International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Detecting Suicidal Tendencies through Machine Learning and Deep Learning Techniques

B.Sindhuja
Student
Hyderabad Institute of Technology and Management (HITAM)
Sindhujabejgam11@gmail.com

K.Shilpa
Student
Hyderabad Institute of Technology and Management (HITAM)
Kshilpaa01@gmail.com

A.Satya Sindhu
Student
Hyderabad Institute of Technology and Management (HITAM)
Sindhuattili14@gmail.com

K.Shiva Shankar vara prasad
Student
Hyderabad Institute of Technology and Management (HITAM)
Shivashankarvaraprasad17@gmail.com

Ms.Ila Chandana Kumari (Guide)
Associate Professor
Hyderabad Institute of Technology and Management (HITAM)
Ilachandanap.cse@hitam.org

Abstract: *This paper proposes a method to detect signs of suicide in individuals by analyzing their social media posts. The study focuses on publicly accessible Reddit data and employs machine learning algorithms and text representation techniques such as TF-IDF and Word2Vec to classify posts as either suicidal or non-suicidal based on their textual features. The models evaluated in this study are a Convolutional Neural Network (CNN) and a Random Forest model. The importance of early detection and prevention of suicide attempts is highlighted, and social media is identified as a potential platform to identify individuals at risk of suicide. The study demonstrates that the CNN model performs better than the Random Forest model in analyzing textual features to detect signs of suicide. The findings suggest that analyzing social media posts can be an effective approach to identifying individuals at risk of suicide and intervening before a suicide attempt occurs.*

Keywords: *Suicide prevention, social media, Text analysis, Machine learning, Convolutional Neural Network, Random Forest model.*

Introduction: Suicide is a significant concern, causing about 700,000 deaths globally each year, with those in their twenties and thirties being the most vulnerable, according to the World Health Organization (WHO). Depression, anxiety, shock, guilt, and anger can all contribute to suicidal ideation, which is the act of considering taking one's own life. However, the stigma attached to seeking professional help often leads individuals to express their intentions on social media platforms. Researchers debate the connection between risk factors for suicide and the transition from suicidal thoughts to actual attempts. Machine learning (ML) and natural language processing (NLP) techniques are increasingly used to deduce suicidal intent from social media content. In the past, ML algorithms have been applied to detect suicidal ideation in tweets using small datasets, resulting in low accuracy. To improve classification accuracy, appropriate annotation rules must be applied to large datasets, and deep learning (DL) models must be trained. This paper presents a method for detecting suicidal ideation by analyzing users' social media posts using text representation techniques such as TF-IDF and Word2Vec and combining DL and ML algorithms for classification. We use publicly accessible Reddit data, and our models are evaluated using accuracy, precision, recall, and F1-score metrics.

Related works

Suicidal recognition

The topic of suicide is complex and influenced by various factors. While social media provides a platform for individuals to express their thoughts and emotions, caution should be exercised when using it to identify individuals with suicidal ideation. It is crucial to provide support and resources to those who have attempted suicide as they may be at risk for future attempts. The use of linguistic characteristics and suicide notes in analyzing social media posts has limitations and may not work well with vast amounts of diverse data. Natural Language Processing (NLP) is a promising approach that allows for the identification of patterns and trends in large amounts of data, helping researchers understand the relationship between language use and mental health outcomes. NLP, along with sentiment analysis and machine learning techniques, is becoming increasingly popular for the study of mental health on social media. Deep learning methods such as LSTM and CNN are particularly useful in analyzing large amounts of data and identifying patterns that traditional statistical methods may miss.

Proposed work

1. Introduction

This project proposes a deep learning model that combines convolutional neural networks for suicidality detection. The main objectives are to evaluate the performance of the proposed model compared to the Random Forest machine learning model, conduct two different experiments with text features to test the performance of the proposed models, and analyze the suicide and non-suicide posts in the dataset using the tkinter tool. The dataset used in this project is a publicly available Reddit dataset downloaded from the Kaggle website, which includes suicidal and non-suicidal posts.

Data Preprocessing & Enhancement:

The data preprocessing and enhancement phase aims to filter textual posts to eliminate noise before applying feature extraction and embedding techniques. This phase includes stop word removal, punctuation, lowercase, tokenization, and lemmatization. The label encoder is used to preprocess the dataset, and the attributes are divided into features and labels. The dataset is then divided into training data and test data. For the training dataset, the Sklearn library is used. The post-padding sequence method is used to ensure that all text sequences in the dataset have equal real value vectors for use in the neural network DL.

Classification model:

After obtaining word embeddings for each post content using TF-IDF and Word2Vect, supervised ML algorithms-namely Random Forest Classifier (RFC) and a hybrid CNN DL algorithm-were used for classification. The performance of RFC using label-encoder word features is compared to that of CNN using Word2Vec word embeddings

Random Forest Model:

Random Forest is a supervised ML technique commonly used in classification and regression tasks. Its implementation has several features, including sparse awareness and a block architecture that facilitates the parallel creation of the tree. Repeated training with previously adjusted data can improve the algorithm's performance.

System Architecture:

The proposed system architecture adopts a three-tier software architecture. The third tier (middle tier server) is between the user interface (client) and the data management (server) components. This middle tier provides process management where business logic and rules are executed and can accommodate hundreds of users by providing functions such as queuing, application execution, and database staging. The three-tier architecture provides increased performance, flexibility, maintainability, reusability, and scalability while hiding the complexity of distributed processing from the user. This architecture is effective for net-centric information systems and internet applications. Its advantages include separating functionality from presentation, clear

separation, changes limited to well-defined components, running on WWW, and effective network performance.

DESIGNS

To conduct studies and analyses of an operational and technological nature and to promote the exchange and development of methods and tools for operational analysis as applied to defense problems.

Logical design: Logical design focuses on creating an abstract representation of the system's data flows, inputs, and outputs. This is typically done through modeling, including the use of Entity Relationship Diagrams.

Physical design: Physical design, on the other hand, is concerned with the actual input and output processes of the system. This includes how data is input, verified, processed, and displayed as output. It involves three sub-tasks: User Interface Design, Data Design, and Process Design. User Interface Design deals with how users interact with the system, while Data Design focuses on how data is represented and stored within the system. Finally, Process Design deals with how data moves through the system and how it is validated, secured, and transformed.

It's important to note that physical design in this context does not refer to the tangible physical design of an information system. Rather, it involves the design of a user and product database structure processor and control processor. The H/S personal specification is created for the proposed system, and documentation describing the design is produced for use in the next phase.

IMPLEMENTATION AND RESULTS METHODOLOGIES

Modules

- Dataset
- Importing the necessary libraries
- Splitting the dataset
- Building the model
- Apply the model and plot the graphs for accuracy and loss
- Accuracy on test set
- Saving the Trained Model
- Comparison

RESULTS

Our project aimed to detect suicidal tendencies using two different algorithms: CNN and Random Forest. The final graph displayed the accuracy, precision, recall, and f-score of both algorithms, represented on the x-axis with variant colors. Upon analysis and comparison, it was found that CNN outperformed Random Forest in terms of results.

Figure 1: Suicidal Tendency Detection

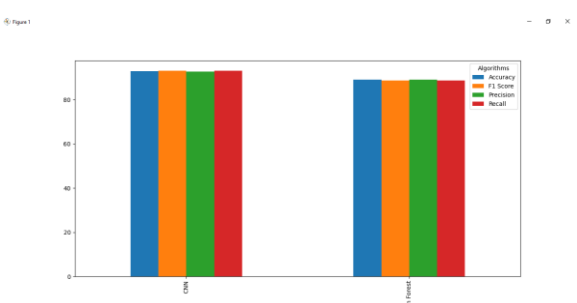

Figure 2: Comparison Graph

FUTURE WORK

In our project, due to time constraints, we were unable to explore several potential adaptations, tests, and experiments, particularly those involving real data which can be time-consuming. Moving forward, there is scope for deeper analysis of specific mechanisms and exploring new proposals using different methods.

While our current focus has been on detecting suicidal factors using CNN for better accuracy, we recognize that detection alone is not enough, and that prevention is crucial. Thus, our next goal is to take the project to

the next level by providing assistance to people who have been identified with suicidal behavior. We aim to develop a talking bot that can act as a therapist, friend, or psychiatrist and listen to people's concerns and worries. By incorporating this bot into our project, we hope to offer a comprehensive mental health package.

CONCLUSION

Therefore, early detection of suicidal ideation is crucial for prevention and intervention. Our proposed system can be used as a screening tool for identifying individuals at risk of suicidal ideation, enabling healthcare providers to provide timely intervention and support. The development of a therapy bot as a future extension of our project can provide emotional and mental support to individuals in need, contributing to a more comprehensive approach towards mental health. Overall, this project highlights the importance of using ML and DL techniques to address mental health issues and provides a foundation for further research and development in this field.

REFERENCES:

1. Sheridan, R. P. (2013). Using random forest to model the domain applicability of another random forest model. Journal of chemical information and modeling, 53(11), 2837-2850.

2. Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised leaning. International journal of computer science, 1(2), 111-117.

3. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

4. Lan, T., Hu, H., Jiang, C., Yang, G., & Zhao, Z. (2020). A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification. Advances in Space Research, 65(8), 2052-2061.

5. Szanto, K., Gildengers, A., Mulsant, B. H., Brown, G., Alexopoulos, G. S., & Reynolds, C. F. (2002). Identification of suicidal ideation and prevention of suicidal behavior in the elderly. Drugs & Aging, 19, 11-24.

6. Neethu, M. S., & Rajasree, R. (2013, July). Sentiment analysis in twitter using machine learning techniques. In 2013 fourth international conference on computing, communications and networking technologies (ICCCNT) (pp. 1-5). IEEE.

7. Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 53(6), 4335-4385.

8. Tang, D., Qin, B., & Liu, T. (2015). Deep learning for sentiment analysis: successful approaches and future challenges. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(6), 292-303.