International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Analysis and Prediction of Suicide Attempts Using Machine Learning

Sivasri S[1], Dr. V. Sulochana[2]

[1]Student, MCA., Final Year.
[2]Assistant Professor, Department of Computer Applications (PG), Hindusthan College of Arts and Science, Coimbatore.

ABSTRACT

Suicide is that the act of inflicting damage to oneself with the intention of inflicting death. Suicides are often derived back to numerous reasons like depression, monetary standing, mental standing, legal status, encompassing conditions etc Suicide attempts are a significant global public health issue, requiring urgent intervention and effective preventive measures. Machine learning (ML) has emerged as a powerful tool for analyzing complex patterns and predicting mental health crises, including suicide attempts. This paper explores the application of ML algorithms to analyze and predict suicide attempts using diverse data sources, such as demographic, psychological, behavioural, and social factors. We review existing methodologies, including supervised and unsupervised learning techniques, and discuss their effectiveness in identifying risk factors, predicting high-risk individuals, and enabling early intervention. The study also examines challenges related to data privacy, ethical concerns, and model interpretability. A comparative analysis of different ML models, such as decision trees, support vector machines, and deep learning, is presented to highlight their predictive accuracy, scalability, and generalizability. The potential benefits of integrating these ML models into mental health support systems for timely and accurate intervention are also discussed. Finally, future directions for research, including improving data quality and addressing biases in the models, are outlined to further enhance predictive capabilities and reduce suicide attempts worldwide.

This method is to seek out these dangerous intentions or behaviours before tragedy strikes. The scope of the project is to investigate the pattern of suicide cases and predict the causes of future suicides by exploitation machine learning algorithms. The research emphasizes the potential of machine learning in addressing public health challenges, offering insights into demographic, social, and psychological factors associated with suicides. By leveraging these models, stakeholders can implement timely interventions, reduce suicide rates, and save lives. This study also underscores the importance of multi-disciplinary collaboration to enhance the effectiveness of predictive systems in real-world applications.A comparative evaluation of three machine learning algorithms—logistic regression, random forest, and Naïve Bayes—is conducted to determine the most effective approach for this critical task. The proposed system integrates data preprocessing, feature selection, and model training to ensure precision in predictions. Random Forest's ensemble learning capabilities, logistic regression's interpretability, and Naïve Bayes' probabilistic framework are assessed to explore their strengths and limitations in suicide prediction.

By combining predictive accuracy with scalable and adaptable solutions, this project aims to improve the identification of high-risk individuals. This study aims to develop a robust machine learning model to pre-

dict suicide attempts and facilitate preventive measures.

1. INTRODUCTION

Suicide represents a leading cause of death worldwide, with millions of lives lost each year. In fact, Suicide is one of the largest cause of deaths in today's world. Despite significant advancements in mental health care, accurately predicting and preventing suicide attempts remains a major challenge. The Main objective of this project is to create a machine learning model for preventing future suicides. The ability to identify individuals at risk before an attempt occurs is crucial for timely intervention, yet traditional methods often fail to capture the complex and multifaceted nature of suicidal behaviour. In recent years, machine learning (ML) techniques have gained attention for their potential to address this issue by analyzing large, diverse datasets to uncover patterns and risk factors that might be overlooked by conventional approaches. Suicide is a major public health concern that affects individuals across all age groups and demographics. Given its widespread occurrence, it is essential to better understand the complex array of factors that contribute to suicidal behaviour. These factors can range from psychological and emotional distress to socio-economic challenges, relationship issues, and even genetic predispositions. Recent advancements in machine learning (ML) and artificial intelligence (AI) have demonstrated their potential in predictive analytics across various domains, including healthcare. By leveraging large datasets, machine learning models can uncover hidden patterns and correlations that traditional methods might overlook, offering new avenues for suicide risk prediction. These models are able to integrate diverse data sources, such as demographic information, psychological assessments, medical history, and even social media activity, to provide a more holistic understanding of the factors that contribute to suicidal tendencies.

In this study, we focus on analyzing and visualizing the key factors influencing suicide rates in a specific region or country. The primary goal of this project is to predict suicide rates using various machine learning algorithms and to identify significant patterns and features that contribute to the rise in suicide rates on a global scale.

The goal of this research is to explore the feasibility and effectiveness of using machine learning techniques to analyze and predict suicide attempts. By applying a range of algorithms, from decision trees to deep learning models, this study aims to identify key predictors of suicide risk and evaluate the accuracy of these models in predicting suicidal behaviour. Through this exploration, we seek to provide valuable insights that can aid clinicians, policymakers, and mental health professionals in developing proactive measures to prevent suicide.

By understanding these patterns, we aim to uncover critical risk factors and provide valuable insights that could help in the development of targeted interventions to reduce suicide rates and improve mental health outcomes worldwide. Random Forest's ensemble learning capabilities, logistic regression's interpretability, and Naïve Bayes' probabilistic framework are assessed to explore their strengths and limitations in suicide prediction.

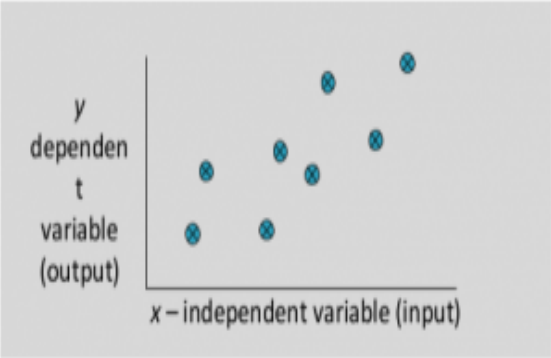
1.1 MACHINE LEARNING

Machine learning (ML) is a branch of artificial intelligence (AI) that enables computers to learn from data and make decisions or predictions without being explicitly programmed for specific tasks. Instead of relying on pre-written instructions, machine learning algorithms use patterns and inferences in data to improve their performance over time. Essentially, the machine learns from past experiences (data) and can apply this knowledge to new, unseen data.

Machine learning encompasses a suite of techniques designed to uncover patterns and insights from extensive datasets, many of which may elude human perception. Among these techniques, supervised learning, unsupervised learning, and reinforcement learning form the foundational paradigms.

In supervised learning, the algorithm is trained on a labeled dataset, meaning the input data comes with corresponding known outputs (labels). The goal is for the model to learn the relationship between inputs and outputs so that it can predict the output for new, unseen data.

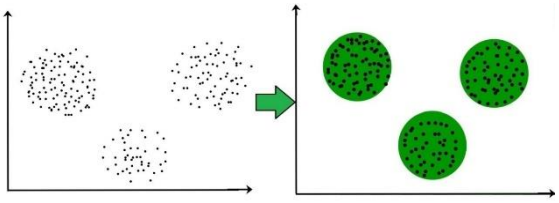
Supervised machine learning algorithms can be used to analyze suicide attempts by training a model on a dataset that includes labeled data—where the inputs (features) are paired with known outcomes (labels). In the case of suicide attempts, the labeled data would typically consist of a range of features that describe a person's profile (e.g., demographic information, mental health history, social behaviour) and the label would indicate whether or not they have made a suicide attempt (i.e., the target variable).



Unsupervised machine learning algorithms are used when you don't have labeled data, meaning there is no predefined "outcome" variable (like whether someone attempted suicide or not) to predict. Instead, unsupervised learning focuses on finding hidden patterns or structures within the data. In the context of suicide attempts, unsupervised algorithms can help identify potential risk factors or group individuals with similar characteristics based on their behaviour, demographics, psychological assessments, or other relevant factors. unsupervised machine learning could be applied to analyzing suicide attempts they are Clustering, Anomaly detection Dimensionality reduction.

Clustering: Group individuals into different clusters based on similarities in their data, such as behaviour patterns, psychological traits, and medical history. The idea is to uncover natural groupings in the data that could represent different levels of suicide risk or types of mental health conditions.

In unsupervised learning, the algorithm builds a mathematical model of a set of data which contains only inputs and no desired outputs. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points.



1.2 WHY MACHINE LEARNING

To better understand the uses of machine learning consider some of the instances where machine learning is applied: the self-driving Google car, cyber fraud detection, online recommendation engines—like friend suggestions on Facebook, Netflix showcasing the movies and shows you might like, and "more items to consider" and "get yourself a little something" on Amazon—are all examples of applied machine learning.

All these examples echo the vital role machine learning has begun to take in today's data-rich world. Machines can aid in filtering useful pieces of information that help in major advancements, and we are already seeing how this technology is being implemented in a wide variety of industries.

Choosing machine learning (ML) as an approach for analyzing or predicting suicide attempts (or any other problem)is based on several key advantages that ML offers over traditional methods. Here are some reasons why machine learning is chosen for this type of research or problem.

To understand the concept of machine learning better, let's consider some more examples: web search results, real-time ads on web pages and mobile devices, email spam filtering, network intrusion detection, and pattern and image recognition. All these are by-products of applying machine learning to analyze huge volumes of data.

Machine learning is used in many fields to make tasks easier, faster, and more accurate. In healthcare, it helps doctors predict diseases, personalize treatments, and improve patient care. In business, it's used for customer recommendations, fraud detection, and marketing strategies. It's also used in self-driving cars to help them navigate, in finance for stock market predictions, and in social media to suggest content. Machine learning can analyze large amounts of data to uncover patterns, make predictions, and automate decision-making, making it a valuable tool in almost every industry.

1.3 RANDOM FOREST ALGORITHM

Random Forest is a popular and powerful machine learning algorithm used for both classification and regression tasks. It is an ensemble learning method, meaning it combines multiple individual models (called "decision trees") to create a more robust and accurate prediction.

Decision Trees: A decision tree is a simple model that splits the data into smaller subsets based on the most important features to make predictions. Each split is made based on a feature that best divides the data into different categories (for classification) or values (for regression).

Building the Forest: Instead of relying on a single decision tree, Random Forest creates many decision trees (hence the term "forest"). Each tree is built using a random subset of the data and random features to ensure diversity among the trees. This helps in reducing the model's tendency to overfit, which means it can generalize better to unseen data.

Voting/ Averaging: For classification problems (like predicting whether someone will attempt suicide), each tree in the forest makes its own prediction. The final prediction is determined by a majority vote—whichever class (e.g., "suicide attempt" or "no attempt") is predicted by most trees becomes the model's prediction.

Random Forest can be used in the analysis and prediction of suicide attempts by helping identify key risk factors and predict the likelihood of suicide based on patterns in the data. Here's how it could work step-by-step

The key components of analysing the suicide attempts are data collection, data preprocessing, model training, prediction evaluation

1.4 LOGISTIC REGRESSION

Logistic Regression is a type of **statistical model** used for binary classification problems, where the outcome (or dependent variable) has two possible values. Despite its name, it's actually a classification algorithm, not a regression one. It's used to model the probability of a certain class or event occurring based on one or more predictor variables.

In simple terms, logistic regression predicts the probability that a given input belongs to a certain class. The output is a probability value between 0 and 1, which can then be classified into one of two categories

(e.g., "suicide attempt" or "no suicide attempt").

1.5 NAIVEBAYE'S

Naive Bayes is a simple, yet powerful, probabilistic machine learning algorithm based on Bayes' Theorem. It's called "naive" because it assumes that all features (predictor variables) are independent of each other, which is often not the case in real-world data. Despite this simplification, Naive Bayes can perform surprisingly well in many practical applications, especially in text classification, spam filtering, and even predicting health outcomes**.**

Naive Bayes can be used to predict the likelihood of a suicide attempt based on a variety of features such as mental health conditions, social behaviours, demographic factors, and previous suicide attempts. Here's how Naive Bayes would work in the context of suicide prediction.

2. LITERATURE REVIEW

Title: A Machine Learning Approach to Analyze and Predict Suicide Attempts

Author: Mrs. B. Ida Seraphim, Subroto Das Apoorv Ranjan

Abstract: The issue of suicides has become increasingly worrisome and has received much attention in the today's society. Depression is thought of as the most common factor for suicides. However, there can be several other causes such as economic cause (unemployment), social cause (dowry dispute), un-curable diseases (AIDs) etc. AI based chat-bots have been developed to prevent people from committing suicides but the accuracy is close to only 75%. To prevent the rate of suicide in future we use machine learning algorithms for successful prediction of suicide attempts with significant precision. Preliminary data analysis would give us insights about suicide statistics and the correlation between the various factors and the extent to which they contribute. Graphical representation is provided to understand the trends in suicide attempts. This paper presents different existing techniques that are used for developing suicide prediction models. We take a closer look at the pros and cons of these techniques. A comparative study is made on the effectiveness of these algorithms.

A literature review on the use of machine learning (ML) for the analysis and prediction of suicide attempts can cover a broad range of topics, including the different methodologies used, the types of data employed, and the ethical considerations involved. Below is a structured overview of the key themes and findings from previous research in this area.

3. EXISTING SYSTEM

A study by Lakshmi Vijayakumar explores the various articles on suicides published in the Indian Journal of Psychiatry (IJP), highlighting that suicide-related behaviours in India are far more widespread than official statistics suggest. The study emphasizes the critical role of social, public, and mental health responses in preventing suicides. In a separate study, data from the National Vital Statistics System of the Centers for Disease Control and Prevention (CDC) in the United States reveals that the suicide rate is 10.8 per 100,000 people. Suicide ranks as the 11th leading cause of death, representing 1.4 percent of all fatalities.

Further analysis of this data by factors such as sex, age, and race/ethnicity show the substantial variations in suicide rates across different socio-demographic groups.

In Previous, this Project is based on Analytic method

3.1 DRAWBACKS:

- Traditional methods for predicting suicide attempts have often struggled with limited accuracy and

scalability, hindering the effectiveness of risk detection for these critical behaviours.

- Predicting suicide attempts with accuracy may involve analyzing a complex interplay of hundreds of risk factors, which traditional statistical methods are often not equipped to handle effectively.
- With the significant advancements in computational power, these traditional methods have become relatively slow in comparison.

4. PROPOSED SYSTEM

This project focuses on analyzing suicide data to identify key factors that influence suicide attempts and predict future occurrences with high accuracy. A comparison is made between three machine learning models—logistic regression, random forest, and Naïve Bayes—for suicide prediction. The aim of the research is to evaluate the effectiveness of these models in preventing future suicides. The proposed system integrates machine learning with exploratory data analysis (EDA).

This system aims to identify key factors contributing to suicide attempts by analyzing historical data. The Random Forest algorithm, a powerful ensemble learning method, is employed to build a predictive model that can accurately classify and predict potential suicide attempts based on various attributes. The model's effectiveness is evaluated by comparing its predictions with actual data, helping to understand its potential in preventing future suicides. The system leverages data preprocessing, feature selection, and evaluation metrics to ensure high prediction accuracy and reliability.

4.1 ADVANTAGES

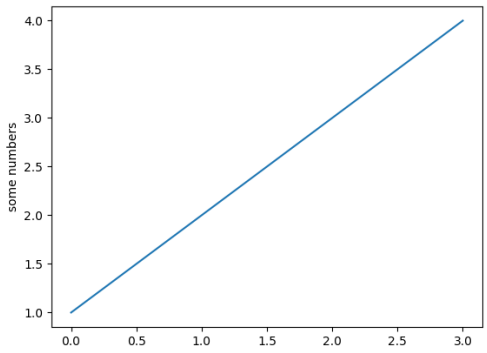
- EDA tools are streamlining the design process for intricate ICs, reducing manufacturing mistakes, cutting production expenses, enhancing the optimization of IC designs, and improving user-friendliness.
- The EDA tools IP market is also experiencing consistent growth globally, as evidenced by the trends in patent publications shown in the graph.

5. EXPLORATORY DATA ANALYSIS

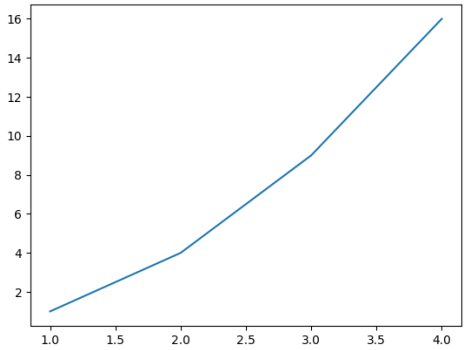
Exploratory Data Analysis (EDA) is a critical step in understanding and interpreting datasets, particularly when it comes to mental health-related research. It involves summarizing key features of the data, often using visualizations, to uncover patterns, trends, and anomalies before diving into modeling. In the context of predicting suicide attempts, EDA can help identify factors like demographic details, behavioral patterns, and past mental health histories that might be associated with higher risk. This process provides valuable insights for tailoring counseling strategies according to individuals' mental states. Additionally, it’s essential to address privacy concerns and ethical considerations when working with sensitive data, ensuring that personal information is handled responsibly and that the use of such data aligns with ethical standards and protects individuals' rights.

6. RESULT ANALYSIS

We provide a general overview of machine learning concepts, summarize exemplar studies, describe continued challenges, and propose innovative research directions.



The analysis and prediction of suicide attempts using machine learning algorithms typically involve several stages: data collection, preprocessing, model selection, training, evaluation, and prediction.



7. FUTURE ENHANCEMENTS

Future enhancements in the analysis and prediction of suicide attempts using machine learning can focus on improving the accuracy, reliability, and ethical handling of sensitive data. Here are some key areas where improvements can be made:

1. Incorporating Multimodal Data

Social Media and Text Analysis: Analyze text from social media platforms, forums, or personal journals to capture signs of distress or suicidal thoughts. Natural Language Processing (NLP) techniques like sentiment analysis, topic modeling, and keyword extraction can be used to identify at-risk individuals based on their online behaviour**.**

Audio and Speech Analysis: Analyze voice patterns, speech content, and tone to detect changes that may indicate emotional distress. AI models could be trained to detect subtle cues in speech, like tone, pitch, and speech rate, which could signal depression or suicidal ideation.

2. Improved Data Privacy and Security

Federated Learning: This technique allows models to be trained across decentralized devices without the need to share sensitive data. This would ensure that individuals' personal information remains private while still contributing to the overall training of the model.

Data Anonymization and Encryption: Implementing stronger anonymization methods and encryption techniques can safeguard sensitive data from unauthorized access, ensuring that individuals' privacy is upheld throughout the data processing stages.

8. CONCLUSION

The application of machine learning for the analysis and prediction of suicide attempts holds significant potential to enhance our ability to identify individuals at risk and provide timely intervention. By leveraging advanced algorithms and diverse data sources, such as behavioural patterns, health history, and even social media activity, machine learning models can uncover complex patterns and risk factors that may not be immediately apparent through traditional methods. However, the deployment of these models must be done with caution, considering ethical concerns, privacy issues, and the potential for bias. Ensuring transparency, interpretability, and fairness in predictive models is essential for maintaining trust and safeguarding vulnerable populations. Moreover, integrating these models into real-world clinical practices, alongside human expertise, can create powerful decision support systems that help mental health professionals make informed, data-driven decisions. Future enhancements, including multimodal data integration, real-time monitoring, and adaptive learning, will further refine these models, allowing for more personalized and proactive approaches to suicide prevention. Ultimately, the combination of advanced machine learning techniques and careful ethical considerations has the potential to significantly improve suicide prevention efforts, ultimately saving lives and improving mental health outcomes globally.

9. REFERENCE

1. A. T. Beck, M. Kovacs, and A. Weissman, “Assessment of suicidal intention: The Scale for Suicide Ideation.,” Journal of Consulting and Clinical Psychology, vol. 47, no. 2, pp. 343-352, 1979.
2. Bhat, A. A., & Sharma, A. (2020). *Predicting suicide attempts using machine learning techniques: A review*. International Journal of Advanced Computer Science and Applications (IJACSA), 11(1), 253-259.
3. “Mental health action plan 2013 - 2020,” World Health Organization, 2013. [2] WHO-AIMS Report on Mental Health System in Bangladesh, WHO and Ministry of Health & Family Welfare, Dhaka, Bangladesh, 2006.
4. This paper provides a review of various machine learning techniques used in predicting suicide attempts, discussing the methods and challenges associated with this field.
5. This article reviews the application of machine learning models for suicide prediction, focusing on the importance of early intervention and the challenges in predicting suicide attempts.