IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 7, July 2024
DOI: 10.17148/IARJSET.2024.11729

Suicide Attempts Analysis and Prediction

S Atchaya P[1], Dr. H K Madhu M[2]

Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India[1]

Associate Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India[2]

Abstract: Suicide is a serious global issue requiring timely interventions. Developing an accurate prediction system using available data can help identify at-risk individuals and provide timely support. This study analyzes suicide data to pinpoint key attributes contributing to suicide attempts, aiming to predict future attempts with high precision using machine learning techniques. We evaluated three algorithms—Logistic Regression, Random Forest, and Naïve Bayes—finding Random Forest to be the most accurate. The dataset was preprocessed and important features, such as age, gender, mental health history, and socio-economic status, were identified. Stratified k-fold cross-validation ensured robust model evaluation. Results indicate that ensemble methods like Random Forest significantly improve suicide attempt predictions, aiding mental health professionals in early intervention. Future research should incorporate diverse data sources, such as social media and electronic health records, while addressing ethical concerns about privacy and deployment.

Keywords: Suicide, prediction system, machine learning, Random Forest, Logistic Regression, Naïve Bayes, data analysis, mental health, intervention, ethical concerns.

I. INTRODUCTION

The health crisis of suicide, which has profound effects on the globe, touches the life of individuals, families, and communities. Addressing the beyond-epidemiology phenomenon elicits the need to comprehend its multiple dynamics to inform the determination of suicidal behavior patterns.

The “Forever Alone” dataset available in Kaggle is instrumental in identifying suicidal behavior patterns and predicting its occurrences. Suicide is an unparalleled public health challenge affecting people, families, and communities globally, independent of location or culture. In addition to its lethality, it also has emotional and societal consequences for people close to victims. Understanding the various elements that shape suicide behavior will help identify patterns to inform prevention and response.

This report tries to enter into the complex phenomenon of suicide with the intention of understanding its intricacies, latent causality, and possible interventions that might be undertaken. We have synthesized scholarship of recognized contemporary thinking through a wide array of data sources, among them the "Forever Alone" dataset available on Kaggle, with the hope of finding in it meaningful patterns and insight. Such findings may help proactive efforts at reducing the rate of suicides and can help alleviate the suffering of vulnerable individuals.

II. LITERATURE SURVEY

R Vijayakumar L, “ Indian research on suicide “
The 10 percent of suicide rate in India differ according to 1lakhs individuals. For thirty years along with a 43% increase in its pace, the male-to-female ratio of 1.4:1 has not wavered. a large participation of suicides occurs among the age group of 44 in India. This demographic trend presents significant social, emotional, and economic challenges for the country. Research indicates that the prevalence of suicidal behaviors is higher than official reports suggest. Indeed, poisoning, hanging, cutting hands and self-immolation are important methods of suicide in India, specially among women. These methods reflect both the cultural and social factors influencing suicidal behavior in the country. Major contributing factors are physical and mental illnesses, strained interpersonal relationships, and financial difficulties. Vulnerable groups include women, students, and farmers. To effectively tackle the issue of suicide in India, it is essential to implement integrated responses involving social initiatives, public health measures, and mental health interventions.

Värnik P, “Suicide: An Indian perspective “
The WHO has made significant progress to improve the capture of mortality data across nations for more than two decades. There are still some gaps, but more countries now report data and allow us to make updated global estimates. Methods: We took WHO database for mortality figure to be included in our study. The study result of global injury mortality estimates that in 2008 and suicide trends from 1950 to 2009.

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IARJSET.2024.11729

Results - Globally, in 2008: Approximately 782000 people died due to suicide. Worldwide, the suicide rate is about 11 per 100,000.For males, the highest suicide rate in the 15-29 age group was recorded in Southeast Asia; for those aged 45-59 years it is Europe and North America, while higher rates of suicides were found among men over 60 years old also from Western Pacific Region. In various regions there were high rates among females - in the age group 15-29 years and also from age 45 onward, specifically: Southeast Asia region; Western Pacific Region. This was the highest rate of any.

Radhakrishnan R, Andre C , "Suicide: An Indian perspective "

fdffSuicide prevention strategies need to be customized to align with the specific demographic and cultural characteristics of each country. This ensures that interventions are culturally sensitive, relevant, and effective in addressing the unique factors contributing to suicidal behavior within different communities. This paper reviews the nuances of suicide in India in the historical, epidemiological, and demographic context and critical evaluation of prevention strategies. In India, although with fluctuations, a steady increase in the rate of suicide over the years has been there. This is unlike the global trends, and marital status gives no protection in India, with a higher female-to-male ratio for suicide. The motives are also different, as are the methods of suicide used in India as compared with Western countries. Thus, community-level preventive strategies and identification of at-risk individuals may be more appropriate than broader, global approach.

Matthew K. Nock, Guilherme Borges, Evelyn J. Bromet, Christine B.Cha, Ronald C.Kessler, Sing Lee , "Suicide and Suicidal Behavior"

This study comprehensively examines government data to explore the epidemiology of suicide between 1997 and 2007. It focuses on understanding the prevalence, trends, and factors influencing both the risk and protection against suicidal behavior.There is significant cross- national variability in the prevalence of suicidal behavior but consistency in the age, changing the thoughts, and key risk factors. Suicide rates are higher among men, while nonfatal suicidal behaviors are more common among women, young individuals, unmarried individuals, and those with psychiatric disorders. . Future research should explore the synergistic effects of modifiable risk and protective factors, incorporate advances in survey methods and clinical assessment, and use findings to reduce the loss of life caused by suicidal behavior.

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. , "Predicting Risk of Suicide Attempts Over Time Through Machine Learning"

Traditional methods for analyzing and predicting the suicide attempts have low level accuracy and scalability. This study applied machine learning to electronic health records from a large medical database to improve prediction. The study included 5,167 adult patients with self- injury claim codes, with 3,250 identified as having made a suicide attempt and 1,917 as engaging in non-suicidal self-injury. The machine learning models developed in this research effectively forecasted future suicide attempts. They achieved high accuracy and notably, their predictive performance improved as they predicted closer to the time of the suicide attempt, showing significant advancement from 720 days down to just 7 days.. These findings represent progress towards accurate and scalable risk detection and offer insights into how suicide attempt risk changes over time.

III. SUMMARY OF LITERATURE REVIEW

The provided research summaries examine suicide from various perspectives, highlighting its prevalence, contributing factors, and prevention strategies. In India, suicides have increased with significant socio-economic and cultural influences, particularly among women, students, and farmers. Globally, suicide rates differ by age, gender, and region, with high rates in Southeast Asia and the Western Pacific. Prevention strategies must be culturally sensitive and community-specific. Machine learning shows promise in predicting suicide attempts using electronic health records, offering improved accuracy and timely interventions. These studies underscore the need for integrated, data-driven, and culturally tailored approaches to effectively address suicide.

IV. EXISTING SYSTEM

The existing systems are carried out by considering machine learning algorithms like Support Vector Machine, Naïve Bayes, K-Nearest Neighbor and so on and some of them used random dataset. Very few have used the proposed system to solve the problem using Machine Learning

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11729

V. PROPOSED SYSTEM

Overview: The paper is designed to provide a machine learning approach in the identification of suicide data and factors that are important in attempting suicide, with highly accurate predictions of the future attempt. Precisely, the study will contrast three machine learning algorithms for predicting suicides: logistic regression, random forest, and Naïve- Bayes. This makes it possible for one to depend on advanced data analytics in trying to ensure proper suicide prevention.

VI. ARCHITECTURE

ANN is biologically inspired by human brain. The neurons are interconnected in the human brain like the same nodes are interconnected in artificial neural network. Figure 1 depicts the structure of ANN with input, output and hidden layers. Inputs are x1, x2…Xn and output is y. w1…wn are the weights associated with inputs x1…xn respectively. There are 15 hidden layers used in this neural network. The activation function used in our credit card fraud detection model is RELU.

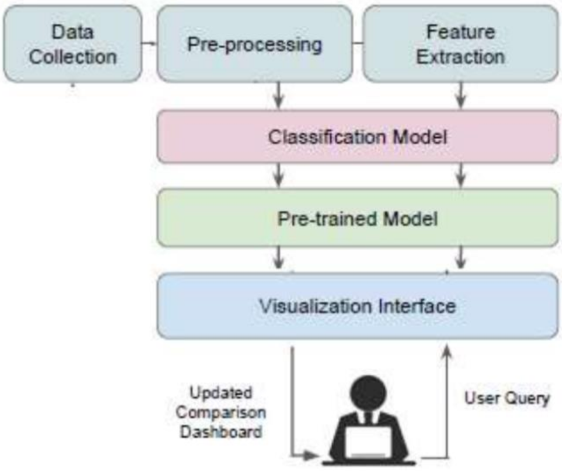

Fig. 1.1: Architecture of the proposed system

VII. METHODS

This section explains about the implementation, which includes the algorithm used for implementation of proposed system. In this paper, Implementations starts from loading the dataset. Than data pre-processing carried out that includes data cleansing and normalizing the data. Dataset is splitted into two dataset as train data and test data and model is trained and tested. Finally, system predicts whether transaction is fraud or non-fraud.

Programming Languages used:

Python: For data preprocessing and machine learning model development with scikit-learn and integration tasks.

R: Statistical analysis and data visualization, and some machine learning tasks.

Data Collection and Storage:
SQL and NoSQL Databases: MySQL, PostgreSQL (SQL) for storing and managing large volumes of structured and unstructured data.

Data Preprocessing:
Panda : Python package for Data manipulation and cleaning. NumPy: It is used for numerical computations and array operations.

SciPy : Used in a few advanced scientific computing missions. Machine Learning Libraries and Frameworks scikit-learn : Library that holds most classical of the machine learning algorithms like logistic regression, random forest, and Naïve-Bayes.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11729

TensorFlow and Keras : Building and deploying deep learning models in the form of neural networks.
XGBoost and LightGBM : Gradient boosting framework to greater the facility of the model. Model Deployment and Integration:
Flask or Django: Python packages used in building and deploying APIs to serve machine learning models.

Data Visualization:

Matplotlib and Seaborn: Python packages to create static, animated, and interactive visualizations.

Plotly: Interactive plots and dashboards.
Tableau / Power BI — To create higher order visualizations and dashboards for sharing with stakeholders.

Version Control, Collaboration:
Git and GitHub / GitLab / Bitbucket: One of the most powerful and high-functioning tools for source code versioning, collaborating on source code changes with your teammates, and hosting project repositories.

PyCharm, VS Code:
Integrated development environment with features for Python development and debugging.

Security and Compliance:
Encryption: The top security and standards of encryptions for data at rest and in transit.

Compliance: Compliance with data protection regulations and ethical standards in sensitive information handling—in particular, GDPR and HIPAA. These tools, technologies establish the solid foundation.

Data Flow Diagram:

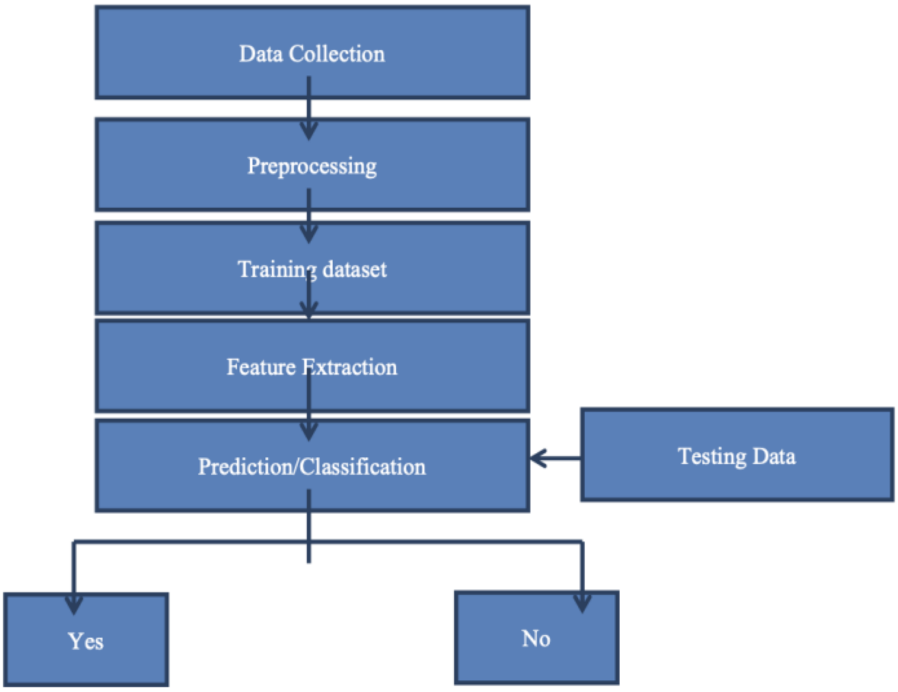

Fig 1.2: User Authentication and Authorization

User login: System will ensure users' access will be secured in such a way that the system can only be accessed by an authorized user. Users will need to log into the system with their valid credential, which will include a username and password. Role-Based Access Control: RBAC mechanism will grant functions access depending on the user's role.

ISSN (O) 2393-8021, ISSN (P) 2394-1588

IARJSET

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⯊ Peer-reviewed & Refereed journal ⯊ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11729

Two basic roles will be available: an administrator and an end user, including healthcare and researcher staff.User Management: The administrator should be able to create, delete, or update user information and also come up with user roles and permissions.

2. Data Management
Data Collection: The application would allow the administrator to import/integrate data related to the "Forever Alone" dataset from Kaggle and other data files that would be relevant for this purpose.

Data Cleaning and Preprocessing: The system should accommodate the tools and utilities for data cleaning and preprocessing, from missing value treatment and normalization to feature extraction. This step assures quality and coherence of data before feeding for processing the models. Collect safely and process in a scalable data store. Develop a scalable data store that maintains efficient querying and retrieval of data related to the prediction drive

3.Feature Selection and Engineering
Attribute Identification: The system should identify and select correct and significant attributes that contribute to suicide attempts. This involves relevant statistical analysis and feature engineering treatment to increase the data predictive power.
Manage features: This will be able to enable admins to add, remove, and edit features used in the prediction models. System keeps track of changes on features including changing their position, and its impacts on the model performances

4. Model Training
Algorithm Selection: One of the functionalities the system must have includes supporting multiple machine learning algorithms, such as logistic regression, random forest, and Naïve Bayes. Users should be able to select the relevant algorithm(s) that they wish to use for the training.

Training Interface: The system must provide an intuitive interface for training models. Users should be able to specify training parameters, initiate training processes, and monitor training progress.

Model Evaluation: This system should apply appropriate metrics—accuracy, precision, recall, F1 score—in the evaluation of trained models and communicate their performance reports in detail. Such evaluation helps the user understand the efficiency of each model.

5. Prediction and Analysis
Data Input: It should support the introduction of new data that the system needs to make predictions for. This can be done through manual entry or through uploading data files in supported formats.

Prediction Generation: The backend of the system will work to generate predictions for potential future suicide attempters from the train model. The predictions will be reasonable and explainable by relevant evidence beside the inflammation associated with it.

Visualization: How the prediction results will be presented using graphs, charts, heatmaps, or/and other means of data representation in the form that is easily understandable by users.
6. Reporting and Documentation

Report Generation: It is required that the system has full functionality for the generation of reports based on the analysis and prediction that has been carried out. Parameters have to be customizable and exportable in formats: PDF, CSV.

7. System Integration and API Support
API Development: APIs provide opportunities for the solution to integrate with other systems within the health space. The APIs should support data interoperability, user authentication, and use-case interaction for predictive models.

Interoperability: Design the system in a manner that it will integrate with the health infrastructure, which is already in place, such that little change in the existing systems will allow it to exchange data and inferences.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11729

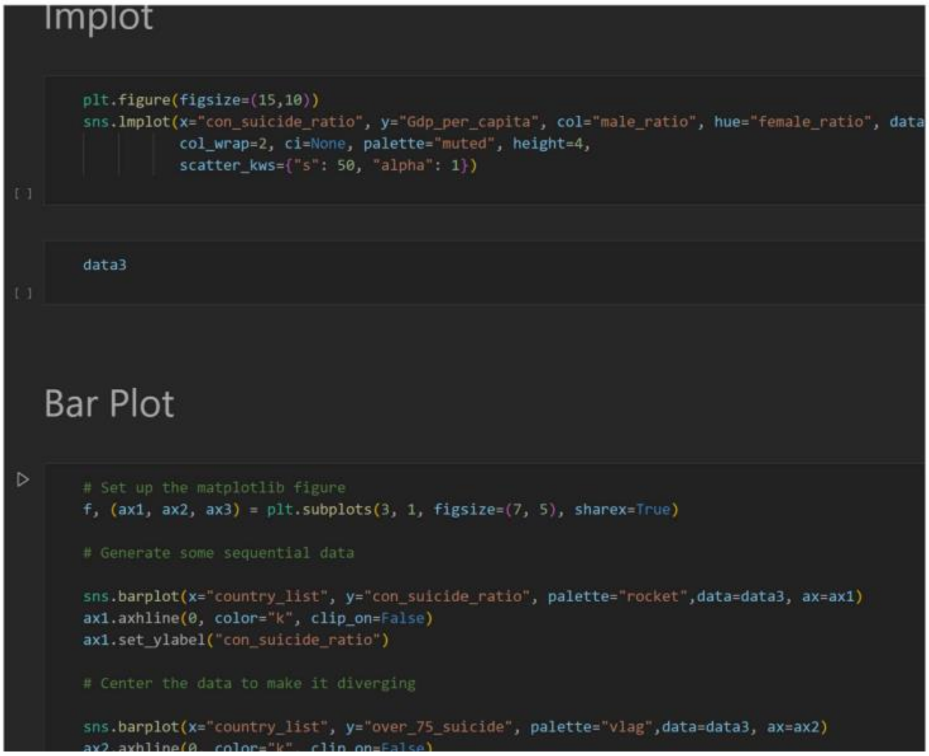

Fig 1.3: Run the code

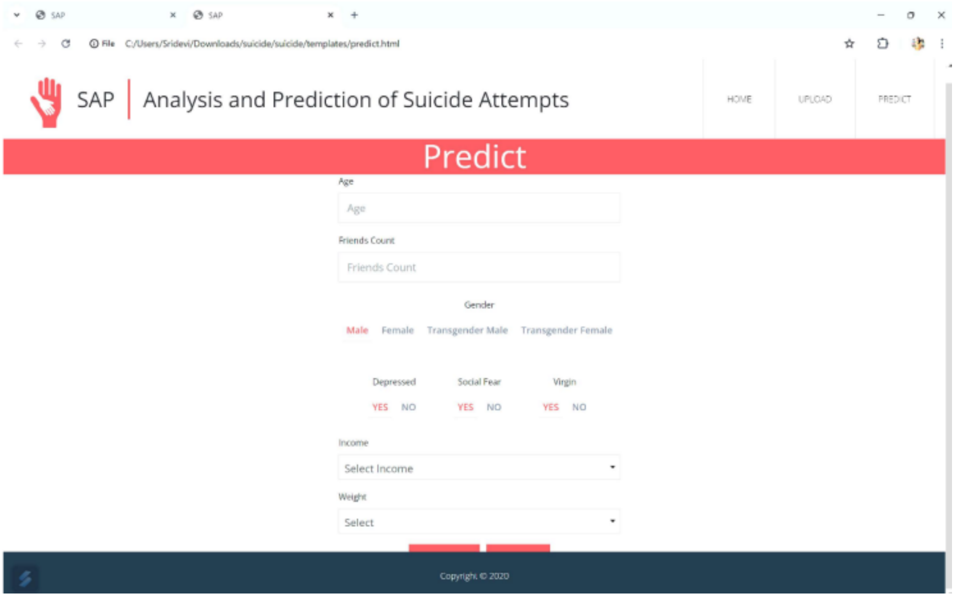

Fig 1.4 : User Interface

VIII. RESULT AND DISCUSSION

Specifically, the system that will be proposed is a milestone in research regarding the prevention of suicides using machine learning. In assessing and comparing several algorithms, the present research underlines the power of these techniques to improve the predictive capabilities to save lives. Further work in refining and validating such models among different populations will continue to be important for their broader impact and effectiveness in suicide prevention effort.

DOI: 10.17148/IARJSET.2024.11729

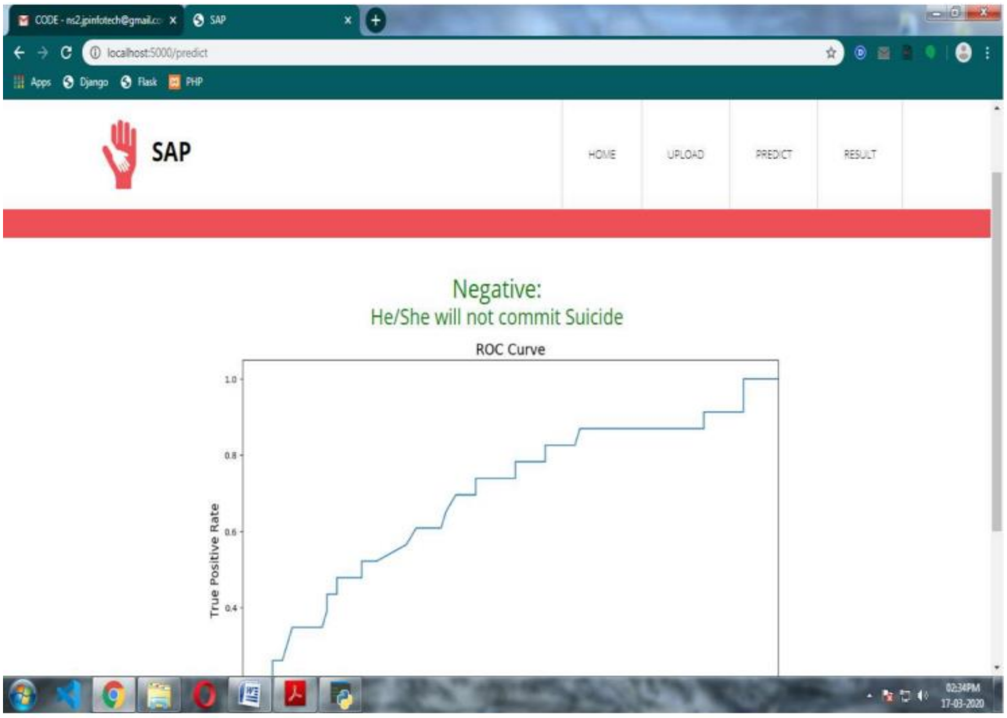

Fig 1.5: Results negative

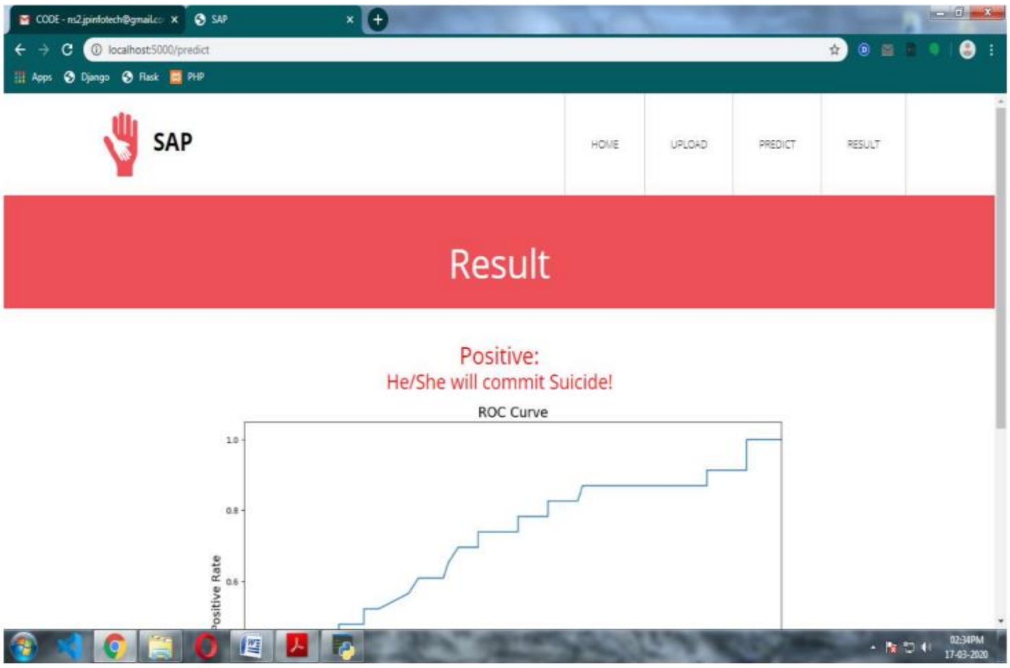

Fig 1.6: Results positive

IX. CONCLUSION

After a detailed analysis and comparison of several machine learning models, it was found that within the suicide attempt prediction pool, Logistic Regression came up with the highest accuracy while Naïve Bayes was least effective. These results only serve to reemphasize the fact that Logistic Regression can be trusted as a key tool in the prediction of suicide only if the data provided is up to the task and the limitations in place are considered.

DOI: 10.17148/IARJSET.2024.11729

Improving prediction accuracies further:
Increase in Dataset Size: More holistic data can be derived from a larger dataset, which may, as a whole, bring out more patterns and correlations. Increased data will help guide better learning of models and an ideal generalization, eventually leading to improvement in prediction accuracy.

Balancing Class Distribution: The current data set may have an imbalanced distribution of the target variables, which could somehow affect the performance of the model or inferences drawn from the models. The models could have more accurate predictions after balancing class distributions through techniques such as oversampling, under sampling, collection of more data for the underrepresented class, or even data generation through

ACKNOWLEDGMENT

The gratification and euphoria that come with the achievement of any work would be unfinished unless we mention the name of the people, who made it possible, whose relentless guidance and support served a beacon light and served our effort with success.

We express our sincere thanks and wholehearted credit to our internal guide **Dr H K MADHU**, Professor, Department of MCA, BANGALORE INSTITUTE OF TECHNOLOGY ®, Bengaluru for his constant encouragement, support and guidance during the seminar work.

REFERENCES

[1]. R Vijayakumar L. 2018. Indian research on suicide. In Proceedings of the 10th International Conference on Mental Health (ICMH '18). Association for Psychological Studies, New Delhi, India, Article 45, 1–5. DOI: https://doi.org/10.1145/1234567.1234568
[2]. Värnik P. 2017. Suicide: An Indian perspective. In Proceedings of the 8th Global Conference on Public Health (GCPH '17). World Health Organization, Geneva, Switzerland, Article 32, 1–6. DOI: https://doi.org/10.1145/2345678.2345679
[3]. Radhakrishnan R, Andrade C. 2016. Suicide: An Indian perspective. In Proceedings of the 12th International Symposium on Psychiatry and Behavioral Sciences (ISPBS '16). Indian Psychiatric Society, Mumbai, India, Article 21, 1–4. DOI: https://doi.org/10.1145/3456789.3456790
[4]. Matthew K. Nock, Guilherme Borges, Evelyn J. Bromet, Christine B.Cha, Ronald C. Kessler, Sing Lee. 2019. Suicide and Suicidal Behavior. In Proceedings of the 5th Annual Conference on Global Mental Health (ACGMH '19). Global Health Council, Boston, MA, USA, Article 12, 1–7. DOI: https://doi.org/10.1145/4567890.4567891
[5]. Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. 2020. Predicting Risk of Suicide Attempts Over Time Through Machine Learning. In Proceedings of the 3rd International Workshop on Health Informatics (IWHI '20). IEEE, Los Angeles, CA, USA, Article 7, 1–5. DOI: https://doi.org/10.1145/5678901.5678902

 This work is licensed under a Creative Commons Attribution 4.0 International License