Exploratory analysis of suicidal intensity within depression, dissect social media post

Md Iftekharul Mobin[1,2*], Second Author[2,3†] and Third Author[1,2†]

[1*]Department, Organization, Street, City, 100190, State, Country.
[2]Department, Organization, Street, City, 10587, State, Country.
[3]Department, Organization, Street, City, 610101, State, Country.

*Corresponding author(s). E-mail(s): iftekhar.mobin@gmail.com;
Contributing authors: iiauthor@gmail.com; iiiauthor@gmail.com;
†These authors contributed equally to this work.

Abstract

Social media's Text datasets have emerged as one of the acceptable option for assessing depression and suicide emotion among Natural Language Processing (NLP) research experts. This study uses Reddit's social media datasets for exploratory data analysis to estimate the degree of suicidal thoughts within depressed person's post. The objective is to determine if a depressed individual has a suicidal tendency, determine the degree of the intensity or the opposite. This study presents an unsupervised feature analysis using the LDA topic model of the Reddit C-SSRS dataset followed by supervised classification to quantize significance. Latent suicidal intents, cross-topic co-occurrence patterns, and dominant high-impact keywords of suicide are revealed from unsupervised LDA modeling. Furthermore, Supervised machine learning classifier models are applied to determine the severity of suicide tendencies. To get the best results, cutting edge tex embedding vectorization techniques and machine learning estimators are applied. Statistical measurements depicts the degree of suicide intensity within depressed label post. From the analysis it is revealed that suicidal tendency within depression people post is extremely high. Depressed person's post showed 60% similarities in various categories of suicidal intensity.

Keywords: Suicide and Depression, NLP, Unsupervised LDA model, exploratory analysis, feature extraction

1 Introduction

Suicide is a major cause of death globally. In india, USA and many other countries large number of population dies because of suicide [21, 43]. Only In USA approximately 46,000 people committed suicide in 2020 [43]. Depression and suicide this two phenomenon are closely connected with each other. Depression triggers suicidal risk. Several studies showed that depression patients are very prone to suicidal attempt [22, 33, 47]. According to [43] more over 50% of people who opted suicide also fit the criteria for severe depression, around 4% of those who diagnosed as depressed, have records of suicidal attempts. To what extent of depression level triggers suicidal risk is a scrutiny.

Clinical depression severity estimation methods rely on interview based interrogation session where patient confront with psychologist. During the interrogation session patient are reluctant about expressing their thoughts. It is common phenomenon that emotionally distressed individual hides their feelings to others. More-often patients prefer not to disclose their emotions, often reluctant to seek help from psychotherapists, or doctor. It is difficult to anticipate a patient's psychiatric status from traditional interview-based diagnostic. Also, It is hard to quantify the level of depression during suicidal attempt. It may varies based on various factors like society, religion, family bonding, emotional maturity and many others factors. Due to lack of confidence, fear of death, religious obligations, and societal stigma against this act, even severe depressed person may not consider making an attempt at suicide. But they seek empathy consciously or unconsciously in the social sites like twitter, reddit and facebook [18]. Shen et.al in [39] and Xu et al. [51], depicted how online users debate topics connected to depression in social networks and what is their language patterns. Choudhury et al. in 2013 [19] showed that there is possibility of detecting and diagnosing depression via social media. In 2013 park et. al [34] 14 Twitter users are directly interviewed, to investigate the depressed behaviors in social media users. With the aforementioned research, it is clearly revealed that social media depression detection is not only possible but promising result can be observed. Severity level of depression can be determined by analyzing social media activities. Through data visualization we can explore various facts and clues among this two emotions. Our research focus on detection of suicidal tendency within a depressed person's post. Find out important features, explore different facts and hidden underlying information of depression and suicide.

2 Literature Review

Extensive research has been conducted before about depression and suicide. In [40] 4,882 medical students were surveyed on the basis of demographic and clinical records via WeChat app. Survey is conducted on specific demographics of population, Mostly statistical machine learning methodologies are applied to determine suicide attempt risk, features and intensities within collected samples. In 2020 Chancellor et. al demonstrated literature review of the mental health status prediction using social media data [17]. More than 75 studies of social media's Text data analysis for depression or suicide were taken into consideration between 2013 to 2018. It provides a detailed overview of data collection sources, data annotation methods, pre-processing and feature selection, model selection followed by accuracy estimation, cross validation and models' benchmarks for mental illness. In 2022 Zhang et. al claimed that 399 scientific research papers were reviewed and mental illness related research is increasing gradually [53]. There were other review papers on these two issues in which similar topics are analyzed such as Castilla et. al in 2020 [16] and Malhotra et. al 2022 [30]. All of these review based research studies analyzed mostly social media's Text data, and discussed NLP tools and techniques for depression and suicide analysis.

2.1 Multi-modal features analysis

Visual impact on individuals to detect depression also has been studied in some papers [52]. Multimodal data samples are used along with social website post such as: Instagram images are taken into consideration [16, 17]. Along with status of the social website post, Electronic Health Record (EHR) has also been taken into consideration by zheng et. al in [54] and Paulo et. al in [31] in 2020. Lang He et. al conducted research of Audio visual features [23] aiming how facial expression and voice can be used as an input features to determine mental illness effectively in 2021. It is observed that Text based expression depicts mental illness more clearly compared to other features and is dominant among researchers to detect mental issues effectively. In this study we will be focusing on the text based features only for mental illness and suicidal pattern detection.

2.2 Instrument for measuring Severity

From the very beginning questionnaire based suicidal/depression intensity measurements tool were available. These scale are applied for setting the questionnaire during the interrogation. This process provides weights to answers replied by individuals. Most renowned scales are CES-D, PHQ-9, DSM-5, DASS-21 [21, 37], Beck's Depression Inventory BSS [13], Columbia Suicide Severity Rating Scale (C-SSRS) [24, 36] etc. Most of these scales are based on predefined specific number of multiple choice questions having specific weights. [14, 21, 25, 26, 45, 50]. According to the answer feedback from the patients severity and symptoms are decided based on cumulative weight. Furthermore, statistical models are applied to investigate patterns [39, 40]. Till date many researchers are using questionnaire based measurement scales to determine suicidal symptoms [27]. In [47] research study conducted in Vantaa, Finland, for the age group of 20 to 69 years with 1119 primary-care patients using (PRIME-MD) questionnaire. Suicidal behavior was conducted, suicidal ideation was assessed using the Scale for Suicidal Ideation (SSI), and suicide attempts were then assessed using medical records. In 2014, Vuorilehto [46] examined how several assessment techniques, such as the SSI, BDI, and HAM-D, perform when predicting the incidence of suicidal thoughts in patients with depressive disorder at Vantaa Primary Care. About 153 patient were investigated for about six months to determine suicidal attempt. The study investigates whether variations in assessment tools and methodologies lead to differing estimates of suicidal ideation rates. In [20] collected dataset of 2181 redditors post which contains posts related to suicidal ideation, behavior, or attempt. Then assisted by professional practitioners, the psychiatrists prepared a dataset of only 500 redditors using tool called C-SSRS [36].

2.2.1 Comparative Analysis of scales

For suicide detection, mental health professionals typically use specialized assessments like the Columbia-Suicide Severity Rating Scale (C-SSRS) in which specific questions are asked related to suicidal ideation and behavior. These assessments are designed to evaluate an individual's risk of suicide and provide a framework for intervention and support.

- **C-SSRS** consists of a series of questions that aim to gather information about an individual's current and past experiences with suicidal ideation (thoughts), behaviors, and rescue factors etc.
  - (i) **Suicidal Ideation** The first set of questions aims to gauge the frequency, intensity, duration, and controllability of the individual's suicidal thoughts. Patient asked to describe how often they think about suicide, how intense these thoughts are, how long they last, and whether they feel they can control them.
  - (ii) **Intensity of Ideation** It assess desire to act on suicidal thoughts and whether there's a specific plan or intent to carry out a suicide attempt, this section deals whether the individual has desires.
  - (iii) **Suicidal Behavior** This part of the scale addresses any suicide-related behaviors that the individual may have engaged in, such as making a plan, preparing to attempt suicide, or actually attempting suicide.
  - (iv) **History of Suicide Attempt** If the individual has previously attempted suicide, this section assesses the methods used and how medically dangerous the attempt was and understanding the past attempt for evaluating risk.
- **DSM-5** [21] provides a set of diagnostic criteria that mental health professionals use to determine if a person's symptoms align with a specific disorder such as mood disorders, anxiety disorders, psychotic disorders, and more. Each category includes specific diagnostic criteria that must be met for a formal psychiatric diagnosis.
- **DASS-21**, or Depression, Anxiety, and Stress Scale-21, is a self assessment tool [56] commonly used to measure and assess the severity of symptoms in clinical psychology. It is a shorter version of the original DASS, which includes 42 items. The DASS-21 is a widely used instrument

for evaluating mental health status of individual's emotion. Depression part scale evaluates the presence and severity of depressive symptoms, including feelings of hopelessness, low self-esteem, lack of interest in activities. The anxiety dimension measures symptoms related to generalized anxiety, including nervousness, restlessness, and excessive worry. The stress dimension assesses the presence of symptoms related to stress, such as tension, irritability, and difficulty in relaxation.

- **BDI** Beck Depression Inventory [13] consists of 21 questions regarding the users' physiological and mental states. It contains question about patient's sleeping pattern, sadness, appetite, physical problems like interest tiredness, stomach problem, sex interest etc.
- **DSM** The Diagnostic and Statistical Manual of Mental Disorders [49] offers nine different types of depressive indications, including low mood and impaired interest. Before making a final judgment, clinicians typically determine if these symptoms have been prevalent throughout time.
- **PHQ-9** having 9 different criteria having question regarding energy, sleepiness, enthusiasm etc and 5 different criteria is mentioned mild, moderate, minimal severe and severe depression.
- **SSI** scale clinical assessment tool used to measure the severity of suicidal ideation [40]. It includes questions about the frequency, duration, controllability, deterrents, and reasons for the suicidal thoughts. It also focuses on the intensity of the suicidal thoughts, measuring how strong and compelling they are to the patients. It measures differentiate between the individual's desire to die versus their desire to continue living.

SSI, BDI, HEM-D, C-SSRS these standards scale have been successfully validated and used for many years in real-world situations, without a doubt. However, these scale might not completely encompass introvert patient behaviors and symptoms like current social media. In this study we will be focusing on the text based dataset collected from the social media sites only for mental illness and suicidal pattern detection. This study conducted exploratory analysis, depicted various latent topics, correlation between the facts and dominant key factors that represent suicide and depression resulting observe the underlying relation between this two emotions from Text dataset.

2.3 Feature Selection and exploration

Feature exploration has substantial impact on exploratory analysis, enabling the researchers to accurately understand data patterns. In 2020 chancellor. et. al [17] published an excellent overview of 75 studies published in between 2013 and 2018. It depicts the methods of data collection strategies, data annotation, pre-processing, feature selection, exploration, model development for classification, accuracy verification for models etc. LIWC, LDA, LSA, n-gram analysis etc are used as features analysis tools in previous research in [35, 44]. In 2017 Shen et. al [39] prepared a feature rich dictionary comprised of social media's profile features, visual emotional features, topic-level features, and domain-specific features. Then that dictionary features are used to train machine learning models to detect the depression on Twitter. Others have used machine text Summarization based feature extraction strategy followed by classification for depression detection is applied in [55]. In 2015 burnap [15] built a set of classifiers using lexical, and psychological features extracted from Twitter posts. Then baseline classifiers are updated by building an ensemble classifier using the Random Forest algorithm and a Maximum Probability voting classifier which improved accuracy.

In NLP research features are converted to corresponding vectors referred as embedding. Typical word embedding approaches TF-IDF, Word2Vec, CNN–BiLSTM demonstrations are available in [12]. Using LIWC features, XGBoost ML together surpasses the accuracy of CNN–BiLSTM in [12]. As classifiers most dominant machine learning approaches are XGBoost, SVM, Random Forest, other machine learning regression models. Since, source of dataset varies and various methodology followed in several researches LDA, n-gram based TF-IDF and Word2Vec reveals most frequent statistical methods.

3 Dataset

For psychological research apart from clinical domain, Text based samples are mostly collected from Twitter, Reddit [44], facebook, weibo etc websites donated by various institutions or researchers [38]. In 2021 the Computational Linguistics and Clinical Psychology CLPsych 2021 workshop organized a Task challenge for detecting suicidal risk [29]. It facilitated participants with authentic dataset for predicting suicide risk from social media Twitter. The dataset for the task includes information who attempted suicide or succeeded along with some control who have not. After collecting dataset from social sites, proper annotation is crucial for training machine learning classifier models. In[28] research study used Twitter post collection API for collecting Tweets and collected Tweets of size 2509 were obtained, of which 216 post were found relevant by 3 Expert psychologists. Furthermore, LIWC (Linguistic Inquiry and Word Count) [35] linguistic feature analysis dictionary the degree of positive and negative emotions Tweets were evaluated and results are statistically presented.

3.1 Training dataset

This research study incorporates the dataset prepared by professionals published in [20] as a training dataset. The dataset was created by shing and Gaur et. al. in 2018 [42] and 2019 [20] and is comprise of 500 posts that were filtered and extracted from 2181 Reddit posts. The dataset is then annotated by a professional practitioner psychiatrist, and divided into five categories using the criteria stated in the Columbian Suicide Severity Rating Scale (C-SSRS) [20]. This dataset introduced 5 label classification of suicidal intensities which was no risk, low risk, moderate risk, and high risk categories before.

(i) Indicator - Indicates suicidal symptoms in the post

(ii) Ideation - Having suicidal Ideas
(iii) Behavior - Having some suicidal symptoms Behavior
(iv) Attempt - Suicidal attempt in the post
(v) Supportive - Someone shows empathy and condolence for a suicidal post

Document samples based on lengths in various categories are depicted in figure 1. Supportive category is not taken into account in the analysis section since supporting group does not belong to examined specimen individual.

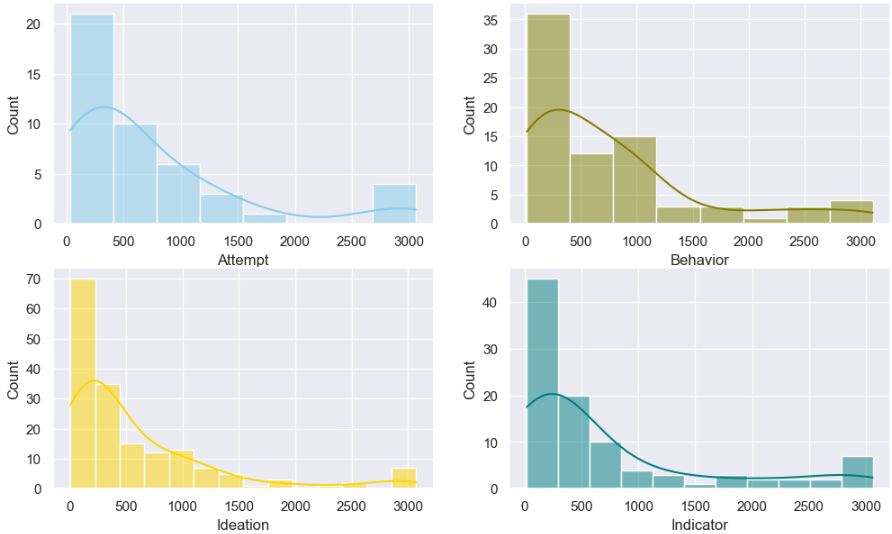


Fig. 1: Document size length in each class of suicidal category and samples frequency

3.1.1 Dataset Imbalance Handling

From figure 1 it is observed that number of samples in the attempt dataset is very low compared to other categories. To increase the dataset size balancing samples of each class, synthetic dataset is prepared. Text Data augmentation [57] is applied. Standalone categorical features of suicide categories are provided in [20]. Categorical features are mixed together shuffled and then grouped into fixed sized tokens. Random Shuffling is applied with various seed. It is considered as sample document for specific category. This process continues until we found desired number of samples. Then synthetic samples are combined with other samples to prepare train classifier dataset.

3.2 Testing/Validation Dataset

For testing dataset is collected from kaggle. It is an opesource dataset publicly available in kaggle collected from reddit website by a pushshift API contained suicide and depression category. The dataset contains posts of “SuicideWatch” and “depression” subreddits. “SuicideWatch” and “depression” posts were collected from roughly 2008 to 2021. The datasets comprised of 232,074 post annotated for binary classification as suicidal or non-suicidal Depression in [12] for detecting suicidal ideation. In this research trained classifier is applied to detect class on this two category. Main objective is to determine the suicide categories (indicator, ideation, behavior, attempt, supportive) within this dataset.

4 Methodology

In this research supervised and unsupervised both type of data modeling is conducted. Unsupervised LDA modeling explores the key features, depict coherent terms related to specific suicide category. Supervised model is applied to train machine learning classifier to determine suicidal severity categories. Trained classifier, detects the suicidal risk within given post which contains depression post. Thereby analyzing result statistically suicidal tendency significance in depression post can be observed. This whole process is depicted in Figure 2. Exploratory analysis started with dataset pre-processing. Most common techniques of NLP for data pre-processing is applied mentioned in the section 4.1.

4.1 Data Pre-processing

Social media dataset are Text data which needs data pre-processing, cleaning, feature extraction and data mining related NLP tasks.

(i) Dataset noises such as: unnecessary quotes, special characters, punctuation etc and stop words are removed. Then further pre-processing is conducted which are followed for standard data cleaning process for NLP task.
(ii) Morphological analysis is conducted to retrieve root words, process involves stemming, lemmatization etc.
(iii) Sentences are divided into equal-length fragments, and null word padding is applied to keep sample documents length same. Thereby corpus is prepared for further analysis.
(iv) Features are passed through a process in which features are converted to corresponding IDs and sentences which contain a series of IDs represented as vector. The embedding is another term that is frequently used in relation to vector text analysis.

4.2 Unsupervised modeling

Unsupervised modeling of text data refers to the techniques to analyze and extract meaningful patterns or structures from text without using labeled data. Hence, without target labels from the pre-processed dataset tokenized sample data is obtained.

4.2.1 N-gram

N-grams are generated (unigrams, Bi-grams, trigrams, etc.). To observe the most common n-grams. Word Clouds are prepared where the size of each n-gram is proportional to its frequency, gives a visually appealing representation of the most frequent n-grams, helps identify patterns of co-occurring n-grams.

4.2.2 Latent Dirichlet Allocation (LDA)

LDA topic modeling [1–4] is used in this research scope that considers documents are mixed of topic based on distribution of words. The objective is to explore the hidden topic and the topic-word distributions, presumably describe the sample documents. An expression of the LDA model is described below:

$$P(\theta_d, z, w \| \alpha, \beta) = P(\theta_d \| \alpha) \prod_{n=1}^{N} P(z_{d,n} \| \theta_d) P(w_{d,n} \| z_{d,n}, \beta) \tag{1}$$

Where $w_{d,n}$ the n_{th} word in document d, $z_{d,n}$ the topic assigned to the n_{th} word in document d, α, β are the Dirichlet LDA model parameters. controls per-document topic distribution, and per topic word distribution. θ_d represent the topic distribution. $P(\theta_d \| \alpha)$ Dirichlet distribution representing the document-topic distribution, $P(z_{d,n} \| \theta_d)$ is the word topic assignment for the n_{th} word in document d, $P(w_{d,n} \| z_{d,n}, \beta)$ is the distribution representing the observed word given a topic.

Different techniques have been developed to perform topic modeling in the unsupervised topic modeling domain of Natural Language Processing (NLP), having their own strengths and limitations [9–11]. Apart from LDA, Mallet LDA, Structural Topic Model (STM), Hierarchical Dirichlet Process (HDP), Non-Negative Matrix Factorization (NMF), Latent Semantic Analysis (LSA) etc are also prevailing and can be considered for comparative research study.

4.2.3 Topic models comparative analysis

While some variations of LDA, Mallet LDA is considered for large corpus processing and analysis [7, 9, 11]. It focuses on scalability, If large corpus needs to analyze, Mallet LDA might be more suitable. LDA in general can still be efficiently applied to moderately sized corpora. Analyzing topics within the context of metadata, STM could be a better fit. Hierarchical Dirichlet Process (HDP) can be useful when we cannot guess the number of topics in advance. As a baseline model LDA is often considered one of the most prominent choices. In [7] LSI, NMF and LDA are compared in terms of coherence and similarity measures for social media dataset and in their analysis LDA is observed most effective measures. In this study Reddit corpus LDA applied to determine dominant topics and explore correlations or co-occurrences. LDA off-the-selves python libraries produces interpretative results for exploratory topic analysis. The identified topics are represented as distributions over words, making it easy to assign meaningful labels to topics. Hence, LDA serves as a baseline for topic modeling in this research. However, how many topics are ideal it is needed to determine (see section 4.2.4) that represents topic modeling quality.

4.2.4 LDA model coherence

Coherence score estimates optimal number of topics. In this research scope coherence is measured to determine how coherent topic terms are [8]. Using this score quality of the topics produced by LDA is assessed and ensures that the topics generated are statistically significant. Coherence C_{topic} can be expressed as follows

$$C_{topic} = \sum_{i=1}^{N} \frac{1}{N(N-1)} \sum_{j=1}^{i} PMI(w_i, w_j) \tag{2}$$

Where, $PMI\,(w_i, w_j)$ represent pointwise mutual information statistical association between two words occurring together. PMI score indicates that the two words are more closely related within a topic. $PMI\,(w_i, w_j)$ represents

$$PMI\,(w_i, w_j) = \log \tfrac{P(w_i, w_j)}{P(w_i)P(w_j)} \tag{3}$$

where $P\,(w_i, w_j)$ is joint probability of occurrence of words w_i and w_j.

4.3 Supervised Modeling

Various off-the-selves Machine learning and Deep Learning models are used for supervised classification of suicide and depression in several studies [16, 17, 53]. There is conception that Deep learning models performs better than standard machine learning approaches. However, traditional machine learning models can outperform deep learning methods when extracted or filtered features are fitted into the training process. In this research cleaned features are feed into machine learning models for classification.

4.3.1 Classification

For classification mostly NLP techniques are applied for the annotated dataset collected from various online social media platforms sus as: Twitter, Reddit, Facebook, instagram, Weibo [48] etc. After that various deep learning and machine learning models are trained for classification of suicide and depression [30]. Then trained models are applied to determine the correct class of given samples. Investigation about passed research techniques, features, datasets, and performance metrics is discussed in [17, 53].

In this research pre-processed dataset converted to vector using cutting edge vectorizers. Feature vector is feed into machine learning classifier which is trained to determine suicide intensity label in depression post. Various vectorization methods are present [12, 41, 48]. TF-IDF and Word2Vec vectorizers are applied here. Following approach is applied in this study for supervised classification

(i) Pre-processed and useful features are used from Reddit's 500 post CSSR dataset for classifier Training
(ii) Used count vectorizer and TFIDF transformer to generate vectors for the dataset
(iii) Trained classifier to determine the categories of various suicidal intensities

Research methodology is portrayed in figure 2

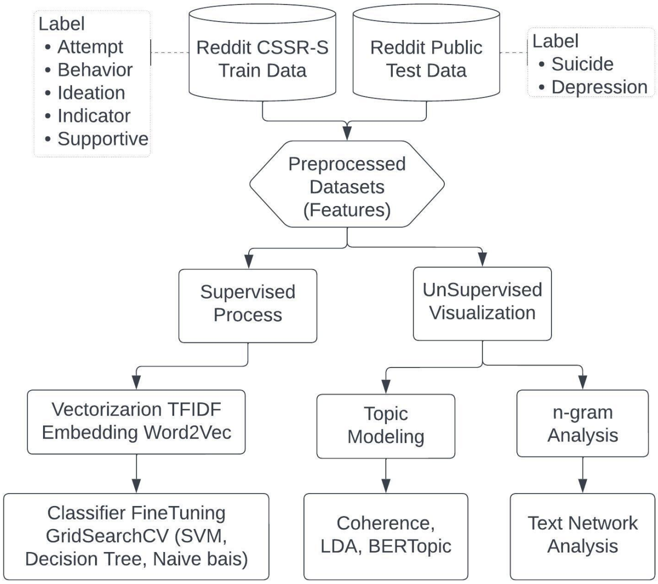
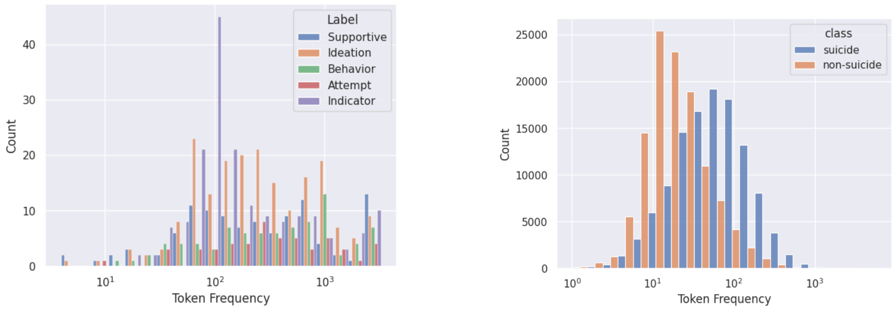


Fig. 2: Research Methodology overview

5 Exploratory analysis of features

After pre-processing document samples contains only noise free features. Document length frequency and features distribution is depicted in Figure 3. From the frequency distribution we can see some of the document sizes are very large. Hence, larger sentences are chopped and one sentence become multiple sentences of features keeping the label same.



(a) Training dataset distribution

(b) Testing dataset distribution

Fig. 3: Train and Test dataset distribution

5.1 Training dataset exploration

Uni-gram Word cloud is showed in figure 4 to depict each category and influence of dominant keywords based on frequency. Visualizing the word cloud it is revealed there are correlation between Ideation and Indicator, Behavior and Attempt categories.

5.1.1 Coherence estimation for LDA model

To calculate the coherence score gensim library provides range of options such as $u_{mass}, c_v, c_{uci}, c_{npmi}$. u_{mass} and c_vThese two methods are most popular. For given topic with words $\{w_1, w_2, w_3, \ldots.., w_n\}$ a fixed context window size is provided (default size 10 words) then coherence score is calculated using an equation $\sum_{j=1}^{i} PMI\left(w_i, w_j\right)$ which provides negative coherence score. c_v can be expressed as

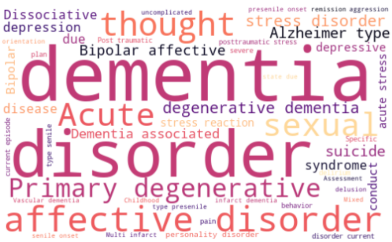
$$c_v = \frac{1}{N(N-1)} \sum_{j=1}^{i} similarity\left(w_i, w_j\right) \tag{4}$$

in which $similarity\left(w_i, w_j\right)$ represent the pairwise similarity between terms based on $PMI\left(w_i, w_j\right)$ scores. c_v provides a positive coherence score. Higher coherence values (higher than 0.5) indicate that the topics are moderately coherent and representative of meaningful themes within the text data.

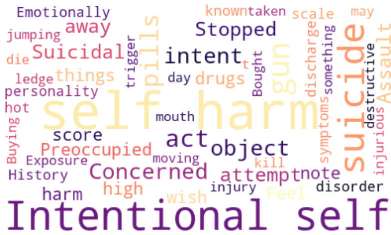
5.1.2 Keywords importance visualization using LDA

LDA driven results of various categories are visualized in this study to observe the latent topic using relative importance measurements depicted in figure 6, 7, 8, and 9 and pyldvis library in figure 10, 11, 12 and 13.

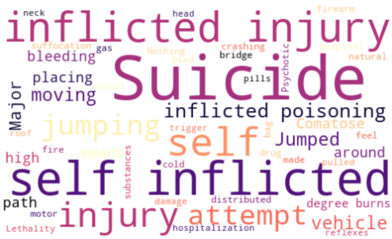
(a) Indicator category word high impact words are anxiety, disease, disorder, nervous etc



(b) Ideation category word cloud high impact words are dementia, affective disorder, stress, bipolar personality



(c) Behavior category word cloud high impact words are self-harm, suicidal, intentional etc.



(d) Attempt category word cloud high impact keywords suicide, attempt, inflicted injury etc.

Fig. 4: Train Reddit C-SSRS dataset word cloud distribution represents that there are overlapping features between Indicator with Ideation and Behavior with Attempt categories

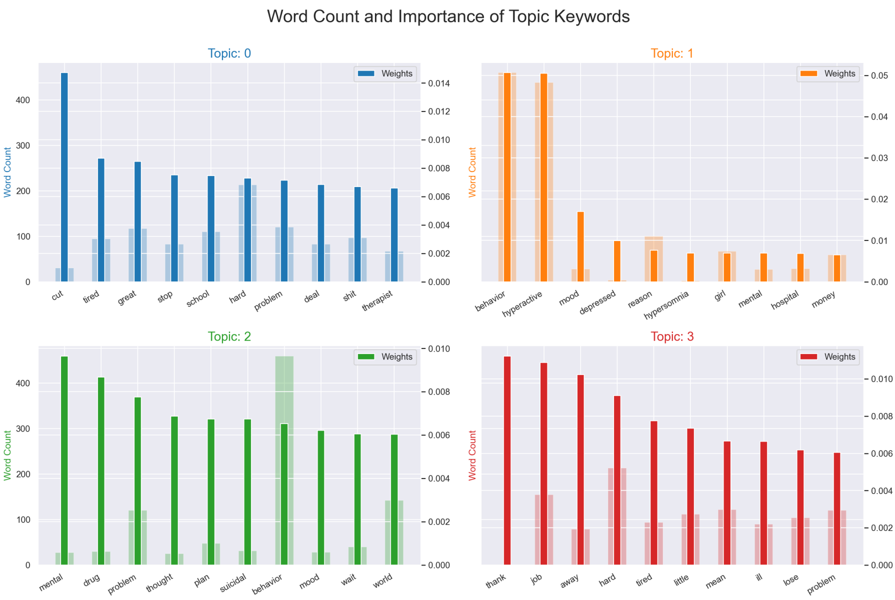
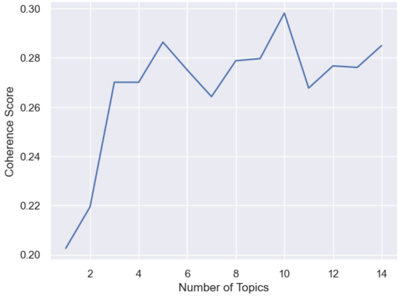
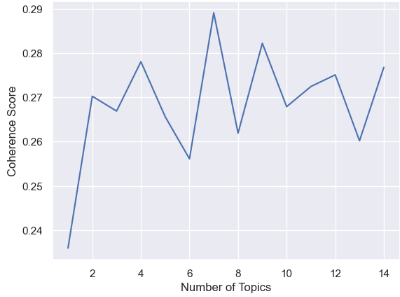


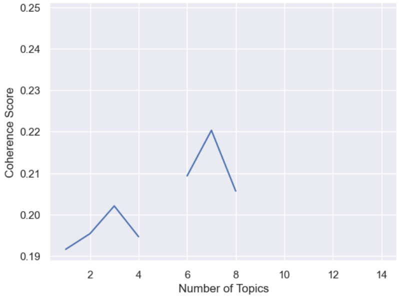
Fig. 6: Indicator category frequency vs LDA based relative Importance



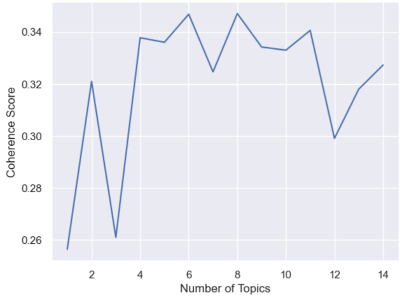
(a) Indicator category word cloud



(b) Ideation category word cloud



(c) Behavior category word cloud



(d) Attempt category word cloud

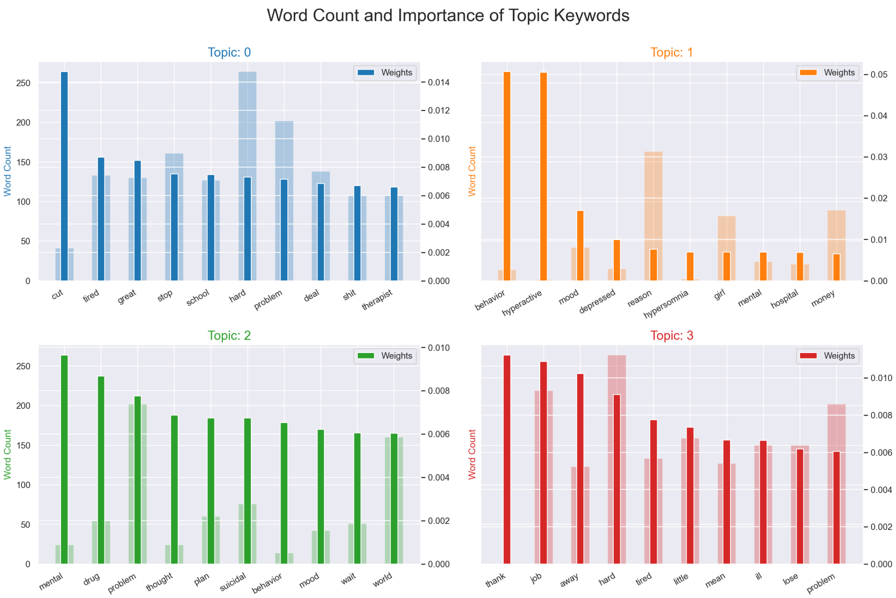


Fig. 7: Ideation category frequency vs LDA based relative Importance

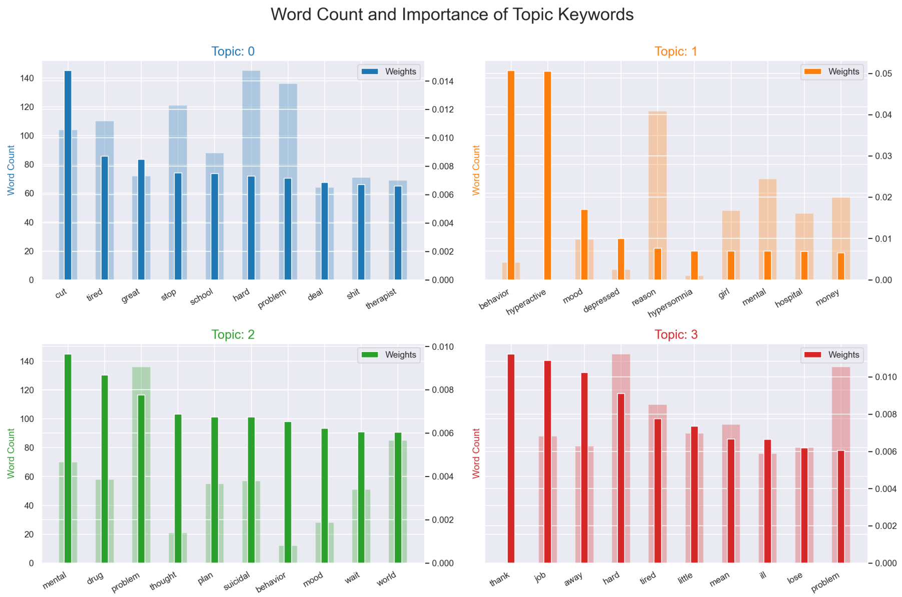


Fig. 8: Behavior category frequency vs LDA based relative Importance

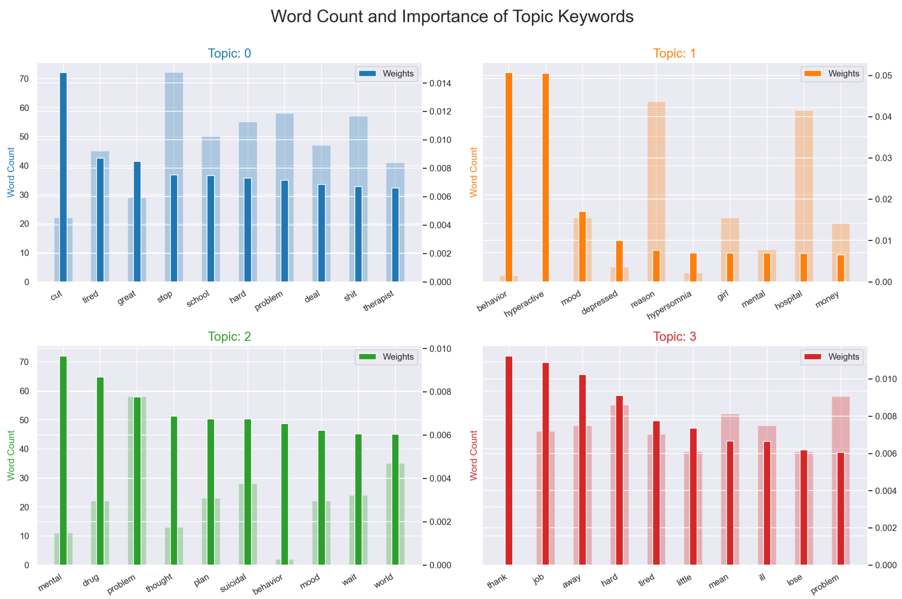


Fig. 9: Attempt category frequency vs LDA based relative Importance

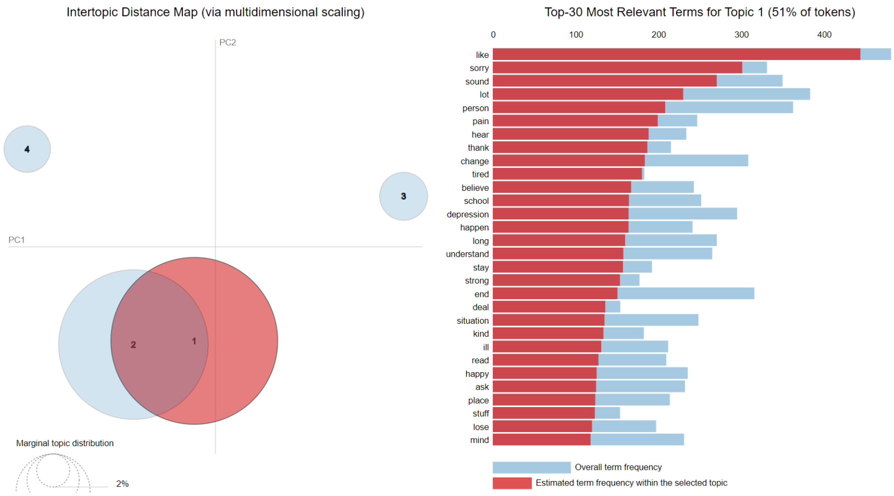


Fig. 10: Indicator categories LDA's salient terms and topic visualization

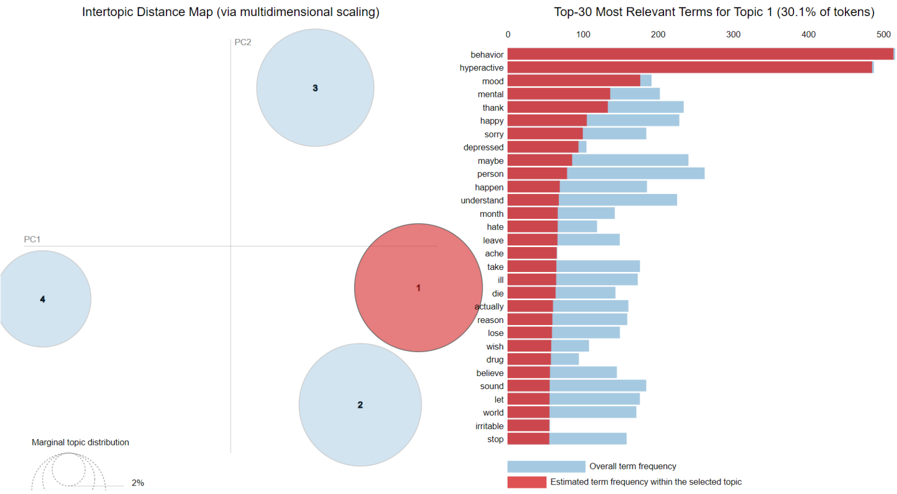


Fig. 11: Ideation categories LDA's salient terms and topic visualization

5.1.3 Combined topic modeling using BERTopic

In this research scope BERTopic topic modeling applied which leverages deep learning pretrained BERT and c-TFIDF to create dense clusters of topics. Here BERTopic applied for the combined samples. Hence we can observe topics' distinctive feature characteristics, how distinctive indicator, Ideation, Behavior and Attempt categories features are.

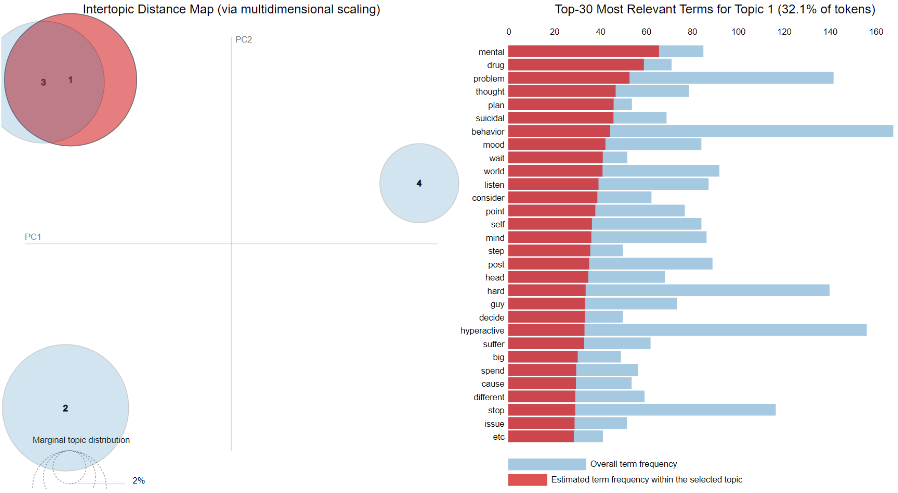


Fig. 12: Behavior categories LDA's salient terms and topic visualization

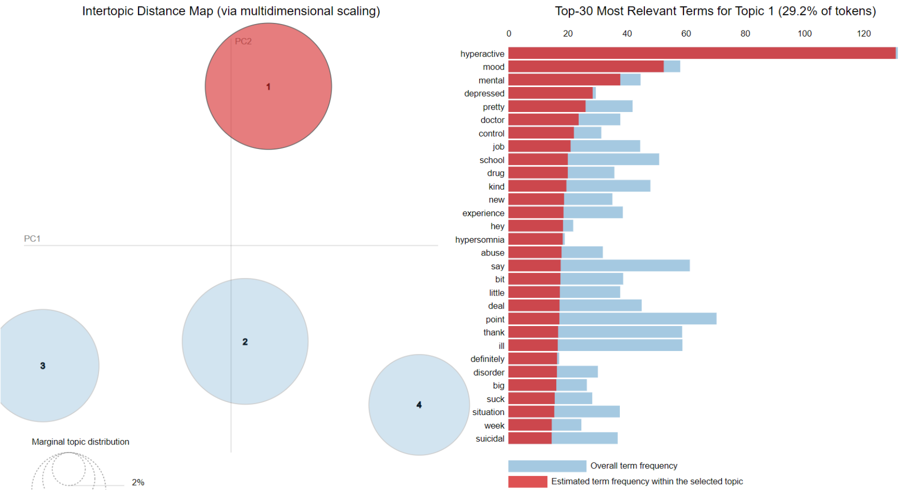


Fig. 13: Attempt categories LDA's salient terms and topic visualization

5.2 Test dataset exploration

Frequency based comparison between two categories is conducted for depression and suicide for Test dataset in Figure 15a and 15b using Uni-gram based wordcloud. From this figure we can see that top ranked words those are occurring frequently tend to have slang and abusive terms compared to suicidal category.

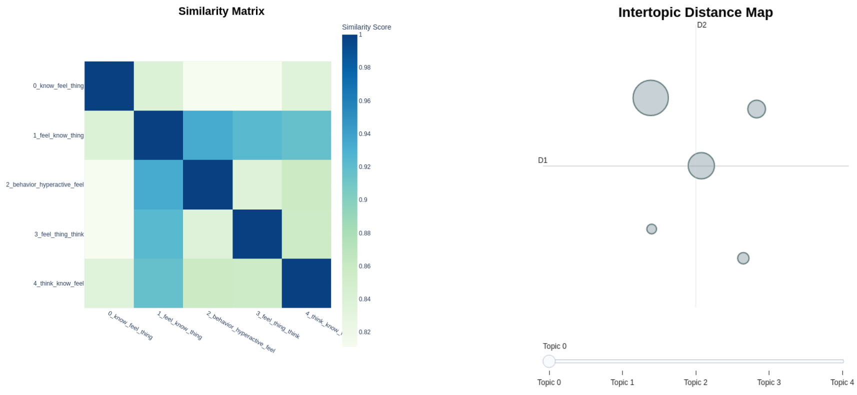
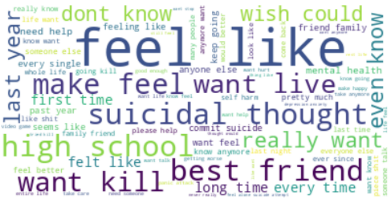


Fig. 14: Inter distance topic similarities

(a) Wordcloud in Depression category



(b) Wordcloud in Suicide category

5.2.1 Uni-gram Token distribution visualization

In Test dataset samples a large number of documents are having low features. In figure 16 the frequency of tokens in each class is depicted. Short sentence does not carry enough features which reveals does not carry enough information to be classified confidently by classifier algorithms. We started reducing the numbers of samples based on document length in Test dataset. By reducing the samples based on numbers of tokens present in a document (see Figure 16). Documents length versus category frequency information is showed in this chart. This charts explains if we filter out the shorter comments suicide post become dominant class and depression post become outnumbered.

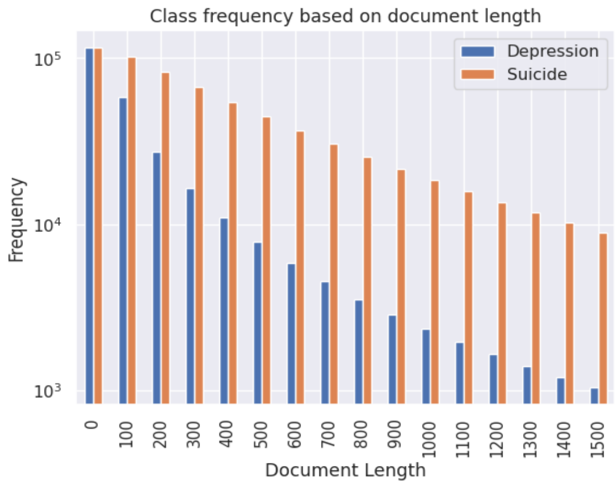


Fig. 16: Test Dataset token frequency in different Document Length

The length of document and term frequency within the corpus is visualized in Figure 14. From the distribution we can see that some of the document length are excessive long and contains more than 1000 tokens (within Train and Test both Dataset). Depression class document length are usually shorter in length. Depression document length are tend to be smaller than suicide document length. The difference showed an exponential pattern as length of document increases. Test dataset Reddit data distribution among depression and suicide class distribution ratio was equal. Filtering the class from figure 16 an interesting fact is revealed that depressed people does not want to comment very long. From figure 16, figure 15a and 15b the following comments can be inferred

1. short statements likely to be more depression category
2. Depressive statements tend to have slang and abusive words
3. Suicidal thinking people's post having very high frequency of "kill" "die" these type of words or phrases.
4. Rather suicidal depressed people want to share their thoughts with others using longer post.

It is also interesting that there are many words have high frequency such as depression or depressed but belongs to suicide class. One important fact is revealed here is that we can see although suicide, suicidal these words has high frequency in Suicide class but depression, depressed also occurred in parallel with high frequency. However, it does not reveals any direct clues in terms of hypothetical relationships between the two category. It is difficult find pattern in which we can determine the depression and suicidal thought.

5.2.2 Bi-gram features relation exploration

Bi-gram is analyzed for depression and suicide both categories. There are some Bi-grams which showed very high frequency. We called this special Bi-grams since. Special Bi-grams in the suicidal and depression both categories appeared highly frequent matter. Special Bi-grams are "mental health", "feel like", "make feel", "high school", "best friend", "really want", "suicide thought", "friend family" showed high occurrences in the test dataset. Frequently occurred Bi-grams and its pattern in the corpus is explored. We want to analyze how these words have impact with its neighboring words depicted in Figure 17 and 18. To explore the impact of special bi-grams on the samples, special bi-gram terms containing samples are filtered from dataset. After that using lebel encoder Bi-grams are encoded as integers and then chord diagram is generated. We can observe meaningful relationship within the samples between the Bi-gram features.

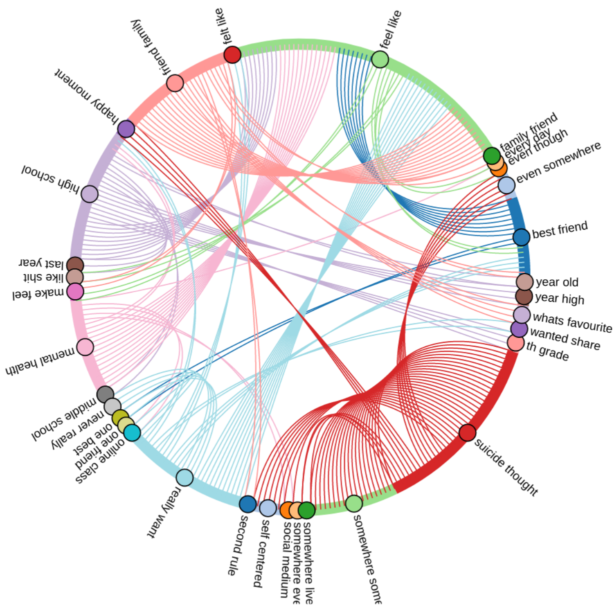


Fig. 17: Depression Chord diagram

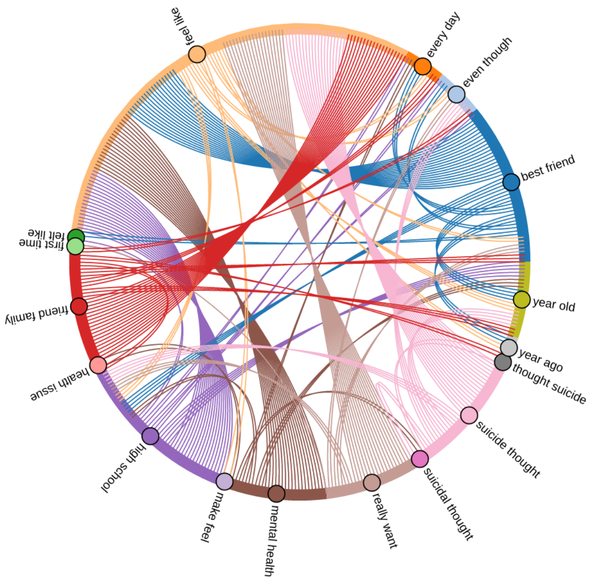


Fig. 18: Depression Chord diagram

From figure 17 and 18 two chord diagram interesting observation can be inferred (see Table 1).

Table 1: Inference from chord diagrams

Depression	**Suicide**
self centered person is depressed	have mental health issue
having suicidal thought	share though with friends (high school friends, Best friends, family members)
want to go somewhere to live	having suicidal thoughts
spend happy moments	Friend family make feel better

Tri-grams or above did not reveals much meaning information, and therefore excluded for further experimental consideration.

6 Classification Results

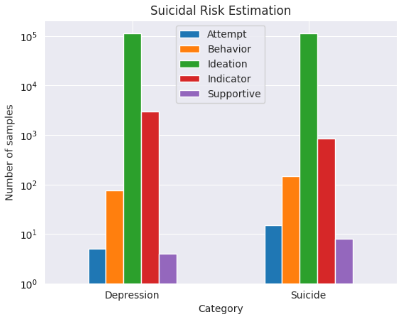
How much depression can trigger suicidal thoughts is an interesting question. In this study classifier is trained on the suicidal intensity dataset. Then trained classifier is applied on the Depression/Suicide class dataset to investigate suicidal intensity in depression. Machine learning classification algorithms used for the experiments are mentioned in the Table 2. The hyper-parameters settings for the classifiers are mostly sklearns' default settings. Classifiers are applied for the TFIDF vectorizer embedding (see results in figure 19a) and also for Word2vec pretrained vectorizer model. Gridsearch technique of sklearn library is used in Two steps. From the selected classifiers to determine the best classifiers default parameters are applied and SVM showed most promising results. Then, to achieve highest accuracy hyper-parameters are feed into grid search. Using various set of parameters, from the experiments results we found almost 60% accuracy for SVM model. From various set of values gridsearch for SVM SVC we found degree=2, gamma=0.7, kernel=rbf showed the highest accuracy.

Table 2: Applied Machine Learning Classifiers and Parameters

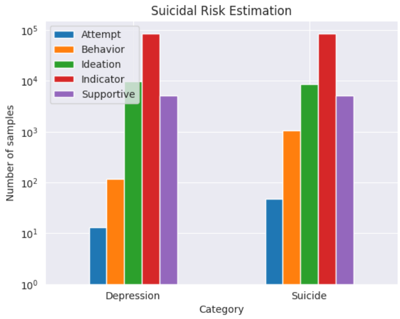
Classifiers	**Hyper-Parameters [Some Specified and rest are default sklearn params]**
K Nearest Neighbors	neighbors=5, weights=uniform
SVM (SVC, Linear, SGD)	C=0.1 $\sim$ 1.0, kernel=rbf/Linear/poly, degree=1 $\sim$ 3
Gaussian Process Regressor	α=1^{-10}
Decision Tree	criterion=gini, splitter=best, min split=2
Random Forest Ensemble	estimators=100, criterion=gini, min split=2, min leaf=1, max features=sqrt
Multi-layer perceptron	solver=Adam, α=1, hidden layer=15
AdaBoost Ensemble	estimators=50, learning rate=1.0, boosting algorithm=SAMME.R
Naive Bayes (GaussianNB)	smoothing=1^{-9}

6.0.1 Suicidal Intensities visualization

From the results we can see that suicidal ideation between depression and suicidal categories number of samples are very similar (figure 19). Within depression more number of samples are showed suicidal indicator category compared to suicide which is an interesting result. Suicidal behavior and attempt is comparatively high within the suicidal category than depression. Hence, figure 19a result seems to be pretty obvious, except for suicidal ideation category. Also for the suicidal indicator symptoms are higher within the depression category.



(a) SVM classifier applied for TFIDF vectorizer



(b) SVM classifier applied for Glove Word2Vec Pretrained model vectorizer

Fig. 19: Visualizing suicide intensities within Depression/Suicide class

For the word2vec vector embedding scenario supportive and indicator categories results are almost similar in depression or suicide both classes. There is slight difference is shown for suicidal ideation and within suicide class, suicidal ideation is slight higher. Except the behavior and attempt category for the rest categories depression and suicide showed almost similar number of samples.

7 Discussion

From the result it is revealed that suicide categories shown within depression and suicide class vividly. Specially suicidal ideation, indicator showed similar patterns. The number of samples within depression and suicide is almost similar for this two categories. Hence, we can infer depressed person comments showed suicidal ideation and suicidal indicating symptoms. Suicidal behavior and attempt showed higher number of samples within the suicide category compared to depression category. All these results seems very logical results. Although from the results mathematical formulas are not derived in this research study since results are susceptible to chosen classifier, chosen dataset, pretrained models vectors or embedding provided to the classifier.

8 Conclusion

Suicidal risk estimation task and classification samples to determine suicidal risk within social websites and blogs, techniques are discussed before. According to suicidal category previous work has been done before. However, to what extent depression level triggers suicidal risk is not yet discussed before. Also it is difficult to determine since depression and suicide categorical variables are independent factor. There is not any underlying correlation. Several research conducted to segregate which post is suicidal and which one is depression various classifiers are proposed. Extensive work has been done to improve the

classification accuracy by adopting most powerful vectorization techniques that uses cutting edge NLP models BERT and its various variants. Research has also been conducted on how much severity label of suicide within a post is studied.
The input format for the above table is as follows:

References

[1] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11):15169–15211, June 2019.

[2] Aakansha Gupta and Rahul Katarya. PAN-LDA: A latent Dirichlet allocation based novel feature extraction model for COVID-19 data using machine learning. *Computers in Biology and Medicine*, 138:104920, November 2021.

[3] Sergey Nikolenko. SVD-LDA: Topic Modeling for Full-Text Recommender Systems. In Obdulia Pichardo Lagunas, Oscar Herrera Alcántara, and Gustavo Arroyo Figueroa, editors, *Advances in Artificial Intelligence and Its Applications*, volume 9414, pages 67–79. Springer International Publishing, Cham, 2015. Series Title: Lecture Notes in Computer Science.

[4] M. Selvi, K. Thangaramya, M. S. Saranya, K. Kulothungan, S. Ganapathy, and A. Kannan. Classification of Medical Dataset Along with Topic Modeling Using LDA. In Vijay Nath and Jyotsna Kumar Mandal, editors, *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering, pages 1–11, Singapore, 2019. Springer.

[5] Ike Vayansky and Sathish AP Kumar. A review of topic modeling methods. *Information Systems*, 94:101582, 2020.

[6] Aly Abdelrazek, Yomna Eid, Eman Gawish, Walaa Medhat, and Ahmed Hassan. Topic modeling algorithms and applications: A survey. *Information Systems*, page 102131, 2022.

[7] Ashjan Alotaibi and Najwa Altwaijry. A comparison of topic modeling algorithms on visual social media networks. In *2022 2nd International Conference on Computing and Information Technology (ICCIT)*, pages 26–31, 2022.

[8] David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing*, pages 262–272, 2011.

[9] Aly Abdelrazek, Yomna Eid, Eman Gawish, Walaa Medhat, and Ahmed Hassan. Topic modeling algorithms and applications: A survey. *Information Systems*, page 102131, 2022.

[10] Xing Yi and James Allan. A comparative study of utilizing topic models for information retrieval. In *Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6-9, 2009. Proceedings 31*, pages 29–41. Springer, 2009.

[11] Ike Vayansky and Sathish AP Kumar. A review of topic modeling methods. *Information Systems*, 94:101582, 2020.

[12] Theyazn HH Aldhyani, Saleh Nagi Alsubari, Ali Saleh Alshebami, Hasan Alkahtani, and Zeyad AT Ahmed. 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, 2022.

[13] A Beck and M Kovacs. weisman a. *Handbook ofpsychiatric measures: Beck Scale for Suicide Ideation. American Psychiatric Association*, 2000.

[14] Aaron T Beck, Calvin H Ward, Mock Mendelson, Jeremiah Mock, and John Erbaugh. An inventory for measuring depression. *Archives of general psychiatry*, 4(6):561–571, 1961.

[15] Pete Burnap, Walter Colombo, and Jonathan Scourfield. Machine classification and analysis of suicide-related communication on twitter. In *Proceedings of the 26th ACM conference on hypertext & social media*, pages 75–84, 2015.

[16] Gema Castillo-Sánchez, Gonçalo Marques, Enrique Dorronzoro, Octavio Rivera-Romero, Manuel Franco-Martín, and Isabel De la Torre-Díez. Suicide risk assessment using machine learning and social networks: a scoping review. *Journal of medical systems*, 44(12):205, 2020.

[17] Stevie Chancellor and Munmun De Choudhury. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*, 3(1):43, 2020.

[18] Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. What about mood swings: Identifying depression on twitter with temporal measures of emotions. In *Companion proceedings of the the web conference 2018*, pages 1653–1660, 2018.

[19] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages 128–137, 2013.

[20] Manas Gaur, Amanuel Alambo, Joy Prakash Sain, Ugur Kursuncu, Krishnaprasad Thirunarayan, Ramakanth Kavuluru, Amit Sheth, Randy Welton, and Jyotishman Pathak. Knowledge-aware assessment of severity of suicide risk for early intervention. In *The world wide web conference*, pages 514–525, 2019.

[21] Jana M Havigerová, Jiří Haviger, Dalibor Kučera, and Petra Hoffmannová. Text-based detection of the risk of depression. *Frontiers in psychology*, 10:513, 2019.

[22] Keith Hawton, Carolina Casañas i Comabella, Camilla Haw, and Kate Saunders. Risk factors for suicide in individuals with depression: a systematic review. *Journal of affective disorders*, 147(1-3):17–28, 2013.

[23] Lang He, Mingyue Niu, Prayag Tiwari, Pekka Marttinen, Rui Su, Jiewei Jiang, Chenguang Guo, Hongyu Wang, Songtao Ding, Zhongmin Wang, et al. Deep learning for depression recognition with audiovisual cues: A review. *Information Fusion*, 80:56–86, 2022.

[24] Thomas E Joiner Jr, M David Rudd, and M Hasan Rajab. The modified scale for suicidal ideation: Factors of suicidality and their relation to clinical and diagnostic variables. *Journal of abnormal psychology*, 106(2):260, 1997.

[25] Sören Kliem, Anna Lohmann, Thomas Mößle, and Elmar Brähler. German beck scale for suicide ideation (bss): psychometric properties from a representative population survey. *BMC psychiatry*, 17(1):1–8, 2017.

[26] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. The phq-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9):606–613, 2001.

[27] Kuiliang Li, Xiaoqing Zhan, Lei Ren, Nan Liu, Lei Zhang, Ling Li, Ting Chen, Zhengzhi Feng, and Xi Luo. The association of abuse and depression with suicidal ideation in chinese adolescents: a network analysis. *Frontiers in psychiatry*, 13, 2022.

[28] Yolanda López-Del-Hoyo and Pedro Cerbuna. Exploring the risk of suicide in real time on spanish twitter: Observational study. *JMIR Public Health and Surveillance*, 8(5), 2022.

[29] Sean MacAvaney, Anjali Mittu, Glen Coppersmith, Jeff Leintz, and Philip Resnik. Community-level research on suicidality prediction in a secure environment: Overview of the clpsych 2021 shared task. In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, pages 70–80, 2021.

[30] Anshu Malhotra and Rajni Jindal. Deep learning techniques for suicide and depression detection from online social media: A scoping review. *Applied Soft Computing*, page 109713, 2022.

[31] Paulo Mann, Aline Paes, and Elton H Matsushima. See and read: detecting depression symptoms in higher education students using multimodal social media data. In *Proceedings of the International AAAI Conference on Web and social media*, volume 14, pages 440–451, 2020.

[32] Laura Martinengo, Louise Van Galen, Elaine Lum, Martin Kowalski, Mythily Subramaniam, and Josip Car. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC medicine*, 17(1):1–12, 2019.

[33] Alexander McGirr, Johanne Renaud, Monique Seguin, Martin Alda, Chawki Benkelfat, Alain Lesage, and Gustavo Turecki. An examination of dsm-iv depressive symptoms and risk for suicide completion in major depressive disorder: a psychological autopsy study. *Journal of affective disorders*, 97(1-3):203–209, 2007.

[34] Minsu Park, David McDonald, and Meeyoung Cha. Perception differences between the depressed and non-depressed users in twitter. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages 476–485, 2013.

[35] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.

[36] Kelly Posner, Gregory K Brown, Barbara Stanley, David A Brent, Kseniya V Yershova, Maria A Oquendo, Glenn W Currier, Glenn A Melvin, Laurence Greenhill, Sa Shen, et al. The columbia–suicide severity rating scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American journal of psychiatry*, 168(12):1266–1277, 2011.

[37] Lenore Sawyer Radloff. The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401, 1977.

[38] Esteban A Ríssola, Seyed Ali Bahrainian, and Fabio Crestani. A dataset for research on depression in social media. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*, pages 338–342, 2020.

[39] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, Wenwu Zhu, et al. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844, 2017.

[40] Yanmei Shen, Wenyu Zhang, Bella Siu Man Chan, Yaru Zhang, Fanchao Meng, Elizabeth A Kennon, Hanjing Emily Wu, Xuerong Luo, and Xiangyang Zhang. Detecting risk of suicide attempts among chinese medical college students using a machine learning algorithm. *Journal of affective disorders*, 273:18–23, 2020.

[41] Nisha P Shetty, Balachandra Muniyal, Arshia Anand, Sushant Kumar, and Sushant Prabhu. Predicting depression using deep learning and ensemble algorithms on raw twitter data. *International Journal of Electrical and Computer Engineering*, 10(4):3751, 2020.

[42] Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, pages 25–36, 2018.

[43] Om P Singh. Startling suicide statistics in india: Time for urgent action, 2022.

[44] Michael M Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. Detection of depression-related posts in reddit social media forum. *IEEE Access*, 7:44883–44893, 2019.

[45] Julio C Tolentino and Sergio L Schmidt. Dsm-5 criteria and depression severity: implications for clinical practice. *Frontiers in psychiatry*, 9:450, 2018.

[46] M Vuorilehto, HM Valtonen, T Melartin, P Sokero, K Suominen, and ET Isometsä. Method of assessment determines prevalence of suicidal ideation among patients with depression. *European Psychiatry*, 29(6):338–344, 2014.

[47] MS Vuorilehto, Tarja K Melartin, and ET Isometsä. Suicidal behaviour among primary-care patients with depressive disorders. *Psychological Medicine*, 36(2):203–210, 2006.

[48] Xiaofeng Wang, Shuai Chen, Tao Li, Wanting Li, Yejie Zhou, Jie Zheng, Qingcai Chen, Jun Yan, Buzhou Tang, et al. Depression risk prediction for chinese microblogs via deep-learning methods: content analysis. *JMIR medical informatics*, 8(7):e17958, 2020.

[49] Owen Whooley. Diagnostic and statistical manual of mental disorders (dsm). *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society*, pages 381–384, 2014.

[50] Janet BW Williams. A structured interview guide for the hamilton depression rating scale. *Archives of general psychiatry*, 45(8):742–747, 1988.

[51] Ying Xu, Juan Qi, Yi Yang, and Xiaozhong Wen. The contribution of lifestyle factors to depressive symptoms: A cross-sectional study in chinese college students. *Psychiatry research*, 245:243–249, 2016.

[52] Jiayu Ye, Yanhong Yu, Qingxiang Wang, Wentao Li, Hu Liang, Yunshao Zheng, and Gang Fu. Multi-modal depression detection based on emotional audio and evaluation text. *Journal of Affective Disorders*, 295:904–913, 2021.

[53] Tianlin Zhang, Annika M Schoene, Shaoxiong Ji, and Sophia Ananiadou. Natural language processing applied to mental illness detection: a narrative review. *NPJ digital medicine*, 5(1):46, 2022.

[54] Le Zheng, Oliver Wang, Shiying Hao, Chengyin Ye, Modi Liu, Minjie Xia, Alex N Sabo, Liliana Markovic, Frank Stearns, Laura Kanov, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Translational psychiatry*, 10(1):72, 2020.

[55] Hamad Zogan, Imran Razzak, Shoaib Jameel, and Guandong Xu. Depressionnet: learning multi-modalities with user post summarization for depression detection on social media. In *proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 133–142, 2021.

[56] Julie D Henry and John R Crawford. The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British journal of clinical psychology*, 44(2):227–239, 2005.

[57] Markus Bayer, Marc-Andre Kaufhold and Christian Reuter A survey on data augmentation for text classification *ACM Computing Surveys*, 55(7):1–39, 2022.