School of Engineering & Technology

Asian Institute of Technology

AT84.02: Business Intelligence and Analytics

Date: 18/04/2023

Analysis and Prediction System for Global Mental Health Disorder

Submitted To:
Dr. Vatcharaporn Esichaikul

Submitted By:
Ms. Lam Thi My Duyen - st123346
Mr. Adesh Shrestha - st123197
Mr. Nguyen Thai Anh - st122910
Mr. Nguyen Hong Duc - st122934

TABLE OF CONTENT

I. BACKGROUND

Mental health is an integral part of our general health and well-being and a basic human right. Having good mental health means we are better able to connect, function, cope and thrive. However, the faster the development of society develops, the more people suffer from mental disorders. Furthermore, in spite of the increasing proportion of people with mental issues, mental health services seem to have not spent worthy care for those issues, which leads to the gap between patients and treatment resources and condition (figure 1). Considerably, the covid-19 pandemic has created a global crisis for mental health. In 2019, specifically, 301 million people globally were living with anxiety disorders; and 280 million were living with depressive disorders (including both major depressive disorder and dysthymia). In 2020, these numbers rose significantly at approximately 26% and 28% respectively as a result of the COVID-19 pandemic. As a result, mental health services have been severely disrupted and the treatment gap for mental health conditions has widened.

These challenges make authorities must put all their efforts into it with the aim of enhancing the quality of citizens' mental health. Thus, our project focuses on providing authorities with history-related data on health disorders and also making predictions to assist governments and relevant organizations in tackling increasing mental health disorders.

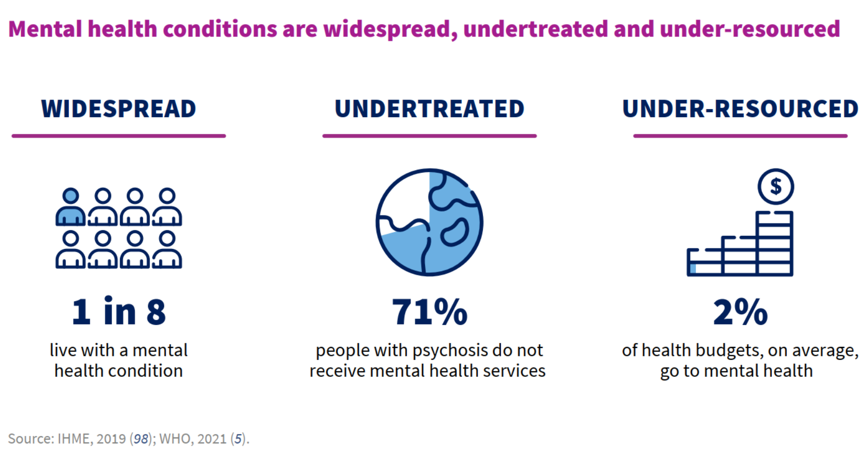


Figure 1: Mental health conditions are widespread, undertreated and under-resourced

II. OBJECTIVE

The principal objective of this project is to provide several reports and a dashboard to our end-users regarding the information about Mental health disorder among adults across various countries. Since our selected subject matter debates chiefly about mental health disorder, the dashboard will display information regarding the major types of disorders which can be found globally. Moreover, another scope of this project is also to provide actionable predictions and information to help users know about the major types of mental health disorder and predictions regarding them over the coming years. Users will be able to obtain reports about their countries' medical health information which will be very valuable for them.

III. DESCRIPTION

Datasets

To effectively analyze the impact of the mental health issue, it is important to have access to a consistent, worldwide and regularly updated dataset. For this reason, we selected Mental Health from Our World in Data, which provides worldwide information on the prevalence of various mental and substance use disorders, depression among adults across different countries. Furthermore, this dataset is open-source, which is freely available to the public.

For more detail, the data is sourced from *the Institute of Health Metrics and Evaluation (IHME), Global Burden of Disease (GBD), World Health Organization (WHO) International Classification of Diseases (ICD), World Health Organization (WHO) Global Health Observatory (GHO)*, and this data covers the years 1990 to 2019. The dataset provides information on the estimated prevalence of different types of mental and substance use disorders, depression among adults, who are male and female at averaged-age, including major depressive episodes, serious mental illness, and substance use disorders. It appears that the dataset is well-structured and includes clear definitions for each of the variables. However, it is always a good idea to perform additional research and data cleaning as necessary to make sure the data is accurate and suitable for your intended use. For instance, in some columns, there is missing data, which is related to the rate of some disorder types corresponding to each year. Overall, this dataset may be useful for researchers, data scientists, and others interested in exploring the prevalence of mental and substance use disorders as well as depression's impact across the nations.

Project Scope

Who we are

We are a Non-Profit Organization dedicated to analyzing the situation of mental health issues and providing predictions based on our findings. Our goal is to help communities and governments make informed decisions about the issues and its impact on society. Through our rigorous research and analysis, we aim to provide up-to-date information and insights on different types of mental and substance use disorders, their impact on public health, and the exploration of measures taken to neutralize their effects. By leveraging cutting-edge data analytics and machine learning techniques, we strive to provide accurate and reliable predictions that can help inform public policy and aid in the response to the issues.

Our end users

Our target end users are the Ministries of Health in related countries, who are tasked with managing and controlling mental health issues. Our aim is to support these organizations in their efforts to address the mental health challenges posed by the various mental and substance use disorders. By providing them with accurate, timely and actionable information, we hope to assist them in making informed decisions and implementing effective strategies to address the mental health impact of these issues. We believe that by working together with these organizations, we can make a positive advance on the lives of people affected by the different mental issues and help mitigate these consequences of this global crisis.

IV. EXPECTED OUTPUT

The expected output for this project is a Decision Support System which will present information and analytics about the global mental health disorder from the year 1990 to 2019. Moreover, this system will be able to present relevant information about several mental health disorders such as schizophrenia, bipolar disorder, eating disorder, drug use, depressive disorder, anxiety disorder and so forth. Additionally, the system will also present relevant information based on countries which will aid the users and researchers to filter and analyze statistics by using selected key features. Furthermore, this system will also predict possible number of cases regarding mortality rates and disorder rates based on the collected dataset. The sample screenshot attached below is an example of how the system would look once it has been fully constructed and developed.

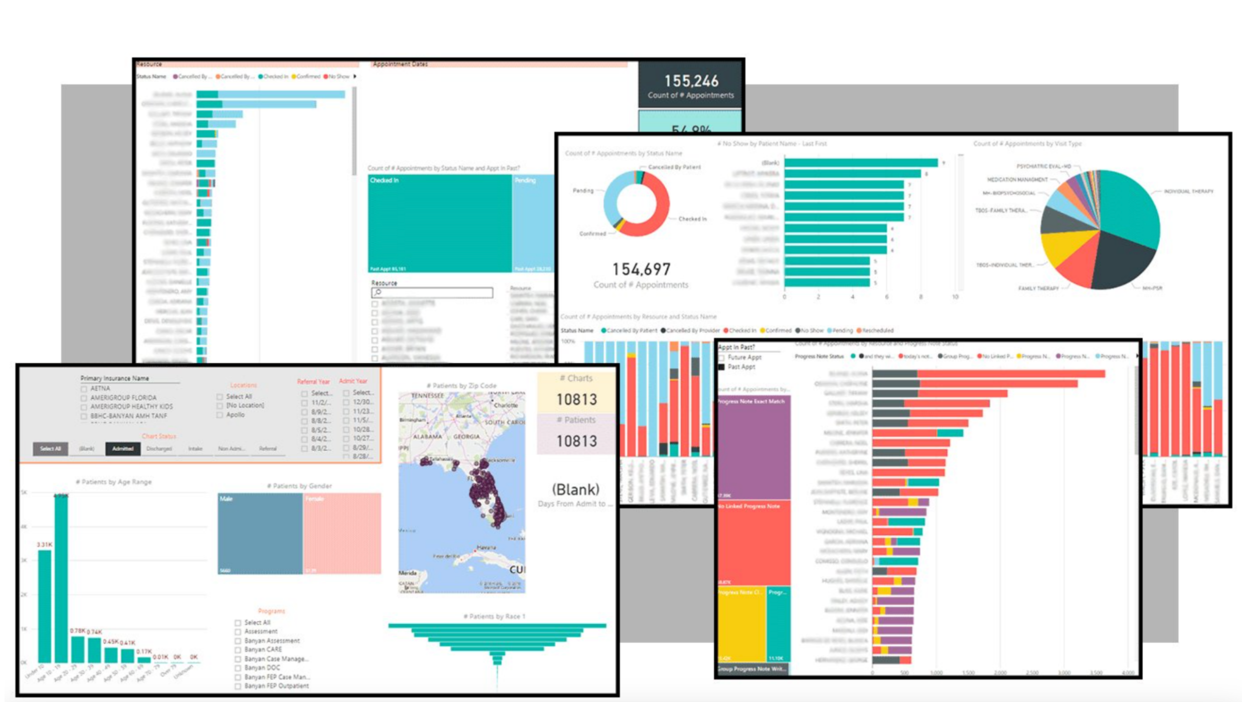
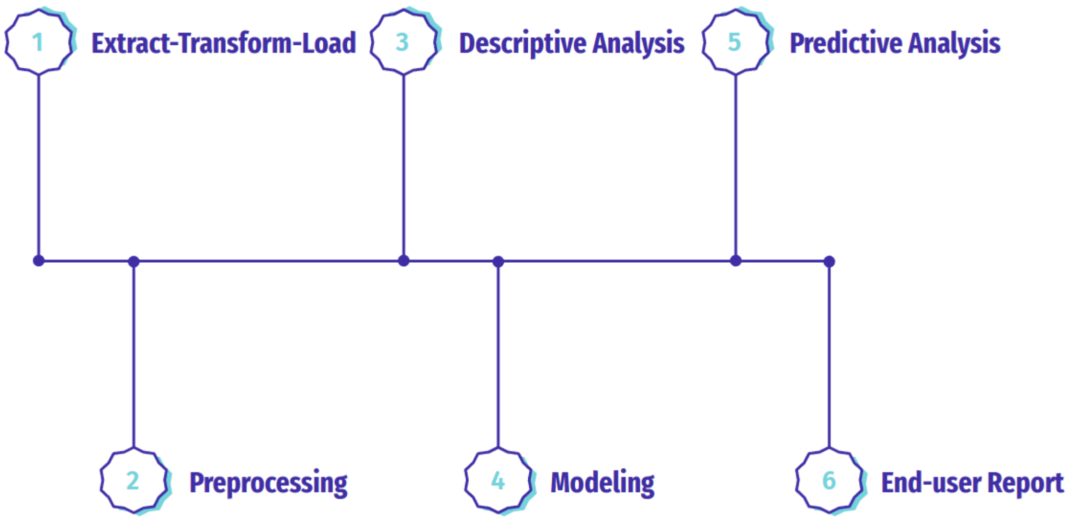


Figure 2: Sample of Business Intelligence/Decision Support System

V. ANALYTIC PIPELINE

The analytic pipeline for our entire project is shown as below. This process includes six major steps as follows:



1. Extract-Transform-Load (ETL): In this first step, data is extracted from Mental Health, we used in total three datasets: number of people with mental health and substance use disorders, deaths from mental health and substance use disorders, and suicide rate from mental health and substance use disorders. Then they are transformed into a structured format suitable for analysis, and successfully loaded to prepare for the following steps. The goal of ETL is to ensure that the data is accurate, complete, and consistent.
2. Preprocessing: During the preprocessing stage, the data is processed to make it ready for analysis. This involves a range of tasks such as handling missing values, managing outliers, normalizing or scaling the data, and adding related variables or features (population) that can provide further insights into our analytics as well as data reduction. Additionally, we restrict our geographical scope to ASEAN countries, particularly Thailand, Vietnam, Laos, Cambodia, and Myanmar in the Predictive Analysis session. Preprocessing ensures that the data set is optimized for analysis.

3. Descriptive Analysis: During this stage, we examine and visualize the data to uncover insights and detect patterns. Descriptive analytics aids in comprehending the data and its features, such as the value distribution, the connections between variables, and trends over time. This phase may include generating histograms, scatterplots, and other visualized dashboards which use Tableau along with Power BI to discover patterns and correlations in the datasets.
4. Modeling: In the modeling phase, we construct machine learning (ML) models to forecast future results or detect patterns in the data. This includes selecting appropriate models, training them on the data, and evaluating their performance by test-loss. Then, we compared multiple models before selecting the most effective one. More detail, in our case, we picked two models: Linear Regression and Long-Short-Term-Memory (LSTM). Further overview of these two models are provided in the Predictive Analytics session.
5. Predictive Analysis: In this stage, the chosen models are used to make predictions based on the datasets. The predictions are used for various purposes which depend on each dataset, such as identifying trends in the number of people who suffer from mental health and substance use disorders, forecasting mortality rates, and uncovering self-harm rates for the upcoming years until 2023.
6. End-user report: In this section, we need to consider both the design and delivery of the report. Initially, we create a report or dashboard that presents the analysis results to the end user. This process involves choosing appropriate visual aids and summarizing the main findings. Subsequently, we must provide the end user with a format that is simple to understand and act upon, emphasizing the importance of creating interactive dashboards. Furthermore, providing end-users with access to the raw data for further analysis along with suggesting recommendations based on key findings are critical aspects of this project. Finally, the report may require regular updates to reflect changing data or the organization's requirements.

VI. DESCRIPTIVE ANALYSIS

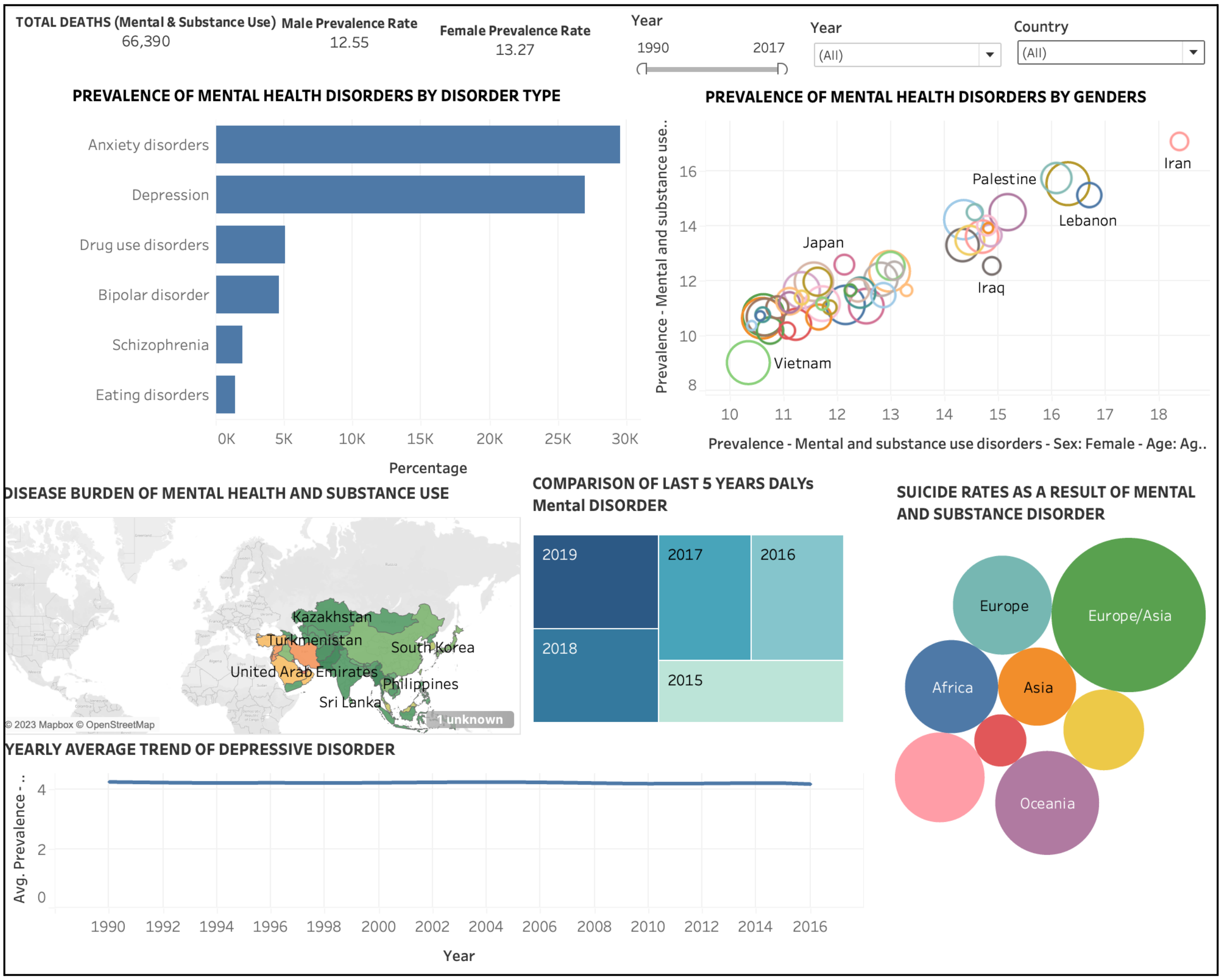


Figure 3: Global Mental Health & Substance Use Disorder Dashboard

The picture above illustrates the overall Mental Health Disorder dashboard created using Tableau software to perform descriptive analysis. As represented in the picture, several levels of analysis have been implemented to better understand and describe the data cited by **Our World in Data** and estimated by the Institute for Health Metrics and Evaluation. The generated dashboard provides analytic reports regarding different disorders across countries among adults (male, female).

The dataset contains information regarding various mental health disorders such as anxiety disorder, depression, drug use, bipolar disorder and so forth based on countries spread across continents focused towards prevalence rate for both male and female genders. Moreover, the produced dashboard presents information regarding total deaths as a result of mental health disorders and substance use along with male and female prevalence rates. Additionally, a filter option has been added to the dashboard which allows users to view various levels of analyzed data based on years and countries.

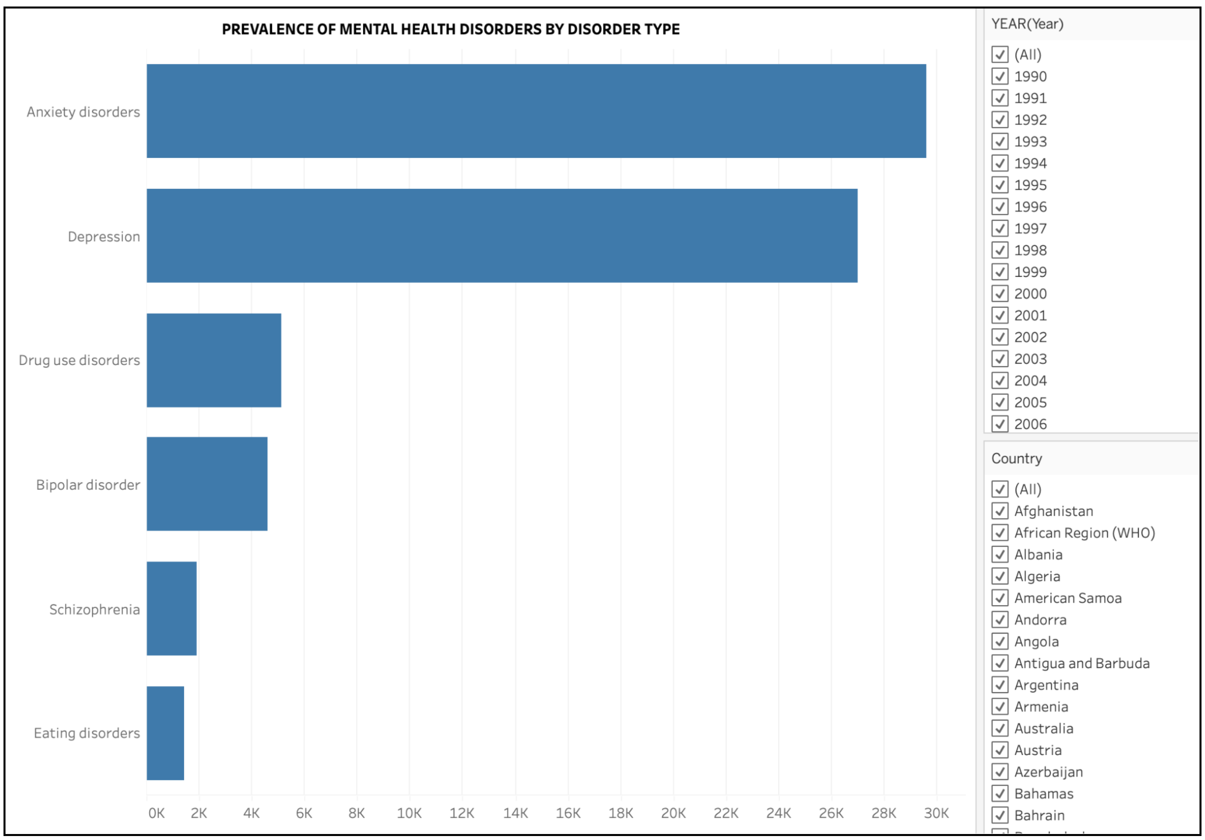


Figure 4: Disorder Type for Mental Health Disorder Prevalence

The bar graph visualization in **Figure 4** represents prevalence of mental health disorders based on the disorders type -

- Anxiety disorders
- Depression
- Drug use disorders

- Bipolar disorders
- Schizophrenia
- Eating disorders

Among the disorders listed above, anxiety disorder is generally found to be the most widespread across most of the countries from the years 1990 to 2019 followed by depression. Meanwhile, eating disorders is the least occuring disorder. Drug use disorders, bipolar disorders and schizophrenia are on average out of the mentioned disorders in the given dataset.

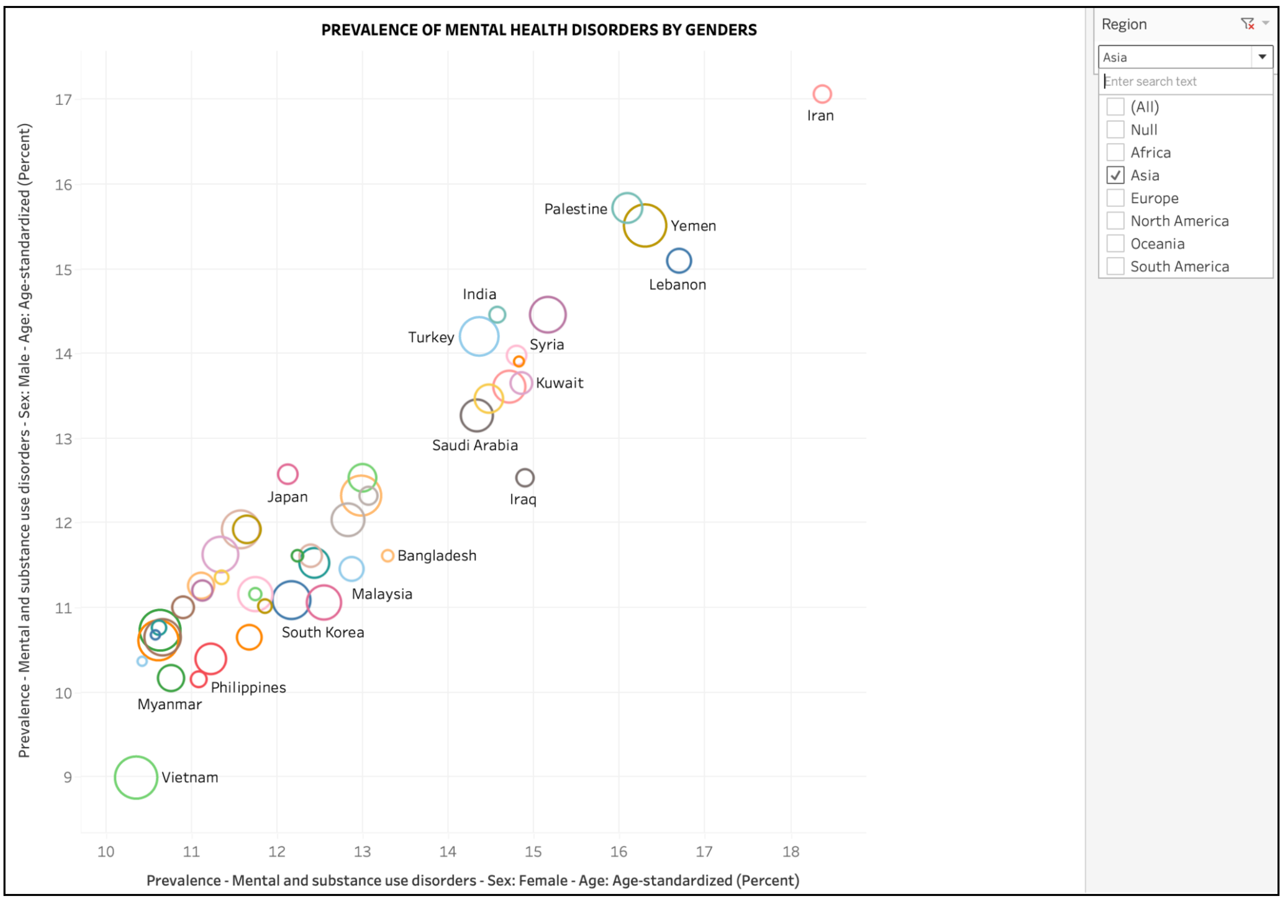


Figure 5: Prevalence of Mental Health Disorder based on Gender

The visualization represented in **Figure 5** shows the prevalence of mental health disorder based on genders across Asia out of which SouthEast Asian and Middle East countries such as South Korea, Saudi Arabia, Turkey have the highest prevalence rates.

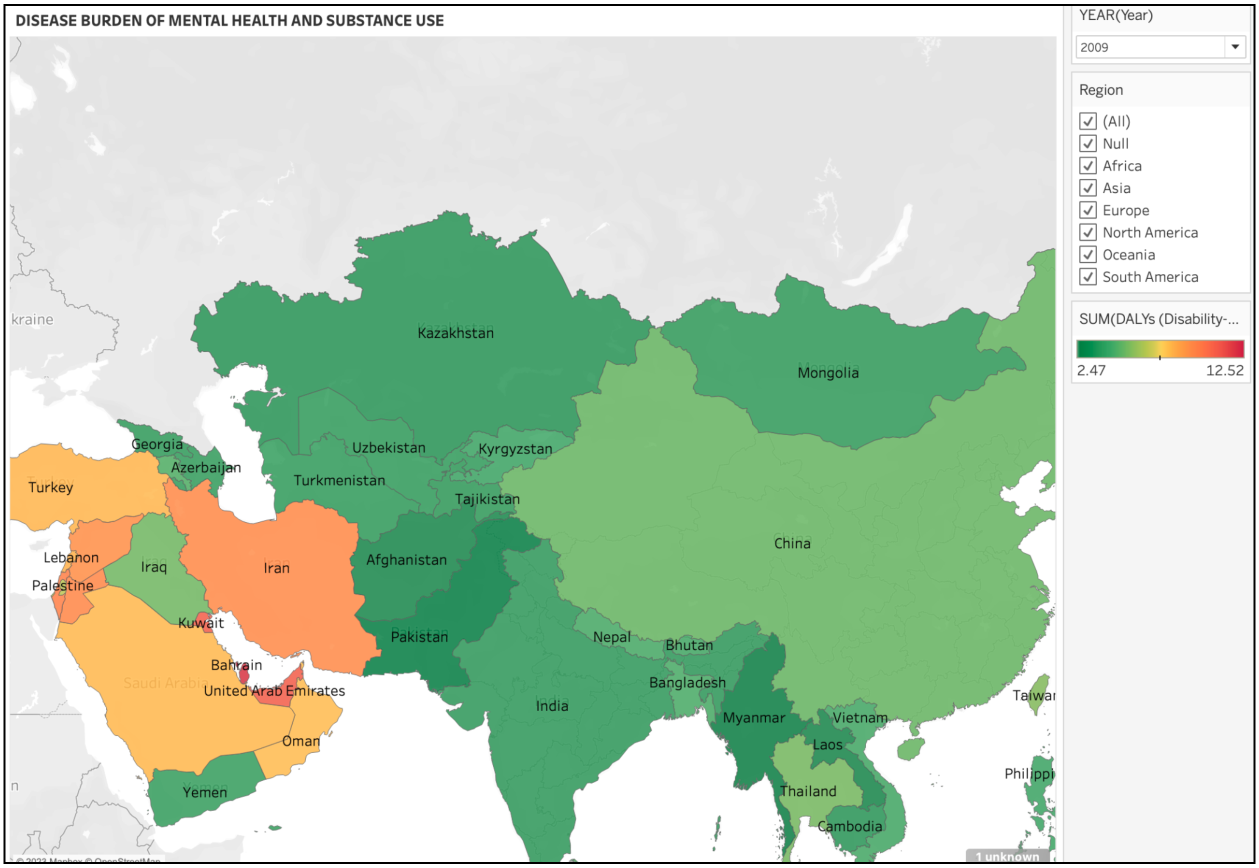


Figure 6: Disease Burden of Mental Health and Substance Use

The geographical map in **Figure 6** shows the disease burden of mental health and substance use across continents where Middle East countries such as Iran, Turkey, United Arab Emirates, Lebanon etc. have the most disease burdens due to Disability Adjusted Life Years disorder.

Figure 7 shows the line chart of the yearly average trend of depressive disorder from the year 1990 which depicts that the average prevalence rate is almost constant throughout the years although the rate was found to be decreased during the last few years.

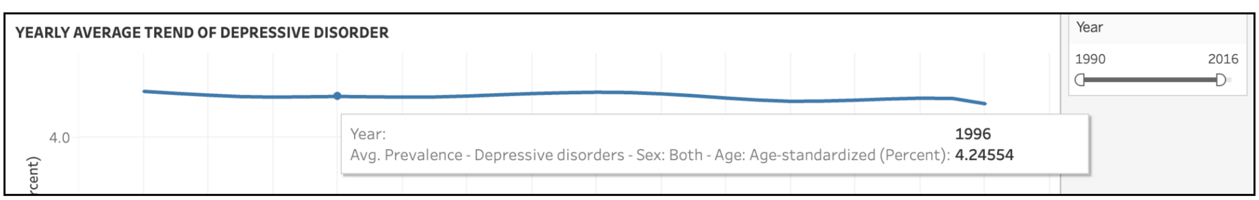


Figure 7: Yearly Average Trend of Depressive Disorder

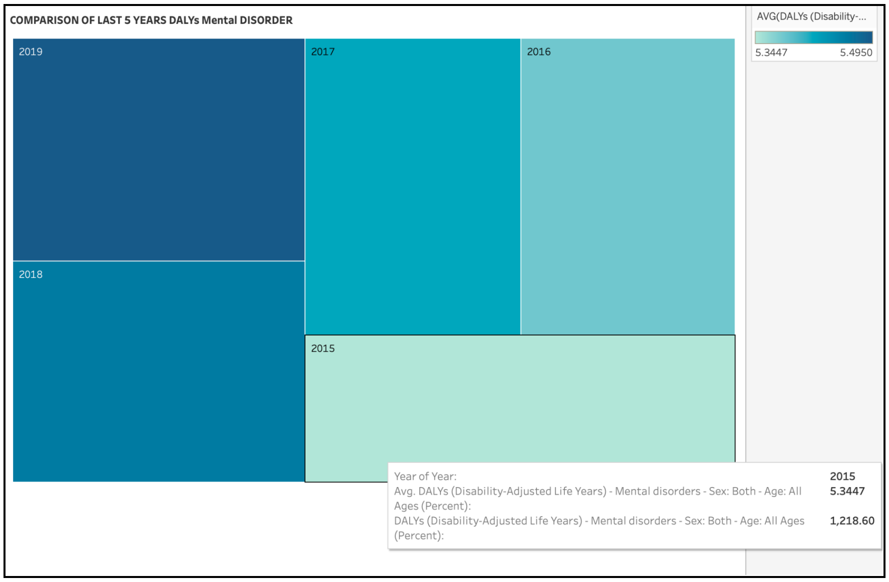


Figure 7: Comparison of Disability Adjusted Life Years Mental Disorder

The treemap in **Figure 7** shows a 5 year comparison of Disability Adjusted Life Years (DALY) Mental Disorder between the years 2015 to 2019 where it is visible that the DALY rate is increasing year after year.

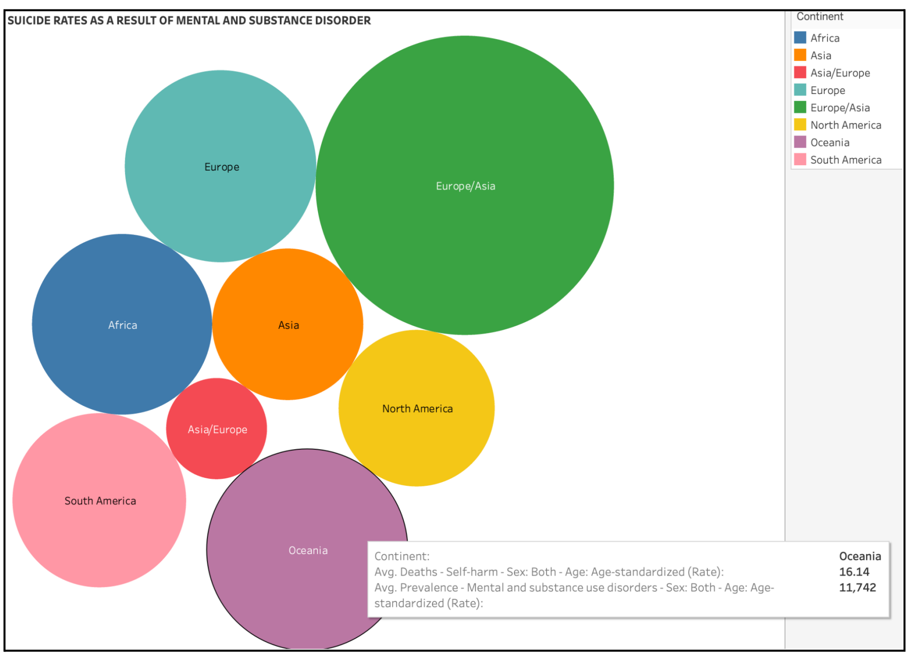


Figure 8: Suicide & Death Rates across Continents

The bubble map in Figure 8 presents suicide and death rates due to mental health disorders across continents. Asian and European countries are prone to suicides as a result of mental health disorders starting from the years 1990.

VII. PREDICTIVE ANALYSIS

Quick overview of two chosen models are described as below:

- Linear regression is a simple and widely used technique that models the relationship between a dependent variable and one or more independent variables. In the case of time series analysis, the dependent variable is usually the variable of interest that is being forecasted over time, and the independent variable(s) are usually the historical values of the dependent variable and possibly other relevant variables. Linear regression assumes a linear relationship between the variables, and it uses historical data to forecast future values. Linear regression is easy to interpret and computationally fast, making it a good choice for short-term forecasting. However, it may not be accurate if the relationship between the variables is not linear or if there are outliers in the data.
- Long-Short-term-Memory (LSTM) is a type of neural network that is specifically designed for processing sequences of data, such as time series. LSTM is able to learn long-term dependencies in the data and is particularly effective when there are complex patterns in the data. LSTM uses a combination of cell states and gates to selectively remember or forget information from previous time steps, allowing it to model complex temporal relationships in the data. LSTM is able to handle non-linear relationships between the variables and is less sensitive to outliers than linear regression. However, LSTM is more complex and computationally expensive than linear regression, and it may require more data to train effectively.

To begin, we identify a number of potential key factors that can be used for forecasting. Next, we narrow our focus to ASEAN countries, specifically Thailand, Vietnam, Laos, Cambodia, and Myanmar. We create a new DataFrame by selecting the columns 'Vietnam', 'Thailand', 'Laos', 'Cambodia', 'Myanmar', and 'Year', which contain information on the number of cases of mental health and substance use disorders in these selected countries over a certain time period.

Subsequently, we divide our DataFrame into two subsets for the purpose of training and validating our LSTM and Linear Regression models. The training set consists of information on mental health and substance use disorder cases in selected ASEAN countries from 1990 to 2014, while the validating set contains data from 2015 to 2019. This process is repeated for both mortality rates and self-harm rates. Finally, we evaluate the performance of our models using the validation set through the test loss. The results are presented in the table below:

Test-Loss	**Linear Regression**	**LSTM**
Number of Cases	0.3705	0.0608
Mortality Rates	0.5210	0.3156
Self-harm Rates	3.2280	1.1066

Looking at the table above, we can see that the test loss for the LSTM model is significantly lower than that of the linear regression model for all three cases. This suggests that the LSTM model is better at predicting the outcomes for these cases.

Furthermore, if we look at the nature of the data, we can see that it is time series data, which makes it well-suited for LSTM models that can capture temporal dependencies. On the other hand, linear regression models assume that there is a linear relationship between the input and output variables, which may not be appropriate for time series data.

Therefore, based on the lower test loss and the nature of the data, it seems reasonable to choose the LSTM model for predicting the outcomes of these cases.

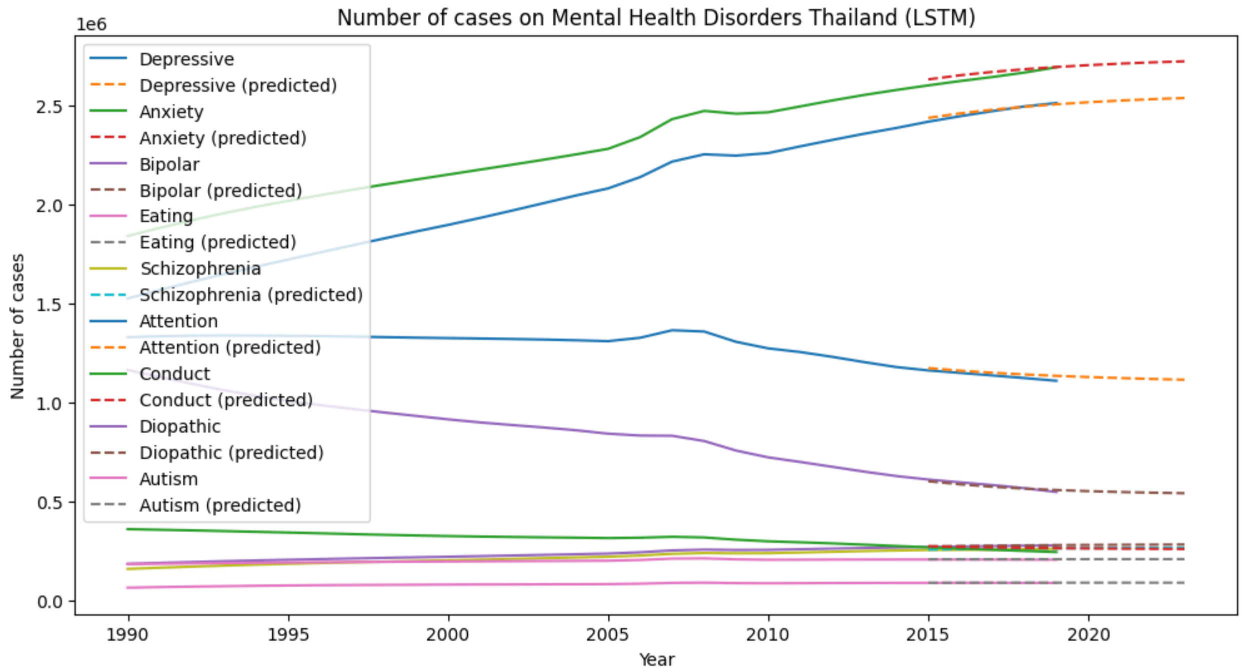


Figure 9: Predicting the number of cases of Thailand Mental Health Disorder

According to **Figure 9**, the LSTM model has successfully predicted the future behaviors on the number of cases with mental health and substance use disorders until 2025, with highly accurate results that closely match validating data. In Thailand, the future trends on most disorder types are expected to remain stable, while the number of cases with diopathic and attention disorders are more likely to decrease slightly. On the other hand, depression and anxiety are forecasted to increase slightly and they are currently the two most prevalent symptoms among all disorders with a very high number of cases.

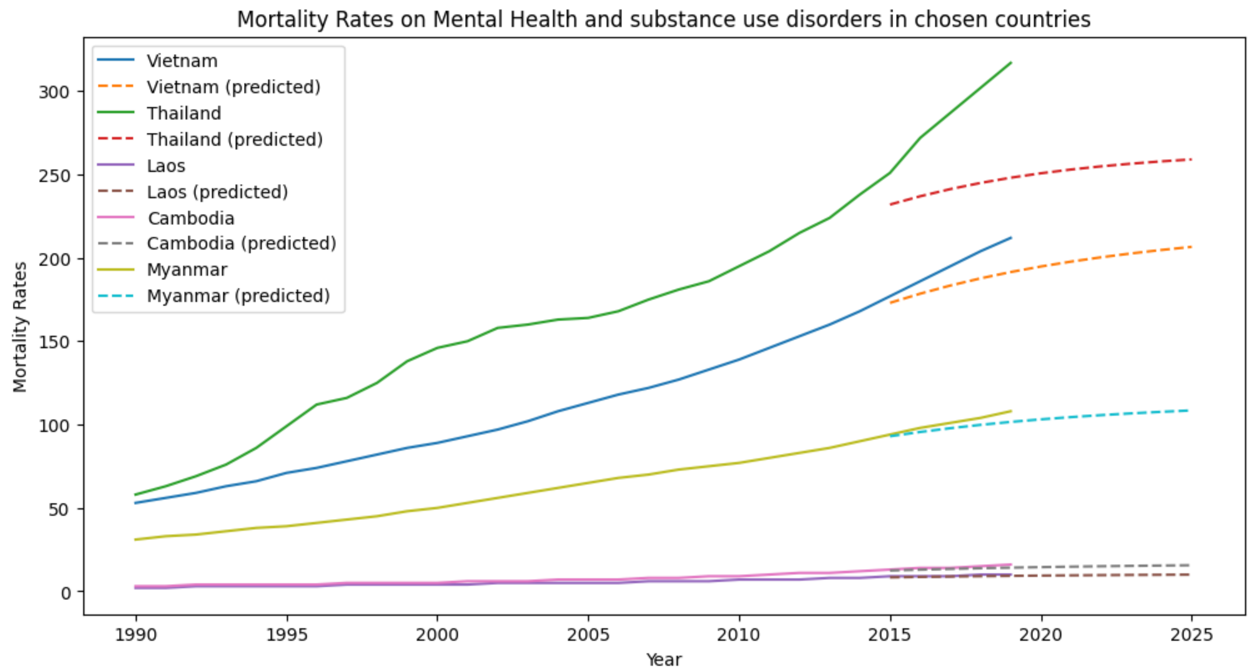


Figure 10: Predicting mortality Rates on Mental Health and substance use disorders in chosen countries

Based on the findings presented in **Figure 10**, it appears that the LSTM model's predictions for mortality rates in five selected countries are accurate in two countries, Laos and Cambodia. However, the predictions for Vietnam and Thailand are not as precise, with actual data differing from the model's forecast. It is possible that expanding the time range to include years before 1990 could improve the model's predictions for these countries. Additionally, Vietnam and Thailand have much higher mortality rates compared to the other three countries, and these rates are expected to continue to increase in the coming years. Meanwhile, Laos and Cambodia are expected to maintain their current levels, while Myanmar's rates are expected to slightly increase.

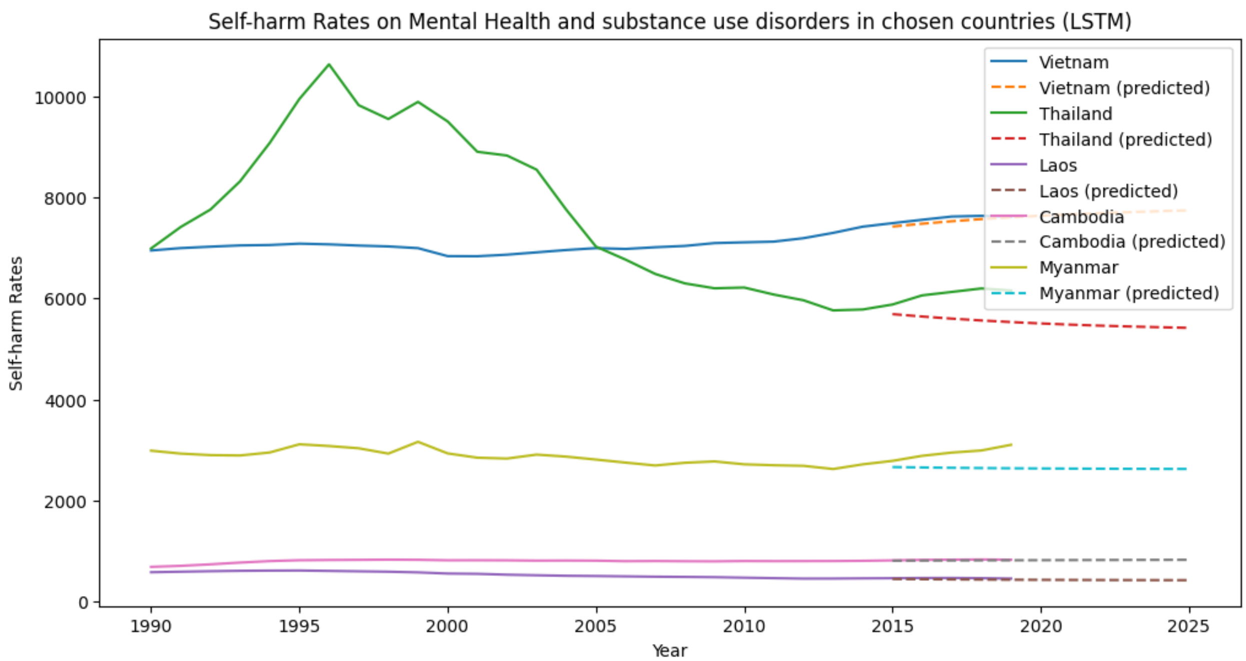


Figure 11: Predicting self-harm Rates on Mental Health and substance use disorders in chosen countries

According to **Figure 11**, the LSTM model predicts future trends of self-harm rates in five selected countries. The model is particularly effective in forecasting the trends in Vietnam, Laos, and Cambodia, but its predictions for Myanmar and Thailand are less accurate compared to actual data. It may be possible to improve the predictions by expanding the time range to include years before 1990. Generally, the self-harm rates are expected to decrease slightly until 2025 in all countries except Vietnam, where there is a slight upward trend.

VIII. SUGGESTED SOLUTION

Based on the descriptive analysis and predictive analysis above, we can foresee that the number of people who have mental issues are likely to increase constantly in at least the next few years. Thus, before jumping directly into our suggested solutions, we would like to provide the audience with a snapshot of key gaps which are contributing to the increasing number of patients.

Gap Issue	Description
Governance gap	Inadequate policies, plan and laws and misplaced priorities: two out of every three dollars spent on mental health goes to running psychiatric hospitals.
Resource gap	Scarce workforce: in low-income countries, there are fewer than one mental health worker per 100 000 population. Digital divide: most households in least developed countries do not have internet access
Services gap	Poor treatment coverage: 71% of people with psychosis do not receive mental health services. Limited range and quality of services: Few countries provide psychosocial interventions in primary care.

From some mentioned key gaps, here is a plethora of potential and possible solutions which we believe it would have a notable positive impact on current situation:

1. **Advancing public health:**

   People living with severe mental health conditions die 10–20 years earlier than the general population, most often through unrecognized and untreated physical health conditions. Therefore, Investing in mental health can greatly reduce suffering and improve the quality of life, social functioning and life expectancy of people with mental health conditions. It can both close the vast care gap that exists for mental health conditions and move significantly closer towards universal health coverage.

2. **Organizing mental health workshops, and short-term courses for young people:**

Mental health issues, especially depression, is having a rejuvenation trend which leads to a significant proportion of adolescents having a mental problem without having comprehensive awareness and knowledge of it. Thus, mental health workshops or short-term courses can provide the necessary knowledge and skills for them to self-protect from potential mental risks. The below table provides us with adverse factors and protective factors with each level of each individual, where we can take advantage of that information to design suitable information transformation approaches for young people.

Level	Adverse Factors	Protective Factors
Individual attributes	Low self-esteem	Self-esteem, confidence
	Cognitive/emotional immaturity	Ability to solve problems & manage stress or adversity
	Difficulties in communicating	Communication skills
	Medical illness, substance use	Physical health, fitness
Social circumstances	Loneliness, bereavement	Social support of family & friends
	Neglect, family conflict	Good parenting/family interaction
	Exposure to violence/abuse	Physical security & safety
	Low income & poverty	Economic security
	Difficulties or failure at school	Scholastic achievement
	Work stress, unemployment	Satisfaction & success at work
Environmental factors	Poor access to basic services	Equality of access to basic services
	Injustice & discrimination	Social justice, tolerance, integration
	Social & gender inequalities	Social & gender equality
	Exposure to war or disaster	Physical security & safety

3. **Building public awareness and interest:**

Building public awareness and interest through education, campaign is essential to transform and scale mental health care. If the general public does not know about or is not interested in mental health issues, they are less likely to take responsibility for self-care, to seek appropriate help when they aren't well, or to prioritize access to quality mental health care for all.

VIII. CONCLUSION

In conclusion, along with the development of society, mental health problems also tend to increase at a considerable level. This project aims to not only deliver the audience with a better understanding of the seriousness of the current mental health situation but also try to figure out feasible solutions to contribute to controlling this trend through education, propagation, and campaign. We are also really aware that shifting public attitudes and tackling stigma towards mental health issues is not easy. But experience shows it is possible if the authorities and global health organizations truly take into account that matter step by step.

VIII. LIMITATION

One of the limitations of our project in the topic of mental health and substance use disorders is the lack of available data for the testing set, which may result in an unreliable evaluation of the performance of the models. Without a sufficient amount of data, there is a risk of overfitting or underfitting the models, which can lead to inaccurate predictions and conclusions.

Furthermore, the absence of relevant datasets on mental health-related factors, such as substance use rates, hospital admission rates as well as recovering rates, can limit the ability to perform prescriptive analysis and pinpoint the key factors that contribute to mental health and substance use disorders. These data points can provide valuable insights into the prevalence and severity of these disorders, as well as the effectiveness of healthcare systems and government policies in addressing them.

Therefore, the lack of available testing data and relevant datasets can be a significant limitation for the project on providing detailed recommendations to minimize the impacts of mental health and substance use disorders. Moreover, It may hinder the ability to accurately identify the root causes of these disorders and develop effective strategies for prevention and treatment.

IX. FUTURE WORK

For detail, due to limited availability of relevant datasets, exploration of alternative data sources can be considered. One of the challenges faced in data analysis is the limited availability of relevant datasets. In some cases, the datasets that are available may not be sufficient to provide the insights needed to make informed decisions. Thus, it may be necessary to explore alternative data sources.

Secondly, we can enhance the analysis capabilities by improving our dashboard using more different related features in various approaches. Dashboard is a great tool to present data insights to stakeholders. However, to make the dashboard more effective, it is essential to incorporate a wide range of related features in various approaches. For instance, these features can be presented by charts, graphs, and other visualizations that help to highlight key findings.

Moreover, we also improve comprehension of our report by re-selecting key factors as well as adding more components if possible, such as hospital admissions, recovering rates for further investigating the current state of healthcare services in order to identify areas where improvements can be made. Report is an important means of communicating data insights to stakeholders, so it is essential to ensure that our report is comprehensive and easy to understand. This can be achieved by selecting key factors that are relevant to the audience and presenting them in a clear and concise manner. For example, in the healthcare industry, hospital admissions and recovery rates could be important factors to consider in assessing the overall state of healthcare services. By including these components in the report, stakeholders can gain a better understanding of the current state of healthcare services and identify areas where improvements can be made.

Furthermore, we may experiment with different models to achieve better performance on prediction. This can involve testing different algorithms, adjusting hyperparameters, or incorporating additional features into the model. By experimenting with different models, we can identify the best approach for predicting future trends or outcomes of chosen key features. This can help stakeholders to make more informed decisions and improve their business operations.

REFERENCES

- TenEleven. (n.d.). *Behavioral Health EHR/EMR Outcomes Dashboards*. [online] Available at: https://10e11.com/features/behavioral-health-ehr-dashboards/ [Accessed 11 Feb. 2023].
- www.kaggle.com. (n.d.). *Mental Health Dataset*. [online] Available at: https://www.kaggle.com/datasets/programmerrdai/mental-health-dataset?select=prevalence-by-mental-and-substance-use-disorder.csv [Accessed 12 Feb. 2023].
- [Mental Health - Our World in Data](https://ourworldindata.org/mental-health#world-health-organization-who-international-classification-of-diseases-icd): https://ourworldindata.org/mental-health#world-health-organization-who-international-classification-of-diseases-icd
- Cai, H., Yuan, Z., Gao, Y., Sun, S., Li, N., Tian, F., Xiao, H., Li, J., Yang, Z., Li, X., Zhao, Q., Liu, Z., Yao, Z., Yang, M., Peng, H., Zhu, J., Zhang, X., Gao, G., Zheng, F. and Li, R. (2022). A multi-modal open dataset for mental-disorder analysis. *Scientific Data*, [online] 9(1), p.178. doi:https://doi.org/10.1038/s41597-022-01211-x.
- “World mental health report: Transforming mental health for all.” 2022. World Health Organization (WHO). https://www.who.int/publications/i/item/9789240049338.