



# A Survey paper on database: Suicides in India

<sup>1</sup>Sarika H. Zile, <sup>2</sup>Pratik Khodwe.

<sup>1</sup>Assistant Professor, <sup>2</sup>Student.

<sup>1</sup>Department of Computer Science and Engineering,  
<sup>1</sup>Maharashtra Institute of Technology, Aurangabad(MH), India

**Abstract:** Suicide is an important worldwide public health issue, and India bears a large percentage of the global suicide burden. This study examines suicide patterns in India from 2001 to 2012 using an open-access dataset of 237,519 entries. The dataset contains detailed information on demographic, temporal, and cause-specific factors, including state, year, cause categorization, specific reasons, gender, age group, and total suicide counts. The study executes meticulous analysis to illustrate variations in suicide rates among demographic groups and geographies, which are influenced by socioeconomic and cultural factors. Notable findings include gender and age-related vulnerability, as well as the most usual explanations of suicide. The research offers practical insights to help academics, legislators and mental health practitioners create focused treatments by integrating visual designs with geographical breakdowns. The outcomes points out how valuable it is to use methods based on data to tackle the complexities of suicide prevention, setting the stage for more successful tactics in India.

**IndexTerms -** Suicide prediction, suicide prevention, demographic trends, suicide datasets, mental health analysis, socioeconomic factors, regional disparities, machine learning, data privacy, ethical considerations.

## I. INTRODUCTION

### INTRODUCTION

Suicide has serious repercussions for both people and communities, making it an imminent risk to public health worldwide. The persistently elevated rates of suicide in India clarify how urgently preventative measures are needed. Reducing the suicide burden involves determining people with elevated risk and risk factors. By examining past data and comprehending patterns, forecast for suicide systems provide a chance to move from reactive to proactive tactics. These platforms may boost the distribution of mental health resources and provide information for early treatments.

This study used the "Suicides in India from 2001-2012" dataset, which contains thorough groupings, chronological, and cause-related information. Analysing this dataset reveals trends and connections, which may be used to design effective suicide mitigation frameworks for India's varied population.

### NEED OF THE STUDY.

Suicide prediction is a crucial but problematic field owing to the complexities of suicidal behaviour, which is influenced by a variety of psychological, social, and economic variables. Traditional preventative strategies frequently fail to detect early warning signals, leaving missed opportunities for intervention. Using past statistics to create prediction models provides a proactive solution to this problem. These methods enable the identification of high-risk people or groups, laying the groundwork for evidence-based preventative initiatives.

The "Suicides in India from 2001-2012" dataset is an invaluable resource for analysing suicide patterns in India. The set of data includes demographic, chronological, and cause-related knowledge about enabling a thorough examination of the underlying causes influencing suicides. By detecting patterns and risk factors, the information helps to construct prediction models that can influence region-specific treatments. Integrating findings from this study into suicide prevention systems improves their effectiveness and immediateness.

#### 1.1 Motivation

The concerning climb in suicide rates in the middle of the world, most significantly in India, serves as a strong call to action for researchers, politicians, and mental health researchers. India's overwhelming contribution to the worldwide suicide burden underscores the critical need to identify and treat the underlying causes of this public health catastrophe. Despite breakthroughs in mental health awareness, traditional preventative strategies frequently fail to identify at-risk persons in time to act successfully. Using extensive datasets like "Deaths by suicide in India from 2001-2012," researchers can identify crucial patterns and connections that would otherwise be concealed in fragmented data. This research is inspired by the ability to turn historical data into practical insights, allowing for data-driven policies that address the unique demands of India's various demographic and regional

communities. Such an approach not only develops prediction systems, but also helps to achieve the larger aim of lowering the social and human toll of suicide.

## 1.2 Objectives

1. Identify patterns: Examine relationships between suicide rates and demographic parameters including gender, age, and geography.
2. Analyse trends: Look for major swings in suicide rates across time and geography.
3. Propose Predictive Features: Identify essential characteristics, such as cause categories, age groups, and geographical discrepancies, to inform suicide prediction systems.
4. Support prevention efforts: Provide actionable insights to help politicians and mental health practitioners create data-driven suicide prevention strategies.

## LITERATURE REVIEW

### Data Sources and Analytical Techniques

Korrapati et al. (2018) stressed the need of combining many data sources, such as social media platforms, healthcare records, and demographic databases, to increase suicide prediction accuracy. Suicide risk patterns have been identified using techniques such as Support Vector Machines (SVM) and sentiment analysis. Similarly, DeLeo et al. (2013) demonstrated the importance of demographic and psychological characteristics in improving predictive model performance.

### Social and Psychological Factors

Strong social capital and community involvement were associated with a lower risk of suicide, according to Helliwell and Putnam (2004), indicating the protective function of social support networks. According to Patel et al. (2017), cultural and societal influences have significant effects on the prevalence of suicide in India. Their studies emphasized the significance of family dynamics and digital footprints in identifying communities that are at risk.

### Machine Learning in Suicide Prediction

Recent advances in machine learning have allowed for the creation of complex prediction models. Chavez et al. (2019) proved the effectiveness of neural networks in detecting suicidal thoughts using text data from social media. However, Mishra and Kuri (2022) identified ethical issues with implementing such systems, such as privacy problems and the possibility of algorithmic prejudice.

### Challenges

Principal problems in the prediction of suicide include data collecting biases, suicide underreporting, and the ethical implications of exploiting sensitive data. As Schell and Schafer (2021) emphasized, the lack of longitudinal research restricts the capacity to trace temporal changes in suicide risk, emphasizing the importance of ongoing and extensive data gathering.

### Dataset Overview:

The dataset "Suicides in India" is a comprehensive compilation of data on suicide instances in India during a 12-year period, with a total of 237,519 entries. It provides key information for each entry, such as state, year, gender, age group, cause type, and total suicides, allowing for a comprehensive analysis of suicide trends and contributing variables. The dataset is an important tool for studying how suicide rates change across demographic, historical, and socioeconomic aspects, providing insights into susceptible populations and locations. Researchers may use this data to find patterns that will help them develop more effective suicide prevention techniques for India's unique socio-cultural setting.

### Demographic Variables:

The demographic characteristics in this dataset (gender, age group, and state) are critical for assessing suicide rate discrepancies between populations. Gender enables researchers to investigate gender-based disparities in suicide rates, revealing if some genders are more vulnerable in particular places or historical periods. Age group statistics can assist determine which age groups are more at risk, which is especially essential for studying suicide patterns among teenagers, adults, and the elderly. State data is especially useful for finding regional discrepancies in suicide rates, as certain states may have greater or lower suicide rates owing to economic, social, or cultural variables. This enables more tailored preventive initiatives at the regional level, addressing the various difficulties encountered in different sections of the country.

### Temporal Data:

The year-wise data is essential for identifying temporal trends in suicide rates. By tracking the number of suicides per year, researchers can observe how the suicide rate has evolved over time, including identifying years with spikes in suicides, which could be linked to socio-economic changes, political events, or societal crises. The temporal data can also help assess the effectiveness of public health campaigns and mental health interventions, providing a historical context for evaluating their impact. Moreover, trends in year-wise data can reveal the broader shifts in suicide causes and demographics over time, contributing to a deeper understanding of the evolving nature of suicide risk factors in India.

### Cause Categories:

The dataset includes cause categories that specify the reasons behind suicides, such as family problems, financial distress, and other social or psychological factors. This categorical information is invaluable for identifying the key drivers of suicide across different regions and demographic groups. Understanding the specific causes behind suicides can inform more focused prevention efforts, such as financial counselling programs, support for victims of domestic violence, or targeted mental health initiatives.

Additionally, analysing the cause types across different years, genders, and age groups allows for the identification of trends and emerging issues, such as the increasing role of social media or economic pressures that may be influencing suicide rates.

### Size and Accessibility:

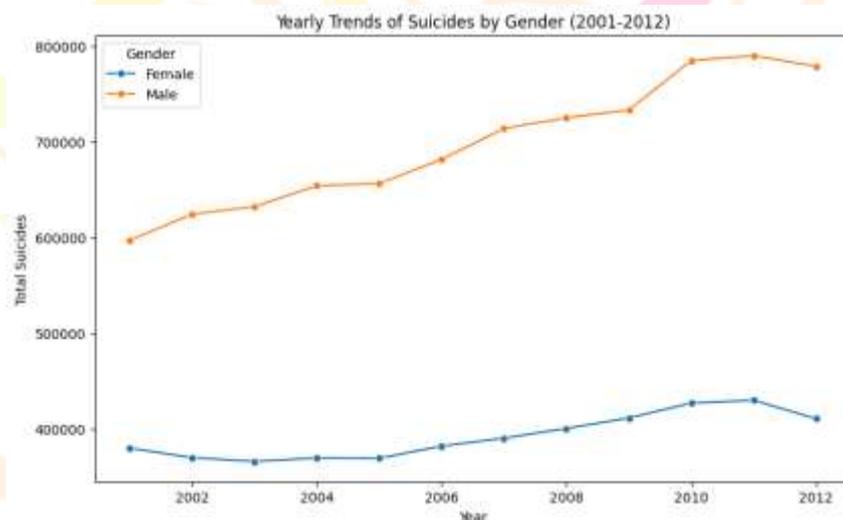
The large sample size of 237,519 records provides a robust foundation for statistical analysis and machine learning modelling. This extensive dataset allows researchers to draw more reliable conclusions about trends and patterns in suicide behavior. The dataset's public availability further enhances its value, as it provides an accessible resource for a wide range of stakeholders, including public health researchers, policymakers, and mental health professionals. However, it is important to note that while the dataset offers extensive demographic, temporal, and cause-related data, it lacks nuanced psychological and behavioural data, such as mental health status, prior suicide attempts, and other individual risk factors that could improve the accuracy and depth of predictive models. Therefore, for more advanced suicide prediction systems, it is crucial to supplement this dataset with additional information on psychological factors and behavioural indicators.

In summary, the "Suicides in India from 2001-2012" dataset offers a comprehensive and valuable resource for understanding suicide trends in India, but further data expansion is necessary to improve the predictive power of models and inform more effective intervention strategies.

### Data Visualization:

The analysis of this dataset can provide invaluable insights for real-world suicide prevention systems:

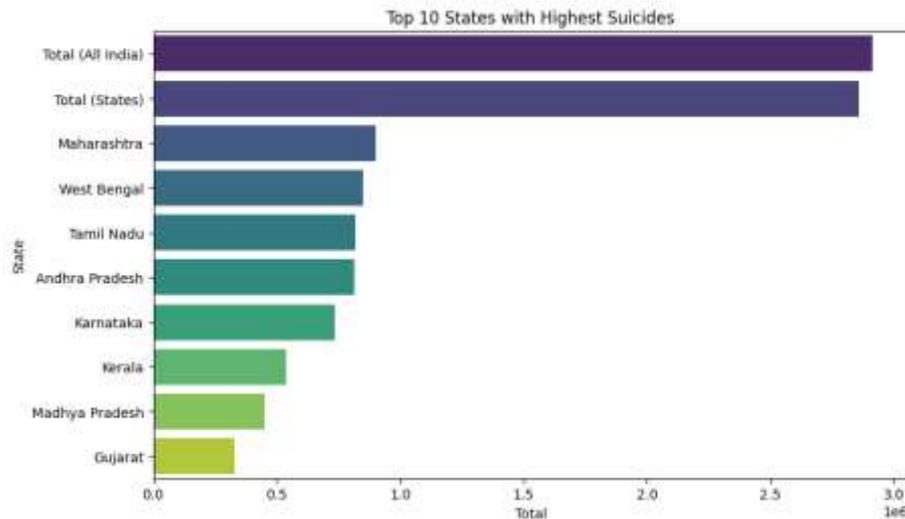
- **Predictive Features:** By identifying key features such as age, gender, state, and cause categories (e.g., economic distress, family problems), and predictive models can be developed to assess suicide risk. For instance, certain age groups or gender-based patterns may emerge, which can inform intervention strategies.
- **Trend Analysis:** Temporal analysis of suicides by year, as demonstrated by the yearly trends plotted in our previous analysis, can help predict future suicide rates and identify times of heightened risk (e.g., after economic downturns or societal crises).



**Figure 1: Yearly Trends in Suicides (2001-2012)**

The plot above reveals a noticeable fluctuation in suicide rates, with certain years showing higher spikes. Identifying such patterns helps policymakers and mental health professionals allocate resources more effectively.

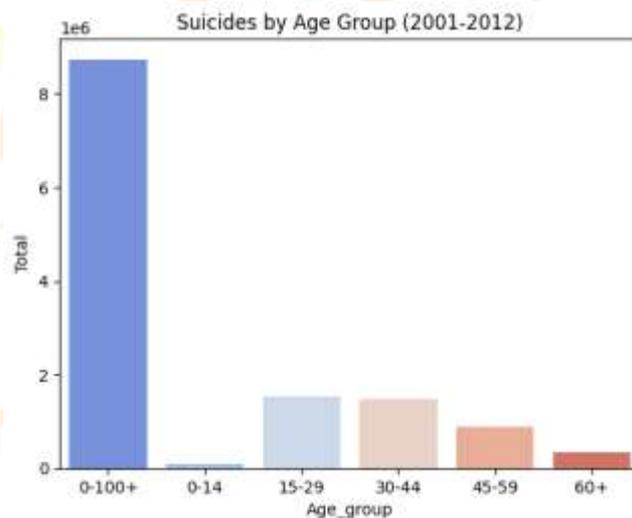
- **State-Wise Analysis:** Suicides by state which reveals significant regional disparities, is critical for targeted suicide prevention efforts. Certain states, particularly those with higher suicide rates, can benefit from localized intervention programs.



**Figure 2: Top 10 States with Highest Suicides**

As shown in Figure 2, Maharashtra and Tamil Nadu emerge as states with the highest suicide rates, indicating that regional factors such as economic distress, social issues, and mental health resources need to be considered when developing prevention strategies.

- **Age Group Trends:** By examining suicide rates across different age groups, we can identify vulnerable populations. It showed that younger age groups (15-29 years) often experience higher suicide rates, which is critical for designing age-specific intervention programs.



**Figure 3: Suicides by Age Group (2001-2012)**

Figure 3 highlights the higher frequency of suicides in the 15-29 age group, pointing to the need for mental health support targeting this demographic.

**Ethical Considerations:** Ethical considerations are paramount when working with sensitive data related to suicide. Data privacy must be a priority to ensure that individual identities remain anonymous and protected, preventing any risk of re-identification. This is particularly important given the sensitive nature of the subject. Additionally, bias mitigation is crucial to avoid skewed insights, as biases in data collection or representation can lead to inaccurate predictions and unfair conclusions about certain demographic groups. Another critical aspect is prediction sensitivity, which involves carefully managing the societal impact of false positives. Incorrectly labeling individuals as at risk can lead to stigma, unnecessary stress, and alarm, undermining the credibility of predictive systems. Addressing these ethical challenges is essential to develop responsible and effective suicide prediction systems that prioritize both accuracy and compassion.

**Challenges and Insights:** Suicide analysis in India faces several challenges that complicate data-driven prevention efforts. One significant issue is underreporting, particularly in rural areas, where stigma and cultural sensitivities may prevent accurate documentation of suicide cases. Additionally, feature extraction from the dataset poses technical difficulties, especially when handling categorical variables and addressing missing values, which are critical for building reliable models. Furthermore, bias in classification arises due to inconsistent categorization of suicide causes across states, potentially distorting overall trends and insights. Despite these challenges, the dataset offers valuable insights that can guide prevention strategies. It enables the identification of key predictive features, such as gender and age group, which are vital for assessing suicide risk.

### Conclusion and Future Directions:

The analysis of the "Suicides in India from 2001-2012" dataset highlights its immense potential as a tool for understanding suicide trends and informing targeted prevention strategies. Through detailed demographic and regional insights, this dataset reveals

significant disparities in suicide rates, identifying vulnerable groups such as specific age demographics and high-risk states. These findings provide a foundation for designing proactive interventions that address the unique challenges faced by different populations. However, the study also underscores the limitations of the existing dataset, particularly the absence of behavioral and psychological factors, which are critical for a deeper understanding of the underlying causes of suicide. Future directions for research and development in this domain focus on enhancing data quality and analytical techniques. Expanding data collection to include real-time behavioral patterns, psychological metrics, and unstructured data, such as social media activity, can significantly improve the accuracy of predictive models. Leveraging advanced machine learning methods capable of handling such complex data can further enhance predictive capabilities, enabling timely identification of at-risk individuals. Moreover, establishing comprehensive ethical frameworks is essential to address concerns about privacy, bias, and the equitable application of predictive systems. By prioritizing these areas, suicide prediction systems can evolve into powerful tools that not only save lives but also foster a broader societal commitment to mental health and well-being.

## II. ACKNOWLEDGMENT

We extend our heartfelt gratitude to the department of Computer Science and Engineering Maharashtra Institute of Technology for their continuous support and resources that made this research possible. We also acknowledge the availability of the open-access "Suicides in India from 2001-2012" dataset, which served as a cornerstone for this research. Lastly, we appreciate the encouragement of our families and colleagues, whose support was invaluable in completing this work.

## REFERENCES

- [1] Korrapati, R., Nuthalapati, K., & Thenmalar, S. (2018). A Survey Paper on Suicide Analysis.
- [2] Pannu, A., Sharma, P., & Gupta, S. (2020). Suicide Prediction in India Using Longitudinal Data: A Statistical Approach.
- [3] DeLeo, D., Bertolote, J., & Lester, D. (2013). Suicide and the Role of Mental Health.
- [4] Patel, V., Araya, R., & Chatterjee, S. (2017). Mental Health in Developing Countries: Challenges and Solutions.
- [5] Kumar, N., & Gupta, A. (2015). Analysis of National Crime Records Bureau (NCRB) Dataset: Patterns of Suicides in India. *Indian Journal of Social Psychiatry*, 31(2), 141-146.
- [6] World Health Organization (WHO). (2014). Preventing Suicide: A Global Imperative.
- [7] Zile, S. H., Kasar, S. L., & Deshmukh, D. R. (2023). Review on Machine Learning-Based Hybrid Model for Suicide Attempt Prediction.

