e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

CONTENT ANALYSIS OF MESSAGES IN SOCIAL NETWORKS AND IDENTIFICATION OF SUICIDAL TYPES

Mr. Ateeb Aejaz*1, Mr. Mohammed Zafar Owais*2, Prof. Gurappa Kalyani*3

*1,2Students Department of Information Science & Engineering PDA College of Engineering Kalaburagi. India

*3Assistant Professor, Department of Information Science & Engineering PDA College of Engineering Kalaburagi, India

DOI : https://www.doi.org/10.56726/IRJMETS41914

ABSTRACT

This project describes content analysis of text with to identify suicidal tendencies and types. This article also describes how to make a sentence classifier that uses a neural network created using various libraries created for machine learning in the Python programming language.

Attention is paid to the problem of teenage suicide and “groups of death” in social networks, the search for ways to stop the propaganda of suicide among minors. Analysis of existing information about so-called “groups of death” and its distribution on the Internet. Individuals who suffer from suicidal ideation frequently express their views and ideas on social media. Thus, several studies found that people who are contemplating suicide can be identified by analyzing social media posts.

However, finding and comprehending patterns of suicidal ideation represent a challenging task. Therefore, it is essential to develop a machine learning system for automated early detection of suicidal ideation or any abrupt changes in a user’s behavior by analyzing his or her posts on social media. The system leverages the Naive Bayes algorithm, which assumes independence between features, to learn the associations between textual content and different suicidal types. The dataset is preprocessed, and features are extracted from the text using techniques such as CountVectorizer. The dataset is then split into training and testing sets for model evaluation.

I. INTRODUCTION

Suicide ideation expressed in social media has an impact on language usage. Many at-risk individuals use social forum platforms to discuss their problems or get access to information on similar tasks. The key objective of our study is to present ongoing work on automatic recognition of suicidal posts. We address the early detection of suicide ideation through deep learning and machine learning-based classification approaches applied to Reddit social media.

II. SYSTEM BLOCK DIAGRAM

The system architecture for identifying suicidal types using a Naive Bayes classifier consists of several components working together to process data, train the classifier, and classify social media content. The following is a high-level overview of the system architecture:

Pre, processing: n the context of identifying suicidal types using a Naive Bayes classifier, preprocessing plays a crucial role in cleaning and transforming the collected social media data to make it suitable for analysis. Here are some common preprocessing steps involved:

- Text Cleaning: Remove any irrelevant information, such as URLs, special characters, or non-alphanumeric symbols, that may not contribute to the classification task. This step helps to eliminate noise from the data.
- Tokenization: Split the text into individual tokens, typically words or n-grams (contiguous sequences of words). Tokenization allows the classifier to process text on a more granular level.
- Stop Word Removal: Exclude commonly used words, known as stop words (e.g., "and," "the," "is"), as they typically do not provide much discriminatory power for classification tasks. Removing stop words reduces the dimensionality of the data and can improve classification performance.
- Lowercasing: Convert all text to lowercase to ensure consistency in word representations. This step prevents the model from treating the same word in different cases as separate features.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

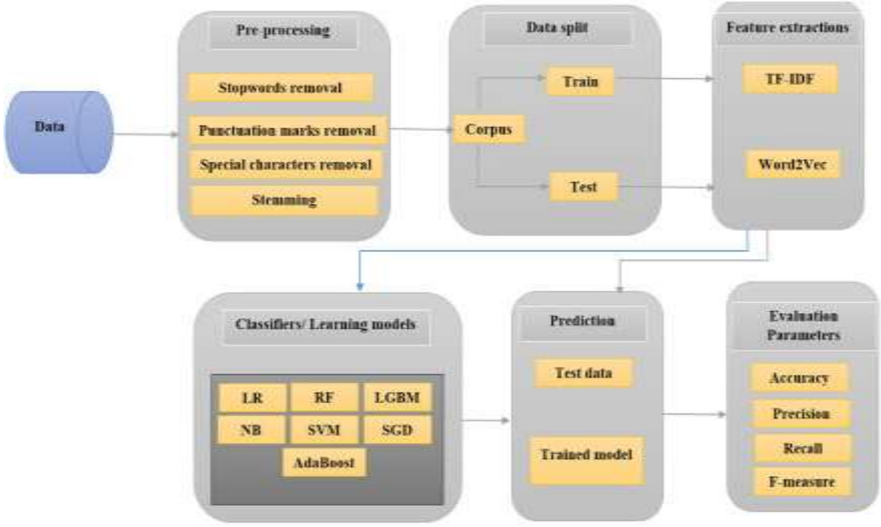

Figure 1: System architecture

Data split: In the identification of suicidal types using a Naive Bayes classifier, a common practice is to split the available data into three main subsets: the training set, the validation set, and the test set. The purpose of each subset is as follows:

- Training Set: The training set is used to train the Naive Bayes classifier. It comprises a significant portion of the available data, typically around 70-80%. The classifier learns from this data by estimating the probabilities of different features given each suicidal type.
- Test Set: The test set is used to evaluate the final performance of the trained classifier. It serves as an unbiased estimate of the model's accuracy on unseen data. The test set should not be used during the training or tuning process to avoid overfitting. Typically, it represents the remaining 10-20% of the data.

Feature extraction: In the context of identifying suicidal types using a Naive Bayes classifier, feature extraction involves transforming the preprocessed textual data into numerical representations that can be used as input for the classifier. Here are some common techniques for feature extraction:

- Term Frequency-Inverse Document Frequency (TF-IDF): TF-IDF calculates the importance of a word in a document by considering its frequency within the document and across the entire corpus. It downweights words that occur frequently across documents and assigns higher weights to words that are specific to a particular document.
- Word Embeddings: Word embeddings are dense vector representations of words that capture semantic relationships between words. Techniques like Word2Vec, GloVe, or FastText learn these representations by training neural network models on large text corpora. These embeddings can capture word similarities and contextual information, which can be valuable for classification tasks.

Classifier /Learning model: In the process of identifying suicidal types using a Naive Bayes classifier, the Naive Bayes algorithm itself is a learning model used for classification. However, there are other classifier learning models that can be employed as alternatives or in conjunction with Naive Bayes. Here are a few commonly used classifier learning models in this context:

- Support Vector Machines (SVM): SVM is a powerful and widely used classifier for both binary and multiclass classification tasks. It aims to find an optimal hyperplane that separates different classes in a high-dimensional feature space. SVM can handle non-linear classification tasks through the use of kernel functions.
- Decision Trees: Decision trees create a hierarchical structure of decision rules based on the feature values. Each internal node represents a decision based on a specific feature, and each leaf node represents a class label. Decision trees are interpretable, can handle both numerical and categorical features, and can capture complex decision boundaries.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

Evaluation Parameter: In the identification of suicidal types using a Naive Bayes classifier, several evaluation parameters can be used to assess the performance and effectiveness of the classifier. Here are some commonly used evaluation parameters:

- Accuracy: Accuracy measures the overall correctness of the classifier's predictions. It calculates the percentage of correctly classified instances out of the total number of instances. However, accuracy alone might not be sufficient if the dataset is imbalanced or if certain types of suicidal individuals are more prevalent than others.
- Precision: Precision measures the proportion of correctly predicted positive instances (true positives) out of the total instances predicted as positive. It provides an indication of the classifier's ability to avoid false positive predictions, i.e., correctly identifying the relevant suicidal types without incorrectly including other non-relevant types.
- Recall (Sensitivity): Recall measures the proportion of correctly predicted positive instances (true positives) out of the total actual positive instances. It indicates the classifier's ability to correctly identify all instances of a particular suicidal type without missing any (false negatives). Higher recall implies a lower likelihood of missing cases of that suicidal type.
- F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall, taking both false positives and false negatives into account. The F1 score is particularly useful when the dataset is imbalanced or when there is an uneven distribution of suicidal types.

III. IMPLEMENTATION

A. Load Dataset: Loads the dataset i.e. tweets.csv which consists of tweets of the users.

Figure 2: Data set

B. Pre, processing: Applying preprocessing , which clears the null data. Null data, also known as missing data, refers to the absence or lack of values in a dataset for one or more variables or observations. It occurs when the data is not recorded, is lost, or is not available for certain instances in the dataset.

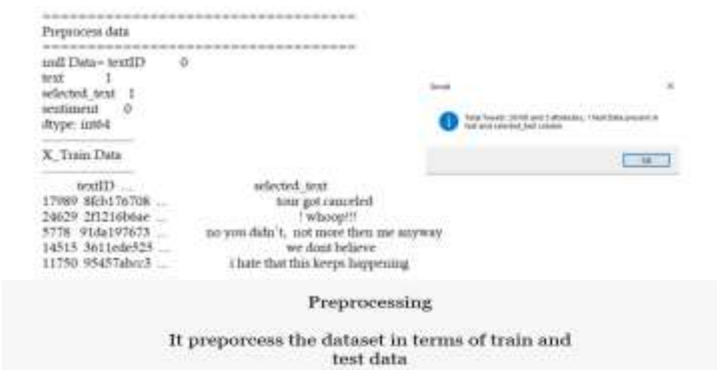

Figure 3: Pre processing

C. Apply NB classifier: Applies NB model for classification purpose, it divides the data into train and test in the ratio 70:30 70% accepts train data and 30% test data. The dataset is split into features (X) and labels (y) where X contains the text data and y contains the corresponding suicidal types.

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

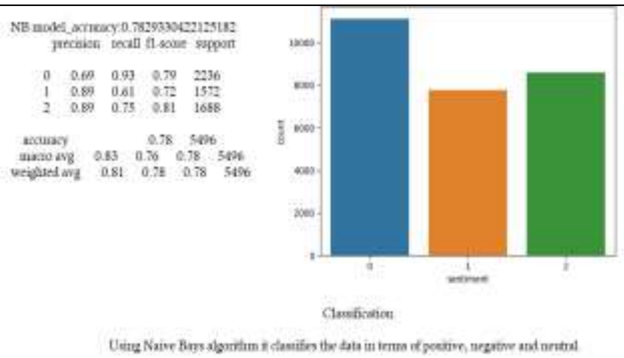

Figure 4: Graphical representation of data

IV. RESULTS

Figure 5: Detection of text messages

V. CONCLUSION

In conclusion, the identification of suicidal types using a Naive Bayes classifier offers a valuable approach to detect individuals at risk of self-harm or suicide by analyzing their social media content. The classifier leverages the Naive Bayes algorithm to learn patterns and associations between text features and different suicidal types.

Through preprocessing, data splitting, feature extraction, and classifier learning, the system is trained on a labeled dataset to recognize specific suicidal types based on textual information. The prediction phase then applies the trained classifier to new or unseen social media content, assigning probabilities and ultimately classifying the content into the most likely suicidal type. By implementing this approach, several applications and benefits can be realized, including early intervention, mental health monitoring, crisis hotline support, research insights, and enhanced safety measures on social media platforms. The classifier contributes to the identification and prioritization of individuals at risk, enabling timely and targeted interventions to prevent self-harm and promote mental well-being.

VI. REFERENCES

[1] World Health Organization. National Suicide Prevention Strategies: Progress, Examples and Indicators; World Health Organization: Geneva, Switzerland, 2018. [Google Scholar]

[2] Beck, A.T.; Kovacs, M.; Weissman, A. Hopelessness and suicidal behavior: An overview. JAMA 1975, 234, 1146–1149. [Google Scholar] [CrossRef] [PubMed]

[3] Silver, M.A.; Bohnert, M.; Beck, A.T.; Marcus, D. Relation of depression of attempted suicide and seriousness of intent. Arch. Gen. Psychiatry 1971, 25, 573–576. [Google Scholar] [CrossRef] [PubMed]

[4] Klonsky, E.D.; May, A.M. Differentiating suicide attempters from suicide ideators: A critical frontier for suicidology research. Suicide Life-Threat.Behav. 2014, 44, 1–5. [Google Scholar] [CrossRef] [PubMed]

[5] Pompili, M.; Innamorati, M.; Di Vittorio, C.; Sher, L.; Girardi, P.; Amore, M. Sociodemographic and clinical differences between suicide ideators and attempters: A study of mood disordered patients 50 years and older. Suicide Life-Threat.Behav. 2014, 44, 34–45. [Google Scholar] [CrossRef] [PubMed]

[6] DeJong, T.M.; Overholser, J.C.; Stockmeier, C.A. Apples to oranges?: A direct comparison between suicide attempters and suicide completers. J. Affect. Disord. 2010, 124, 90–97. [Google Scholar] [CrossRef] [PubMed]