iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** XI **Month of publication:** November 2020

DOI: https://doi.org/10.22214/ijraset.2020.32185

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue XI Nov 2020- Available at www.ijraset.com

Psychological Guidance

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

Abstract: ***India, given its vast population, has an astonishing number of young minds. The ever-growing competition, peer pressure, intense competition, a great deal of work, or lack thereof, coupled with the business-minded tutoring culture and decline in employment opportunities have led to a surge in stress and other types of psychological and emotional distress amongst the students. The main objective of this paper is to provide a safe anonymous environment where people with common psychological issues such as depression, anxiety, relationship issues, fear of failures, and low self-esteem can work on these issues with the help of verified psychologists and psychiatrists and proven medical healthcare techniques to improve the condition of mental strain amongst students.***

Keywords: CBT, Depression, Abuse Detection, Hate Speech Detection, Semantic Sentence Matching, LSTM, BERT, Spacy

I. INTRODUCTION

Depression is a serious and severe medical ailment that affects how you feel, how you think, and how you behave negatively. Thankfully, it is also treatable. Depression triggers feelings of disappointment and causes a lack of interest in things a person once enjoyed. It can contribute to several physical and emotional issues and can affect one's ability to perform at his/her workplace and home.

India being a prime educational hub for students all around the world also ranks highest in teenage depression in the world. This can be caused due to a plethora of reasons, majorly caused due to succumbing to the intense amount of pressure being laid onto them by their parents and society as a whole. Some other factors involved behind the depression of students are bad eating habits, not having a socially interactive lifestyle and a balanced workout schedule, pressure after completion to find a fitting career, financial obstacles, and shattered relationships. Often, depression is mistaken for unhappiness, and people perceive it as a weak point in their lifetimes. Unfortunately, amongst Indian students, depression has been one of the leading causes of death.

Studies suggest students in Indian Universities face depression on varying levels; 37.7%, 13.1%, and 2.4% of the students suffer from moderate, severe, and very severe depression. The dissertation emphasized the need for immediate mental health support services for about 15.6% of the students suffering from severe or very severe depression. It also concludes that sharing personal problems with others and doing daily exercises were associated with positive mental health. [1] According to a survey by Lokniti-CSDS, 4 out of 10 students went through depression. The study conducted in the 15-34-year age group also found that one in four young people moderately suffered from depression, isolation, and feelings of worthlessness whereas 6% of them had thoughts of suicide at least once [2]. The methodology proposed in this paper attempts to tackle the problem of mental issues amongst students by providing a safe anonymous environment wherein the student can openly discuss his/her problem with other students suffering from the same problem as well as psychologists who can provide solutions to their problems. The paper also tackles the problem of online abuse detection and similar question matching. The main aim of this paper is to elevate positive mental growth amongst students in India.

II. LITERATURE SURVEY

In 2014, Tan Qiao Yi Michelle et al. [3] carried out a Google Play Store and Apple App Store analysis of Cognitive Behavioural Therapy-related applications to investigate the existing state of advancement of CBT smartphone apps, their breadth of coverage, and the extent of adoption. They analysed apps to identify facets of app-supported CBT. They first studied the Android and Apple app stores using Cognitive Behavioural Therapy-related terms such as CBT, cognitive counselling, cognitive restructuring, CBT strategies, behavioural therapy and exposure therapy. They then installed apps and looked at their significance to their research and ruled out non-functional or error-prone applications. Of the 37 apps surveyed, only 15 of them had real support for CBT interventions. Several types of cognitive restructuring exercise were sponsored by most, with only one app providing a behavioural experiment. Among the apps surveyed, they saw data collection as the most common feature which consisted of characteristics such as reporting on mood, monitoring sleep, journals, and worksheets. 23 of the 37 applications made it possible for users to record their mood and feelings. A popular feature was User Education, given by about 30 per cent of the apps. They found the apps to be predominantly self-help, in essence, users are provided with tips and coping strategy resources to allow users to identify their problems.

©IJRASET: All Rights are Reserved

In 2019, Ariel Teles et al. [4] evaluated upwards of 553 applications, 405 of which were from the Google Play Store, upwards of 125 applications from the Apple App Store, and 21 applications from both markets were listed as duplicate apps. In terms of their cost, apps were categorized into free, charged, and partial. A total of 157 were free of cost, 13 were paying, and 46 charged for premium services. With the following findings, they described the type of a developer: 167 apps created by commercial companies, fifty built by people, eight apps maintained by mental health professionals, six non-commercial organisations, and five university apps. The main findings reflect the evolution of the applications' purposes and roles, showing that applications aim to meet customer needs, but do not always take into account the need for face-to-face meetings with mental health professionals. As a result, the findings of this study may be an opportunity to recommend ways to identify and facilitate the involvement of qualified practitioners in the process of treatment and monitoring.

In 2017, Palak Baid et al. [5] in their paper analysed varying techniques for sentiment analysis of movie reviews. They analysed the sentences classifying them into positive and negative sentences using various classification methods. They made use of the Naïve Bayes Classifier, K-Nearest Neighbour Classifier, and Random Forest Classifier. The Naïve Bayes classifier achieved an accuracy of 81.45%, the Random Forest classifier achieved an accuracy of 78.65%, and the K-Nearest Neighbour classifier achieved an accuracy of 55.30%. After experimenting with the data on each of these algorithms they observed that Naïve Bayes out-performs other algorithms on the dataset achieving the highest accuracy.

In 2015, Humera Shaziya et al. [6] in their experiment, used WEKA to translate pre-processed text files into arff files, stemming, stop word elimination, tokenization, and selection of features, to estimate the execution of the selected function. They used text categorization to automatically translate reviews into one or more than one predefined category or group, taking into consideration that a single review may express both positive and negative sentiment. The earlier work done by them has categorized sentences into either positive or negative sentiments, and in their later work, they considered the semantics of the sentences. They made use of Naïve Bayes and SVM. On the evaluation following results were obtained: accuracy of 85.1% on Naïve Bayes and 84.5% on SVM. Experimental findings indicate that the influence of an optimal set of features increases the accuracy of both classifiers. However, regarding the accuracy of the classifier, Naïve Bayes shows to be better than SVM.

In 2015, Aditya Thyagarajan et al. [7] in their paper, studied the Siamese adaptation of Long Short-Term Memory (LSTM) network for labelled data that comprised of pairs of variable-length sequences. They made use of Word2vec to convert the sentences into vectors. Their model makes use of LSTM to read in these word-vectors that represents the input sentence. The similarity between these representations was used as a predictor of similarity. They made use of the SICK dataset that contains 9927 English sentence pairs with 5000 training sentences and 4927 testing sentences. A relatedness label corresponding to the average relatedness that was measured by 10 different individuals was labelled for each pair. They also provide their model with inputs that cannot be inferred due to the small number of training sentences. Their model takes input word-vectors that is pre-trained on an external corpus. They used the 300-dimensional word2vec embedding that can capture intricate word relationships. In all three measurement metrics, their methodology produced better results than contemporary semantic relatedness schemes.

In 2018, Lin Yao et al. [8] in their paper demonstrated how text similarity could be accomplished using a Long Short Term Memory (LSTM) encoder. The experiment was mainly divided into three stages: pre-processing, training, and evaluating. The pre-processing stage uses the word2vec method to vector every word. They proposed a new normalization algorithm to avoid gradient vanishing problems in the process of backward propagation. They used LSTM to sequence information, whereas cosine distance was used to calculate vector distance between short texts in evaluating stages. They used the MSR Paraphrase Corpus dataset and Quora Question Pairs dataset. The MSR dataset contains 5,801 pairs of sentences extracted from the web while Quora Question Pairs contains 404,302 pairs of questions. They proposed a novel algorithm which helps in computing unlabelled short text similarity by making use of an LSTM encoder. Their experiment confirms their algorithm got higher accuracy on both datasets.

As per the research done on existing web and app projects, the bulk of the applications were more focused on artificial intelligence-based solutions to combat problems like depression and anxiety. Although this may work to a certain degree, research carried out on such applications confirms that the people using such applications don’t as such gain from it. As human sentiments are complicated, most artificial intelligence-based projects don’t solve the principal problems. Another type of application was based on recording your daily routine and uses worksheets. Such applications provide worksheets that significantly influence the life of the person and are not feasible in a real-life scenario. Other apps found were based upon proven methods to combat such mental issues, such as more positivity, calming music, and some positive quotes to provide a proper environment. While these applications do help, they do not solve the core problem which exists even after using such applications. These provided temporary solutions to those problems.

©IJRASET: All Rights are Reserved

TalkLife was one such application found to provide a safe social community [9]. On further researching this application, by reading reviews of the users, there were some serious drawbacks. The main problem was the fact that many features were blocked by a paywall, hence restricting many people from actual psychological help. Another significant problem includes the lack of anonymity. Due to this, people usually answered questions of the people they found more appealing or attractive judging by their profile, which deprived other people of getting solutions to their problems. Another shortcoming was the lack of actual guidance from psychologists and psychiatrists, which caused many people to be unsatisfied with their answers.

III. PROPOSED METHODOLOGY

The plan is to build an application coupling advantages of different types of applications while combating the disadvantages of those applications by providing solutions to enhance such applications. The main target is students who mostly lack the means required to combat their problems, much more than adults. Furthermore, solving the difficulties of people at a younger age hopefully means they can have a better stress-free life in their future.

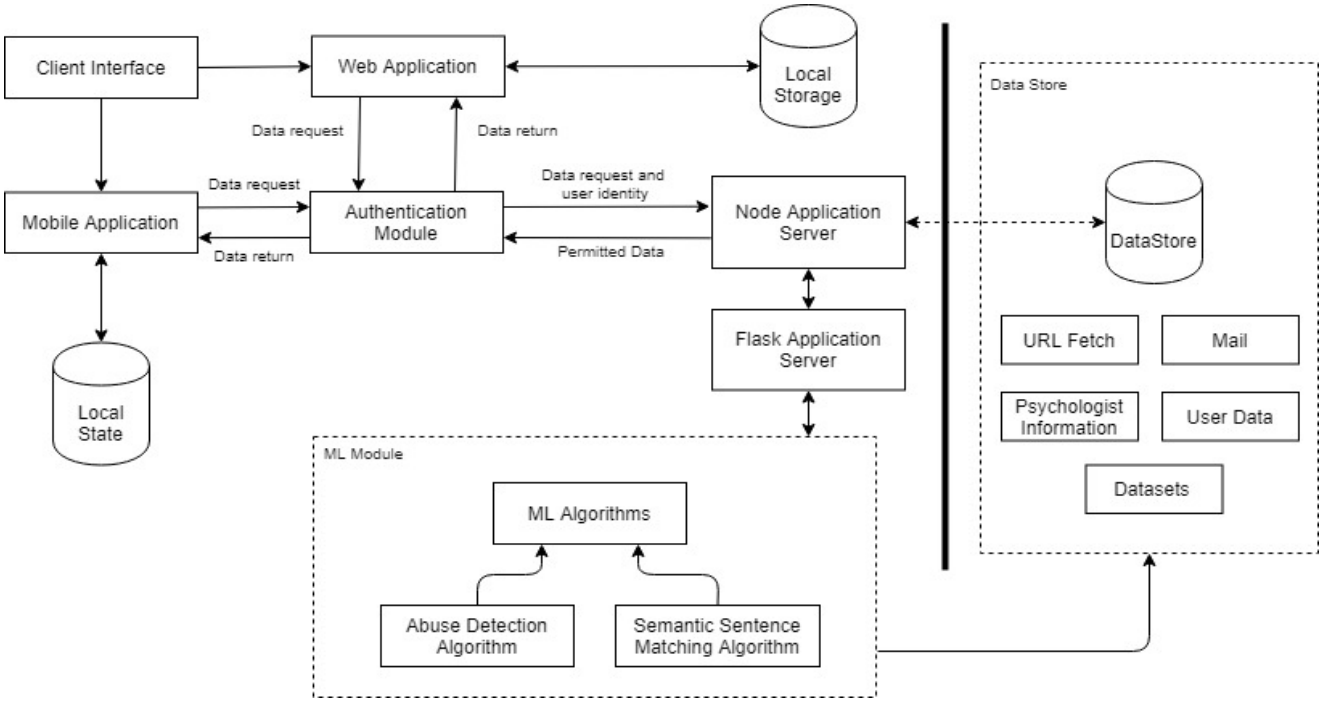


Fig. 1 Proposed System Architecture

The plan is to provide a cross-platform anonymous environment for students where they can freely express their problems on both web and Android/iOS platforms. For cross-platform mobile development, the aim is to use Flutter to provide a seamless user experience across mobile platforms. As for web applications, React will be used as it enables the usage of cleaner abstraction methods and more reusable components as well as better cross-browser compatibility.

The mobile and web application will first go through an authentication module that will verify the user and save their token locally that will be used to call the web server for all further API requests. To ensure security and availability at all times, the data will be stored in the cloud and will be handled by the webserver. All user posts will go through a moderation process to ensure hate and abuse-free communication via the moderation module. This will include automatic as well as manual moderation as per the requirement. The proposed system will also show the user about similar questions that are already asked using semantic sentence matching.

The proposed system can be mainly divided into two modules: Abuse Detection and Semantic Sentence Matching. In this section, these modules are explained in detail.

A. Abuse Detection

Abuse detection is the mechanism of detecting the use of any slang language used by an individual. As the proposed system allows users to post their opinions on any topic or answers various questions, abuse detection is required to check whether language is appropriate or not.

©IJRASET: All Rights are Reserved

For performing abuse detection, the plan is to use Machine learning by using different classification algorithms to check whether a user is using appropriate language or not. Algorithms such as Naïve Bayes, Support Vector, Random Forest, and Logistic regression provide accurate results, with three different datasets that are IMDB movie reviews [10], the Sentiment140 dataset [11], and a Twitter dataset [12]. The IMDB dataset consists of 50,000 movie reviews classified as positive and negative. The Sentiment140 dataset consists of 1,600,000 tweets classified as positive and negative, and the Twitter dataset consists of 28,402 tweets classified into three different categories: Hate Speech, Offensive language, and Neither. These tweets are categorized based on votes given by some number of individuals.

The next task is to use various data pre-processing techniques that eliminate the noise and unwanted data that is present in the before mentioned dataset. It will also refine the data by removing all undesirable words, punctuations, numbers, and emoji that need to be filtered out.

After implementing different algorithms on all these datasets, the algorithm with the highest accuracy will be incorporated into the proposed system. The algorithm will classify the sentence given by the user, and only non-abusive sentences will be posted.

B. Semantic Sentence Matching

Semantic similarity is one of the applications of Natural Language Processing (NLP). It is used to find the similarity between two phrases or sentences. Semantic similarity helps to learn the contextual similarity of two or more phrases that help to determine whether they have the same meaning or not. The proposed system makes use of semantic similarity to determine the resemblance between two questions due to which the user can optimize his/her search and get the desired output quickly.

For applying semantic similarity, the plan is to make use of Spacy [13] and the BERT model [14]. Spacy is an open-source library that is used for various NLP applications. These various named entities are identified and replaced by category tokens. The similarity is determined by comparing word vectors or word embedding, multi-dimensional meaning representations of a word. Spacy comes with a corpus of English words that can be used for finding the similarity between the phrases. This corpus consists of common English words with their vectors. BERT [14] is another model that is to be used. It performs the bidirectional training of transformers and creates a popular attention model for language modelling. As BERT is a bidirectional model, the accuracy with the BERT is more as compared to other models. It also has a deeper knowledge of language context that was not possible with the models trained in a single direction. The datasets to be used are the Quora question-answer dataset [15] and dataset provided by SNLI (Stanford Natural Language Inference) corpus [16]. The Quora dataset consists of two sentences that show whether sentences are similar or not by giving binary values. The dataset from the SNLI corpus classifies the sentence into three categories: Contradiction, Entailment, and Neutral, and gives the similarity scores between two sentences based on those classes. In the end, the one which performs best will be implemented.

IV. CONCLUSION

The proposed system is meant to reduce the workload of field experts. It is meant to provide an interconnected platform for psychologists and other users to help each other out. It will deal as a medium for the vast majority of students affected by mental health not only in India but all over the world. Moreover, it helps them to overcome these issues using proven scientific techniques and valid experts in the field of psychology. Abuse detection is possible to implement using datasets [10][11][12] and classification algorithm models. Some shortcomings as of now for the system might include the accuracy of the abuse detection and semantic sentence matching algorithms that can be further improved with time. Semantic sentence matching can be implemented using Spacy [13] and BERT [14] and datasets [15][16]. The system will heavily rely on its users in the initial phase. But once there's enough traffic, the algorithms can be further improved based on the database. Since the system will be modular, newer features can be easily added to the existing system. Thus, on successful implementation of the proposed system, it will be reliable, eccentric, and invaluable to students facing psychological issues.

REFERENCES

[1] (2017) Study Finds 37 Per Cent University Students In India Suffer Moderate Depression. [Online]. Available: https://www.ndtv.com/education/world-health-day-2017-study-finds-37-per-cent-university-students-in-india-suffer-moderate-depressio-1679142/

[2] (2020) Student suicides rising, 28 lives lost every day. [Online]. Available: https://www.thehindu.com/news/national/student-suicides-rising-28-lives-lost-every-day/article30685085.ece/

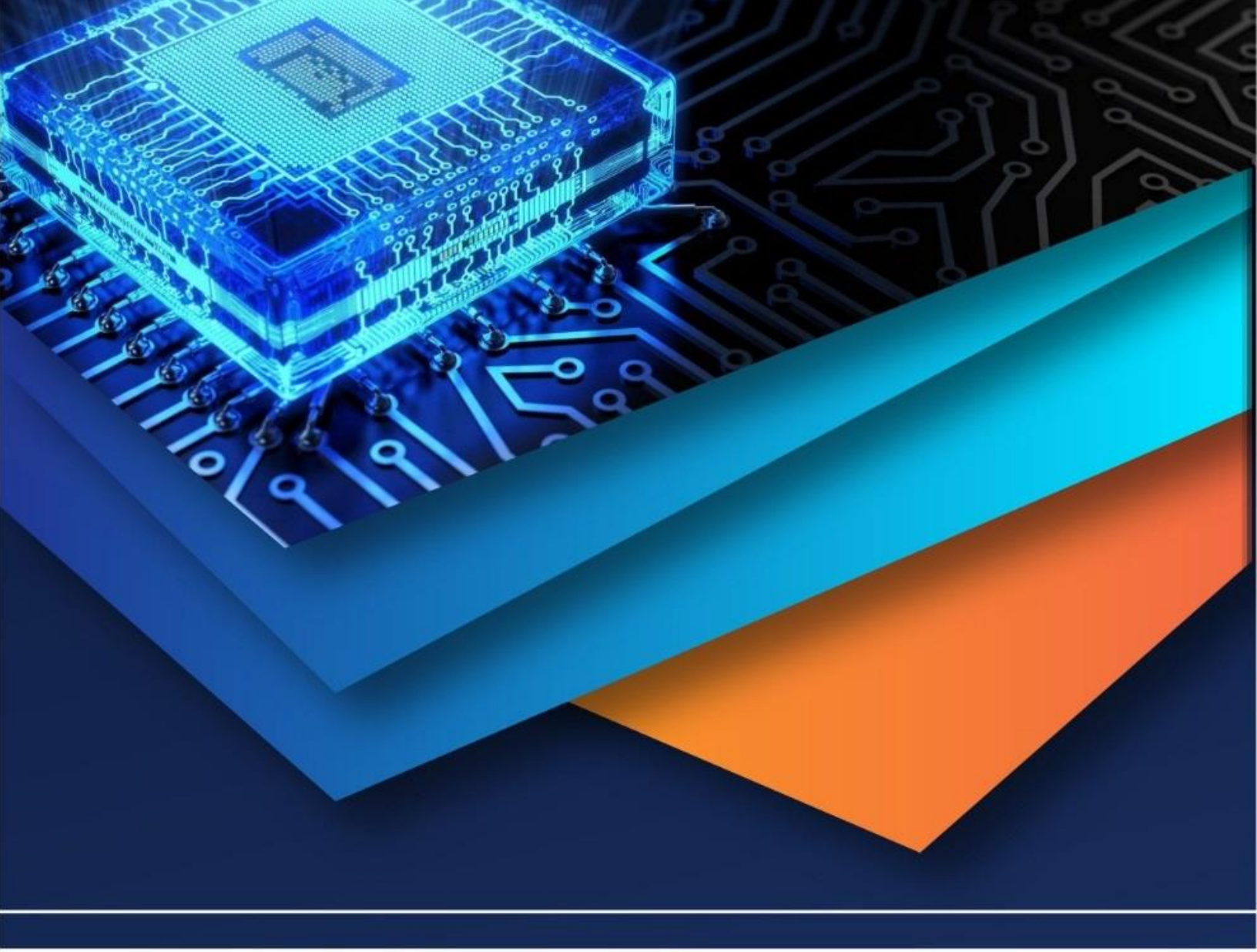
[3] T. Q. Y. Michelle, S. Jarzabek and B. Wadhwa, "CBT Assistant: MHealth App for psychotherapy," *2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite (GHTC-SAS), Trivandrum*

[4] A. Teles et al., "Mobile Mental Health: A Review of Applications for Depression Assistance," *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain*

©IJRASET: All Rights are Reserved

[5] Palak Baid, Apoorva Gupta and Neelam Chaplot, "Sentiment Analysis of Movie Reviews using Machine Learning Techniques", *2017 International Journal of Computer Applications*
[6] Humera Shaziya et al, "Text Categorization of Movie Reviews for Sentiment Analysis", *2015 International Journal of Innovative Research in Science, Engineering and Technology*
[7] Aditya Thyagarajan, "Siamese Recurrent Architectures for Learning Sentence Similarity", *2015 Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*
[8] L. Yao, Z. Pan and H. Ning, "Unlabeled Short Text Similarity With LSTM Encoder", in *IEEE Access, vol. 7, pp. 3430-3437, 2019*, doi: 10.1109/ACCESS.2018.2885698.
[9] (2020) Talklife. [Online]. Available: https://web.talklife.co/
[10] (2019) IMDB Dataset of 50K Movie Reviews. [Online]. Available: https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews/
[11] (2017) Sentiment140 dataset with 1.6 million tweets. [Online]. Available: https://www.kaggle.com/kazanova/sentiment140/
[12] (2019) Hate speech and offensive language twitter dataset. [Online]. Available: https://github.com/t-davidson/hate-speech-and-offensive-language/tree/master/data/
[13] (2020) Word Vectors and Semantic Similarity spaCy. [Online]. Available: https://spacy.io/usage/vectors-similarity/
[14] (2018) BERT Explained: State of the art language model for NLP. [Online]. Available: https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270/
[15] (2017) Quora Question Pairs. [Online]. Available: https://www.kaggle.com/c/quora-question-pairs/
[16] (2020) Stanford Natural Language Inference Corpus. [Online]. Available: https://github.com/MohamadMerchant/SNLI/

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)