**Please use this template for your paper – this is the title**

Prof.Ajay Talele, Rohan More, Hemant Nandane, Aditya Patil, Rahul Kumare

**Department of Computer Science and Engineering (Artificial Intelligence)**

**Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India**

**ABSTRACT**

**A study is provided with the goal of identifying the primary factors that influenced the suicide rate in specific Indian regions between 2001 and 2012 and then applying those elements to forecast the suicide rate going forward. Making decisions about education and renovation of current facilities, among other things, in the impacted regions can be aided by this suicide prediction. The characteristics of the study mainly pertain to the segment of the populace that experiences the greatest impact from suicides. By establishing suicide prevention centres and focusing on the impacted community with specialized suicide preventive measures like mental counselling, steps can be taken to lower the number of suicides in the areas.**

**For data analytics purposes, the Indian government keeps a public database of all suicide instances that have been recorded in the country's various regions. The demo-graphical characteristics of each of the states under consideration, in addition to the suicide rate, were utilized as features in the model's formulation. The linear model established here may be used to forecast the number of suicides for any given time in the future. This linear model can be used to assess the efficacy of various policy decisions as well as suicide preventive initiatives implemented by NGOs or the government.**

**Keywords—Suicide, Linear Regression, causes, Prediction**,

1. Introduction

In today’s world because of various reasons number of suicide increases day by day. India saw 475 suicide per day in year 2023. WHO(world health organization) reports that 17% of the worldwide suicide victims are residents of India. The results of WHO's multisite intervention study on suicidal behaviours on suicide prevention have shown that early identification and assistance of those who are at risk can reduce the few NGOs works on that. Under Section 309 of the Indian Penal Code, anyone who attempts suicide or does any action that could lead to the commission of such an offense faces a simple sentence.   
 It has been determined that legally prohibiting suicides with a sentence of up to a year in jail, a fine, or both is ineffective. Hospitals and physicians frequently refuse to provide emergency care to those who have tried suicide, citing potential legal issues. This is a serious issue that the nation is dealing with, and the top authorities are ignoring it.   
Identification of persons who are at risk of suicide attempts is the first stage in the public health approach to suicide prevention, according to the United States National Strategy for Suicide Prevention. The purpose of this essay is to pinpoint the various demographic segments that have a major impact on the suicide rate in a certain area. Next, these features are put to use.  
  
to estimate the future suicide rate for the specified area. This is accomplished by analyzing the Pearson correlation between the various population features and the frequency of suicides. Next, using linear regression, a linear predictive model is created by identifying the features that have a significant linear association.  
Research has been done to forecast the suicide rates of several demographic groups, including women, teenagers, and students. For this prediction, numerous algorithms have been proposed. Grey Theory application has been suggested by Mondal and Pramanik to forecast the number of suicides in women.

II. Dataset relation

The purpose of this paper is to identify state-level characteristics that are correlated with suicide rates. To do this, a collection of state characteristics that affect the amount of   
in the state's suicide rate.   
It is plausible that an individual's marital status has an impact on suicide. People who are bereaved or divorced typically have to make significant life changes. If they can't handle these changes, they can make an attempt at suicide. Thus, the proportion of married classes within a community may impact the total number of suicides within that population. Marital status datasets have been taken into consideration for every state based on this intuition. The attributes are broken down by age and include the number of married men, the number of divorced women, and 13 additional features. Information about the Ages 15 to 29 are taken into account for the research. Everywhere in the globe, education is crucial.   
People with education typically have more social maturity. However, those who lack literacy may experience personal issues as a result of to a lack of educational resources. Any worsening of these mental health issues may result in a suicide attempt. This intuition suggests that the number of individuals in various educational classes can be used to estimate the number of suicides that will occur in that group.   
The number of individuals in each literary category is included in the educational status databases.

III personal correlational ad modelling

1. *Pearson Correlation*

A method for determining the relationship between two sets of data is correlation; the Pearson correlation, commonly known as the Pearson   
The linear relationship is best described by the product moment correlation [8]. There is always a correlation coefficient between -1 and 1. If there are two variables, X (independent) and Y (dependent), then Y is predicted to produce a line with a negative slope when plotted with respect to X. A value of -1 shows that there is a perfectly negative relationship. On the other hand, a value of 1 denotes a perfect positive association, which suggests the creation of a straight line with a positive slope.   
When X and Y have no association, it can be assumed that X and Y vary independently of one another.

1. *Modelling regression*

The statistical technique known as linear regression is used to model the connection between a dependent variable an done or more separate variables [9]. The procedure is known as Simple Linear Regression when dealing with a single independent variable and Multivariate Linear Regression when dealing with several independent variables. By (1), the variables are modelled.   
where x = x1, x2, x3, xn, and so on are the independent variables, and n is the number of independent variables, and y = f(x) + c (1) is the dependent variable.   
By using the traditional ordinary least squares method to the significant features, a linear model was created to predict the number of suicides. OLS is a statistical technique for estimating a linear regression model's unknown parameter. This involves reducing the sum of squares of the deviations between the answers that were observed in the dataset and those that a linear function of a number of explanatory variables anticipated. In this work, a linear model was developed using the nine relevant features derived from Pearson correlation. With remarkable precision, this linear model can forecast the number of suicides. Figure 3 presents the model's partial regression plot. The model's fit to the plot is quite good. Near the prediction line is where all the features are located. Sections before this one go into greater detail regarding the model's subtleties and outcomes.

1. *Results*

*Matlab* was used to do the Pearson correlation calculation and OLS regression implementation. Table II provides an overview of the findings.   
The R-squared is the first value in Table II. The statistical measure of the data's proximity to the fitted regression line is called R-squared [10]. It is sometimes referred to as the multiple determination coefficient for multiple regression or the coefficient of determination. By (2), it is defined. R2 is equal to 1 . residual sum squares   
sum total of squares(2) The sum of squares representing the differences between the actual and anticipated values is known as the residual sum of squares. The difference between the actual values and the mean of the actual values is represented by the total sum of squares.

**4. CONCLUSIONS**

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science and Engineering, M S Ramaiah Institute of Technology Bangalore for its invaluable support in this study**.**

**REFERENCES**