for longer periods (ILO, 2020).

In Sri Lanka, it is even harder for people with disabilities, specifically because there are no resources directed towards occupational integration, which leads to the aforementioned issues (Perera & Nandasiri, 2020). Why do people think that hiring a PWD is not a good idea? Well according to ILO "Researchers highlight some of the biases that employers have towards hiring disabled employees owing to their productivity and other issues that affect their work" Hence people with disability are exposed to higher unemployment times since in a lot of them opportunities are restricted and ill-equipped while

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

still being lacking in provisions that improve employment chances of them (Gunasekara, 2021).

The unemployment predictive modelling field has progressed with the coming of machine learning (ML). The global labour markets are undergoing significant transformations in the context of Industry 4.0, which renders traditional statistical models obsolete. Moreover, these traditional regression, or survival analysis-based models are overly reliant on strong bounding distributions, which are unable to capture the non-linear associations and embedded patterns within complex multi-faceted datasets.

Predictive modelling using Machine Learning, on the other hand, presents a versatile and powerful approach because of its speed in processing large volumes of data, parsing through layers of complexity, and improving accuracy rates in forecasting. The uniqueness of this approach stems from its non-parameterised image of the world, as grade distribution does not place bounds and limits on the data, allowing outstanding models to explore non-linear relationships.

The expanding use of machine learning techniques in unemployment projection has been highlighted in more recent articles, with an emphasis on the capacity to combine numerous datasets, assess changes over time, and enhance policymakers' choices. To develop strategic predictive actions, these models make use of recalled knowledge obtained from earlier labour market research and economic data. LSTM networks, neural networks, hybrid algorithms, and ensemble approaches like Random Forest and XG Boost are a few of the predictive models that use machine learning algorithms to address unemployment. All of these models have demonstrated excellent levels of accuracy and prediction.

A different, commonly used method in predicting unemployment uses regression analysis and neural networks. These models are particularly effective when trying to design suggests that artificially intelligent models for neural networks are useful in places where economic and social predictions and factors to determine expectations of unemployment economies are volatile, and working. These networks flexibly deal with non-linear neural dependencies associated with the transformations of macroeconomic factors into employment trends. This allows greater flexibility in dealing with the non-linear complexities of many phenomena. In contrast to traditional regression models that have linear bounds, neural networks can account for non-linear dependencies. Hence, their flexibility in dealing with these complexities is more predictive of the rapidly fluctuating state of the economy.

The use of hybrid models has also become advanced in accuracy within the context of unemployment prediction since they combine various machine learning models. Such models that involve decision trees, deep learning models, and ensembles outdo single-method models. Alharbi and Al-Alawi (2024) show that random forest and XG Boost models surpass other hybrid models significantly.

Unemployment rates are affected by several factors and should be considered for setting up predictive models. Over macroeconomic indicators like economic growth and inflation, which impact labour market trends. Based on the expansion or contraction of economies, the types and quantity of jobs will vary accordingly and should be included in any unemployment prediction model (Chukwuere, 2024). By processing vast datasets of historical economic growth patterns, inflation rates, and employment statistics, machine learning algorithms, especially deep learning models, can yield insights that help policymakers with real-time insights into future labour market trends.

Education levels and technological progress are another key factor in predicting unemployment. There is a well-established relationship between education and work that shows that higher education levels are associated with lower unemployment. Yet, with swift technological advancements across industries, the availability of jobs shifts and workers are compelled to adapt their skills to maintain financial stability (Chukwuere, 2024). Machine learning techniques, especially hybrid models, have also been playing an important role in inspecting the effects of technological advancement on workforce trends, detecting skill mismatches, and predicting potential job displacement due to automation.

However, machine learning has shown potential as a game-changer in predictive modelling for the studies of unemployment, providing unprecedented precision and adaptability in the analysis of increasingly complex labour market patterns. Methods such as LSTM networks, neural networks, and a shame of the above-mentioned and also ensemble learning models such as Random Forest and (extreme) gradient boosted trees (XG Boost) outperformed any other algorithms in terms of unemployment rate prediction. Fusing multiple datasets & incorporating fundamental economic and demographic factors, these models help policymakers move away from anecdotal musings and instead make targeted workforce planning and policies addressing unemployment and related issues.

Moving forward, issues regarding data availability, algorithmic bias, and model interpretability remain challenges to be addressed to effectively leverage the potential of machine learning approaches to unemployment prediction. We recommend that moving forward, new models are developed that include as many socioeconomic variables as possible, relating to both local populations and those deploying models, and better data quality to improve relevance, as well as ethical AI frameworks that ensure both fairness and transparency of models. Lastly, the incorporation of real-time data and adaptive machine learning models can be investigated to make unemployment prediction models more sensitive to the rapid changes in the labour market.

VI. Methodology

The primary data source for this research is the national labour force surveys conducted annually in Sri Lanka, which provide systematic information on employment and demographics. And other factors on an individual level. There is a focus on the working age groups and includes a subset of the unemployed who are actively seeking work. The dataset is cross-sectional for each year but combined to create a panel in 2017, 2018, and 2019, allowing the researcher to triangulate unemployment duration in only a matter of a few years when repeated cross-section data are available.

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

Variables and Operational Definitions

Unemployment Duration is the main outcome variable and was defined in terms of a person who tried to search for a job but failed to find it, and the period or the duration of such failure, i.e. weeks or months. This can also be termed as a time to an event and is a metric used in survival analysis where the event is either getting employment or other factors that lead to exit from the workforce.

Independent Variables:

- Demographic Characteristics: These are essential baseline factors and goals for analysing the variations in the amount of time a person is unemployed.
- Age: Age is a continuous variable measured in years, so it implies that one would expect to have a relative effect on any of the job opportunities offered in the market, as young people may have different employment needs than older job applicants do.
- Marital Status: Applicable to individuals who have either of the categories, such as single, married, divorced or widowed. An individual's marital status has a psychological and financial impact on the need for and determination to work.
- Educational Attainment: Refers to the ordinal measures on the highest level of education acquired by an individual. Education is a significant determinant of people's chances of being employed, their skills, and their competitiveness in the job market.
- Disability Conditions: It is one of the variables of interest because a person with a disability is likely to face different challenges in the labour market as compared to a person without.
- Employment Search Efforts: This group includes activities performed in the last four weeks to seek work and is reported as binary variables, yes or no, for each action performed by an individual.
- Employment Strategies: Training program participation is divided as follows:
  - Participation in formal job training programs: Reporting whether the person underwent any such job training is a binary variable.
  - Participation Non-currently cited reasons: Pertaining such categorical data that addresses reasons why the individual is absent, such as economic issues or unavailability.
- Expectations regarding employment: The appropriation of a particular variable describes individuals' subjective perceptions and preparedness regarding workforce participation and their engagement in employment.
- Job readiness: Evaluated based on a scale designed to elicit an individual's sense of job preparedness, which will influence the motivation and chances to seek employment.
- Sector preferences: Labelled variables showing whether individuals are keen on a certain category of job, such as private, public employment or self-employment.

Data Preprocessing

While the source of data remains the national labour force surveys. It, however, begs the need for a comprehensive preprocessing section that emphasises the constituents of data quality and integrity. In this research, such missing variables will be overcome by employing imputation techniques. For instance, some variables are more continuous, and these variables deal with mean imputation or median, whereas categorical ones' deal with nearest-neighbour or mode imputation. Analysts will also adjust the problem of outliers, for instance, data points that fall outside the interquartile range (IQR) or those with z-scores above 3 will be dealt with through transformation or exclusion. Moreover, for continuous variables, normalising or scaling techniques will be adopted to prepare the data for machine learning models, thereby ensuring that no single variable dominates the result. The major steps of data preparation are:

1. Data Cleaning: Any discrepancies in survey data are corrected, including, but not limited to, typographical errors and incorrect data entry. They are complementarily considered to be outliers in cases where they are suspected to be extreme in value and need to be screened, whether to be included, transformed or disregarded entirely.
2. Missing Data Replacement Strategies: Data acquisition processes, however mundane, may sometimes lead to data unavailability, for instance, the absence of responses or failure of respondents to complete the survey at hand. In such scenarios where a dataset contains numerous missing fields (e.g, age), it is only logical to use measures of central tendency within the age scope, like mean or median, for the variables affected, whereas validation of mode or nearest neighbour values could be sufficient for categorical variables, e.g., marital status. If huge amounts of data are missing, it is best practice to select either list-wise deletion or multiple imputation to maintain the strength of the dataset.
3. Consolidating Data across Years: It is pertinent to note that data has been cross-sectional, hence the three-year timelines, 2017, 2018, and 2019, must be combined to form a panel data set to maintain the integrity of the data. Therefore, it is significant that the parameters are adjusted to maintain equivalence across several years, thus providing a true indication of the length of time individuals remained unemployed.
4. Transformation of Non-Numeric Attributes: Machine learning is done with non-numeric values like the marital status and the type of disability of the participant, thus they need to be transformed into numeric values. For the most part, this entails one hot encoding of the categorical variables so that they enter the algorithm in the anticipated binary format.
5. Scaling and Normalisation: Variables like age and how long someone remains unemployed are examples of continuous variables which require normalisation or scaling before they are run through machine learning models. This is particularly significant concerning models that are influenced by the scale of the variable, especially the models that are distance-based.
6. Censoring: In survival analysis, some participants in the research may still be unemployed by the study's endpoint, which results in censored data. This process precedes the

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

analysis, where the intention is to ensure that censored data is classified appropriately so that the aim of the analysis, which is to model the time-to-event outcomes accurately, is not compromised.

Analytical Approach

In the context of this study, the use of an analytical approach combines survival analysis and machine learning to assist in the simulation and prediction of the duration of unemployment among individuals willing to work actively in Sri Lanka. They would seem to believe that the weaker and the stronger performers do not differ in their average non-linear relationships during the span of this study. This analytical method is developed in 2 steps. The first one deals with traditional methods of survival analysis. The second one deals with methods based on predictive machine learning models. All these forms of estimation not only predict the length of time persons will remain unemployed but also understand the factors affecting it as well.

Survival analysis is a technique that is used specifically for time-to-event data, the duration of unemployment for this study. This modelling would be ideal for us, considering that it focuses on unemployment duration bias, where some people are not looking for work at the end of the study period. The model will also take into consideration the popularisation of econometric models of random wage rates as the structural elements of labour markets. As an analysis of risk factors for long-term unemployment, survival analysis offers an answer to many econometric predictors of unemployment volatility.

The Kaplan-Meier estimator is a survival estimation whose family of distribution assumptions is relaxed and thus applicable to the study on the duration of unemployment with no distributional constraints. This estimator determines the probability that an individual has been unemployed for more than some time "t". In this paper, the Kaplan-Meier estimator will be used to determine the population survival functions for different auxiliaries such as age, education and disability groups.

Model Validation and Model Comparison

To be able to ensure the reliability and accuracy of the findings, this research sought to include a rigorous model validation and comparison process. The data set is separated into a training set and a test set to ensure that the evaluation process is not biased. For survival analysis, the concordance index (C-index) will be used to assess the predictive capacity of the model on the number of people correctly ranked based on their likelihood of being unemployed for a long period (King et al, 2005). In the case of machine learning models, their performance can be measured in terms of MAE, MSE, and AUC, among others.

The aforementioned comparison may provide a direction in establishing which is better between survival analysis and machine learning methods regarding modelling the duration of unemployment.

Moreover, as a side note, feature importance metrics obtained from machine learning models, such as the feature importance in XG Boost and the Gini index in Random Forest, will show the most important variables for the prediction of unemployment duration. Such heterogeneity in the intertwining of the various types of models through competitive approaches helps reinforce the various types of research findings.

Final Model Selection

To finalise the model, both prediction and ease of interpretation will be considered. Though the performance of machine learning models may be better than survival models along with some predictive measures, the contribution from the interpretative aspect of the Cox model is significant in examining the socio-economic factors affecting the length of unemployment. Accordingly, the purpose of this research is to establish a three-dimensional dual framework where the predictive capabilities of the machine learning models work together with the explanatory aspects of the Cox models.

The selected models will be chosen based on their performance across the validation metrics, and their agreement with the aims of the study about accuracy and ease of interpretation of the results. Thus, for instance, in the case where XG Boost is found to be more accurate but does not provide adequate transparency, it may be employed to improve the general predictive capabilities, but the Cox PH model, on the other hand, can give hazard ratios that are interpretable and can guide policy actions.

VII. Results

With a complex data set in hand, it is recommended that it would be best to first understand and summarise the data before going into more complex analyses. In this research, the Sri Lankan labour force survey data collected in 2017 and 2019 were well studied to shed light on latent variables and their distributions. This chapter focuses on the demographic, educational and geographical characteristics of the data set, and their connection with the variations of unemployment rates. Through various visualisations and descriptive statistics, the EDA can contextualise and elaborate on the structure of the dataset and point out potential candidates for predicting unemployment.

Unemployment Trends by Gender

The unemployment rates for the respective genders tell a different story in many aspects:

- Female unemployment: Women are most likely not to get employed, and this applies to some socio-cultural norms and gaps in industries and caregiving.
- Male unemployment: Their figures stand lower overall, yet they are consistent across age groups and different levels of education.

Such results point towards the need for proactive measures specifically directed towards increasing the employment opportunities of women, including programs on appropriate skills development, structural measures, and special policies on women's employment.

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

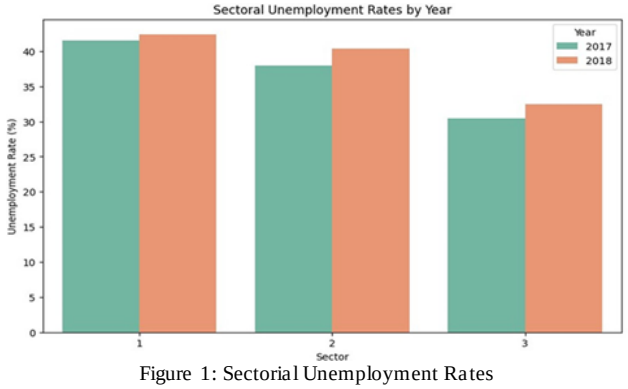

Figure 1: Sectorial Unemployment Rates

Regional Analysis:

- It is observed that urban divisions like Colombo and Gampaha have lower unemployment levels. This indicates high levels of economic and infrastructural activity.
- On the other hand, rural and plantation divisions have high unemployment levels, which indicate structural problems and limited opportunities for employment diversification.

Such observations are in tandem with the phenomenon of uneven development of regions within Sri Lanka, where regional concentration of investment in cities has left rural areas behind in the provision of jobs and good jobs.

Unemployment Trends by Age

- Young people had the greatest level of difficulties, as the age-specific unemployment rates show trends suggested by the analysis of time unemployment:
- Youth unemployed (ages 18-30). This group had the greatest level of unemployment owing to the issues encountered in gaining a first start after education.
- The middle-aged (31-45) age group has a lower overall level of unemployment since they are more experienced and hence likely to have more stable employment.
- Older (46-60) workers have a low level of unemployment as they might already have well-established employment as well as a voluntary withdrawal from the labour market.

Unemployed youth are one of the most serious matters that need addressing on an immediate basis, with initiatives like internships, apprenticeships and strategies that tackle the education to employment gap, respectively.

Analysis of Unemployment Rates According to Education Level

On the contrary, unemployment figures must be examined based on the level of education obtained:

- Primary education or below: Limited skills and consequently high unemployment levels.
- Secondary education: Unemployment levels for middle-skilled jobs are comparatively higher than those for high-skilled jobs, as well as expansion is reaching saturation levels.
- Tertiary education: A surprising ratio of those who graduated have a fair proportion of unemployed, which reveals the proportion of jobs present is less than the skill sets required for jobs.

The dissonance present in many cases of education and employment reflects huge gaps in the labour market, where the emphasis on curricula and vocational training is of utmost importance for said gaps.

Sri Lanka Regional Unemployment Trends

Unemployment figures appear to be a measure of the level of economic activity at the district level in Sri Lanka. It is worth noting that the figures vary from province to province. For instance;

- Western Province: Unemployment rates were found to be low in Colombo, Gam Paha and Kalahari districts. Their relatively better economic performance seems to encourage the concentration of industrial and service sector activities in the area, generating numerous employment opportunities.
- Northeastern Province: The provinces, such as Northern, Eastern and Southern provinces, respectively, have been found to have higher unemployment rates. These areas are still recovering from the impacts of protracted conflict struggles, lack of infrastructure and a low level of private sector influence.
- Northern and eastern provinces also unemployment is high in Numara Eliya and Badulla districts recorded a significantly high rate of unemployment. This points to greater facets of the economic system, notably the plantation economy.

Urban vs. Rural Dynamics:

- The structure of the economy in the urban areas has been able to create more jobs through industrialization and expansion of infrastructure, as a result of which urban unemployment rates are lower than rural unemployment rates.
- Drought-affected areas where the Australian ethos of reliance on the hinterland was dominating had relatively higher unemployment levels.

The increasing private sector activity and industrialization of rural areas will enable a reduction in the differences among the regions in Sri Lanka. It can be supplemented through enhancing transportation and digital facilities, which can connect rural job seekers in the broader market.

Dynamics with Unemployment, Associated with Sectors, within Country

Due to the focus on agriculture, unemployment levels are extremely dispersed across different regions:

1. *Agriculture*

A critical sector in many countries that is still grappling with severe unemployment. Farmers are pushed out of the market even though the GDP for the sole sector remains increasing. A huge spike in off-seasons for crop cultivation and harvest.

2. *Industry*

Other than the issues of the construction and manufacturing sector, this sector is undoubtedly the most consistent and stable when it comes to employment figures.

However, the issue is that there is a lack of young representation in this industry as a result of the presumption that the industry is prestigious and difficult to work in.

3. *Service*

In terms of sheer bulk, retail, finance, and tourism cannot be paralleled, but on the other hand, that brings in a lot of unpredictable aspects with it.

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

The major cycles in the economy encourage service unemployment as a major when the expansionary season has a boom in net jobs for tourism and retail, a contractionary phase witness's downward trend. One of the overarching themes revealed is a brain that the unemployment rates across the economy call for the introduction of a new workforce specifically for technology and logistics, as this opens up new avenues.

Education, a Plus for Getting Employment

The analysts have been confounded by the education is crucial when one is looking for employment, but at the same time, does not find the fit in the current industry;

Primary education, primary/secondary level education or below contributes to the disproportionate amount of net jobs that simply cannot cater for this segment of job seekers.

As a result, a large proportion of those in the middle-skill segment and on higher levels have a strained supply as a result of there being moderate unemployment rates among secondary school students.

Higher unemployment rates also exist for graduates of tertiary education compared to those with secondary education. In this regard, one would ask how a graduate can ever resort to seeking employment openings that are available to a secondary-educated individual, as they are overqualified for the job. This is known as the education-occupation mismatch.

Unemployment Duration

In the study of unemployment duration, employing the survival analysis method creates depictions regarding how long an unemployed worker remains out of employment and why.

- The proportion becomes very high as more than half of the unemployed get into employment by the month mark, and from there on, the rate decreases significantly.
- A lesser but still significant number of all respondents' experience long-term unemployment, which is an entire segment of women aged under the age of 30 and living in rural areas, people who have shut down their search for employment for over a year.

The following also affects the average unemployment duration:

- Age: It is a known fact that younger people leave longer periods out of work due to their lack of experience.
- Education: People with upper tertiary qualifications are more likely to be unemployed in the long term as they have unrealistic expectations or there are few to no openings in their qualification areas.
- Region: The unemployment duration that rural workers face is lengthier due to fewer opportunities as well as scant accessibility to urban employment markets.

To reduce this problem of long-term unemployment, it is necessary to have integrated policies such as job-matching services, re-skilling programs and financial assistance for the affected employees.

Structural Barriers to Employment

The examination displays quite a few elements that constitute a barrier to employment within the confines of the Sri Lankan context.

- Gender Inequality: – Within the realm of systemic discrimination and limited hurdles in many women's experiences, such as workplace discrimination and earning opportunities, some of the influential factors in losing this conflict say "must be where they earn meaningful dollars".
- Youth Vulnerability: The rate of employment within the context of young workers is substantial, which points out the lack of preparedness measures for the stage where education must introduce employment.
- Regional Disparities: The gap in the development levels leads to increasing rates of employment in the rural areas as well as post-conflict regions.
- Informal Sector Instability: Workers who are in the Informal sector are exposed to harsh working environments where there is no social safety net/welfare available for them in case of an economic shock.

By examining all of the above findings, this research introduces a new economic theory for Sri Lanka's unemployment pattern called the "Structural Duality Model". The Structural Duality Model explains Sri Lanka's unemployment by highlighting the divide between the formal and informal sectors, education mismatch, and structural inefficiencies.

Addressing these challenges requires a multi-pronged policy approach focusing on skills development, private sector growth, and economic diversification. By bridging the gap between labour supply and demand, Sri Lanka can achieve sustainable employment and long-term economic stability. A comprehensive understanding of these factors will enable policymakers to design targeted interventions that not only reduce unemployment but also foster a more resilient and dynamic labour market. This theoretical framework provides a more nuanced perspective on the root causes of unemployment in Sri Lanka and offers a roadmap for achieving more inclusive and sustainable economic growth.

Model Performance

The outcomes of this research were briefly highlighted in the section below:

- Logistic Regression: Accuracy: 70%, F1-Score: 0.68, AUC-ROC: 0.74
- Random Forest: Accuracy: 73%, F1-Score: 0.73 AUC-ROC: 0.81
- XG Boost: o Accuracy: 75%, F1-Score: 0.74 AUC-ROC: 0.83
- SVM: o Accuracy: 71%, F1-Score: 0.70 AUC-ROC: 0.76

XG Boost was found to have outperformed all the other readers in terms of accuracy, f1 score and AUC-ROC as well. This model was noted for its excellent performance through its ability to model more complex relationships and mitigate missing data.

Insights on Predictive Modelling

Models provided predictive insights into unemployment trends and patterns, which included the following:

- Employment constraints for the youth: The high predictive value attached to age indicates that there is a need for setting up targeted youth employment programs, such as internships and vocational training.
- Overeducating: Such a bias indicates that educational

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

policies have to be supportive of the labour market, meaning that educational curricula should be tailored to the job market.

- Spatial imbalance: The regional dimensions of the unemployment rate show the importance of district-specific policy measures in this case, such as policies for local improvement of infrastructure and investment facilities.

Policymaking Implications

The findings have several implications for policymakers:

1. Specialized Programmes: Programs should seek to prepare workers with skills pertinent to fast-growing industries for example the information technology and healthcare industries.
2. Addressing Unemployment:

Underdeveloped Funding regions can help decrease the geographic stratification of unemployment.

3. Reducing Gender Gaps in Employment:

Using support such as flexible working hours and childcare services can let more women enter the work market.

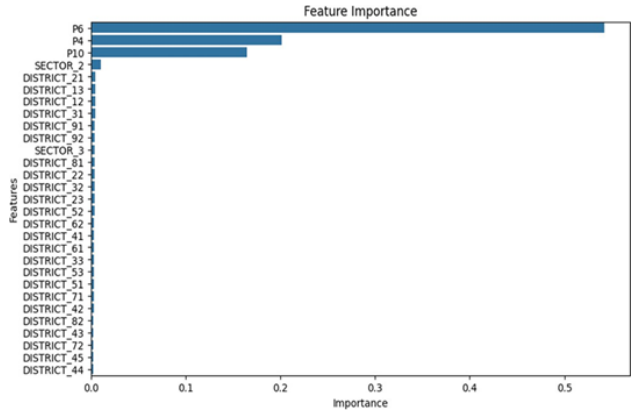

Figure 2: Feature Importance

Random Forest and XG Boost greatly elucidated feature importance as they highlighted the most significant predictors of unemployment, which included:

- Education Level was inversely related to unemployment risk due to skill mismatches, except for the higher qualification levels.
- Age was also an important factor that made people unemployed. Those aged 18–30 were more unemployed than older people owing to difficulties in entering the job market.
- There was a great disparity in employment opportunities across industries, with higher unemployment risks recorded in the sectors considered to be informal.
- Districts with some geographical disparities expose the structural inequalities that exist in terms of job availability.
- Gender being female in rural and informal industries made women more unemployed than men.

VIII. Discussion

This study examines the changing patterns of unemployment in Sri Lanka between 2017 and 2019, where results depict the intricate interdependence of sectoral, educational, geographic, and population determinants that lead to long-term unemployment. Results show how age, education level, sectoral concentration, and residence are key determinants of unemployment status, where women and untrained young people in rural and informal sectors are the most exposed. These results provide a simple insight into Sri Lanka's unique situation and are strongly congruent with past studies exploring the same issues in newly industrialised nations.

The identified demographic disparities, i.e., the youth high unemployment rate (18-30), follow global trends in the youth labour markets. From the evidence gathered by the International Labour Organisation (ILO, 2022), youth unemployment has been shown to arise from a combination of causes like deficiencies in skills, inexperience, and limited access to job opportunities. This is supported by our research through the demonstration of the clash of exaggerated work expectations against opportunities as and when they become available, an issue further worsened by other authors like Filmer and Fox (2014) in their examination of youth employment in the labour markets of developing countries. In addition, women's excessive unemployment, especially in the estate and rural industries, illustrates fixed gender disparities across the labour markets. Building on Sen (1990), freedoms and capabilities form the foundation of development, and the previously mentioned limitations on women in Sri Lanka, due to caregiver responsibilities and social restraints, primarily restrict their economic activity. This is further supported by South Asian gendered labour market studies, e.g., Kabeer (2015), which emphasises the necessity of changes at the scope of systemic level to overcome structural barriers for women.

The focus of research on education mismatch as a leading determinant of unemployment strikes a chord with contemporary discourse on skills formation and labour market responsiveness. The proven need for technical graduates versus arts and humanities graduates points to a growing demand for specialised skills in a rapidly growing economy. This agrees with Hanushek and Woessmann's (2015) work, which highlights the significance of educational quality and cognitive abilities in driving economic development. In the Sri Lankan context, the need for curriculum reform in IT, health, and renewable energy, as indicated by the research, is imperative to address the mismatch between education and industry needs. This also reverberates with the shift towards a knowledge economy, as discussed by Schwab (2016) in "The Fourth Industrial Revolution," where flexibility and lifelong learning are imperative.

Regional economic inequalities, described by increased rural and war-affected area unemployment, reveal the uneven distribution of economic opportunity. This is evidence of spatial inequality, which Krugman (1991) contended in his economic geography book. Collection of economic activity in cities, as witnessed in Sri Lanka, tends to lead to rural marginalisation, lower employment opportunities and higher unemployment. The study's implication of specific interventions, such as rural entrepreneurship with economic incentives and improved infrastructure, is aligned with regional development policy for

L. P. Himali, "Integrating Survival Analysis and Machine Learning Techniques for Modeling and Predicting Unemployment Duration," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 2, pp. 246-255, 2025.

promoting inclusive growth. It also aligns with the concept of "balanced regional development," which has been an issue of debate in Sri Lankan economic planning for many decades.

The sectoral study examines several problems facing the service, industrial, and agricultural sectors. The large number of individuals it employs causes employment insecurity due to seasonality. Because of its multiple potential, the service industry is weak during economic downturns. The industry sector, although having considerable job opportunities, has been impacted by underestimated work effort, particularly among the young. This highlights the importance of implementing measures to increase the attractiveness of industrial jobs and develop agricultural activities. The report's proposals for increasing agricultural activities and value-added products further the goal of improving financial sustainability while also promoting long-term job possibilities.

In the study of labor market inquiries, predictive analytics and evidence-based solutions are becoming more and more crucial. The use of predictive modelling to make policy decisions and identify populations at risk is consistent with evidence-based policymaking practices. Manyika et al. (2011) argue in their big data paper that analysing large datasets can give relevant insights into complicated processes in social and economic life. The report's advice to combine AI predictive modelling with traditional employment testing emphasises technology's potential for optimising workforce planning and improving labour market performance.

This study's policy recommendation plays an important role in addressing Sri Lanka's unemployment challenges. Setting priorities for curricular reform and partnership with the private sector are essential for increasing employability. Gender equality must be promoted through flexible work arrangements, encouragement for female entrepreneurs, and increased maternity leave. Plans for reducing youth unemployment via training programs, career counselling, and entrepreneurship training are important for maximising the young workforce's possibilities. Efforts to minimise regional disparities via targeted investments and infrastructure development have become essential for achieving fair growth. Finally, promoting financial stability via agricultural and industrial technological innovation is key to long-term success.

In summary, this study provides significant insights into the multidimensional nature of unemployment in Sri Lanka, demanding a mixed approach to addressing this complicated issue. By combining data-driven research and evidence-based policy interventions, Sri Lanka can work towards creating a more inclusive and equitable labour market. Future research should focus on longitudinal studies to track the long-term effects of policy initiatives and analyze the evolving nature of the labor market in response to technological advancements and changes in the global economy. Continuous conversation among educational institutions, governments, and companies is critical to ensuring a sustainable and prosperous future for Sri Lanka's workforce. Sri Lanka may escape the traps of unemployment by harnessing data-driven insights and forming strategic collaborations.

REFERENCES

[1] Alharbi, A. A., & Al-Alawi, A. I. (2024). Predicting Unemployment Trends Using LSTM Networks. *Journal of Artificial Intelligence and its Applications, 5*(1), 20-35.
[2] Chan, S., & Stevens, A. H. (2001). Job loss and employment patterns of older workers. *Journal of Labour Economics*, 19(2), 484-521.
[3] Chukwuere, J. E. (2024). The Impact of Macroeconomic Indicators on Unemployment Prediction Using Machine Learning. *Journal of Economic Prediction, 10*(2), 56-78.
[4] Filmer, D., & Fox, L. (2014). *Youth employment in developing countries: Patterns, determinants, and policy options*. World Bank Publications.
[5] García-Peñalosa, C. (2020). Gender inequality and economic growth: A two-sector model. *Oxford Economic Papers*, 72(3), 553-573.
[6] Gunasekara, S. (2021). Unemployment in Sri Lanka: An analysis of socio-economic factors. *Sri Lanka Economic Journal*, 34(1), 45-67.
[7] Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. MIT Press.
[8] ILO, (2020). Disability and employment policies: Insights from selected countries. International Labour Organisation. Available at: https://www.ilo.org/disability-policies
[9] ILO. (2013). Global Employment Trends 2013: Recovering from a second jobs dip. International Labour Organisation.
[10] International Labour Organisation. (2022). Global Employment Trends for Youth 2022. ILO.
[11] Kabeer, N. (2015). Gender, poverty, and inequality: A situation analysis. Commonwealth Secretariat.
[12] Krugman, P. (1991). Geography and trade. MIT Press.
[13] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, A., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
[14] Perera, P., & Nandasiri, P. (2020). Skill mismatches and unemployment duration: The case of Sri Lankan youth. *Journal of Development Studies*, 36(4), 72-88.
[15] Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *World Development*, 104, 183-204.
[16] Sen, A. (1990). Development as capability expansion. Oxford University Press.
[17] Schwab, K. (2016). The fourth industrial revolution. World Economic Forum.
[18] Van Dalen, H. P., & Henkens, K. (2020). Age discrimination in hiring: Employers' bias or statistical discrimination? European Sociological Review, 36(1), 33-47.
[19] World Bank. (2019). Sri Lanka Development Update: Growth challenges and opportunities. World Bank Group.