5-2024

A Study of the Impact Education Level Has on Unemployment Rates for Youths in Emerging Market Countries

Alia AlDhaheri
aa3207@rit.edu

A Study of the Impact Education Level Has on Unemployment Rates for Youths in Emerging Market Countries

by

Alia AlDhaheri

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

May 2024

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Thesis Approval

Student Name: Alia AlDhaheri

Graduate Capstone Title**:** A Study of the Impact Education Level Has on Unemployment Rates for Youths in Emerging Market Countries

Graduate Thesis Committee:

Name: Dr. Sanjay Modak **Date:**

Chair of the Committee

Name: Dr. Khalil Al Hussaeni **Date:**

Member of Committee

Acknowledgments

I am deeply grateful to the professors of the Master of Science in Professional Studies: Data Analytics program for their tremendous support, guidance, and dedication throughout my academic journey. Their expertise, encouragement, and willingness to constantly accommodate our needs have shaped my academic and personal growth. Coming from a finance background, I am proud to have persevered through challenges, pushed my boundaries, and reached this significant milestone in my academic journey. Through my professors' guidance, I have gained the knowledge, skills, and confidence to tackle the challenges that came with entering the Data Analytics field.

Abstract

Economic stress endured by countries globally can be due to numerous socioeconomic factors, including unemployment, measured by the unemployment rate expressed in percentage (%). Various micro and macro socioeconomic variables, including the quality of education and vocational training in the countries and the availability of it, cause high unemployment rates. This study assesses the impact education has on unemployment rates and focuses on youths in emerging market countries to determine the nature of the relationship between the two variables. In addition, the study aims to predict which continents will have increasing unemployment rates for youths in the next five years across the Americas (including South and North American countries), Africa, Asia, and Europe. CRISP-DM framework will be followed by a methodology that includes data mining techniques such as correlation analysis, regression analysis, decision tree, k-means cluster analysis, and time series forecasting analysis, all enabled by R studio, and using Tableau for geographic visualizations. Using the data sourced from theglobaleconomy.com, the data mining techniques were applied to achieve the study goals. The findings concluded that the relationship between education level and unemployment rate for youths in emerging market countries is inverse. Generally, higher education levels, especially secondary education, led to lower unemployment rates due to better employment prospects. However, due to the considerable variation in the results of the findings, the relationship could be more significant, meaning education level does not significantly impact unemployment rates, and many other factors come into play.

Keywords: Unemployment, Education, Youth, Emerging Markets, Socioeconomics.

Table of Contents

ACKNOWLEDGMENTS II

ABSTRACT III

LIST OF FIGURES V

LIST OF TABLES VI

CHAPTER 1 7

INTRODUCTION 7

1.1 BACKGROUND AND STATEMENT OF PROBLEM 7

1.2 PROJECT GOALS 9

1.3 AIMS AND OBJECTIVES 10

1.4 RESEARCH METHODOLOGY 11

1.5 LIMITATIONS OF THE STUDY 12

CHAPTER 2 13

LITERATURE REVIEW 13

2.1 LITERATURE REVIEW 13

2.2 LITERATURE REVIEW KEY TAKEAWAYS 23

CHAPTER 3 24

PROJECT DESCRIPTION 24

3.1 PROJECT DESCRIPTION BRIEF 24

3.2 DATASET 25

3.3 DATA DICTIONARY 26

CHAPTER 4- ANALYSIS 29

4.1 DATA PRE-PROCESSING 29

4.2 CORRELATION ANALYSIS 30

4.3 REGRESSION ANALYSIS 31

4.4 CLASSIFICATION DECISION TREE 32

4.5 CLUSTER ANALYSIS 34

4.6 FORECASTING 40

4.7 COMPARATIVE ANALYSIS OF STUDY FINDINGS 46

CHAPTER 5- CONCLUSION 47

5.1 CONCLUSION 47

5.2 RECOMMENDATIONS AND FUTURE WORK 50

REFERENCES 51

List of Figures

Figure 1 CRISP DM Framework (Source IBM) 11
Figure 2 Unemployment rate of Chinese urban residents based on educational level (2003) (%) 15
Figure 3 Box plot of unemployment rate by education level (1995-2012) (%) 15
Figure 4 Individual Classification Tree. 22
Figure 5 Correlation Analysis – Correlation Heatmap 30
Figure 6 Regression Analysis Results 31
Figure 7 Classification Decision Tree 32
Figure 8 RMSE of Classification Decision Tree Model 33
Figure 9 K-Means Cluster Analysis for Primary Education Enrollment 34
Figure 10 K-Means Cluster List for Primary Education Enrollment 34
Figure 11 K-Means Cluster Analysis for Secondary Education Enrollment 35
Figure 12 K-Means Cluster List for Secondary Education Enrollment 35
Figure 13 K-Means Cluster Analysis for Tertiary Education Enrollment 37
Figure 14 K-Means Cluster List for Tertiary Education Enrollment 37
Figure 15 Average Youth Unemployment Rate 1990 – 2022 41
Figure 16 Highest Average Youth Unemployment Rate 1990 – 2022 42
Figure 17 ARIMA Model Point Forecasts 43
Figure 18 Forecasted Youth Unemployment Rates by Continent (%) 44
Figure 19 ARIMA Model Evaluation 45

List of Tables

Table 1 Average unemployment rate by education level (%)..14
Table 2 Theglobaleconomy Education and Employment Dataset - Data Dictionary..........................26

Chapter 1
Introduction

1.1 Background and Statement of Problem

Unemployment is when a person, usually aged 15 to 64 (ages might vary slightly between countries), actively wants to find a job but cannot. This excludes those not currently working due to retirement, pursuing higher education, or unable to do so due to a disability. Unemployment is a crucial indicator of a country's economic health since it represents the working population's ability to secure employment and contribute to the country's economic productive production. An extremely high unemployment rate implies lower total economic output. In contrast, an extremely low unemployment rate might signal an overheating of the economy where it is likely to produce near total capacity, leading to inflationary pressures. Unemployment is classified into various categories; voluntary and involuntary unemployment are the two broadest categories. When unemployment is voluntary, it indicates that the person left their job voluntarily in pursuit of new employment opportunities. When it's involuntary, it suggests that the person was fired and needs to find new employment (Hayes, 2023). The unemployment rate is the most used metric of unemployment. It is computed by dividing the number of unemployed people by the total number of people in the labor force (typically those aged 15 – 64 years old) and multiplied by 100% to be expressed in percentage (%) terms.

$$unemployment\ rate(\%) = \frac{Number\ of\ unemployed\ people}{Labor\ force}\ x\ 100\%$$

The impact of unemployment is closely felt by the population and the national economy, causing families to struggle with little to no household income, which increases social hardships and causes less consumer spending, leading to economic recessions and setbacks (Corporate Finance Institute, n.d.). With the rapid advancements in technology, education, and research and the disappearance of many jobs that were once crucial to a functioning society and thriving economy, unemployment rates have fluctuated significantly, with an average global unemployment rate of 6.12% from 2003 to 2022 (statista, 2023). And a total number of 205.25 million unemployed people worldwide in 2023 (statista, 2023). That said, unemployment is a definite issue caused by numerous factors regardless of whether the global economy is enduring a crisis.

These factors could include economic stress and poverty, political corruption and war, restrictions on specific job fields, and a lack of education and training. The latter point motivates this study.

Many emerging market countries, especially Middle Eastern countries such as the United Arab Emirates, The Kingdom of Saudi Arabia, and Qatar, have invested billions into their public education systems to increase the education standards among their citizens and enhance their employment prospects. They have developed numerous educative and training programs to improve the employment of their citizens further and manage unemployment rates. The initiatives undertaken by emerging market countries usually target youths because they represent the country's future working population. Nowadays, securing employment is becoming extremely challenging, particularly for youth entrants, which impacts them financially, psychologically, socially, and health-wise (Parliament of Australia, n.d.). Therefore, studying the possible impact of education on unemployment rates for youths in emerging market countries is crucial.

1.2 Project Goals

This project aims to study the relationship between education level and unemployment rate to explore and suggest possible solutions to address unemployment issues among youths in emerging market countries and identify areas for further study. The goal is to assess the degree of impact education has on youth unemployment rates in emerging market countries and whether attaining higher education degrees decreases unemployment rates for youths in emerging market countries. Finally, identify which of the following continents, Americas (including South and North American countries), Europe, Asia, and Africa, are predicted to have increased youth unemployment rates in the next five years to analyze possible socio-economic causes and suggest appropriate solutions.

The results of this study are essential in this era of rapid advancements in technology and artificial intelligence. Successful publication of the project's conclusion will enable countries to reassess their public education systems to support economic decisions that will increase employment opportunities for upcoming generations.

1.3 Aims and Objectives

This study aims to generate valuable insights to help policymakers make well-informed decisions when developing and implementing employment-support programs and education and vocational training budgets for youths representing future generations. The findings of this study will provide a strong reference point to recommend appropriate solutions and identify areas requiring further analysis. The outcomes of this study will guide emerging market countries' governments while establishing policies relevant to youth employment, educational support, and vocational training programs and in managing and stabilizing unemployment rates for the upcoming young generations.

1.4 Research Methodology

The methodology followed in this thesis includes the CRISP-DM Framework, which is detailed below:

Figure 1 CRISP DM Framework (Source IBM)

Cross-Industry Standard Process for Data Mining (CRISP-DM) is a structured framework to process data mining efforts in the industry. It consists of six crucial steps guiding analysts; the following six steps are followed in this thesis:

1. Business understanding: Defining the study's problem statement, objectives, and aims.
2. Data understanding: Understanding that socioeconomic data is most appropriate for this study, understanding that this data is sourced primarily from annual population surveys of open-sourced nations that could come in multiple smaller data sets.
3. Data preparation: Many nations need to produce the socioeconomic data needed for this study regularly, so most datasets available need to be completed and have missing values. Also, irrelevant attributes should be removed to make the dataset more manageable.
4. Modeling: Use of R studio to perform statistical summary analysis to study the data closely, correlation analysis, regression analysis, decision tree, k-means cluster analysis, and time series forecasting analysis. Use of Tableau to provide geographical visualizations.
5. Evaluation: evaluating and analyzing the models' results to draw valuable conclusions and achieve the study's objective.
6. Deployment: publishing the results and analysis of findings in this thesis document.

1.5 Limitations of the Study

This study aims to study the relationship between two key variables: education level and unemployment rate. Like any socioeconomic study, we encounter the challenge of various external social, microeconomic, and macroeconomic influences, which undoubtedly impact outcomes, even if they are not directly accounted for in the study. However, considering every single factor poses a considerable challenge due to the increasing complexity of the models and lack of focus, thereby limiting the comprehensiveness of our findings. Consequently, not all factors affecting our study are fully considered, constraining the depth of our results.

Another area for improvement is data completeness and quality. While searching for suitable datasets for this study, a significant obstacle arose in finding one that included all the key variables essential for the analysis. Moreover, those datasets that included these variables often needed to be completed, with significant gaps in periodic data for most countries in consideration. This issue stems from the fact that the key variables crucial for this study are typically sourced from population surveys, which tend to be outdated, irregularly updated, and presented incompletely. Additionally, these surveys generate hundreds of datasets containing numerous population-related information, further complicating the selection of the most suitable ones for our study.

Additionally, the generalizability of the analysis's findings may be limited due to factors such as the quality of education and vocational training that are not considered in this study.

Chapter 2
Literature Review

2.1 Literature Review

Nickell (1979) utilizes data relevant to the United Kingdom to establish the relationship between education and unemployment duration. The study concludes that attaining higher qualifications decreases unemployment duration by 12%, suggesting an inverse relationship between education and unemployment. However, since the focus of the study is the unemployment "duration," we cannot generalize that education level and unemployment "rate" have an inverse relationship.

Mincer (1991) believes that workers who are educated hold three primary advantages over workers who are less educated in the labor market, which are; 1) higher wages, 2) a higher chance of getting promoted in occupation and income, and 3) employment stability. This implies that a relationship does exist between education and unemployment since educated employees have higher employment stability, which means a lower chance of unemployment.

Riddell et al. (2011) state that other unobserved factors associated with both variables may produce positive relationships between education and re-employment rates of unemployed workers. The study examines the effects of education on the probability of becoming unemployed and the likelihood of re-employment at the survey date conditional on being unemployed in the previous period using longitudinal data from the 1980 to 2005 Current Population Survey (CPS) and the 1980 Census. According to the findings, education considerably enhances re-employment success among the unemployed. Evidence has also been found that higher postsecondary education reduces unemployment but does not diminish the likelihood of unemployment.

A study made by Puspadjuita (2018), applies descriptive analysis techniques and a multiple linear regression model to the 2000 population census labor force (BPS) in Indonesia data to analyze the relationship between the unemployment rate (dependent variable (Y)) and independent variables like; senior high school workforce education through the multiple linear regression model applied in this study. The linear regression model results proved that a positive relationship exists between the unemployment rate and high school workforce education; a regression coefficient value of 0.08097 indicates that for every 1% additional high school workforce, the unemployment rate will increase by 0.08%. The Indonesian population is considered high, and every year, more students are completing their formal high school education; if graduates do not find suitable jobs post-graduation, the unemployment rate will increase.

The study conducted by Couppie et al. (2000) includes representations of averages calculated based on Eurostat's Labor Force Surveys (LFS) for the periods of 1993 to 1997 regarding unemployment rates (%), diploma levels and vulnerability to unemployment (%), etc. in selected European countries; Denmark, United Kingdom, France, and Italy. Survey results provide the average unemployment rate (%) divided by each diploma level category: higher-education graduates, secondary school graduates, and no diploma for juniors and seniors across the four selected countries. We get the following results when computing the average unemployment rate (%) across the four selected countries.

Table 1 Average unemployment rate by education level (%)

Juniors (people who concluded their studies within the past 5 years)	Higher-education graduates = 14.75% Secondary school graduates = 21.5% No diploma = 30.75%
Seniors (people who concluded their studies over 5 years ago and are less than 50 years old)	Higher-education graduates = 4% Secondary school graduates = 8% No diploma =12.5%

The results suggest that the unemployment rate decreases as you increase the education level, with higher-education graduates having the lowest unemployment rates in both categories and no diploma holders having the highest unemployment rates in both categories. The source suggests this is due to the difficulties juniors face entering the labor force, such as criteria for specific qualification requirements or employers opting for more experienced (senior) candidates.

Zhongchang et al. (2007) used data from the China Labor Statistical Yearbook 2004 to plot a line graph displaying unemployment rates of different educational levels, from the illiterate to the graduate level. Notably, the graph shows an inverted V shape with a peak unemployment rate of almost 11% for senior high school. The unemployment rate is collectively high for educational levels of junior high school, senior high school, junior college, and undergraduate but falls very closely to 0% for graduate level. The relationship between educational level and the unemployment rate cannot be established from the graph's results.

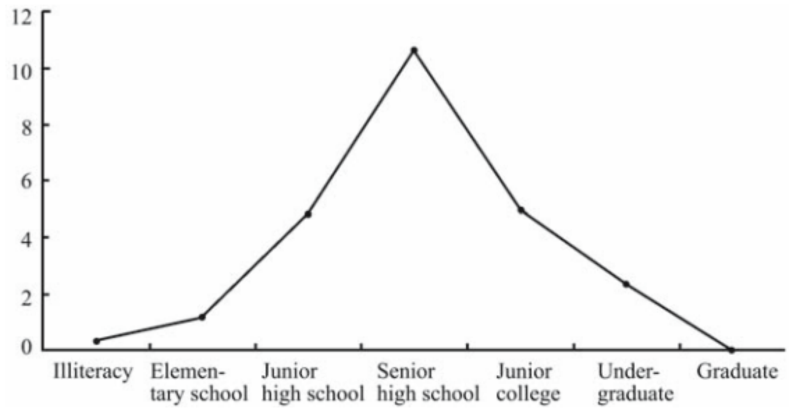

Figure 2 Unemployment rate of Chinese urban residents based on educational level (2003) (%)

Regression modeling is then applied, including independent variables, employment rate of urban residents, educational structure, and educational scale, among others, to determine the relationship between employment and academic level. The result of the model indicates that there is no clear correlation between educational structure and jobs. This is because other factors, such as the labor force market's supply and demand, significantly impact the variables. Therefore, it is difficult to establish the relationship between educational level and employment.

A study that analyzed the shifts in the labor market's educational composition in South Africa between 1995 and 2012 suggests that a person's education level is closely correlated with the unemployment rate. More specifically, those with higher education levels are in more demand for employment and therefore conclude that the relationship between education and unemployment is inverse; the higher the education level, the lower the unemployment rate will be (Bhorat et al., 2016).

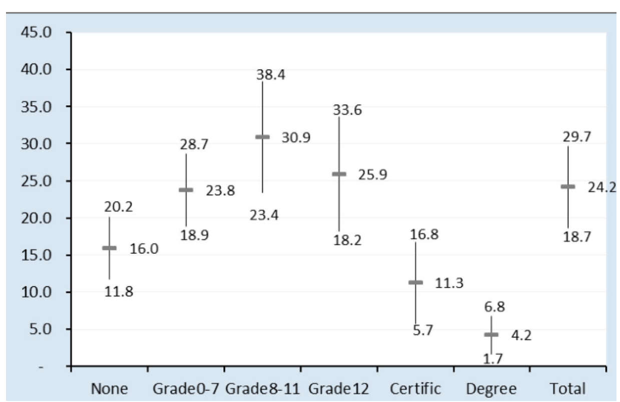

Figure 3 Box plot of unemployment rate by education level (1995-2012) (%)

Bugudui (2016) uses multivariate techniques to analyze unemployment in Romania from 1996 to 2014 for four age groups. Multivariate techniques are statistical methods measuring relationships between 2 (or more) variables, such as principal component analysis (PCA) and cluster analysis used in this study. The use of PCA is justified in the study mainly to reduce dimensionality to contain the focus survey variables since, with more variables, the significance of each variable may be diminished. Therefore, we cannot establish its significance and relationship to the response variable. The study applies cluster analysis to the age groups studied to detect similarities between the different periods analyzed, the level of education, and whether unemployment was short- or long-term according to each age group.

Jantawan et al. (2013) built a classification model that predicts graduate employment status: employed, unemployed, or others. Ten classification models have been applied, including five versions of decision trees and five versions of the Naïve Bayes classifier using the WEKA program. In addition, a classification accuracy table was generated to compare the accuracies of methods used and display statistical learning computations such as mean absolute error (MAE) and root mean squared error (RMSE). The RMSE values of all methods were plotted to allow for the interpretation of results.

A statistical analysis study by a student at the University of Iowa explored numerous factors concerning the unemployment rate, such as CPI, burglary, public debt, trade balance, average hourly salary, etc. The initial hypotheses of each factor were clearly defined at the start, and a criterion for rejecting and accepting each factor was also determined. To select the best fit of the linear regression model, methods like forward and backward selections, r-squared computations, and MSE plots have been applied to each variable. Following the applications of such techniques, the linear regression model is built and used to determine the relationship between unemployment and the independent variables in consideration. Interesting to see that only 1 hypothesis was proven to be true out of the nine independent variable hypotheses and the average gas price variable has proven to have a direct relationship with the unemployment rate even though the initial hypothesis was "unsure" of the type of relationship between average gas price and unemployment (Collins, 2009).

Pompei and Selezneva (2021) investigated whether structured, objective, and formal education significantly affected unemployment and inactivity in the European Union countries between 2006 and 2010. The study asserts that people strive to bolster their education level to attain the qualifications to get their desired jobs. The researchers used a regression approach that analyzed data to determine the association between education level and employment status. The outcome of the analysis showed that additional years of education reduces the risk of being unemployed relative to the probability of being an employee as the co-efficient changed from -0.068 to -0.162. The authors link this outcome to formal education's capacity to motivate students or workers to attain firm-specific knowledge that enables them to thrive in specific sectors. Thus, the study states that education level improves the probability of employment.

Ramzan et al. (2018) investigated the effects of education on some economic variables related to unemployment in Pakistan between 1972 and 2012. They hypothesized that inflation, economic growth, wages, employment, and education were the primary factors that affected unemployment regulation within the timeframe. Ramzan et al. (2018) used a multivariate regression equation to find the nature of the relationship between education and unemployment in Pakistan, an emerging economy. The authors regressed unemployment by wages, inflation, education, GDP, population, and money supply to establish the relationship between the dependent and independent variables. The result was that education negatively correlated to the unemployment rate, but the relationship was insignificant.

Mpendulo and Mang'unyi (2018) explored the education level of the youth and the unemployment rate in four South African municipalities. The researchers used a quantitative approach to conduct the study. Specifically, they utilized a cross-sectional research design to prepare and distribute questionnaires to collect data. In addition to developing various regression models, Mpendulo and Mang'unyi (2018) employed chi-square data, crosstabulation, and correlation analyses to establish the relationship between education and unemployment. Below is the regression model used to confirm the correlation between the study variables:

$Yt = \beta + \beta1\ X1 + \beta2\ X2 + \beta3\ X3 + \beta4\ X4 + u$

The results showed a positive correlation between educational level and unemployment. Specifically, the Spearman rank correlation (r) between the two variables was 265**; sig = 0.003 (Mpendulo and Mang'unyi, 2018). In addition, the R^2 and adjusted R^2 derived from the regression analysis were 0.395 and 0.349, respectively. The outcomes suggest that the people in the region were acquiring more knowledge but still found it hard to get jobs. This means that as the education level increases, youth unemployment also increases.

Cristescu's (2017) goal was to establish the effect of education on unemployment, particularly in Europe. South European countries have been experiencing a declining employment rate over the past decade. Hence, to redress the issue, Cristescu (2017) conducted a study that used panel data to establish the effects of education attainment on the unemployment rate in Southern European countries. Cristescu (2017) used EViews software to execute the regression analysis and confirm the relationship between education and unemployment. The study highlighted that the rising unemployment rate mainly arose from educational deficiencies. The associated coefficient of -0.5447 validates this assertion. In addition to the gross domestic product (GDP) and foreign direct investment (FDI), Cristescu (2017) notes that an increase in the level of public spending on academics, as well as a rise in graduation rate in these nations, would curb unemployment.

Suyanto et al. (2019) wanted to determine how quality education and excellent health impact the unemployment rate in East Java, a province in Indonesia. Suyanto et al. (2019) argue that the increase in East Java's population is among the primary causes of unemployment in the country. Suyanto et al. (2019) used descriptive statistics to analyze the secondary data from East Java BPS (population survey). The analysis results showed that education has a significant positive effect on the unemployment rate in East Java. Concisely, the coefficient increased by 0.343%.

Maneejuk and Yamaka (2021) researched to understand how education impacts economic growth in ASEAN-5 nations. They used multiple educational metrics, such as students' enrolment in higher education and public expenditure on the sector, to assess higher education and its economic impact. They utilized the nonlinear regression model to ascertain the effect of higher education on the countries' economic growth, which includes unemployment. They found that strategic government spending on education positively influences economic growth. Given that employment is associated with good economic growth, it would be viable to deduce that a higher level of education reduces unemployment.

Alawad et al. (2020) experimented to establish the primary determinants of youth unemployment in Jordan, an emerging country. . The authors acknowledged that the rate at which unemployment is increasing in the country and the region is alarming, necessitating decision-makers to develop systems that identify and counteract its causes. Alawad et al. (2020) used a logistic regression model to analyze data derived from the Jordan Labor Market Panel Survey. They found that the increasing level of education determined the employment rate of the residents in Jordan. For example, 25%, 31%, and 42% of the people with primary, secondary, and post-secondary education, respectively, were employed. This outcome shows that the higher the level of education people have, the higher the probability that they will get jobs in Jordan. Overall, the study shows that education bolsters people's marketability, increasing their chances of employment.

Egessa et al. (2021) conducted in-depth research backed by statistical analysis to understand the determinants of youth unemployment in Uganda. The authors acknowledge that unemployment has been increasing steadily in Uganda, posing severe conflicts with the residents' quality of life. Hence, the study aimed to examine the extent to which demographic variables like gender, education, and age determine youth unemployment in the country. Egessa et al. (2021) used a binary logistics regression approach to analyze data from the Uganda Bureau of National Statistics. Below is the regression model to establish the correlation between dependent and independent variables.

$$\text{logit}(p) = \log(p/1-p) = + \beta 0 + \beta 1XI + \beta 2X2 + \beta 3XI3 + \beta 4X4 + \ldots + \beta\ kXk$$

Where:

- P – unemployment
- (p/1-p) – the odd of a young person experiencing unemployment
- X1-Xk – variables of interest.

The research outcome indicated that residence, gender, education, and age directly correlated with unemployment. It was shown that the probability of a young Ugandan being unemployed is 0.154 greater for individuals with a primary education than those without any formal education. In addition, the analysis highlights that the likelihood of a citizen being unemployed is 0.229 higher for people with secondary education than those without classroom experience. Egessa et al. (2021) assert that this discrepancy could be due to poor quality education, skills mismatch, and job availability, among other variables.

Bojadjieva et al. (2022) examine school-to-work transitions and youth employability in Southeastern European countries, emphasizing Bosnia, Herzegovina, Croatia, Montenegro, North Macedonia, Serbia, and Slovenia. Using panel data and regression models over 2009-2019, it assesses the impact of different education levels on youth employability. Results suggest an ambiguous effect of education on youth employment. Challenges persist in youth employability and school-to-work transitions, urging comprehensive educational policies and improved labor market functioning.

Singh et al. (2019) wanted to understand how education determines Kerala's employment rate and social status. They acknowledge that the area continues to have an alarming unemployment rate despite the enhanced education, particularly among the youth. The authors used data from 2004 to 2012 derived from the National Survey Sample Office, an affiliate of the Ministry of Statistics and Program, to determine the correlation between the research variables. They used regression analysis to find the association between Kerala's unemployment rate and education levels. The study showed a negative correlation between education and unemployment for all categories in the rural and urban areas. This outcome suggests that people who pursue higher education are likelier to be employed than those with essential learning. Despite this result, the unemployment rate in Kerala is still high even after the youths enhance their education level. The article notes that the negative output could be due to the inability of the state to create a favorable environment to attract more companies that foster job creation. Hence, in addition to the people getting the required education, the government should establish good corporate or regulatory conditions that would attract local or foreign direct investments (FDIs) to create more job opportunities for the inverse relationship established in the study to be practical.

Oktafianto et al. (2019) wanted to investigate and present the factors that determine unemployment in specific regions in Indonesia. They introduced spatial aspects to objectively identify the determinants of unemployment in the different areas of Indonesia. They used the Spatial Durbin Models (SDM) between

2000 and 2017. The outcome showed that higher education in various regions diminished the unemployment rate significantly.

Examining the period from 1985 to 2006, Hérault et al. (2012) studied macroeconomic conditions, education, and employment outcomes among Australian youth. Notably, gender and education-level disparities emerge, with young men lacking secondary education experiencing heightened unemployment risks during economic downturns. Females with a university degree exhibit a 1.2% increase in full-time work probability for each percentage point rise in GDP growth. Conversely, those with a Year 12 education become less likely to work, emphasizing the complex relationship between economic conditions and educational choices. The research comprehensively explains how macroeconomic factors mold educational and labor market prospects for Australia's youth.

Biagi and Lucifora (2005) investigated the impact of demographic and education changes on European unemployment rates from 1980 to 2000. They determined that shifts in population age structure and educational levels significantly influence unemployment. The findings indicate that structural changes in population age and educational shifts play a crucial role. Youth unemployment rates vary, with Germany, Norway, Portugal, Sweden, and the UK having lower rates, while France and Italy show rates ranging between 22-38%. Demographic changes mainly affect young workers, while educational shifts reduce unemployment among the more educated.

Erdem and Tugcu (2012) explored the relationship between higher education and unemployment in Turkey. The research addresses a gap in existing literature by explicitly investigating the impact of higher education on unemployment in Turkey. The authors emphasize the challenge of aligning the education system with labor market needs in the context of the country's economic transformation since the 1980s. Critical variables in the study include the unemployment rate and the indicator of higher education, which is measured as the total number of individuals completing higher education each year. They concluded that higher education curbed unemployment in the country.

The study by Rahmawati and Putri (2021) explores the impact of education-related variables on the Open Unemployment Rate (TPT) in Banten Province, Indonesia. The authors conducted regression analysis on panel data from 2010-2019. The findings indicate that the literacy rate has no significant effect on TPT, while government expenditure on education, the Gross Enrollment Rate of Senior High School/equivalent, and the proportion of Senior High School workforce and above have significant and adverse effects on TPT. The study emphasizes the crucial role of education in reducing unemployment and highlights the positive impact of government spending on the education sector.

The research by Alcin (2021) highlights the persistent challenge of youth unemployment in Spain and Turkey, emphasizing the role of higher education. Despite expectations, the study finds no causal relationship between higher education enrollment and the youth unemployment rate in both countries. Turkey's Johansen Cointegration Test results indicate no cointegration vectors at the 5% significance level. Therefore, it can be concluded that Turkey's youth unemployment rate and higher education schooling rate series do not have a long-term stable relationship. The Johansen Cointegration Test results for Spain show no cointegration vectors at the 5% significance level. These findings suggest that, similarly to Turkey, there is no long-term stable relationship between youth unemployment and higher education schooling rate in Spain.

Lam et al.'s (2009) study emphasizes the global challenge of youth unemployment, with an International Labor Office report revealing that the youth (15-24) constitute 47% of the world's unemployed population. In South Africa, where unemployment has persisted for decades, the issue worsens with lengthy unemployment durations. The March 2005 Labor Force Survey exposes that 42% of South African youths (15-24 years) leave their studies to enter the labor market, with over 60% having less than a matric qualification. The high unemployment rate prompts questions about why many quit school prematurely. Resource constraints are a likely factor, hindering the pursuit of further studies and delaying entry into the job market. Data from the Cape Area Panel Study reveals the complex dynamics of young people's transitions from school to work, emphasizing the need for careful analyses in understanding the complex relationship between education and unemployment.

The study conducted by Blasques et. al (2021) examines the connection between education participation and unemployment rates nationwide. The study uses cluster analysis to determine how similar the various forms of schooling are to one another in terms of how they relate to macroeconomic cycles. The study starts by looking at a scoring model that filters the unemployment rate's conditional expectation. Then, a multivariate model that regresses the student population on the dynamic macroeconomic factor is considered in the study. The clustered loading matrix is then estimated using the k-means approach. According to the data, students enrolled in part-time programs present a stronger correlation with unemployment than full-time students, suggesting that full-time students have a greater likelihood of finding employment.

Celbiş, M. G. (2022) uses a variety of machine learning (ML) techniques to identify the primary individual-level characteristics associated with unemployment. Tree-based classification models, bootstrap aggregation, random forests, etc., are used to forecast unemployment status. The findings indicate that the labor market disparities brought about by parental education levels, gender, age, and educational attainment can be reduced or maintained by providing access to training programs.

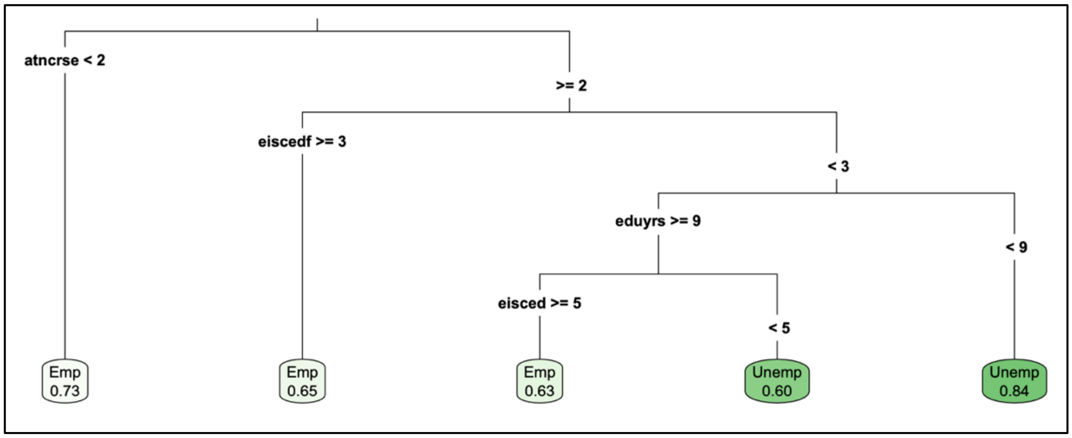

Figure 4 Individual Classification Tree.

It is projected that people in rural areas who have made efforts to advance their knowledge and abilities, by enrolling in classes and training programs (represented by the variable atncrse), will find employment. The findings suggest that participation in these events is not the only way people can acquire the information and skills required for work; a person is more likely to find a job if their father (eiscedf) has at least an "upper secondary" degree. Only persons with at least nine years of education and an advanced vocational degree are anticipated to be employed among those whose fathers have lower levels of education (i.e., eiscedf< 3). Suggesting unemployment can be reduced by attaining general knowledge transferred from others, not only by achieving higher education degrees.

A'rifian et al. (2019) analyzed graduate employment data from the Ministry of Higher Education for 2017 using data mining techniques in a comparative analysis of graduate employment in Malaysia. The study created three predictive models; decision trees, logistic regression, and artificial neural networks, and evaluated each one's ability to forecast the employment sectors for graduates. Compared to the other two models, the study finds that the artificial neural network performs the best at predicting graduates who found employment in the private sector. However, the study is focused on assessing the performance of the models. It does not provide the methodology behind the predictive models and analysis on whether education impacts unemployment rates.

2.2 Literature Review Key Takeaways

- Riddell et al. (2011) suggest that other external variables may play either a positive or negative role in the relationship between education level and unemployment rate and must be taken into consideration when a study on the relationship between the two variables is conducted. Zhongchang et al. (2007) suggest factors like the labor force market's supply and demand significantly impact the variables in consideration. Therefore, it is difficult to establish the relationship between educational level and employment.
- Debatable results among the studies reviewed, such as the study by Puspadjuita (2018), indicate that attaining higher levels of education does not necessarily reduce the unemployment rate, while other studies like the study done by (Bhorat et al., 2016) suggest that the relationship between education and unemployment is an inverse relationship; the higher the education level, the lower the unemployment rate will be.
- Developing nations present an inverse relationship between unemployment and education (Erdem &Tugcu, 2012; Pompei & Selezneva, 2021; Cristescu, 2017). These countries have feasible structures that create diversified markets, presenting job opportunities for their youths.
- According to (Mpendulo & Mang'unyi, 2018; Egessa et al., 2021), there is a positive correlation between unemployment and educational levels in emerging nations. The more youths acquire higher levels of education, the higher the probability of unemployment. This outcome is mainly due to the incapacity of their countries to create adequate job opportunities to meet the massive supply in the labor market.
- Overall, quality education combats unemployment (Pompei & Selezneva, 2021). It creates a skillful workforce, allowing them to get good jobs in the local and international job markets. Most of the studies in the literature review support Pompei and Selezneva's (2021) assertion.

Chapter 3
Project Description

3.1 Project Description Brief

The following three objectives will guide the analysis.

1. Determine the type of relationship between education level and unemployment rate for youths in emerging market countries, whether positive or negative.
2. Assess the strength of the relationship between education level and unemployment rate for youths in emerging market countries to determine whether attaining higher education degrees decreases unemployment rates.
3. Identify which of the following continents: Europe, Asia, Africa, and the Americas (including South and North America) are predicted to have increasing unemployment rates for youths in the next five years.

Initially, it was proposed that the youth unemployment rates for the different continents would be predicted for the next three years; 2023 to 2025, but as the forecasting models were applied, the predictions for three years proved to be inefficient in showing a clear trend to determine whether youth unemployment rates will increase or decrease in the respective continents. Therefore, it was decided that the period of the predictions would increase from three years to five years. Also, the Americas were included in addition to the other continents to ensure regional inclusivity when conducting the forecasting.

3.2 Dataset

The dataset used for this study is sourced from theGlobalEconomy.com, a platform that provides economic and business-related datasets for 200 countries since 2012. The platform is run by a team of qualified economists headed by Ph.D. economist Neven Valev to simplify the data collection process for research and teaching purposes. They collect, collate, and regularly update the data sets provided, sourced directly from publications of official organizations. Each data set downloaded from the platform is unique since it allows users to select what data they want regarding country names, indicators, and periods. The data set initially proposed to be used in this study, sourced from the World Bank data, was proven difficult to work with as the dataset format was challenging. It was extremely time-consuming to preprocess the data set for the analysis. So, a decision was made to dismiss and replace the data set. This data set provides the same data but in a better structure that is easier to work with.

The data set used in this study contains data from 196 countries spanning 32 years from 1990 to 2022. It includes 24 socio-economic attributes that are categorical and numeric, with over 6,600 observations recorded. The following data dictionary details the characteristics of the data set.

3.3 Data Dictionary

This section provides a data dictionary table detailing the attributes in the dataset and their respective types, measures, and descriptions.

Table 2 Theglobaleconomy Education and Employment Dataset - Data Dictionary

No	Attribute	Type / Measure	Description
1	Country	Categorical	Name of country.
2	Code	Categorical	Code of country.
3	ContinentCode	Categorical	Code of the continent the country is located in.
4	Year	Numeric	The year the value of the indicator was recorded in.
5	Unemployment rate	Numeric (in %)	The general rate of unemployment
6	Unemployment rate for females	Numeric (in %)	The rate of unemployment for female population.
7	Unemployment rate for males	Numeric (in %)	The rate of unemployment for male population.
8	Youth unemployment ages 15-24	Numeric (in %)	The rate of unemployment for youth category (ages 15 to 24 years old).
9	Labor force participation rate	Numeric (in %)	Represents the percentage of the total working-age population that is currently employed or actively looking for employment.
10	Female labor force participation rate	Numeric (in %)	Represents the percentage of female working-age population that is currently employed or actively looking for employment.

11	Male labor force participation rate	Numeric (in %)	Represents the percentage of male working-age population that is currently employed or actively looking for employment.
12	Employment in agriculture percent of total employment	Numeric (in %)	Represents the percentage of working-age population that is employed in the agriculture industry.
13	Population size in millions	Numeric	The total size of the population.
14	Poverty percent of population	Numeric (in %)	Represents the percentage of the total population that is extremely poor / in poverty.
15	Public spending on education percent of GDP	Numeric (in %)	Government expenditure on education institutes represented as a percentage of GDP divided by different education levels (primary, secondary and tertiary).
16	Literacy Rate	Numeric (in %)	General rate of literacy which represents the percentage of people with the ability to read and write.
17	Female literacy rate ages 15-24	Numeric (in %)	Rate of literacy for females aged 15 to 24 years old
18	Male literacy rate ages 15-24	Numeric (in %)	Rate of literacy for males aged 15 to 24 years old
19	Youth literacy rate ages 15-24	Numeric (in %)	General rate of literacy for youth category (ages 15 to 24 years old)
20	Primary school completion rate	Numeric (in %)	Percentage of children that completed primary school (based on respective age category).

21	Primary school enrollment percent of all eligible children	Numeric (in %)	Represents the percentage of children enrolled and currently completing primary school from the total of all eligible children (4 to 11 years old).
22	Secondary school enrollment percent of all eligible children	Numeric (in %)	Represents the percentage of youth enrolled and currently completing secondary school from the total of all eligible youth (11 to 17 years old).
23	Tertiary school enrollment percent of all eligible children	Numeric (in %)	Represents the percentage of youth enrolled and currently completing tertiary / higher education from the total of all eligible youth. This includes any education pursued post-high school level such as diplomas, master's and doctoral.
24	Education service price index, world average = 100	Numeric	Represents the changes in the education service prices for multiple education levels during the period.

Chapter 4- Analysis

4.1 Data Pre-processing

The data pre-processing stage aims to generate a new data set that is cleaner and easier to work with throughout the analysis stage.

4.1.1 Column Selection

In this step, the following attributes have been selected and retained in a new clean dataset.

1. "Country"
2. "Year"
3. "Youth unemployment ages 15 – 24"
4. "Female literacy rates ages 15 – 24"
5. "Male literacy rates ages 15 – 24"
6. "Youth literacy rates ages 15 – 24"
7. "Public spending on education percent of GDP"
8. "Primary school enrollment percent of all eligible children"
9. "Secondary school enrollment percent of all eligible children"
10. "Tertiary school enrollment percent of all eligible children"

4.1.2 Addressing Missing Values

In this step, missing values showing as "NA" values were dropped from the data set using the drop_na() function.

4.1.3 Retaining only Emerging Market Countries

This study is focused on emerging market countries, so only the data of emerging market countries must be kept in the final cleaned dataset. Emerged/developed market countries, impoverished countries, and countries with unclear market classifications have been excluded to focus the analysis on strongly performing emerging market countries.

4.1.4 Renaming Columns

Used the colnames() function in R Studio to rename columns into shorter names to facilitate analysis.

4.2 Correlation Analysis

Correlation analysis was conducted to determine the nature and strength of the relationship between education level and unemployment rate for youths in emerging market countries and the remaining variables. A correlation heatmap using the heatmaply() function was produced to visualize the correlation between all variables.

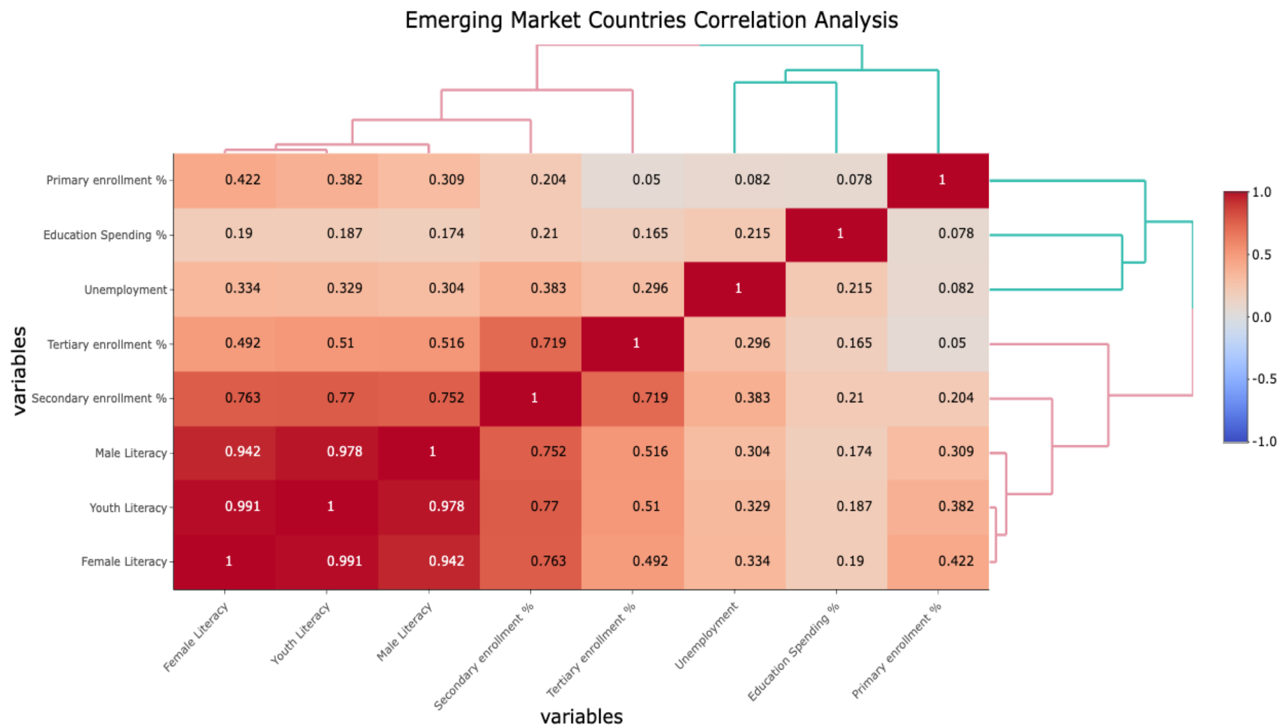

Figure 5 Correlation Analysis – Correlation Heatmap

The variable "Unemployment" represents the unemployment rate, and it is apparent that overall, the unemployment rate positively correlates with variables like youth literacy rate, female and male youth literacy rates, and all levels of education (primary, secondary, and tertiary). This positive correlation is weak due to the low correlation values close to 0. As youth literacy rate and education level increase, the youth unemployment rate is expected to increase as well, but in a weak and unreliable way.

4.3 Regression Analysis

Regression analysis was performed to closely study the relationships between youth unemployment rate (dependent variable) and youth literacy rate, primary education enrollment, secondary education enrollment, and tertiary education enrollment (independent variable).

```
Regression Analysis Results

Call:
lm(formula = Unemployment_Rate ~ Youth_LRate + Primary + Secondary +
    Tertiary, data = emerging_economies_filtered)

Residuals:
    Min      1Q  Median      3Q     Max
-20.506  -8.312  -2.379   6.175  42.353

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.27889   12.30181  -1.405   0.1612
Youth_LRate   0.24410    0.14149   1.725   0.0855 .
Primary       0.06772    0.15331   0.442   0.6590
Secondary     0.03475    0.06620   0.525   0.6000
Tertiary      0.08319    0.03896   2.135   0.0336 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.66 on 293 degrees of freedom
Multiple R-squared:  0.1324,    Adjusted R-squared:  0.1205
F-statistic: 11.18 on 4 and 293 DF,  p-value: 1.882e-08
```

Figure 6 Regression Analysis Results

Youth literacy rate, primary education, secondary education, and tertiary education enrollment all have positive coefficient estimates, suggesting that all independent variables are associated with higher levels of unemployment rate. It is observed that the p-value associated with the t-statistic of the coefficient estimate (Pr(>|t|)) of tertiary education enrollment is 0.0336, which is statistically significant at the 0.05 level. This means that tertiary education, the highest level of education enrollment, has a significant positive relationship with youth unemployment rates. Meanwhile, the other independent variables yielded Pr(>|t|) values much higher than 0.05, indicating that they are not statistically significant. A multiple r-squared value of 0.1324 means that the independent variables explain approximately 13.24 % of the variance in the dependent variable (unemployment rate). In contrast, the remaining variation in the unemployment rate is not explained by the independent variables included in the regression model. F-statistic value of 11.18 with a p-value of about 1.882e-08 indicates that the model as a whole is statistically significant, and there is a positive relationship between the independent variables and the dependent variable.

4.4 Classification Decision Tree

In this section, a classification decision tree was conducted to explore the relationship between secondary and tertiary education levels and unemployment rates to understand how education level impacts employability and deduce whether increasing education level increases or decreases unemployment rates for youths in emerging market countries. The decision tree focused on Secondary and Tertiary education levels only since they are considered higher levels of education and include the study's target group. The data was split into 70% train data and 30% test data, rpart.plot() function was used to visualize the decision tree, and the model has been evaluated by computing its root mean square error (RMSE).

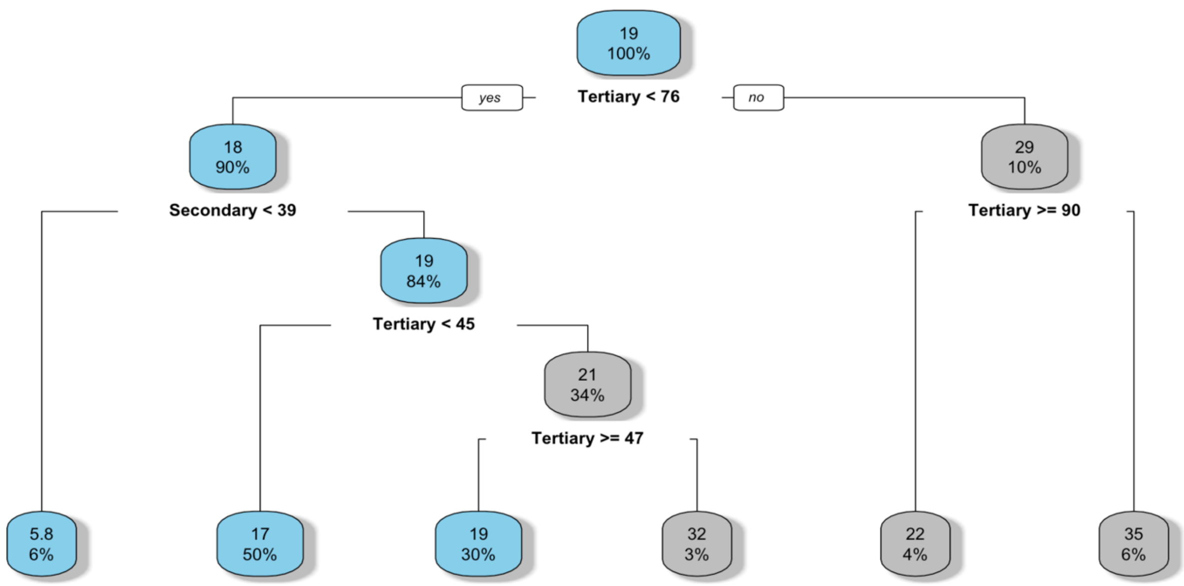

Figure 7 Classification Decision Tree

4.4.1 Classification decision tree results analysis:

Reading the decision tree from top to bottom, starting with tertiary education enrollment, the decision tree results concluded that if the tertiary education enrollment is less than 76% and secondary education enrollment is less than 39%, the youth unemployment rate is extremely high at 90%. If secondary education enrollment is at 39% or more but tertiary education enrollment is less than 45%, the youth unemployment rate remains high at 84%. Tertiary education enrollment of less than 45% reduces the youth's unemployment rate to 50% and is further reduced to 34% for tertiary education enrollment between 45% and 47%. Tertiary education enrollment higher than 47% significantly drops the youth unemployment rate to 3%, indicating strong employment prospects. Finally, the right split of the decision tree shows that for very high tertiary education enrollment of 90% or more, the youth unemployment rates are very low at 10%, 4%, and 6%. They also indicate strong employment prospects with high tertiary education.

In conclusion, the decision tree shows a clear inverse relationship between higher education levels (tertiary education enrollment) and youth unemployment rates. Higher percentages in secondary and tertiary education enrollment are associated with lower youth unemployment rates. Tertiary education enrollment significantly affects employment outcomes, particularly when the enrollment exceeds 45% and youth unemployment rates are at their lowest when tertiary education enrollment is 76% and above. This means that countries with higher enrollment rates in secondary and tertiary education levels have significantly lower unemployment rates for youths than countries with lower enrollment rates in secondary and tertiary education levels, where youth unemployment rates are incredibly high.

4.4.2 Classification decision tree - model evaluation:

The root mean square error (RMSE) is a commonly used metric that evaluates the performance of regression models, including decision trees. It measures the average of errors between the predicted values and the actual values to support assessing the accuracy of predictions made by the decision tree model to assess the reliability and performance of the predictive model.

The RMSE computed was at 11.46, meaning the predicted unemployment rates of the decision tree model were 11.46 units away from the actual unemployment rates. Due to the high variation of unemployment rates in the data set, RMSE of 11.46 is considered low and indicates good predictive performance.

```
[1] "RMSE: 11.4605479092143"
```

Figure 8 RMSE of Classification Decision Tree Model

4.5 Cluster Analysis

In this section, K means clustering is applied to each education level (primary, secondary, and tertiary) to mine continent-based patterns and understand the similarities and differences across different regions worldwide. The Ggplot() function has been used to visualize the clusters, and a cluster list was printed to identify which countries fall into each cluster. The 4 and 6 clusters were seen as appropriate in capturing the respective countries and clearly visualizing the different clusters for each education level.

4.5.1 Primary Education

K-means clustering was applied to primary education enrollment percent with a choice of 4 clusters.

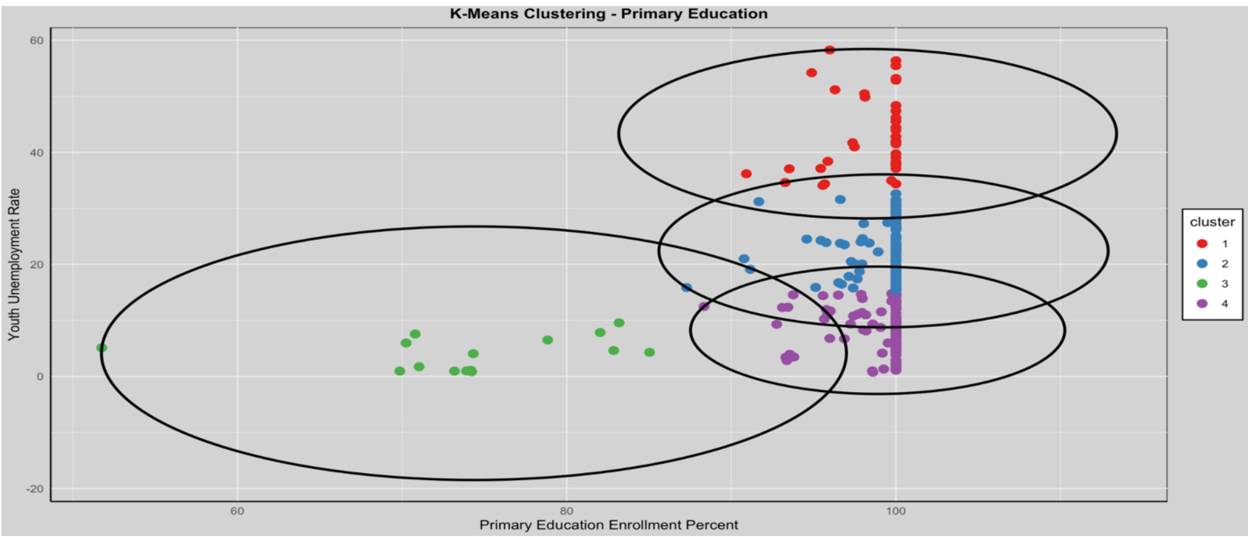

Figure 9 K-Means Cluster Analysis for Primary Education Enrollment

```
Cluster 1 :  Argentina, Bulgaria, Egypt, Georgia, North Macedonia, Palestine, Saint Lucia, Serbia,
South Africa, Spain, Swaziland
Cluster 2 :  Albania, Armenia, Belize, Bhutan, Brazil, Bulgaria, Cape Verde, Chile, Colombia,
Egypt, Georgia, Greece, Guyana, India, Indonesia, Iran, Italy, Kazakhstan, Latvia, Lithuania,
Mauritius, Moldova, Mongolia, Morocco, Palestine, Panama, Paraguay, Romania, Rwanda, Serbia,
Slovenia, Spain, Sri Lanka, Suriname, Tajikistan, Tunisia, Turkey, Ukraine
Cluster 3 :  Ethiopia, Guatemala, Ivory Coast, Pakistan
Cluster 4 :  Afghanistan, Armenia, Azerbaijan, Bangladesh, Belarus, Bhutan, Cambodia, Ecuador,
Ethiopia, Ghana, Guatemala, Indonesia, Ivory Coast, Kazakhstan, Laos, Latvia, Lithuania, Mexico,
Moldova, Mongolia, Oman, Panama, Paraguay, Peru, Philippines, Poland, Qatar, Thailand, United Arab
Emirates, Vietnam, Zimbabwe
```

Figure 10 K-Means Cluster List for Primary Education Enrollment

Cluster Analysis for Primary Education Enrollment:

There is no clear pattern to deduce the relationship between primary education enrollment and youth unemployment rate in the different clusters. Most countries have incredibly high primary education enrollment percentages, which is close to 100%, but there is significant variability in the youth unemployment rates. This is possibly due to primary education being the lowest education level available for individuals. So, most countries have made primary education available for the most part, and this does not significantly impact an individual's employability, and other socioeconomic variables are present.

4.5.2 Secondary Education

K-means clustering was applied to secondary education enrollment with a choice of 6 clusters.

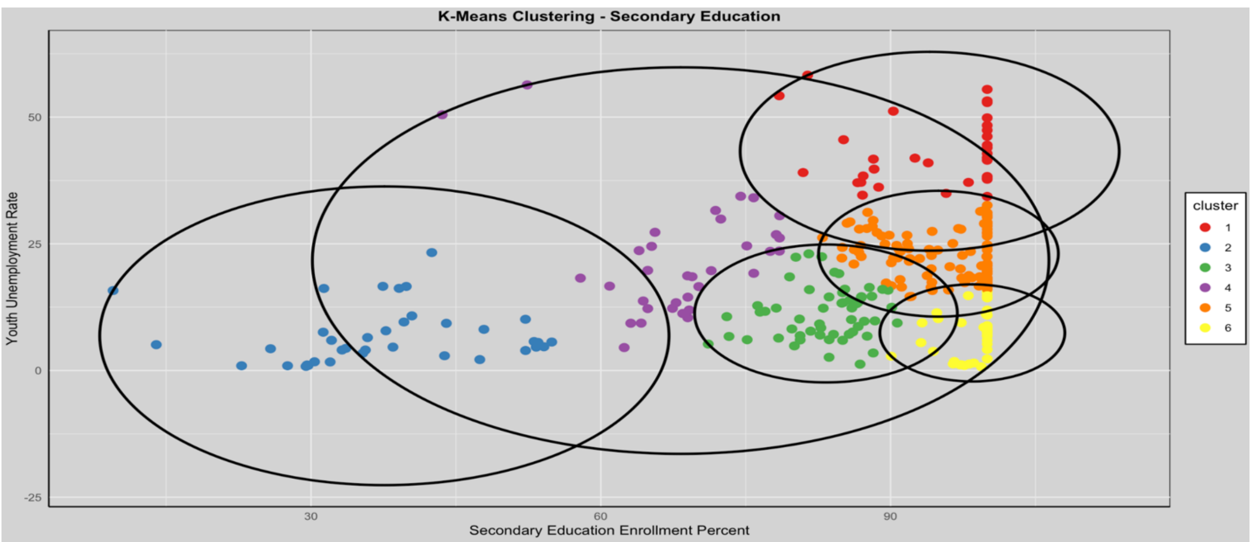

Figure 11 K-Means Cluster Analysis for Secondary Education Enrollment

Cluster 1 : Argentina, Bulgaria, Georgia, North Macedonia, Palestine, Saint Lucia, Serbia, South Africa, Spain
Cluster 2 : Afghanistan, Bangladesh, Bhutan, Cambodia, Ethiopia, Guatemala, Ivory Coast, Laos, Pakistan, Rwanda, Zimbabwe
Cluster 3 : Armenia, Azerbaijan, Bhutan, Bulgaria, Ghana, Indonesia, Iran, Mexico, Moldova, Oman, Panama, Paraguay, Peru, Philippines, Thailand, Turkey, Ukraine
Cluster 4 : Albania, Bangladesh, Belize, Colombia, Egypt, Ghana, India, Indonesia, Iran, Laos, Mauritius, Mongolia, Morocco, Panama, Paraguay, Suriname, Swaziland, Tajikistan, Tunisia
Cluster 5 : Albania, Armenia, Bhutan, Brazil, Bulgaria, Cape Verde, Chile, Colombia, Georgia, Greece, Guyana, Indonesia, Iran, Italy, Kazakhstan, Latvia, Lithuania, Mauritius, Mongolia, Morocco, Palestine, Romania, Serbia, Slovenia, Spain, Sri Lanka, Tajikistan, Tunisia, Turkey, Ukraine
Cluster 6 : Azerbaijan, Belarus, Brazil, Ecuador, Indonesia, Kazakhstan, Latvia, Lithuania, Mexico, Moldova, Peru, Poland, Qatar, Thailand, United Arab Emirates, Vietnam

Figure 12 K-Means Cluster List for Secondary Education Enrollment

Cluster Analysis for Secondary Education Enrollment:

Cluster 1 (red) represents countries with high secondary education enrollment correlating strongly with higher unemployment rates, including mainly European countries. In contrast, cluster 5 (orange), directly under cluster 1, shows significant variability in the high secondary education enrollment unemployment rate. This cluster includes mainly European and Asia countries, which comprise 73% of the countries in this cluster. Meaning attaining secondary education has a varied impact on youths' unemployment rates, suggesting the relationship is unclear. Cluster 6 (yellow) clusters countries with extremely high secondary education enrollment and low to moderate unemployment for youths with minimal variation in the unemployment rate. This cluster includes mainly countries in Asia (about 37% of countries) and Europe (about 37% of countries) and does not include African countries. This could indicate the strong alignment between education systems and job market needs in the countries included in this cluster. Cluster 2 (blue) represents countries with low secondary education enrollment and medium to high youth unemployment rates. Countries in Asia comprise approximately 55% of the countries in this cluster, and Africa 36%, with no European countries included. These countries need help with lower secondary education enrollment levels and higher unemployment rates, indicating a need for significant investments in educational infrastructure and economic development to create job opportunities. Cluster 3 (green) represents countries with moderate secondary education enrollment percentages and youth unemployment rates. Countries in Asia comprise the majority of countries in this cluster, while countries in the Americas and Europe collectively represent 47% of the countries in this cluster. This cluster indicates a balance between both variables but has the potential for improvement by increasing the quality and availability of secondary education. As for cluster 4 (purple), the countries in this cluster are characterized by moderate to high secondary education enrollment with moderate to high unemployment rates for youths. The countries' observations have high variability, presenting an unclear relationship between both variables.

Compared to primary education enrollment, secondary education enrollment has a more direct and stronger correlation with unemployment rates for youths. Higher secondary education levels are generally associated with lower unemployment rates, as seen in clusters 3 and 6. Countries located in Asia have varying relationships between the two variables, and this could be due to differing educational and economic policies of different parts of Asia, the various sizes of the emerging market countries, and differing overall education infrastructure. European countries generally have high secondary education enrollment correlated with low or high youth unemployment rates. African countries fall mainly in the cluster with low secondary education enrollment and medium to high youth unemployment rate, represented by cluster 2. Finally, countries in the Americas are found mainly in cluster 3, countries with moderate secondary education enrollment percentages and low to moderate unemployment rates for youths.

4.5.3 Tertiary Education

K-means clustering was applied to tertiary education enrollment percent, which is the highest level of education in this study. Six clusters were chosen.

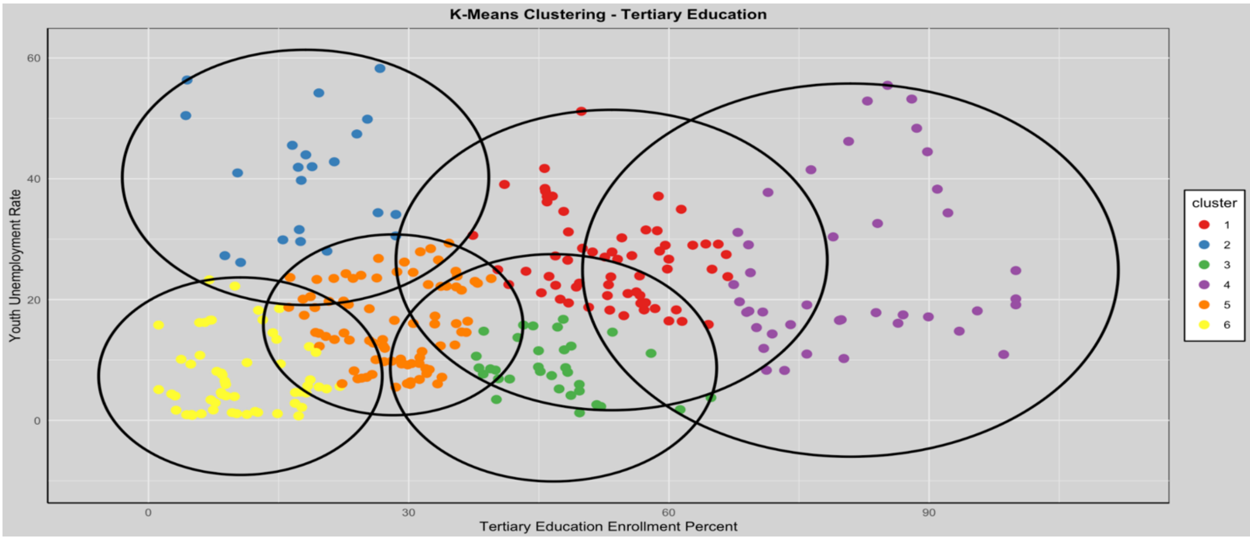

Figure 13 K-Means Cluster Analysis for Tertiary Education Enrollment

```
Cluster 1 :  Albania, Argentina, Armenia, Brazil, Bulgaria, Chile, Colombia, Georgia, Greece, Iran, Italy,
Latvia, Lithuania, Mongolia, Morocco, Palestine, Panama, Romania, Serbia, Slovenia, Spain, Turkey, Ukraine
Cluster 2 :  Albania, Bhutan, Cape Verde, Egypt, Mauritius, North Macedonia, Saint Lucia, South Africa,
Suriname, Swaziland, Tajikistan, Tunisia
Cluster 3 :  Armenia, Brazil, Ecuador, Indonesia, Kazakhstan, Mexico, Moldova, Mongolia, Panama, Paraguay,
Philippines, Thailand, United Arab Emirates
Cluster 4 :  Belarus, Bulgaria, Chile, Georgia, Iran, Latvia, Lithuania, Peru, Poland, Spain, Turkey,
Ukraine
Cluster 5 :  Albania, Armenia, Azerbaijan, Bangladesh, Cape Verde, Chile, Colombia, Egypt, Greece, India,
Indonesia, Iran, Kazakhstan, Mauritius, Mexico, Moldova, Mongolia, Morocco, Paraguay, Peru, Philippines,
Romania, Sri Lanka, Tajikistan, Tunisia, Vietnam
Cluster 6 :  Afghanistan, Bangladesh, Belize, Bhutan, Cambodia, Ethiopia, Ghana, Guatemala, Guyana, Ivory
Coast, Laos, Mexico, Morocco, Oman, Pakistan, Paraguay, Qatar, Rwanda, Zimbabwe
```

Figure 14 K-Means Cluster List for Tertiary Education Enrollment

For tertiary education enrollment, it is apparent that there is high variation between the observations in each cluster, meaning that despite being clustered together based on critical similarities, the observations captured in each cluster are different, and their rates vary. This could be due to inherent diversity where the emerging market countries in consideration have similar tertiary education enrollment percentages but differ significantly in how they impact youth unemployment rate due to factors like economic structure, technological advancement, or their labor laws. Clusters 2 (blue), 5 (orange), and 6 (yellow) all have lower percentages of tertiary education enrollment, but their respective unemployment rates for youths vary greatly. Countries clustered in cluster 3 (green) have low to moderate tertiary education enrollment and, as a result, have low to moderate youth unemployment rates. Countries located in Asia represent approximately 54% of countries in this cluster.

In comparison, countries in the Americas represent approximately 38% of countries in this cluster. However, these countries vary significantly in their population sizes and economic and job market structure, differentiating them from one another. Moreover, no countries located in Africa have been observed in this cluster. Cluster 1 (red) shows a great degree of variation between its observations and is also classified as low to moderate tertiary education enrollment similar to cluster 3 (green) but exemplifies higher youth unemployment rates as compared to cluster 3 (green). European countries comprise 52% of the countries in cluster 1 (red), while countries in Asia and the Americas collectively represent about 43% (21.7% each). Although cluster 1 (red) is similar to cluster 3 (green) regarding low to moderate tertiary education enrollment, its youth unemployment rates are higher and could be due to excessive mismatches between education level and job market demands in the countries assigned to this cluster. Finally, countries clustered in cluster 4 (purple) have significantly high levels of tertiary education enrollment, but their respective youth unemployment rates vary significantly. Despite very high tertiary education rates, the variability in youth unemployment indicates that while some graduates find appropriate job opportunities, others do not. This could also be caused by mismatches between education level and job market demands, as well as by oversaturation in specific job fields/industries or other economic factors inhibiting job creation in these countries.

4.5.4 Comparison of clustering analysis for primary, secondary, and tertiary education levels

The general trend identified for primary education level had the weakest correlation between education enrollment and youth unemployment rate. This is mainly because primary education, while essential for developing basic literacy and foundational skills, can only partially impact employment opportunities in modern emerging market economies. Variability between observations in clusters was very low, suggesting that primary education does not drastically change employment prospects, but it does set a baseline educational standard across emerging market countries. Secondary education showed a stronger correlation between education enrollment and youth unemployment rates than primary education as it begins to equip youths with more specialized skills and knowledge that are directly linked to employment requirements, increasing their employment prospects. However, moderate variability was observed between countries showing high secondary education enrollment rates and lower youth unemployment rates.

In contrast, other countries with high secondary education enrollment rates have moderate to high youth unemployment rates. This indicates gaps in economic structures, mismatches between education and job market demands, or possibly education quality. Finally, the highest level of education, tertiary education, exemplified the most complex relationship between its education enrollment rate and youth unemployment rate. Higher levels of tertiary education generally correlated with better employment outcomes but also presented the highest variability, making it difficult to conclude the relationship between the two variables and mine patterns relevant to different geographic regions and continents. This high variability suggests a significant mismatch in some regions between the level of education and suitable job opportunities. High tertiary education is generally associated with low youth unemployment rates in some countries. In contrast, in other countries, a high tertiary education level is associated with moderate to high youth unemployment rates, meaning higher levels of education do not guarantee better employment outcomes for youths in emerging market countries.

4.6 Forecasting

In this section, the study aims to predict which continent will see increasing youth unemployment rates in the next five years – 2023 to 2027. Since the data available is up to 2022, the model will forecast for 2023 and beyond. The continents are the Americas (including North and South America), Africa, Asia, and Europe. This geographic context will provide insightful information to allow policymakers and key stakeholders to focus on the regions facing an increasing unemployment issue, understand the factors responsible, and work on developing the appropriate solutions. Geographic context facilitates the identification of the factors responsible for the rising unemployment rate for youths through comparison of the socioeconomic situations and relevant policies of the regions that are predicted to have stable unemployment rates or decreasing unemployment rates for youths in their respective emerging market countries.

4.6.1 Geographical Analysis

Using Tableau, we visualize the average youth unemployment rates for the years 1990 to 2022 geographically to identify areas of emerging market countries with higher and lower than average youth unemployment rates.

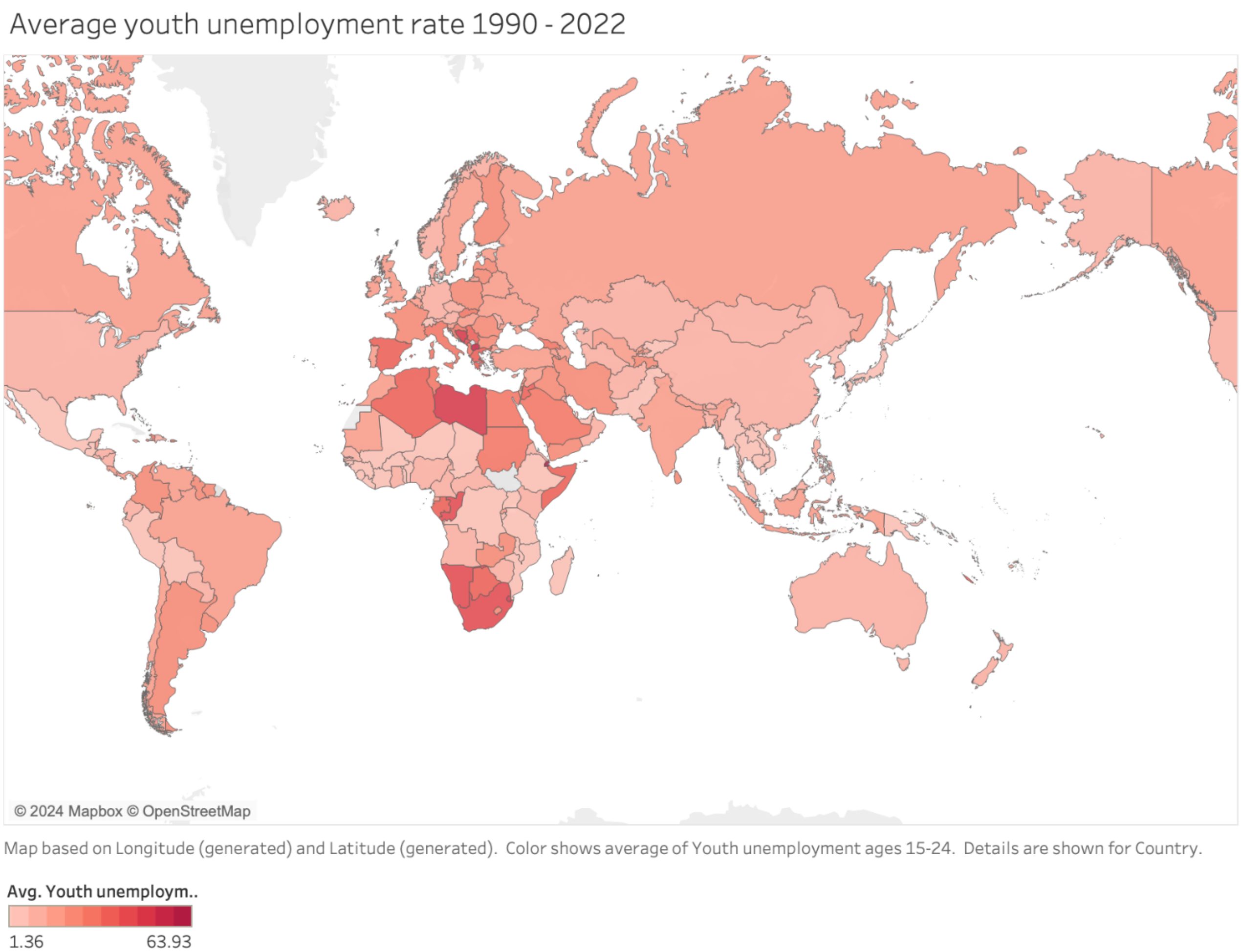

Figure 15 Average Youth Unemployment Rate 1990 – 2022

Countries with lighter shades have lower average unemployment rates for youths, while countries with darker shades have higher average unemployment rates. It is apparent that countries in Africa, Europe, and Asia have relatively darker shades, indicating higher average unemployment rates for youths, and Africa has the darkest shades, indicating countries in Africa have the highest unemployment rates for youths.

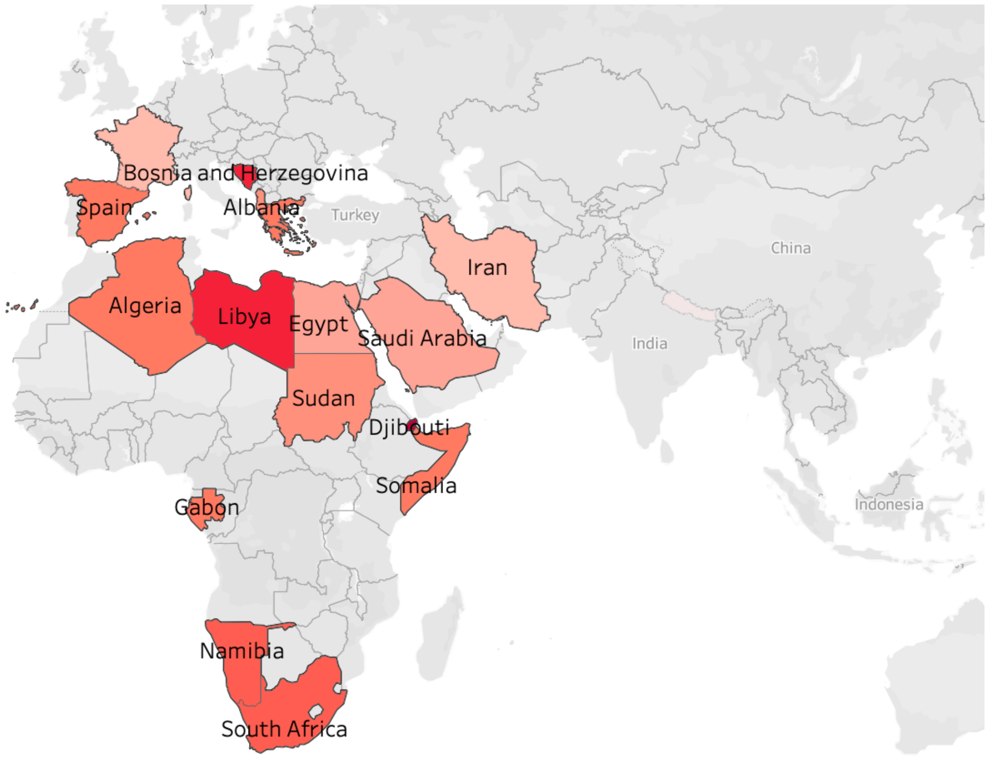

Figure 16 Highest Average Youth Unemployment Rate 1990 – 2022

Looking closer at the highest average rates of unemployment for youths, we find that Djibouti has the highest average unemployment rate for youths at an average value of 63.93% for the years 1990 to 2022. Bosnia and Herzegovina have an average youth unemployment rate of 48.85%, Spain has an average rate of 34.88%, and Greece has an average rate of 33.23%. While Libya has an average rate of 47.51%, Namibia has an average rate of 40.17%, and South Africa has an average rate of 39.35%. We can conclude that Africa has the highest unemployment rates for youths and Europe comes in second place.

4.6.2 Time Series Forecasting

This section includes predicting the youth unemployment rates of the African, American, Asian, and European continents for the next five years (2023 to 2027) to determine which continent will see increasing unemployment rates for youths in emerging market countries.

The time series model used was Autoregressive Integrated Moving Average (ARIMA), which effectively captures patterns and trends in time series data and provides highly reliable forecasts, particularly for short to medium-term predictions, as it is in this case. The model was produced by creating subsets of the continents, including their respective countries' data. We then faced multiple challenges when applying time series forecasting models, including various observations for each year in the dataset due to different countries' unemployment rates. We aggregated the data by taking the mean unemployment rate for each year. Then another challenge posed was that some continents' subsets had gaps in the time series, meaning the time series is incomplete, with many years having no recorded observations. This issue was solved through tidyr package to complete the sequence of years for each continent to ensure no gaps in the time series sequence. In addition to the zoo package which provided the linear interpolation of the missing years by estimating the values of the missing years by assuming that the relationship between the known observations is linear and computing the slope between the known observations that are recorded before and after the gap years to calculate the youth unemployment rates for the missing years in between. Once the data sets were aggregated and complete, we applied the ARIMA time series forecasting model to predict the youth unemployment rates for the next five years for each continent in consideration. The following results were produced.

continent <chr>	Year <dbl>	.mean <dbl>
Africa	2023	24.70302
Africa	2024	25.00013
Africa	2025	25.10241
Africa	2026	25.13762
Africa	2027	25.14974
Americas	2023	26.38296
Americas	2024	27.14444
Americas	2025	27.90593
Americas	2026	28.66741
Americas	2027	29.42889
Asia	2023	22.71149
Asia	2024	19.09362
Asia	2025	17.19625
Asia	2026	16.20119
Asia	2027	15.67934
Europe	2023	31.61670
Europe	2024	32.13634
Europe	2025	32.35469
Europe	2026	32.44643
Europe	2027	32.48498

Figure 17 ARIMA Model Point Forecasts

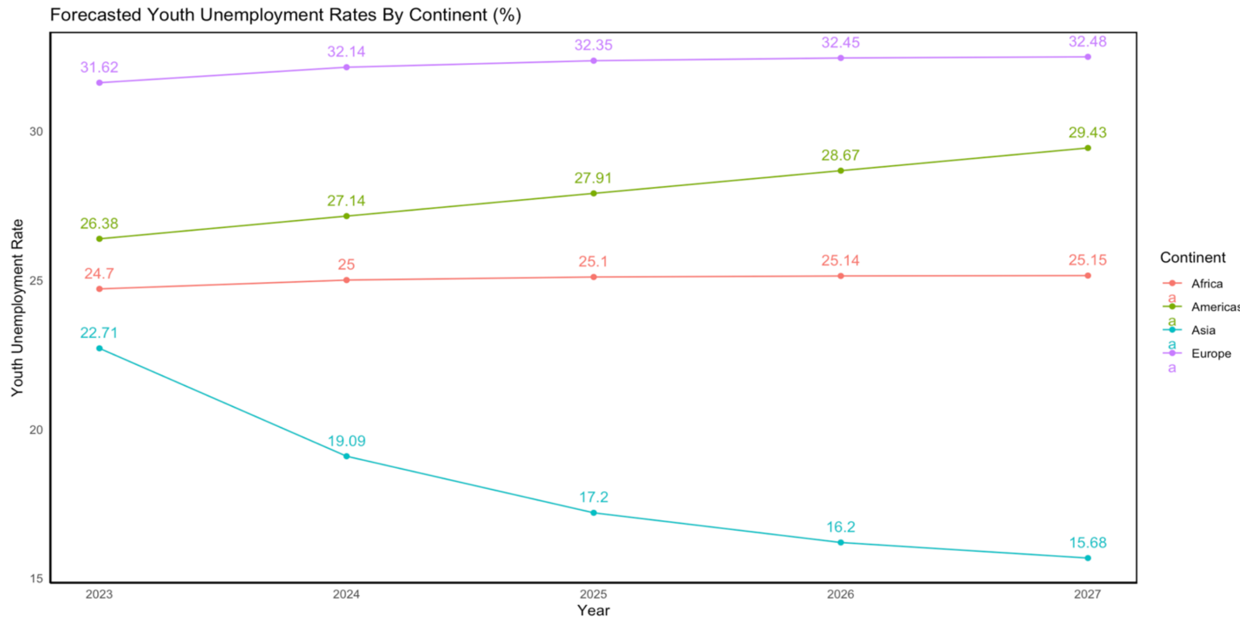

Figure 18 Forecasted Youth Unemployment Rates by Continent (%)

Time series forecasting analysis:

Based on the predicted youth unemployment rates of the time series forecasting model, we immediately see that Asia is expected to have a consistent drop in its youth unemployment rates in its respective emerging market countries from the year 2023 to the year 2027, decreasing by about 15.91% in 2024 compared to 2023. All the remaining continents exhibit an increase in unemployment rates for youths in the respective emerging market countries. Africa shows a constant positive increase in its youth unemployment rate, increasing at significantly small rates by approximately 1.21% in 2024, 0.40% in 2025, 0.16% in 2026, and 0.04% in 2027. The Americas shows a consistent increase in youth unemployment rate for its emerging market countries, with approximately a 3% increase every year from 2023 to 2027. Finally, Europe has the highest forecasted youth unemployment rates for the respective emerging market countries, which are estimated to be above 30%.

In comparison, the remaining continents' youth unemployment rates are estimated at less than 30%. Although Europe exemplifies the highest youth unemployment rates when compared to other continents, Europe's youth unemployment rates are expected to increase at a very low rate through the years to 2027, by increasing approximately 1.64% in 2024, 0.65% in 2025, 0.31% in 2026 and 0.09% in 2027. To conclude, the continent predicted to have increasing unemployment rates for youths in emerging market countries is the Americas (including North and South America) since its youth unemployment rate is expected to increase by approximately 3% each year from the year 2023 to 2027.

ARIMA Model Evaluation:

In this section, we explore key statistics and parameters to evaluate the ARIMA model's performance in time series forecasting and understand how well it fits the data.

continent	Year	.mean	.model	sigma2	log_lik	AIC	AICc	BIC	ar_roots	ma_roots
<chr>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<list>	<list>
Africa	2023	24.70302	model	100.025012	−114.40047	234.8009	235.6898	239.1029	<cplx [1]>	<cplx [0]>
Africa	2024	25.00013	model	100.025012	−114.40047	234.8009	235.6898	239.1029	<cplx [1]>	<cplx [0]>
Africa	2025	25.10241	model	100.025012	−114.40047	234.8009	235.6898	239.1029	<cplx [1]>	<cplx [0]>
Africa	2026	25.13762	model	100.025012	−114.40047	234.8009	235.6898	239.1029	<cplx [1]>	<cplx [0]>
Africa	2027	25.14974	model	100.025012	−114.40047	234.8009	235.6898	239.1029	<cplx [1]>	<cplx [0]>
Americas	2023	26.38296	model	7.921713	−66.25104	136.5021	136.9821	139.1665	<cplx [0]>	<cplx [0]>
Americas	2024	27.14444	model	7.921713	−66.25104	136.5021	136.9821	139.1665	<cplx [0]>	<cplx [0]>
Americas	2025	27.90593	model	7.921713	−66.25104	136.5021	136.9821	139.1665	<cplx [0]>	<cplx [0]>
Americas	2026	28.66741	model	7.921713	−66.25104	136.5021	136.9821	139.1665	<cplx [0]>	<cplx [0]>
Americas	2027	29.42889	model	7.921713	−66.25104	136.5021	136.9821	139.1665	<cplx [0]>	<cplx [0]>
Asia	2023	22.71149	model	26.074558	−72.30273	150.6055	151.8055	154.1396	<cplx [1]>	<cplx [0]>
Asia	2024	19.09362	model	26.074558	−72.30273	150.6055	151.8055	154.1396	<cplx [1]>	<cplx [0]>
Asia	2025	17.19625	model	26.074558	−72.30273	150.6055	151.8055	154.1396	<cplx [1]>	<cplx [0]>
Asia	2026	16.20119	model	26.074558	−72.30273	150.6055	151.8055	154.1396	<cplx [1]>	<cplx [0]>
Asia	2027	15.67934	model	26.074558	−72.30273	150.6055	151.8055	154.1396	<cplx [1]>	<cplx [0]>
Europe	2023	31.61670	model	95.933572	−117.48904	240.9781	241.8352	245.3753	<cplx [1]>	<cplx [0]>
Europe	2024	32.13634	model	95.933572	−117.48904	240.9781	241.8352	245.3753	<cplx [1]>	<cplx [0]>
Europe	2025	32.35469	model	95.933572	−117.48904	240.9781	241.8352	245.3753	<cplx [1]>	<cplx [0]>
Europe	2026	32.44643	model	95.933572	−117.48904	240.9781	241.8352	245.3753	<cplx [1]>	<cplx [0]>
Europe	2027	32.48498	model	95.933572	−117.48904	240.9781	241.8352	245.3753	<cplx [1]>	<cplx [0]>

Figure 19 ARIMA Model Evaluation

Based on the evaluation of the results of the above key statistics and parameters, we evaluate the ARIMA model performance as follows. Americas exhibits the best model performance with the lowest variance in residuals and the best efficiency metrics (AIC, BIC). The simplicity and stability of the model are indicated by the absence of complex roots. Asia stands in the middle with moderate variance and efficiency metrics. The presence of complex roots might require a closer look to ensure the model's stability. As for Africa and Europe, both show signs of less efficient models with higher variance and the highest log-likelihood and efficiency metrics. The complex roots in their AR and MA components suggest that these models may capture general cyclical behaviors but not optimally. The results suggest that the Americas model was the best fit and Asia’s generally performs adequately while models of Africa and Europe must be revised in terms of specifications to improve the fit of the model.

4.7 Comparative Analysis of Study Findings

4.7.1 Analysis of the Relationship between Education Level and Youth Unemployment Rates

To determine the nature of the relationship between education level and youth unemployment rates for emerging market countries, 4 data mining techniques were used: correlation analysis, regression analysis, classification decision tree analysis, and cluster analysis. The data mining techniques produced varied results, and the correlation analysis and regression analysis showed a positive relationship between education level and youth unemployment rates in emerging market countries. As you increase the youth education level, the unemployment rate also rises. The results of both data mining techniques suggest that the relationship is generally seen as insignificant, weak, and unreliable. However, the classification decision tree results and the cluster analysis performed for each education level suggest that the relationship between education level and youth unemployment rates is significantly inverse. As you increase the youth's education level, unemployment rates decrease. This is very apparent in the decision tree generated in which, according to the data, higher tertiary education enrollment significantly affects unemployment rates for youths, as the rates are at their lowest when tertiary education enrollment is 76% and above. In addition, cluster analysis results showed that secondary education level has a stronger correlation between education enrollment and youth unemployment rates than primary education. Still, when the education level is increased to tertiary education, there is significant variation in the clusters, with some showing positive relationships between both variables and others showing negative relationships. We can conclude that youths with secondary education have better employment prospects and lower unemployment rates than those with primary education or basic literacy skills. Still, as the youths obtain higher graduate-level degrees, the relationship is ambiguous, and other economic factors have more significant effects on the employment prospects of these graduates.

4.7.2 Analysis of Time Series Forecasting Findings

The geographic visualizations generated by Tableau gave us a factual view of the continents with the highest unemployment rates for youths in the respective emerging markets: Africa and Europe. It was concluded that Africa was the continent with the highest observed youth unemployment rates in emerging market countries, with Djibouti having an average of 63.93%, Namibia having an average rate of 40.17%, and South Africa having an average rate of 39.35%. After applying the time series forecasting model to predict which continent will continue to see higher unemployment rates for youths in emerging market countries, it was deduced that the Americas (including North and South American countries) will continue to increase by approximately 3%. At the same time, Africa will remain mostly stable, rising at a very low rate each year.

Chapter 5- Conclusion

5.1 Conclusion

This study was guided by three main objectives, including determining the nature of the relationship between education level and unemployment rate for youths in emerging market countries and assessing the strength of that relationship. It also identified which continent is predicted to have increasing youth unemployment rates in the next five years. Various data mining techniques were applied to draw the necessary conclusions, which are as follows.

5.1.1 The Impact of Education Level on Youth Unemployment Rates in Emerging Market Countries

The data mining techniques provided varied results regarding the impact education level has on youth unemployment rates in emerging market countries. Although initially, the correlation analysis and regression analysis data mining techniques showed a weak positive relationship between education level and youth unemployment rates in emerging market countries, it was then proved that higher education levels, precisely secondary education level, provide better employment outcomes and are therefore associated with lower unemployment rates for youths in emerging market countries. However, as the education level increased past the secondary education level, represented by the tertiary education level, the unemployment rate for youths in emerging markets had high variability based on the k-means clustering data mining technique. Still, it showed significantly lower unemployment rates based on the classification decision tree model. We can interpret that education level impacts youth unemployment rates in emerging market countries in an inverse manner, as the education level increases as the youths develop their knowledge and skills and attain higher education degrees, the unemployment rate generally decreases but not to a high degree as there is significant variability in the results of the data mining techniques and also due to the weak correlation associated between the variables. This could be mainly due to the mismatches between higher education graduates and the limited number of suitable job opportunities available to accommodate them, possibly because the emerging market countries in consideration could need more infrastructure to create and foster job opportunities for youths.

5.1.2 Assessing the Nature of the Relationship between Education and Youth Unemployment Rates

The analysis showed that the relationship between education level and youth unemployment rates is more complex than our initial thoughts. It is a complex relationship with various dimensions and angles that must be considered. This includes the quality of education and the structure of education systems for different countries. The constitution of education systems differs significantly in various regions and across different cultures, many education systems focus on developing technical skills of youths that directly enhance their employment prospects, while many others are dynamic being flexible to meet evolving job market demands and focus on equipping youths with the skills and knowledge required to meet job market demands. On the other hand, many education systems have poorer performance and rigid curriculums that do not meet current and evolving job market demands, which, as a result, lowers youths' employment prospects, increasing the unemployment rate for youths. The availability and affordability of education are other factors that need to be considered. This affects the percentage of youths enrolled in higher education and their ability to increase their employment prospects.

5.1.3 Geographic Assessment of Forecasted Unemployment Rates by Continent

During the time series forecasting stage of the study, we were able to forecast the predicted youth unemployment rates for the Americas, Africa, Asia, and Europe continents for the years 2023 to 2027, focusing on emerging market countries. We successfully determined that Americas (including South and North America) emerging market countries are predicted to have increasing unemployment rates for youths for the next five years from 2023 to 2027, an increase of approximately 3% per year. Meanwhile, other continents like Africa and Europe will see an increase in youth unemployment rates each year by minimal rates. Despite many emerging market countries' efforts in developing and implementing policies to maintain youth unemployment rates and foster job creation, the respective emerging market countries in the Americas, Africa and Europe need to study the socioeconomic causes of the gap between graduates and unemployment and are urged to re-asses their relevant policies to ensure suitable education curriculums and vocational training is in place to support the employment of youths to avoid this increasing and persistent issue. Asia is the only continent predicted to see a significant decrease in its youth unemployment rates from 2023 to 2027, and this could be due to the considerable investments from its respective emerging market countries in education infrastructure and the development of many employment training programs to ease the transition of youths from students to employees and enhance their employment outcomes. Countries in Asia are leading in preparing their youths with evolving job market demands as Singapore, South Korea, China, Japan, India United Arab Emirates, and Saudi Arabia have successfully developed education curriculums with a heavy focus on artificial intelligence, machine learning, data science and STEM (science, technology, engineering, and mathematics) courses. The failure of emerging market countries in the Americas, Africa, and Europe to improve their education infrastructure and work on reducing the mismatch between graduates and job opportunities will only lead to a bigger unemployment problem for the upcoming youth generation, which will be more challenging to control shortly.

5.2 Recommendations and Future Work

Unemployment is inevitable; it cannot be eliminated but should be closely monitored and regulated to maintain the overall health and productivity of the economy. Youths face adversity with unemployment since job markets are rapidly evolving, and youths seeking employment need adequate skills and knowledge to enhance their job prospects, which is attained mainly through education degrees. After a deeper look into the complex relationship between education level and unemployment rates for youths and deducing the degree of impact education level has on unemployment rates, we suggest re-assessing the education curriculums in different regions around the world to match the specific job market needs of that region. Education curriculums should be flexible and dynamic in meeting the constantly changing demands of the job market and meeting the needs of newer skill sets sought by different industries. Vocational and on-the-job training should be heavily implemented in education curriculums worldwide to equip youths with the skillsets they need for their future jobs to ensure better employment prospects. Finally, relevant authorities and stakeholders should develop and implement policies to balance the oversaturation in specific job fields/industries.

A recommendation for future work and research on this topic would be to consider the education quality, infrastructure, and the level of vocational and on-the-job training provided to youths in different regions worldwide to assess the direct relationship between youth unemployment rates and their respective education levels.

References

1. Alawad, A. S., Kreishan, F., & Selim, M. (2020). Determinants of youth unemployment: Evidence from Jordan. *International Journal of Economics and Business Administration*, *8*(4), 152-165.
2. Alçin, s., Sahin, B. E., & Hamzaoglu, N. M. (2021). Analysis of the relationship between education and youth unemployment: evidence from Turkey and Spain. *Journal of Life Economics*, *8*(2), 185-192.
3. Bhorat et al., (2016). *Higher education, employment and economic growth: Exploring the interactions.* Taylor and Francis Online.
4. Biagi, F., & Lucifora, C. (2005). Demographic and education effects on European unemployment: Economic factors and labor market institutions. IZA Discussion Papers, No. 1806, Institute for the Study of Labor (IZA), Bonn.
5. Blasques, Francisco, et al. "Dynamic Factor Models with Clustered Loadings: Forecasting Education Flows Using Unemployment Data." *International Journal of Forecasting*, Elsevier, 23 Mar. 2021, www.sciencedirect.com/science/article/pii/S0169207021000261.
6. Bojadjieva, D. M., Cvetanoska, M., Kozheski, K., Mujčinović, A., & Gašparović, S. (2022). The impact of education on youth employability: The case of selected southeastern European countries. *Youth & Society*, *54*(2_suppl), 29S-51S.
7. Bugudui, E. (2016). The Use of Multivariate Techniques for the Unemployment Analysis. *Academic Journal of Economic Studies*.
8. Celbiş, M. G. (2022, June 2). *Unemployment in Rural Europe: A Machine Learning Perspective*. Applied Spatial Analysis and Policy. https://doi.org/10.1007/s12061-022-09464-0
9. Collins, R. (2009). *Factors related to the unemployment rate: A statistical analysis.* University of Northern Iowa.
10. Corporate Finance Institute. (n.d.). *Unemployment*. Retrieved from Corporatefinanceinstitute.
11. Couppie et al., (2000). *The Situation of Young Labour-Market Entrants in Europe.* Institution of Education Sciences.
12. Cristescu, A. (2017). The impact of education on the unemployment rate in the Southern European Model. *Romanian Journal of Regional Science*, *11*(1), 62-75.

13. Egessa, A., Nnyanzi, J. B., & Muwanga, J. (2021). Determinants of youth unemployment in Uganda: The role of gender, education, residence, and age. *IZA Journal of Labor Policy*, *11*(1), 1-29.
14. Erdem, E., & Tugcu, C. T. (2012). Higher Education and Unemployment: A cointegration and causality analysis of the case of Turkey. *European Journal of Education*, *47*(2), 299-309.
15. Hayes, A. (2023, August 9). *What is Unemployment? Understanding Causes, Types, Measurement*. Retrieved from Investopedia: https://www.investopedia.com/terms/u/unemployment.asp
16. Hérault, N., Kostenko, W., Marks, G., & Zakirova, R. (2012). The effects of macroeconomic conditions on the education and employment outcomes of youth. *Australian Journal of Labour Economics*, *15*(1), 17-36.
17. IBM. (2021, August 17). *Introduction to CRISP-DM*. Retrieved from IBM: https://www.ibm.com/docs/en/spss-modeler/saas?topic=guide-introduction-crisp-dm
18. Jantawan et al., (2013). TThe Application of Data Mining to Build Classification Model for Prediciting Graduate Employment. *International Journal of Computer Science and Information Security*, 7.
19. Lam, D., Leibbrandt, M., & Mlatsheni, C. (2009). Education and youth unemployment in South Africa. In *Labor markets and economic development* (pp. 90-109). Routledge.
20. Maneejuk, P., & Yamaka, W. (2021). The impact of higher education on economic growth in ASEAN-5 countries. *Sustainability*, *13*(2), 520.
21. Mincer, J. (1991). Education and Unemployment. *Working Paper No. 3838*.
22. Mpendulo, G., & Mang'unyi, E. E. (2018). Exploring relationships between education level and unemployment. *Journal of Social Sciences (COES&RJ-JSS)*, *7*(2), 86-102.
23. Nickell, S. (1979). Education and Lifetime Patterns of Unemployment. *Journal of Political Economy*.
24. Nur Iman Natasha Binti A'rifian, et al. (2019) "A Comparative Study on Graduates' Employment in Malaysia by using Data Mining." Journal of Physics: Conference Series 1366 012120.
25. Oktafianto, E. K., Achsani, N. A., & Irawan, T. (2019). The determinant of regional unemployment in Indonesia: The Spatial Durbin Models. *Journal of Economics*, *8*(2), 179-194.
26. Parliament of Australia. (n.d.). Chapter 2: Consequences of Unemployment - Parliament of Australia .
27. Pompei, F., & Selezneva, E. (2021). Unemployment and education mismatch in the EU before and after the financial crisis. *Journal of Policy Modeling*, *43*(2), 448-473.

28. Puspadjuita, E. (2018). Factors that Influence the Rate of Unemployment in Indonesia. *International Journal of Economics and Finance*, 8.

29. Rahmawati, A., & Putri, P. I. (2021). The effect of education variable on the open unemployment rate. *Efficient: Indonesian Journal of Development Economics*, *4*(2), 1160-1173.
30. Ramzan, M., Abbas, M., Aslam, S., & Farooq, U. (2018). Can education be a solution to the unemployment? *WALIA Journal*, *34*(1), 71-78.
31. Riddell et al., The impact of education on unemployment incidence and re-employment success: Evidence from the U.S. labour market, Labour Economics, Volume 18, Issue 4, 2011, Pages 453-463.
32. Singh, A., Aggarwal, M., & Jain, P. (2019). Education and unemployment in rural and urban Kerala. *Available at SSRN 3464399*.
33. Suyanto, S., Purnomo, B., & Standsyah, R. E. (2019). Impact of education and health on the unemployment rate and economy of East Java. Management Journal, *9*(2), 26-31.
34. statista. (2023, February). *Number of unemployed persons worldwide from 1991 to 2024*. Retrieved from statista: https://www.statista.com/statistics/266414/unemployed-persons-worldwide/
35. statista. (2023, September). Global unemployment rate from 2003 to 2022. Retrieved from statista: https://www.statista.com/statistics/279777/global-unemployment-rate/
36. Zhongchang et al., (2007). The relationship between education and employment: A theoretical analysis and empirical test. Higher Education Press and Springer-Verlag.