Developing a Text Classification Model for Identifying Suicidal Ideation using Natural Language Processing Techniques

A Project Report
Submitted in partial fulfillment of the
academic requirements for the award of the degree
of
Bachelor of Engineering
in
COMPUTER SCIENCE & ENGINEERING

Under the Supervision of

Mr. Sumit Gupta
Assistant Professor
Department of Computer Science and Engineering
University Institute of Technology
The University of Burdwan

By

Arya Manikya Sinha (Roll no.-20191040, Regn. No.- A7560 of 2019-20)

Debjyoti Prodhan (Roll no.-20191059, Regn. No.- A7553 of 2019-20)

Souvik Modak (Roll no.-20191057, Regn. No.- A7552 of 2019-20)

Nirnay Ghosh (Roll no.-20191061, Regn. No.- A7554 of 2019-20)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN
GOLAPBAG (NORTH), BURDWAN-713104, WEST BENGAL, INDIA

JULY 2023

UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN

CERTIFICATE OF APPROVAL

This is to certify that the project entitled ***"Developing a Text Classification Model for Identifying Suicidal Ideation using Natural Language Processing Techniques"*** *is successfully carried out by* ***Mr. Arya Manikya Sinha*** *(Roll no.- 20191040, Regn. no.- A7560 of 2019-20),* ***Mr. Debjyoti Prodhan*** *(Roll no.- 20191059, Regn. No.- A7553 of 2019-20),* ***Mr. Souvik Modak*** *(Roll no.- 20191057, Regn. No.- A7552 of 2019-20) and* ***Mr. Nirnay Ghosh*** *(Roll no.- 20191061, Regn. No.- A7554 of 2019-20) of University Institute of Technology, The University of Burdwan, Burdwan, West Bengal, India, under my guidance and supervision towards the partial fulfillment of the academic requirements for the award of the degree of Bachelor of Engineering (B.E.) in Computer Science & Engineering (CSE).*

Signed by:

____

Mr. Sumit Gupta

Project Supervisor & Assistant Professor
Department of Computer Science & Engineering
University Institute of Technology
The University of Burdwan
Burdwan-713104, West Bengal, India

Date:

Place:

UNIVERSITY INSTITUTE OF TECHNOLOGY
THE UNIVERSITY OF BURDWAN

CERTIFICATE

This is to certify that the project entitled ***"Developing a Text Classification Model for Identifying Suicidal Ideation using Natural Language Processing Techniques"*** *submitted by* ***Mr. Arya Manikya Sinha*** *(Roll no.- 20191040, Regn. no.- A7560 of 2019-20),* ***Mr. Debjyoti Prodhan*** *(Roll no.- 20191059, Regn. No.- A7553 of 2019-20),* ***Mr. Souvik Modak*** *(Roll no.- 20191057, Regn. No.- A7552 of 2019-20) and* ***Mr. Nirnay Ghosh*** *(Roll no.- 20191061, Regn. No.- A7554 of 2019-20) to University Institute of Technology, The University of Burdwan, Burdwan, West Bengal, India, is a record of bonafide Project work carried out by them towards the partial fulfillment of the academic requirements for the award of the degree of Bachelor of Engineering (B.E.) in Computer Science & Engineering (CSE).*

We wish them every success in their future endeavors.

Mr. Sumit Gupta
Project Supervisor & Assistant Professor
Department of Computer Science & Engineering
University Institute of Technology
The University of Burdwan
Burdwan-713104, West Bengal, India

Dr. Souvik Bhattacharyya
In-Charge & Assistant Professor
Department of Computer Science & Engineering
University Institute of Technology
The University of Burdwan
Burdwan-713104, West Bengal, India

ACKNOWLEDGEMENT

This project work in the domain of 'Machine Learning' entitled "**Developing a Text Classification Model for Identifying Suicidal Ideation using Natural Language Processing Techniques**", has truly helped us shape our knowledge in this branch of study. Machine Learning has always been our interest but after indulging in this project, we feel connected to the real word problems and how they can be solved.

First and foremost, we express our sense of gratitude to our respected and learned supervisor, **Mr. Sumit Gupta**, Assistant Professor of Computer Science & Engineering Department, University Institute of Technology, The University of Burdwan for his patience, enthusiasm, insightful comments, invaluable suggestions, helpful information, practical advice and unceasing ideas which have helped us tremendously at all times in our research and writing of this thesis.

We are also grateful to **Dr. Souvik Bhattacharyya**, In-charge, Computer Science & Engineering Department, University Institute of Technology, The University of Burdwan and all other lecturers and staff of CSE Department for their invaluable support, kindness, hospitality in the making of our final year project work. We would also like to thank our fellow classmates to help us at each step and for sharing wonderful moments.

Lastly, we are very thankful to our teammates for their kind co-operation and help in completing this thesis. We are also thankful to our parents for being our backbone in the process.

Arya Manikya Sinha
Roll. No.- 20191040
Regn. No.- A7560 of 2019-20

Debjyoti Prodhan
Roll. No.- 20191059
Regn. No.- A7553 of 2019-20

Souvik Modak
Roll. No.- 20191057
Regn. No.- A7552 of 2019-20

Nirnay Ghosh
Roll. No.- 20191061
Regn. No.- A7554 of 2019-2

TABLE OF CONTENTS

LIST OF ACRONYMS

<u>SHORT FORM</u>:	***<u>FULL FORM</u>***:
OSN	Online Social Networks
ML	Machine Learning
DL	Deep Learning
DD	Depression Detection
SI	Suicidal Ideations
RF	Random Forest
BN	Bayesian Network
RN	Relation Network
SVM	Support Vector Machine
NB	Naive Bayes
DT	Decision Tree
ADAB	AdaBoost
BoW	Bag of Words
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
SGD	Stochastic Gradient Descent classifier
LR	Logistic Regression
MNB	Multinomial Naive Bayes
LSTM	Long-Short-Term Memory
Bi-LSTM	Bi-directional LSTM
GRU	Gated Recurrent Unit
BiGRU	Bi-directional GRU
CNN	Convolutional Neural Network

RNN	Recurrent Neural Network
VADER	Valence Aware Dictionary for sEntiment Reasoning
CLSTM	Contextual LSTM
LMT	Logistic Model Tree
SMA	Social Media Analysis
SRA	Suicide Risk Assessment
WEKA	Waikato Environment for Knowledge Analysis
ET	Extra Tree
SMO	Sequential Minimal Optimization
KARA	Knowledge Aware Risk Assessment
MLP	Multilayer Perceptron
TF-IDF	Term Frequency – Inverse Document Frequency
LIWC	Linguistic Inquiry and Word Count
ICD	International Classification of Diseases
UMLS	Unified Medical Language System
PAM	Partitioning Around Medoids
HCA	Hierarchical Cluster Analysis
JRIP	Joint Reserve Intelligence Program
CRF	Conditional random fields
WHO	World Health Organization
POS	Part-of-speech
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
API	Application Program Inter
MDP	Markov Decision Process
BERT	Bidirectional Encoder Representations from Transformers

LIST OF FIGURES & TABLES

Abstract

Suicide has become a serious social health issue in the modern society. Suicidal intent is people's thoughts about committing suicide. The suicide of a person is a tragedy that deeply affects families, communities, and countries. According to the standardized rate of suicides per number of inhabitants worldwide, in 2022 there has been approximately 903,450 suicides and 18,069,000 unconsummated suicides, affecting people of all ages, countries, races, beliefs, social status, economic status, sex, etc. Depression is a prevalent mental disorder that can affect productivity in daily activities and might lead to suicidal thoughts or attempts. That means depression and suicide are related to each other. So, the aim of this project is to detect suicidal intent through depression detection by creating ML and DL models. Two proposed systems are developed for detecting suicidal intent from textual data. Proposed System 1 employs ML and DL approaches with features such as Bag-of-Words (unigram), TF-IDF, and bigram representations. Four ML classifiers (SVM, NB, LR, and RF) and four DL classifiers (LSTM, Bi-LSTM, CNN, and BERT) have been utilized. The evaluation of Proposed System 1 achieves accuracies of 82%, 87%, 77%, 73%, 71%, and 94% for the Reddit, Life_Corpus, CEASE, SWMH, SDCNL, and SDD datasets, respectively. Proposed System 2 incorporates a hybrid approach, combining LSTM, SVM, and the VADER Lexicon. The evaluation of Proposed System 2 yields accuracies of 97.4%, 97.14%, 99.2%, 91.5%, 94.4%, and 93.7% for the Reddit, Life_Corpus, CEASE, SWMH, SDCNL, and SDD datasets, respectively. Comparing the proposed systems with existing works, it is found that they outperform previous research works for the Reddit, Life_Corpus, CEASE, and SWMH datasets, achieving accuracy scores of 97.4%, 97.14%, 99.2%, and 91.5% respectively. Although the proposed model achieves promising results for the SDCNL and SDD datasets with accuracy scores of 94.4% and 94%, further research is needed to surpass the results of previous works.

Keywords: Machine Learning, Deep Learning, Depression, Suicide, Depression Detection, Suicidal Ideation, VADER, Text Classification.

Chapter 1

Introduction

Through text anyone can express their feelings and show their intentions towards anything. From these texts and their features many meaningful outcomes can be derived. Text analysis is an approach by which many kinds of feelings like emotion, happiness, excitement, depression level, mental state etc. can be identified.

Suicide has been around for as long as human society, ranking among the top 13 causes of death in all ages worldwide and continues to challenge our collective wisdom. As per the world Health Organization (WHO), suicide is a primary cause of death among individuals between 15-29 years old across the world. 8,00,000 of people commit suicide every year leading to increase in suicidal ideation [1]. However, an individual person suicide plays an unsocial act that has overwhelming impact towards relations and families. Several suicidal demises are inevitable and very significant to know the behaviour and the way how individual communicate thoughts and depression for inhibiting such deaths. Suicidal avoidance predominantly focuses on monitoring and observation of suicidal efforts and self-harm tendencies.

1.1 Suicide:

Suicide is defined as the intentional taking of one's own life. Prior to the late nineteenth century, suicide was legally defined as a criminal act in most Western countries. In the social climate of the early 2000s, however, suicidal behaviour is most commonly regarded and responded to as a psychiatric emergency.

The approach to understanding suicide must be multidisciplinary, involving psychologists, psychiatrists, toxicologists, physiologists and physicians, because suicide is a multi-faceted and complex event. As seen in Fig 1.1, the premature and violent death of the victims has negative repercussions in society and should be prevented whenever possible [2].

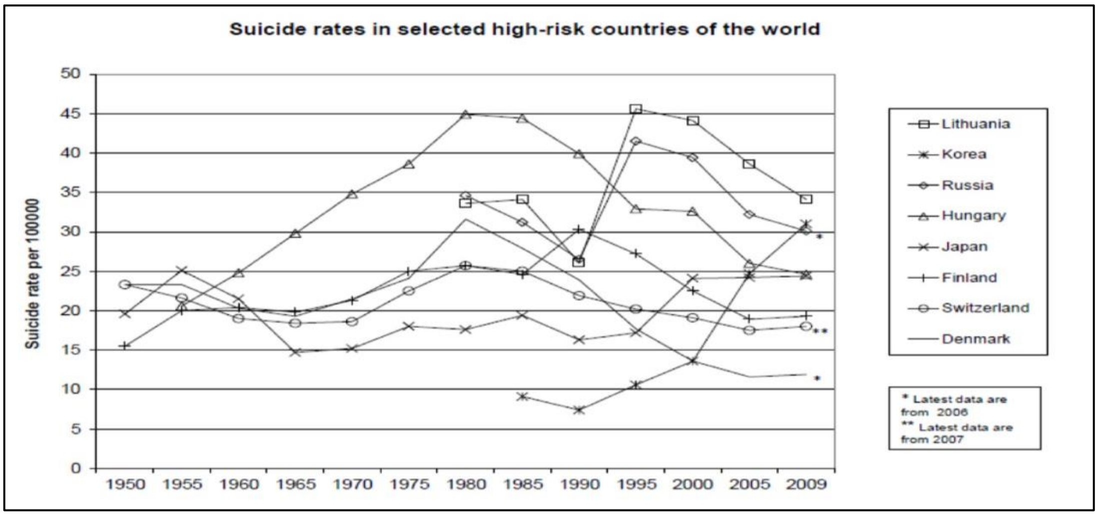

Fig 1.1: Suicide rates in selected high-risk countries of the world [2].

1.2 Deaths due to Suicide in India:

In India, pesticides, firearms, self-hanging, jumping off bridges and in front of trains are the major means by which suicide is attempted. Policies limiting access to pesticides, firearms and putting barriers on bridges and railway platforms could be some of the preventive options. In addition, counselling services and creating destigmatized platforms for discussion around these taboo subjects could be considered.

Over 153 thousand deaths due to suicides were recorded in India in 2020. Furthermore, majority of suicides were reported in the state of Maharashtra, followed by Tamil Nadu. The number of suicides that year had increased from the previous year. Some of the causes for suicides in the country were due to professional problems, abuse, violence, family problems, financial loss, sense of isolation and mental disorders [3].

1.3 Depression:

Depression is a major human blight. Globally, it is responsible for more 'years lost' to disability than any other condition. This is largely because so many people suffer from it — some 350 million, according to the World Health Organization — and the fact that it lasts for many years. Yet depression is widely undiagnosed and untreated because of stigma, lack of effective therapies and inadequate mental-health resources. Almost half of the world's population lives in a country with only two psychiatrists per 100,000 people [4].

Lots of things can increase the chance of depression, including the following:

- **Abuse.** Physical, sexual, or emotional abuse can make you more vulnerable to depression later in life.

- **Age.** People who are elderly are at higher risk of depression. That can be made worse by other factors, such as living alone and having a lack of social support.

- **Certain medications.** Some drugs, such as isotretinoin, the antiviral drug interferon-alpha, and corticosteroids, can increase your risk of depression.

- **Conflict.** Depression in someone who has the biological vulnerability to it may result from personal conflicts or disputes with family members or friends.

- **Death or a loss.** Sadness or grief after the death or loss of a loved one, though natural, can increase the risk of depression.

- **Gender.** Women are about twice as likely as men to become depressed. No one's sure why. The hormonal changes that woman go through at different times of their lives may play a role.

- **Genes.** A family history of depression may increase the risk. It's thought that depression is a complex trait, meaning there are probably many different genes that each exert small effects, rather than a single gene that contributes to disease risk. The genetics of depression, like most psychiatric disorders, are not as simple or straightforward as in purely genetic diseases such as Huntington's chorea or cystic fibrosis.

- **Major events.** Even good events such as starting a new job, graduating, or getting married can lead to depression. So can moving, losing a job or income, getting divorced, or retiring. However, the syndrome of clinical depression is never just a "normal" response to stressful life events.

- **Other personal problems.** Problems such as social isolation due to other mental illnesses or being cast out of a family or social group can contribute to the risk of developing clinical depression.

- **Serious illnesses.** Sometimes, depression happens along with a major illness or may be triggered by another medical condition.

- **Substance misuse.** Nearly 30% of people with substance misuse problems also have major or clinical depression. Even if drugs or alcohol temporarily make you feel better, they ultimately will aggravate depression [5].

1.4 Suicidal Ideation:

Suicidal Ideations (SI), often called suicidal thoughts or ideas, is a broad term used to describe a range of contemplations, wishes, and preoccupations with death and suicide. There is no universally accepted consistent definition of SI, which leads to ongoing challenges for clinicians, researchers, and educators. For example, in research studies, SI is frequently given different operational definitions. This interferes with the ability to compare findings across studies and is frequently mentioned as a limitation in meta-analyses associated with suicidality. Some SI definitions include suicide planning deliberations, while others consider planning to be a discrete stage [6].

It is evident that suicidal ideations present in a "waxing and waning manner", so the magnitude and characteristics of SI fluctuate dramatically. It is critically important for healthcare professionals to recognize that SI is a heterogeneous phenomenon. It varies in intensity, duration, and character. As there is no "typical" suicide victim, there are no "typical" suicidal thoughts and ideations. Unfortunately, healthcare records often document SI in a binary yes/no fashion, although it encompasses everything from fleeting wishes of falling asleep and never awakening to intensely disturbing preoccupations with self-annihilation fuelled by delusions [6].

1.5 Motivation:

After surveying as many as twenty-two research papers in this domain, it was found that a large number of features are being used by researchers, along with different techniques, to rectify the predicted result to the best level. Furthermore, different techniques in Machine Learning are employed for detecting suicidal ideation in different types of problems. Therefore, the aim of our project is to identify suicidal intention through the detection of the depression level using different Machine Learning techniques.

1.6 Contribution:

So, in this thesis, a classifier was proposed that can determine the depression level from the several texts collected from online or manually. By this, it can be checked whether the text has a suicidal intention or not.

1.7 Organization of thesis:

In this section, the organization of the next portion of the thesis will be described.

Chapter 2 is "Literature Survey". In this chapter 22 research papers were analysed on Machine Learning based on this work domain. The papers were taken from over a decade and critically analysed in terms of accuracy, methodologies used, limitations and so on. After that 19 research papers were analysed on Deep Learning for the work process. Those papers also were taken from over a decade and critically analysed in terms of accuracy, methodologies used, limitations and so on.

Chapter 3 is "Machine Learning". Where various Machine Learning techniques were elaborated.

Chapter 4 is "Deep Learning". In this chapter Deep Learning is elaborated and various Deep Learning techniques were explained.

Chapter 5 is "Depression Detection & Suicidal Ideation". Where depression was explained and process of depression detection were also elaborated.

Chapter 6 is "Sentiment analysis and VADER". Where sentiment analysis and different types of lexicon-based sentiment analyser were discussed precisely.

Chapter 7 is "Proposed Methodology". In this chapter the architecture of the proposed method, proposed flowchart and the overall working principle was described. So, this chapter is the base of this project work and outlines the major methods that were used.

Chapter 8 is "Implementation". In this chapter the main focus is to describe which datasets were used for training and testing of the proposed method and the feature parts were also discussed. The coding environment was also presented. Hence, this chapter is more about the 'how' rather than 'what's' the functionality of our proposed system.

Chapter 9 is "Results". Where all results of the project were presented in tabular form.

Chapter 10 is "Conclusion". In this chapter the main achievement of this research was described.

Chapter 11 is "Future Scope". In this chapter, it was described in which sections will be worked on in the future to make the classifier more precise. The shortcomings of our work which is done till now and that which could be further improvised have also been added.

Finally, the thesis winds up with the reference section listing all the resources studied and analysed to do this project work.

Chapter 2

Literature Survey

2.1: Overview of related works on Machine Learning based Suicidal Ideation:

2.1.1: *Discussion on Existing Suicidal Ideation Systems based on Machine Learning Techniques:*

In this section the research papers were individually analysed to get their importance.

In paper [7] authors propose a new method that automatically detects suicidal users through their created profiles in Online Social Networks and creates global profiles allowing to build a comprehensive individual user profile. Authors have extracted several types of features from the posting content of users to build a complete profile that contribute to high suicidal user prediction. Authors have used supervised Machine Learning techniques to classify the profiles between suicidal and non-suicidal. This paper the implemented method performed well on the profiles of the selected dataset.

Researchers [8] in this study focused on determining a semi-supervised method to populate the Life Corpus, using a bootstrapping technique, to automatically detect and classify texts extracted from social networks and forums related to suicide and depression based on initial supervised samples. To carry out the experiments authors use two different classifiers: Support Vector Machine (SVM) and Rasa. Authors achieve a macro F1-value of 0.80.

The study in paper [9] offers a comparative analysis of multiple Machine Learning and Deep Learning models to identify suicidal thoughts from the social media platform Twitter. The main purpose of this is to achieve better model performance and to recognize early indications with high accuracy and avoid suicide attempts. Researchers applied text pre-processing and feature extraction approaches such as Count Vectorizer and word embedding, and trained several Machine Learning models (Random Forest (RF), Support Vector Machine (SVM), Stochastic Gradient Descent classifier (SGD), Logistic Regression (LR), and Multinomial Naive Bayes (MNB)) and Deep Learning models (Long-Short Term Memory (LSTM), Bi-directional LSTM (BI-LSTM), Gated Recurrent Unit (GRU), Bi-directional GRU (BIGRU), and combined model of CNN and LSTM (C-LSTM)). The dataset is collected from live tweets using Python Tweedy API. In this study the RF model can achieve the highest classification score among Machine Learning algorithms, with an accuracy of 93% and an F1 score of 0.92. However, training the Deep Learning classifiers with word embedding increases the performance of ML models, where the Bi-LSTM model reaches an accuracy of 93.6% and a 0.93 F1 score.

Researchers In paper [10] propose a transformable based model named transformer RNN which can effectively extract contextual and long-term dependency information by using a transformer encoder and a Bi-directional Long Short-Term Memory (Bi-LSTM) structure. In this paper authors also extract linguistics and psychological features from the dataset. In this paper the authors achieved 95.0%, 94.9% and 94.9% performance in P, R and F1-score.

Gupta et al. in paper [11] focuses on analysing Reddit posts to identify users with poor mental health conditions who are on the verge of inflicting self-harm or committing suicide. In the process, Machine Learning models are built using six different classification techniques (J48, Logit Boost, Naïve Bayes, Random Forest, Sequential Minimal Optimization (SMO), and Support Vector Machine (SVM)) and Sentiment Analysis is performed to extract features that delineate the emotional mental state of an online user. Naïve Bayes classifier is the most effective and performs better with a precision value of 71.40%, thus showing an affirmative cue in solving the task of suicide risk assessment and depression detection.

In this paper [12] authors investigate whether it is possible to automatically identify suicide notes from other types of social media blogs in two document-level classification tasks. The first task aims to identify suicide notes from depressed and blog posts in a balanced dataset, whilst the second experiment looks at how well suicide notes can be classified when there is a vast amount of neutral text data, which makes the task more applicable to real-world scenarios. Furthermore, researchers perform a linguistic analysis using LIWC (Linguistic Inquiry and Word Count). Authors have presented a learning model for modelling long sequences in two experiment series and achieved an f1-score of 88.26% in experiment-1 and a f1-score of 96.1% in experiment-2.

The study of paper [13] is to develop a domain knowledge-aware risk assessment (KARA) model to improve our ability of suicide detection in online counselling systems. The authors obtain the largest known de-identified dataset from an emotional support system established in Hong Kong, comprising 5682 Cantonese conversations between help-seekers and counsellors. Of those, 682 conversations disclosed crisis intentions of suicide. The authors construct a suicide-knowledge graph, representing suicide-related domain knowledge as a computer-processible graph. The authors get to the point that KARA significantly outperformed standard NLP models, demonstrating good translational value and clinical relevance.

Paper [14] exhibits different techniques to comprehend suicidal ideation through online user contents in particularly by considering twitter data for past last two years as an objective of early detection by means of sentiment analysis and supervised leaning methods. The authors analyse the text descriptions and users' language exposes rich knowledge that can be utilize as a primary cautioning system for suicidal detection. To identify tweets exhibiting suicidal ideation, several features are extract and a set of features are proposed for training the model over the dataset by using ensemble and baseline classifiers. Based on the outcome of the baseline classifier; improved ensemble random forest (RF) algorithm.

Researches in paper [15] collect and validate suicide-related terms from the U.S. English language in 2018–2019. By validating clinical and lay terms with people on the front lines of suicide prevention, the study provides a necessary foundation for lexical analyses of suicide communication on social media. The author's survey validated common terms used to communicate about suicide.

Caicedo et al in [16] makes a systematic analysis of 28 supervised classifiers using different features of the corpus Life to detect messages with suicidal ideation and depression to know if these can be used in an automatic prevention online system. In the paper the Life Corpus, used in this research, is a bilingual text corpus (English and Spanish) oriented to the detection of suicide ideation. This corpus was constructed by retrieving texts from several social networks and its quality was measured using mutual annotation agreement. The researchers get a F-measure of 0.7148.

Paper [17] presents a methodology and experimentation using social media as a tool to analyse and detect the suicidal ideation in a better way, thus helping in preventing the chances of being the victim of this unfortunate mental disorder. According to the proposed work, the data is collected from Twitter, then pre-processed and annotated manually. Finally, various Machine Learning and ensemble methods like Naïve Bayes (NB), Multinomial Naïve Bayes (MNB), Decision tree (REPTree and J48), Logistic Regression (LR) and Support Vector Machine (SVM) are used to automatically distinguish Suicidal and Non-Suicidal tweets. Here researchers get a promising result, where Random Forest is the most effective and performs better with the accuracy of 98.5%, precision of 98.7% and recall value 98.2%.

Authors In paper [18] present an automated conversational platform that was used as a preliminary method of identifying depression associated risks. The platform was developed to understand conversations using Natural Language Processing (NLP) via Machine Learning Technique. It is a two-phased platform, the initial intent recognition phase would analyse conversation and identify associated sentiments into four categories of 'happy', 'neutral', 'depressive' and 'suicidal' states. In the final emotion nurturing phase, the platform continued with supportive conversations for the first three states while triggering a local call to a suicide prevention helpline for 'suicidal' state as a preventive measure. It has an overall accuracy score of 76%.

Paper [19] Mental Health America designed ten questionnaires that are used to determine the risk of mental disorders. The authors use the techniques SNOMED, ICD, UMLS, and Data Med. The authors use the semantic feature for their work in this paper. The authors use the dataset of Reddit and Manual depressed disorder people. In this paper the suicide indicator performed well in this paper however the numeric data is not given.

In this paper [20] authors present an approach to categorize potential suicide messages in social media which is based on unsupervised learning. This approach has five phases: the first two phases correspond to data acquisition and pre-processing where texts available in a corpus for suicide detection were taken and converted into a structured format; in the third phase, similarity between texts are computed using semantic similarity measures; traditional clustering algorithms were used to identify categories of potential suicide messages in the fourth phase; and, in last phase, using validation metrics researchers verified the usefulness of this approach to replicate the allocation of text into categories as in the original corpus data. This approach is promising for the case of binary annotations.

Du et al [21] main motive is to investigate techniques for recognizing suicide related psychiatric stressors from Twitter using Deep Learning based methods. The dataset method that is used in this project are convolutional neural networks (CNN), recurrent neural networks (RNN), named entity recognition (NER). The result of this paper is F-1 measure of 83%. The results indicate the advantages of Deep Learning. Uses of keyword in this project are Suicide, Mental health, Psychiatric stressors, social media, Deep Learning.

In paper [22] authors present a new approach that uses the social media platform Twitter to quantify suicide-warning signs for individuals and to detect posts containing suicide-related content. The main originality of this approach is the automatic identification of sudden changes in a user's online behaviour. To detect such changes, they combine natural language processing techniques to aggregate behavioural and textual features and pass these features through a martingale framework, which is widely used for change detection in data streams.

Researchers In paper [23] focused on validating the use of Machine Learning algorithms for Twitter data against empirically validated measures of suicidality in the US population. Using ML, the Twitter feeds of 135 Mechanical Turk participants were compared with validated, self-report measures of suicide risk. The result of suicidal rate in 92% of cases (sensitivity: 53%, specificity: 97%, positive predictive value: 75%, negative predictive value: 93%). Using social media data, evidence for suicidality can be measured in nonclinical populations.

In paper [24] authors examine whether the level of concern for a suicide-related post on Twitter could be determined based solely on the content of the post, as judged by human coders and then replicated by Machine Learning. From 18th February 2014 to 23rd April 2014, Twitter was monitored for a series of suicide-related phrases and terms using the public Application Program Interface (API). This study demonstrated that it is possible to distinguish the level of concern among suicide-related tweets, using both human coders and an automatic machine classifier. It has an overall accuracy score of 76%.

Nikfarjam et al [25] present a method for sentiment analysis of suicide notes submitted to the i2b2/VA/Cincinnati Shared Task 2011.900 suicide notes were labelled with the possible emotions that they reflect. It proposes a new approach to extract the sentence's clauses and constitutive grammatical elements and to use them in syntactic and semantic feature generation. It reaches a precision of 41.79 with a recall of 55.03 for an f-measure of 47.50.

In paper [26] the author creates a system to detect emotions in suicide notes. In this case, there are three hybrid approaches use for distinguishing between the different categories. The first approach considers single label multi-class classification, where SVM and CRF classifiers are trained. The second approach trains individual binary classifiers (SVM and CRF) and, the third approach is a combination of binary and multi-class classifiers (SVM and CRF) trained on different subsets of the training data. Finally, the result is F1 score of 45.6%, whereas our best Recall (43.6%) was obtained using the third system.

In paper [27] the author created a sentiment classification system that can identify subjective and objective categories. They developed a hybrid system using SVM. The system consists of three types of classification-based systems: spanning n-gram, bag-of-n-gram, pattern matching. The project is assessed by the overall micro-averaged precision, recall and F-measure. The result is F-measure of 0.59. The results of this project indicated that classifying fine-grained sentiments at sentence level is a non-trivial task. Uses of keyword in this project are sentiment analysis, suicide note, spanning n-gram, web data.

Sohn et al [28] implement three systems that have been trained on suicide notes provided by the I2B2 challenge organizer—a Machine Learning system, a rule-based system, and a system consisting of a combination of both. The author's Machine Learning system trains on re-annotated data in which apparently inconsistent emotion assignment is adjusted. Then, the Machine Learning methods by RIPPER and multinomial Naïve Bayes classifiers, manual pattern matching rules, and the combination of the two systems were tested to determine the emotions within sentences. The authors produce a micro-average F-score of 0.5640.

Table 2.1 presents the comparison of existing research works on detecting suicidal tendencies using ML approaches in a tabular manner.

Table 2.1: Comparison of selected research work on suicidal ideation based on ML

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
An across online social networks profile building approach: Application to suicidal ideation detection [7]	Mbarek et al (2022)	automatically detects suicidal users through their created profiles in Online Social Networks.	not using network embedding techniques and Deep Learning techniques. It does not handle fake profiles.	Random forest (RF), Bayes Net (BN), SVM, Decision Tree (DT) and Adaboost (AdaB). Tuser algorithm is also used to search for their matched profiles in Twitter.	Emotional features, Stylometric features, Temporal features, Timeline features, Account features	Twitter (TM), YouTube (YT), Tumblr (TM)	F1-score: 85.4%
Bootstrapping semi-supervised annotation method for potential suicidal messages [8]	RWA Caicedo et al (2022)	automatically detect and classify texts collected from social media related to suicide and depression-based samples.	This system will not be able to detect suicidal users in social networks in other languages	SVM classifier (using the features BoW, BoW+Embeddings, Tf/Idf, and Tf/Idf + Embeddings) and Rasa (with default features extraction system)	Bag of Words (BoW) features, Lexical & Syntactical features	Life corpus, Reddit corpus, Department of Criminal Justices	Macro F1-score: 0.78–0.81

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning [9]	Haque et al (2022)	To identify suicidal thoughts from the social media platform Twitter	improve the model's performance and produce a practical online application for clinical psychologists and healthcare practitioners	NLTK, Random Forrest (RF), Support Vector Machine(SVM), Stochastic Gradient Descent classifier (SGD), Logistic Regression (LR), and Multinomial Naive Bayes (MNB), Long-Short Term Memory (LSTM), Bi-directional LSTM (BI-LSTM), Gated Recurrent Unit (GRU), Bi-directional GRU (BIGRU), combined model of CNN and LSTM (C-LSTM), Keras, Tweepy API, VADER, TextBlob, CountVectorizer and word embeddings.	Sentimental, Emotional, n-gram Features	Twitter subreddit "SuicideWatch" (SW)	F1 score: 0.93

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Automatic identification of suicide notes with a transformer-based deep learning model [10]	Zhang et al (2021)	classifying suicide notes collected from social media.	collect more precise data from different social media and groups of people. Semi-supervised and unsupervised approaches are not used. It is not directly interpretable so not suitable for clinical decision-making process	Linguistic Inquiry and Word Count software (LIWC 2015). Techniques used: J48, Naive Bayes, Bayes Net, LMT, CNN, Bi-LSTM, Bi-LSTMAttention, DLSTMAttention, Transformer	linguistic and psychological features	Kaggle's Suicide Notes	F1-score: 94.9%
Machine Learning-Based Social Media Analysis for Suicide Risk Assessment [11]	Gupta et al (2021)	This paper analyzes Reddit posts to identify users who are on the verge of inflicting self-harm or committing suicide.	Other social media texts are not implemented, Deep Learning techniques are not explored	J48, LogitBoost, Naïve Bayes, Random Forest, Sequential Minimal Optimization (SMO), Support Vector Machine (SVM)	Sentimental, Emotional Features	Reddit C-SSRS Suicide Dataset	Precision: 71.40%

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Hierarchical Multiscale Recurrent Neural Networks for Detecting Suicide Notes [12]	Schoene et al (2021)	automatically identify suicide notes from other types of social media and clasify them	This paper does not follow linguistic patterns on these type of textual data so the accurate classification is not possible here. In future these linguistic differences could be helpful for analysis of mental health issues online.	RNNs, Dilated LSTM, Maximum Entropy, Bi-LSTM	Linguisitc, Sentimental, Emotional features	Texas Department of Criminal Justices (2019) , Tumbler(2013): Suicide notes, blogger.com	F1-score: 96.10%(Experiment-2) F1-score: 88.26%(Experiment-1)
Detecting suicide risk using knowledge-aware natural language processing and counseling service data [13]	Xu et al (2021)	The objective of this study is to develop a domain knowledge-aware risk assessment (KARA) model to improve our ability of suicide detection in online counseling systems.	the proposed model is a black box because the inner decision process of neural networks is too complex to be translated into transparent rules.	KARA,Bi-LSTM,MLP	semantic	5682 Cantonese conversations between help-seekers and counselors. Of those 682 conversations disclosed crisis intentions of suicide.	F-score: 0.870(KARA) F-score: 0.791(Bi-LSTM)

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Suicidal ideation prediction in twitter data using Machine Learning techniques [14]	kumar et al(2020)	to analyse the data from online social media mainly twitter and detect suicide ideation	The paper uses small dataset for their experiment thats way they get that amount of accuracy	Logistic regression, NB, RANDOM FOREST, XGBoost	TF-IDF,LIWC,stastical feature	TWITTER	F-score: 0.99(RF)
Social Media and Suicide: A Validation of Terms to Help Identify [15]	Parrott et al (2020)	In this paper they find What terms are commonly used when people communicate about suicide and phrases from the American English language were compiled.	The lexicon did not capture international phrases. It also did not document less direct language, such as expressions of emotion.	98 terms related to suicide were collected from online, academic, and other sources. Mental health professionals and members of the electronic mailing list of the American Association of Suicidology were asked to validate terms.	sampling approaches emailing practicing	Internet-based resources such as blogs. message boards. news sites.	The survey validated common terms used to communicate about suicide.

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Assessment of supervised classifiers for the task of detecting messages with suicidal ideation [16]	R.W. Acuna Caicedo et al (2020)	The objective of this project is to detect the suicide indication from the received texts form several social networks.	A limitation in this research is that, because it is oriented to computer science, we do not delve into the area of psychology, which is addressed in other research related to the Life research platform.	28 supervised classifier algorithms. Life Corpus developed by the research group from Natural Language Processing and Information Systems, ascribed to the University of Alicante.	Bag of Words (WORD) Bag of Stems (STEM) Bag of Lemmas (LEMMA) Bag of SYNSETS (SYNSET) Bag of POS (POS)	A bilingual corpus (English and Spanish) oriented to suicide. Texts from several social networks like Twitter,Weibo and Netlog.	F-score: 0.7148.
Detection of Suicidal Ideation on Twitter using Machine Learning & Ensemble Approaches [17]	S.T.Rabani et al (2020)	Used to automatically distinguish Suicidal and Non-Suicidal tweets..	The connectivity between suicidal users is not explored, Blog posts will be investigated later, Multi-class classification is not used, Deep Learning algorithms are not used	WEKA tool, Naïve Bayes (NB), Multinomial Naïve Bayes (MNB), ZaroR, Decision tree (REPTree and J48), Logistic Regression (LR), SMO and Support Vector Machine (SVM)	TFIDF BoW Manual Features	Twitter Application Programming Interface (API)	Accuracy: 98.5%
Recognizing Suicidal Intent in Depressed Population using NLP: A Pilot Study [18]	Hassan et al(2020)	It focuses on developing an effective diagnostic tool for depression who demonstrate the need for clinical treatment.	the nature of VPA-DR conversation initiation technique.	VPA-DR, Dialogflow Machine Learning (ML) & SVM	Emotional feature	Kaggle dataset	Accuracy: 76%

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Question Answering for Suicide Risk Assessment using Reddit [19]	Alambo et al(2019)	To determine suicidal intend by asking them some designed question	The error margin a predicted Reddit post's suicide risk severity level differs from the annotator's label.	SNOMED, ICD, UMLS, and DataMed	Semantic features	Reddit, Manual depressed disorder people	The suicide idendicator performed well in this papper however the neumeric data is not given
An Unsupervised Learning Approach for Automatically to Categorize Potential Suicide Messages in Social Media [20]	Parraga-Alava et al (2019)	objective is to categorize suicide messages collected from social media which is based on unsupervised learning.	in future the authors should use more robust similarity metrics and clustering algorithm to improve match rates. And use more texts	NLTK (Natural Language Toolkit),"tm" libraries, k-means, Partitioning Around Medoids (PAM), Hierarchical clustering	semantic features	Life Corpus	F1-score: 87%
Extracting psychiatric stressors for suicide from social media using Deep Learning [21]	Du et al (2018)	To detect early prevention of suicidal behaviors and suicide.	Highly imbalanced dataset,lack data for exact mental health status	Extra Trees (ET), Random Forest (RF), Logistics Regression (LR) and (SVM)	linguistic, lexical, syntactic, n-grams,	Twitter, Social media,Mental health status	F1-score: 83%,

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Detection of Suicide-Related Posts in Twitter Data Streams [22]	Vioulès et al(2018)	Social media platform Twitter to quantify suicide-warning signs for individuals and to detect posts containing suicide-related content.	the results of the full methodology run on only two Twitter users.	multinomial Naïve Bayes, Sequential Minimal Optimization (SMO) with a poly kernel, C4.5 decision tree (J48), nearest neighbor classifier (IB1), multinomial logistic regression, rule induction (Jrip), Random Forest, SMO with a Pearson VII universal kernel function (PUK).	n-grams, symptoms, pronouns, and swear components.andbehavioral	twitter	F1-score: 0.334
Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality [23]	Braithwaite et al (2016)	The objective of this study is to validate the use of Machine Learning algorithms for Twitter data against empirically validated measures of suicidality in the US population.	This study has a number of limitations as well as strengths.Novel research ideas are often tested. The results provide strong evidence i.e. reliably able to differentiate those who are clinically significantly suicidal from those who are not.	Tree Learning Algorithm	Internet-Based Profile and Linguistic	social media; twitter	Accuracy: 92%

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Detecting Suicidality on Twitter [24]	O'Dea et al (2015)	To design and implement an automated computer classifier that could replicate the accuracy of the human coders using recall and precision metrics.	unclear whether the 'strongly concerning' tweets were genuine statements of suicidality.	SVM, Logistic Regression	Statistical analysis & Semantic features	twitter	Accuracy: 76%
A Hybrid system for emotion extraction from suicide notes [25]	Nikfarjam et al(2012)	present a method for sentiment analysis of suicide notes submitted to the i2b2/VA/Cincinnati Shared Task 2011.	the rules were not enough to cover the test cases for some emotion categories (hopefulness, blame, anger and abuse) which caused their result to become zero.	rule-based engine and trained a Support Vector Machine (SVM)	syntactic, clausal, TF-IDF and semantic features	suicide notes (around 600 training notes and 300 test notes)	F-score: 47.50%
Three Hybrid Classifiers for the Detection of Emotions in suicide notes [26]	Liakata et al (2012)	The objective of this study is to creating a system able to detect emotions in suicide notes.	A drawback in using SVMs is that one cannot easily model the sequence of categories in a message with- out introducing errors.	JRip and SMO &LibSVM and CRFSuite,	content-based ,structural,ngram and unigram	social media; twitter;	F1-score: 45.6%

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Suicide Note Sentiment Classification: A Supervised Approach Augmented by Web Data [27]	Xu et al (2012)	To create a sentiment classification system for the Fifth i2b2/VA Challenge Track 2, which can identify thirteen subjective categories and two objective categories.	No-label sentences are labeled& Multiple labels compete with each other.	linear SVM,POS patterns,	Computing and ranking , selection through LiveJournal corpus	Suicide note, spanning n-gram, web data	F-score: 0.59.
A Hybrid Approach to Sentiment Sentence Classification in Suicide Notes [28]	Sohn et al(2012)	This paper describes the sentiment classification system. Challenge. The sentiment classification task is to assign any pertinent emotion to each sentence in suicide notes.	The major advantage we observed was that it generalized well across the problem space, when provided with sufficient training data.	MNB, RIPPER, Token normalization, Classifier Ensemble, Corpus re-annotation, GENIA tagger, SVM	sentiment classification	The training set consists of 600 actual suicide notes and the gold standard annotation of the emotional sentences.	F-score: 0.5640.

2.1.2: *Summary finding and Gap analysis of Machine Learning based Research:*

Papers from the year 2012 to 2022 have been taken into account in this study. The following table (Table 2.2) gives a report of the number of papers analysed so far.

Table 2.2: Year wise count of selected research papers on Machine Learning

Year	No of papers
2022	3
2021	4
2020	5
2019	2
2018	2
2016	1
2015	1
2012	4
Total =	22

In this section, different research papers were described that were analyzed before proposing our own classifier.

Lexical, Syntactical and Emotional features extraction is a very important task for the first level of screening. From text data if it is needed to detect depression level or suicidal ideation then these features are very useful because they work well in very less time complexity and without using complex algorithms. The following papers [7], [8], [9], [11], [12], [18], [21], [25] give stress on the lexical, syntactical and emotional features of the dataset (Either training or in test dataset)

According to the literature survey it is seen that Stylometric, Sentimental, Linguistic, Semantic, TF-IDF features are useful for prediction of suicidal ideation more precisely. The following papers [10], [13], [14], [17] emphasize these kinds of features. The paper [7], [9], [12] not only uses lexical, syntactical and emotional features but also uses stylometric, temporal, timeline, sentimental, linguistic features. N-gram, statistical, BoW, Psychological features also produces a promising result while detecting suicidal tendencies from textual dataset. The following papers [7], [8], [9], [10], [14], [16], [17], [22], [23] use this kind of feature in their research along with many previously discussed features like lexical, syntactical, sentimental, linguistic, emotional features.

The structural features are not good enough for depression detection or detection of suicidal ideation. This feature is used in paper [26] and it does not give the expected result.

In Machine Learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. The following papers [7], [8], [11], [17], [18], [21], [24], [25], [27], [28] use SVM.

In Machine Learning, Naive Bayes' and Multinomial Naive Bayes' classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong independence assumptions between the features. The following paper [9], [10], [11], [14], [17], [22], [28] uses Naive Bayes'.

Logistic regression (LR) is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. The following paper [9], [14], [17], [21], [22], [24] uses logistic regression.

J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open-source Java implementation of the C4.5 algorithm. The following paper [10], [11], [17], [22] uses J48.

The Long-Short Term Memory (LSTM) Network model is a special kind of Neural Networks which are generally capable of understanding long term dependencies. The LSTM model was generally designed to prevent the problems of long-term dependencies. The LSTM Network models generally have capacity to remove or add data carefully which is regulated by a special structure known as gates. The following paper [9], [10], [12], [13] uses Bi-LSTM, C-LSTM and Di-LSTM.

Random Forest (RF) is a powerful and versatile supervised Machine Learning algorithm that grows and combines multiple decision trees to create a forest-like structure. It can be used for both classification and regression problems. The following paper [7], [9], [11], [14], [21], [22] uses Random Forest.

Sequential minimal optimization (SMO) is an algorithm for solving the quadratic programming (QP) problem which is required during the training of support-vector machines (SVM). SMO is an iterative algorithm for solving the optimization problem. SMO breaks problems into a series of smallest possible sub-problems, which are then solved analytically. The following paper [11], [17], [22], [26] uses the SMO algorithm.

The following attributes have been collected: Year, Title, Objective, limitation, technique,

Corpus, Features, and Result from the given papers. It is very evident from the frequency distribution of the papers based upon a specific year that most of the papers are from the past 3 years. So the proposed method is compared with the newly invented methods.

In the graph (Fig 2.1) that follows next, the maximum and minimum accuracy levels reported in existing works for a specific year have been demonstrated, via blue dots and red squares respectively.

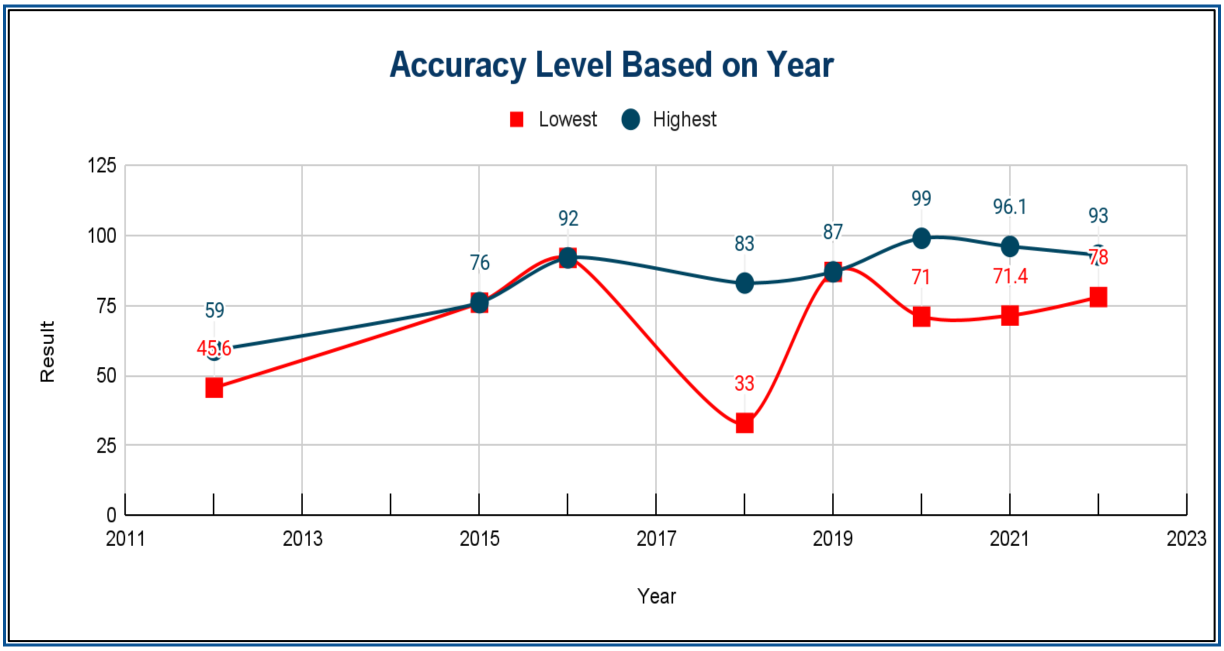

Fig 2.1: Analysis of accuracy scores obtained from ML-based existing works

Thus, it can be inferred that in 2020, the maximum percentage in terms of accuracy was acquired. However, depending upon the technique and dataset size the conclusion can be changed.

The tabularization of these papers based on their techniques is represented below.

From the above table it can be easily inferred that most of the researchers went for the SVM. It is not only one of the most reliable methods but also is capable of producing good results. Next follows the Naive Bayes method which is also used in various papers. According to the Literature Survey Semantic, Sentimental, Emotional and TF-IDF features are very much helpful to detect Depression and identify Suicidal Ideation, these features are used in most of the papers. Although structural and stylometric features cannot solely help to identify Suicidal Intent, they are used in quite a few papers.

From previously analysed papers, the limitations of the project can be easily eliminated. Some papers only introduced one or two classifiers for their work, while others used small datasets or included unwanted features. Therefore, the focus of our project work will be on how to avoid these kinds of limitations.

2.2: Overview of related works on Deep Learning based Suicidal Ideation:

2.2.1 *Discussion on Existing Suicidal Ideation Systems based on Deep Learning Techniques:*

In this section the research papers were individually analysed to get their importance.

In paper [29], the study focuses on automatic recognition of suicidal posts. The authors explore the performance of LSTM-CNN, CNN, combined class of deep neural networks as their proposed model for detection of suicide ideation tasks to improve the state-of-the-art method. To detect suicide ideation, classification models are built on a Reddit social media dataset. The accuracy obtained is 93.8% for LSTM-CNN hybrid model.

Du et al. [30] conducted a study aiming to identify suicide-related psychiatric stressors from Twitter utilizing Deep Learning methods. The researchers employed Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) models for their analysis. Various features were incorporated in their approach, encompassing lexical features, syntactic features, context features, distributional representation of words, and domain knowledge features. The outcomes of their investigation revealed that the CNN model achieved an accuracy of 74%, while the Bi-LSTM model attained an accuracy of 72%.

Researchers in the paper [31], present a hybrid model designed to identify depression by analyzing the textual posts of users. The primary objective of this research is to detect signs of depression at an early stage by examining the posts made by Reddit users. The model incorporates various features including Trainable Embedding, Glove Embedding, Word2Vec Embedding, Fastest Embedding, and Metadata. The evaluation of the model yielded promising results, with an f1 score of 81% achieved using Word2Vec Embedding in combination with Metadata.

The focus of work [32], is to develop different Deep Learning models for the purpose of emotion detection. The authors accomplished this by creating a meticulously annotated corpus called CEASE, which consisted of suicide notes in English. They then proceeded to train and evaluate various Deep Learning models using this carefully curated dataset. The study involved the development and training of three fundamental Deep Learning models: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) networks utilizing pre-trained word embeddings. The obtained results indicated that the CNN achieved an accuracy of 59.54%, followed by the GRU with 58.70%, and the LSTM with 58.08%.

The authors of the paper [33], address the longstanding challenge of accurately predicting suicide attempts, which has remained close to chance levels for many years. The study primarily concentrates on investigating different Deep Learning architectures such as convolutional neural networks, long short-term memory networks, and neural network synthesis. Notably, the study focuses on exploring a set of features that have not been previously examined for suicide risk assessment or screening. The reported results of the study indicate F1 scores of .57 and .48, demonstrating the effectiveness of the proposed approach.

The objective of research paper [34] was to introduce an automated detection and prediction system designed to identify severe depression by analyzing sentiments and emotions expressed across various platforms, including social networks, blogs, emails, and textual notes. To accomplish this, the authors established a sequence of procedures for extracting characteristics from notes that reflect the emotional state of individuals. The Linguistic Inquiry and Word Count (LIWC) software was employed to identify emotionally charged words and cognitive processes present in the language used. The findings of the study demonstrated an accuracy of 86.61% using sentiment analysis and linguistic features in conjunction with the logistic tree regressor algorithm.

Main focus of the work [35] was on the classification of suicide notes obtained from social media. However, the study had certain limitations, such as the need for more precise data collection from diverse social media platforms and various groups of individuals. Additionally, the paper did not employ semi-supervised or unsupervised approaches and lacked direct interpretability, making it unsuitable for clinical decision-making processes. To address these shortcomings, the authors proposed the use of a Transformer-based Recurrent Neural Network (Transformer RNN) to automatically identify suicide notes. The outcome of this research demonstrated a promising F1-score of 94.9% for the Transformer RNN model.

Haque et al. [36], focuses on employing a Deep Learning approach for classifying suicide versus depression. The authors proposed a framework that utilizes word embedding models to convert textual posts into numerical word embeddings. Specifically, the techniques employed in this study involved BERT embeddings coupled with various neural network architectures, namely CNN, a fully-dense neural network, and Bi-LSTM neural network, as well as GUSE embeddings with a fully-dense neural network. For dimensionality reduction, UMAP was utilized, and the clustering algorithm involved K-Means and GMM. The results of the paper demonstrated an impressive AUC score of 98.18% for the guise-dense model with UMAP and K-Means.

The primary objective of paper [37] was to assess the precision of diagnostic codes and employ text-mining techniques to extract symptom profiles and functional impairments from electronic health records. To conduct this study, the researchers utilized the Integrated Medical Database at National Taiwan University Hospital (NTUH-IMD) spanning from January 1, 2006, to September 30, 2016. Additionally, they investigated the accuracy of text mining in diagnosing major depressive disorders. The findings of the study indicated F-scores ranging from 0.774 to 0.753.

The researchers in paper [38] aimed to improve the prediction sensitivity of suicide risk in young individuals by developing and validating a deep graph neural network (GNN) model. They incorporated multi-dimensional questionnaires and suicidal ideation to enhance the model's performance. Although the paper acknowledged limitations in predicting major depressive episodes due to small datasets, the researchers successfully developed a specialized GNN model for predicting acute suicidal ideation within a two-week timeframe. The model demonstrated improved sensitivity compared to baseline models and was validated using an external test set. The results showed a sensitivity of 76.3%, specificity of 83.4%, accuracy of 83.3%, and an area under the curve (AUC) of 0.878 (95% CI, 0.855–0.899) by employing an ensemble of GIN models with different sampling methods.

The authors of paper [39], developed deep neural network models to predict suicide risk from user-generated social media texts. They incorporated theory-driven risk factors and employed advanced Machine Learning techniques. The dataset included CWE representation, ANN modeling, psychodiagnostic measures, and TF-IDF analysis of everyday language. The results showed an AUC of 0.621 with a 95% confidence interval of 0.576 to 0.657.

In research paper [40], the aim was to develop a Deep Learning model for identifying a user's mental state using their posting information. The authors collected posts from mental health communities on Reddit, including subreddits such as r/depression, r/Anxiety, r/bipolar, r/BPD, r/schizophrenia, r/autism, and r/mentalhealth. The obtained result was an impressive F1-score of 96.96%.

Researchers of paper [41], utilized CNN, RNN, LSTM, and GRU models to discreetly detect depressive symptoms in social media data, specifically using Arabic tweets, Twitter tweets, and KAGGLE datasets. The study achieved a testing accuracy of 87.23%. Shetty et al. (2020) employed sentiment analysis (SA) to analyze Twitter tweets.

Paper [42] presents a methodology for developing a suicidal ideation detection system. The approach utilizes publicly available Reddit datasets, word-embedding techniques (TF-IDF and Word2Vec) for text representation, and a combination of Deep Learning and Machine Learning algorithms for classification. By incorporating LIWC features, the system achieves an accuracy of 91.5%.

Authors of paper [43] explores three interrelated tasks: depression detection, sentiment analysis, and investigating their significance in analyzing the mental state of individuals. The study examines 66 suicide notes, comprising 33 authentic and 33 fabricated ones, employing discourse analysis techniques to identify distinguishing features. The paper reports an accuracy of 75.34% in differentiating genuine notes from fake ones.

The researchers in the paper [44], assessed the viability of detecting suicidal thoughts using various datasets and advanced Deep Learning models such as RNN, CNN, and Attention-based models. The study replicated existing social media-based models for detecting suicidal ideation and examined different configurations of features and models across multiple datasets to evaluate their effectiveness. The results highlighted the performance of different models and feature combinations for each dataset.

Ghosh et al. [45] present an end-to-end VAD-assisted transformer-based multi-task network for detecting emotion and its intensity in suicide notes. The introduced method consistently outperforms existing state-of-the-art approaches on various datasets, showcasing its ability to generalize to related tasks. The paper reports an F1 Score of 86.86% as the result.

The authors of the paper [46] improve text representation using sentiment scores and latent topics and proposes relation networks for detecting suicidal ideation and mental disorders with associated risk indicators. However, it has limitations in predicting low-risk suicidal ideation, showing similar performance to other baselines. The techniques employed include LSTM, Bi-LSTM, CNN, Fast Text, RCNN, SSA, and RN. The reported accuracy in the paper for the Reddit SWMH dataset using relation networks is 0.64.

The primary objective of the paper [47] is to construct a predictive model for identifying suicidal intent in social media posts. The utilized features were Word2Vec and GloVe, while the techniques and datasets included Logit, CNN, LSTM, BERT, Electra, and a suicide and depression detection dataset. The obtained result showed an accuracy of 0.97 using Electra.

Table 2.3 presents the comparison of existing research works on detecting suicidal tendencies using DL approaches in a tabular manner.

Table 2.3: Comparison of selected research work on suicidal ideation based on DL

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Detection of Suicide Ideation in Social Media Forums Using Deep Learning [29]	Tadesse et al (2020)	Early detection of suicide ideation through Deep Learning and Machine Learning classification approaches	Data deficiency and annotation bias	RF,NB,SVM, XGBOOST,LSTM, CNN	Statistics,TF-IDF,BoW,Word2vec	Reddit's SuicideWatch BBS	Accuracy= 93.8% (LSTM-CNN combined)
Extracting psychiatric stressors for suicide from social media using Deep Learning [30]	Du et al (2018)	To investigate techniques for recognizing suicide related psychiatric stressors from Twitter using Deep Learning	–	CNN, Bi-LSTM	GloVe embedding, MIMIC embedding, GloVe Twitter embedding,	No specific dataset is given but they used dataset of suicide-related tweets collected from Twitter streaming data	Accuracy- CNN (74%), Bi-LSTM (72%)
Early Depression Detection from Social Network Using Deep Learning Techniques [31]	Shah et al (2020)	To proposed a hybrid model that can detect depression by analyzing user's textual posts.	Used only one Deep Learning classifier in this paper	Bi-LSTM	Trainable Embedding, GloveEmbedding, Word2VecEmbedding, FastextEmbedding, Metadata	Reddit dataset	f1 score - 81% (Word2VecEmbed+Meta)

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
CEASE, a Corpus of Emotion Annotated Suicide notes in English [32]	Ghosh et al (2020)	To develop various Deep Learning models to perform emotion detection.	Divided the emotion in very vast category that's why they got less accrucy	CNN, GRU, LSTM	GloveEmbeddin g	manual scavenging of actual suicide notes from the internet.	Accuracy- CNN (59.54%), GRU(58.70%), LSTM(58.08%)
An Investigati on of Deep Learning Systems for Suicide Risk Assessmen t [33]	Morales et al (2019)	To show the ability to predict suicide attempts has been near chance for decades.	Skip-gram when facing out-of-vocabulary words.	SVM,CNN,K NN,RF,Naiv e Bayes,LSTM	BoW,POS,Word Embeddings,To ne,syntax,NER	Reddit & subreddit's SuicideWatch	F1-score of .57 and .48 respectively.
Deep Learning Algorithm for Suicide Sentiment Prediction [34]	Boukil et al(2019)	To present an automate d detection and predictio n system depressio n through analyzing sentiment s and feelings	–	Naive Bayes,CNN, KNN	BoW,TF-IDF	Suicide & Nonsuicidial notes,several Websites (Facebook, Twitter etc)	Accuracy=86.61%(Sentiment analysis and linguistic features with the logistic tree regressor algorithm)

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Automatic identification of suicide notes with a transformer-based deep learning model [35]	Zhang et al (2021)	classifying suicide notes collected from social media.	collect more precise data from different social media and groups of people. Semi-supervised and unsupervised approaches are not used. It is not directly interpretable so not suitable for clinical decision-making process	J48, Naive Bayes, Bayes Net, LMT, CNN, Bi-LSTM, Bi-LSTMAttention, DLSTMAttention, Transformer RNN	linguistic and psychological features	Last Statements: Texas Department of Criminal Justices (2019) Suicide Notes: Kaggle's Suicide NotesSuicide notes Neutral posts: ten subreddits	F1-score: 94.9% (TransformerRNN)
Deep Learning for Suicide and Depression Identification with Unsupervised Label Correction [36]	Haque et al (2021)	suicide versus depression classification method through a Deep Learning approach.	There is no way to evaluate which posts are accurate.	BERT embeddings with a CNN (bert-cnn), BERT with a fully-dense neural network (bert-dense), BERT with a Bi-LSTM neural network (bert-Bi-LSTM), and GUSE with a fully-dense neural network (guse-dense). UMAP K-Means and GMM	(TFIDF, CVec, HVec)	Reddit C-SSRS Dataset + Pang and Lee Movie Review Dataset	AUC: 98.18 (guse-dense with UMAP-KMeans)

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records [37]	C.-S. Wu, et al. (2019)	evaluate the accuracy of diagnostic codes and use text mining to extract symptom profile	Clinical and research implications	CNN, Bi-LSTM	Words,Part-of-speech tags,Symptom- and function-dictionary in the CRF model.	Integrated Medical Database,psychiatric diag-nosis (ICD-9-CM: 290–319 or ICD-10-CM: F00-F99)	F-scores of 0.774–0.753.
Deep graph neural network-based prediction of acute suicidal ideation in young adults [38]	Choi et al (2021)	To developed and validated a deep graph neural network model that increased the prediction sensitivity of suicide risk in young using multi-dimensional questionnaires	Prediction of major depressive episodes using small datasets	SVM,CNN,LR,GNN	PHQ-9,PHQ_5, PHQ_6, and PHQ_8,SMOTE-NC,PHQ_2 and STAI-S,pseudo-labels for the MaDE section	CNG the presence of MDD,KAIST,Gachon University Hospital,Samsung Medical Center in Seoul,Seoul National University	Accuracy = 90.90% (Combined SVM,GIN-MaDE model)

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Deep neural networks detect suicide risk from textual facebook posts [39]	Ophir et al (2020)	To predict suicide risk from Facebook postings directly and a Multi-Task Model, which included hierarchical.	Offine, external validations of suicide risk,mental health conditions in general and suicide risk in particular	ML,NLP,ELMo,ANN	CWE representation method, ANN modeling, psychodiagnostic measures, and analysis of everyday language,TF-IDF	Facebook	[AUC=0.621, 95% CI: 0.576, 0.657]
A Deep Learning model for detecting mental illness from user content on social media [40]	Kim et al (2020)	Developed a Deep Learning model to identify a user's mental state based on his/her posting information.	Can't explain the difference between socio-demographic and regional. In future study, adopting an ensemble approach with multiple binary classifcation models,	Synthetic minority over-sampling,CNN	TF-IDF,Continuous bag-of-words representation (CBoW) models	six mental-health-related subreddits, r/depression, r/Anxiety, r/bipolar, r/BPD, r/schizophrenia, r/autism and r/mentalhealth	F1-score of (96.96%)
An optimized Deep Learning approach for suicide detection through Arabic tweets [41]	Baghdadi et al (2020)	To analysis of social media data allows for the discreet detection of depressive symptoms	–	CNN, RNN,LSTM, and GRU	No Features used	Arabic tweets dataset,Twitter tweets,KAGGLE datasets	

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models [42]	Aldhyani et al (2022)	Proposed a methodology based on experimental research for building a suicidal ideation detection system using publicly available Reddit datasets,.	Handling words that are not in the selected vocabulary size as maximum features, it is nevertheless optimal for NLP tasks.	CNN,Bi-LSTM,SVM, NLTK,ML XG Boost	textual,LIWC,TF-IDF,Word Embedding	Reddit,50,000 tweets,Russian social networking platform	91.5% accuracy using LIWC features,
A Multitask Framework to Detect Depression, Sentiment and Multi-label Emotion from Suicide Notes [43]	Ghosh et al(2021)	Focuses on learning three closely related tasks, viz. depression detection, sentiment citation, and to investigate their impact in analysing the mental state of the victims.	Does not allow to capture other secondary emotions for the sentences that also carry vital emotional traits.	NLP,DL,ML	No Features used	900 suicide notes,The CEASE,Ekman's notion of emotion,4932 instances and annotated for depression, sentiment and emotion labels.	75.34 % accuracy

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
A Quantitative and Qualitative Analysis of Suicide Ideation Detection using Deep Learning [44]	Long et al (2022)	Evaluated the feasibility of detecting suicidal ideation using multiple datasets and different state-of-the-art Deep Learning models.	Not generalisable ,No standard or perfect solution for selecting and preparing features	CNN,RNN,BI-LSTM	TF-IDF,300 features rotation forest	Twitter,consisting of 660 tweets,UMD Reddit Suicidality,BERT,RoBERT	Accuracy range of 45%-60%
VAD-assisted multitask transformer framework for emotion recognition and intensity prediction on suicide notes [45]	Ghosh et al (2023)	The main objective of this paper to introduces an end-to-end VAD-assisted transformer-based multi-task network for detecting emotion and its intensity in suicide notes.	Small, imbalanced dataset and the lack of a measure that assesses the history of suicide attempts.	CNN,LSTM, SVM	word embeddings,VAD	suicide notes, CEASE-v2.0,VAD-BERT	F1 Scores 86.86%

PAPER TITLE [ID]	AUTHORS (YEAR)	OBJECTIVE	LIMITATIONS	TECHNIQUES	FEATURES	DATASET	REPORTED RESULT
Suicidal ideation and mental disorder detection with attentive relation networks [46]	Ji et al (2021)	Enhances text representation with lexiconbased sentiment scores and latent topics and proposes using relation networks to detect suicidal ideation and mental disorders with related risk indicators.	Fails in predicting lowrisk suicidal ideation, with a similar performance to other baselines	LSTM,Bi-LSTM,CNN, Fast Cash,RCNN, SSA,RA	word embeddings,TF-IDF	Twitter,UMD reddit,Reddit SWMH	Accuracy = 0.64 (Reddit SWMH, With RN)
Suicidal text detection [47]	Yi et al (2022)	To build a predictive model to detect suicidal intent in social media posts	–	Logit,CNN,LSTM,Bert,Electra	Word 2 vec,glove	Suicide and depression detection dataset	Accuracy = 0.97 (electra)

2.2.2: *Summary finding and Gap analysis of Deep Learning:*

Papers from the year 2018 to 2023 have been taken into account in this study. The following table (Table 2.4) provides a report on the number of papers analysed so far.

Table 2.4: Year wise count of selected research papers on Deep Learning

Year	No of papers
2023	1
2022	3
2021	5
2020	6
2019	3
2018	1
Total =	19

In this section, different research papers were described that were analysed before the proposal of the classifier.

According to the literature survey it is seen that TF-IDF features are useful for prediction of suicidal ideation more precisely. The following papers [29], [34], [35], [36], [39], [40], [42], [44], [46] emphasize these kinds of features.

BoW, Psychological features also produces a promising result while detecting suicidal tendencies from textual dataset. The following papers [29], [33], [34], [35], [39], [40] use this kind of feature in their research along with many previously discussed features like lexical, syntactical, sentimental, linguistic, emotional features.

Word 2vec word embedding is creating vectors of the words that are distributed numerical representations of word features. The following papers [29], [31], [33], [42], [45], [47] emphasize these kinds of features.

Glove Embeddings are a type of word embedding that encode the co-occurrence probability ratio between two words as vector differences. It is an algorithm for obtaining vector representations for words. The following papers [30], [31], [32], [37] emphasize these kinds of features.

CNN is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. The following paper [29], [30], [32], [33], [34], [35], [36], [37], [38], [40], [41], [43], [44], [45], [46], [47] used CNN.

The Long-Short Term Memory (LSTM) Network model is a special kind of Neural Networks which are generally capable of understanding long term dependencies. The LSTM model was generally designed to prevent the problems of long-term dependencies. The LSTM Network models generally have capacity to remove or add data carefully which is regulated by a special structure known as gates. The following paper [29], [32], [33], [36], [41], [46] uses LSTM.

In Deep Learning, Bi-LSTM are learning algorithms that analyze data used for classification and data analysis. It is a bidirectional LSTM (Bi-LSTM) layer learns bidirectional long-term dependencies between time steps of time series or sequence data**.** The following papers [30], [31], [35], [36], [37], [44], [46] use Bi-LSTM.

BERT is designed to help computers understand the meaning of ambiguous language in text by using surrounding text to establish context. It is used to understand the users' search intentions and the contents that are indexed by the search engine. The following papers [36], [47] use BERT.

The following attributes have been collected: Year, Title, Objective, Limitation, Technique, Corpus, Features, and Result from the given papers. It is evident from the frequency distribution of the papers based on a specific year that most of the papers are from the past 5 years. Therefore, our method is compared with the newly invented methods.

In the graph that follows below, the result or accuracy level maximum and minimum have been structured for a specific year, where blue dots and red squares show the range of accuracy level for that specific year.

The tabularization of these papers based on their techniques is represented below.

From the above table it can be easily inferred that most of the researchers went for the Bi-LSTM, LSTM, CNN, BERT. It is not only one of the most reliable methods but also is capable of producing good results. Next follows the Naive Bayes method which is also used in various papers. According to the Literature survey BoW, TF-IDF, word2vec embedding, Glove embedding features are very much helpful to detect Depression and identify Suicidal Ideation, these features are used in most of the papers. Although structural and stylometric features cannot solely help to identify Suicidal Intent, they are used in quite a few papers.

The Pie chart (Fig 2.2) depicts the four commonly used DL classifiers in order viz. CNN, Bi-LSTM, LSTM and BERT that have been employed to detect depression and presence of suicidal tendency.

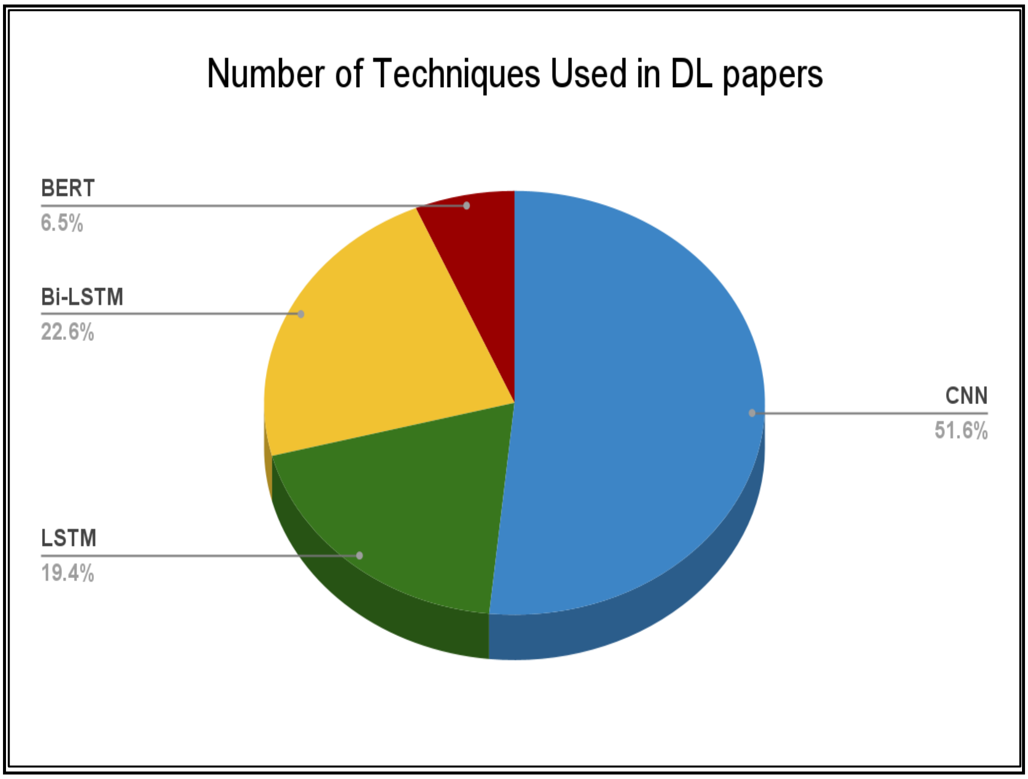

Fig 2.2: Four most used DL techniques obtained from DL-based existing work

Chapter 3

Machine Learning

3.1: Overview of Machine Learning:

Machine Learning (**ML**) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks.[48] It is seen as a part of artificial intelligence. Machine Learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so.[49] Machine Learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.[50]

3.2: Types of Machine Learning:

Based on the methods and way of learning, Machine Learning is divided into mainly four types, which are:

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Semi-Supervised Machine Learning
4. Reinforcement Learning

1. Supervised Machine Learning: Supervised Machine Learning is a type of Machine Learning where a model is trained on a labeled dataset, which consists of a set of input data and corresponding correct outputs. The goal is for the model to learn the relationships between the input data and the output labels, so that it can make predictions on new, unseen data.

In supervised learning, the training data consists of a set of examples, each of which consists of an input and an output. The model is trained by feeding it the input data, and comparing its predicted output to the correct output. The model then adjusts its internal parameters to minimize the difference between its predicted output and the correct output. This process is repeated for each example in the training dataset, and the model continues to learn and improve its performance as it processes more and more examples.

There are many different algorithms and techniques used in supervised learning, including linear regression, logistic regression, support vector machines, and decision trees. These algorithms can be used to solve a wide range of tasks, such as image classification, spam detection, and natural language processing.

One of the main advantages of supervised learning is that it can learn to make highly accurate predictions when the input data and output labels have a strong relationship. However, it requires a large amount of labelled training data in order to learn this relationship, and it may not be able to generalize to new, unseen data if the relationships between the input and output data are more complex [51].

Supervised Machine Learning can be classified into two types of problems, which are given below [52]:

- Classification (SVM, RF)
- Regression (SLR, DT)

2. Unsupervised Machine Learning: Unsupervised Machine Learning is a type of Machine Learning where a model is not given any labelled training data. Instead, the goal is for the model to discover the underlying structure of the data by finding patterns and relationships within the data itself.

There are many different algorithms and techniques used in unsupervised learning, including clustering, anomaly detection, and dimensionality reduction.

Clustering algorithms try to group together data points that are similar to one another, based on some measure of similarity. For example, a clustering algorithm might group together customers with similar purchasing habits, or group together genes with similar expression patterns.

Anomaly detection algorithms are used to identify data points that are unusual or unexpected, and may indicate some kind of problem or issue. For example, an anomaly detection algorithm might be used to identify fraudulent credit card transactions, or to detect mechanical failures in a manufacturing process.

Dimensionality reduction algorithms are used to reduce the complexity of the data by transforming it into a lower-dimensional space. This can make it easier to visualize the data and to identify patterns and relationships within it.

One of the main advantages of unsupervised learning is that it does not require any labelled training data, which can be expensive and time-consuming to collect. However, it can be more challenging to evaluate the performance of an unsupervised learning model, and it may be more difficult to apply the results of the model to real-world tasks [53].

Unsupervised Learning can be further classified into two types, which are given below [54]:

- Clustering (K-Means)
- Association (Apriori)

3. Semi-supervised Machine Learning: Semi-supervised learning is a type of Machine Learning that involves using a mix of labelled and unlabelled data to train a model. The model is able to make use of the additional unlabelled data to improve its performance, but it still requires some labelled examples in order to learn the task.

Semi-supervised learning is often used when it is expensive or time-consuming to label a large dataset, but a small amount of labelled data is still available. It can also be useful when the data is naturally biased and there are not enough examples of certain classes to train a supervised learning model.

There are many different algorithms and techniques used in semi-supervised learning, including self-training, co-training, and multi-view learning. These algorithms can be used to solve a wide range of tasks, such as image classification, natural language processing, and speech recognition.

One of the main advantages of semi-supervised learning is that it can make use of large amounts of unlabelled data to improve the model's performance, while still requiring only a small amount of labelled data. However, the quality of the model's predictions may depend on the quality of the small amount of labelled data, and it may not be as accurate as a fully supervised learning model [54].

4. Reinforcement Machine Learning: Reinforcement learning is a type of Machine Learning that involves training an agent to take actions in an environment in order to maximize a reward. The agent learns by interacting with the environment, receiving rewards or punishments based on its actions, and adjusting its behaviour accordingly.

In reinforcement learning, an agent is trained to learn a policy, which is a function that maps states of the environment to actions. The goal is for the agent to learn a policy that maximizes the cumulative reward over time.

There are many different algorithms and techniques used in reinforcement learning, including Q-learning, SARSA, and actor-critic methods. These algorithms can be used to solve a wide range of tasks, such as robot control, game playing, and recommendation systems.

One of the main advantages of reinforcement learning is that it can learn to optimize a long-term reward by taking into account the sequence of actions and their consequences. However, it can be challenging to define an appropriate reward function, and the learning process can be slow and require a large amount of trial and error [55].

Reinforcement learning is categorized mainly into two types of methods/algorithms [56]:

- Positive Reinforcement Learning
- Negative Reinforcement Learning

3.3: Some Important Machine Learning Classifiers:

- **NAIVE BAYES:** Naive Bayes is a simple but surprisingly powerful algorithm for predictive modelling. Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values. It is called naive Bayes or idiot Bayes because the calculation of the probabilities for each hypothesis is simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value P (d1, d2, d3|h), they are assumed to be conditionally independent given the target value and calculated as P(d1|h) * P(d2|H) and so on. This is a very strong assumption that is most unlikely in real data, i.e., that the attributes do not interact. Nevertheless, the approach performs surprisingly well on data where this assumption does not hold. [57]

- **SVM:** A Support Vector Machine (SVM) is a discriminative classifier formally defined by separating hyperplanes. In other words, given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay on either side. In Machine Learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.[58]

- **LOGISTIC REGRESSION**: Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. [59]

- **RANDOM FOREST:** Random Forest method is a type of supervised learning method used for classification of data. In Random Forest Algorithm, the dataset is divided into several individual trees and forms a class with similar functionalities. Then from the different classes, the final class is selected by voting the majority. The main advantage of any decision tree-based algorithm like Random Forest is the clear and understandable prediction rules that may be generated from training dataset.[60]

3.4: Some Important Features:

- **LEXICAL FEATURES:** Lexical semantics (also known as lexicon semantic), is a subfield of linguistic semantics. The units of analysis in lexical semantics are lexical units which include not only words but also sub-words or sub-units such as affixes and even compound words and phrases. Lexical units make up the catalog of words in a language, the lexicon. Lexical semantics looks at how the meaning of the lexical units correlates with the structure of the language or syntax. Lexical items contain information about category (lexical and syntactic), form and meaning. The semantics related to these categories then relate to each lexical item in the lexicon. Lexical items can also be semantically classified based on whether their meanings are derived from single lexical units or from their surrounding environment. [61]

- **SYNTACTIC FEATURES:** A syntactic category is a type of syntactic unit that theories of syntax assume. Word classes, largely corresponding to traditional parts of speech (e.g., noun, verb, preposition, etc.), are syntactic categories. In phrase structure grammars, the phrasal categories (e.g., noun phrase, verb phrase, prepositional phrase, etc.) are also syntactic categories. Dependency grammars, however, do not acknowledge phrasal categories (at least not in the traditional sense). Word classes considered as syntactic categories may be called lexical categories, as distinct from phrasal categories. The terminology here is by no means consistent, however. Many grammars also draw a distinction between lexical categories (which tend to consist of content words, or phrases headed by them) and functional categories (which tend to consist of function words or abstract functional elements, or phrases headed by them). The term lexical category therefore has two distinct meanings. Moreover, syntactic categories should not be confused with grammatical categories (also known as grammatical features), which are properties such as tense, gender, etc. [62]

At least three criteria are used in defining syntactic categories:

- The type of meaning it expresses
- The type of affixes it takes
- The structure in which it occurs

- **SEMANTIC FEATURES:** Semantic features represent the basic conceptual components of meaning for any lexical item. An individual semantic feature constitutes one component of a word's intention, which is the inherent sense or concept evoked. The linguistic meaning of a word is proposed to arise from contrasts and significant differences with other words. Semantic features enable linguistics to explain how words that share certain features may be members of the same semantic domain. Correspondingly, the contrast in meanings of words is explained by diverging semantic features. For example, father and son share the common components of 'human', 'kinship', 'male' and are thus part of a semantic domain of male family relations. They differ in terms of 'generation' and 'adulthood', which is what gives each its individual meaning. The analysis of semantic features is utilized in the field of linguistic semantics, more specifically the subfields of lexical semantics, and lexicology. One aim of these subfields is to explain the meaning of a word in terms of their relationships with other words. In order to accomplish this aim, one approach is to analyze the internal semantic structure of a word as composed of a number of distinct and minimal components of meaning. [63]

- **STYLOMETRIC FEATURES:** Stylometry or the study of measurable features of (literary) style, such as sentence length, vocabulary richness and various frequencies (of words, word lengths, word forms, etc.), has been around at least since the middle of the 19th century. These applications are usually based on the belief that there exist such conscious or unconscious elements of personal style that can help detect the mental condition at that time the text was written. [64]

Table 3.1 enlists various categories of ML Techniques along with various examples and applications of each technique for solving real world problems.

Table 3.1: Examples and applications of Machine Learning:

Category of ML	Technique	Example	Application
Supervised	Classification	SVM	Text classification
	Regression	SLR	Error Detection
Unsupervised	Clustering	K-MEANS	Image segmentation
	Association	APRIORI	Grocery store
Semi-supervised	Merging clustering	-	Image analysis
	Classification algorithms	-	Speech analysis
Reinforcement	Positive	-	Video games
	Negative	-	Text mining

Chapter 4

Deep Learning

4.1: Overview of Deep Learning:

Deep Learning is a subfield of artificial intelligence (AI) that uses artificial neural networks, which are inspired by the structure and function of the human brain, to learn from large amounts of data. It is a powerful tool for solving complex problems in a variety of domains, including computer vision, natural language processing, speech recognition, and robotics.

Deep Learning has revolutionized many industries and has been used to develop self-driving cars, improve medical diagnoses, and detect fraud in financial transactions, among other applications. Its ability to learn from large amounts of data and to generalize to new situations has made it a key technology in the AI toolkit.

The success of Deep Learning is due in part to the availability of large datasets, powerful computing resources, and new algorithms that enable faster and more accurate training of neural networks. The development of Deep Learning has also been driven by breakthroughs in hardware, such as graphical processing units (GPUs) and tensor processing units (TPUs), which can accelerate training and inference tasks.

While Deep Learning has achieved impressive results in many domains, it is not without limitations. Deep Learning models can be computationally expensive to train and require large amounts of data. They can also be prone to overfitting, where the model performs well on the training data but poorly on new data. Despite these challenges, Deep Learning remains a rapidly advancing field with many exciting opportunities for research and innovation.

4.2: Types of Deep Learning:

There are several types of Deep Learning that are commonly used in practice. Here are some of the most important types of Deep Learning:

1. **Convolutional Neural Networks (CNNs):**

CNNs are commonly used in computer vision applications, such as image classification and object detection. They are designed to identify spatial patterns in data, such as the presence of edges or shapes, and to use those patterns to make predictions.

Convolutional neural networks (CNNs) are a type of Deep Learning algorithm that are commonly used in computer vision tasks, such as image classification, object detection, and image segmentation. CNNs are designed to identify spatial patterns in data, such as the presence of edges, shapes, or textures, and to use those patterns to make predictions.

The architecture of a CNN consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply a set of learnable filters to the input image to extract features, such as edges and shapes, at multiple scales. The pooling layers down sample the output of the convolutional layers, reducing the size of the feature maps and making the network more computationally efficient. The fully connected layers take the output of the convolutional and pooling layers and use it to make predictions about the input image.

The training of a CNN involves adjusting the weights of the filters in the convolutional layers and the weights of the fully connected layers to minimize a loss function. The loss function measures the difference between the predicted output of the network and the true output, and the weights are adjusted using backpropagation and stochastic gradient descent.

CNNs have achieved impressive results in a variety of computer vision tasks, including image classification, where they have outperformed traditional Machine Learning algorithms and achieved near-human level accuracy. They have also been used for object detection, where they can identify the location and class of multiple objects in an image, and image segmentation, where they can separate an image into regions of interest.

Despite their success, CNNs have some limitations, including a tendency to overfit the training data and a requirement for large amounts of training data. However, ongoing research is addressing these limitations and improving the performance of CNNs in a variety of tasks.

2. **Recurrent Neural Networks (RNNs):**

RNNs are commonly used in natural language processing and speech recognition applications. They are designed to process sequences of data, such as words in a sentence or audio samples in a speech signal, and to use context from previous inputs to make predictions.

Recurrent neural networks (RNNs) are a type of Deep Learning algorithm that are commonly used in natural language processing, speech recognition, and time series analysis. RNNs are designed to process sequences of data, such as words in a sentence or audio samples in a speech signal, and to use context from previous inputs to make predictions.

The architecture of an RNN includes a set of recurrent connections that allow the network to maintain an internal state that captures information about the sequence it has seen so far. The output of the network at each time step depends not only on the current input but also on the previous state. This allows the network to capture long-term dependencies in the input sequence and to use that information to make predictions.

The training of an RNN involves adjusting the weights of the recurrent connections and the output connections to minimize a loss function. The loss function measures the difference between the predicted output of the network and the true output, and the weights are adjusted using backpropagation through time and stochastic gradient descent.

RNNs have achieved impressive results in a variety of tasks, including natural language processing, where they have been used for language modelling, machine translation, and text generation. They have also been used for speech recognition, where they can recognize spoken words and transcribe them into text, and time series analysis, where they can predict future values based on historical data.

Despite their success, RNNs have some limitations, including a tendency to suffer from the vanishing gradient problem, where the gradients of the loss function with respect to the weights become very small, making it difficult to learn long-term dependencies. However, there are several variants of RNNs, such as long short-term memory (LSTM) networks and gated recurrent units (GRUs), that have been developed to address these limitations and improve the performance of RNNs in a variety of tasks.

3. **Generative Adversarial Networks (GANs):**

GANs are a type of unsupervised learning algorithm that can generate new data that is similar to a given dataset. They consist of two neural networks that are trained together: a generator network that creates new data, and a discriminator network that tries to distinguish between the generated data and the real data.

Generative adversarial networks (GANs) are a type of Deep Learning algorithm that are used for generating new data samples that are similar to a given dataset. GANs are composed of two neural networks, a generator and a discriminator, that are trained together in a game-like framework.

The generator network takes random noise as input and produces a synthetic data sample, such as an image or a sound. The discriminator network takes a data sample, either real or synthetic, and predicts whether it is real or synthetic. During training, the generator and the discriminator are trained together in a adversarial setting where the generator tries to produce synthetic data samples that are similar to the real data samples and the discriminator tries to distinguish between the real and synthetic data samples.

The training of a GAN involves adjusting the weights of the generator and the discriminator networks to minimize a loss function that captures the adversarial nature of the training. The loss function includes both a generator loss, which measures how well the generator is able to produce synthetic data samples that fool the discriminator, and a discriminator loss, which measures how well the discriminator is able to distinguish between the real and synthetic data samples.

GANs have achieved impressive results in a variety of tasks, including image generation, where they can produce realistic images of faces, animals, and scenes that are difficult to distinguish from real images. They have also been used for video generation, where they can generate new video sequences that are similar to a given dataset, and text generation, where they can generate new text samples that are similar to a given corpus of text.

Despite their success, GANs have some limitations, including a tendency to produce mode collapse, where the generator produces a limited set of data samples that do not cover the full range of the input data distribution, and a requirement for large amounts of training data. However, ongoing research is addressing these limitations and improving the performance of GANs in a variety of tasks.

4. **Autoencoders:**

Autoencoders are a type of unsupervised learning algorithm that are used for dimensionality reduction and feature extraction. They consist of an encoder network that compresses the input data into a lower-dimensional representation, and a decoder network that reconstructs the input data from the compressed representation.

Autoencoders are a type of neural network used for unsupervised learning that are designed to learn a compressed representation of input data. The goal of an autoencoder is to learn a low-dimensional representation of the input data that captures the most important features or patterns in the data. Autoencoders can be used for tasks such as data compression, denoising, and image generation.

The architecture of an autoencoder typically includes an encoder network and a decoder network. The encoder network takes the input data and compresses it into a low-dimensional representation, while the decoder network takes the low-dimensional representation and reconstructs the original input data.

An autoencoder involves adjusting the weights of the encoder and decoder networks to minimize a reconstruction loss, which measures the difference between the input data and the reconstructed output. The loss function encourages the encoder to learn a compressed representation of the input data that can be used to reconstruct the original input data with minimal error.

Autoencoders have a variety of applications in computer vision, natural language processing, and signal processing. For example, they can be used for image compression, where they learn a compressed representation of an image that can be stored using less memory than the original image. They can also be used for denoising, where they learn to remove noise from input data, and for anomaly detection, where they learn to detect unusual or anomalous input data.

In addition, autoencoders can be used for image generation, where they can generate new images by sampling from the low-dimensional representation learned by the encoder. Variational autoencoders (VAEs) are a type of autoencoder that have been specifically designed for image generation, and they have been used to generate realistic images of faces, animals, and scenes.

5. **Deep reinforcement learning:**

Deep reinforcement learning is a combination of Deep Learning and reinforcement learning, which is a type of Machine Learning algorithm that learns through trial and error. It is commonly used in robotics and game-playing applications, where an agent learns to make decisions based on rewards or penalties received from the environment.

Deep reinforcement learning (DRL) is a type of Machine Learning algorithm that combines deep neural networks with reinforcement learning techniques. Reinforcement learning is a type of Machine Learning where an agent learns to take actions in an environment to maximize a reward signal.

In DRL, a deep neural network is used to approximate the value function or the policy function of the reinforcement learning algorithm. The value function estimates the expected total reward that an agent will receive over time from a given state, while the policy function maps states to actions.

The combination of deep neural networks with reinforcement learning allows DRL algorithms to learn complex policies in high-dimensional state and action spaces. This has enabled DRL algorithms to achieve impressive results in a variety of tasks, including game playing, robotics, and natural language processing.

One of the most famous examples of DRL is AlphaGo, a program developed by DeepMind that was able to defeat the world champion in the game of Go. AlphaGo used a combination of deep neural networks and reinforcement learning to learn a policy function that was able to predict the next move in a game of Go.

DRL has also been used for robotics, where agents are trained to perform tasks such as grasping objects, navigating through environments, and manipulating objects. In natural language processing, DRL has been used for tasks such as dialogue generation and machine translation.

Despite its successes, DRL has some limitations, including the need for large amounts of training data, the difficulty of tuning hyper parameters, and the risk of over fitting. However, ongoing research is addressing these limitations and improving the performance of DRL algorithms in a variety of tasks.

6. **BERT:**

BERT (Bidirectional Encoder Representations from Transformers) is a Deep Learning model for natural language processing (NLP) developed by Google. It is a pre-trained language model that can be fine-tuned for a variety of NLP tasks, such as question answering, sentiment analysis, and language translation.

BERT is based on the Transformer architecture, which is a neural network architecture that uses self-attention mechanisms to process input data. Unlike traditional language models that process data sequentially, BERT processes the data in a bidirectional manner, taking into account the context on both sides of each word or token in the input.

The pre-training process for BERT involves training the model on a large corpus of text data, such as Wikipedia articles or news articles. During this process, the model learns to predict missing words or tokens in the input text, based on the context provided by the surrounding words or tokens.

After pre-training, the BERT model can be fine-tuned for specific NLP tasks by adding a task-specific output layer and training the model on task-specific data. Fine-tuning allows the model to adapt to the specific language patterns and vocabulary of the task, which can improve performance compared to training a model from scratch.

BERT has achieved state-of-the-art results on several NLP benchmarks, including the GLUE benchmark and the Stanford Question Answering Dataset (SQuAD). Its success has led to the development of several variations, including RoBERTa, ALBERT, and ELECTRA.

7. **LSTM:**

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) that is designed to handle the vanishing gradient problem in traditional RNNs. It was first proposed by Hochreiter and Schmid Huber in 1997 and has since become a popular model in natural language processing (NLP) and other sequence-to-sequence learning tasks.

LSTM networks use a memory cell to store information about previous inputs and outputs, allowing them to learn long-term dependencies in sequential data. The memory cell is controlled by three gates: the input gate, the forget gate, and the output gate.

The input gate controls whether to add new information to the memory cell, based on the current input and the previous output. The forget gate controls whether to remove information from the memory cell, based on the current input and the previous output. The output gate controls whether to output information from the memory cell, based on the current input and the previous output.

The gates in an LSTM network are trained to learn which information to store and which information to discard, based on the task at hand. This allows the network to selectively remember or forget information over long sequences, making it particularly useful for NLP tasks such as sentiment analysis and language translation.

LSTM networks have been shown to perform well on a wide range of NLP tasks, including language modelling, named entity recognition, and machine translation. They are also widely used in speech recognition, time series analysis, and other sequence-to-sequence learning tasks.

8. **Bi-LSTM:**

Bi-LSTM (Bidirectional Long Short-Term Memory) is an extension of the LSTM architecture that processes input sequences in both forward and backward directions. Unlike traditional LSTMs, which can only access past inputs in the sequence, Bi-LSTMs can access both past and future inputs, allowing them to capture more context and make more accurate predictions.

The forward and backward LSTMs in a Bi-LSTM network operate independently, but their outputs are combined at each time step to produce a single output. This combination can be a simple concatenation of the two outputs or a more complex operation such as a weighted sum.

Bi-LSTMs are particularly useful for tasks that require a deep understanding of the input sequence, such as natural language processing (NLP) and speech recognition. In NLP, for example, Bi-LSTMs can be used for tasks such as named entity recognition, sentiment analysis, and machine translation.

One of the advantages of Bi-LSTMs is that they are less prone to the vanishing gradient problem than traditional LSTMs, as the backward LSTM can capture long-term dependencies that may be missed by the forward LSTM. This makes Bi-LSTMs more effective at handling long input sequences and capturing complex relationships between the inputs.

Overall, Bi-LSTMs have shown to be effective models in a variety of sequence-to-sequence learning tasks, and they have become a popular choice for researchers and practitioners in the field of Deep Learning.

4.3: Advantages of Deep Learning:

1. **Deep Neural Networks:** Deep Learning is based on the use of deep neural networks, which are neural networks with many hidden layers. The use of deep neural networks allows for the automatic learning of complex features and patterns in data.

2. **Representation Learning:** Deep Learning models are designed to automatically learn representations of data that are useful for a given task. These learned representations can capture complex and abstract features of the data that are difficult to design by hand.

3. **End-to-End Learning:** Deep Learning models are often designed to learn directly from raw data, without the need for hand-crafted features or preprocessing. This allows for end-to-end learning, where the entire system is trained on a task, rather than relying on separate components.

4. **Scalability:** Deep Learning models can scale to handle large and complex datasets, as well as large and complex models. This allows for the modelling of complex tasks, such as image and speech recognition, that would be difficult to accomplish with traditional Machine Learning methods.

5. **Transfer Learning:** Deep Learning models can be pretrained on large datasets and then fine-tuned on smaller datasets for a specific task. This allows for the transfer of knowledge from one task to another, which can improve performance and reduce the amount of training data required.

6. **Regularization:** Deep Learning models often use regularization techniques, such as dropout and weight decay, to prevent overfitting and improve generalization performance.

7. **Nonlinearities:** Deep Learning models use nonlinear activation functions, such as the rectified linear unit (ReLU), to allow for the modelling of complex and nonlinear relationships in data.

Note that these are just some of the important features of Deep Learning. The field is constantly evolving and new techniques and features are being developed. The choice of technique and feature depends on the specific problem and the available data.

Chapter 5

Depression Detection & Suicidal Ideation

5.1: Overview of Depression Detection & Suicidal Ideation:

Depression and suicide are two of the most critical mental health problems worldwide. It is estimated that depression affects over 264 million people worldwide, while suicide is responsible for over 800,000 deaths annually. Suicide ideation is one of the most critical symptoms of depression and is often difficult to detect. The use of Deep Learning techniques can help to identify and predict depression and suicide ideation in individuals.

Suicidal ideation refers to thoughts or contemplation about taking one's own life. It is a serious and concerning issue that is often associated with mental health conditions such as depression, anxiety, bipolar disorder, or other psychiatric disorders. Suicidal ideation can range from fleeting thoughts about death to persistent and intense desires to end one's life. Individuals experiencing suicidal ideation may feel overwhelmed by emotional pain, hopelessness, and a belief that ending their life is the only way to escape their suffering. They may exhibit signs of withdrawal, social isolation, changes in mood or behaviour, giving away possessions, or expressing feelings of worthlessness or being a burden to others.

It is essential to take suicidal ideation seriously and provide support to individuals experiencing these thoughts. If you or someone you know is struggling with suicidal ideation, it is crucial to seek help immediately from a mental health professional, helplines, or emergency services in your country. Many resources and organizations are available to provide assistance, counselling, and interventions to prevent suicide and support mental well-being.

In this chapter, the role of Machine Learning and Deep Learning in depression detection and suicide ideation has been discussed. The details about the challenges involved in developing such a model's architecture (Fig 5.1) and some of the solutions that have been proposed to address them are highlighted here.

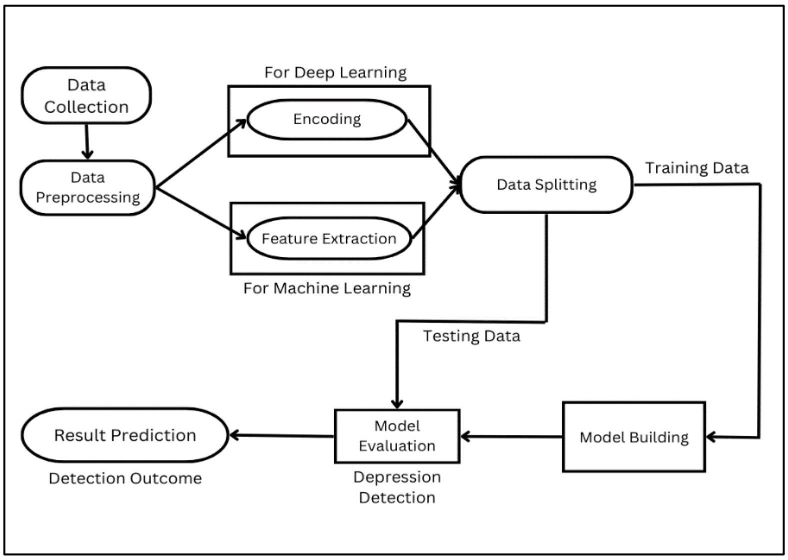

Fig 5.1: Architecture of Depression Detection

5.2: Phases of Depression Detection using ML & DL:

1. **Data Collection:** The data collection phase for depression detection involves gathering data from various sources to build a dataset. This can include surveys, questionnaires, medical records, online forums, social media platforms, mental health apps, and wearable devices. The data is collected with the aim of identifying patterns and indicators of depression. It is important to obtain proper consent and permissions, ensure data privacy and confidentiality, and comply with ethical guidelines. The collected data is then pre-processed, cleaned, and annotated with depression labels if necessary. The dataset is validated for quality and integrity before being used for model development and training.
2. **Data Preprocessing:** Data preprocessing is a crucial phase in depression detection that involves transforming raw data into a suitable format for analysis and modelling. It encompasses various steps to prepare the data for further processing. Initially, the data is cleaned by removing irrelevant or redundant information, handling missing values, and resolving inconsistencies. If the data includes textual content, text preprocessing techniques like lowercase conversion, tokenization, stop word removal, and stemming or lemmatization are applied. By performing these preprocessing steps, the data is transformed into a clean, consistent, and properly formatted representation, laying the groundwork for accurate and reliable depression detection models.

3. **Encoding:** Data encoding plays a crucial role in depression detection, as it involves transforming raw data or textual information into a format suitable for analysis by machine learning algorithms. The process encompasses several steps to convert the data into numerical or categorical representations that capture the essential information for identifying patterns related to depression. Initially, text preprocessing is performed, where the text is cleaned by removing unwanted characters, punctuation, and special symbols, and transformed to lowercase. Following that, tokenization breaks down the text into smaller units, such as words or n-grams, facilitating structured representation. Additionally, if the data contains categorical variables, encoding methods like one-hot encoding or label encoding are employed. By performing data encoding, the raw data is transformed into a suitable format that enables the machine learning models to detect patterns and identify indicators of depression effectively.

4. **Feature Extraction:** In the feature extraction phase for depression detection using text data, the aim is to extract meaningful and informative features from textual content that can help distinguish between depressive and non-depressive patterns. Various techniques are employed to capture relevant linguistic characteristics, sentiment, and semantic information from the text. These features can include word frequencies, lexical and syntactic patterns, sentiment scores, topic modelling, and linguistic features like part-of-speech tags. Additionally, advanced natural language processing techniques such as word embeddings, semantic similarity measures, and discourse analysis may also be utilized to capture deeper linguistic and contextual information. The extracted features from the text data serve as valuable inputs for subsequent modelling and classification algorithms, enabling the development of accurate and effective depression detection models. The choice of features and their representation play a crucial role in capturing the distinct patterns and cues related to depression, ultimately enhancing the performance and accuracy of the detection system.

5. **Data Splitting:** In the data splitting phase for depression detection using text data, the dataset is divided into training, validation, and testing sets to develop and evaluate the depression detection model. To ensure an unbiased evaluation, the widely used technique of k-fold cross-validation is employed. In k-fold cross-validation, the dataset is partitioned into k subsets or folds. The model is then trained and evaluated k times, with each fold serving as the test set once while the remaining folds are used for training. This allows for a comprehensive assessment of the model's performance and generalization capabilities. Throughout the process, the text data undergoes preprocessing steps such as tokenization and normalization, followed by feature extraction to convert it into a suitable format for the model. Stratified splitting is employed to maintain a balanced representation of depressive and non-depressive samples across the folds. By leveraging k-fold cross-validation, we obtain reliable performance estimates and gain insights into the model's robustness, enabling us to identify and address any overfitting or underfitting issues. Ultimately, this approach provides a more accurate assessment of the model's effectiveness in detecting depression using text data.

   K-fold cross-validation is a widely used technique in machine learning and model evaluation. It is a robust method for assessing the performance of a predictive model on a dataset. In k-fold cross-validation, the dataset is divided into k equal-sized subsets or folds. The model is trained and evaluated k times, where each fold is used as the test set once while the remaining folds are used for training. This process helps to address the issue of overfitting by providing multiple independent evaluations of the model's performance. It ensures that the model is tested on different subsets of the data, allowing for a more reliable estimation of its generalization ability. The results from each fold are then averaged to obtain an overall performance metric, such as accuracy or mean squared error. K-fold cross-validation helps to assess the model's stability and robustness, as well as to identify potential issues such as high variance or bias. By providing a comprehensive evaluation, k-fold cross-validation enables researchers and practitioners to make informed decisions about model selection, hyperparameter tuning, and generalization to unseen data.

6. **Model Building:** The model building phase for depression detection using text data involves selecting a suitable machine learning algorithm, such as logistic regression, support vector machines (SVM), random forests, or neural networks, and training it on labelled text data. Additionally, techniques like k-fold cross-validation can be employed to evaluate the model's performance and ensure its generalization ability. Hyperparameter tuning is performed to optimize the model's configuration, and performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess the model's effectiveness in detecting depression based on text data. This iterative process aims to develop a reliable and accurate model for depression detection.

7. **Model Evaluation:** The model evaluation phase for depression detection using text data involves testing the trained models on unseen data to assess their performance. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to measure the model's effectiveness in correctly classifying instances as depressive or non-depressive. Additional metrics like AUC-ROC and AUC-PR provide insights into the model's discrimination ability across different classification thresholds. Cross-validation techniques are employed for robust evaluation, and analysing model limitations and potential sources of errors helps

inform improvements. This phase plays a vital role in assessing model performance and reliability, guiding future refinements for more accurate depression detection using text data.

8. **Result Prediction:** In the result prediction phase for depression detection using text data, the focus is on utilizing the extracted features to make predictions about an individual's depressive state. This phase involves training a machine learning or statistical model using the extracted features and a labelled dataset of individuals with known depression statuses. The model learns the underlying patterns and relationships between the features and the corresponding depression labels. The result prediction phase aims to provide actionable insights about an individual's depressive state based on their text data. The predicted results can be used to identify individuals at risk of depression, prioritize interventions, or provide personalized support and treatment recommendations. It is important to evaluate the performance and accuracy of the prediction model using appropriate evaluation metrics to ensure its reliability and generalizability to real-world scenarios.

Chapter 6

Sentiment Analysis and VADER

6.1: Sentiment Analysis:

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that involves identifying and extracting subjective information from textual data to determine the sentiment or emotion associated with it. It is widely used in various applications, including social media monitoring, customer feedback analysis, market research, and brand reputation management. [65]

6.2: Some Commonly Used Lexicons for Sentiment Analysis:

1. **The Affective Norms for English Words (ANEW) lexicon** is a widely used resource in the field of sentiment analysis and emotion research. It is a collection of emotional ratings for a large number of English words, providing information about the valence (positive or negative), arousal (level of activation), and dominance (level of control) associated with each word.

   The ANEW lexicon was developed by psychologists Bradley and Lang in 1999, and it has been used in various studies to investigate the emotional content of texts and to analyse the sentiment expressed in different contexts.

   The lexicon contains over 1,000 words, and each word is rated by a large number of participants on a scale from 1 to 9 for valence, arousal, and dominance. Valence represents the positivity or negativity of the emotional response, arousal reflects the level of activation or excitement, and dominance refers to the perceived level of control or influence.

   Researchers and practitioners often use the ANEW lexicon as a resource for sentiment analysis and emotion detection tasks. By associating words in a text with their corresponding emotional ratings from the ANEW lexicon, sentiment analysis models can estimate the overall sentiment expressed in the text based on the emotional valence of the words present.

   The ANEW lexicon has been referenced and utilized in numerous studies and publications related to sentiment analysis and emotion research. [66]

2. **SentiWordNet** is another lexicon often used in sentiment analysis and opinion mining tasks. It is a lexical resource that assigns sentiment scores to synsets (sets of synonymous words) in WordNet, a widely used lexical database.

   SentiWordNet provides a numerical sentiment score for each synset, indicating the degree of positivity, negativity, or neutrality associated with the words in that synset. The sentiment scores range from -1 to 1, where negative values represent negativity, positive values represent positivity, and zero represents neutrality.

The sentiment scores in SentiWordNet are derived by combining information from WordNet, which provides synsets and their definitions, and WordNet-Affect, which maps synsets to emotion categories. By leveraging this information, SentiWordNet allows sentiment analysis algorithms to associate sentiment scores with individual words or synsets in a text.

SentiWordNet has been widely used in various sentiment analysis applications, including sentiment classification, opinion mining, and emotion detection. It provides a valuable resource for quantifying the sentiment expressed in text documents, enabling the development of sentiment analysis models and algorithms.

To access SentiWordNet, you can visit the official website of the project, which provides the lexicon along with its documentation and guidelines for usage. It is worth noting that SentiWordNet is based on WordNet 3.0, so newer versions of WordNet may not be incorporated into SentiWordNet. [67]

3. **SenticNet** is another lexical resource used in sentiment analysis and opinion mining. It is a semantic network that associates concepts with sentiment polarity and other affective information. Unlike SentiWordNet, which focuses on synsets and WordNet, SenticNet provides a broader coverage of affective information at the concept level.

   In SenticNet, each concept is represented by a set of semantic and affective features, including polarity, pleasantness, attention, sensitivity, and aptitude. These features capture different aspects of affective and emotional content associated with the concepts.

   The sentiment polarity in SenticNet represents the degree of positivity or negativity of a concept and ranges from -1 to 1, where negative values indicate negativity, positive values indicate positivity, and zero represents neutrality. The other affective features in SenticNet provide additional information related to emotions, attention, and other affective dimensions associated with the concepts.

   SenticNet is constructed using a combination of techniques, including semantic analysis, sentiment analysis, and machine learning. It leverages resources such as WordNet, ConceptNet, and WordNet-Affect to build a comprehensive semantic network with affective information. [68]

4. **The VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon** is a widely used sentiment analysis tool that focuses on analyzing the sentiment of social media texts. It is a pre-built lexicon that contains a large number of words and their associated sentiment scores.

   VADER provides sentiment scores for individual words, as well as for phrases and idioms, allowing for more accurate sentiment analysis in informal and colloquial language often found in social media posts, tweets, and online reviews. The sentiment scores in VADER range from -1 to 1, where negative values indicate negative sentiment, positive values indicate positive sentiment, and zero represents neutral sentiment.

   One notable aspect of VADER is its consideration of both polarity and intensity in sentiment analysis. It not only detects whether a word expresses positive or negative sentiment but also takes into account the intensity of that sentiment. This feature helps capture nuances and contextual meanings in sentiment analysis.

VADER also incorporates rules and grammatical patterns to improve sentiment analysis accuracy. For example, it can interpret the use of capital letters, punctuation, and degree modifiers to adjust sentiment scores accordingly.

VADER is widely used in various sentiment analysis applications, especially for social media data analysis, due to its effectiveness in handling informal language and short texts. It is available as part of the Natural Language Toolkit (NLTK), a popular Python library for natural language processing. [69]

6.3: Some popular Sentiment Analysis methods:

1. **Fine-Grained:** This sentiment analysis model helps you derive polarity precision. You can conduct a sentiment analysis across the following polarity categories: very positive, positive, neutral, negative, or very negative. Fine-grained sentiment analysis is helpful for the study of reviews and ratings. For a rating scale from 1 to 5, you can consider 1 as very negative and five as very positive. For a scale from 1 to 10, you can consider 1-2 as very negative and 9-10 as very positive.

2. **Aspect-Based**: While fine-grained analysis helps you determine the overall polarity of your customer reviews, aspect-based analysis delves deeper. It helps you determine the particular aspects people are talking about. Let's say; you're a mobile phone manufacturer, and you get a customer review stating, "the camera struggles in artificial lighting conditions." With aspect-based analysis, you can determine that the reviewer has commented on something "negative" about the "camera."

3. **Emotion Detection:** As the name suggests, emotion detection helps you detect emotions. This can include anger, sadness, happiness, frustration, fear, worry, panic, etc. Emotion detection systems typically use lexicons – a collection of words that convey certain emotions. Some advanced classifiers also utilize robust machine learning (ML) algorithms. It's recommended to use ML over lexicons because people express emotions in a myriad of ways. Take this line, for example: "This product is about to kill me." This line may express feelings of fear and panic. A similar line – this product is killing it for me – has an entirely different and positive meaning. But the word "kill" might be associated with fear or panic in the lexicon. This may lead to inaccurate emotion detection.

4. **Intent Analysis:** Accurately determining consumer intent can save companies time, money, and effort. So many times, businesses end up chasing consumers that don't plan to buy anytime soon. Accurate intent analysis can resolve this hurdle. The intent analysis helps you identify the intent of the consumer – whether the customer intends to purchase or is just browsing around. If the customer is willing to purchase, you can track them and target them with advertisements. If a consumer isn't ready to buy, you can save your time and resources by not advertising to them. [70]

Chapter 7

Proposed Methodology

The proposed methodology of this research work is divided into two categories. One of this is for Suicidal Ideation using Machine Learning and Deep Learning classifiers and other one is by using hybrid model.

7.1: Proposed System 1:

7.1.1: Architectural framework of the Proposed System:

In the given diagram the architecture of the Proposed System 1 is presented.

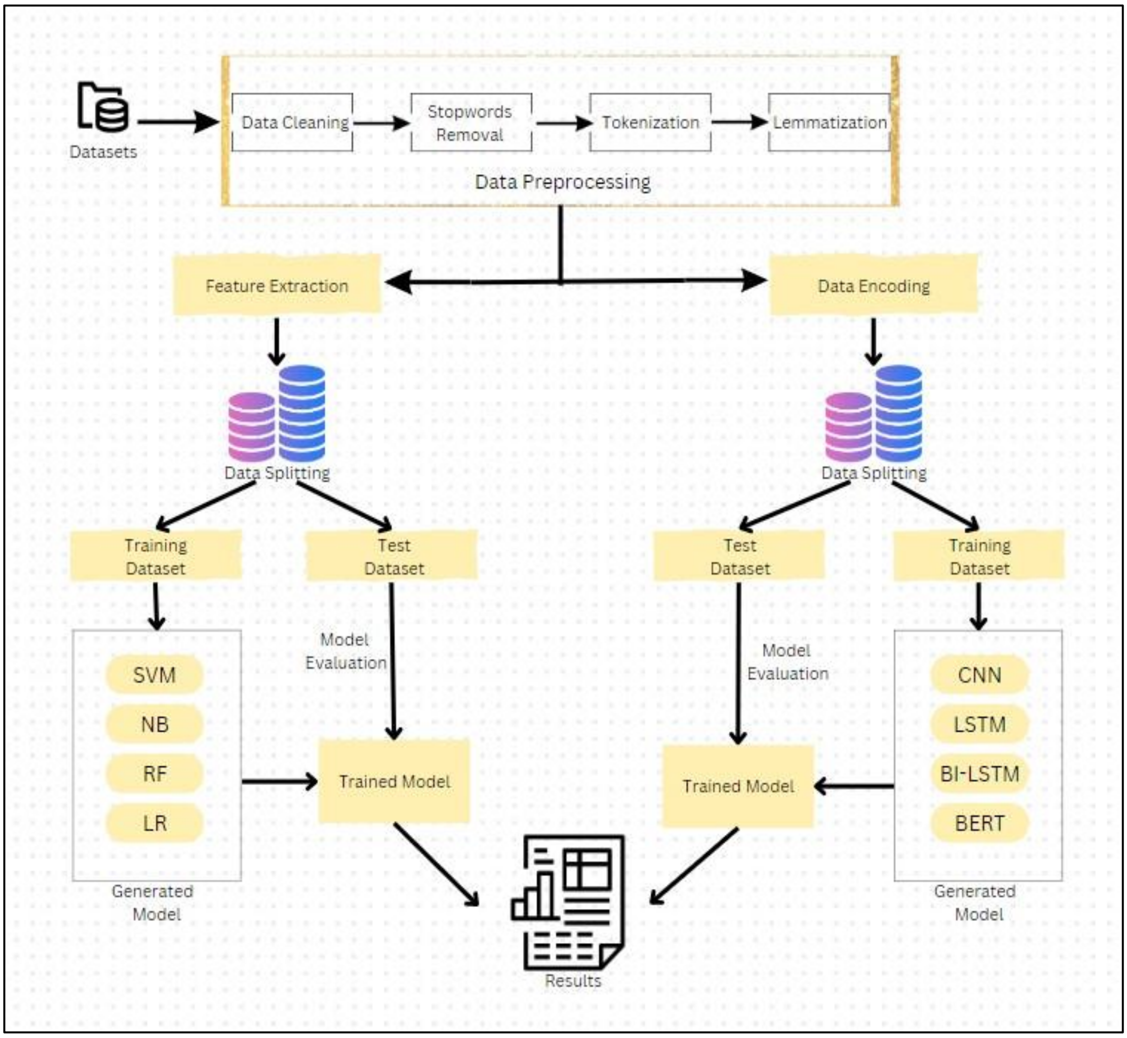

Fig 7.1: Proposed System 1 architecture using ML & DL approach [71]

In Fig 7.1, the architectural framework of the proposed suicidal ideation system has been presented which incorporates four ML classifiers viz. SVM, NB, RF, LR and four DL classifiers viz. CNN, LSTM, Bi-LSTM, BERT. To conduct experimentation using ML approach, the feature extracted include TF-IDF (unigrams & bigrams) and BoW (unigrams & bigrams). To conduct experimentation using DL approach, one-hot encoding technique is used prior to building the classification model. Further details about the text classification approach have been discussed in the subsequent sections.

7.1.2: Flowchart of ML part of Proposed System 1:

In the given diagram (Fig 7.2) the flowchart of the ML part of Proposed System 1 is presented.

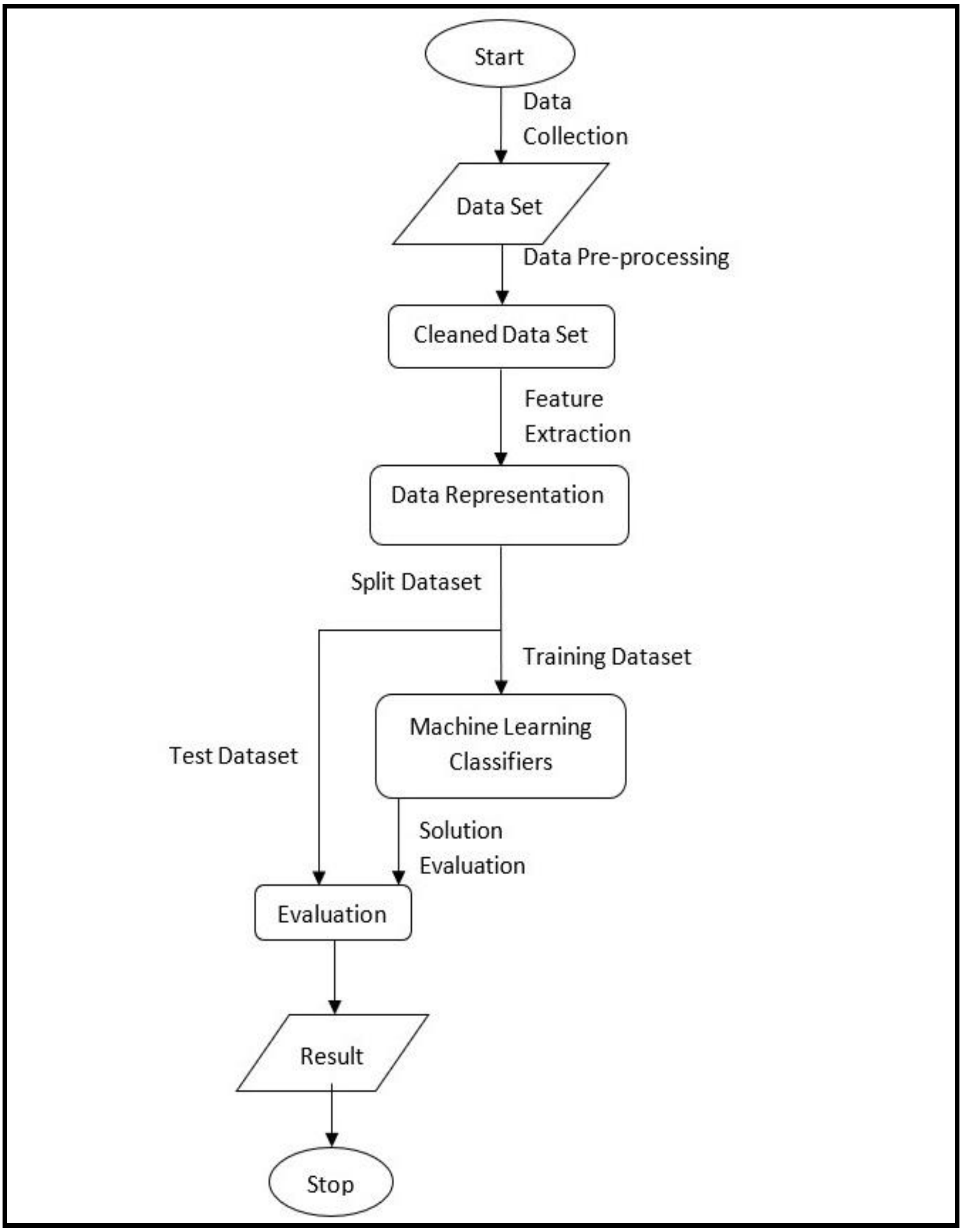

Fig 7.2: Flowchart of ML part of Proposed System 1

7.1.3: Flowchart of DL part of Proposed System 1:

In the given diagram (Fig 7.3) the flowchart of the DL part of Proposed System 1 is presented.

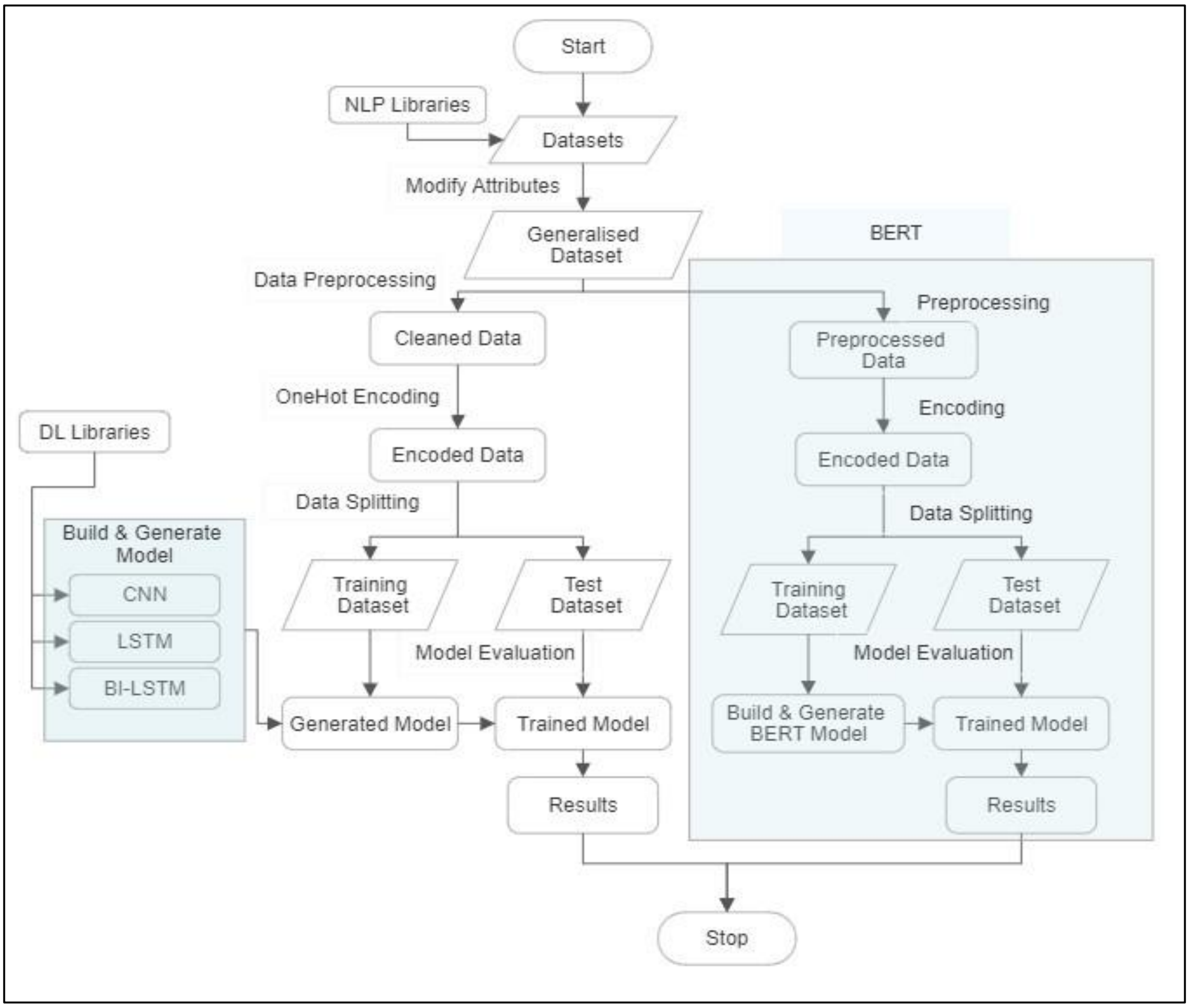

Fig 7.3: Flowchart of DL part of Proposed System 1

7.1.4: Workflow of Proposed System1:

The working of our Proposed System1 is explained via different phases which are discussed as follows-

Step 1. Data Preprocessing: Collect relevant data related to suicidal ideation, such as text or social media posts, interviews, or medical records. For this work, six datasets are collected from various sources. Next clean the data by removing irrelevant information, special characters, and potentially sensitive information. Normalize the text data by converting it to lowercase, removing stop words, and applying lemmatization techniques.

Step 2. Feature Extraction: Extract meaningful features from the preprocessed texts. Some common approaches include Bag-of-Words (BoW); which represents the text as a frequency count of individual words or n-grams, TF-IDF, which weighs the importance of words based on their frequency in a document and the entire corpus, Sentiment Analysis; which analyzes the emotional tone of the text using sentiment lexicons or ML models.

Step 3. Encoding: Convert the extracted features into numerical representations that Deep Learning algorithms can process. If needed, perform additional encoding steps like one-hot encoding or label encoding for categorical variables.

Step 4. Data Splitting: n this part the dataset was splitted into train and test sub-units. The train unit is for training the classifiers in model and the test unit is for testing purposes for the results.

(a) In this project the datasets were splitted in 80:20 ratio for training and testing accordingly.

(b)To obtain a normalized result, k-fold cross-validation is often employed.

Step 5. Model Selection and Training: Apply various Machine Learning models on the training data. The models such as SVM, LR, RF and NB are traditional ML algorithms,while CNN, LSTM, Bi-LSTM, and BERT are DL models commonly used for text classification. Train each model using the training data and evaluate their performance using the validation set. Tune hyperparameters for each model to optimize their performance.

Step 6. Evaluation: Assess the performance of each model based on evaluation metrics such as accuracy, precision, recall, and F1-score. Select the model with the best performance on the validation set.

Step 7. Testing and Deployment: Assess the final model's performance on the testing set to obtain a more accurate estimate of its generalization ability.

7.2: Proposed System 2:

7.2.1: Architecture of Proposed System 2:

In Fig 7.4, the architectural framework of the novel hybrid suicidal ideation system has been presented which incorporates LSTM, and SVM classification approaches and has been trained using sentiment scores generated via VADER lexicon.

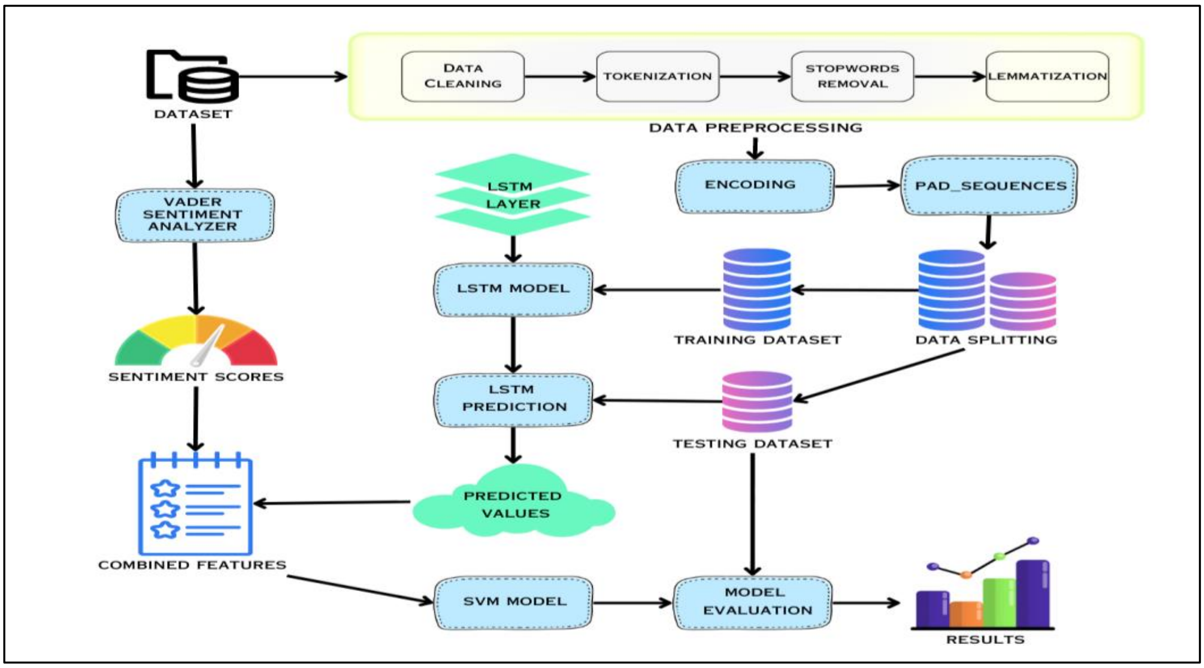

Fig 7.4: Architecture of Proposed System 2

7.2.2: Algorithm of Proposed System 2:

In the given steps the proposed algorithm of the Proposed System 2 is presented.

Step 1: Import the required libraries: “pandas”, “numpy”, “matplotlib.pyplot”, “seaborn”, “nltk”, “tensorflow” and others.

Step 2: Mount the Google Drive to access the dataset.

Step 3: Read the dataset (CSV file) using “pd.read_csv()” and store it in a DataFrame (‘df’).

Step 4: Define a mapping for the class labels.

Step 5: Preprocess the text data:
- (a) Download the stopwords using “nltk.download('stopwords')”.
- (b) Initialize a lemmatization using “WordNetLemmatizer()”.
- (c) Iterate over the text data in the DataFrame and perform the following steps:
- (d) Remove non-alphabetic characters using regex.
- (e) Convert the text to lowercase.
- (f) Split the text into words.
- (g) Remove stopwords and perform lemmatization on each word.
- (h) Join the processed words to form a sentence and append it to the “corpus” list.

Step 6: Perform one-hot encoding on the processed corpus using “one_hot()” and store the encoded representations in the “onehot_repr” list.

Step 7: Pad the one-hot encoded sequences to ensure they have the same length using “pad_sequences()”.

Step 8: Download the VADER sentiment intensity analyzer using “nltk.download('vader_lexicon')”.

Step 9: Create a deep learning model using TensorFlow and Keras:
- (a) Define the model architecture with an Embedding, Dense and LSTM layer.
- (b) Compile the model with binary cross-entropy loss and the Adam optimizer.

Step 10: Split the data into training and testing sets using “KFold” from “sklearn.model_selection”.

Step 11: Perform k-fold cross-validation on the training data:
- (a) Fit the LSTM model on the training data.
- (b) Make predictions using the trained LSTM model.
- (c) Calculate sentiment scores for text in each fold using VADER sentiment analysis.
- (d) Combine the LSTM predictions with sentiment scores.
- (e) Train a LinearSVC classifier on the combined features.
- (f) Calculate the accuracy score for the classifier.
- (g) Append the accuracy score to the “accuracy_scores” list.

Step 12: Calculate the mean accuracy score from the accuracy_scores.

7.2.3: Workflow of Proposed System 2:

The working of our Proposed System 2 is explained via different phases which are discussed as follows-

1. Data Collection: Dataset has a huge role on depression detection or suicidal ideation. There are lots of datasets available in online sources. For the project, firstly the idea is to focus on the online dataset available. In future work the implementation of the manual handwritten texts can also be possible. From the help of the all the analysed research papers the datasets from Kaggle, GitHub, reddit, life corpus etc were collected. Finally, the five datasets were shortlisted for the project work. The datasets were named as– "500 Reddit Post", "SDCNL", "Suicide and Depression Detection", "Life_Corpus" and "SWMH".

2. Reset Index: Various datasets contain various types of indexes. From those the all indexes may not be used. To overcome these problems the reset index part is proposed where the extra indexes of the dataset were removed. After that the different labels were changed in the form of our required type. As the various datasets contain various types of indexes, the reset index part for each dataset is different. The various libraries were used for this part. For example- pandas, NumPy etc. After this process the data is renamed or presented as re-established data.

3. Data Pre-processing: The data pre-processing unit is an addition of four different sub-units. The first part is "Data cleaning", the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. The first part is "Stopwords removal", the process of removing the words that occur commonly across all the documents. Where "nltk.corpus.stopwords('english')" is used. The second part is "Tokenization", the process of replacing sensitive data with unique identification symbols that retain all the essential information about the data without compromising its security. Where the whole dataset is splitted then turned into tokens and finally arranged accordingly. The third and final part of data pre-processing is "Lemmatization'', the process of grouping together different forms of the same word. For example, Lemmatization of the words 'good', 'better' and 'best' would return 'good'. For the process the "WordNetLemmatizer()" is used.

4. Encoding: After data preprocessing unit the preprocessed data goes for encoding. encoding is the process of putting a sequence of characters (letters, numbers, punctuation, and certain symbols) into a specialized format. In this part "onehot encoding" is implemented. For this the "one_hot" is imported. One hot encoding is a technique used to represent categorical variables as numerical values in a Deep Learning model.

5. VADER Lexicon: The VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon is imported from NLTK.

6. Model Building: In this part the backbone of the LSTM classifier is constructed. In this section all layers of the LSTM model are created. In the end of this model the models for each classifier were compiled.

7. Data Splitting: k-fold cross-validation process by splitting the data into training and testing sets for each fold. The training and testing data are then used for further analysis, such as training the model on the training data and evaluating its performance on the testing data.

8. Model Evaluation: In this scenario, an LSTM model is trained using the training data, X_train, and their corresponding labels, Y_train. The model undergoes 10 epochs of training, with a batch size of 64. Subsequently, the trained LSTM model is employed to make predictions on the testing data, X_test, using the predict () method. To gauge the sentiment of the text in the test set, sentiment scores are computed using the polarity_scores() method from the SentimentIntensityAnalyzer (SID) object, referred to as sid. These scores represent the overall sentiment polarity of each text, capturing its positive or negative connotation. To create a comprehensive feature matrix, the LSTM predictions (lstm_predictions) and the sentiment scores (test_sentiment_scores) are combined using np.column_stack() from NumPy, which horizontally stacks the arrays. The resulting combined_features array contains both sets of features. Next, an SVM model is trained using the combined feature matrix (combined_features) and the corresponding labels (Y_test). The SVM model learns to classify instances based on the combined features. Using the trained SVM model, predictions for the labels of the combined feature matrix are generated. Finally, the accuracy of the SVM predictions is calculated by comparing them to the true labels, providing a measure of the model's performance.

Chapter 8

Implementation

8.1: Dataset Description:

Selected classifiers have been trained and tested on various datasets (Fig 8.1). The description of the datasets is provided below:

1. *The Reddit C-SSRS Suicide dataset (Reddit dataset)* is a dataset consisting of 500 Reddit posts. It contains three columns, namely "User," "Post," and "Label." The "User" column contained usernames and was removed since it was unnecessary for our purposes. The "Post" column contains the text messages of the posts. Each post was originally labeled with five different labels: "Ideation," "Supportive," "Attempt," "Behavior," and "Indicator." However, the labels were modified to meet the requirements by replacing "Ideation," "Attempt," "Behavior," and "Indicator" with the value 1, while "Supportive" was replaced with 0. An 80:20 train-test ratio was adopted for the experiments, with 400 posts used for training and 100 posts for testing. Furthermore, 5-fold and 10-fold cross-validation were applied for training and testing.

2. The *SDCNL dataset* comprises 1894 rows and 14 columns, including "Unnamed: 0," "Unnamed: 0.1," "title," "selftext," "author," "num_comments," "is_suicide," "url," "selftext_clean," "title_clean," "author_clean," "selftext_length," "title_length," and "megatext_clean." However, for the project, only the columns "selftext_clean" and "is_suicide" are required. The "selftext_clean" column contains the text messages, while the "is_suicide" column contains corresponding labels represented by 0 and 1. To conduct the experiment, an 80:20 train-test ratio was adopted, and 5-fold and 10-fold cross-validation was employed for training and testing.

3. The *Life_Corpus* dataset is comparatively small, consisting of 273 rows and 2 columns. The first column, labeled as "text," contains the text messages, while the second column, labeled as "cls," contains the corresponding labels. The labels were converted to numeric values, where "No risk" was represented as 0 and "Risk" was represented as 1. The dataset was split into a training set and a testing set with a ratio of 80:20. Additionally, 5-fold and 10-fold cross-validation techniques were employed for training and testing the classifiers.

4. The *Suicide and Depression Detection (SDD) dataset* comprises posts extracted from the "SuicideWatch" and "depression" subreddits of the Reddit platform. It consists of 232,074 rows and 3 columns, namely "Unnamed: 0", "text", and "class". For this experiment, the "Unnamed: 0" column, which is unnecessary, was removed. The "text" column contains the text messages, while the "class" column contains corresponding labels categorized as "non-suicide" and "suicide". The "non-suicide" labels were converted to 0, while the "suicide" labels were converted to 1. To manage the size of the dataset, we selected the last 30,000 rows for our experiment, splitting them into training and testing sets with an 80:20 ratio across all eight selected classifiers. However, when applying 5-fold and 10-fold cross-validation for

training and testing using DL classifiers, only the last 10,000 rows were utilized. On the other hand, when applying 5-fold and 10-fold cross-validation for training and testing using ML classifiers, the entire 30,000 rows were considered.

5. The *CEASE dataset* comprises a cluster of 15 text files containing text messages expressing different emotions. These text files include "abuse_alltext.txt," "anger_alltext.txt," "blame_alltext.txt," "fear_alltext.txt," "forgiveness_alltext.txt," "guilt_alltext.txt," "happiness_peacefulness_alltext.txt," "hopefulness_alltext.txt," "hopelessness_alltext.txt," "information_alltext.txt," "instruction_alltext.txt," "love_alltext.txt," "pride_alltext.txt," "sorrow_alltext.txt," and "thankfulness_alltext.txt." The messages from these files were merged into a single text file named "Total_dataset.txt." Subsequently, sentiment intensity analysis was performed on each message using the VADER lexicon. The depressive messages were labeled as 1, while non-depressive messages were labeled as 0. The dataset consists of 2393 rows and 3 columns: "Unnamed: 0," "text," and "Sentiment." The "text" column contains all the text messages, and the sentiments were assigned in the "Sentiment" column. The "Unnamed: 0" column was removed as it was unnecessary for the experiment. For the experiment, an 80:20 train-test ratio was employed, and both 5-fold and 10-fold cross-validation techniques were applied for training and testing purposes.

6. The *SWMH dataset* was collected from mental health-related subreddits and was split into three parts: "train.csv", "test.csv", and "val.csv". For the experiment, "train.csv" was used for training, while "test.csv" was used for testing. Each .csv file consisted of two columns, namely "text" and "label". The "text" column contained all the messages, and the "label" column contained five different labels: 'self.Anxiety', 'self.bipolar', 'self.depression', 'self.SuicideWatch', and 'self.offmychest'. The "train.csv" had 34,823 rows, and the "test.csv" had 10,883 rows. These two datasets, "train.csv" and "test.csv", were merged to create a single dataset called "merged_dataset.csv". Subsequently, the dataset underwent 5-fold and 10-fold cross-validation for training and testing purposes.

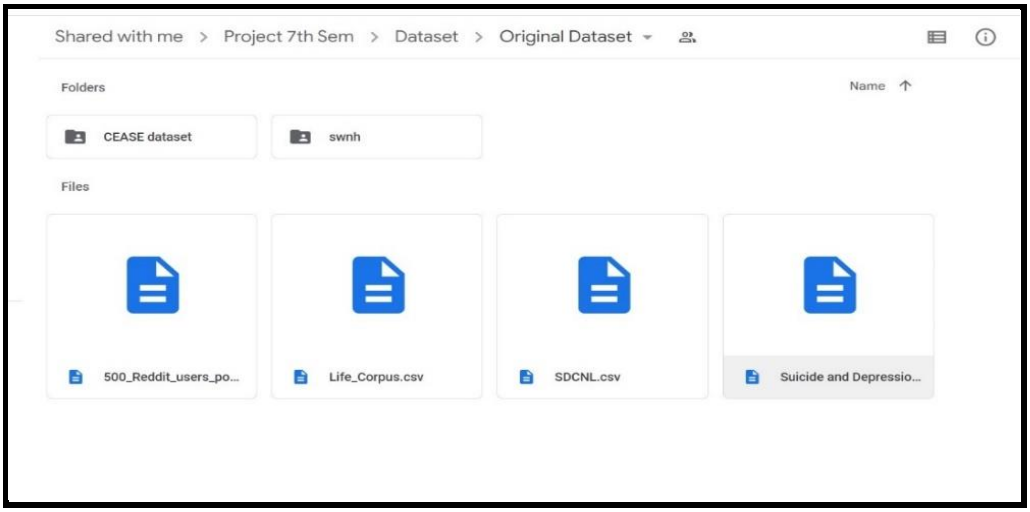

Fig 8.1: Datasets used in this Project

8.2: Feature Extraction:

In the following those features are listed which were used for Identifying Suicidal Intent through Depression Detection from Texts using Machine Learning:

1. *Bag of Word(unigram) / BoW(1):* Bag of words is a technique used in natural language processing to represent a text document as a collection of individual words, without considering the order or structure of the text. A unigram, also known as a "1-gram," is a sequence of one word in a text document. Therefore, a bag of words unigram representation of a text document is a set of all the unique words in the document, without considering their order or frequency. Bag of words unigram representation is a simple but effective way to convert text data into a format that can be used for Machine Learning algorithms. However, it does not capture the order or structure of the text, which can be important in some applications such as language translation or sentiment analysis. And to implement this 'CountVectorizer' from sklearn.feature_extraction.text' and 'ngram_range' as '(1,1)' is used.

2. *Bag of Word(bigram) / BoW(2):* Unlike unigrams, which focus on individual words, bigrams capture contextual information by examining adjacent word sequences. By converting text data into this representation, it becomes suitable for Machine Learning algorithms. However, it's important to note that bag of words (bigram) still does not account for overall grammatical structure or word order. The implementation involves using the 'CountVectorizer' module from 'sklearn.feature_extraction.text' and specifying 'ngram_range' as '(2, 2)' to indicate the use of bigrams

3. *TF-IDF(unigram) / TF-IDF(1):* TF-IDF stands for Term Frequency-Inverse Document Frequency, a commonly used weighting scheme in information retrieval and natural language processing. In a TF-IDF unigram model, each document is represented as a bag of words, and the frequency of each term (or unigram) in the document is calculated. The term frequency (TF) is then normalized by the maximum frequency of any term in the document to avoid bias towards longer documents. The inverse document frequency (IDF) measures the rarity of a term across all documents in the corpus. It is calculated by taking the logarithm of the total number of documents in the corpus divided by the number of documents containing the term. The final TF-IDF weight of a term in a document is the product of its normalized term frequency and its inverse document frequency.

   TF = count of term in a document / number of words in that document

   IDF = $\log_2$(Total number of documents in a corpus / number of documents in which the word appears).

   And to implement this 'TfidfVectorizer' is used from 'sklearn.feature_extraction.text' with 'max_features=10000' and 'ngram_range' as '(1,1)' is used.

4. *TF-IDF(bigram) / TF-IDF(2):* TF-IDF (Bigram) is a technique in natural language processing that represents text documents numerically by combining the concepts of term frequency-inverse document frequency (TF-IDF) and bigrams. TF-IDF measures the importance of a word in a document relative to its frequency in a corpus, while bigrams capture pairs of adjacent words. TF-IDF (Bigram) combines these approaches to assess the significance of bigrams in a document, considering both their frequency in the document and rarity in the

corpus. By implementing the 'TfidfVectorizer' from 'sklearn.feature_extraction.text' with 'max_features=100000' and 'ngram_range' set to '(2,2)', this technique generates a nuanced representation that captures the contextual meaning of words in the document.

After evaluating all the Machine Learning classifiers using these three features one by one, different possible combinations of these three features are also being used to evaluate the classifiers. Those combinations were given below:

- TF-IDF(1) + BoW(1)
- TF-IDF(1) + BoW(2)
- TF-IDF(1) + TF-IDF(2)
- TF-IDF(2) + BoW(1)
- TF-IDF(2) + BoW(2)
- BoW(1) + BoW(2)
- TF-IDF(1) + TF-IDF(2) + BoW(1)
- TF-IDF(1) + TF-IDF(2) + BoW(2)
- TF-IDF(1) + BoW(1) + BoW(2)
- TF-IDF(2) + BoW(1) + BoW(2)
- TF-IDF(1) + TF-IDF(2) + BoW(1) + BoW(2)

8.3: Coding Environment:

Table 8.1: Coding environment and system details

OS Name	Windows 10 Home Single Language
Version	22H2(10.0.19045)
OS Manufacturer	Microsoft Corporation
System Type	x64-based PC
Processor	Intel(R) Core(TM) i3-8130U CPU @ 2.20GHz 2.20 GHz
BIOS Version/Date	F.43/16-11-2022
Baseboard Manufacturer	Intel Corporation
Platform	GOOGLE COLAB
System Directory	C:\WINDOWS\system64
Installed Physical Memory (RAM)	4.00 GB
Total Physical Memory	3.93 GB
Available Physical Memory	618 MB
Total Virtual Memory	12.00 GB
Available Virtual Memory	4.56 GB
Page File Space	5.56 GB
Device name	LAPTOP-S15FQ825

- Python is used as the programming language in the GOOGLE COLAB platform.

Chapter 9

Results

In this chapter, all the results of the proposed system have been illustrated.

Reddit Dataset:

Table 9.1: Accuracy of Reddit Dataset using ML approach

Features	ML Model											
	SVM			NB			RF			LR		
	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold
BoW(1)	69	78	76	73	82	82	77	78	78	69	78	78
TF-IDF(1)	73	80	80	77	78	78	77	78	78	77	79	79
BoW(2)	74	77	78	51	53	53	77	78	78	74	78	78
TF-IDF(2)	77	78	78	77	78	78	77	78	78	77	78	78
TF-IDF(1) + BoW(1)	69	77	77	73	82	81	77	78	78	69	79	78
TF-IDF(1) + BoW(2)	74	77	77	50	54	53	77	78	78	73	78	78
TF-IDF(1) + TF-IDF(2)	75	80	81	77	78	78	77	78	78	77	79	79
TF-IDF(2) + BoW(1)	69	77	77	75	81	81	77	78	78	69	78	78
TF-IDF(2) + BoW(2)	74	77	78	55	59	58	77	78	78	74	78	78
BoW(1) + BoW(2)	74	79	79	62	67	66	77	78	78	73	80	80
TF-IDF(1) + TF-IDF(2) + BoW(1)	68	78	77	76	80	81	77	78	78	70	78	78
TF-IDF(1) + TF-IDF(2) + BoW(2)	74	77	77	57	60	59	77	78	78	73	78	77
TF-IDF(1) + BoW(1) + BoW(2)	74	79	79	61	67	67	77	78	78	73	80	80
TF-IDF(2) + BoW(1) + BoW(2)	74	79	79	62	70	68	77	78	78	73	80	80
TF-IDF(1) + TF-IDF(2) + BoW(1) + BoW(2)	74	80	79	66	70	68	77	78	78	73	81	80

Table 9.1 shows that the "10-fold" and "5-fold" cross-validation approaches consistently yield better results compared to the 80:20 split. Additionally, when considering the different feature types, it is evident that BoW and TF-IDF(1) + BoW(1) provide the highest accuracy for the given dataset in the NB classifier.

SDCNL:

Table 9.2: Accuracy of SDCNL Suicide Dataset using ML approach

Features	**ML Model**											
	SVM			**NB**			**RF**			**LR**		
	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold
BoW(1)	68	66	65	66	67	66	67	63	63	67	68	67
TF-IDF(1)	68	69	68	65	66	67	66	66	64	70	71	69
BoW(2)	56	56	56	59	59	59	59	52	54	59	59	59
TF-IDF(2)	60	59	58	63	60	60	58	52	54	60	56	56
TF-IDF(1) + BoW(1)	69	67	65	66	67	66	68	65	65	68	68	68
TF-IDF(1) + BoW(2)	62	61	61	60	60	60	63	62	62	62	62	62
TF-IDF(1) + TF-IDF(2)	70	70	68	69	68	67	65	62	62	71	70	69
TF-IDF(2) + BoW(1)	66	67	66	68	67	66	62	59	59	68	68	68
TF-IDF(2) + BoW(2)	56	56	56	62	59	59	56	54	54	60	59	59
BoW(1) + BoW(2)	67	66	64	65	65	64	66	59	58	71	69	68
TF-IDF(1) + TF-IDF(2) + BoW(1)	66	67	66	68	66	66	68	63	61	68	68	68
TF-IDF(1) + TF-IDF(2) + BoW(2)	63	61	60	61	60	60	62	61	61	62	62	62
TF-IDF(1) + BoW(1) + BoW(2)	69	66	66	66	65	65	66	62	62	71	69	68
TF-IDF(2) + BoW(1) + BoW(2)	67	66	65	65	64	64	65	58	59	70	68	68
TF-IDF(1) + TF-IDF(2) + BoW(1) + BoW(2)	69	66	66	66	65	65	67	63	62	71	69	68

From table 9.2, it can be observed that the "80:20" cross-validation approaches consistently yield better results compared to the "10-fold" and "5-fold" split. Additionally, when considering the different feature types, it is evident that the combination of TF-IDF(1) and BoW(1) provides higher accuracy for the given dataset in the LR classifier.

SDD Dataset:

Table 9.3: Accuracy of SDD Dataset using ML approach

Features	ML Model											
	SVM			NB			RF			LR		
	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold
BoW(1)	91	90	90	90	90	90	81	80	80	93	92	92
TF-IDF(1)	93	93	93	91	90	90	83	83	83	92	92	92
BoW(2)	90	90	90	91	90	90	79	78	79	91	90	90
TF-IDF(2)	92	91	91	89	89	90	80	80	80	89	90	90
TF-IDF(1) + BoW(1)	92	91	91	91	91	91	82	82	82	93	92	92
TF-IDF(1) + BoW(2)	92	91	91	91	91	91	80	81	80	92	91	92
TF-IDF(1) + TF-IDF(2)	94	93	93	92	91	91	81	81	81	93	93	93
TF-IDF(2) + BoW(1)	92	91	91	91	91	91	81	81	80	93	92	92
TF-IDF(2) + BoW(2)	91	90	90	91	91	91	80	79	79	91	90	90
BoW(1) + BoW(2)	92	92	92	91	91	91	78	79	79	93	93	93
TF-IDF(1) + TF-IDF(2) + BoW(1)	93	92	92	92	92	92	82	81	81	93	92	92
TF-IDF(1) + TF-IDF(2) + BoW(2)	92	91	91	92	91	91	80	80	80	92	91	92
TF-IDF(1) + BoW(1) + BoW(2)	92	92	92	92	91	91	80	80	80	93	93	93
TF-IDF(2) + BoW(1) + BoW(2)	92	92	92	91	91	91	81	80	80	93	93	93
TF-IDF(1) + TF-IDF(2) + BoW(1) + BoW(2)	93	92	92	92	92	92	80	80	80	93	93	93

Table 9.3 shows that the "10-fold" and "5-fold" cross-validation approaches consistently yield better results compared to the 80:20 split. Additionally, when considering the different feature types, it is evident that "TF-IDF(1) + TF-IDF(2)" provides the highest accuracy for the given dataset in the SVM classifier.

Life_Corpus:

Table 9.4: Accuracy of Life_Corpus Suicide Dataset using ML approach

Features	ML Model											
	SVM			NB			RF			LR		
	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold
BoW(1)	76	77	79	75	77	77	75	79	81	76	78	79
TF-IDF(1)	76	80	82	75	72	73	75	80	81	80	74	77
BoW(2)	71	75	75	36	40	42	67	67	67	71	80	80
TF-IDF(2)	67	67	67	65	64	63	67	67	67	67	67	67
TF-IDF(1) + BoW(1)	78	77	78	78	78	80	78	81	82	76	78	79
TF-IDF(1) + BoW(2)	73	75	76	38	43	43	76	79	81	73	81	81
TF-IDF(1) + TF-IDF(2)	78	80	82	76	73	73	73	78	81	76	72	76
TF-IDF(2) + BoW(1)	78	77	78	78	77	78	76	79	79	76	78	79
TF-IDF(2) + BoW(2)	71	75	76	40	50	51	67	67	67	71	81	80
BoW(1) + BoW(2)	80	81	80	60	65	66	80	77	81	78	81	81
TF-IDF(1) + TF-IDF(2) + BoW(1)	78	77	79	78	78	80	75	78	82	76	78	79
TF-IDF(1) + TF-IDF(2) + BoW(2)	73	75	76	47	52	54	80	75	78	73	81	81
TF-IDF(1) + BoW(1) + BoW(2)	80	80	80	67	69	68	76	80	81	78	81	82
TF-IDF(2) + BoW(1) + BoW(2)	80	80	80	69	76	74	76	78	80	78	81	82
TF-IDF(1) + TF-IDF(2) + BoW(1) + BoW(2)	80	80	80	71	76	76	78	81	81	78	81	82

Table 9.4 shows that the "10-fold" and "5-fold" cross-validation approaches consistently yield better results compared to the 80:20 split. Additionally, when considering the different feature types, it is evident that all combinations of TF-IDF(1) provide the highest accuracy for the given dataset in the LR classifier.

CEASE:

Table 9.5: Accuracy of CEASE Suicide Dataset using ML approach

Features	ML Model											
	SVM			NB			RF			LR		
	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold
BoW(1)	76	77	78	73	72	72	63	65	65	76	77	77
TF-IDF(1)	78	77	77	75	73	74	62	65	65	72	73	73
BoW(2)	70	73	72	63	63	63	62	65	65	71	71	71
TF-IDF(2)	72	72	72	67	68	68	62	65	65	65	67	67
TF-IDF(1) + BoW(1)	75	76	77	67	72	72	62	66	66	77	77	77
TF-IDF(1) + BoW(2)	73	75	75	68	67	66	62	65	65	72	73	74
TF-IDF(1) + TF-IDF(2)	79	77	78	73	72	72	67	65	65	77	74	74
TF-IDF(2) + BoW(1)	77	78	79	73	71	72	63	65	65	78	77	77
TF-IDF(2) + BoW(2)	69	72	72	62	61	62	62	65	65	71	71	71
BoW(1) + BoW(2)	76	77	78	70	68	69	62	65	65	76	77	77
TF-IDF(1) + TF-IDF(2) + BoW(1)	78	78	78	72	71	72	63	65	65	77	77	77
TF-IDF(1) + TF-IDF(2) + BoW(2)	73	75	75	67	65	65	62	65	65	73	74	74
TF-IDF(1) + BoW(1) + BoW(2)	77	78	78	71	69	70	63	65	65	77	77	78
TF-IDF(2) + BoW(1) + BoW(2)	76	77	78	69	67	68	62	65	65	77	77	77
TF-IDF(1) + TF-IDF(2) + BoW(1) + BoW(2)	76	78	78	70	68	69	63	65	65	77	77	78

Table 9.5 shows that the "10-fold" and "5-fold" cross-validation approaches consistently yield better results compared to the 80:20 split. Additionally, when considering the different feature types, it is evident that TF-IDF(2) + TF-IDF(1) and TF-IDF(2) + BoW(1) provides the highest accuracy for the given dataset in the SVM classifier.

SWMH:

Table 9.6: Accuracy of SWMH Suicide Dataset using ML approach

Features	ML Model											
	SVM			NB			RF			LR		
	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold
BoW(1)	48	76	76	47	78	78	53	70	69	48	79	79
TF-IDF(1)	48	80	80	47	77	77	52	75	74	50	81	82
BoW(2)	52	74	74	47	77	77	53	56	56	51	77	77
TF-IDF(2)	51	77	78	52	78	78	53	62	62	53	79	79
TF-IDF(1) + BoW(1)	50	76	76	48	78	78	53	73	73	51	79	79
TF-IDF(1) + BoW(2)	52	75	75	47	77	78	53	62	62	51	79	79
TF-IDF(1) + TF-IDF(2)	49	79	80	48	79	79	53	70	70	50	82	82
TF-IDF(2) + BoW(1)	51	76	77	47	78	79	53	68	69	49	79	80
TF-IDF(2) + BoW(2)	50	74	75	47	77	78	53	59	59	50	77	78
BoW(1) + BoW(2)	50	77	77	47	78	78	53	62	62	51	80	80
TF-IDF(1) + TF-IDF(2) + BoW(1)	49	77	77	48	78	78	53	72	72	48	79	80
TF-IDF(1) + TF-IDF(2) + BoW(2)	52	75	76	47	78	78	53	64	63	51	79	79
TF-IDF(1) + BoW(1) + BoW(2)	54	77	77	47	78	78	53	66	66	54	80	80
TF-IDF(2) + BoW(1) + BoW(2)	51	77	77	47	78	78	53	64	63	51	80	80
TF-IDF(1) + TF-IDF(2) + BoW(1) + BoW(2)	51	77	77	47	79	79	53	66	66	51	80	80

Table 9.6 shows that the "10-fold" and "5-fold" cross-validation approaches consistently yield better results compared to the 80:20 split. Additionally, when considering the different feature types, it is evident that TF-IDF(1) and TF-IDF(1) + TF-IDF(2) provide the highest accuracy for the given dataset in the LR classifier.

Table 9.7: Accuracy of selected datasets using DL approach

Dataset	Deep Learning									
	LSTM			Bi-LSTM			CNN			BERT
	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20	5-Fold	10-Fold	80:20
Reddit Dataset	82	91.2	93	82	78.4	78.4	81	78.4	78.4	78
SDCNL	67	93.5	95	62	63.2	61.1	66	59.4	60.7	55
SDD Dataset	88	92.5	92	87	85	86	86	86	88	80
Life_Corpus	85	90	92.8	85	68.2	70	87	67	57.4	85
CEASE	76	92.8	93.2	77	67.2	68	77	64.4	65	68
SWMH	71	87	91	73	73	74	68	70	71	60

Table 9.7 shows that LSTM outperformed other classifiers using 80:20 split, 5-fold and 10-fold cross-validation. While using the SDD Dataset, the last 30,000 rows were taken for an 80:20 split, and the last 10,000 rows were taken for 5-fold and 10-fold cross-validation.

Total Result Table:

Table 9.8: Best Accuracy of selected datasets using ML & DL approach

Datasets	Machine Learning				Deep Learning			
	SVM	NB	RF	LR	BERT	LSTM	Bi-LSTM	CNN
Reddit Dataset	81	82	78	81	78	**93**	82	81
SDCNL	70	69	68	71	55	**95**	63	66
SDD Dataset	**94**	92	83	93	80	92	87	88
Life_Corpus	81	80	81	81	85	**93**	85	87
CEASE	79	75	67	78	68	**93**	77	77
SWMH	80	79	75	82	60	**91**	74	71

It can be observed from Table 9.8 that LSTM outperforms other classifiers in the Reddit, SDCNL, Life_Corpus, CEASE, SWMH datasets with 93%, 95%, 93%, 93% and 91% accuracy value respectively whereas SVM outperforms other classifiers in the SDD dataset with accuracy value of 94% with LSTM just close behind in terms of performance with accuracy value of 92%.

Result of Proposed System 2:

Table 9.9: Accuracy of selected datasets using Proposed System 2

Model	Datasets					
	Reddit Dataset	SDCNL	SDD Dataset	Life_Corpus	CEASE	SWMH
Proposed System 2	97.4	94.4	93.7	97.14	99.2	91.5

The table 9.8 shows the accuracy scores of Proposed System 2 on all six selected datasets

Reddit Dataset: The proposed system achieved a performance score of 97.4 on this dataset, indicating a high level of accuracy in understanding and generating responses based on Reddit data.

SDCNL Dataset: The system achieved a performance score of 94.4 on this dataset. The specifics of this dataset are not clear, but the system's performance suggests it can handle the data with good accuracy.

SDD Dataset: The proposed system achieved a performance score of 93.7 on this dataset. Again, the specifics of this dataset are not mentioned, but the system's performance indicates a reasonably accurate understanding and generation of responses.

Life_Corpus: The system achieved a performance score of 97.14 on the Life_Corpus dataset. This dataset likely contains data related to various aspects of life, and the system demonstrates a high level of accuracy in handling such content.

CEASE: The proposed system achieved a performance score of 99.2 on the CEASE dataset. The nature of this dataset is not described, but the high performance score suggests that the system excels in understanding and generating responses based on this particular data.

SWMH: The system achieved a performance score of 91.5 on the SWMH dataset. The details of this dataset are not provided, but the system's performance indicates a moderate level of accuracy.

Overall, Proposed System 2 performs well across different datasets, with high scores on several datasets like Reddit, Life_Corpus, and CEASE, showcasing its effectiveness in understanding and generating responses in various domains. However, more information about the datasets and the specific evaluation metrics used would provide a clearer understanding of the system's performance.

Comparison Table:

Table 9.10: Comparison of accuracy between existing research work and proposed work

Datasets	Authors	Technique / Features	Reported Result
Reddit Dataset	Gupta et al. [11]	NB, Semantic Features	71
	Proposed System 1	LSTM	82
	Proposed System 2	Hybrid model	97.4
SDCNL	Proposed System 1	LR, TF-IDF(1), TF-IDF(2)	71
	Proposed System 2	Hybrid model	94.4
	Haque et al [36]	guse-dense with UMAP-KMeans	**98.18**
SDD Dataset	Proposed System 2	Hybrid model	93.7
	Proposed System 1	SVM, TF-IDF(1), TF-IDF(2)	94
	Yi et al [47]	Electra	**97**
Life_Corpus	Caicedo et al [8]	Rasa	0.49 ± 0.02 (Macro f1)
	Parraga-Alava et al [20]	Hierarchical clustering. average	0.79 (F1)
	Proposed System 1	CNN	87
	Caicedo et al [16]	POS, SYNSETS, lemma, word RandomCommitte	0.958 (F1)
	Proposed System 2	Hybrid model	**97.14**
CEASE	Ghosh et al [32]	CNN	59.54
	Proposed System 1	CNN	77
	Proposed System 2	Hybrid model	**99.2**
SWMH	Ji et al [46]	RN	64
	Proposed System 1	Bi-LSTM	73
	Proposed System 2	Hybrid model	**91.5**

Table 9.10 showcases that the proposed systems manage to perform better when compared to other existing works for four datasets viz. Reddit, Life_Corpus, CEASE and SWMH with accuracy score of 97.4%, 97.14%, 99.2%, and 91.5% respectively. On the other hand, for SDCNL and SDD, the proposed model also gives promising results with 94.4% and 94% accuracy scores but is unable to touch the results of the previous research work.

Chapter 10

Conclusion

The main focus of the project work is to identify suicidal intent through depression detection from texts using Machine Learning and Deep Learning techniques. To accomplish this, multiple previously conducted research works were analysed, and based on the gap analysis, suitable features, ML and DL techniques were identified. Additionally, six different datasets were collected from these research papers.

Subsequently, one workflow was proposed for detecting suicidal intent from textual data based on ML and DL approach in Proposed System 1. The proposed method involved the construction of a system architecture and a flowchart. Furthermore, four features, namely BoW (unigram), TF-IDF, BoW (bigram), and TF-IDF (bigram), were extracted. These individual features and combinations of them were utilized as inputs for four different ML classifiers: SVM, NB, LR, and RF. The model also employed four DL classifiers, namely LSTM, Bi-LSTM, CNN, and BERT. Where data encoding was performed. After evaluating the model 1, it gives the accuracy of the datasets viz. Reddit, Life_Corpus, CEASE , SWMH, SDCNL and SDD as 82%, 87%, 77%, 73%, 71% and 94% respectively.

Another workflow was proposed for detecting the detecting suicidal intent from textual data based on a hybrid approach in Proposed System 2. The proposed method involved the construction of a system architecture and a flowchart. Where the LSTM , SVM and VADER Lexicon is used and combined in the proposed model. After evaluating the model 2, it gives the accuracy of the datasets viz. Reddit, Life_Corpus, CEASE , SWMH, SDCNL and SDD as 97.4%, 97.14%, 99.2% , 91.5%, 94.4 and 93.7 respectively.

As the comparison, the proposed systems manage to perform better when compared to other existing works for four datasets viz. Reddit, Life_Corpus, CEASE and SWMH with accuracy scores of 97.4%, 97.14%, 99.2% and 91.5% respectively. In other hand, for SDCNL and SDD the proposed model also gives promising results with a 94.4% and 94% accuracy score but unable to touch the result of the previous research work. It means some more research work has to be done on those datasets.

Chapter 11

Future Work

As part of the future improvement plan, the following tasks have been identified:

- Firstly, the implementation of word embedding techniques such as "Word2vec," "GloVe embedding," "FastText," and "ELMo embedding" is planned.

- Additionally, testing our models on more datasets, including CEASE v2.0, will be conducted.

- Subsequently, a user-friendly interface will be developed to enable ordinary people to detect the presence of depression in a given text.

- Finally, handwritten notes will be collected to create an image dataset, and the proposed approach will be implemented for detecting suicidal tendencies.

REFERENCES

[1]: Kessler, R. C., Bernecker, S. L., Bossarte, R. M., Luedtke, A. R., McCarthy, J. F., Nock, M. K., ... & Zaslavsky, A. M. (2019). The role of big data analytics in predicting suicide. Personalized psychiatry, 77-98.

[2]: Mohanty, S., Sahu, G., Mohanty, M. K., & Patnaik, M. (2007). Suicide in India–A four year retrospective study. Journal of forensic and legal medicine, 14(4), 185-189.

[3]: Staista.com. Number of suicides in India from 1971 to 2021 (July 10, 2023). Retrieved from: https://www.statista.com/statistics/665354/number-of-suicides-india/

[4]: Smith, K., & De Torres, I. B. C. (2014). A world of depression. Nature, 515(181), 10-1038.

[5]: Bruce, D., (April 25, 2023). Depression Guide. Retrieved from: https://www.webmd.com/depression/guide/causes-depression

[6]: Harmer B, Lee S, Duong TVH, Saadabadi A. Suicidal Ideation. 2022 May 2. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 33351435.

[7]: Mbarek, A., Jamoussi, S., & Hamadou, A. B. (2022). An across online social networks profile building approach: Application to suicidal ideation detection. Future Generation Computer Systems, 133, 171-183.

[8]: Caicedo, R. W. A., Soriano, J. M. G., & Sasieta, H. A. M. (2022). Bootstrapping semi-supervised annotation method for potential suicidal messages. Internet Interventions, 100519.

[9]: Haque, R., Islam, N., Islam, M., & Ahsan, M. M. (2022). A comparative analysis on suicidal ideation detection using NLP, machine, and Deep Learning. Technologies, 10(3), 57.

[10]: Zhang, T., Schoene, A. M., & Ananiadou, S. (2021). Automatic identification of suicide notes with a transformer-based Deep Learning model. Internet interventions, 25, 100422.

[11]: Gupta, S., Das, D., Chatterjee, M., & Naskar, S. (2021). Machine Learning-Based Social Media Analysis for Suicide Risk Assessment. In Emerging Technologies in Data Mining and Information Security (pp. 385-393). Springer, Singapore.

[12]: Schoene, A. M., Turner, A., De Mel, G. R., & Dethlefs, N. (2021). Hierarchical multiscale recurrent neural networks for detecting suicide notes. IEEE Transactions on Affective Computing.

[13]:Xu, Z., Xu, Y., Cheung, F., Cheng, M., Lung, D., Law, Y. W., ... & Yip, P. S. (2021). Detecting suicide risk using knowledge-aware natural language processing and counseling service data. Social Science & Medicine, 283, 114176.

[14]: Rajesh Kumar, E., Rama Rao, K. V. S. N., Nayak, S. R., & Chandra, R. (2020). Suicidal ideation prediction in twitter data using Machine Learning techniques. Journal of Interdisciplinary Mathematics, 23(1), 117-125.

[15]: Parrott, S., Britt, B. C., Hayes, J. L., & Albright, D. L. (2020). Social media and suicide: a validation of terms to help identify suicide-related social media posts. Journal of Evidence-Based Social Work, 17(5), 624-634.

[16]: Caicedo, R. W. A., Soriano, J. M. G., & Sasieta, H. A. M. (2020). Assessment of supervised classifiers for the task of detecting messages with suicidal ideation. Heliyon, 6(8), e04412.

[17]: Rabani, S. T., Khan, Q. R., & Khanday, A. M. U. D. (2020). Detection of suicidal ideation on Twitter using Machine Learning & ensemble approaches. Baghdad Science Journal, 17(4), 1328-1328.

[18]: Hassan, S. B., Hassan, S. B., & Zakia, U. (2020, November). Recognizing Suicidal Intent in Depressed Population using NLP: A Pilot Study. In 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 0121-0128). IEEE.

[19]:Alambo, A., Gaur, M., Lokala, U., Kursuncu, U., Thirunarayan, K., Gyrard, A., ... & Pathak, J. (2019, January). Question answering for suicide risk assessment using reddit. In 2019 IEEE 13th International Conference on Semantic Computing (ICSC) (pp. 468-473). IEEE.

[20]: Parraga-Alava, J., Caicedo, R. A., Gómez, J. M., & Inostroza-Ponta, M. (2019, November). An unsupervised learning approach for automatically to categorize potential suicide messages in social media. In 2019 38th International Conference of the Chilean Computer Science Society (SCCC) (pp. 1-8). IEEE.

[21]: Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., & Xu, H. (2018). Extracting psychiatric stressors for suicide from social media using Deep Learning. BMC medical informatics and decision making, 18(2), 77-87.

[22]: Vioules, M. J., Moulahi, B., Azé, J., & Bringay, S. (2018). Detection of suicide-related posts in Twitter data streams. IBM Journal of Research and Development, 62(1), 7-1.

[23]: Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., & Hanson, C. L. (2016). Validating Machine Learning algorithms for Twitter data against established measures of suicidality. JMIR mental health, 3(2), e4822.

[24]: O'dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on Twitter. Internet Interventions, 2(2), 183-188.

[25]: Nikfarjam, A., Emadzadeh, E., & Gonzalez, G. (2012). A hybrid system for emotion extraction from suicide notes. Biomedical informatics insights, 5, BII-S8981.

[26]: Liakata, M., Kim, J. H., Saha, S., Hastings, J., & Rebholz-Schuhmann, D. (2012). Three hybrid classifiers for the detection of emotions in suicide notes. Biomedical informatics insights, 5, BII-S8967.

[27]: Xu, Y., Wang, Y., Liu, J., Tu, Z., Sun, J. T., Tsujii, J., & Chang, E. (2012). Suicide note sentiment classification: a supervised approach augmented by web data. Biomedical informatics insights, 5, BII-S8956.

[28]: Sohn, S., Torii, M., Li, D., Wagholikar, K., Wu, S., & Liu, H. (2012). A hybrid approach to sentiment sentence classification in suicide notes. Biomedical informatics insights, 5, BII-S8961.

[29]: Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using Deep Learning. Algorithms, 13(1), 7.

[30]: Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., & Xu, H. (2018). Extracting psychiatric stressors for suicide from social media using Deep Learning. BMC medical informatics and decision making, 18(2), 77-87.

[31]: Shah, F. M., Ahmed, F., Joy, S. K. S., Ahmed, S., Sadek, S., Shil, R., & Kabir, M. H. (2020, June). Early depression detection from social network using Deep Learning techniques. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 823-826). IEEE.

[32]: Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2020, May). Cease, a corpus of emotion annotated suicide notes in English. In Proceedings of the 12th Language Resources and Evaluation Conference (pp. 1618-1626).

[33]: Morales, M., Dey, P., Theisen, T., Belitz, D., & Chernova, N. (2019, June). An investigation of Deep Learning systems for suicide risk assessment. In Proceedings of the sixth workshop on computational linguistics and clinical psychology (pp. 177-181).

[34]:Boukil, S., El Adnani, F., Cherrat, L., El Moutaouakkil, A. E., & Ezziyyani, M. (2019). Deep learning algorithm for suicide sentiment prediction. In Advanced Intelligent Systems for Sustainable Development (AI2SD'2018) Vol 4: Advanced Intelligent Systems Applied to Health (pp. 261-272). Springer International Publishing.

[35]: Zhang, T., Schoene, A. M., & Ananiadou, S. (2021). Automatic identification of suicide notes with a transformer-based Deep Learning model. Internet interventions, 25, 100422.

[36]: Haque, A., Reddi, V., & Giallanza, T. (2021). Deep Learning for suicide and depression identification with unsupervised label correction. In Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part V 30 (pp. 436-447). Springer International Publishing.

[37]:Wu, C. S., Kuo, C. J., Su, C. H., Wang, S. H., & Dai, H. J. (2020). Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records. Journal of affective disorders, 260, 617-623.

[38]: Choi, K. S., Kim, S., Kim, B. H., Jeon, H. J., Kim, J. H., Jang, J. H., & Jeong, B. (2021). Deep graph neural network-based prediction of acute suicidal ideation in young adults. Scientific reports, 11(1), 1-11.

[39]: Ophir, Y., Tikochinski, R., Asterhan, C. S., Sisso, I., & Reichart, R. (2020). Deep neural networks detect suicide risk from textual facebook posts. Scientific reports, 10(1), 16685.

[40]: Kim, J., Lee, J., Park, E., & Han, J. (2020). A Deep Learning model for detecting mental illness from user content on social media. Scientific reports, 10(1), 1-6.

[41]: Baghdadi, N. A., Malki, A., Balaha, H. M., AbdulAzeem, Y., Badawy, M., & Elhosseini, M. (2022). An optimized Deep Learning approach for suicide detection through Arabic tweets. PeerJ Computer Science, 8, e1070.

[42]: Aldhyani, T. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. (2022). Detecting and analyzing suicidal ideation on social media using Deep Learning and Machine Learning models. International journal of environmental research and public health, 19(19), 12635.

[43]: Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2022). A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes. Cognitive Computation, 1-20.

[44]: Long, S., Cabral, R., Poon, J., & Han, S. C. (2022). A Quantitative and Qualitative Analysis of Suicide Ideation Detection using Deep Learning. arXiv preprint arXiv:2206.08673.

[45]: Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2023). VAD-assisted multitask transformer framework for emotion recognition and intensity prediction on suicide notes. Information Processing & Management, 60(2), 103234.

[46]: Ji, S., Li, X., Huang, Z., & Cambria, E. (2022). Suicidal ideation and mental disorder detection with attentive relation networks. Neural Computing and Applications, 34(13), 10309-10319.

[47]: A.S.S. Wen, G.J. Yi, L.Z. Hui, L. Xiao, Q.Y. Zhen, "Suicidal Text Detection," Retrieved from: https://github.com/gohjiayi/suicidal-text-detection, 2021.

[48]: Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.

[49]: Artificial Intelligence in Design '96. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170. doi:10.1007/978-94-009-0279-4_9. ISBN 978-94-010-6610-5.

[50]: Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; Arvin, F., "Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning" IEEE Transactions on Vehicular Technology, 2020.

[51]: Russell, Stuart J.; Norvig, Peter (2010). Artificial Intelligence: A Modern Approach (Third ed.). Prentice Hall. ISBN 9780136042594.

[52]: Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). Foundations of Machine Learning. The MIT Press. ISBN 9780262018258.

[53]: Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2. ISBN 978-0-07-042807-2.

[54]: Jordan, Michael I.; Bishop, Christopher M. (2004). "Neural Networks". In Allen B. Tucker (ed.). Computer Science Handbook, Second Edition (Section VII: Intelligent Systems). Boca Raton, Florida: Chapman & Hall/CRC Press LLC. ISBN 978-1-58488-360-9.

[55]: Alex Ratner; Stephen Bach; Paroma Varma; Chris. "Weak Supervision: The New Programming Paradigm for Machine Learning". hazyresearch.github.io. referencing work by many other members of Hazy Research. Archived from the original on 2019-06-06. Retrieved 2019-06-06.

[56]: van Otterlo, M.; Wiering, M. (2012). Reinforcement learning and markov decision processes. Reinforcement Learning. Adaptation, Learning, and Optimization. Vol. 12. pp. 3–42. doi:10.1007/978-3-642-27645-3_1. ISBN 978-3-642-27644-6.

[57]: Rish, Irina (2001). An empirical study of the naive Bayes classifier (PDF). IJCAI Workshop on Empirical Methods in AI.

[58]: Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks". Machine Learning. 20 (3): 273–297.

[59]: Cramer, J. S. (2002). The origins of logistic regression (PDF) (Technical report). 119. Tinbergen Institute. pp. 167–178. doi:10.2139/ssrn.360300.

[60]:https://www.researchgate.net/publication/343384461_DEPRESSION_DETECTION_USING_INTELLIGENT_ALGORITHMS_FROM_SOCIAL_MEDIA_CONTEXT_-_STATE_OF_THE_ART_TRENDS_AND_FUTURE_ROADMAP

[61]: Famer, Pamela B.; Mairal Usón, Ricardo (1999). "Constructing a Lexicon of English Verbs". Functional Grammar (in English) 23 (illustrated ed.). Walter de Gruyter. p. 350. ISBN 9783110164169

[62]: Bach, E. 1974. Syntactic theory. New York: Holt, Rinehart and Winston, Inc.

[63]: Bussmann, Hadumod (1996). Trauth, Gregory; Kazzazi, Kerstin (eds.). Routledge dictionary of language and linguistics. Translated by Trauth, Gregory; Kazzazi, Kerstin (2nd completely revised ed.). London: Routledge. ISBN 0-415-02225-8.

[64]: Baayen, H. "Analyzing Linguistic Data: A Practical Introduction to Statistics using R. Cambridge". In Cambridge UP 2008

[65] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135.

[66]: Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

[67] Esuli, A., & Sebastiani, F. (2006). SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC-2006). Retrieved from https://dl.acm.org/doi/10.5555/1610075.1610088

[68]: Cambria, E., & Hussain, A. (2012). SenticNet: A publicly available semantic resource for opinion mining. In Proceedings of the 5th International Conference on Human System Interaction (HSI). IEEE. Retrieved from https://ieeexplore.ieee.org/document/6266204

[69]: Hutto, C.J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14)

[70]: analyticsinsight.net. Types of Sentiment Analysis and How Brands Perform Them. (November 12, 2020). Retrieved from: https://www.analyticsinsight.net/types-of-sentiment-analysis-and-how-brands-perform-them/

[71]: Gupta, S., Sinha, A.M., Prodhan, D., Modak, S. & Ghosh, N., (2023). Developing a Text Classification Model for Identifying Suicidal Ideation using Natural Language Processing Techniques. In National Conference On Research Advancements and Innovations in Computing, Communications and Information Technologies (RAICCIT 2023), Kolkata, West Bengal, India [Presented & Publication Awaited].