Analysis and prevention of suicides using World Health Organization (WHO) data

Ronanki Bharath	Janaki S.	Bhanu Narra
bharath_ramakrishna@srmuniv.edu.in	janaki_sampath@srmuniv.edu.in	bhanu_mohan@srmuniv.edu.in
SRM Institute of Science and Technology, Chennai, Tamil Nadu	SRM Institute of Science and Technology, Chennai, Tamil Nadu	SRM Institute of Science and Technology, Chennai, Tamil Nadu

Rishabh Gupta	S. Rama
rishabhgupta_vinod@srmuniv.edu.in	rama.s@vdp.srmuniv.ac.in
SRM Institute of Science and Technology, Chennai, Tamil Nadu	SRM Institute of Science and Technology, Chennai, Tamil Nadu

ABSTRACT

Youth committing suicides is the major public health issue in the world. It is the third leading death reason in the United States of America for ages between 13-18. Many adults who have suicidal thoughts or make a decision of committing suicide plan never approach any professional for help or advice. Most of the cases, people attempting suicide don't express it a lot in person due to the judgmental attitude of society. We have come with an idea of using WHO published data on suicide. With this we evaluate results to recognize and reduce the suicidal risk of the age group committing more suicides. We can use the same data to target the right audience through social media and campaigns to prevent it and reduce the cases. Linear regression and Random Forest algorithms are used which can predict the possible cases of suicides. We included a prediction algorithm called fbprophet. Once the analysis of the dataset is completed, we visualize the data and the group of audience committing more suicides will be targeted on social media. In the part of prevention, the audience is targeted on Facebook by using a parameter like the age, sex, and region. For this Facebook provides a paid medium through which we can provide awareness and motivation for targeted people.

Keywords— *WHO Data, Prediction, Fbprophet, Suicide, Linear Regression, Random Forest*

1. INTRODUCTION

Every year around eight lakh people are trying to kill themselves and even more who are trying to commit suicide. Most of the families of the people committing suicides are getting affected by this. In a survey, suicide was the second most cause of death among 15–29-year old throughout the world in 2016. Suicide doesn't just occur in countries with high-income, it is a global phenomenon that occurs almost in all the places in the world. In fact, more than 80 per cent of the suicides occurs globally in low-class & middle-class countries in 2016. Suicide is a serious problem for public health [2]. However, suicides are preventable on the evidence-based. Suicides can be prevented. There are many steps that can be taken on the population as a whole & at the individual level for preventing the suicides and suicide attempts. There are people who attempted suicide more than once, these are the cases the patient loses hope and gets back to the suicidal condition.

There are many causes that make people to do suicide:
- **Mental disorder:** Nervousness, Bipolar illness, Depression, Hallucination
- **Excruciating Experience**: Physically abused, sexually abused, War
- **Addiction to the drug:** People addicted to drugs are more common in going into a depression. Most of the people use drugs to escape the pain. This addiction increases the feeling of depression.

We cannot track every individual and read their minds, therefore we perform a general analysis finding the pattern in which the suicide occurs. Suicides occur and we try to prevent them but we lack the system for the prevention. So as a proactive method, we have prediction algorithms that are used to predict the future pattern of the suicides and based on which we target the right audience and save their life. The targeting would treat the psychological systems of clinical depression and improves the ability of persons who consider committing suicide to deal effectively and positively with life and other hardships. [1].

2. EXISTING SYSTEM

In this paper, there are two cases in suicidal risk assessment which are examined and checked for differences. The first segment which is interview segment consists of five open-ended questions and the answers are tested whether they are having the potential to arrange non-suicidal patients with suicidal repeaters and suicidal repeaters with non-suicidal repeaters between the age of 13-18. When significant differences are resolved from data obtained after the study of the raw data, they are used in automatic classification experiment. Verbal and non-verbal behavior of clinical-patients. To differentiate between people committing suicides repeatedly acoustic information features will be useful. By doing this highest accuracy can be achieved. Altogether the paper stated that the question and answer part of the interview can help estimate the suicidal risk of adults. It results with the analysis of the suicidal with non-suicidal & the suicidal with repeatedly suiciding. This paper showed that the ubiquitous question and answer in the interview part can help to estimate the suicidal risk of adults. The verbal behavior of clinicians and the young patients are also important as well as estimating the potential of re-attempting suicidal non-verbal behavior information, particularly the acoustic features is also important for the patients. In the differentiation of the suicidal and the non-suicidal patients the UQ case unveils favorable results with an accuracy of 88.3%. Things lacking in the referred paper are, in certain situations regression testing is taking, Targets only the youth, need physical interaction with the patient, has to record all the patient's encounter.

3. PROPOSED SYSTEM

In this system, we are analyzing the suicide dataset taken from WHO which will help in preventing suicide cases. Most of the category committing suicides are teenagers and adults, hence we are targeting audience in Facebook which will make easy to reach more people at less time. The dataset was taken from WHO will contain many duplicate values and missing values this can cause an error to the result thus we do data wrangling. After cleaning the data, the dataset is used for analysis. The analysis is done by plotting graphs between parameters like year vs sex, age, and region. we use linear regression algorithm which can predict the possible cases of suicides. Linear Regression algorithm is based on supervised learning. Also, fbprophet is used to predict the suicides rate in the future years based on the past rate. Our paper's novel approach is using the Random Forest algorithm which obtains results more accurate than linear regression.

This dataset is used to target the right audience through social media targeting. For Facebook targeting, we used the analyzed data using which we advertise motivational quotes to targeting group of people based on parameters we got after analysis. Using linear regression, we perform regression tasks. Based on independent variables regression models a target the prediction value. We used it for finding out the relation between data values and statistics.
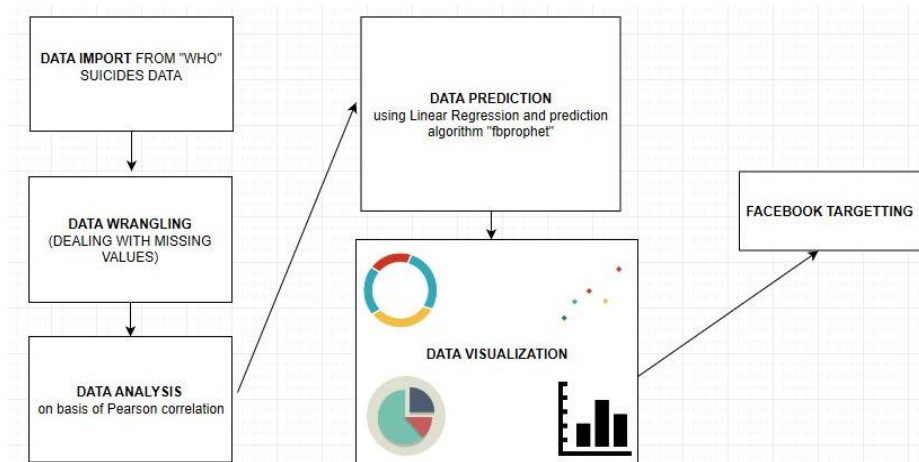

Fig. 1: System architecture

Things that we overcame in our paper are, target all age groups, Fbprophet and Random forest for an accurate result, no need any physical interaction or record any conversation.

4. MODELLING

4.1 Correlation

Correlation is a technique of measuring how to relate two sets of data. Based on the dataset we have plotted the country and age parameter in Y-axis and the number of suicides in X-axis.
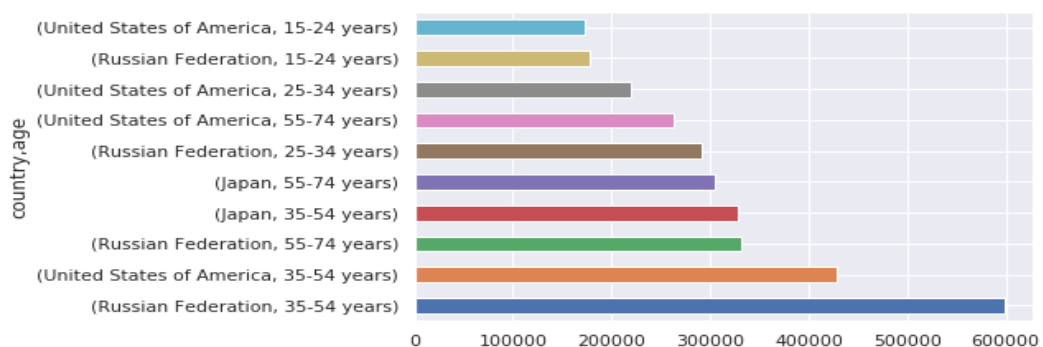
Fig. 2: No. of suicides vs. country and age

4.2 Linear regression

Linear regression is a machine learning algorithm. Basically, it is based on superintended learning and does tasks related to regression tasks. Mostly it is used for finding out the relationship between values and statistics. Linear regression executes the tasks that predict the value of dependent variable let it be(p) based on the hypothesis given independent variable let it be(q). So, linear regression is used to finds out the linear relationship between q (input) and p(output) and being a regression task, it is said to be linear regression. We compare in our project between the independent variable which is a year and predicting the dependent variables which are age, year, gender etc.

Table 1: Result summary for linear regression

MSE	469681.7148632005
RMSE	685.3332874326188
R^2	0.4289879644224436

R-squared in the above table is the coefficient of determination. It is defined as (1)

$$R^2 = 1 - (\text{residual} \sum \text{of squares}/ \sum \text{of squares})$$

The residual $\sum$ of squares is the difference between the original and predicted values. Total $\sum$ of squares is the difference between the original values and the mean of the original values. The value of R^2 lie between 0 and 1. 1 indicates that the line of regression fits the data perfectly. As the R^2 value is 0.428, it indicates that the line of regression has a low accuracy of fit.

4.3 Random Forest Regressor

Random forest is an ensemble technique which is capable of performing both regression and classification tasks together by using the multiple decision trees. The basic idea of using this in the project is to combine the multiple decision trees achieved after processing the data that determines the final output rather than depending on individual decision trees.

Table 2: Result summary for random forest regressor

| MSE	146476.54759546957
RMSE	382.7225464948068
R^2 SCORE	0.8219222316729469

4.4 Fbprophet algorithm

The prophet is a procedure or an algorithm that forecasts data of time series where trend fit with yearly, weekly and daily seasonality. It is accurate and fast. It performs better than any other approach. Forecasts are retrieved in just a few seconds.
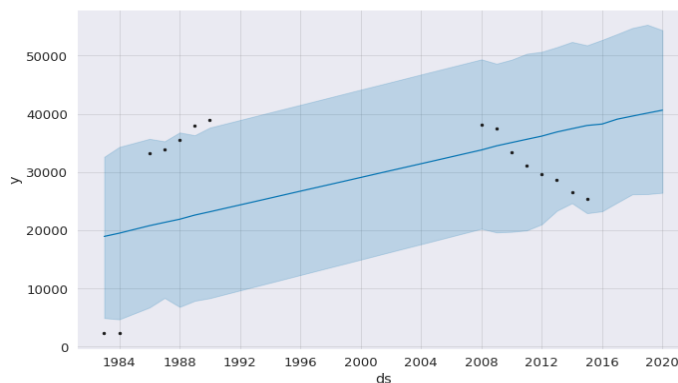
Fig. 3: Predicted suicide number for future till 2020

4.5 Facebook targeting

The result from the above analysis is taken and we have targeted the audience through social media Facebook. This is a premium account where we can target the right people to save them from attempting suicide.
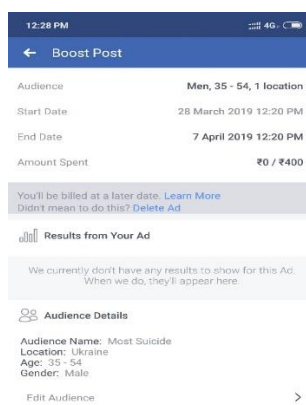
Fig. 4: Facebook targeting the right audience

5. CONCLUSION
Thus, in this paper, we find the group of people who are more prone to suicidal activities through analysis of the data acquired from WHO and help them in motivating themselves. This help to reduce suicidal activities.

6. ACKNOWLEDGEMENT
The authors gratefully acknowledge Prof. S. Rama from the SRM Institute of Science and Technology for her contribution in assisting us with the design development of the system and making her available for guidance in creating this work.

7. REFERENCES
[1] https://en.wikipedia.org/wiki/Suicide_prevention
[2] https://en.wikipedia.org/wiki/Suicide_in_India
[3] Assembling and Using a Cellular Dataset for Mobile Network Analysis and Planning by Paolo Di Francesco, Francesco Malandrino, Luiz A. DaSilva
[4] Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics by Anindya Ghose and Panagiotis G. Ipeirotis, Member, IEEE
[5] Analyses of the Clustering Coefficient and the Pearson Degree Correlation Coefficient of Chung's Duplication Model by Duan-Shin Lee, Cheng-Shang Chang and Hao-Neng Chang
[6] Regression Analysis of Automatic Measurement Systems by Michael J. Flynn, Member, IEEE, Shahram Sarkani, and Thomas A. Mazzuchi
[7] https://en.wikipedia.org/wiki/Random_forest