An Efficient Suicide based Dataset using Machine Learning Algorithms

Sidharth raj. R.S, A. John Paul Praveen, S. Balaji

Abstract-A study is presented on analyzing the major factors that affect the number of suicides in different parts of India from year 2000 to 2012 and using them to predict the number of suicides in the future. By analyzing the data and predicting the major causes of suicides it can help government to know which part of population is most affected, so that the government can provide required steps to avoid suicides. The Indian government records the database of each suicide occurs in India. Along with the age-group, cause of death, state of victim, this data was made public by crime branch bureau of the data analytics purpose. Relationship will be made between the different features of suicide so that a linear relationship can be formed with the help of linear regression and other machine learning algorithms will be used to develop a model for the prediction of number of suicides in the future. It has been found that the results obtained by machine learning algorithms are more accurate when compared with the traditional algorithms.

KEY WORD– Suicide, Analysis, Linear regression, Machine Learning Algorithms

I. INTRODUCTION

Suicide is one of the serious issues that Indian government is confronting. Around 800000 individuals end it all worldwide consistently. of these 17% are occupants of India. The nations shares the worldwide number of suicide passings has been rising, both for people. Now is the ideal time India to handle suicide ,the quiet executioner ,as a general wellbeing emergency .The male to female suicide proportion has been about 2:1 in India. On a normal an all out number of suicides in India every day are 300. As per the Suicides reports in India and National Crime Records Bureau, the absolute number of suicides in India according to 2014 insights is 1, 09,456.

In India the suicide rates expanding [1] step by step .as financial occasions posted that India has 2.3 lakhs passings per year.it expanded from 1 lakh to 2.3 lakh presently. Information discharged as of late by the online diary Lancet Public Health, as a component of its Global Burden of Disease Study (1990 to 2016), uncovered, just because, the quantum of the issue India is confronting: in 2016, it had the most astounding number of suicide passings. The a lot of the

* Correspondence Author

Sidharthraj. R.S, Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India. E Mail - sidarthraj93@gmail.com

A. John Paul Praveen, Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.
 E Mail - johnpaul.embedded@gmail.com

S. Balaji, Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India. E Mail - bala.sripathy@gmail.com

worldwide number of suicide passings has been rising, both for people. One of each three ladies kicking the bucket by suicide, the report uncovered, is from india.

The Indian Government classifies a death as suicide if it meets the following three criteria:

- It is an unnatural death
- The intent to die originated within the person, there is a reason for the person to end his or her life.
- In most cases the person writes the reason in a suicide note or it remains unspecified.

There are various reasons due to which a person may commit suicide. It may be due to education burden, family problem, financial problem, health status, etc. There may be certain states where particular reason for suicide is higher than other reason. We need to find relation between reason of suicide and the state in which it is committed. So that government will focus on particular reason resulting in minimizing that factor of suicide. For example, if in particular state the maximum suicides are committed by farmers due to low agriculture production or higher input and lower output [2] in agriculture than the government can focus more on particular field thus making more schemes that would help farmers of particular state resulting in reduced number of suicides in that state. Another feature is age group.

The dataset is divided in different age1 groups. So that we can find relationship between age and number of suicides which helps us to know that particular age group population is mostly affected by which suicide reason. For example, if in particular age group the major reason of suicide is failure in examination than more counselling institutes should be open in that state to do counselling of all people belonging to that age group resulting in reduction of suicides[3]

Suicide avoidance is a term for the aggregate endeavors of nearby resident associations, wellbeing experts and related experts to lessen the rate of suicide. Other direct mediations to stop a looming suicide, techniques additionally include treating the mental side effects of despondency, giving advising to the individual, improving the adapting methodologies of people who might some way or another genuinely think about suicide, diminishing the pervasiveness of conditions accepted to establish hazard factors for suicide, and giving individuals trust in a superior life by settling current issues. The first venture in general wellbeing way to deal with suicide anticipation is to distinguish the individuals who are at the danger of suicide endeavors. To distinguish this that which populace is at more serious hazard this investigation would be valuable.

This should be possible by making a connection between various highlights[4] and the quantity of individuals submitting suicides. The

relationship can be utilized to build up a straight relationship so number of suicides because of a specific reason specifically age bunch inside a specific state can be anticipated.

II. LITERATURE SURVEY

Dmytro Karamshuk, Francesshaw, Julie Brownlie, Nishanth Sastry

With the ascent of online life, a huge measure of new essential research material has turned out to be accessible to social researchers, however the sheer volume and assortment of this make it hard to access through the conventional methodologies: close perusing and nuanced understandings of manual subjective coding and investigation. This paper embarks to cross over any barrier by creating semi-robotized swaps for manual coding through a blend of publicly supporting and AI, seeded by the improvement of a cautious manual coding plan from a little example of information. To demonstrate the guarantee of the methodology, we endeavor to make a nuanced order of reactions on twitter to a few late prominent passing by suicide. Through these, we demonstrate that it is conceivable to code naturally over an enormous dataset to a high level of precision (71%), and examine the more extensive potential outcomes and entanglements of utilizing Big Data strategies for sociology.

By using this method, we are not able to analysis large suicide death rates. It shows drawback for uneducated people and those who are not using social media. we have seen high suicides in adults.so the students under 16 or 18 ages may not have access to mobiles and also so many farmers in our India uneducated but they are most intelligent persons most of the farmers are not active in social media.as we are seeing increasing farmers death.so we can't able predict in these group of people.

Sachil kumar , Anoop K.verma, Sandeep Bhattacharya, Shiuli Rathore

Creators of this Sachil Kumar ,Anoop K. verma, Sanded Bhattacharya, Shiuli Rathore have taken just lucknow ,Uttar Pradesh suicide dataset. There study presents assessing suicide cases in Luck now ,capital of Uttar Pradesh and probably the biggest city in India, between Jan 2008 and October 2012.There were 5204 cases with 2946 male and 2258 female exploited people, between ages 20 and 60 years. Suicide rates extended structure 21.55 to 24.23 per 100,000 populace. The technique for suicide was distinctive among male and female exploited people, as male unfortunate casualties attempted to utilize more vicious strategies than females. The study demonstrated that suicide rates have expanded.

III. EXISTING SYSTEM

Analysis of Suicide Victim Data for the Prediction of Number of Suicides in India (Publisher IEEE 2017). Suicide Victim Data was analyzed for the Prediction of Number of Suicides in India. Linear regression algorithm was used for prediction. Data of different places was used for prediction. Data set used was of made public by crime branch bureau.

Drawbacks of Existing System

The existing system has considered only 12 states and given results on basis of them. Union Territories were not taken into consideration during study. Age group between 15-29 was only considered for study. No proper simulation and results were explained. Only 2011 data has been used for prediction which is less efficient. [5]

THE EXISTING SYSTEM CONSIDERED THESE STATES
- Maharashtra
- Kerala
- Gujarat
- Rajasthan
- Bihar

IV. PROPOSED SYSTEM

The proposed system considers all the states and even union territories also.It considers all Age Groups of the people.[6]

Advantages Of Proposed System

As we considered all states and union territories, we can prevent suicides in all over India. It helps the people in all over India.[7]

The Proposed System Considers Some Extra Places Too
- Chhattisgarh
- Odisha
- Assam
- Haryana
- Delhi (UT)
- Jharkhand
- Tripura
- Pondicherry
- Himachal Pradesh
- Uttarakhand
- Goa
- Jammu & Kashmir
- Sikkim
- A & N Islands
- Arunachal Pradesh
- Meghalaya
- Chandigarh
- Mizoram
- D & N Haveli
- Manipur
- Nagaland
- Daman & Diu
- Lakshadweep

V.MODULES SPECIFICATION

A. Data Visualization:

Data visualization plays a very important role in data analytics .data visualization means data is represented in graphical or pictorial form. Our human brain understands information very easily and effectively when data is shown as graphs. [8]
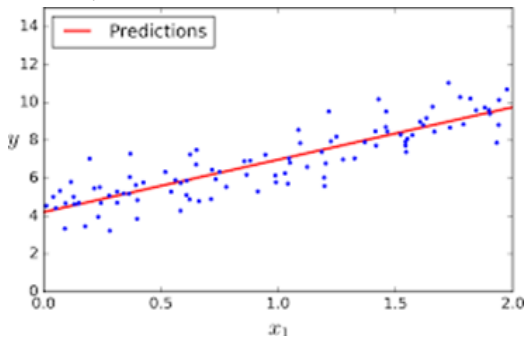
B. Linear Regression

"Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent and independent variable". On the core concept of linear regression[9], the data is modeled using a straight line.

➤ Linear regression is used with continuous variable here we find correlation between x and y.
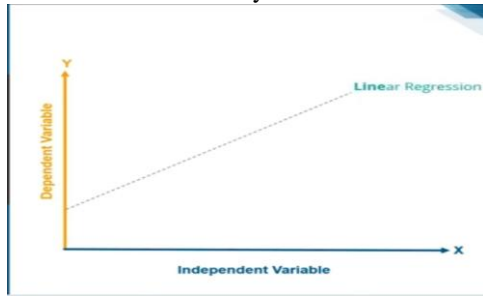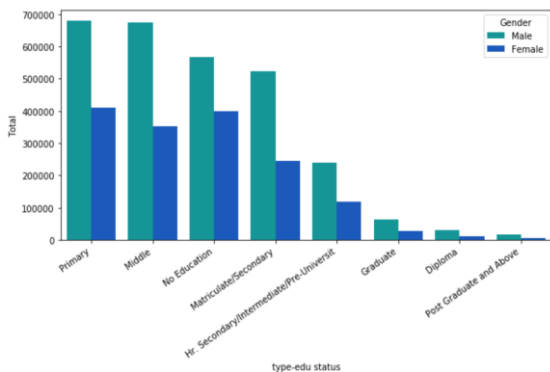

Fig.1 Linear regression

With simple linear regression as shown in Fig1 we can model our data as follow

$y = C + M * X$.

**Fig.2 Linear regression**

VI. RESULTS

Number of suicides appear to be concentrated towards Lower education level. Most of the people who have committed suicides have education level below Matriculate/Secondary. This is represented in fig 4.

Fig.5 Bar Graph representing Number of suicides of different education level people

### A. Number of suicides on social status

Social status also tends to have great affect over suicides. Married people commit more suicides as compared to never married, widower, separated and divorced.
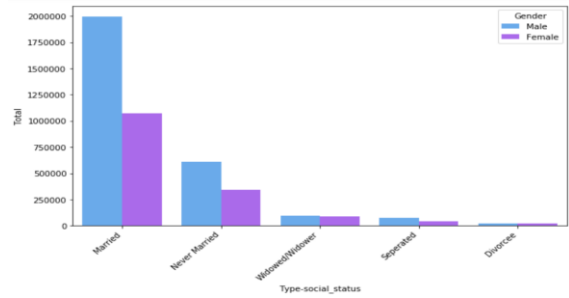
Fig.5 Bar graph representing number of suicides on social status

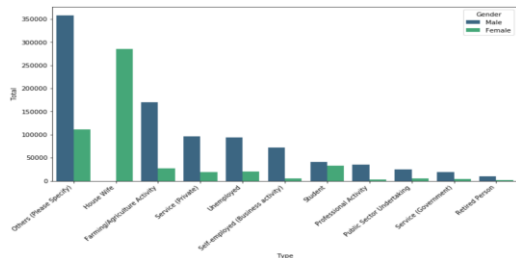### B. Number of suicides because of different causes

While most of the causes of the suicides are not known, the three major causes among the known cases are Family problems, prolonged illness and mental illness. This is represented in fig 6
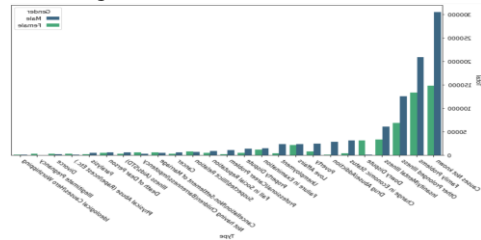
Fig.6 Bar graph representing number of suicides based on causes.

C. Number of suicides because of dowry disputes

According to these data men seem to be badly effected by unemployment, property dispute, poverty, drug abuse or addiction and change in Another major concern is number of women who have committed suicides due to dowry disputes is much higher when compared to men. This is represented in fig 8.

Fig.8 Bar Graph representing Number of suicides because of dowry disputes

According to the data, most of the females who have committed suicides are house wives. The percentage of female suicides in all other categories is much lower than the percentage of male suicides. While this is an indication of lower representation by women in professional careers, it also reiterates the importance of girl education and women empowerment. The graph shows that financially independent women are much mentally stronger.

It is disheartening to see that farmers who feed the rest of the country are the ones who are more committing suicides than any other profession.

Followed by farmers, it's the unemployed and private sector employees who are most effected. It's surprising to know that the number of suicides among the unemployed and the private sector employees is almost the same. This also might be due to higher pressure in private sector when compared to government sector jobs.

Number of suicides of different age group in different states: Based on state and age group we can see 15-29 is the most vulnerable age in all states except Kerala. Maharashtra is the state with most number of suicides.
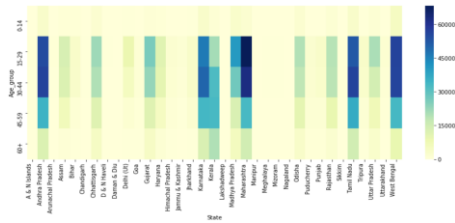

Fig.9 Heat Map representing Number of suicides of different age group in different states

D. Representing total number of suicides increasing with time:

From 2001 to 2012 the percentage increase in suicides is 24.8% which is very scary. This is represented in fig 10.
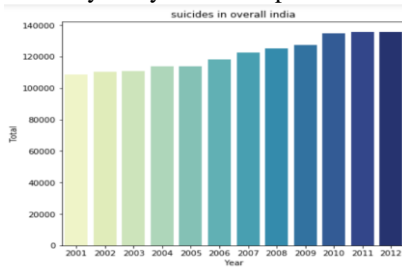

Fig. 10 Suicides over all India

VII. CONCLUSION

The results obtained gives us the clear vision about what type of population is highly affected by this problem. Government should take the preventive measures in bringing down number of suicides in our country by giving more attention on the population that is mostly affected in their respective states. It's not only the government but even it is job of us also to work hand in hand with the government and help them in reducing suicides from our nation by providing counseling to the population of the respective states who are greatly affected by it.

FUTURE ENHANCEMENT

In future by this study for any group of population in given states the number of suicides can be predicted. This can be used as a reference for evaluating the effectiveness of the preventive measures and policies that government took for reduction of suicides.

REFERENCES

1. Kumaravel A., Meetei O.N.,An application of non-uniform cellular automata for efficient cryptography,2013 IEEE Conference on Information and Communication Technologies, ICT 2013,V-,I-,PP-1200-1205,Y-2013
2. Kumarave A., Rangarajan K.,Routing alogrithm over semi-regular tessellations,2013 IEEE Conference on Information and Communication Technologies, ICT 2013,V-,I-,PP-1180-1184,Y-2013
3. Dutta P., Kumaravel A.,A novel approach to trust based identification of leaders in social networks,Indian Journal of Science and Technology,V-9,I-10,PP--,Y-2016
4. Kumaravel A., Dutta P.,Application of Pca for context selection for collaborative filtering,Middle - East Journal of Scientific Research,V-20,I-1,PP-88-93,Y-2014
5. Kumaravel A., Rangarajan K.,Constructing an automaton for exploring dynamic labyrinths,2012 International Conference on Radar, Communication and Computing, ICRCC 2012,V-,I-,PP-161-165,Y-2012
6. Kumaravel A.,Comparison of two multi-classification approaches for detecting network attacks,World Applied Sciences Journal,V-27,I-11,PP-1461-1465,Y-2013
7. Tariq J., Kumaravel A.,Construction of cellular automata over hexagonal and triangular tessellations for path planning of multi-robots,2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2016,V-,I-,PP--,Y-2017
8. Sudha M., Kumaravel A.,Analysis and measurement of wave guides using poisson method,Indonesian Journal of Electrical Engineering and Computer Science,V-8,I-2,PP-546-548,Y-2017
9. Ayyappan G., Nalini C., Kumaravel A.,Various approaches of knowledge transfer in academic social network,International Journal of Engineering and Technology,V-,I-,PP-2791-2794,Y-2017
10. Kaliyamurthie, K.P., Sivaraman, K., Ramesh, S. Imposing patient data privacy in wireless medical sensor networks through homomorphic cryptosystems 2016, Journal of Chemical and Pharmaceutical Sciences .
11. Kaliyamurthie, K.P., Balasubramanian, P.C.An approach to multi secure to historical malformed documents using integer ripple transfiguration 2016 Journal of Chemical and Pharmaceutical Sciences 9
12. A.Sangeetha,C.Nalini,"Semantic Ranking based on keywords extractions in the web", International Journal of Engineering & Technology, 7 (2.6) (2018) 290-292
13. S.V.GayathiriDevi,C.Nalini,N.Kumar,"An efficient software verification using multi-layered software verification tool "International Journal of Engineering & Technology, 7(2.21)2018 454-457
14. C.Nalini,ShwtambariKharabe,"A Comparative Study On Different Techniques Used For Finger – Vein Authentication", International Journal Of Pure And Applied Mathematics, Volume 116 No. 8 2017, 327-333, Issn: 1314-3395
15. M.S. Vivekanandan and Dr. C. Rajabhushanam, "Enabling Privacy Protection and Content Assurance in Geo-Social Networks", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 49-55, April 2018.
16. Dr. C. Rajabhushanam, V. Karthik, and G. Vivek, "Elasticity in Cloud Computing", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 104-111, April 2018.
17. K. Rangaswamy and Dr. C. Rajabhushanamc, "CCN-Based Congestion Control Mechanism In Dynamic Networks", International Journal of Innovative Research in Management, Engineering and Technology, Vol 3, Issue 4, pp. 117-119, April 2018.
18. Kavitha, R., Nedunchelian, R., "Domain-specific Search engine optimization using healthcare ontology and a neural network backpropagation approach", 2017, Research Journal of Biotechnology, Special Issue 2:157-166
19. Kavitha, G., Kavitha, R., "An analysis to improve throughput of high-power hubs in mobile ad hoc network" , 2016, Journal of Chemical and Pharmaceutical Sciences, Vol-9, Issue-2: 361-363
20. Kavitha, G., Kavitha, R., "Dipping interference to supplement throughput in MANET" , 2016, Journal of Chemical and Pharmaceutical Sciences, Vol-9, Issue-2: 357-360

21. Michael, G., Chandrasekar, A.,"Leader election based malicious detection and response system in MANET using mechanism design approach", Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
22. Michael, G., Chandrasekar, A.,"Modeling of detection of camouflaging worm using epidemic dynamic model and power spectral density", Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .
23. Pothumani, S., Sriram, M., Sridhar, J., Arul Selvan, G., Secure mobile agents communication on intranet,Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg No S32-S35, 2016
24. Pothumani, S., Sriram, M., Sridhar , Various schemes for database encryption-a survey, Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg NoS103-S106, 2016
25. Pothumani, S., Sriram, M., Sridhar, A novel economic framework for cloud and grid computing, Journal of Chemical and Pharmaceutical Sciences, volume 9, Issue 3, Pg No S29-S31, 2016
26. Priya, N., Sridhar, J., Sriram, M. "Ecommerce Transaction Security Challenges and Prevention Methods- New Approach" 2016 ,Journal of Chemical and Pharmaceutical   Sciences, JCPS Volume 9 Issue 3.page no:S66-S68 .
27. Priya, N.,Sridhar,J.,Sriram, M."Vehicular cloud computing security issues and solutions" Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016
28. Priya, N., Sridhar, J., Sriram, M. "Mobile large data storage security in cloud computing environment-a new approach" JCPS Volume 9 Issue 2. April - June 2016
29. Anuradha.C, Khanna.V, "Improving network performance and security in WSN using decentralized hypothesis testing "Journal of Chemical and Pharmaceutical Sciences(JCPS) Volume 9 Issue 2, April - June 2016 .

AUTHORS PROFILE

Sidharthraj. R.S Assistant Professor, Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, India.

A. John Paul Praveen Assistant Professor, Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, India.

S. Balaji Assistant Professor, Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, India.