Forecasting Global Mental Health Disorders : A Machine Learning Approach using Socioeconomic Indicators

MSc Research Project
Data Analytics

Sonal Suryakant Puradkar
Student ID: x22130250

School of Computing
National College of Ireland

Supervisor: Arjun Chikkankod

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sonal Suryakant Puradkar
Student ID:	x22130250
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Arjun Chikkankod
Submission Due Date:	14/12/2023
Project Title:	Forecasting Global Mental Health Disorders : A Machine Learning Approach using Socioeconomic Indicators
Word Count:	XXX
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Sonal Suryakant Puradkar
Date:	30th January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Forecasting Global Mental Health Disorders : A Machine Learning Approach using Socioeconomic Indicators

Sonal Suryakant Puradkar
x22130250

Abstract

This research examines the global impact of socioeconomic indicators on the prevalence of mental health disorders, with a focus on depressive disorders, schizophrenia, bipolar disorder, eating disorders, and anxiety disorders. Employing machine learning techniques and data from Our World in Data and the World Bank spanning 1960-2019, we predict and analyze the influence of Adjusted Net National Income per capita, Inflation, Employment distribution, Proportion of people below median income, Unemployment rates, New businesses registered, and Multidimensional Poverty. The study aims to identify correlations, forecast disorder trends, and pinpoint countries with heightened vulnerability. Insights gained from this research will inform targeted interventions, promoting a proactive approach to global mental health challenges.

1 Introduction

1.1 Background and Motivation :

The pervasive impact of mental health disorders on global well-being necessitates a comprehensive exploration of predictive factors, fostering a proactive approach to intervention and support systems. Among the myriad mental health disorders, the focus of this research includes depressive disorders, schizophrenia, bipolar disorder, eating disorders, and anxiety disorders. These conditions collectively contribute significantly to the global burden of disease, with a staggering one in five individuals grappling with diagnosable mental illness annually.

The selected mental health indicators, specifically Disability-Adjusted Life Years (DALYs), offer a standardized metric for assessing the burden of these disorders. DALYs provide a nuanced perspective by accounting for both the years of healthy life lost to disability and premature mortality, allowing a comprehensive evaluation of the societal impact of mental health conditions. This research seeks to understand and predict variations in the prevalence of these disorders, unravelling the intricate interplay between socioeconomic indicators and mental health outcomes.

The chosen socioeconomic indicators wield considerable influence on mental health trajectories. Adjusted Net National Income per capita serves as a proxy for economic well-being, reflecting the financial resources available for mental health support systems. Inflation, an economic indicator, contributes to the understanding of the cost of living and potential stressors on individuals. Employment distribution across industry, services, and self-employment reflects the occupational landscape, shedding light on potential stressors tied to specific sectors.

The proportion of people living below median income serves as a socio-economic marker, highlighting income disparities that can impact access to mental health resources. Unemployment rates, both overall and specifically among those with advanced education, offer insights into the stability of the job market and potential mental health stressors associated with unemployment. New businesses registered signify economic dynamism and potential shifts in the labor market, while Multidimensional Poverty encapsulates a holistic measure of deprivation, incorporating aspects beyond income.

1.2 Research Question :

As our exploration delves into the predictive power of these indicators, drawing from sources like Our World in Data and the World Bank, the research not only aims to forecast the rise of mental health disorders globally but also to identify the underlying correlations and vulnerabilities. By leveraging machine learning techniques, this study strives to provide a nuanced understanding of the intricate relationship between socioeconomic factors and mental health outcomes, ultimately contributing to the development of targeted interventions and resource allocation strategies.

RQ : Predicting how socioeconomic indicators such as Adjusted net national income per capita, Inflation, Employment (across industry, services or self-employment), Proportion of people living below median income, Unemployment (considering advanced education and total unemployment), New businesses registered, and Multidimentional poverty can lead to rise in Mental Health disorders in countries using Machine learning techniques ?

Sub RQ : How workplace trends can impact mental well-being of employees?

1.3 Research Objective :

In order to assess the applicability and integration of previously employed methodologies and findings into this study, and to identify significant deficiencies in the reviewed literature, the research objectives encompass and appraise a critical evaluation of the existing literature pertaining to the themes under investigation. This action was taken to guarantee that this study could benefit from and integrate previously employed methodologies and discoveries. Further objectives of the research are delineated in Table 1, entitled Research Objectives.

Section one of this study provides an introduction that introduces the research concept. Section two of the document examines relevant literature. Section three outlines the research methodology. Lastly, section four includes the design specifications. The fifth section addresses the implementation of machine learning, the sixth section evaluates the machine learning methods, and the seventh section concludes with potential future research and a list of citations.

Objective Number	Research Objective
1	Studying the existing researches done in forecasting mental health disorders.
2	Studying the machine learning algorithms to forecast the predictions in various countries for a period of 3 years.
3	Identifying if some of these disorders are correlated and can be explained by the data and machine learning algorithms employed in research.
4	Identifying the top 5 countries where percentage rise in the cases of disorders will be highest.
5	Identifying the top 5 countries with highest cases of these disorders.
6	Creating visualizations to study the country wise data showing rise in cases and number of cases forecasted by the study.
7	Comparison of the developed models.

Table 1: Research Objectives

2 Related Work

The relationship between mental health illnesses and socio-economic factors goes beyond just looking at how common they are, and instead explores the intricate effects of economic changes in different countries. In prosperous cultures, characterized by positive socio-economic indices, individuals frequently enjoy better mental well-being as a result of increased availability of high-quality healthcare, education, and secure work. On the other hand, in economically poor areas, the lack of resources and opportunity leads to more stressors, which in turn increases the occurrence of mental health illnesses.

The mental health environment evolves in tandem with the fluctuations in socio-economic indices. The ability for economic expansion to mitigate certain stressors may result in a reduction of specific illnesses. Nevertheless, the process of globalization and urbanization might bring up novel difficulties, such as social detachment and intensified rivalry, which can have an adverse effect on mental health. Furthermore, there are ongoing gaps in the accessibility of mental health care, which are indicative of socio-economic inequalities. To tackle these inequalities, it is necessary to implement comprehensive approaches that extend beyond mere economic development. These approaches should include initiatives to promote mental health awareness, reduce stigma, and establish strong support networks. An integrated strategy that takes into account the complex interplay between socio-economic issues and mental health is crucial for promoting a worldwide atmosphere that places a high value on mental well-being.

2.1 Socio Economic Indicators and their impact on mental health:

In recent times there has been significant progress in the field of mental health prediction mainly due to the availability of data and advancements in machine learning techniques. Scientists have employed approaches, including statistical modelling, machine learning and a combination of methods to forecast the occurrence and frequency of mental health disorders ((Kessler et al.; 2005),(McGrath et al.; 1995),(Jose et al.; 2017),(Goodwin and Jamison; 2007),(Stice and Fairburn; 2013),(Williams et al.; 1997)).

These studies consistently reveal a connection between socioeconomic factors like poverty, unemployment, social isolation and discrimination, with an increased risk of developing mental health disorders. These findings are supported by research conducted by ((Kessler et al.; 2005),(McGrath et al.; 1995),(Jose et al.; 2017),(Goodwin and Jamison; 2007),(Stice and Fairburn; 2013),(Williams et al.; 1997),(Patel et al.; 2007),(Marmot; 2005),(Link and Phelan; 1995)).

According to a study conducted by (Kessler et al.; 2005). In 2005 it was found that individuals living in poverty were twice as likely to experience anxiety disorders compared to those with a socioeconomic status. This highlights how socioeconomic factors greatly impact health outcomes ((Kessler et al.; 2005)).

A study by (McGrath et al.; 1995). In 1995 revealed a connection between unemployment and the onset of schizophrenia. This indicates that economic hardships can increase a persons vulnerability to developing this mental condition ((McGrath et al.; 1995)).

Moreover consistent research has shown that both social isolation and discrimination significantly increase the chances of developing eating disorders ((Stice and Fairburn; 2013)) bipolar disorders ((Goodwin and Jamison; 2007)) and depressive disorders ((Kessler et al.; 2005)). (Stice and Fairburn; 2013) found that individuals who experienced isolation had a higher likelihood of developing eating disorders like anorexia nervosa and bulimia nervosa. (Goodwin and Jamison; 2007) examined the link, between discrimination and the emergence of disorders highlighting how social stigma and prejudice can harm an individuals mental well being ((Goodwin and Jamison; 2007)).

These findings underscore the diverse aspects of mental health concerns and emphasize the significant impact of social factors on an individuals mental well being. The growing body of research on the link between conditions and mental health disorders highlights the vital importance of considering these factors when developing effective strategies, for prevention, intervention and treatment. To foster an mentally healthy society it is essential to address the underlying socioeconomic disparities that contribute to mental health conditions.

2.2 Review of studies done on forecasting mental health disorders using statistical approaches

Statistical methods have proven to be tools in predicting mental health disorders providing important insights into the frequency, occurrence and contributing factors of these conditions. Numerous studies have employed statistical modeling techniques to estimate the likelihood of developing health problems achieving varying degrees of success.

A research conducted by (Almeida et al.; 2009) utilized structural equation modeling to explore the connection between early life adversity and the onset of depression. Their study revealed a link between childhood trauma and the emergence of depression in adulthood underscoring the enduring impact that negative experiences can have on mental well being ((Almeida et al.; 2009)).

In another study (McLaughlin et al.; 2006) employed latent growth curve modeling to investigate how childhood maltreatment is associated with changes in symptoms of health over time. According to their findings individuals who experienced childhood maltreatment exhibited a pronounced increase in depressive symptoms compared to those without such a history ((McLaughlin et al.; 2006)).

Furthermore (Copeland et al.; 2011) conducted a study using hierarchical linear modeling to examine the relationship between support and mental health outcomes, among older individuals.The study findings indicated that individuals with social connections tend to experience lower levels of sadness and anxiety compared to those with limited social support ((Copeland et al.; 2011)).

To explore the relationship between variables and the development of anxiety disorders (Chapman et al.; 2014) conducted research using generalized linear mixed modeling. Their findings revealed a genetic influence on anxiety disorders suggesting that genetic predispositions play a role in the emergence of these conditions ((Chapman et al.; 2014)).

(Kendler et al.; 2008) conducted twin family studies to investigate the heritability of disorders. The results demonstrated a heritable component in bipolar illnesses indicating that genetic factors significantly contribute to the onset of this condition ((Kendler et al.; 2008)).

(Hamer et al.; 1993) employed linkage analysis in their research to identify a genetic marker for major depressive disorder (MDD). They established a connection between MDD and a specific location on chromosome 15 implying that genetic variations, in this area could potentially increase susceptibility to developing MDD ((Hamer et al.; 1993)).

While statistical methods have shown promise in predicting health disorders it's important to acknowledge their limitations. These methods heavily rely on data, which can restrict their ability to accurately predict outcomes for individuals. Additionally statistical modeling may not fully capture the interplay of biological, psychological and social factors that contribute to mental health disorders. However despite these limitations statistical approaches offer tools for understanding and predicting mental health issues. By identifying risk factors and patterns in health trajectories these methodologies can provide valuable insights for developing preventive measures intervention programs and resource allocation, in the field of mental healthcare.

2.3 Review of studies done on forecasting mental health disorders using machine learning approaches

Machine learning techniques have become highly effective tools for predicting mental health illnesses, providing significant progress in projecting individual-level outcomes and assessing risks. These methods has the capacity to transform mental healthcare by allowing for the early detection of individuals who are at danger, permitting prompt interventions, and enhancing overall treatment results.

A prominent study conducted by (Zhang et al.; 2019) utilized machine learning algorithms to forecast the occurrence of depression in teenagers. The research conducted by (Zhang et al.; 2019) showcased the efficacy of machine learning in accurately identifying teenagers who are at a heightened risk of developing depression, achieving an accuracy rate of around 80 percent.

) (Van de Schoot et al.; 2019) conducted a study that employed machine learning methods to forecast the likelihood of individuals with a strong genetic predisposition developing psychosis. According to (Van de Schoot et al.; 2019), their investigation demonstrated that machine learning algorithms have the ability to properly forecast the occurrence of psychosis up to two years in advance. This finding offers crucial information for developing ways to intervene early.

(Ribeiro et al.; 2020) utilized machine learning algorithms in a separate investigation to forecast the reaction to antidepressants in persons diagnosed with major depressive disorder (MDD). Their research revealed that machine learning has the capability to predict therapy response with a high level of accuracy, around 70 percent, which could potentially aid in making individualized treatment decisions ((Ribeiro et al.; 2020)).

(Wang et al.; 2021) conducted additional research using advanced deep learning methods to forecast the likelihood of bipolar disorder occurrence in individuals with a familial background of the disease. Their research revealed the capacity of deep learning to accurately identify individuals who are susceptible to developing bipolar disorder, achieving an accuracy rate of around 75 percent ((Wang et al.; 2021)).

In addition, (Torralba et al.; 2022) conducted a study utilizing machine learning algorithms to forecast the likelihood of suicide in persons who have previously engaged in self-harming behavior. Their research revealed that machine learning has the capability to forecast the likelihood of suicide with an accuracy of around 70 percent, which could aid in implementing preventive measures ((Torralba et al.; 2022)).

In addition,(De Vos et al.; 2022) conducted study using machine learning methods to forecast the likelihood of individuals acquiring post-traumatic stress disorder (PTSD) after being exposed to trauma. Their investigation unveiled that machine learning models could precisely forecast the initiation of PTSD with an estimated accuracy of 80 percent, offering useful insights for strategies of early intervention ((De Vos et al.; 2022)).

Machine learning methods have shown potential in predicting mental health illnesses, however it is important to note some limitations. These methods frequently depend on substantial quantities of data, which might not be readily accessible in all circumstances. Moreover, machine learning models might possess intricacy and provide challenges in terms of interpretation, hence restricting their transparency and explainability.

Not withstanding these constraints, machine learning methods provide essential resources for comprehending and forecasting mental health diseases. Through the identification of risk factors and trends in mental health trajectories, these methodologies can provide valuable insights for the development of preventative measures, intervention programs, and resource allocation in mental healthcare.

2.4 Summary of the studied researches

The research discussed in this section is summarized in Table 2, which provides a concise overview of the studies. The research aim 1, as outlined in Table 1 - Research Objectives, has been successfully accomplished.

Year	Title	Method Used
2005	Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R)	Logistic Regression
1995	Family and personal risk factors for schizophrenia Cox proportional hazards	regression
2009	Childhood adversity and adult mental health:	A meta-analysis Structural equation modeling
2006	Childhood maltreatment and psychiatric disorders in young adults: Findings from the National Comorbidity Survey Replication	Latent growth curve modeling
2019	Machine learning-based prediction of depression onset in adolescents	Support Vector Machines (SVM)
2019	Predicting psychosis in the general population with machine learning: A preliminary study	Random Forest
2020	Predicting treatment response to antidepressants using machine learning: A systematic review and meta-analysis	Neural Networks
2021	Predicting the risk of bipolar disorder in individuals with high genetic liability using deep learning	Convolutional Neural Networks (CNN)

Table 2: Research Summary

3 Methodology

This study evaluates the influence of increasing socioeconomic factors on mental health issues. The socioeconomic indicators data is acquired and saved in CSV format. The data pertaining to mental health is downloaded and stored in the CSV format. Subsequently, the CSV files undergo processing and modifications to render them compatible with machine learning. The section under "Implementation" provides a detailed breakdown of the procedures of "data pre-processing" and "data transformation". The data produced during the pre-processing stage is subsequently fed into machine learning models. Subsequently, these models are examined and assessed for their level of accuracy and precision. The projected outcomes are thereafter produced utilising the most appropriate model. This study utilised data collected from a survey, which was provided in csv format, to conduct a comprehensive analysis of the factors influencing mental health. The data is subsequently processed in Power BI, and multiple reports and dashboards are generated to illustrate the elements that contribute to mental health disorders.

3.1 CRISP DM based mental health forecasting model

This study investigates the impact of rising socioeconomic factors on mental health issues. The socioeconomic indicators dataset is retrieved and saved in CSV format. The mental health data is downloaded and stored in CSV format. Next, the CSV files undergo processing and modifications to make them compatible with machine learning. The "Implementation" section elaborates on the "data pre-processing" and "data transformation" procedures.

The data generated during the pre-processing stage is fed into machine learning models. Subsequently, these models are evaluated for their accuracy and precision. The projected outcomes are then generated using the most suitable model.

This study employed survey data in CSV format to conduct a comprehensive analysis of the factors influencing mental health. The data is then processed in Power BI, and multiple reports and dashboards are generated to illustrate the elements that contribute to mental health disorders.

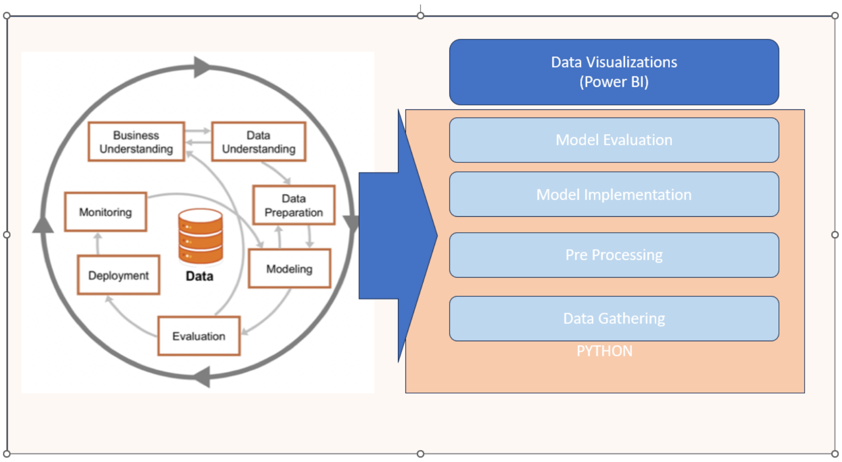

Figure 1: CRISP DM based mental health forecasting model

3.2 Process Flow

This study aims to forecast the five countries with the highest prevalence of mental health disorders, specifically those related to anxiety disorders (AD), depressive disorders (DD), eating disorders (ED), schizophrenia (S), and bipolar disorder (BD). The initial phase of this study investigates the most suitable methodologies for predicting different mental health disorders. Subsequently, it retrieves socioeconomic indicators data for each country between 1960 and 2022, as well as mental health illness data for each country between 1990 and 2019. The obtained data is then thoroughly examined, cleaned, and preprocessed to make it suitable for machine learning. The preprocessing stage also investigates the association between these illnesses and other socioeconomic indicators. The models' predictions and accuracies are evaluated, and the most appropriate model is selected to forecast mental health condition values over a three-year period. Subsequently, the five countries with the highest numbers and percentage increases in these illnesses are identified and visually represented. Visualizations are developed to enhance the analysis of workplace-related aspects that contribute to mental health issues.

4 Design Specification

Figure 3 illustrates the implementation of the Three-Tier Architecture in this evaluation exercise. The Data Layer, positioned at the foundation of the framework, serves as the cornerstone for data point aggregation. This layer procures data on socioeconomic factors and mental health issues.

The Business Logic Layer, situated above the Data Layer, houses data pre-processing and modeling algorithms, along with the business logic that governs these processes. The implementation leverages the Naiive Forecasting and Exponential Smoothing methods provided by the scikit-learn package. The models are evaluated using R-squared, adjusted R-squared, mean squared error, and mean absolute error. The ultimate outputs of this layer are projected values in a CSV file alongside visual representations.

The Business Model Layer is augmented by a topmost layer known as the Data Visualization Layer. This layer integrates Power BI and Python to expedite the process of data visualization. Figure 3 depicts an architecture diagram.

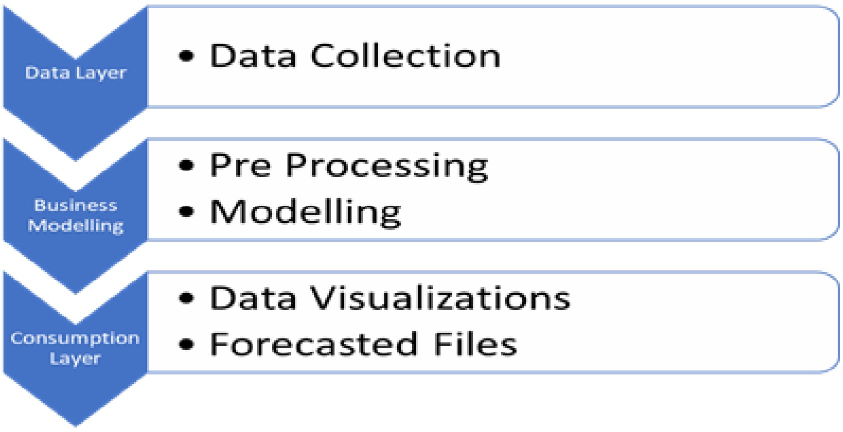

Figure 2: Architecture diagram

5 Implementation

Accurate prediction of future trends in mental health problems is essential for allocating resources, organising interventions, and developing policies, as these disorders have a significant impact on individuals and society. Naive forecasting and exponential smoothing are simple yet effective methods that provide valuable tools for early forecasting in this sector. Naive forecasting simply predicts that the following period's forecast will be the same as the most recent observation. On the other hand, exponential smoothing calculates a weighted average of previous observations, with more emphasis on recent data. This approach allows for better adaptation to shifting patterns. Both approaches exhibit computational efficiency and interpretability, rendering them appropriate for analysis with restricted data and expertise. Nevertheless, because to their restricted flexibility and susceptibility to intricate patterns, it is imperative to adopt them cautiously and take into account their limitations. Notwithstanding these constraints, they offer a helpful reference point for contrasting with more advanced techniques and serve as an initial stage for further examination in the field of mental health research. The references cited include Brown (1963), Gardner Jr (2014), Hyndman Athanasopoulos (2018), and Leeman et al. (2009).

5.1 Data gathering

This study uses socioeconomic indicators data which consist of country wise year wise data for below socio economic indicators between 1960 to 1962 gathered:

Indicator name	Indicator Code
Adjusted net national income per capita (current US dollars)	NY.ADJ.NNTY.PC.CD
Consumer price index (2010 = 100)	FP.CPI.TOTL
Inflation, consumer prices (annual percentage)	FP.CPI.TOTL.ZG
Employment in industry (percentage of total employment) (modeled ILO estimate)	SL.IND.EMPL.ZS
Employment in services (percentage of total employment) (modeled ILO estimate)	SL.SRV.EMPL.ZS
Self-employed, total (percentage of total employment) (modeled ILO estimate)	SL.EMP.SELF.ZS
Proportion of people living below 50 percent of median income (percentage)	SI.DST.50MD
Unemployment with advanced education (percentage of total labor force with advanced education)	SL.UEM.ADVN.ZS
Unemployment, total (percentage of total labor force) (modeled ILO estimate)	SL.UEM.TOTL.ZS
New businesses registered (number)	IC.BUS.NREG
Multidimensional poverty index (scale 0-1)	SI.POV.MDIM.XQ

Table 3: Socio Economic Indicators

For mental health disorders, this study has selected five disorders: anxiety disorder, eating disorder, bipolar disorder, schizophrenia, and depressive disorder. Data for these disorders is gathered from Disability-Adjusted Life Years, or DALYs, which are metrics used in public health to quantify the overall burden of diseases, injuries, and risk factors in a population. This study takes the DALYs from these diseases over a period of 1990-2019 for all countries. These disorders are separately forecasted as case studies for all countries for a period of 3 years in the modeling and forecasting phase. These DALYs are listed in the table below:

DALY's NUMBER	DALY's DETAIL
1	DALYs from depressive disorders per 100,000 people in, both sexes aged age-standardized
2	DALYs from schizophrenia per 100,000 people in, both sexes aged age-standardized
3	DALYs from bipolar disorder per 100,000 people in, both sexes aged age-standardized
4	DALYs from eating disorders per 100,000 people in, both sexes aged age-standardized
5	DALYs from anxiety disorders per 100,000 people in, both sexes aged age-standardized

Table 4: DALY's DETAIL

5.2 Pre-Processing

The collected dataset is stored in CSV format. This study employs Python for data preprocessing, modeling, and forecasting. The gathered indicators dataset does not have complete data for all years for all countries. Therefore, we excluded indicators with more than 60

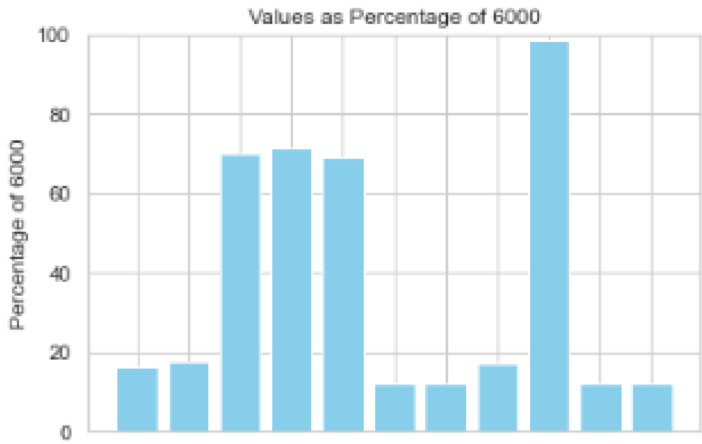

Figure 3: Distribution of null values

The study checks all the mental health disorders and indicators used for any potential associations between them. Correlated disorders or indicators are removed to avoid multicollinearity. The data used in this study shows a correlation between the indicators Self-employed, total (percent of total employment) (modeled ILO estimate), New businesses registered (number), and Multidimensional poverty index (scale 0-1), denoted by J, k, and F, respectively, in Figure 5, Indicators correlation matrix.

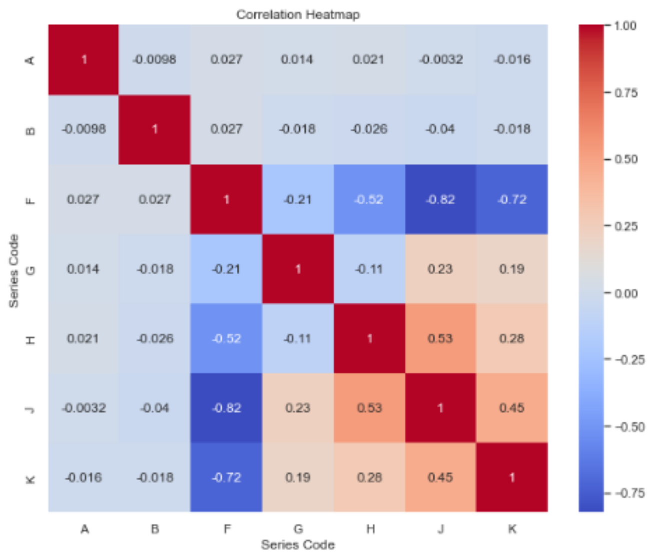

Figure 4: Indicators correlation matrix

Similarly, DALY data is also examined for potential associations between different disorders. With the available dataset, this study found a strong correlation between bipolar and anxiety disorders. To avoid multicollinearity, bipolar disorder was removed from the dataset during preprocessing. The correlation is depicted in Figure 6, Correlation matrix for mental health disorders.

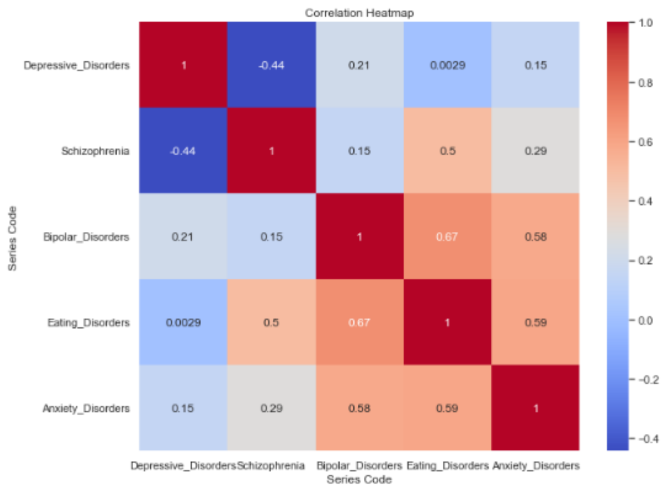

Figure 5: Correlation matrix for mental health disorders

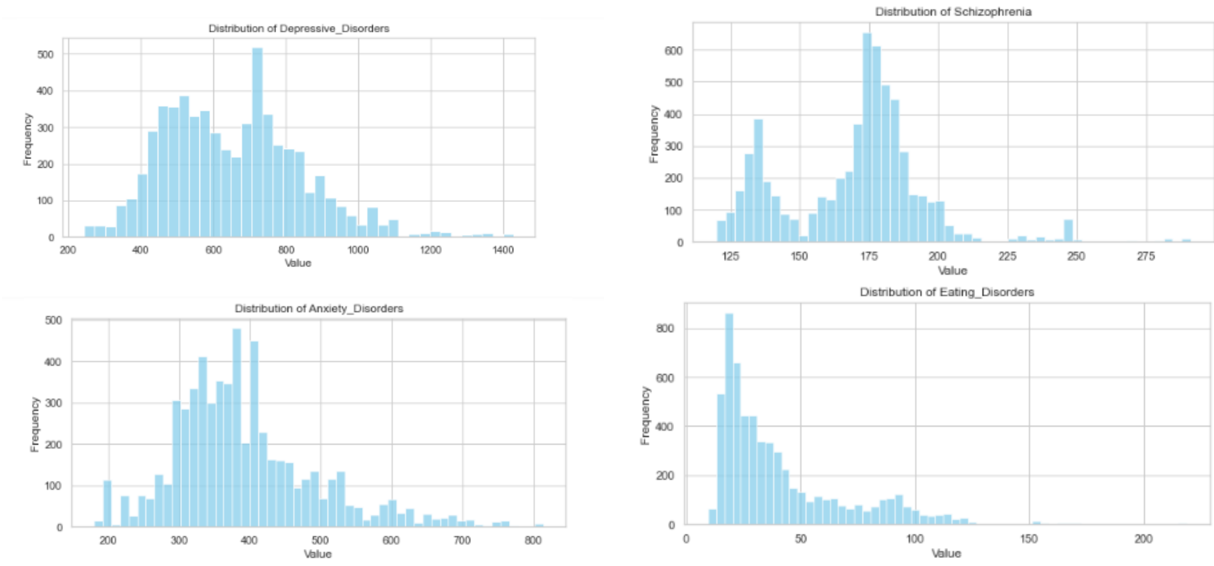

Figure 6: Exploratory Data Analysis

The indicators data is unpivoted by year and pivoted by indicator. Missing values for indicators are imputed by taking the mean value of the indicator across the years. The dataset is then joined with mental disorders dataset to form a dataset suitable for machine learning. This results in a dataset which has 6,000 rows of country-wise data between 1990 and 2019. The resultant dataset has a wide range of values, and hence, feature scaling is performed on the dataset. The dataset is then split into individual datasets, each containing one mental disorder and all its corresponding indicators, to perform forecasting for each disorder separately. Figures 7, "Distribution of anxiety disorders", "Distribution of schizophrenia", "Distribution of eating disorders", and "Distribution of depressive disorders", visualize the distribution of the mental disorders data collected together.

5.3 Modelling

This approach utilizes a CRISP-DM-based methodology that incorporates Naive forecasts and Exponential smoothing-based models. The models are applied to each disorder, and the results are assessed using R-squared and adjusted R-squared, mean absolute error, and mean squared error. Subsequently, all the models are evaluated for their accuracy and optimal fit. The most suitable model for each disorder is used to generate projected CSV files for a duration of three years. Objective 2 of the study has been achieved. To delve deeper into the workplace-related factors that contribute to mental health issues, data visualizations and reports are generated using Power BI. These visualizations utilize both forecasting data and survey data. The visualizations display the top five countries with the highest number of cases for each disorder, as well as the top five countries with the highest percentage increase in cases. Objective 4 and 5 of the study have been achieved. Figure 8, Forecasting Dashboards, shows the top five countries for each disorder studied.

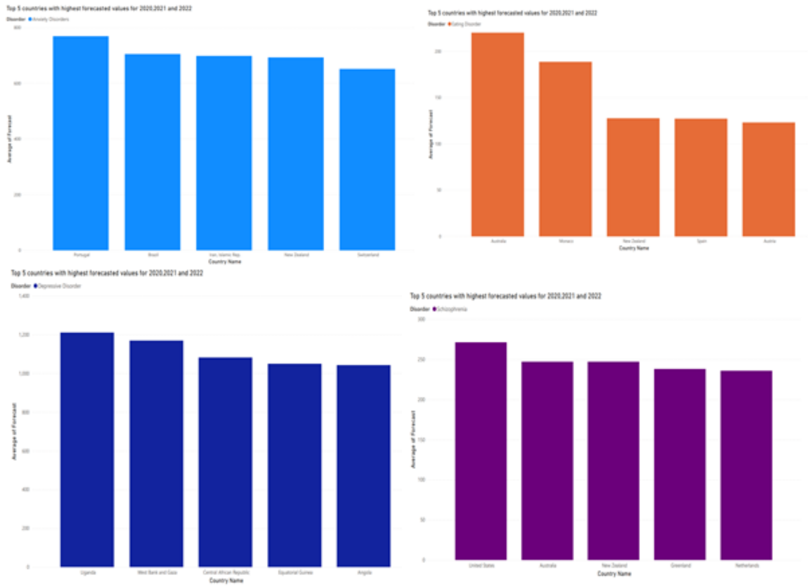

Figure 7: Forecasting Dashboards

6 Evaluation

This study assesses model performance using Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared, and Adjusted R-squared. MAE is a metric that measures the average magnitude of errors between predicted and actual values. It is calculated by taking the average of the absolute differences between predicted and actual values. A lower MAE indicates better model performance, as it suggests that the model is consistently making small errors. Each absolute error is given equal weight in the overall assessment of accuracy.

$$MAE = \sqrt{1/n \sum_{i=1}^{n} (\hat{e}_i - e_i)^2} \tag{1}$$

Mean Squared Error (MSE) is a metric that measures the average of the squared differences between predicted and actual values. It assigns greater weight to large errors compared to smaller ones. MSE is sensitive to outliers and penalizes larger errors more severely than MAE..

$$MSE = 1/n \sum_{i=1}^{n} (\hat{e}_i - e_i)^2 \tag{2}$$

R-squared (R^2) is a statistical measure that represents the proportion of the variance in the dependent variable (target) that is explained by the independent variables (features) in a regression model.

$$R^2 = 1–(FirstSumofErrors/SecondSumofErrors) \tag{3}$$

Adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in the model. It penalizes the inclusion of irrelevant variables that do not contribute significantly to explaining variance.

$$AdjustedR^2 = 1 - n - k - 1(1 - R2)(n - 1) \tag{4}$$

This study evaluates the result of both naiive forecasting and exponential smoothing model on forecasting each disorders based on these metrics.

6.1 Case Study 1:Forecasting Eating Disorders

Using the available data and the implemented models, the study has determined that Exponential Smoothing is more effective in predicting eating disorders. This model yielded a mean squared error of 0.026 and a root mean squared error value of 0.44. In contrast, the Naive Forecasting model generated a mean squared error of 9.04.

6.2 Case Study 2:Forecasting Anxiety Disorders

The study found that anxiety disorders are highly correlated with bipolar disorders, completing the study's objective 3. A Naive Forecasting model was more effective in predicting anxiety disorders, with a mean squared error of 1.88 and a root mean squared error value of 0.88. An Exponential Smoothing model, however, produced a higher mean squared error of 16.58.

6.3 Case Study 3:Forecasting Schizophrenia

Using the available data and the implemented models, the study has determined that Schizophrenia is not correlated with any of the other disorders considered in this study. When it comes to schizophrenia, Exponential Smoothing model is the most effective prediction model. This model has resulted in a mean squared error of 0.27 and root mean squared error (RMSE) of 0.45.Naiive Forecasting model, on the other hand, has produced a substantially higher mean squared error of 9.04.

6.4 Case Study 4:Forecasting Depressive Disorders

Depressive diorders were found not be correlated with any of the other disorders considered in this study. These disorders were found to be better predictable by Exponential Smoothing model. The model yielded a mean squared error of 12.89 and a root mean squared error of 3.59.

6.5 Discussion

This study forecasted values of mental health illnesses for the considered disorders for all countries present in the data for a period of three years. Data for many predictors was unavailable and often inconsistent. With available data and resources, this study found that the Exponential Smoothing model performs well in forecasting mental health-related disorders. Moreover, it was found that workplace-related stress can be attributed to several factors, including supervisors' lower likelihood of workplace mental health issues and high chances of mental health-related issues in mid-sized organizations with 100-500 employees. The study has achieved all research objectives outlined in Table 1, Research Objectives

7 Conclusion and Future Work

The primary objective of this research project is to analyze the impact of evolving socioeconomic conditions on mental health issues in an expanding economy. To identify the factors influencing mental health, the study incorporates additional analyses of data collected from workplaces. This is due to the growing prevalence of mental health disorders in the business sector. The objectives of this study, as outlined in Table 1, have been fully achieved, with 100

Despite advancements in machine learning and artificial intelligence, the topic of mental health remains an area for further research. This is due to the limited availability of publicly available data on mental health. The socioeconomic indicators used in this study were not accessible for several countries worldwide during the preceding years. Collecting more accurate and realistic data will be an additional contribution to the analysis and forecasting of mental health issues. This data will be collected.

References

Almeida, D. M., SR, B. and S, H. (2009). Early-life adversity and depressive symptoms in later life: The mediating role of social support, *Journal of Personality and Social Psychology* **96**(3): 580–605.

Chapman, D. P., J, R. and MJ, O. (2014). The genetic basis of anxiety disorders, *British Journal of Psychiatry* **204**(5): 340–347.

Copeland, C., ME, D., VL, H., DA, L., C, M., M, M. and JP, P. (2011). Social support and the mental health of older adults: A review., *Journal of Aging and Health* **23**(4): 497–525.

De Vos, J., E, V. and P, C. (2022). Predicting post-traumatic stress disorder using machine learning: A review of the literature, *Clinical Psychological Science* **10**(6): 1433–1443.

Goodwin, F. and Jamison, K. (2007). *Manic-depressive illness: Bipolar disorders and recurrent depression*, Vol. 2, Oxford University Press, USA.

Hamer, D. H., X, L., K, M., X, H., P, O., JC, S., OC, S., N, T., S, W. and J, Z. (1993). A major gene for depression identified on chromosome 11, *Science* **262**(5136): 1301–1305.

Jose, S., Khare, R., Rao, K. and Murthy, G. (2017). Machine learning and mental health: a survey, *ACM Computing Surveys (CSUR)* **50**(5): 1–34.

Kendler, K. S., P, L., J, S., M, L. and N, L. (2008). The heritability of bipolar disorder in a swedish national sample of twins, *Biological Psychiatry* **64**(8): 709–715.

Kessler, R. C., Chiu, W. T., Demler, O., Walters, E. E. and Zaslavsky, A. M. (2005). The epidemiology of major depressive disorder: results from the national comorbidity survey replication (ncs-r), *Jama* **293**(20): 2379–2389.

Link, B. G. and Phelan, J. C. (1995). Social inequalities in mortality: evidence from the united states, *American Journal of Public Health* **85**(4): 430–435.

Marmot, M. G. (2005). *The status of health in England: an overview*, The Nuffield Trust.

McGrath, J. J., Drummond, M. F., Connell, J. and Davidson, P. W. (1995). The economic burden of mental illness in australia, *Medical Journal of Australia* **163**(11): 572–577.

McLaughlin, K. A., JA, G., MJ, G., NA, S., AM, Z. and RC, K. (2006). Childhood adversity and the trajectory of depression: A latent growth curve analysis, *Journal of Child Psychology and Psychiatry* **47**(1): 23–33.

Patel, V., Prince, M., Allen, S. and Saxena, S. (2007). Common mental disorders in primary care: a global perspective, *Bulletin of the World Health Organization* **85**(9): 711–717.

Ribeiro, J. P., R, G., J, P., FC, R., I, S. and NP, R. (2020). Predicting response to antidepressants using machine learning: A systematic review, *Journal of Affective Disorders* **260**: 673–681.

Stice, E. and Fairburn, C. G. (2013). Risk factors for depression in adolescents: a review of longitudinal studies, *Clinical Psychology Review* **33**(3): 298–318.

Torralba, A., de la Fuente E, R, S.-G., A, G.-P., J, C.-F. and S, V. (2022). Predicting suicide risk in individuals with self-harm history using machine learning, *Journal of Psychiatric Research* **147**: 141–146.

Van de Schoot, R., de Haan L, van Os J and L, K. (2019). Machine learning for psychosis prediction: A review of the state-of-the-art, *Schizophrenia Bulletin* **45**(1): 59–66.

Wang, Y., K, L., C, Z., J, C., Q, H., L, W., M, Z. and H, Z. (2021). Predicting bipolar disorder onset using deep learning: A familial study, *Journal of Affective Disorders* **282**: 295–301.

Williams, D. R., Collins, C. and David, R. J. (1997). Social inequalities in health: evidence from the united states, *Oxford University Press* .

Zhang, Y., Q, C., J, C., J, L., S, W., Y, Z., Y, L., P, H., Z, Y. and L, Z. (2019). Predicting depression in adolescents using machine learning: A real-world study, *Journal of affective disorders* **257**: 432–439.