iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: IV Month of publication: April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33989

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

Enhance: A Community Platform for Depression

Adnan Hakim[1], Arsh Shaikh[3], Hussain Sadriwalla[2], Prof. Anusha Vegesna[4]
[1, 2, 3, 4]*Department of Information Technology, Dwarkadas J. Sanghvi College of Engineering,* Mumbai, India

Abstract: The internet has become an integral part of our daily lives and is irreplaceable. Not only has it become a necessary force, but it has also practically become a chief modern-day tool for shopping, research, networking and keeping in contact with family and friends across various social media applications. However, all these benefits come with a plethora of psychological issues viz inferiority complex, loneliness, stress and cyberbullying. It has significant implications when it comes to mental and physical well being. People today try keeping such matters to themselves rather than seeking out help. This platform aims to avoid all the irrationals, and abusive behavior found online and provide a safe environment for people to support each other.
Keywords: Abuse Detection, Depression, Hate Speech Detection, CBT, Semantic Sentence Matching, Spacy, BERT, React.js, Web Application, Flutter, Mobile Application

I. INTRODUCTION

There has been an influx of personal views on public platforms with the introduction of social media [1]. While the concept was a democratic one, the modern history of social networks is not an achievement, where the forum offered was supposed to spark healthy debate. It seems that political tensions and echo bubbles have exacerbated the internet's advent.

The implementation of robust abuse detection systems is a critical component of the moderation of online content and plays a fundamental role in creating a free, safe and accessible Internet [2]. In light of recent public pressure, it is of increasing concern to both host platforms and regulators. Detection systems, such as understanding the temporal and geographic dynamics of violence, are also crucial for social science research.

Cyberbullying amplifies the effects of depression and post-traumatic stress disorder in young people, researchers have found [3]. Health professionals have also stressed that it can be even more horrifying in some situations than physical bullying. Individuals with a history of being victimized were rated more likely to be cyberbullied than those who haven't.

The platform proposed in this paper aims to resolve the subject of mental issues among individuals by offering a secure, anonymous environment. The person can freely discuss his/her problem with other individuals suffering from the same issue with qualified counsellors who can provide solutions to their problems. The platform also discusses the topic of identification of internet harassment and related question matching. The principle purpose of this paper is to increase positive mental development among individuals as well as to show ways of identifying harassment and question similarities.

The paper is structured as follows. In section 2, we define the methodologies used for this task. In section 3, we demonstrate our system architecture and components. In section 4, we discuss our dataset, the experimental setup we used for this classification mission, and the results procured. Finally, in section 5, we recap our contributions and provide some insights into this work.

II. LITERATURE SURVEY

A. Abuse Detection

In 2019, Noé Cecillon, Vincent Labatut, Richard Dufour and Georges Linarès [4] proposed a new approach to conduct abuse detection based on the premise that the interactions between users and the content of the messages exchanged convey different details. They used the content [5] and graph-based [6] methods by Etienne Papegnies, Vincent Labatut, Richard Dufour and Georges Linares. that they had previously developed for this purpose. They suggested three ways of merging them and analysing their output on a corpus of chat logs originating in a French online multiplayer game group. They then conduct a review of features to find the most insightful ones and discuss their function. Their contribution is twofold: exploring fusion techniques and, more importantly, identifying discriminatory characteristics for this issue. They then proposed a new strategy that aims to take advantage of both previously mentioned approaches. It was based on the premise that different information was transmitted by text and graph-based features. They could therefore be complementary, and their combination could boost the efficiency of the classification. They demonstrated that the characteristics derived from their content and graph-based methods are complementary and that integrating them enabled results to be improved up to 93.26 (F-measure).

In 2017, Pinkesh Badjatiya, Shashank Gupta, Manish Gupta and Vasudeva Varma [7] evaluated multiple classifiers such as Random Forest, Logistic Regression, Gradient Boosted Decision Trees (GBDTs), SVMs, and Deep Neural Networks (DNNs). In turn, the feature spaces for these classifiers defined by task-specific integrations learned using three architectures of deep learning: FastText, Long Short-Term Memory Networks (LSTMs) and Convolutional Neural Networks (CNNs). They compared feature spaces comprising TF-IDF vectors, char n-grams [8], and Bag of Words vectors (BoWV) as baselines. A dataset made available by the authors of [8] was used, which contained 16K annotated tweets. Amongst those tweets, 3383 branded as sexist, 1972 as racist, and the rest branded as neither sexist nor racist. They used pre-trained word embeddings of GloVe [9] for embedding based methods. This method outperformed the existing ones significantly. Combined with gradient boosted decision trees, embedding learned from deep neural network models led to the best accuracy values.

B. Semantic Sentence Matching

In 2006, Yuhua Li, David McLean, Zuhair Bandar, James O'Shea and Keeley Crockett [10] proposed that text-related research and applications, such as text mining, web page retrieval, and dialogue systems increasingly rely on sentence similarity measures. It provides an algorithm that considers the implied semantic and word order information in the sentences. A structured lexical database and corpus statistics are used to determine the semantic similarity of two sentences. With the use of a lexical database, their approach can model human common sense knowledge, and the inclusion of corpus statistics makes it adaptive to a variety of domains. The proposed procedure can be applied to a wide range of text knowledge representation and discovery applications.

In 2020, Daulet Nurmanbetov [11] proposed that our brains initiate a semantic search, in which we examine the results and look for sentences that match our search query. Since the inception of NLP, machine learning has been attempting to solve the problem of semantic search and has become a separate field of study. With the recent advances in Deep Learning, computers can now reliably surface essential information to us with minimum human intervention. NLP can be used to efficiently capture the context in those words, which is called "embeddings", which are typically a vector of numbers with a set of unique properties. Similar vectors exist for words with similar meanings, enabling the computation of vector similarities. Extending this concept should be able to compute the resemblance between any two sentences in the vector space. Sentence embedding models convert any given sentence into a vector so that the similarity or dissimilarity of any pair of sentences can be computed quickly.

In our own previous work [12], we reviewed different classification algorithms for abuse detection such as Naïve Bayes Classifier, K-Nearest Neighbour Classifier, Random Forest Classifier and Logistic Regression with various datasets. We also reviewed different models for semantic sentence matching such as spaCy [13], LSTM and BERT [14]. We also explored current CBT applications available in the Google Play Store and Apple App Store and their similarities and differences with our proposed system.

III. SYSTEM ARCHITECTURE

Enhance aims to reduce the gap between people and expert psychologists while providing a secure and stress-free environment. People and moderators cannot monitor high amounts of textual data to find abusive content, and hence automation is necessary. To improve the degree of abuse detection, moderators also need to check content falsely considered as abuse to refine the model with more advancements. This method significantly reduces the load of moderators and additionally helps refine the classifier with improved data.

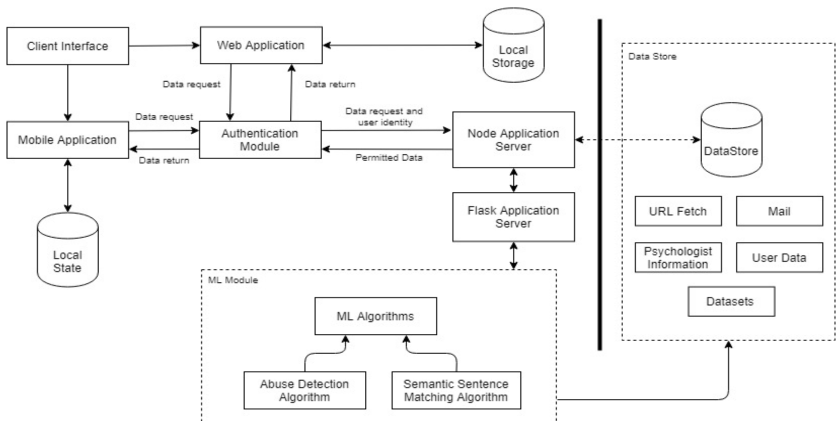


Fig. 1. System Architecture

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

We designed Enhance keeping in mind simplicity and modern design. The Enhance system architecture comprises six main components:

1) The Client Interface consists of a web and a mobile interface. It is the user interface with the help of which user can use the platform.
   a) The Mobile Application built using Flutter, enables users to access the platform on Android and iOS platforms.
   b) The Web Application built using React.js, allows the user to use the platform on the desktop.
2) The Authentication Module verifies the user and checks the data received from the front-end with the database that determines whether the user exists or not, or has the required permission to access/modify the resource. After satisfying this criterion, it allows users to view/modify the data.
3) The Node Application Server handles the data received from the front-end. It helps in storing and retrieving data from the database and also parses the data throughout the application. It helps in server-side scripting.
4) The Flask Application Server handles tasks that require machine learning algorithms, such as abuse detection and similar question search. API requests are made to the Flask server using the Node application server to invoke various ML algorithms and parse data between them.
5) The Data Store is a No-SQL Mongo database to store the data. It consists of data regarding users, psychologists, moderators, questions, answers, reports and other such collections. Data is fetched and stored in the database, and the Node server helps manage the data in the database.
6) The ML Modules consists of ML algorithms used for different tasks in the application. There are two such ML algorithms - Abuse Detection and Similar Question Search. The Flask application server invokes these algorithms.

IV. WORKING OF SYSTEM

A. Abuse Detection

Now we will discuss the dataset used for online abuse detection, data pre-processing and the algorithm used for the classification of the sentence into different classes.

We made use of a Twitter sentences dataset [15] that consists of 24,802 labelled tweets. Each tweet was voted by a certain number of individuals and classified into three categories: "Hate speech", "Offensive Language" and "Neither". As the input data introduced noise, it had to be pre-processed before applying classifications algorithms on it. Therefore, undesirable words, punctuations, numbers, and emoticons needed to be filtered out to refine the data. The data was then processed using different sequences, patterns, and stop words. Various classification algorithms were executed on the data after cleaning and dividing it into training and testing data.

Logistic regression was selected as the primary classification algorithm to conduct the analysis. To improve the model, a Grid search combined with K-fold cross-validation was performed. The dataset was then trained and tested on a 90-10 train-test split. Before fitting it into the logistic regression model, data pre-processing steps were carried out.

The best performing model had an overall precision of 0.87, recall of 0.83, and F1 score of 0.83. The accuracy of the model was 82%. From Fig. 2 and Fig. 3, we can see that there is not much difference between the outputs of both predicted and real data histograms.

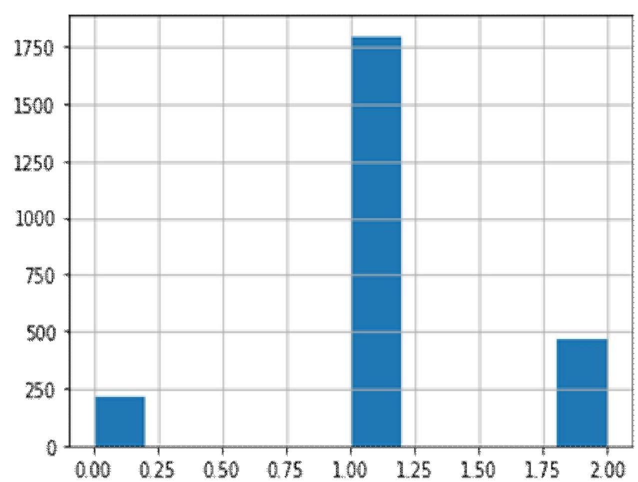


Fig. 2. Predicted Data Histogram

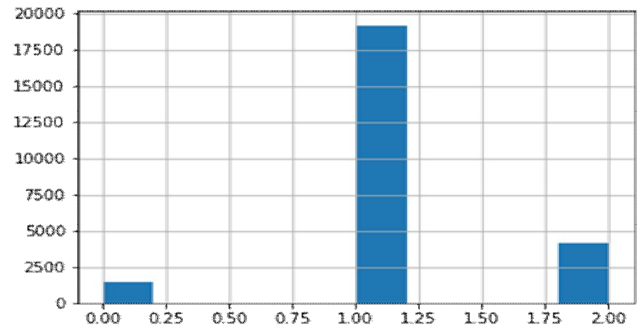


Fig. 3.Real Data Histogram

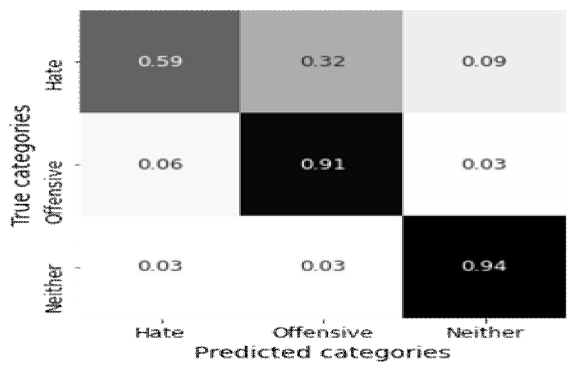


Fig. 4.Confusion Matrix

B. *Semantic Sentence Matching*

We used the BERT model and data from the SNLI (Stanford Natural Language Inference) corpus [16] to predict semantic similarity with transformers. The BERT model takes two sentences as input and provides a similarity score between them and categorizes them into various classes. For training the model, 100,000 sample sentences were used. As for training and validation purposes, 10,000 sentences were used. The sentences are classified into three separate groups in the dataset: Contradiction, Entailment, and Neutral and are assigned similarity scores based on one of these categories.

In our model, we first loaded and pre-processed the data that generated a data generator. In this data generator, we performed tokenization and the procedure of Masked LM. We then built the model and carried out training and tested it against the validation data. One of the most crucial steps of the model, fine-tuning, was executed with a low learning rate. This step helped improve the pre-trained features of the new data. After fine-tuning, we once again retrained the model and then evaluated the model on the test data. The model accuracy was around 79%, then we tested the model against different sentences, and their results are as follows.

TABLE I. Bert Model Test Results

Sentence 1	Sentence 2	Label	Similarity Score
Two women are observing something together.	Two women are standing with their eyes closed.	Contradicti on	0.91
A smiling costumed woman is holding an umbrella	A happy woman in a fairy costume holds an umbrella	Neutral	0.88
A soccer game with multiple males playing.	Some men are playing a sport	Entailment	0.94

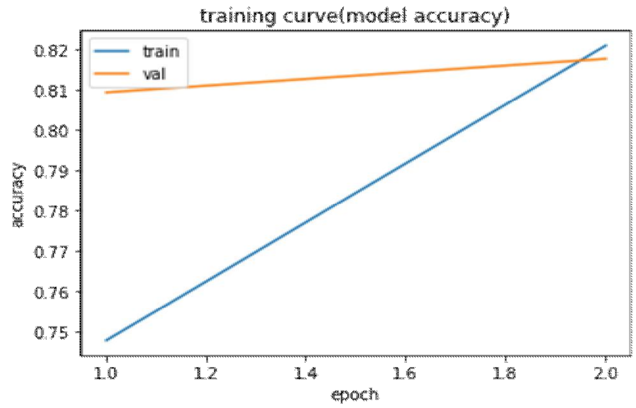


Fig. 5. Training Curve: Model Accuracy

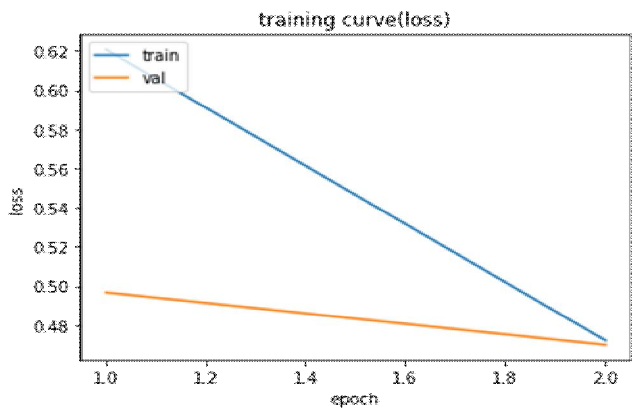


Fig. 6. Training Curve: Loss

V. CONCLUSION

Enhance is an attempt towards creating a community with an emphasis on mutual wellbeing. It aims to provide a networked forum for psychologists and other users to assist one another, further reducing the psychologist's workload. It will serve as a conduit for the vast majority of students worldwide who are affected by mental illness, not just in India. Furthermore, it assists them in overcoming these issues by using validated experimental methods. Initially, the system will depend profoundly on its users. With more traffic, the algorithms can be developed further using the database. Newer functionality can be quickly implemented in the current system since the system is modular. As a result, it would be reliable and priceless to students dealing with psychological issues.

REFERENCES

[1] The impact of cyberbullying on mental health and how to deal with online abuse, January 24, 2020. [Online]. Available: https://www.firstpost.com/health/the-impact-of-cyberbullying-on-mental-health-and-how-to-deal-with-online-abuse-7951381.html/.

[2] Vidgen, Bertram & Harris, Alex & Nguyen, Dong & Tromble, Rebekah & Hale, Scott & Margetts, Helen. (2019), “Challenges and frontiers in abusive content detection,” 80-93. 10.18653/v1/W19-3509.

[3] Online bullying more horrifying, leads to depression in youths, January 22, 2020. [Online]. Available: https://www.outlookindia.com/newsscroll/online-bullying-more-horrifying-leads-to-depression-in-youths/1714700/

[4] Noé Cecillon, Vincent Labatut, Richard Dufour and Georges Linarès (2019), “Abusive Language Detection in Online Conversations by Combining Content-and Graph-Based Features,” 13th International AAAI Conference on Web and Social Media, doi:10.3389/fdata.2019.00008.

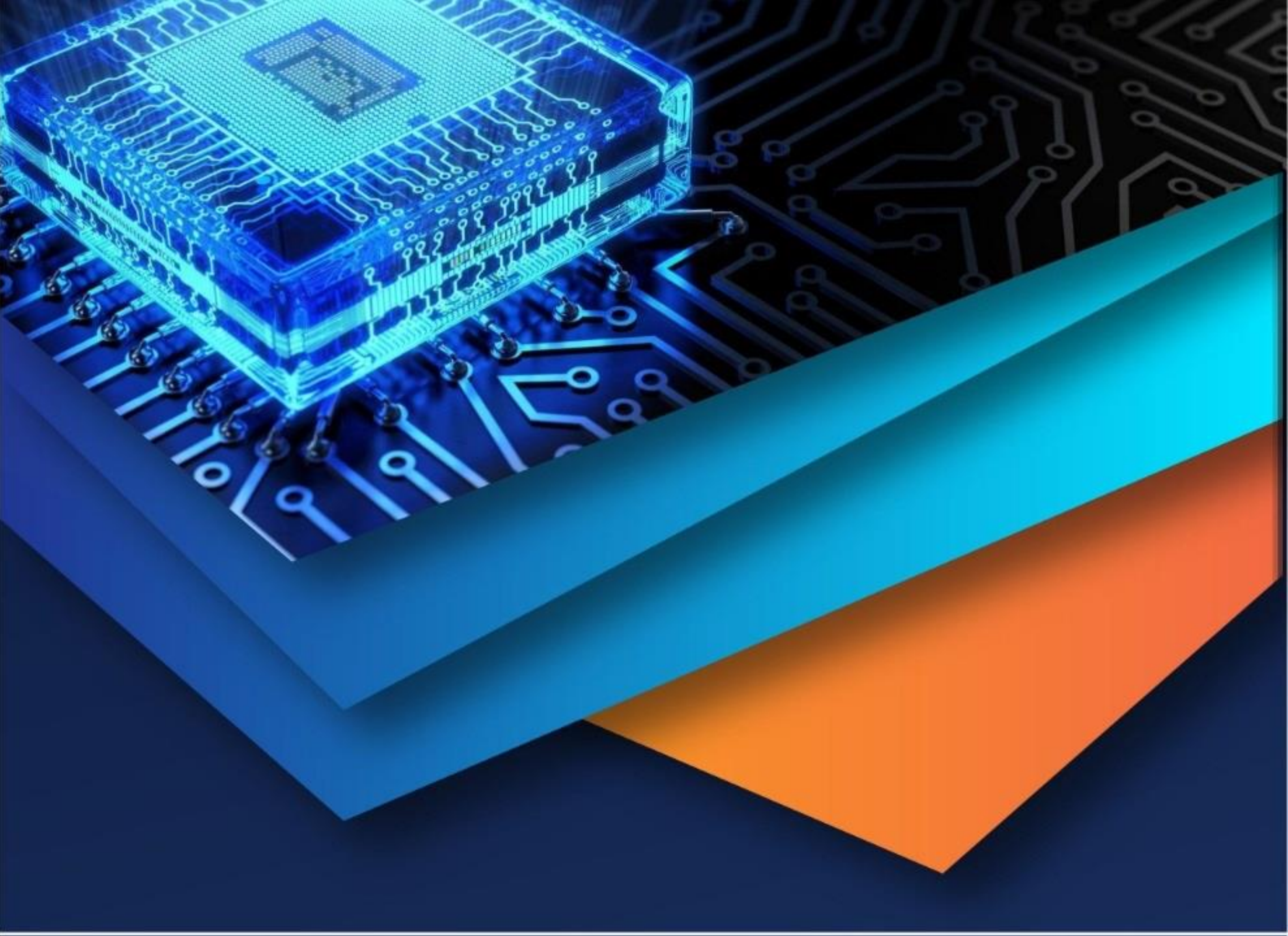
[5] Etienne Papegnies, Vincent Labatut, Richard Dufour and Georges Linarès (2017), “Impact of content features for automatic online abuse detection,” in International Conference on Computational Linguistics and Intelligent Text Processing, volume 10762 of Lecture Notes in Computer Science (Berlin:Springer), 404–419. doi: 10.1007/978-3-319-77116-8-30.

[6] Etienne Papegnies, Vincent Labatut, Richard Dufour and Georges Linarès (2019), "Conversational Networks for Automatic Online Moderation," in IEEE Transactions on Computational Social Systems, vol. 6, no. 1, pp. 38-55, Feb. 2019, doi: 10.1109/TCSS.2018.2887240.

[7] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta and Vasudeva Varma Badjatiya (2017), “Deep Learning for Hate Speech Detection in Tweets”. 10.1145/3041021.3054223.

[8] Zeerak Waseem and Dirk Hovy (2016), “Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter,” 88-93. 10.18653/v1/N16-2013.

[9] Jeffrey Pennington, Richard Socher, and Christopher Manning (2014), "Glove: Global Vectors for Word Representation," 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) 14. 1532-1543. 10.3115/v1/D14-1162.
[10] Yuhua Li, David McLean, Zuhair Bandar, James O'Shea and Keeley Crockett (2006), "Sentence Similarity Based on Semantic Nets and Corpus Statistics," IEEE Transactions on Knowledge and Data Engineering, 18. 1138-1150. 10.1109/TKDE.2006.130.
[11] Daulet Nurmanbetov, Cutting edge semantic search and sentence similarity, May 4, 2020. [Online]. Available: https://towardsdatascience.com/cutting-edge-semantic-search-and-sentence-similarity-53380328c655/.
[12] Adnan Hakim, Arsh Shaikh, Hussain Sadriwalla and Anusha Vegesna (2020), "Psychological Guidance," International Journal for Research in Applied Science and Engineering Technology, 8. 444-448. 10.22214/ijraset.2020.32185.
[13] Word Vectors and Semantic Similarity spaCy, Accessed: March 21, 2020. [Online]. Available: https://spacy.io/usage/vectors-similarity/.
[14] BERT Explained: State of the art language model for NLP, November 10, 2020. [Online]. Available: https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270/.
[15] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber (2017), "Automated Hate Speech Detection and the Problem of Offensive Language," 11th International AAAI Conference on Web and Social Media.
[16] Mohamed Merchant, Stanford Natural Language Inference Corpus (SNLI). [Online]. Available: https://github.com/MohamadMerchant/SNLI/.



10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

SJIFactor

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY