Complete Research Manuscript

ON

E-RISK Classification using state of art NLP and Deep Learning Techniques

Submitted by

Nipam Nayan Gogoi

PRN: 18030242015

MBA ***(DS & DA)***
(2018-2020)

Acknowledgement

I would like to express my sincere gratitude to Professor Dr. Thomas Mandl, Hildesheim University, Germany for his valuable guidance and support for this project for a span of 5 months.

I would also like thank my mentor Dr Anil Jadhav here at SCIT, Pune for his guidance and support in writing this paper.

Abstract

Online social networks empower people to freely and publicly share and communicate their thoughts and feelings with others. The information on social media is a rich source for analyzing sentiments or inferring mental health problems. Identification of mental illnesses from social media tweets and posts can greatly benefit people through early detection of those conditions and suspected patients will contact physicians and clinicians for further care. This work was undertaken to test the efficacy of the Natural Language Processing techniques and Deep Learning Algorithms by developing an end-to-end model to predict these forms of physiological diseases. The data for the experiments were taken from the annual CLEF eRisk competition, which focuses on early detection of signs of depression or anorexia through social media posts or comments. Early risk prediction is a new area of research potentially applicable to a wide range of situations, such as identifying people with mental illness over social media. In this research we have experimented with the psychological disease Anorexia, which is a mental illness involving body shape and weight perception. Anorexia is still under-diagnosed and under-treated, illustrating the need to expand the reach of current methods of screening. We have used different machine learning techniques and analyzed their performance for early risk prediction of anorexia or depression. The techniques involve various classifiers and feature engineering schemes. The TF-IDF model has been used to perform ada boost, random forest, logistic regression and support vector machine classifiers to identify documents related to anorexia or depression in the individual corpora. The performance of Neural Networks has also been observed using GloVe word embeddings. Glove are pre-trained word vectors developed using specific corpora e.g., Wikipedia. The final model implemented using Multi-channeled Deep Convolutional Networks(MCDCNN) gave an accuracy of 89 percent and a recall of 85 percent. Integrating such a model into a software or mobile application will prove quite beneficial, affordable and handy for people to detect any kind of mental illnesses especially among the youths and students.

INTRODUCTION:

Anorexia is a serious psychiatric condition, characterized by a failure to maintain a minimally normal body weight, extreme fear of weight gain, and changes in the perception of body shape and weight. Anorexia has severe physical side-effects and may be associated with multi-organ system disorders. Anorexia is still under-diagnosed and under-treated, illustrating the need to expand the reach of current methods of screening. 70 million people of all ages suffer from anorexia according to the National Eating Disorder Association, USA. A WHO report calls extreme anorexia one of the world's most burdensome disorders. In fact, anorexia can adversely affect chronic health conditions such as heart disease, cancer, diabetes and obesity. A person with anorexia shows one or more symptoms such as rapid weight loss or being noticeably thin, de-pressed or lethargic etc. Some related e-risk classification related topics are NLP, Transfer Learning, and Word Embedding etc. Natural Language Processing is the technology used to help computers understand the natural language of a human being.

Natural language processing, commonly shortened as NLP, is a branch of artificial intelligence that uses the natural language to deal with the interaction between computers and humans. The ultimate aim of the NLP is to interpret, translate, understand and make sense of human languages in a manner that is useful. Word embedding is the collective name for a group of language modeling and feature learning techniques in the natural language processing (NLP) where vocabulary words or phrases are mapped to real number vectors. This requires, conceptually, a mathematical embedding of a space with many dimensions per term into a continuous vector space with a much lower dimension. Methods to produce this mapping include neural networks, reduction of dimensionality in the word co-occurrence matrix, probabilistic models, explainable knowledge base process, and explicit representation of the sense in which words appear. Word and phrase embedding have been shown to boost performance in NLP tasks such as syntactic parsing and sentiment analysis when used as the underlying representation of inputs. Transfer learning is a form of machine learning, in which a model built for a task is reused as the starting point for a second task model. It is a popular approach in deep learning where pre-trained models are used as the starting point for computer vision and natural language processing tasks due to the vast computational and time resources needed to develop neural network models on these issues and the huge skill leaps they provide on related issues. Specifically, we used an efficient Multi-Channel Deep Convolution Neural Networks (MC-DCNN) model, each of which uses 100-dimensional word embeddings from tweets as input and learns features individually. Instead, the MC-DCNN model combines each channel's learned features and feeds them into a Multilayer Perceptron (MLP) to eventually perform classification. We use gradient-based approach to train our MC-DCNN model to estimate the parameters. On two real-world data sets we test the efficiency of our MC-DCNN model chunk-wise and tweet-wise. The experimental results on both data sets indicate that our MC-DCNN model with significant margins outperforms the baseline methods

The goal is to perform a task to detect anorexia early on the risk. The goal is to analyze pieces of evidence sequentially and to identify early signs of anorexia as soon as possible. The role is primarily concerned with assessing Text Mining approaches and therefore focuses on social media written texts. Texts should be handled in the order in which they were produced. In this way, systems that perform this function efficiently could be extended to track user experiences sequentially in forums, social networks or other online media styles.

The main objective of this research is to correctly classify Anorexic and Non-anorexic persons through the use of Modern NLP techniques by

1) Automating the process to detect mental illness from social media tweets.

2) Using advanced deep learning algorithms to build the text classification models.

THEORETICAL BACKGROUND AND RELATED WORK

Early risk prediction is a new area of research theoretically relevant to a wide range of situations, such as the detection of people with mental illness via social media. Online social networks empower people to freely and publicly share their thoughts and feelings with others and to express them. The knowledge on social media is a rich source for assessing emotions or infering mental health problems. The 2018 challenge of the CLEF E-risk focuses on early detection of mental disorder threats using the social media. The main objective of eRisk 2018 is to facilitate discussion on the development of reliable criteria for the evaluation of early risk detection algorithms by addressing questions of assessment methodology, efficacy metrics and other processes relevant to the production of test collections for early detection of depression. It has planned two tasks this year and released two separate companies for individual tasks, and these corporations are created using posts and comments on Reddit, popular social media. The first task is early risk prediction of depression using Reddit posts and comments. The other task is a pilot task and the aim of the task is to recognize anorexia symptoms by using the given corpus of comments and posts about Reddit. Depression is a common illness that negatively affects emotions, thoughts and behaviors and can impact regular activities such as sleeping. It is a major cause of disabilities and many other illnesses. According to WHO (World Health Organization) figures, depression affects more than 300 million people worldwide and care is given in at least 10 per cent of each population. Low diagnosis and depression treatment can aggravate symptoms of heart failure, precipitate functional regression, interrupt social and occupational functioning and lead to increased mortality risk. Early detection of depression is therefore important. Sadly the rates of diagnosis and treatment of depression are quite small among those with medical illness. To be diagnosed with depression, adequate resources must be in place before identifing depression. In the last few years, a lot of groundbreaking work has been done to explore the potential of social media as a resource for early detection of depression or mental illness. The first task of this challenge is mainly to assess the output of various machine learning frameworks regarding the symptoms of depression for possible knowledge extraction from the given corpus of Reddit posts. A collection of posts over the Reddit of a particular person shall be considered as a single document. The corpus is broken down into preparation and test collection. However, the training set is split into two categories: depression and control category, i.e., non-depression. So 10 chunks of the test set were released with each chunk per week over a period of ten weeks. Each test chunk includes a given person's articles. The goal is to decide whether a particular person's posts within a chunk belong to the category of depression. [1] However, most standard and state- of-the-art supervised machine learning models (such as SVM, MNB, Neural Networks, etc.) are not well suited to deal with this scenario. [2]

Symptoms associated with mental illness can be seen on Twitter, Facebook, and web forums, and automated systems are increasingly capable of detecting depression and other mental illnesses. Automated methods of identification may help identify distressed or otherwise at-risk individuals through large-scale passive monitoring of social media, and may complement current screening procedures in the future. [3]

Clinical characteristics of Anorexia Nervosa (AN) and symptoms and signs of AN can differ depending on the severity of the disorder as well as the strategies used to achieve weight loss. Apart from low weight, it is not unusual for the physical test and laboratory values to be unremarkable. Normal findings do not rule out a diagnosis of eating disorder, and a thorough examination with particular attention to symptoms of eating disorders is important for their identification in primary care settings. A clinician who believes that a patient has AN should inquire about the patient's history of weight, explicitly inquiring about the highest and lowest adult weights, the latest form of weight loss and/or nutritional behaviour, menstrual history, and exercise and eating patterns. Some AN patients may be more willing to endorse physical symptoms such as fatigue, as opposed to emotional issues or eating disorder.
Patients with AN may have amenorrhea and may complain of constipation, tiredness, abdominal pain, cold intolerance and/or excess strength. In addition to emaciation, hypotension, hypothermia, dryness of the skin and lanugo may be noteworthy for physical examination. Many individuals are bradycardic and some may experience peripheral edema, especially during weight recovery or as a result of termination of laxative or diuretic assault. Individuals with AN may also experience salivary gland hypertrophy, especially the parotid glands. Patients with self-induced vomiting can have dental enamel loss and cicatrices or calluses on their dorsum. [4]

Anorexic treatment is to treat a patient with low weight and reduced eating accurately; a physician may consider other mental and medical problems that may cause these symptoms. Psychotic disorders and anxiety disorders may include food avoidance and body hallucinations. Clinical problems may also cause weight loss, including gastrointestinal and endocrine disorders, infections, and neoplastic processes. In particular, patients without eating disorders are generally concerned about their weight loss, while patients with AN are concerned about body image disturbance and are reluctant to gain weight [4].

Current Convolutional Neural Networks (CNNs) typically contain two parts. One is a function extractor, which automatically learns features from the raw data.
And the other is a fully-connected trainable MLP which performs classification based on the previous part's learned features. The function extractor usually consists of multiple similar stages, and each stage consists of three cascading layers: filter layer, activation layer, and pooling layer. Each layer's input and output is called function maps. The function extractor usually contains one, two or three such 3-layer stages in CNN's previous work. Deep learning does not need any hand-crafted features of humans, but can learn the hierarchical feature representation from raw data automatically.
In particular, we propose an efficient Multi-Channels Deep Convolution Neural Networks (MC-DCNN) model, each channel of which takes as input a single dimension of multivariate time series and learns individual features. Then the MC-DCNN model combines the learned features of each channel and feeds them to the Multilayer Perceptron (MLP) to eventually perform the classification. We use gradient-based approach to train our MC-DCNN model to

estimate the parameters. We test our MC-DCNN model performance on two sets of real-world data. The experimental results on both data sets show that our MC-DCNN model with significant margins outperforms the baseline methods and has a strong generalization, particularly for weakly labeled data. [6]

Due to the large number of parameters, the deep neural networks (DNN)-based approaches typically require a large-scale corpus; it is difficult to train a network that generalizes well with limited data. Nonetheless, the costs of building the large-scale infrastructure for certain NLP activities are extremely expensive. Such models often require an unsupervised phase of pre-training to deal with this problem. The final model is fine-tuned with a gradient-based optimization with a supervised training criterion. [7]

Multitask learning uses the association between related tasks to enhance the classification of learning tasks in parallel. Motivated by the success of multi-task learning, some NLP models based on neural networks use multitask learning to learn multiple tasks together with the goal of mutual benefit. Such models ' main multi-task architectures are to share certain lower layers in order to determine common features. The remaining layers are divided up into the various specific tasks after the common layers. [7]

Neural models have the primary role of describing the variable-length text as a fixed-length vector. Such models usually consist of a projection layer that maps words, sub-word units, or n-grams to vector representations (often previously trained with unsupervised methods), and then combines them with the different neural network architectures.
There are several types of text processing models, such as the Neural Bag-of-Words (NBOW) model, recurrent neural network (RNN), the neural network that is recursive. Such models take the embedding of words in the text sequence as their input, and summarize its context with a vector representation of a fixed length. Recurring neural networks (RNN) are amongst them one of the most common architectures used in NLP problems because of their recurring structure. [7]

Detection of depression is a major public health concern, as almost 12% of all disorders can be linked to depression. Computational models for the diagnosis of depression need to show not only that they can diagnose depression, but that they can do it early enough to make it feasible for an intervention. Current Depression Detection tests, however, are low in assessing model latency. Latency Weighted F1 metric is among the common metrics used to detect these problems. [8]

ABOUT THE DATA

The Data is taken from the competition organized by the CLEF 2019 Workshop in the project. It is an annual international competition' eRisk 2019: Web early risk prediction.'

ERisk examines the methods, effectiveness metrics and practical applications of early risk detection on the Internet (particularly those related to health and safety). Early detection technologies, especially those related to health and safety, can be employed in different areas. For example, early notices could be sent when an abuser begins to communicate with a child for sexual purposes, or when a possible offender starts making antisocial comments on a blog, website or social network. The main objective of CLEF is to develop a new interdisciplinary research area that could theoretically be extended to a wide range of situations and to many different personal profiles. Examples include suspected pedophiles, stalkers, persons who may fall into the hands of criminal organizations, people with suicidal tendencies or people vulnerable to depression.
There are three tasks in the competition:-Task 1: early detection of anorexia signs; Task 2: early detection of self-harm signs; and Task 3: assessment of the extent of depression signs [8]

The Data consisted of Reddit tweets from 152 users in the training set. Among them 20 were termed as Anorexic (positive) whereas 132 users were termed as Non-Anorexic (negative). Both the positive and negative data were divided into 10 chunks. Each chunk contained the data in XML files. Each XML file inside a chunk consisted of a number of tweets from a particular user.
The Test Datasets were also in 10 chunks. It contained the data from 320 users.
The Golden Truth file was provided for both the training and test datasets. There were 41 users among 320 users who were anorexic in the Test Chunks.

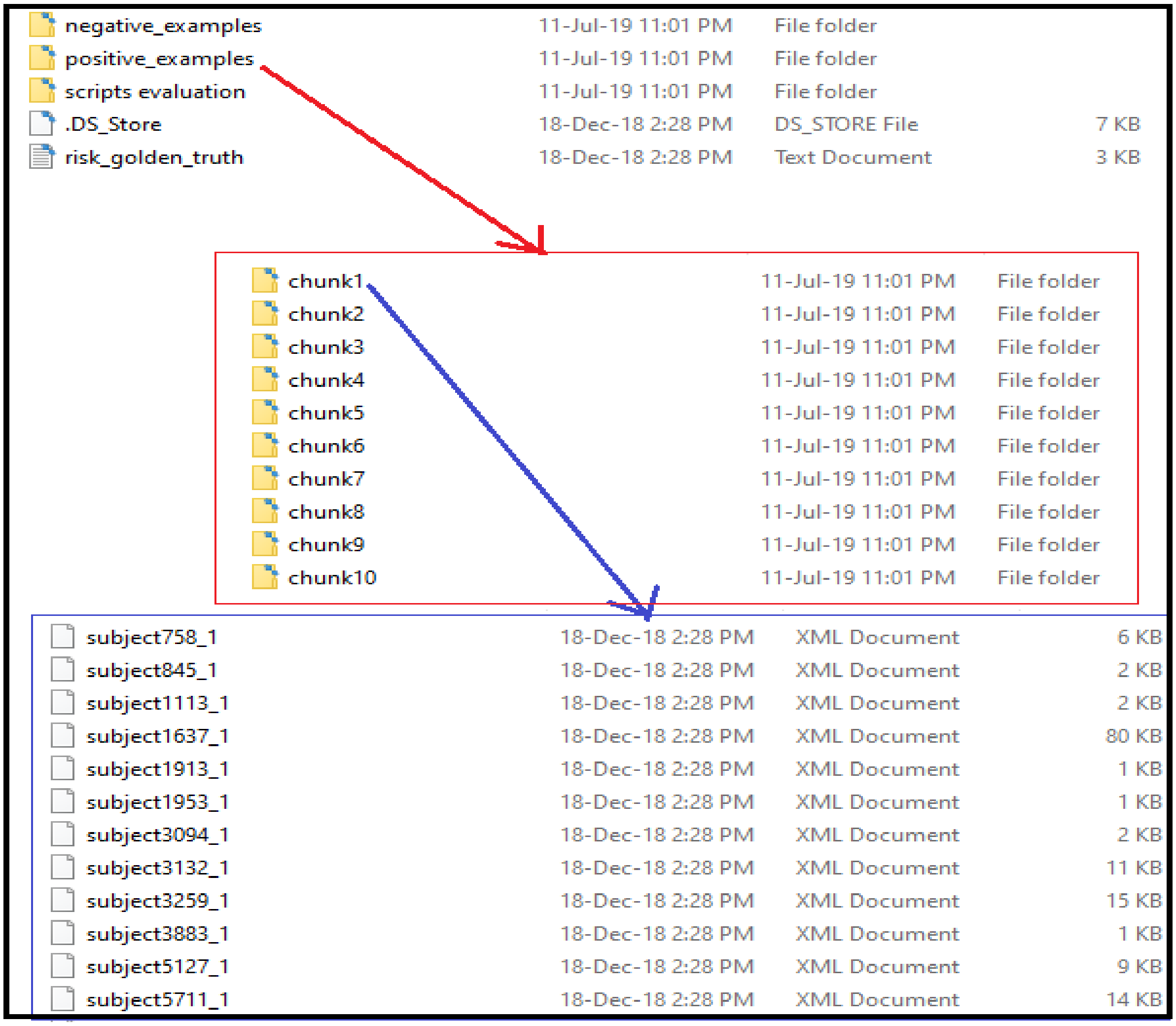


Fig: Data inside Trainset

- <INDIVIDUAL>
<ID>subject758</ID>
- <WRITING>
<TITLE> </TITLE>
<DATE> 2013-10-09 02:26:19 </DATE>
<INFO> reddit post </INFO>
<TEXT> I googled her and ha, that's not me..but interesting. </TEXT>
</WRITING>
- <WRITING>
<TITLE> </TITLE>
<DATE> 2013-10-08 15:08:59 </DATE>
<INFO> reddit post </INFO>
<TEXT> in terms of book genres- fiction, surrealism/realism, horror, suspense. It is somewhat heavy, but it's definitely sort of relieving to read, just in my perspective haha. To each their own I guess! in terms of music I've pretty much been listening to the smiths or girly stuff like Grimes and Crystal Castles.. and a bunch of classical/acoustic/instrumentals from my favorite movies, it's really cliche.. I know. Thanks for distracting me by the way.. It's helping me not think too much about things. </TEXT>
</WRITING>
- <WRITING>
<TITLE> </TITLE>
<DATE> 2013-10-08 14:45:07 </DATE>
<INFO> reddit post </INFO>
<TEXT> Breaking Bad, I suppose. I read a lot of Murakami and I've recently reread Invisible Monsters by Palahniuk. I guess it's pretty small stuff like this that keeps me alive. Either that, or something bad. </TEXT>
</WRITING>
- <WRITING>
<TITLE> I'm a 19 (F) and I guess this is my story. Death seems like a great solution. </TITLE>
<DATE> 2013-10-08 14:07:15 </DATE>
<INFO> reddit post </INFO>
<TEXT> I'm a 19 year old girl. I'm mentally ill. I don't think I've ever fully typed out my life story on a website, or written it down anywhere before, so prepare for a TLDR or skip to the bottom. Ever since I was a young child, domestic abuse was a common thing growing up. My father was abusive towards my family and I, and we had to put it up with it almost every single day. Everyone suffered. My sisters, my mother, even my own relatives suspected what was going on but no one intervened. At age 12, I was molested by my first therapist/counselor at school, I kept it a secret until my early teens. Bullying seemed of the norm or common to me, so I put up with that as well as my grades suffered. I was in denial of what happened for a very long time and I During high school at age 15, I experienced a traumatic experience from an incident with my dad and I had to leave the house the move in with my aunt. CPS and the police were involved, I had to go to court a lot, it was horrible. I developed my eating disorder at this age and I ended up dropping out of school. I was not missed. From then on, everything got worse. Suicide/OD attempts, jumping into traffic, self-harm, binging and purging- I hated myself. I experienced a lot of shitty traumatic events. I

Fig: A sample XML file consisting of multiple tweets (Only text were used from each file)

CONCEPTUAL MODELS / FRAMEWORKS

- **Tools and Technologies used:**

1) Thorough analysis and modeling of data using the Python language.
2) Jupyter Notebook
3) Microsoft Excel

- **Modules used in Python:**

1) nltk
2) numpy
3) pandas
4) TensorFlow
5) pickle
6) sys
7) StringIO
8) Xml.etree.ElementTree
9) Sklearn
10) Sys
11) Matplotlib
12) os

13) gensim

- **Machine Learning and Statistical Techniques used:**

1) Logistic Regression
2) Support Vector Machine
3) Random Forest
4) AdaBoost
5) Naïve Bayes
6) Word Embeddings Glove
7) MCDCNN (Multi Channel Deep Convolutional Neural Networks)
8) Confusion Matrix
9) Recall
10) ROC

RESEARCH DESIGN/METHOD

1) Initially various codes available in Github which were related to the tasks were implemented for learning purposes. [yiuwin][10]
2) After learning various concepts for text classification, efforts were undertaken for doing everything from scratch.
3) **Preprocessing Phase 1:** The Data was already in a very complex format. Two approaches were undertaken to convert the data into a simpler MS Excel format as in Dataframes.
   a) Preprocessing User-wise: Here all the texts from each tweets from every XML file was collected. After that all texts were arranged user-wise. Moreover the texts for each user from all the chunks were combined. So, finally for each user the data was from 10 chunks in the final dataframe. The final dataframe consisted of three columns: **Text, Label and ID.**
   In case of the Train Dataset, the data taken from the positive_examples file were labeled as 1 which meant anorexic and the negative_examples were labeled as 0 which denoted non-anorexic.
   For the Test Dataset for each user the the Label(anorexic or non-anorexic) was collected from a text file named Golden Truth. So the final dataframe for testing also contained three columns: **Text, Label and ID.**
   The train dataset contained 152 rows and the test dataset consisted of 320 rows.
   The datasets were saved as excel file to train the models.

   b) Preprocessing Chunk-wise: Here all the texts from each XML file was collected from each chunks and arranged user-wise except the texts from all the 10 chunks were not combined for each user. Hence each row in both train and test dataset consisted of three columns: **Text, Label and ID.** But the text column contained tweets from 1 chunk only for each user. Here also the Golden truth file was used for filling up the Labels column.

The train dataset contained 1520 rows and the test dataset consisted of 3200 rows. The datasets were saved as excel file to train the models.

	Text	Label	Id
0	I'm 19, almost 20. I use they/them pronoun...	1	subject1113
1	Any tips to improve essay writing speed? I'm...	1	subject1637
2	Adorable little derpface! :D I love him, h...	1	subject1913
3	Hi! Maybe you could wrap them in clot...	1	subject1953
4	I tried to pull off the top two bananas... ...	1	subject3094
5	Nicky because she is just downright a good...	1	subject3132
6	Agreed What's one thing someone should do...	1	subject3259
7	Do you hallucinate?	1	subject3883
8	I've got the same situation. I was on abil...	1	subject5127
9	Yes cats make people fat because dog peopl...	1	subject5711

Fig: Preprocessed form of the datasets

4) All these tasks were done using python modules 'os' and 'xml.etree.ElementTree'.
5) Initially simpler Machine Learning Models were applied using the python module Sklearn.
6) **Preprocessing Phase 2**: In the entire project the text were preprocessed for ML models using two methodologies: TF-IDF (Term frequency Inverse Document Frequency) and Glove Word Embeddings.

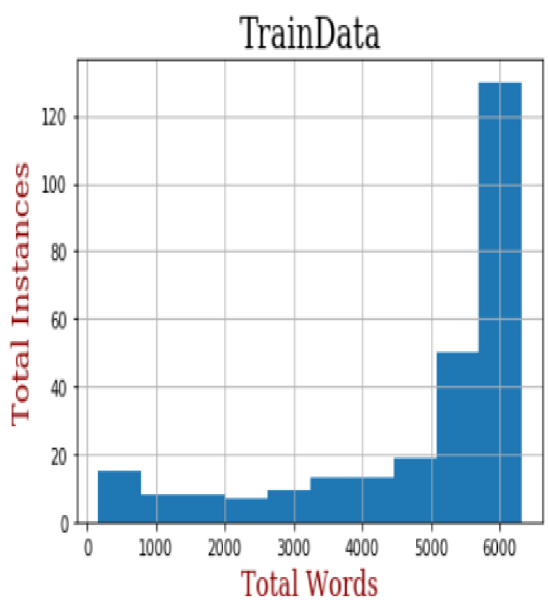
   Word Embeddings was implemented in the final Model and TF-IDF was applied only in the initial phase of the project for simpler ML Models in Sklearn.
   For applying Word embeddings all the other characters except alphabets and numbers were removed from the text in each row.
   After that the text was converted into lower case.
   Moreover Lemmatization was also done to the text after tokenization.
   But Stopwords were not removed in case of word embeddings in order to improve its performance.
7) From the sklearn module ML models such as logistic regression, random forest, adaboost etc were implemented after the texts from the datasets were converted into TF-IDF.
8) Metrics such as Confusion Matrix, Recall, Precision, Accuracy score were used for determining the performance of the Models. Among these Models the best performance was depicted by Logistic Regression.
9) All the models were tested for two scenarios:-
   a) User-Wise
   b) Chunk-wise
10) The User-wise implementation is described below.

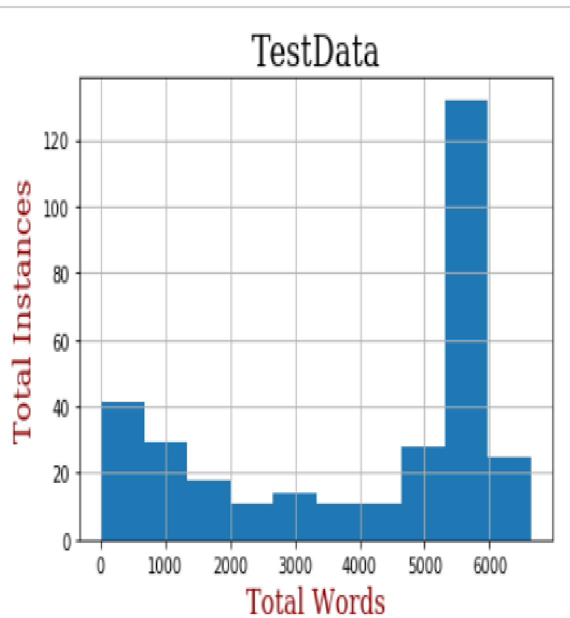
11) The train dataset was very unbalanced consisting of 152 users where only 20 were labeled as 1(anorexic) and 132 were non-anorexic, Oversampling of the rows with label =1 was done by repeating it 7 times. Hence in the final trainset there were 132 non-anorexic and 140 anorexic subjects. However the test dataset was left as it is. It was also unbalanced with only 41 anorexic subjects and 272 non-anorexic subjects.

12) After the preprocessing phases the pretrained Glove Vectors were loaded for the text data. Glove vectors with 100 dimensions which were pre-trained on 6 billion words were used in the project.

Further scope is there to test the models by using other different pretrained Glove vectors such as 27 billion words, 300 dimensions etc.

- **Information Of The Train And Test Datasets**





count	272.000000
mean	4828.805147
std	1677.848533
min	163.000000
25%	4240.750000
50%	5692.000000
75%	5915.000000
max	6320.000000

count	320.000000
mean	3987.731250
std	2156.238002
min	10.000000
25%	1627.500000
50%	5290.500000
75%	5719.500000
max	6642.000000

Fig: Information about train and test dataset

13) Each text row in the dataframes consisted of around 4000-6000 words. While loading the word vectors each word had a 100 dimensional representation as numbers. Hence three

100 dimensional matrices were formed from the text of each row. They are:- Average word vectors, Maximum and Minimum word vectors from each row.

14) Hence after loading the word vectors we have obtained three matrices of size 100*272 from the train dataset and 100*320 for the test dataset.
15) Initially sklearn machine learning models were implemented on matrices obtained using either of the average matrix or minimum matrix or maximum matrix. Satisfactory results were obtained from these models also.
16) One of the main emphasis in this project was to implement deep learning models on the datasets using Tensorflow. It was inspired by a research paper **"Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks"** which implemented multi-channeled deep convolutional networks for Time Series Classification.[5]
17) For the purpose initially the three matrices: average, maximum and minimum word vectors were combined to form a 3D matrix just like the RGV layers of an Image.
18) After that one convolutional layer and one pooling layer is used to extract the features from the matrices.
19) After feature extraction it is feed as input through two fully connected layers to finally predict the output as anorexic or non-anorexic.
20) **The Architecture is described in detail below:**

    In contrast to image classification, the inputs of multichannel text classification are multiple 1D subsequence but not 2D image pixels. The traditional CNN is modified and applied to multichannel text classification task in this way: The 3D matrix is separated into three matrices and feature learning on each series is performed individually. Then we concatenate a normal Multi Layer Perceptron (MLP) at the end of feature learning to do classification. To be understood easily, we illustrate the architecture of MC-DCNN in the figure below. Specifically this is an example of 2-stages MC-DCNN for text classification. It includes 3 channels inputs and the length of each input is 100. For each channel, the input (i.e., the word vector matrix) is fed into a 2-stages feature extractor, which learns hierarchical features through filter, activation and pooling layers. At the end of feature extractor, we flatten the feature maps of each channel and combine them as the input of subsequent MLP for classification. Note that in Figure, the activation layer is embedded into filter layer in the form of non-linear operation on each feature map.

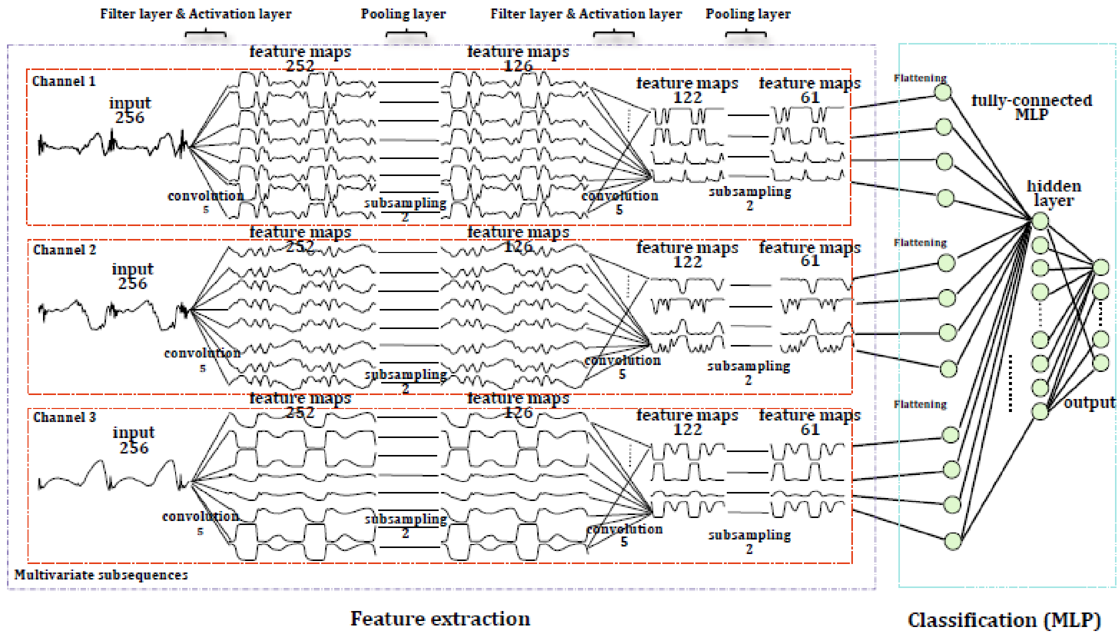


Fig.(Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks): A 2-stages MC-DCNN architecture for text classification. This architecture consists of 3 channels input, 1 filter layer, 1 pooling layer and 2 fully-connected layers.This architecture is denoted as 30(10)-2-128-128-2 based on the template C(Size)-S-H1-H2-O, where C is the number of filters in the convolutional stage,Size denotes the kernel size, S is the subsampling factor, H and O denote the numbers of units in hidden and output layers of MLP.

21) Relu is the activation function used in all the stages and in the pooling layer Max Pooling is used.
22) To estimate parameters of models, gradient based optimization method is utilized to minimize the loss function. Specifically, simple backpropagation algorithm is used to train our MC-DCNN model, since it is efficient and most widely used in neural networks. Stochastic gradient descent (SGD) is adopted instead of full-batch version to update the parameters .Because SGD could converge faster than full-batch for large scale data sets. A full cycle of parameter updating procedure includes three cascaded phases: feedforward pass, backpropagation pass and the gradient applied.
23) To converge fast, decay and momentum strategies are also utilized. The values: momentum = 0.9, decay = 0.0005 and epsilon = 0.01 are set for the experiments.
24) All the above models were applied for the chunk-wise train and test dataset also.

Model Parameters for MCDCNN

```
input_height = 1
input_width = 100
num_labels = 2
num_channels = 3

batch_size = 8
kernel_size = 10
nbr_kernels = 30
num_hidden = 128

learning_rate = 0.001
momentum=0.9
decay_rate = 0.0005
epsilon = 0.01
training_epochs = 30
```

Fig: User-wise Model

```
input_height = 1
input_width = 100
num_labels = 2
num_channels = 3

batch_size = 20
kernel_size = 5
nbr_kernels = 30
num_hidden = 128

learning_rate = 0.001
momentum=0.9
decay_rate = 0.0005
epsilon = 0.01
training_epochs = 30
```

Fig: Chunk-wise Model

RESULTS AND DISCUSSION

- **Observations:**

1. The Logistic Regression was the most consistent model in all the approaches.
2. The TF-IDF models performed better user-wise but in Chunk-wise approaches Word Embedding models performed the Best.
3. However the final model implemented using MCDCNN outperformed all the other models in terms of Anorexia Prediction.
4. More focus was made on Recall because the Test data was a very unbalanced dataset with only 41 subjects as Anorexic out of total 320 test subjects(User-wise approach). Hence precision was expected to be low for all instances. However some of the models had good precision too.

- **Results:**

	TF-IDF TESTED USING SKLEARN MACHINE LEARNING MODELS								
	USER-WISE					CHUNK-WISE			
320 TEST ROWS	Precision	Recall	Accuracy	Total Anorexia Prediction out of 41	3200 TEST ROWS	Precision	Recall	Accuracy	Total Anorexia Prediction out of 410
Logistic Regression	0.9	0.63	0.94	26	Logistic Regression	0.81	0.6	0.93	248
Random Forest	0.9	0.22	0.9	9	Random Forest	0.46	0.37	0.86	150
Adaboost	0.94	0.41	0.92	17	Adaboost	0.86	0.22	0.9	90
Sklearn Neural Network	0	0	0.87	0	Sklearn Neural Network	0.91	0.07	0.88	29
SVM	0.6	0.8	0.91	33	SVM	0.47	0.64	0.86	262

Fig: Results using TF-IDF

	WORD EMBEDDING(GLOVE) MODELS								
	USER-WISE					CHUNK-WISE			
320 TEST ROWS	Precision	Recall	Accuracy	Total Anorexia Prediction out of 41	3200 TEST ROWS	Precision	Recall	Accuracy	Total Anorexia Prediction out of 410
X avg (Log Regression)	0.52	0.78	0.88	32	X avg (Log Regression)	0.46	0.76	0.85	313
X min (Log Regression)	0.48	0.61	0.87	25	X min (Log Regression)	0.32	0.67	0.78	273
X max (Log Regression)	0.56	0.59	0.89	24	X max (Log Regression)	0.32	0.61	0.78	252
X avg (SVM)	0.58	0.8	0.9	33	X avg (SVM)	0.46	0.74	0.86	303
MCDCNN (Final Model)	0.54	0.85	0.89	35	MCDCNN (Final Model)	0.35	0.78	0.78	319

Fig: Results using Word Embeddings

- **Limitations and Future research directions:**

1. Better results could be obtained by using pre-trained Glove Vectors from Reddit data. Moreover the Glove pre-trained vectors trained from 6 billion words (6B) was used in the final model even though 27B, 42B and 840B vectors were available for use. This was done due to system limitations.

2. Increasing the Dimensions of word vectors from 100 to 300 is another approach to improve the predictions. Moreover bag of words approach was implemented in the project which means average vector, min or max vectors of all the words tweeted by a user was used. Another approach could have been taking all the word vectors as input for prediction.

3. Doc2Vec could also have been used for the project which performs better than word vectors in classification according to various research papers.

4. The Chunk based approach could be further improved in the project by taking 2 chunks per user or 3 chunks per user and so on up to 10 chunks. After that an analysis could be carried out to find the optimal number of chunks required per user for better prediction of Anorexia. This could lead to the optimal number of tweets required to find out whether a person is anorexic or not.

- **Application in terms of Business context :**

The methodology and NLP techniques adopted in this project can be easily transferred to perform other tasks related to business such as detection of Customer Attrition, Customer Retention, classification of Customer Complaints and Queries etc and all other such types of tasks wherever text data is available. This will save an organization from manual errors and also save time and increase efficiency of their business activities in the long run. Moreover this model can be easily integrated into an android or IOS app which can be quite handy in detecting mental illnesses from Social media posts.

CONCLUSION

In this paper we propose an Anorexia detection software or application scheme based on NLP and Deep learning, aiming at exploring the feasibility of implementing it in a real life Social Media Analytics scenario. Based on the data set containing tweets of 152 subjects in the training data and 320 test subjects, an end-to-end classification model trained by the MC-DCNN method is developed for Anorexia prediction and detection. All the tweets were transformed into GloVe word vectors during preprocessing. The method we proposed also demonstrated the effectiveness of Deep Neural convolutional layers on Text Data. The experiment results in this study show this approach achieves high sensitivity and high prediction accuracy on the test dataset. In the future work, we will focus on improving algorithm efficiency and precision and also use GloVe Word vectors with higher dimensions and larger number of Word Vectors. Doc2vec is also a feasible option to experiment in future for the same task. Moreover these techniques are easily transferable to detect other mental diseases such as Depression, Bipolar disorder etc.

BIBLIOGRAPHY / REFERENCES

1. Sayanta Paul, Jandhyala Sree Kalyani and Tanmay Basu, Early Detection of Signs of Anorexia and Depression Over Social Media using Effective Machine Learning Frameworks.

2. Sergio G. Burdisso, Marcelo Errecalde, Manuel Montes-y-Gómez , A text classification framework for simple and effective early depression detection over social media streams

3. Sharath Chandra Guntuku, David B Yaden, Detecting depression and mental illness on social media: an integrative review

4. Danila Musante, BA, and Evelyn Attia, MD, Anorexia Nervosa: Role of the Primary Care Physician

5. Johannes C. Eichstaedt, Robert J. Smith, Facebook language predicts depression in medical records

6. Yi Zheng, Qi Liu, Enhong Chen, Yong Ge and J.Leon Zhao, Time Series Classification Using Multi-ChannelsDeep Convolutional Neural Networks.

7. Pengfei Liu, Xipeng Qiu, Xuanjing Huang, Recurrent Neural Network for Text Classification with Multi-Task Learning

8. Farig Sadeque, Dongfang Xu, Steven Bethard, Measuring the Latency of Depression Detection in Social Media

9. Patrice Bellot, Nicola Ferro, Report on CLEF 2018: Experimental IR Meets Multilinguality, Multimodality, and Interaction

10. Yiuwin, eRisk_Anorexia. Available from: https://github.com/yiuwin/eRisk_Anorexia