Sentimental Analysis on Social Media by using Deep Learning

Prof. Pranali Rahul Dandekar[1] , Prof. Yoginee Surendra Pethe[2]

Assistant Professor

Shri Ramdeobaba College of Engineering & Management, Nagpur

dandekarpr@rknec.edu, pethey@rknec.edu

Abstract

Detection of depression through messages sent by a user on social media are often a fancy task because of the recognition and trends in them. In recent years, messages and social media has over up being a really shut illustration of a person's life and his status. this is often an enormous stockpile of information a couple of person's behaviour and might be used for detection of varied mental sicknesses (depression in our case) victimisation tongue process and Deep Learning. This project is regarding constructing a deep learning model victimisation NLP to predict such mental disorders. short-term memory networks square measure well-suited to classifying, process and creating predictions supported statistic knowledge, since there are often lags of unknown length between necessary events during a statistic.

Keywords— Depression, social media, mental illness, deep learning, NLP

INTRODUCTION

Depression as a typical psychological state disorder has long been denned as one illness with a group of diagnostic criteria. It usually co-occurs with anxiety or alternative psychological and physical disorders; and has a bearing on feelings and behavior of the affected people. in keeping with the World Health Organization study, there ar 322 million folks calculable to suffer from depression, cherish four.4% of the worldwide population. In today's world, communication through social media is rising as an enormous deal. They're willing to share their thoughts, stories and their personal feelings, mental states, needs on social network sites , blogging platforms etc.. Receivers use the manuscripts from emails and alternative varieties of social media comments to make correct reasoning and to correct the mistakes. once folks write digitally on social media, their texts are processed mechanically. linguistic communication process techniques are accustomed infer people's mental behaviour.

According to World Health Organization, depression could be a common worldwide folio that affects a colossal quantity of people no matter their age. There are multiple factors that interfere period of time detection and treatment like lack of skilled specialists, social shaming, improper identification so on . The experiment conducted throughout this work needs the text knowledge that the chosen knowledge supply is Twitter wherever folks tweet regarding their feelings, hopes, desires, thoughts, stories and mental states.

 The goals of our analysis are: collect the in public on the market media messages of healthy and self-diagnosed people that contains mixed emotions therefore measure the extracted Twitter knowledge and apply NLTK and deep learning classifiers like LSTM-RNN to predict depressive and anxiety tweets. we are able to hunt for an answer to a performance increase through a correct options choice and their multiple feature mixtures. First, we decide the foremost helpful linguistic options applied for depression identification to characterize the content of the posts. Second, we

have a tendency to analyse the correlation significance, hidden topics and word frequency extracted from the text. we have a tendency to compare the performance results supported 3 single feature sets and their multiple feature mixtures. In our experiment, we have a tendency to use knowledge collected from the Reedit social media platform.

AIMS AND OBJECTIVE

• Identify the foremost effective deep neural spec among some of elect architectures that were with success utilized in tongue process tasks.

• The architectures square measure wont to notice users with signs of mental diseases (depression in our case) given restricted unstructured text knowledge extracted from the Twitter social media platform.

• To investigate the result of depression detection, we have a tendency to propose Deep learning technique as associate economical and ascendable technique.

• The main contribution of this study lies in exploiting a chic, diverse, and discriminating feature set that contains each tweet text and behavioural trends of various users.

• This study are often extended within the future by considering additional deciliter models that live} extremely unlikely to over-fit the used knowledge and notice a additional dependable thanks to measure the features' impact.

III. LITERATURE SURVEY

Michael M. Tadesse , Hongfei Lin , Bo Xu , And Liang Yang : we can significantly improve performance accuracy. The best single feature is bigram with the SupportVector Machine (SVM) classifier to detect depression with 80% accuracy and 0.80 F1 scores. The strengthand effectiveness of the combined features (LIWC+LDA+bigram) are most successfully demonstrated with the Multilayer Perceptron (MLP) classifier resulting in the top performance for depression detection reaching 91% accuracy and 0.93 F1 scores.

P.V. Rajaraman AsimNath ,Akshaya.P.R, ChaturBhuja.G : Messages and social media has ended up being a very close representation of a person's life and his mental state. This is a huge stockpile of data about a person's behaviour and can be used for detection of various mental illnesses (depression in our case) using Natural Language Processing and Deep Learning.

Akshi Kumara, AditiSharmab, AnshikaArorac : This mixed anxiety-depressive disorder is a predominantly associated with erratic thought process,restlessness and sleeplessness. Based on the linguistic cues and user posting patterns, the feature set is defined using a 5-tuple vector <word, timing, frequency, sentiment, contrast>. An anxiety-relatedlexicon is built to detect the presence of anxiety indicators. Tweet frequency and time analyzed for irregularities and opinion polarity analytics is done to find inconsistencies in posting behaviour. Basically model building with the three classifier such as naives bayes, gradient boosting , and random forest

Michael M. Tadesse ; Hongfei Lin ; Bo Xu ; Liang Yang: We can significantly improve performance accuracy. The best single feature is bigram with the Support Vector Machine (SVM) classifier to detect depression with 80% accuracy and 0.80 F1 scores. The strength and effectiveness of the combined features (LIWC+LDA+bigram) are most successfully demonstrated with the Multilayer Perceptron (MLP) classifier resulting in the top performance for depression

detection reaching 91% accuracy and 0.93 F1 scores.

Hoyun Song, Jinseon You, Jin-Woo Chung Jong C. Park : We propose Feature Attention Network (FAN), inspired by the process of diagnosing depression by an expert who has background knowledge about depression. Experimental results demonstrate that FAN shows good performance with high interpretability despite a smaller number of posts in training data. We investigate different aspects of posts by depressed users through four feature networks built upon psychological studies, which will help researchers to investigate social media posts to find useful evidence for depressive symptoms.

Raza Ul Mustafa, Noman Ashraf, Fahad Shabbir, AhmedJaved Ferzund, , Alexander Gelbukh : A sample of their recent tweets collected ranges from (200 to 3200) tweets per person. From their tweets, we selected 100 most frequently used words using Term Frequency-Inverse Document Frequency (TF-IDF). Later, we used the 14 psychological attributes in Linguistic Inquiry and Word Count (LIWC) to classify these words into emotions. Moreover, weights were assigned to each word from happy to unhappy after classification by LIWC and trained machine learning classifiers to classify the users into three classes of depression High, Medium, and Low. According to our study, better features selections and their combination will help to improve performance and accuracy of classifiers.

Kali Cornn : a dataset of scraped Reddit comments, this project aims to classify depression in comments. Focusing on the setting of social media, this project explores methods of machine learning and neural network architectures for identifying depression in digitally shared text entries. This project developed machine learning (logistic regression, support vector machines), a BERT-based model, and neural networks with and without word embeddings (CNN) for this classification task.

IV. DESIGN AND IMPLEMENTATION

A. Proposed System

Mental illness detection in social media will be thought of a posh task, primarily because of the difficult nature of mental disorders. In recent years, this analysis space has began to evolve with the continual increase in quality of social media platforms that became associate degree integral a part of people's life. This shut relationship between social media platforms and their users has created these platforms to replicate the users' personal life with completely different limitations. In such associate degree surroundings, researchers are conferred with a wealth of knowledge relating to one's life. additionally to the amount of quality in distinguishing mental sicknesses through social media platforms, adopting supervised machine learning approaches like deep neural networks haven't been wide accepted because of the difficulties in getting decent amounts of annotated coaching information. because of these reasons, we tend to attempt to establish the foremost effective deep neural spec among a number of hand-picked architectures that were with success utilized in tongue process tasks. we tend to gift a brand new model NLTK with LSTM. The chosen architectures are accustomed observe users with signs of mental sicknesses (depression in our case) given restricted unstructured text information extracted from the Twitter social media platform.
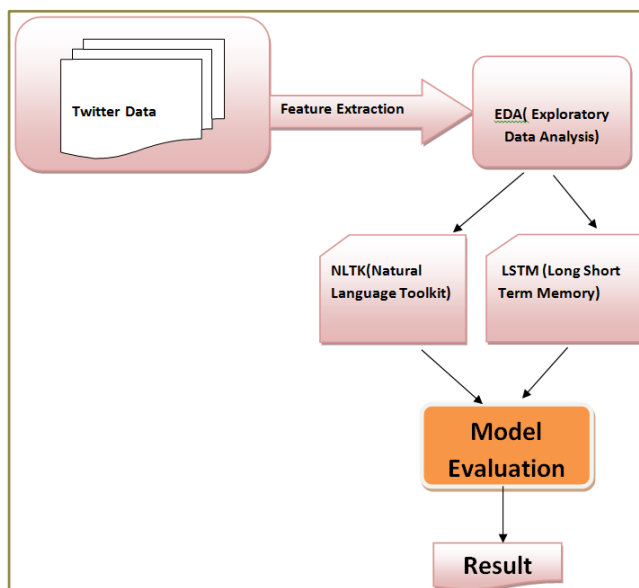
Fig. 1 System flow Diagram

Fig. 2 Data Flow Diagram

Data for the model are extracted from the Twitter web site. information are in unstructured format and it'll have uncountable useless information. thus we'll initial do information improvement. Once we have a tendency to clean the info we'll apply EDA for locating the polarity of the info like positive, neutral, or negative mechanically or several complicated sentiments like happiness, sadness, anger, joy, etc. Once EDA is finished can pass the info to the NLTK model and LSTM model and that we can choose the model giving the simplest accuracy.

Modules

1) information selection: A coaching set is employed to coach the machine learning method to

know the potential relationship between the informative variables and target variable.

2) information Cleaning: information improvement suggests that filtering and modifying your information specified it's easier to explore, understand, and model. Filtering out the elements you do not wish or would like in order that you do not ought to consider or method them.

3) information imputation: Machine learning algorithms need numeric input values, and a worth to be gift for every row and column in a very dataset. As such, it's common to spot missing prices in a very dataset and replace them with a numeric value.

4) information Analysis: a way of knowledge analysis that automates analytical model building. it's a branch of computer science supported the concept that systems will learn from information, determine patterns and create selections with least human intervention.

5) information image: information visualization is that the illustration of {information} or information in a very graph, chart, or alternative visual format. Machine learning makes it easier to conduct analyses like prophetic analysis, which might then function useful visualizations to gift.

6) coaching: Training information is that the information you employ to coach associate degree algorithmic rule or machine learning model to predict the result you style your model to predict.

7) Testing: A take a look at dataset may be a dataset that's freelance of the coaching dataset, however that follows identical likelihood distribution because the coaching dataset.

8) algorithmic rule selection: Machine learning algorithms ar the engines of machine learning, which means it's the algorithms that flip an information set into a model.

B. planned Methodology

1) Linguistic communication Toolkit (NLTK): The linguistic communication Toolkit, or a lot of usually NLTK, may be a suite of libraries and programs for symbolic and applied mathematics linguistic communication process (NLP) for English written within the Python programing language. it had been developed by Steven Bird and Edward Loper within the Department of pc and data Science at the University of Pennsylvania. NLTK includes graphical demonstrations and sample information. it's in the middle of a book that explains the underlying ideas behind the language process tasks supported by the toolkit, and a reference work. NLTK is meant to support analysis and teaching in information science or closely connected areas, together with empirical linguistics, scientific discipline, computer science, info retrieval, and machine learning. NLTK has been used with success as a teaching tool, as a private study tool, and as a platform for prototyping and building analysis systems.

NLTK may be a powerful Python package that gives a collection of numerous natural languages algorithms. It is free, open supply, straightforward to use, massive community, and well documented. NLTK consists of the foremost common algorithms like tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition. NLTK helps the pc to analysis, pre-process, and perceive the written communication.

2) Long Short Term Memory (LSTM): The central role of associate degree LSTM model is control by a memory cell referred to as a 'cell state' that maintains its state over time. The cell

state is that the horizontal line that runs through the highest of the below diagram. It are often envisioned as a transporter through that info simply flows, unchanged. info are often additional to or aloof from the cell state in LSTM and is regulated by gates. These gates optionally let the data flow in and out of the cell. It contains a point-wise multiplication operation and a sigmoid neural internet layer that assist the mechanism.

LSTM stands for long remembering networks, employed in the sphere of Deep Learning. it's a spread of perennial neural networks (RNNs) that are capable of learning semipermanent dependencies, particularly in sequence prediction issues. LSTM has feedback connections, i.e., it's capable of process the whole sequence of knowledge, aside from single information points like pictures. This finds application in speech recognition, AI, etc. LSTM may be a special reasonably RNN, that shows outstanding performance on an outsized style of issues.
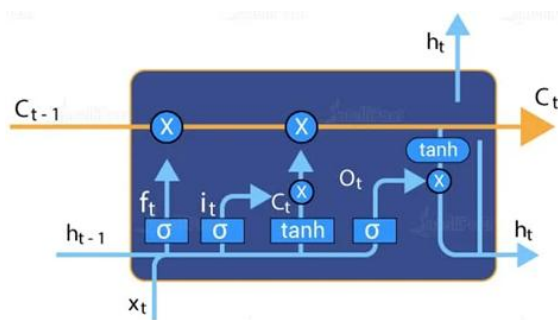

Fig. 3 LSTM Logic

V. CONCLUSION

We have exhibited the capability of using twitter as a tool for measuring and detecting major depression among its users. To give a clear understanding of our work, numbers of research challenges were stated. The algorithms are designed to analyze the tweet for emotion detection as well as for detection of suicidal thoughts among people on social media. The mechanism does analysis of the tweets for prediction of depression without checking the validity of tweets. Social media is an open platform where many people refrain from telling their true emotions that might relate to depression they are facing, and so the model analysis here are mostly based on the prediction from posts using various machine learning algorithms. The main requirement of model is to be perfectly able to predict the result as there are a number of implementations that require verification of data before predicting the thoughts or posts of the person as suicidal or non-suicidal we are developed new model with combination NLTK & LSTM for better Performance result.

REFERENCES

[1] P.V. Rajaraman , Asim Nath ,Akshaya.P.R , Chatur Bhuja.G   Depression Detection of Tweets and A Comparative Test, International Journal of Engineering Research & Technology (IJERT) ,Vol. 9 Issue 03, March-2020

[2] Akshi Kumara, Aditi Sharmab, Anshika Arorac, Anxious Depression Prediction in Real-time Social Data, ICAESMT19, International Conference on Advanced Engineering, Science, Management and Technology – 2019

[3] Michael M. Tadesse , Hongfei Lin , Bo Xu , And Liang Yang, Detection of Depression-

Related Posts in Reddit Social Media Forum, Digital Object Identifier 10.1109/ACCESS.2019.2909180, April 4, 2019

[4] Raza Ul Mustafa, Noman Ashraf, Fahad Shabbir Ahmed, Javed Ferzund,Basit, Shahzad,Alexander Gelbukh, A Multiclass Depression Detection in Social Media Based on Sentiment Analysis, 17th International Conference on Information Technology–New Generations (ITNG 2020) , 12 May 2020

[5] Kali Cornn,Identifying Depression on Social Media, Department of Statistics Stanford University Stanford, CA 94305

[6] Michael Mesfin Tadesse ,ongfei Lin ,Bo Xu, andLiang Yang ,Detection of Suicide Ideation in Social Media Forums Using Deep Learning, State Key Laboratory of Cognitive Intelligence, Algorithms 2020, 13(1), 7,24 December 2019