31

Context-Aware Suicide Rate Prediction for Significant Mental Health Monitoring in Smart Cities

Aravinda Boovaraghavan, V. Maheysh, Anova Pandey, Christy Jackson J.* and Abraham Sudharson Ponraj

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India

Abstract
In this digital age, the pressure to perform and succeed can often lead to depression and, in extreme cases, death. Suicide is defined as the deliberate act of causing one's death. They are increasing at an alarming rate all over the world, and it is important to reduce the number of suicides taking place. Suicide attempts have serious emotional, physical, and financial ramifications. People who attempt suicide and live may have serious injuries that have long-term health repercussions. It is a serious public health problem that can be avoided with timely, evidence-based, and usually low-cost treatments. A multisectoral suicide prevention plan is required for nationwide responses to be effective. An integrated system to perform suicide analysis is needed. Although a single factor cannot be pinpointed, among the common factors leading towards taking the extreme step, some of them can be analyzed through this project. A visualization of these could give us a better understanding of the causes of their suicide. A prediction plot of the overall rate of suicides in the future years is found using regression algorithms. A culminated model was created to predict the total count of suicide based on various factors including state, age group, and year. These results would be helpful to reduce the number of suicide rates in the world and create awareness among people.

***Keywords*:** Suicide, analysis, prediction, R programming, visualization, risk assessment

**Corresponding author*: christyjackson.j@vit.ac.in

O.V. Gnana Swathika and K. Karthikeyan (eds.) Digital Cities, (479–496) © 2025 Scrivener Publishing LLC

31.1 Introduction

Suicide involves the intentional act of causing one's demise. It is a very important public health problem. Suicides are predicted to cost $34.6 billion in medical and job losses [3]. In 2019, 139,123 suicides were recorded in India alone, underlining the fact that it is the world's tenth leading cause of death. India's percentage contribution to global suicide fatalities is significant and growing. India's SDR, particularly for women, is higher than predicted given the country's Socio-Demographic Index level, with significant differences in size and the ratio of males to women between states. India must create a suicide prevention strategy that considers these disparities in order to address this problem [9]. The suicide rate in cities was greater, at 13.9%, than the overall suicide rate in India, which was 10.4% [7].

Predicting suicidal thinking patterns is a challenging classification problem that will involve the investigation of dozens or hundreds of variables [8]. The pressure to perform and achieve in our digital era may frequently lead to melancholy and, in severe situations, death. Suicide affects people of all ages, genders, and ethnicities. It is the result of a mix of factors. Suicide attempts in the past, depression, other mental diseases, or a drug abuse issue are all possibilities. Another predictor includes persistent pain, as well as a family history of mental illness or substance misuse. Physical or sexual abuse in the family is also a risk factor. Although no single element can be identified as the only cause of taking the extreme step, a few will be examined in this study.

The dataset "Suicides in India 2001-2012.csv" was used for this purpose. Here are the attributes: state, year, type code, type, gender, age group, and total. Data cleaning is first performed to process the data, rename it, and split it depending on type code. The data are then extensively shown using various sorts of graphs. Based on these findings, several machine learning models were developed to forecast the suicide rate by year, year and age, year and reasons, year and states, and ultimately by year, age, and state. Linear Regression, Multilinear Regression, Logistic Regression, Lasso Regression, and Support Vector Machines were among the machine learning methods employed. All of these results will be provided to the user in order to shed light on this terrible problem.

31.2 Literature Survey

Suicide is a serious issue that must be addressed. The development of an integrated system for suicide prediction and analysis is required [1].

Blogs, social media, and polls may all be used to gather information. The ways of suicide that are currently in use are investigated. In our technological age, social media is quite important. Tweets are categorized based on gender, location, and age. By examining the presence of words associated with worry, despair, and melancholy, such emotional analysis can be utilized to map out suicidal thoughts. A model that divides clinical notes into word categories and analyzes text using a genetic algorithm, a type of supervised learning approach, may also be used to study clinical notes. Clinical notes including specific phrases such as "anxiety," for example, might be linked to suicidal intentions in patients [6]. Social Network Mental Disorders affect certain social networking users (SNMD). This has a strong association with attempts at suicide. Machine learning techniques, as well as semantic sentiment analysis, have been employed in social network suicide prediction. SVM, Maximum Entropy, and Naive Bayes models are employed for this purpose. The characteristics utilized for ranking include presence, word frequency, Ngrams, negation, and part of speech. Natural Language Processing (NLP)-based models, as well as machine learning approaches, were utilized.

After analyzing the data, it was shown that marital status, educational status, and gender had a substantial link with suicide. Suicide is committed by those who have a decent education. Females are more likely to plot suicide, although male suicide attempts are more likely to result in death. Regression models have also revealed that lonely students who have undergone some type of sexual assault had the highest likelihood of suicide ideation.

Suicide attempts are a substantial public health issue, with an estimated 25 million non-fatal suicide attempts happening globally each year [2]. The extent and gravity of the problem have spurred significant scholarly interest. Traditional methods for predicting suicide attempts and traditional statistical approaches are not well adapted for such investigations; thankfully, machine learning techniques are. The data were created using BioVU Synthetic Derivative, a de-identified data source of clinical EHR data from Vanderbilt University Medical Center. This study simulated many data classes: (a) data encompassing age, gender, race, and ethnicity; (b) claims-based diagnosis; (c) past health care history; and (d) prior suicide attempts.

Random forest is employed due to its high accuracy and ease of implementation. The random forest is an ensemble learning approach that consists of a series of decision trees produced by recursive sampling of bootstrapped predictor data samples. The authors of this paper ran a second set of logistic regression analyses to compare the current ML technique's performance to that of the classic logistic regression approach.

Model performance was assessed using AUC, accuracy, and recall measures in this technique. Brier scores vary between 0 and 1, where 0 signifies impeccable calibration and discrimination. Creating precise and scalable systems for detecting the risk of suicide attempts is vital for addressing these behaviors on a broad level. They made a significant contribution to formulating one such approach through the utilization of machine learning for the analysis of EHR data. This strategy yielded better accurate forecasts of suicide attempts with a longer lead time than traditional methods.

People get unhappy and worried as a result of the ongoing increase in rivalry between individuals in this modern era [3]. Their brains may be blocked as they strive to figure out how to get away. In such cases, people will profit from the available knowledge that is employed in decision-making by examining the attitudes, views, feelings, and emotions of others. An initial study of suicide data is performed in order to determine the many relevant factors that lead to suicide attempts. Python and Anaconda were used for the development of machine learning methodologies, encompassing Logistic Regression, Random Forest, and Naive Bayes, and have been employed to delve into these dimensions. Furthermore, assorted visualization aids, including bar plots, heat maps, and Kdeplot, have been harnessed to elucidate the ensuing dataset.

Suicide attempts are higher among persons with fewer than five friends (77.65%), showing that social relationships play an important role in keeping the mind quiet and sane. Suicide attempts among those aged 15 to 30 account for 87.06% of all attempts. Male suicide attempts outnumber female suicide attempts by a startling 3:1 in the presented dataset, whereas female suicide attempts outnumber male suicide attempts by 71.76%. Income also has a significant effect, since individuals in the lower economic strata are more vulnerable. Surprisingly, most suicide attempts in the 20–30 age group were in the normal weight category, whereas those who were overweight, underweight, or obese were the least impacted. The model with the best accuracy was also linear regression.

For the last 30 years, Beck's Suicide Intent Instrument (SIS) has been the most used psychometric scale for detecting suicide intent in suicidal attempters [4]. The researchers intended to discover if the Suicide Intent Scale had any predictive value in suicidal patients. The secondary purpose was to determine if employing the variables of the Suicide Intent Scale might enhance suicide risk detection, and as a result of the item analysis, a simplified version of the scale was constructed. According to the SIS Factorial Structure, two to four elements have been discovered. Mieczkowski found two components in suicide attempters: the planning subscale and fatal intent, while Mission *et al.* [16] recently published a four-factor SIS solution in suicide attempters. In a study by Antretter *et al.*, [17] just one aspect, the “subjective component” of the SIS, which has items 9 to 14, was well supported; however, in 11 clinical samples, no good model fit for the “objective portion” was found.

In this study, a cohort of 81 individuals who had attempted suicide underwent evaluation utilizing Beck’s Suicide Intent Scale (SIS). All patients were subjected to an investigative process aimed at elucidating the causative factors behind their fatalities. Through the utilization of receiver-operating characteristic (ROC) curves and corresponding tables, optimal cutoff values for both SIS and its variables were ascertained, aiding in the anticipation of suicidal tendencies. Over a span of 9.5 years, seven patients ultimately succumbed to suicide. A key revelation emerged as the mean SIS scores exhibited the capacity to differentiate between those who died by suicide and those who survived. The area under the curve (AUC) stood at 0.74, accompanied by a positive predictive value of 16.7%. Statistically, only the planning scale retained significance. This prompted the selection of four questions to evaluate a streamlined rendition of the SIS concerning its efficacy in predicting suicide. Remarkably, the AUC for this modified version was noted at 0.82, complemented by a positive predictive value of 19%. Acknowledging the inherent complexities in gauging suicide intentions, the study acknowledges the constraint posed by the limited patient pool. Notably, the Suicide Intent Scale emerged as a commendable tool for gauging clinical suicide risk, potentially gaining enhanced predictive ability through a condensed iteration of the scale.

The study was conducted in several parts of the country and covered a variety of sampling approaches and settings [5]. The majority of research endeavors have employed methods such as verbal autopsy (interviewing family members and other acquaintances of the deceased) or amalgamation of diverse data sources pertaining to a demise. These approaches hold widespread acknowledgment for their reliability in categorizing a death as suicide. While certain investigations were embedded within continuous community-based surveillance initiatives, which offered rates aligned with the broader population, alternative studies either omitted the description of their sampling approaches or embraced non-random sample frameworks, thereby introducing challenges in assessing the data’s veracity.

In this research, a closer step has been taken towards a more accurate prediction about risks related to suicide and in the detection of future cases. A visual analysis has also been provided to help understand the new records and analysis for temporary variation in suicide attempt risk. Thus, machine learning algorithms have been implemented for visualizing, analyzing, and predicting the possibilities of suicide attempts in the future.

31.3 Problem Statement

Recent forecasts and analyses have revealed that the ability to anticipate suicide trials has been in jeopardy for decades [10–15]. Due to inaccuracy or poor performance in the past researches, there are not many models who can be sure about their results regarding predictions they have made after analysis. This happens as for a better and precise prediction, there might be a need for complicated mixtures with many hazard factors. In the course of this study, a substantial stride has been marked towards the attainment of scalable and precise suicide risk detection, as well as presented visual analysis to aid comprehend the new data and analysis for transitory changes in the suicide attempt risk. In addition, the goal is to gain a better knowledge of suicide rates and compare the performance of current machine learning algorithms with older methodologies. Thus, machine learning algorithms have been implemented for visualizing, analyzing, and predicting the possibilities of suicide attempts in the future.

31.4 Proposed Methodology

Support Vector Machine

The Support Vector Machine (SVM) emerges as a pivotal supervised learning tool, adeptly addressing both classification and regression quandaries. At its core, the SVM algorithm strives to discern an optimal line or decision boundary, thereby partitioning n-dimensional space into distinct classes to accommodate forthcoming data points effectively. This discerned demarcation, often referred to as a hyperplane or best decision boundary, hinges on the identification of critical points or vectors known as support vectors. In our research, the potency of Linear SVM has been harnessed to handle data instances that exhibit linear separability—situations wherein a solitary straight line can distinctly cleave the dataset into two distinct groups. This facilitated the effective classification of the data, leveraging its inherent linear separability.

Linear Regression

Linear regression stands as a fundamental technique within supervised machine learning. It operates primarily in the realm of regression tasks, wherein prediction values are yielded based on independent variables. The utility of regression models lies in their prowess to predict, as well as to uncover relationships and correlations among variables. Employed as a statistical tool, linear regression's purpose revolves around forecasting a dependent variable's (y) value, hinging on an independent variable's (x) value. This approach yields a discernible linear connection between the input (x) and the output (y), thereby resulting in the determination of the optimal line of fit across the data. This process further entails deriving the regression coefficient(s) that work harmoniously to minimize the cumulative error intrinsic to the model. Linear regression's general mathematical equation is given by the equation (31.1)

$$\boldsymbol{y = cx + d} \tag{31.1}$$

where c and d are called the constant coefficients.

- x is the predictor variable.
- y is the response variable.

Multilinear Regression

It is a statistical technique that uses multiple predictor variables to formulate a final target answer. This algorithm's purpose is to simulate the linear relationship between the independent and dependent variables. For proper prediction by the MLR algorithm, the independent variables should not be too highly correlated with one another, observations used for building the model are selected without any bias and randomly, residuals should have a normal distribution, and the line of best fit should be a straight line rather than a nonlinear curve of some sort.

The linear regression formula is provided by equation (31.2)

$$\boldsymbol{y = c_0 + c_1 * x_1 + c_2 * x_2 + c_n * x_n + w} \tag{31.2}$$

- $\boldsymbol{y}$ is the dependent variable
- $\boldsymbol{x_1}, \boldsymbol{x_2}$, and $\boldsymbol{x_n}$ are the explanatory variables
- $\boldsymbol{c_0}$ is the constant value of $\boldsymbol{y}$ intercept
- $\boldsymbol{c_1}, \boldsymbol{c_2}$, and $\boldsymbol{c_n}$ are the slope coefficients for the explanatory variables
- w is the residual term of the model

Lasso Regression

It is an acronym that stands for "Least Absolute Shrinkage and Selection Operator". It is a classification algorithm that uses shrinkage in simple models, including models with fewer parameters. During this procedure, the data values are shrunk towards a center point, like a mean. The tuning parameter lambda governs the intensity of the L1 regularization penalty. The best lambda value should minimize the Mean Squared Error of the test variable. No parameters are rejected when lambda = 0, and all coefficients are eliminated when lambda = infinity, an increasing number of coefficients are set to zero and rejected and bias increases as lambda increases and finally when variance increases, lambda decreases.

Lasso Regression is given by equation (31.3)

$$\min_{\alpha,\beta} \frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i, \alpha, \beta) \text{ subject to } \|\beta\|_1 \leq t \tag{31.3}$$

Logistic Regression

To forecast a dependent data variable, a logistic regression model examines the correlation between one or more independent factors. Logistic regression serves as a statistical analysis approach to predict a data value by drawing insights from previous dataset observations. Within the machine learning domain, logistic regression has gained significant prominence. This method assists machine learning applications in categorizing incoming data through an algorithm rooted in historical data patterns. The logistic regression technique was applied subsequent to the dataset refinement process. The fitting of the logistic regression model will be achieved using the glm() function, which is typically employed to construct generalized linear models.

Logistic Regression is provided by equation (31.4)

$$y = \frac{e^{(c_0 + c_1 * x)}}{1 + e^{(c_0 + c_1 * x)}} \tag{31.4}$$

where y is the forecasted output and c_0 and c_1 are the bias/intercept term and coefficient, respectively; for a singular input value x, each column within the input dataset incorporates a distinct learned coefficient denoted as "c", which originates from the training data and remains consistently fixed.

31.5 Analysis of Results

31.5.1 Data Cleaning

The Suicide dataset contains about 2,37,519 rows and 7 columns. In the dataset, 5 columns are of the character data type, which is "State", "Type_code", "Type", "Gender", and "Age_group", and the other 2 columns consist of the integer data type, which is "Year" and "Total". The Suicide dataset contains some missing data, while some of the attributes of the columns were unclear. Inconsistency in the dataset with respect to States and Union territory names and some unspecified data in column "Type" was found. So, a data cleaning process was done in these data in order to perform better analysis and prediction. The data cleaning process involved renaming the column values for Union territory for better understanding, rephrasing the causes of death in better manner, dropping unwanted state-titles, dropping the data containing "0" values in the column "Total", dropping the unspecified data in the column "Type", and splitting the data frame into a smaller data frame based on the column "Type_code" for better analysis during the machine learning stage. After cleaning the dataset, it contains about 86576 rows and 7 columns.

31.5.2 Data Visualization

The pie chart as seen in Figure 31.1 represents the percentage of suicides divided with respect to causes of suicides. It is observed that most suicide attempts were done under the "Married" category with a percentage of 25.6%, the second highest is the "By Hanging" category accounting for 11.6%, whereas the least number of suicides represented here were committed by Matriculate or Secondary level education students with 6.43%.

In Figure 31.2 the grouped bar graph displays Age group vs. Total number, men have a larger count of suicides as compared to women in all age groups; the numbers for males are as high as 700k in the 30–44 age group range. Women in the range of 15–29 years of age have committed suicide, close to males touching around 600k. Most of the suicides have occurred in the age range of 15–44.

Tree map for the number of suicides with respect to the top 10 States of India shows us that Maharashtra had the highest number of suicides followed by Andhra Pradesh and West Bengal as seen in Figure 31.3. The least number of suicides in the top 10 states can be observed in the states Odisha, Chhattisgarh, and Gujarat.

The box plot as seen in Figure 31.4 represents the suicide count over the years among people with different education levels in India. With the increase in education, the suicide count drops as indicated by the interquartile range of each plot. Diploma, Graduate, and Post graduate students have the least suicide count, while secondary, no education, middle school, and primary school students have a higher suicide count.

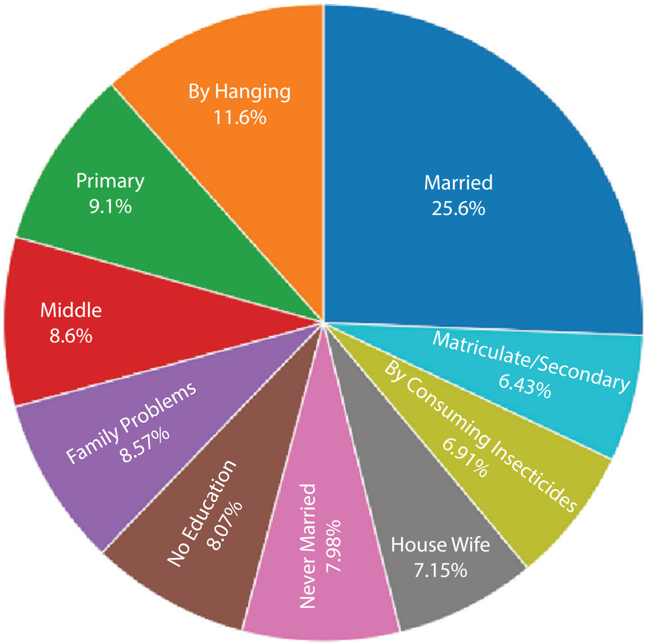

Figure 31.1 Pie chart representing highest percentage of causes.

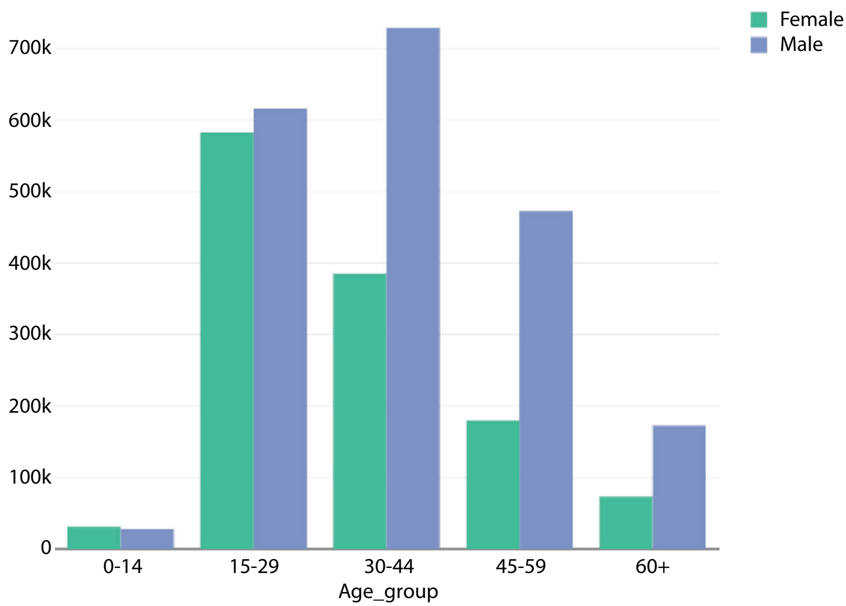

Figure 31.2 Bar chart representing age group vs. total count.

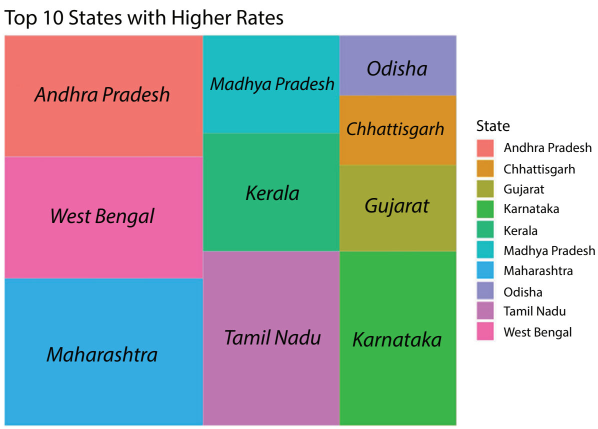

Figure 31.3 Tree map representing state-wise distribution of total number of suicides in India.

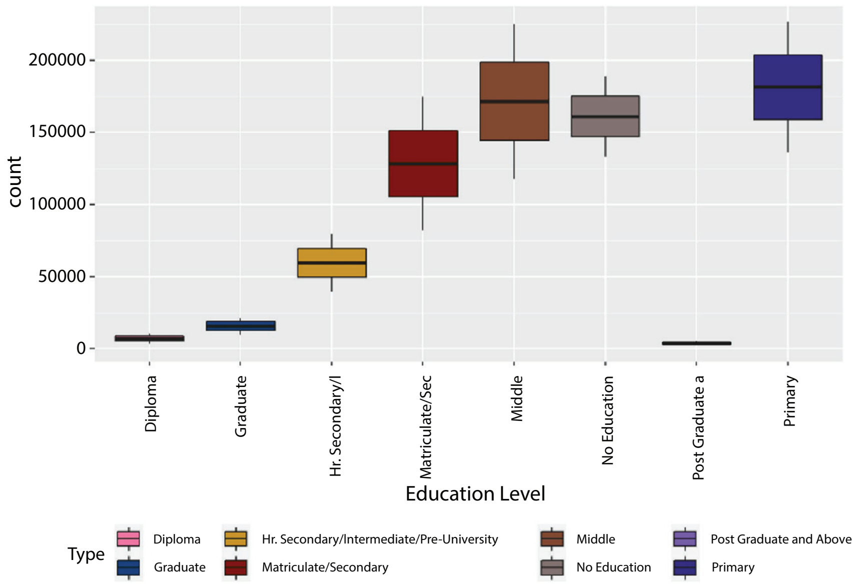

Figure 31.4 Box plot representing education level vs. total count.

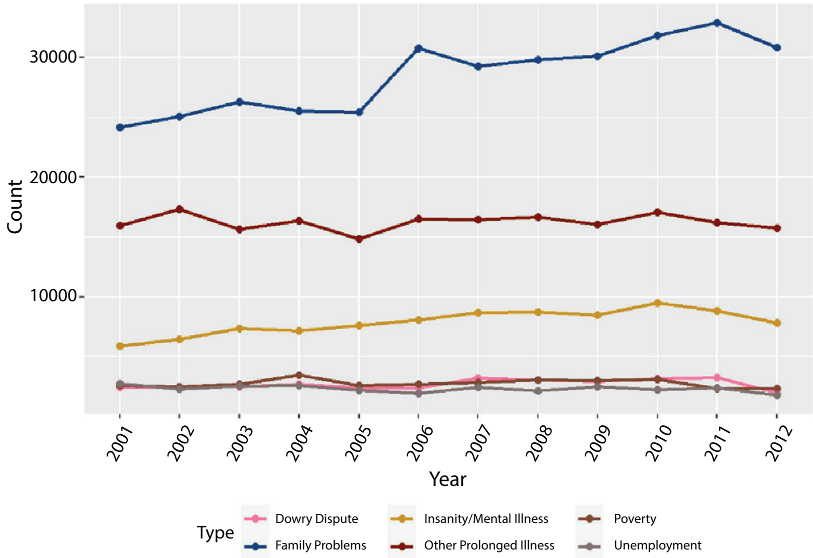

Figure 31.5 Line graph for year vs. total count showing the trends of types.

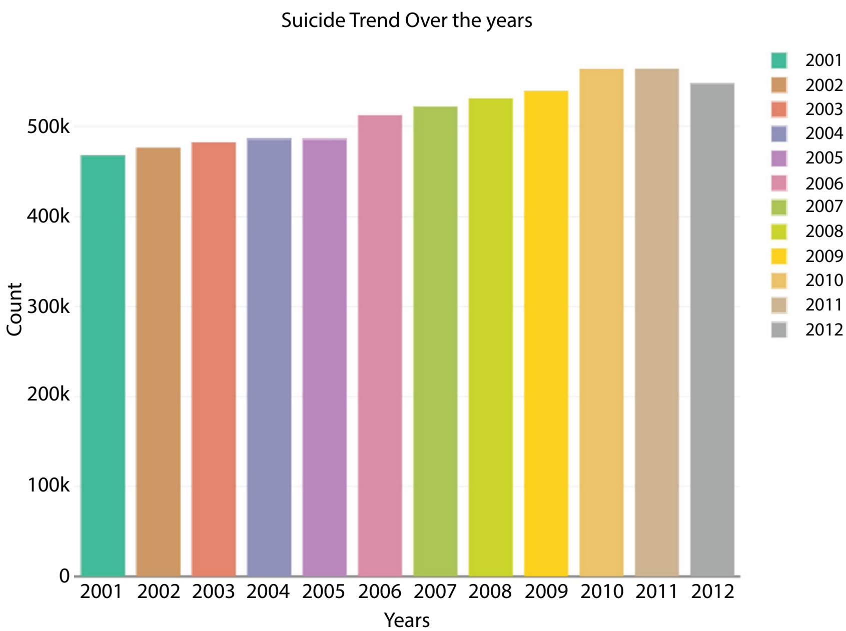

Figure 31.6 Bar graph for suicide trend over the years.

The line graph for Year vs. Total count shows the changes in the number of suicides with year for different types is seen in Figure 31.5. The highest count of suicides has reached 30,000 and above under the "Family problem" type, and it has drastically increased from 2005 to 2006 and was increasing till the year 2011. The trend with least number of suicides can be seen for Unemployment, Poverty, and Dowry; these were always below 10,000 between 2001 and 2012.

The overall suicide count in India is represented in Figure 31.6. It can be observed that there is an increase in suicide count from 2001 to 2011 which is an ominous sign for the upcoming years. For the year 2012, the overall count seems to be lesser than 2011, possibly since the dataset was created before the year got over. At its peak, the count is greater than 500k in 2011.

31.5.3 Machine Learning

Linear Regression for Prediction of Suicide Count from 2013 to 2033

A new data frame was created with Year and the corresponding count of suicides by aggregating it. This was then found to be highly correlated at 0.96. This data frame was then split into training and testing data using the createDataPartition() function. By using the training data, the linear regression model was developed. Here, the total cases are the variable to be predicted while year is the predictor variable. The summary of this model indicates that Year is a highly significant variable with a high t value and a low pr value. The R^2 values obtained are mentioned in Table 31.1. After verifying the accuracy of the model, a prediction was made from the years 2013 to 2033, where the total number of cases increases from 570,195 to 745,528. These values were also plotted in Figure 31.7.

Machine Leaning Models for Age Group, State, and Causes with Respect to Year

After exploration of a linear model with a single predictor and prediction variable, exploration and building of machine learning models were done with two factors taken at a time, namely, age group and year, state and year, and causes and year. Analyzing these factors would give an understanding

Table 31.1 R^2 values for year model.

Prediction dependent attribute	Multilinear regression	Lasso regression
Year	97.53	97.53

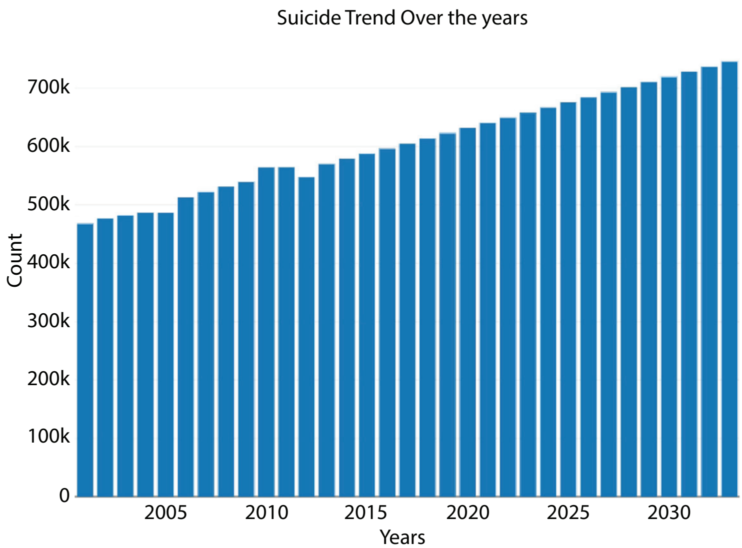

Figure 31.7 Prediction plot for total suicide count from 2013 to 2033.

about the dataset in depth. For each model, filtration was done to obtain a dataset with the total count, year, and the variable factor. After splitting into training and testing data, four machine learning algorithms—multilinear regression, logistic regression, lasso regression, and support vector machine—were applied to the training data. The results obtained and the best performing model in each case are tabulated in Tables 31.2 and 31.3.

Table 31.2 R^2 values for causes and top states model.

Prediction dependent attribute	Multilinear regression	Lasso regression	Logistic regression
Causes, Year	85.75	91.94	91.57
Top States, Year	94.73	92.61	96.35

Table 31.3 R^2 values for age group and culminated model.

Prediction dependent attribute	Multilinear regression	Lasso regression	Logistic regression
Age Group, Year	99.47	92.61	99.47
State, Age Group, Year	99.51	99.50	99.46

A Culminated Model with State, Age Group and Year as Factors
A data frame was created after filtering the total suicide count, with respect to the top 3 states, age group, and year. The value to be predicted here was total suicide count. Using the createDataPartition() method, this data frame was divided into training and testing data. Four models were created using the training data: multilinear regression, lasso regression, logistic regression, and support vector machine model. All these models were trained independently. Table 31.3 shows the comparison among R^2 values. Multilinear Regression performed the best with an accuracy of 99.51%.

31.6 Conclusion and Future Work

The number of suicide cases is escalating year by year. A cause of concern observed during data cleaning is that many suicides are unclassified. Males tend to end their lives as compared to females. People belonging to the age group of 15–44 constitute for most suicide cases. Maharashtra, Andhra Pradesh, and West Bengal are the states having the greatest number of suicides in India. The major factors leading to suicide include family problems, prolonged illness, and insanity. Education is an essential factor in combating suicide, as with the increase in education level, the suicide count drops. The culminated machine learning model indicates that the suicide count varies based on state and age group, but the factor that remains constant is that the suicide count is predicted to go higher and higher every year. Hence, immediate actions must be taken to bring awareness to the general public about mental health issues and ensure that they stay in a healthy state of mind and are able to stay strong when faced with problems so that they never consider the act of suicide.

Future work for this project includes building an even bigger culminated model and testing out other regression algorithms. The dataset used has values only till the year 2012; hence, with an updated dataset, the models, visualizations, and predictions will be even more accurate.

References

1. Agarkhed, J. and Reddy, M., Prediction Model For Preventing Suicide Attempts Using Machine Learning, in: *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, IEEE, pp. 1–6 2021, September. Doi : https://doi.org/10.1109/ICECCT52121.2021.9616815.

2. Walsh, C.G., Ribeiro, J.D., Franklin, J.C., Predicting Risk of Suicide Attempts Over Time Through Machine Learning. *Clin. Psychol. Sci.*, 5, 3, 457–469, Apr. 2017. doi: https://doi.org/10.1177/2167702617691560.
3. Agarwal, T., Dhawan, A., Jain, A., Jain, A., Gupta, S., Analysis and Prediction of Suicide Attempts. *IEEE Xplore*, 451–456, Sep. 01, 2019. https://ieeexplore.ieee.org/abstract/document/8940481 (accessed Aug. 27, 2023).
4. Stefansson, J., Nordström, P., Jokinen, J., Suicide Intent Scale in the prediction of suicide. *J. Affective Disord.*, 136, 1–2167–171, Jan. 2012. doi: https://doi.org/10.1016/j.jad.2010.11.016.
5. Rane, A. and Nadkarni, A., Suicide in India: a systematic review. *Shanghai Arch. Psychiatry*, 26, 2, 69–80, 2014. doi: https://doi.org/10.3969/j.issn.1002-0829.2014.02.003.
6. Poulin, C., Shiner, B., Thompson, P., Vepstas, L., Young-Xu, Y., Goertzel, B., Watts, B., Flashman, L., McAllister, T., Brusic, V., Predicting the Risk of Suicide by Analyzing the Text of Clinical Notes. *PLoS One*, 9, 1, e85733, Jan. 2014. doi: https://doi.org/10.1371/journal.pone.0085733.
7. Ribeiro, J.D., Franklin, J.C., Fox, K.R., Bentley, K.H., Kleiman, E.M., Chang, B.P., Nock, M.K., Letter to the Editor: Suicide as a complex classification problem: machine learning and related techniques can advance suicide prediction - a reply to Roaldset (2016). *Psychol. Med.*, 46, 9, 2009–2010, Apr. 2016. doi: https://doi.org/10.1017/s0033291716000611.
8. Dandona, R., Kumar, G.A., Dhaliwal, R.S., Naghavi, M., Vos, T., Shukla, D.K., Vijayakumar, L., Gururaj, G., Thakur, J.S., Ambekar, A., Sagar, R., Arora, M., Bhardwaj, D., Chakma, J.K., Dutta, E., Furtado, M., Glenn, S., Hawley, C., Johnson, S.C., Khanna, T., Kutz, M., Mountjoy-Venning, W.C., Muraleedharan, P., Rangaswamy, T., Varghese, C.M., Varghese, M., Reddy, K.S., Murray, C.J.L., Swaminathan, S., Dandona, L., Gender differentials and state variations in suicide deaths in India: the Global Burden of Disease Study 1990–2016. *Lancet Public Health*, 3, 10, e478–e489, Oct. 2018. doi: https://doi.org/10.1016/s2468-2667(18)30138-5.
9. Linthicum, K.P., Schafer, K.M., Ribeiro, J.D., Machine learning in suicide science: Applications and ethics. *Behav. Sci. Law*, 37, 3, 214–222, 2019. Doi: https://doi.org/10.1002/bsl.2392.
10. Burke, T.A., Ammerman, B.A., Jacobucci, R., The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *J. Affect. Disord.*, 245, 869–884, 2019. Doi: https://doi.org/10.1016/j.jad.2018.11.073.
11. Coppersmith, G., Leary, R., Crutchley, P., Fine, A., *Natural Language Processing of Social Media as Screening for Suicide Risk*, Biomedical informatics insights, Thousand Oaks, California, USA, 2018, Doi: https://doi.org/10.1177%2F1178222618792860.

12. Cha, C.B., Franz, P.J., Guzmán, E.M., Glenn, C.R., Kleiman, E.M., Nock, M.K., *Annual Research Review: Suicide among youth – epidemiology, (potential) etiology, and treatment*, John Wiley & Sons, Ltd., Hoboken, New Jersey, USA, 2018, Doi: https://doi.org/10.1111/jcpp.12831.
13. Bernert, R.A., Hilberg, A.M., Melia, R., Kim, J.P., Shah, N.H., Abnousi, F., Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. *Int. J. Environ. Res. Public Health*, 17, 16, 5929, 2020. Doi: https://doi.org/10.3390/ijerph17165929.
14. Joshi, M.L. and Kanoongo, N., Depression detection using emotional artificial intelligence and machine learning: A closer review. *Mater. Today Proc.*, 58, 217–226, 2022. Doi: https://doi.org/10.1016/j.matpr.2022.01.467.
15. Ji, S., Pan, S., Li, X., Cambria, E., Long, G., Huang, Z., Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Trans. Comput. Social Syst.*, 8, 1, 214–226, 2020. Doi: https://doi.org/10.1109/TCSS.2020.3021467.
16. Misson, H., Mathieu, F., Jollant, F., *et al.*, Factor analyses of the Suicidal Intent Scale (SIS) and the Risk-Rescue Rating Scale (RRRS): toward the identification of homogeneous subgroups of suicidal behaviors. *J. Affect. Disord.*, 121, 1–2, 80–87, 2010. doi:10.1016/j.jad.2009.05.012.
17. Antretter, E., Dunkel, D., Haring, C., Corcoran, P., De Leo, D., Fekete, S., Hawton, K., Kerkhof, A.J.F.M., Lönnqvist, J., Renberg, E.S., Schmidtke, A., Van Heeringen, K., Wasserman, D., The factorial structure of the Suicide Intent Scale: a comparative study in clinical samples from 11 European regions. *Int. J. Methods Psychiatr. Res.*, 17, 63–79, 2008. https://doi.org/10.1002/mpr.231.