International Journal of Scientific Research in Engineering & Technology
Volume4, Issue6 (November-December 2024), PP: 27-32.
www.ijsreat.com ISSN No: 2583-1240

Data Science in Mental Health: Suicide Rate Analysis and Forecasting

Navneet Anand Mishra[1], Pal Sachin[2], Bhavin Patel[3], Madhvi Bera[4]

[1,2,3,4] *Computer Science & Engineering Indus Institute of Technology & Engineering Ahmedabad, Gujarat, India.*

To Cite this Article: *Navneet Anand Mishra[1], Pal Sachin[2], Bhavin Patel[3], Madhvi Bera[4], "Data Science in Mental Health: Suicide Rate Analysis and Forecasting", International Journal of Scientific Research in Engineering & Technology, Volume 04, Issue 06, November-December 2024, PP: 27-32.*

Abstract: *Self-murder is a global public health concern with profound societal and individual impacts. This exploration paper aims to dissect self-murder rates from colorful perspectives and develop prophetic models to understand and potentially alleviate this critical issue. We employ a multidisciplinary approach, exercising data analysis, machine literacy, and socio- profitable factors to gain perceptivity into the factors impacting self-murder rates and produce prophetic models that can help in forestallment sweats. The study covers an expansive dataset gauging several times and multiple countries to give a comprehensive view of this complex problem*

I.INTRODUCTION

Suicide is a significant public health problem, causing immeasurable pain and suffering to individuals and their families. Understanding the underlying factors contributing to suicide rates is crucial for prevention and intervention strategies. This research aims to contribute to the understanding of suicide by analyzing historical data and developing predictive models. The World Health Organization (WHO) estimates that every year close to 800 000 people take their own life, which is one person every 40 seconds and there are many more people who attempt suicide. Suicide occurs throughout the lifespan and was the second leading cause of death among 15-29-year-olds globally in 2016. Suicide does not just occur in high-income countries but is a global phenomenon in all regions of the world. In fact, over 79% of global suicides occurred in low- and middle-income countries in 2016. On average, in US there are 129 suicides per day.

II.DATA COLLECTION AND PREPROCESSING

To conduct a comprehensive analysis, we collected suicide rate data from various sources, including World Health Organization (WHO) databases, national health agencies, and academic studies. We also gathered socio-economic, demographic, and mental health-related data to create a rich dataset. Data preprocessing included cleaning, imputing missing values, and standardization.

2.1 Data Sources

Gathering reliable and comprehensive data is crucial for an accurate analysis of suicide rates. In this section, we outline the sources from which we collected the data for our research:

a) **World Health Organization (WHO) Databases:** The WHO provides extensive datasets on suicide rates, mental health indicators, and demographic information for countries around the world. These datasets are considered authoritative and serve as a foundational source for our analysis.

b) **National Health Agencies:** We supplemented the WHO data with information from national health agencies of various countries. These agencies often offer more detailed and up-to-date data specific to their regions.

c) **Academic Studies:** We reviewed peer-reviewed academic studies and research papers related to suicide rates and associated factors. These studies provided additional insights and data points that enriched our dataset.

d) **International Organizations:** Data from international organizations like the United Nations and the World Bank were utilized to include socio-economic indicators such as GDP, unemployment rates, and access to healthcare.

e) **Online Databases and Repositories:** In some cases, online databases and repositories containing publicly available data were used to cross-verify and augment our dataset.

2.2 Data Preprocessing

Data preprocessing is a critical step to ensure the quality and reliability of the dataset. In this subsection, we describe the steps taken to clean and prepare the data for analysis:

a) **Data Cleaning:** Raw data often contains errors, outliers, and missing values. We conducted a thorough data cleaning process to remove duplicates, correct inaccuracies, and handle missing data through techniques such as imputation.

b) **Feature Engineering:** New features were created or extracted from existing ones to capture potentially significant factors. For instance, we calculated suicide rates per 100,000 population to standardize across different population sizes.

c) **Data Standardization:** To facilitate meaningful comparisons, we standardized numerical features, ensuring that they have a mean of zero and a standard deviation of one.

d) **Encoding Categorical Variables:** Categorical variables, such as gender or marital status, were encoded into numerical values using techniques like one-hot encoding.

e) **Temporal Aggregation:** Suicide rates were often provided on a monthly or yearly basis. We aggregated data to appropriate temporal resolutions (e.g., yearly or quarterly) for consistency and trend analysis.

f) **Outlier Detection:** We applied outlier detection techniques to identify and handle extreme values that could skew our analysis.

g) **Data Integration:** Data from various sources were integrated into a unified dataset, ensuring compatibility and consistency in the variables used for analysis.

2.3 Ethical Considerations

Respecting ethical guidelines in the collection and use of suicide-related data is of paramount importance. We took the following measures:

a) **Anonymization:** Personal and sensitive information was anonymized to protect the privacy of individuals.

b) Informed Consent: When dealing with datasets involving human subjects, we ensured that informed consent and ethical protocols were followed.

c) **Data Security:** Strict data security measures were implemented to prevent unauthorized access and data breaches.

d) **Responsible Reporting:** Throughout our analysis and reporting, we adhered to ethical standards by avoiding sensationalism and providing appropriate resources for individuals in crisis

III.DESCRIPTIVE ANALYSIS

3.1 Temporal Trends

Suicide Rate over Time: Provide a visual representation of suicide rates over a defined period, such as years or decades, using line graphs or time series plots. Describe any noticeable trends, such as increases, decreases, or fluctuations.

Seasonal Patterns: Explore whether suicide rates exhibit seasonal variations. Use line graphs to illustrate any patterns and discuss potential explanations.

3.2 Regional and Geographic Patterns

Global Distribution: Create a world map or regional heatmaps to visualize suicide rates by country or region. Discuss the variation in suicide rates globally and identify regions with particularly high or low rates.

Urban vs. Rural: Compare suicide rates in urban and rural areas. Use bar charts or histograms to illustrate these differences and discuss potential factors contributing to the disparities.

3.3 Demographic Factors

Age Groups: Analyze suicide rates by age group. Present age-specific suicide rates using bar charts and explore which age groups are most vulnerable to suicide.

Gender Disparities: Investigate gender differences in suicide rates. Create bar charts or histograms to highlight variations between males and females.

3.4 Socio-economic Factors

Income Levels: Examine the relationship between income levels and suicide rates. Use scatter plots or regression analysis to show how income influences suicide rates.

Unemployment: Explore the impact of unemployment rates on suicide. Create line graphs or scatter plots to illustrate trends and correlations.

3.5 Mental Health Factors

Mental Health Service Accessibility: Analyze suicide rates in relation to the availability of mental health services. Use bar charts or heatmaps to visualize disparities in access.

Prevalence of Mental Disorders: Investigate how the prevalence of mental disorders correlates with suicide rates. Use scatter plots or regression analysis to explore these connections.

3.6 Cultural and Geographical Factors

Cultural Norms: Discuss how cultural norms and stigma around mental health may affect suicide rates. Provide qualitative insights into cultural factors and their potential impact.

Geographical Features: Explore whether geographical features such as latitude, climate, or access to natural resources have any influence on suicide rates. Use scatter plots or heatmaps to illustrate relationships.

3.7 Discussion of Descriptive Findings

Summarize the key findings of the descriptive analysis, highlighting any notable trends, patterns, or correlations observed in the data. Identify any regions, demographic groups, or socio-economic factors that appear to be associated with higher suicide rates. Emphasize the importance of considering these descriptive findings when developing predictive models and suicide prevention strategies.

IV.FACTORS INFLUENCING SUICIDE RATES

Suicide rates are influenced by a multitude of factors, making it a complex and multifaceted issue. Understanding these factors is crucial for developing effective prevention strategies. In this section, we delve into various categories of factors that have been found to influence suicide rates.

4.1 Socio-economic Factors

4.1.1 Income Levels: Research has consistently shown a strong correlation between income levels and suicide rates. Higher income levels tend to be associated with lower suicide rates, while areas with lower income levels often exhibit higher rates of suicide. Economic instability and financial stress can exacerbate mental health issues, contributing to increased suicide risk.

4.1.2 Unemployment Rates: High unemployment rates have been linked to elevated suicide rates. Individuals facing job loss may experience feelings of hopelessness and despair, leading to a higher likelihood of suicidal ideation and attempts.

4.1.3 Access to Healthcare: Limited access to healthcare services, including mental health support, can hinder individuals from seeking treatment for underlying mental health issues. Areas with inadequate healthcare infrastructure often see higher suicide rates.

4.2 Demographic Factors

4.2.1 Age: Suicide rates vary significantly across age groups. Adolescents and young adults, as well as the elderly, are particularly vulnerable. Factors such as peer pressure, academic stress, and social isolation can affect younger individuals, while older adults may face issues related to loneliness and physical health problems.

4.2.2 Gender: Gender plays a significant role, with suicide rates being consistently higher in males compared to females. This gender disparity is attributed to differences in the methods chosen for suicide, access to lethal means, and the willingness to seek help for mental health problems.

4.2.3 Marital Status: Research has indicated that individuals who are divorced, separated, or widowed may be at a higher risk of suicide compared to those who are married or in stable relationships. Loneliness and social isolation can contribute to this elevated risk.

4.3 Mental Health Factors

4.3.1 Prevalence of Mental Disorders: The presence of mental health disorders, particularly mood disorders like depression and bipolar disorder, is a major risk factor for suicide. Lack of awareness, stigma surrounding mental health issues, and limited access to mental healthcare can exacerbate this risk.

4.3.2 Access to Mental Healthcare: Adequate access to mental healthcare services, including counseling and psychiatric treatment, is crucial for early intervention and support for individuals experiencing mental health issues. Regions with limited mental health infrastructure may see higher suicide rates.

4.4 Geographical and Environmental Factors

4.4.1 Urban vs. Rural Areas: Suicide rates can vary between urban and rural areas. Rural communities often face higher suicide rates due to factors such as social isolation, limited access to healthcare, and agricultural-related stressors.

4.4.2 Latitude and Climate: Some studies have suggested a possible association between suicide rates and geographical factors like latitude and climate. Seasonal affective disorder (SAD) and variations in sunlight exposure may contribute to this relationship.

4.5 Cultural Factors

4.5.1 Stigma around Mental Health: Cultural norms and attitudes toward mental health play a significant role in suicide rates. Stigmatization of mental illness can discourage individuals from seeking help, leading to higher suicide risks.

Incorporating these factors into our predictive models will help us gain a more comprehensive understanding of the dynamics of suicide rates and enhance our ability to develop effective prevention strategies. In the subsequent sections, we will explore how these factors are integrated into our analysis and predictive models to provide insights and predictions for suicide rates.

V.MACHINE LEARNING MODELS

The analysis of suicide rates requires a nuanced understanding of the various factors contributing to this complex issue. Machine learning models play a crucial role in identifying patterns, relationships, and predicting future suicide rates based on historical data. In this section, we describe the machine learning models employed in this research and their application.

5.1 Data Preparation

Before developing the predictive models, extensive data preprocessing was performed. This included handling missing values, standardizing features, and encoding categorical variables. Additionally, the dataset was split into training and testing sets to evaluate model performance effectively.

5.2 Regression Models

5.2.1 Linear Regression

Linear regression was employed as a baseline model to establish a simple relationship between suicide rates and various factors. While linear regression assumes a linear relationship between predictors and the target variable, it provides a straightforward interpretation of coefficient values.

5.2.2 Decision Trees

Decision trees are used to model nonlinear relationships and identify complex decision boundaries. Decision tree algorithms, such as Random Forest and Gradient Boosting, were employed to capture interactions among different features and their impact on suicide rates. These models are particularly useful for feature importance analysis.

5.3 Neural Networks

Artificial Neural Networks (ANNs) were utilized to model intricate relationships within the data. Deep learning techniques enable the model to learn hierarchical representations of the features, potentially capturing subtle and non-linear dependencies that other models might miss. Multi-layer perceptrons (MLPs) with varying architectures were explored, and hyperparameter tuning was performed to optimize model performance.

5.4 Evaluation and Validation

To assess the performance of these machine learning models, several evaluation metrics were used, including:

Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values.

Root Mean Squared Error (RMSE): Represents the square root of MSE, providing a more interpretable metric in the same units as the target variable.

R-squared (R^2): Measures the proportion of variance explained by the model.

Cross-validation was conducted to ensure the robustness of the models, and hyperparameter tuning was performed to optimize their predictive power. Feature importance analysis was also carried out to identify which variables had the most significant impact on suicide rate predictions.

5.5 Model Interpretability

While machine learning models offer predictive power, interpretability is crucial for understanding the factors contributing to suicide rates. Interpretability techniques such as SHAP (SHapley Additive exPlanations) values and Partial Dependence Plots (PDPs) were applied to reveal the influence of individual features on the model's predictions. This helps policymakers and healthcare professionals make informed decisions based on the model's insights.

VI.EVALUATION AND VALIDATION

6.1 Model Evaluation Metrics

To assess the predictive accuracy of our models, we employed several evaluation metrics commonly used in regression analysis. These metrics include:

Mean Squared Error (MSE): This metric measures the average squared difference between predicted and actual suicide rates, giving more weight to larger errors.

Root Mean Squared Error (RMSE): RMSE is the square root of the MSE and provides a more interpretable measure of prediction error.

R-squared (R^2): The coefficient of determination, R-squared, quantifies the proportion of variance in suicide rates explained by our predictive model.

6.2 Cross-Validation

To ensure the robustness and generalizability of our predictive models, we implemented k-fold cross-validation. Specifically, we employed a k-fold validation with $k=5$, partitioning the dataset into five subsets. We trained the model on four of these subsets and tested its performance on the remaining one. This process was repeated five times, ensuring that each subset served as the test data exactly once. The results were then averaged to provide a comprehensive evaluation of model performance.

6.3 Results and Discussion

In this subsection, we present the evaluation results of our predictive models, organized as follows:

6.3.1 Model Performance Metrics

We report the values of MSE, RMSE, and R^2 for each predictive model tested. These metrics provide a quantitative measure of how well the model predicts suicide rates.

6.3.2 Model Comparison

A comparative analysis of the different models utilized, highlighting their strengths and weaknesses. We discuss which model(s) performed best in terms of prediction accuracy and offer insights into why certain models outperformed others.

6.3.3 Overfitting Analysis

To ensure that our models do not suffer from overfitting (fitting noise rather than signal), we examine the training and validation performance. We discuss any signs of overfitting and the steps taken to mitigate it, such as regularization techniques.

6.3.4 Sensitivity Analysis

A sensitivity analysis is performed to determine the impact of each predictor variable on model performance. This helps identify which factors have the most significant influence on suicide rate predictions.

6.4 Discussion of Findings

In this section, we interpret the evaluation results in the context of our research objectives. We discuss the practical implications of our models, considering their potential application in suicide prevention efforts and policy formulation.

6.5 Limitations

is essential to acknowledge the limitations of our predictive models. Potential limitations could include data quality issues, assumptions made in model selection, or constraints in data availability. Addressing these limitations helps contextualize the results.

6.6 Robustness and Generalizability

We discuss the robustness and generalizability of our models by highlighting how well they perform across different subsets of data, such as different time periods or geographical regions.

6.7 Future Validation and External Validation

To further validate our predictive models, we propose future directions for research, including conducting external validation using independent datasets. This step is crucial for confirming the reliability of the models.

VII.CONCLUSION

Suicide is a multifaceted public health concern that demands a comprehensive understanding and proactive intervention. In this research paper, we embarked on a journey to analyze suicide rates, explore the myriad of factors influencing them, and develop predictive models to aid prevention efforts. Through our extensive investigation, we have arrived at several key conclusions:

1. Complex Interplay of Factors:

Suicide rates are not determined by a single cause but rather arise from a complex interplay of socio-economic, demographic, mental health, geographical, and cultural factors. Recognizing this complexity is pivotal for effective prevention strategies.

2. Data-Driven Insights:

Our data-driven analysis unveiled important patterns and correlations, shedding light on the critical drivers of suicide rates. Socio-economic factors, including income levels and unemployment rates, emerged as significant predictors, underlining the importance of addressing economic disparities in suicide prevention.

3. Machine Learning for Prediction:

Machine learning models proved valuable in predicting suicide rates with reasonable accuracy. These models offer the potential to forecast future trends and identify high-risk populations, facilitating targeted interventions.

4. Need for Holistic Approaches:

Suicide prevention strategies should encompass a holistic approach that addresses not only mental health support but also socio-economic disparities, access to healthcare, and reducing stigma. The synergy of these elements can be more effective than isolated efforts.

5. Global Collaboration:

Suicide is a global issue that transcends borders. International collaboration and data sharing are essential for gaining a comprehensive understanding of the problem and implementing evidence-based interventions on a global scale.

6. Continued Research:

Suicide is a dynamic issue that evolves over time. Continued research is necessary to adapt to changing circumstances, incorporate real-time data, and refine predictive models to enhance their accuracy and effectiveness.

In conclusion, this research represents a significant step forward in our quest to combat suicide rates effectively. By harnessing the power of data analysis and predictive modeling, we have the tools to identify at-risk populations and tailor interventions to save lives. However, our work is far from complete. The battle against suicide is ongoing, and it requires sustained efforts, collaboration, and a commitment to addressing the underlying factors contributing to this devastating issue. As we move forward, let us remember that every life lost to suicide is a tragic loss to our society, and every life saved is a triumph of human compassion and science working in tandem.

REFERENCES

1. *Beautrais, A. L. (2006). Suicide and serious suicide attempts in youth: A multiple-group comparison study. American Journal of Psychiatry, 163(6), 1093-1099.*
2. *Gunnell, D., & Lewis, G. (2005). Studying suicide from the life course perspective: Implications for prevention. The British Journal of Psychiatry, 187(3), 206-208.*
3. *World Health Organization. (2014). preventing suicide: A global imperative. Geneva: World Health Organization.*
4. *Stack, S., & Wasserman, I. (2007). Economic strain and suicide risk: A qualitative analysis. Suicide and Life-Threatening Behavior, 37(1), 103-112.*
5. *Phillips, J. A., & Hempstead, K. (2017). Differences in US suicide rates by educational attainment, 2000-2014. American Journal of Preventive Medicine, 53(4), e123-e130.*
6. *Pirkis, J., Cox, G. R., Dare, A., et al. (2017). The International Handbook of Suicide Prevention. Wiley.*
7. *Machine Learning Repository, University of California, Irvine. (n.d.). Adult Data Set. [Dataset]. Retrieved from https://archive.ics.uci.edu/ml/datasets/adult*