School of Technology

B.Tech – Computer Science & Engineering (Sem-V)

Report

Suicide Rate Prediction

Data Mining | 20CP306P

Team Number – DMP01

Atharva Patel | 19BCP014

Parva Barot | 19BCP092

Vedant Patel | 19BCP138

Abstract:

Classification and prediction of suicide attempts is important for preventing suicide. The project shows the comparison between Male and Females who have committed suicide over a span of 12 Years and have different occupations. In our project, we have selected a dataset in which suicide numbers are given starting from the year 2001 to 2012 in the state of Gujarat. The linear regression model is used in this project. Result shows the number of suicides in respected years and graph the increase/decrease in the rate of suicides following every year. The project compares the number of suicides that occurred between each year along with other details. In recent times suicides cases have been increasing rapidly due to some major reasons such as depression, study pressure, unemployment, etc. The challenge is to help such people undergoing mental issues.

Table Of Contents:

Introduction:

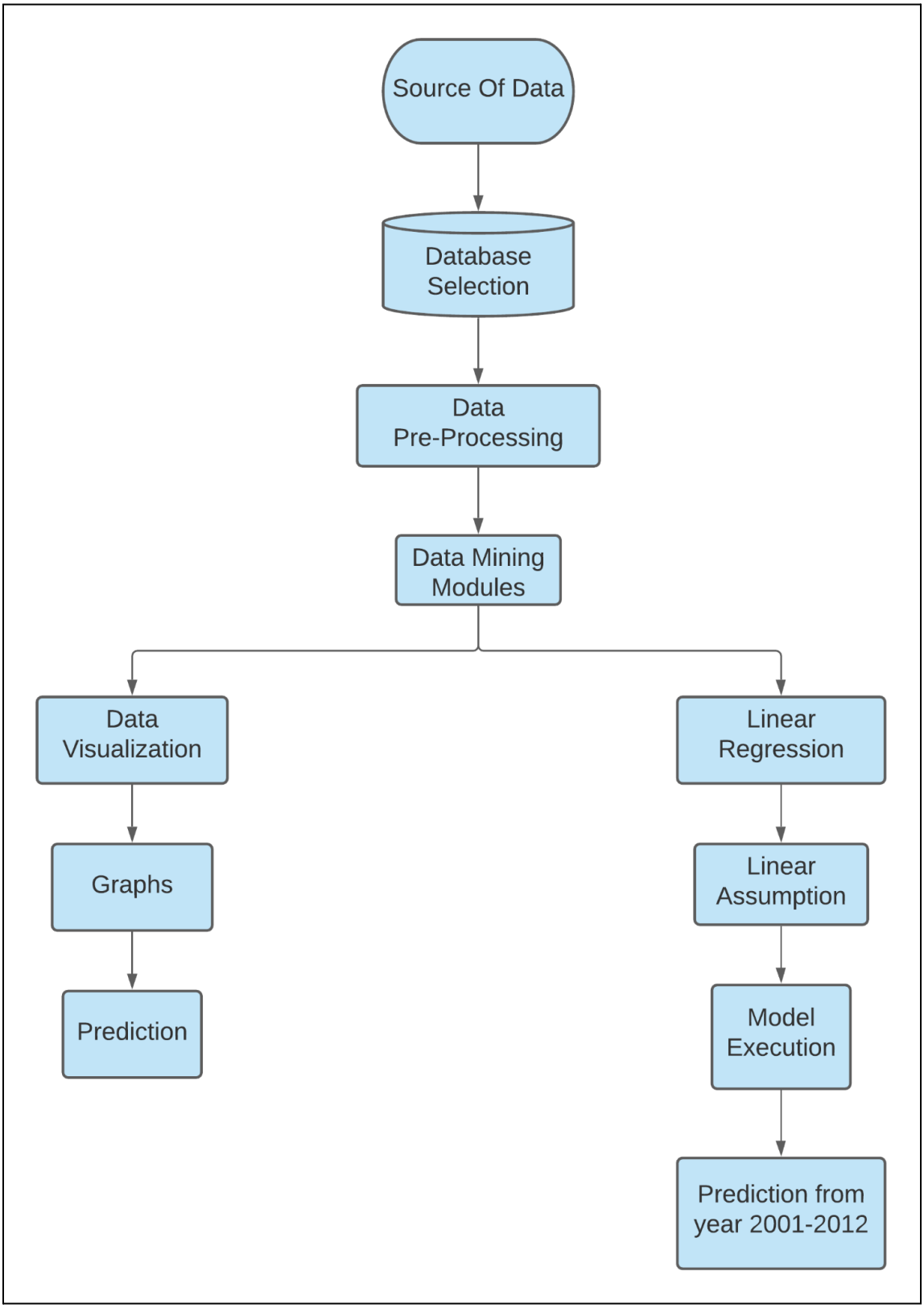
Suicide is a serious public health problem. According to World Health Organization (WHO) every year close to 800 000 people take their own life, i.e. one person attempts suicide every 40 seconds and there are many more people who attempt suicide. Suicide occurs throughout the lifespan and is the second leading cause of death across the world. The main purpose of this project is to bring awareness to people and the government about how suicides are increasing every year due to different reasons and the need to control/take necessary steps. It mainly focuses on the number of deaths caused by suicide attempts in Gujarat state.There are many Societal Problems due to which suicides attempts are done and here we have thrown light on some of the causes for which we can find ways to help people to solve their problems and prevent them from attempting suicide.

Dataset:

The dataset we have used contains the name of the given state. Further, it is classified into two categories Male and Female. The occupation of the deceased's and the years i.e. from 2001-2012 in which the deceased had attempted suicide. The data is collected with help from different websites on google and the overall idea for the dataset was taken from Kaggle.

	state	year	cause	male	female
0	GUJARAT	2001	House Wife	0	1568
1	GUJARAT	2001	Service (Government)	62	8
2	GUJARAT	2001	Service (Private)	365	39
3	GUJARAT	2001	Public Sector Undertaking	92	15
4	GUJARAT	2001	Student	134	130
5	GUJARAT	2001	Unemployed	426	71
6	GUJARAT	2001	Self-employed (Business activity)	255	1
7	GUJARAT	2001	Professional Activity	325	27
8	GUJARAT	2001	Farming/Agriculture Activity	517	77
9	GUJARAT	2001	Retired Person	56	43
10	GUJARAT	2002	House Wife	0	1461
11	GUJARAT	2002	Service (Government)	40	3
12	GUJARAT	2002	Service (Private)	317	33
13	GUJARAT	2002	Public Sector Undertaking	72	8
14	GUJARAT	2002	Student	155	114
15	GUJARAT	2002	Unemployed	467	59

Architecture:



Steps showing how we have coded the actual architecture:

- ➔ Importing necessary libraries like pandas, NumPy & matplotlib.
- ➔ Importing CSV file as a pandas data frame.
- ➔ Finding the values of 'm' and 'c', so for that, we need to find the mean of X & Y values.
- ➔ Calculating m & c using the formulas.
- ➔ Plotting the input points and the regression line.

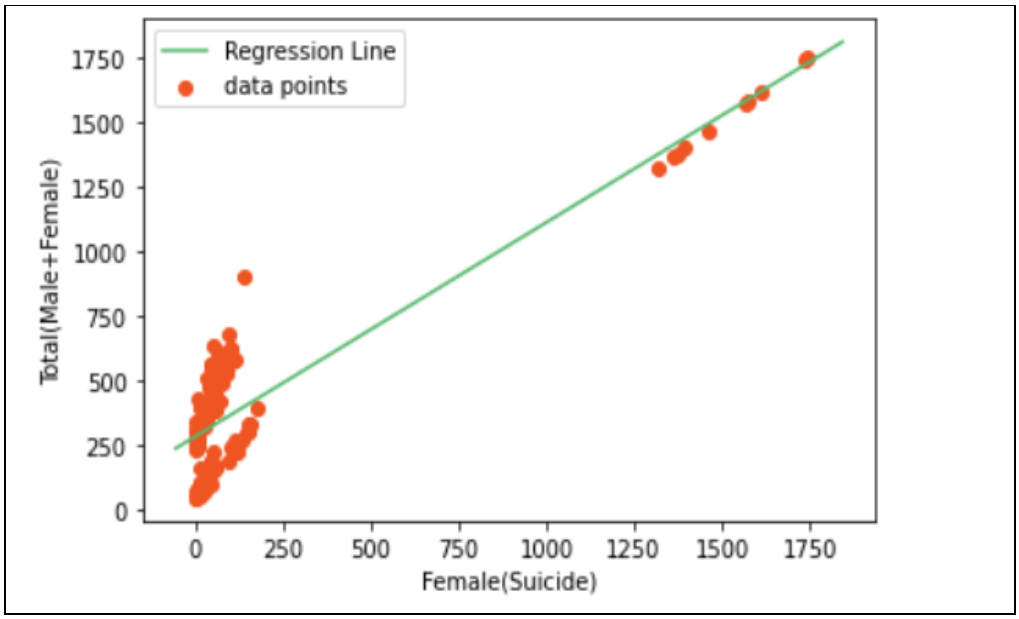
Experiment:

We have used linear regression in our project and also have plotted the bar graphs regarding the total number of suicides every year from 2001-2012 on the basis of gender and their causes.

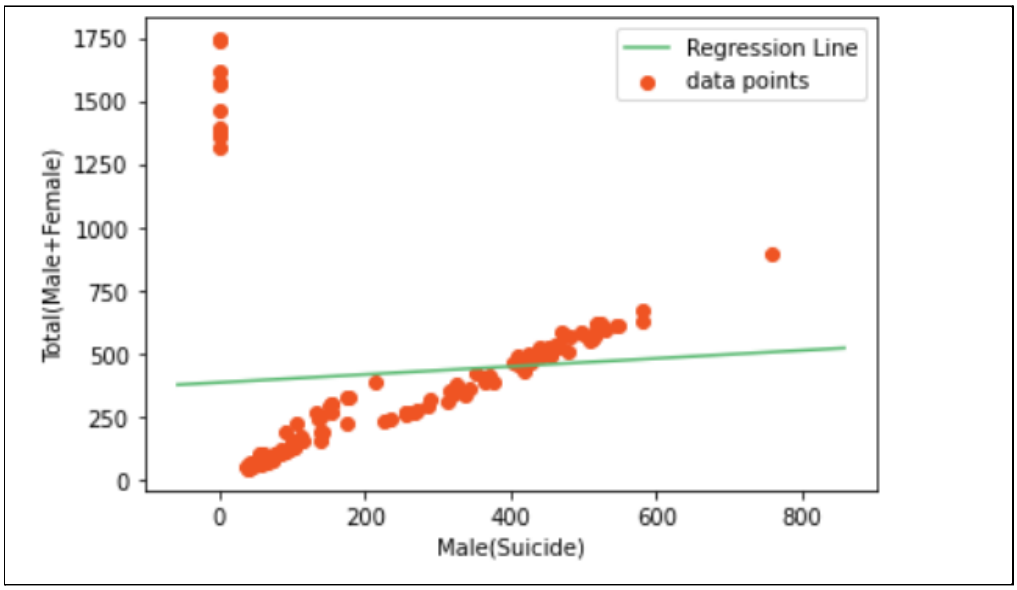
The Python code are available below:
https://colab.research.google.com/drive/1f0v5YZfq4SYfqCik7jV-7NDnpZucMD79?usp=sharing

Result:

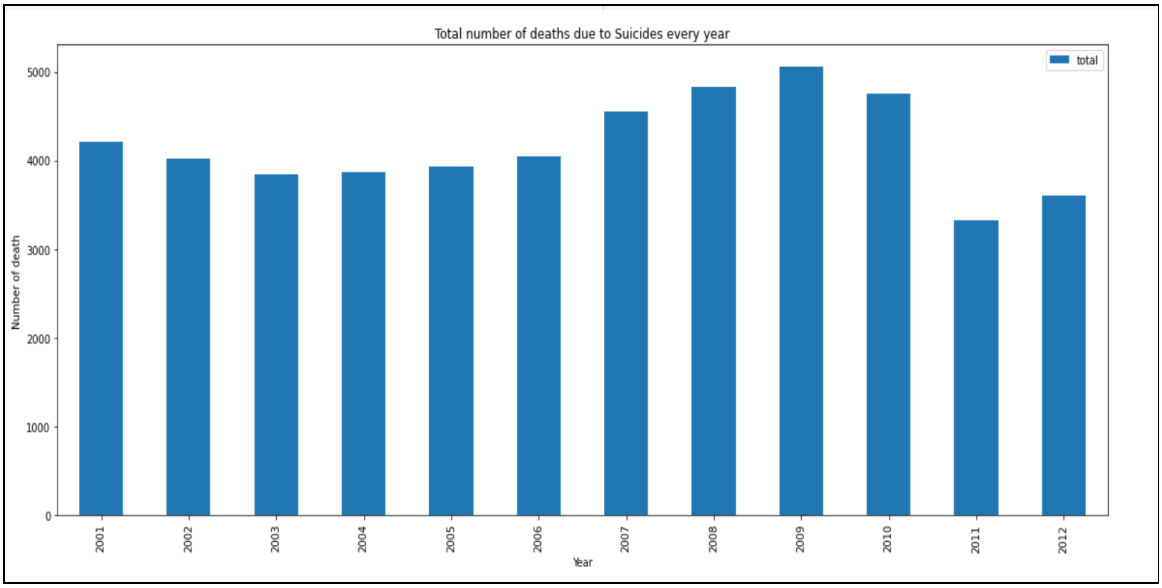
- Using linear regression we have successfully implemented a regression line for both Male and Female deaths.
- The result shows that with time the number of suicides attempts had been increasing for the year 2001-2012.
- The R-squared value for the Female regression graph is 85.909% which is used to determine how good the model is.



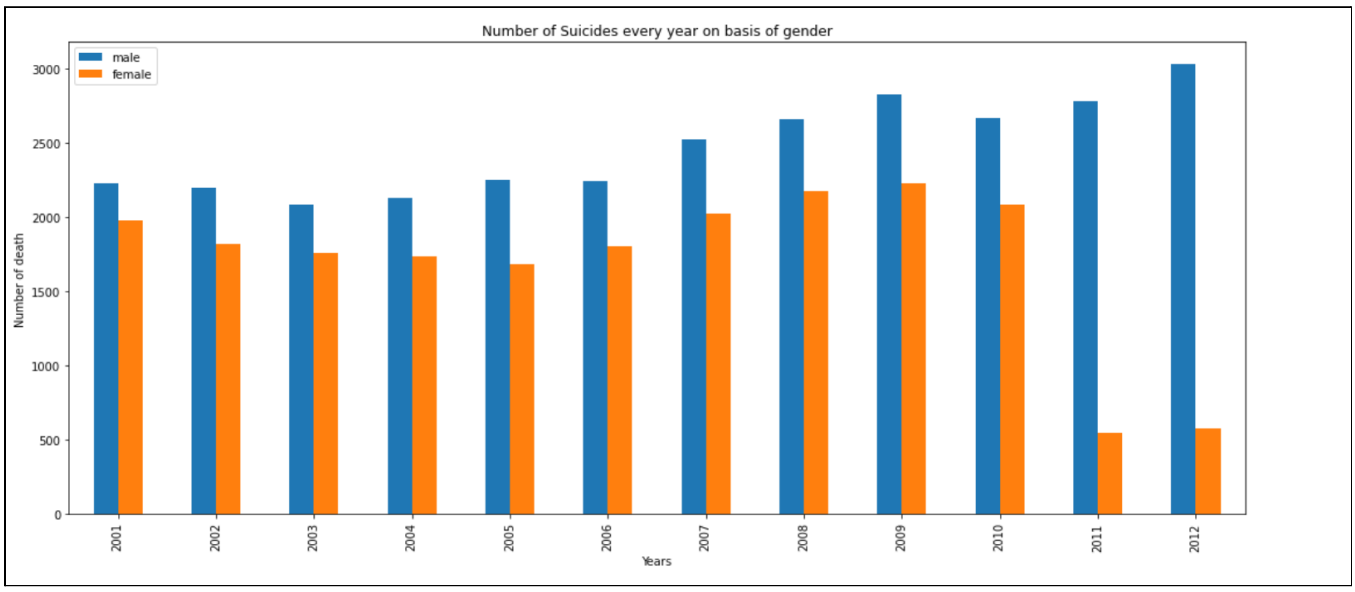
Linear Regression plot for number of female suicides.



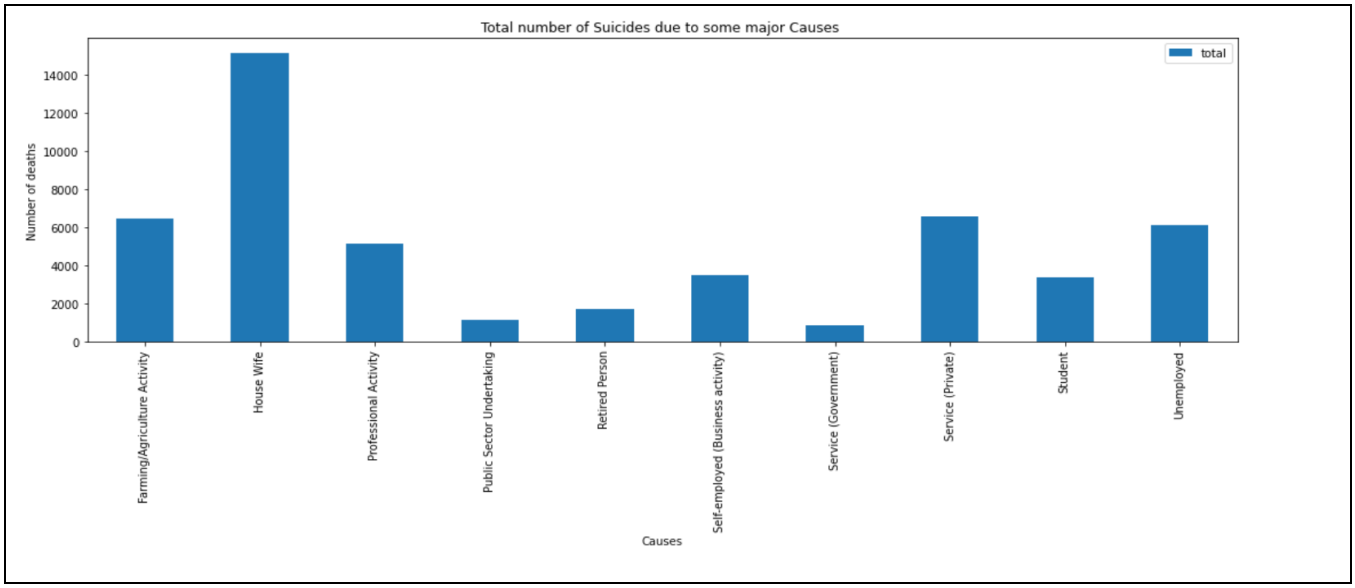
Linear Regression plot for number of male suicides.



Total number of suicides every year



Total number of suicides every year on basis of gender



Number of suicides from 2001-2012 due to some major causes.

Conclusion:

Work done in the project was analyzing the data set concerning the number of suicide attempted by Male and females over 12 Years in the Gujarat state. The number of suicide that occurred have been depicted through the bar graphs above. A Linear Regression model was used to analyze the data.

From graph-plotting and linear regression, it can be highlighted that what are the major causes, responsible for the most number of suicides attempts such as unemployment, farming/Agriculture, etc.

Societal Problems in the project were what causes a person to give up on his/her life. With the help of the project, the major cause can be easily found and can provide help to the person who is suffering from such problems.

Shortcomings in the project are that the linear regression model gives less accuracy i.e. 85%, we can try to use other models that can give more than 90% accuracy for the best predictions so that major causes for the suicide deaths can be highlighted.