STUDY OF TWEET CLASSIFICATION FOR DISASTER MANAGEMENT USING LSTM and ANN

Team Members :

Ravipati Bhanu Chandu - 700728966

Pavan kalyan rao yelamati - 700729168

Hari Chandana Gaddampally - 700726882

Abstract:

Microblogging stages, for example, Twitter give dynamic correspondence channels during mass intermingling and crisis occasions such as quakes, tropical storms. During the abrupt beginning of an emergency circumstance, influenced individuals post helpful data on Twitter that can be utilized for situational mindfulness and other compassionate fiasco reaction endeavors, whenever prepared auspicious and successfully. Preparing social media data represent numerous difficulties, for example, parsing uproarious, brief, and casual messages, taking in data classifications from the approaching stream of messages and arranging them into various classes among others. One of the fundamental necessities of huge numbers of these assignments is the accessibility of information, specifically human-commented on information. In this paper, we present human-explained Twitter corpora gathered during 19 distinct emergencies that occurred somewhere in the range of 2013 and 2015. To show the utility of the explanations, we train machine learning classifiers. In addition, we distribute first biggest word2vec word embedding prepared on 52 million emergency related tweets. To bargain with tweets language issues, we present human-explained standardized lexical assets for various lexical varieties.

Emergency informatics center around the commitment of client created content (UGC) to debacle the board. To use the web-based life information successfully, it is significant to filter out boisterous data from the huge volume of information flow so we could more readily evaluate fiasco harm with this information. Not satisfied with essential watchword based filtration; numerous specialists go to AI for arrangement. In this task, I apply profound learning strategies to address Tweets classification issue in catastrophe the executives' field. The names of Tweets reflect various sorts of fiasco related data, which have diverse potential utilization in crisis reaction. Specifically, ANN is utilized for move learning. The standard ANN design for classification and a few other ANN structures are prepared to contrast and the gauge gradient descent algorithm with pretrained Glove Twitter embeddings. Results show that ANN and ANN-based LSTM accomplish the best outcomes, beating the standard model by 3.29% by and large regarding F-1 score individually. Equivocalness and subjectivity influence the

presentation of these models impressively. In certain models the models can outperform human execution.

Keywords: Natural language processing, Twitter, Disaster response, Classification, LSTM & ANN.

Introduction:

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

Machine learning tasks are classified into several broad categories. In supervised learning, the algorithm builds a mathematical model from a set of data that contains both the inputs and the desired outputs. For example, if the task were determining whether an image contained a certain object, the training data for a supervised learning algorithm would include images with and without that object (the input), and each image would have a label (the output) designating whether it contained the object. In special cases, the input may be only partially available, or restricted to special feedback. Semi-supervised learning algorithms develop mathematical models from incomplete training data, where a portion of the sample input doesn't have labels.

Classification algorithms and regression algorithms are types of supervised learning. Classification algorithms are used when the outputs are restricted to a limited set of values. For a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email. For an algorithm that identifies spam emails, the output would be the prediction of either "spam" or "not spam", represented by the Boolean values true and false. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. Examples of a continuous value are the temperature, length, or price of an object.

In unsupervised learning, the algorithm builds a mathematical model from a set of data which contains only inputs and no desired output labels. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories, as in feature learning. Dimensionality reduction is the process of reducing the number of "features", or inputs, in a set of data.

Reinforcement learning algorithms are given feedback in the form of positive or negative reinforcement in a dynamic environment, and are used in autonomous vehicles or in learning to play a game against a human opponent. Other specialized algorithms in machine learning include topic modelling, where the computer program is given a set of natural language documents and finds other documents that cover similar topics. Machine learning algorithms can be used to find the unobservable probability density function in density estimation problems. Meta learning algorithms learn their own inductive bias based on previous experience. In developmental robotics, robot learning algorithms generate their own sequences of learning experiences, also known as a curriculum, to cumulatively acquire new skills through self-guided exploration and social interaction with humans. These robots use guidance mechanisms such as active learning, maturation, motor synergies, and imitation.

ML is the logical investigation of calculations and factual models that PC frameworks used to play out a particular undertaking without utilizing unequivocal guidelines, depending on examples and derivation. It is viewed as a subset of man-made consciousness. AI calculations assemble a numerical model dependent on test information, known as "preparing information", so as to settle on forecasts or choices without being unequivocally customized to play out the undertaking. AI calculations are utilized in a wide assortment of uses, for example, email sifting and PC vision, where it is troublesome or infeasible to build up a traditional calculation for successfully playing out the errand.

LSTM

Long short-term memory (LSTM) is an artificial Recurrent Neural Network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition], speech recognition and anomaly detection in network traffic or IDS's (intrusion detection systems).

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

ANN (Artificial Neural Network)

A neural network is a method in artificial intelligence that teaches computers to process data in a way that is inspired by the human brain. It is a type of machine learning process, called deep learning, that uses interconnected nodes or neurons in a layered structure that resembles the human brain.

Neural networks can help computers make intelligent decisions with limited human assistance. This is because they can learn and model the relationships between input and output data that are nonlinear and complex.

Motivation:

Twitter give dynamic correspondence channels during mass intermingling and crisis occasions such as quakes, tropical storms. During the abrupt beginning of an emergency circumstance, influenced individuals post helpful data on Twitter that can be utilized for situational mindfulness and other compassionate fiasco reaction endeavors, whenever prepared auspicious and successfully. Preparing social media data represent numerous difficulties, for example, parsing uproarious, brief, and casual messages, taking in data classifications from the approaching stream of messages and arranging them into various classes among others. One of the fundamental necessities of huge numbers of these assignments is the accessibility of information, specifically human commented on information. In this paper, we present human-explained Twitter corpora gathered during 19 distinct emergencies that occurred somewhere in the range of 2013 and 2015. To show the utility of the explanations, we train machine learning classifiers. In addition, we distribute first biggest word2vec word embedding prepared on 52 million emergency related tweets. To bargain with tweets language issues, we present human-explained standardized lexical assets for various lexical varieties.

Significance & Objectives:

It is significant to filter out boisterous data from the huge volume of information flow so we could more readily evaluate fiasco harm with this information. Not satisfied with essential watchword-based filtration; numerous specialists go to AI for arrangement. In this task, I apply profound learning strategies to address Tweets classification issue in catastrophe the executives' field. The names of Tweets reflect various sorts of fiasco related data, which have diverse potential utilization in crisis reaction. Specifically, LSTM is utilized for move learning. The standard LSTM design for classification and a few other LSTM structures are prepared to contrast and the gradient descent algorithm with pretrained Glove Twitter embeddings. Results show that LSTM and Bi-LSTM accomplish the best outcomes, beating the standard model by 3.29% by and large regarding F-1 score individually. Equivocalness and subjectivity influence the presentation of these models impressively. In certain models the models can outperform human execution.

Contribution:

Work	Contributed by
Text processing	Pavan & Bhanu
Evaluation Metrics(Accuracy, Precision, recall, F1-score)	Pavan
Code Implementation(Preprocessing, Fit Evaluate, Predict)	Bhanu
Identification of candidate OOV words	Chandana
Statistical Reports	Chandana

Related Work:

Text Processing:

Writings are lowercased. Non-ascii letters, URLs, @RT: [NAME], @[NAME] are expelled. For LSTM and ANN, an extra [CLS] token is embedded to the start of every content. Writings with length under 4 are discarded. No lemmatization is performed, and no accentuation imprint is evacuated since pre-prepared embeddings are constantly utilized. No stop-word is expelled for fluency reason. The gauge model is a solitary layer gradient descent algorithm neural system with shrouded size 256. The model gets preprepared Glove Twitter 27B embeddings (200d) as info. The stacked final concealed condition of the succession is connected to a completely associated layer to perform SoftMax. We trained all three different kinds of classifiers using the pre-processed data. For the evaluation of the trained models, we used 10-folds cross-validation technique. Table 2 shows the results of the classification task in terms of Area under ROC curve4 for all classes of the 8 different disaster datasets. We also show the proportion of each class in each dataset. Given the complexity of the task i.e., multiclass classification of short messages, we can see that all three classifiers have pretty decedent results. In this case, a random classifier represents an AUC = 0.50 and higher values are preferable. Other than the “missing trapped or found people” class, which is the smallest class in term of proportion across all the datasets, results for most of the other classes are at the acceptable level (i.e. ≥ 0.80).

Identification of candidate OOV words

To identify candidate OOV words that require normalization, we first build initial vocabularies consisting of lexical variations mentioned in the previous section. We use a dictionary available on the web to normalize abbreviations, chat shortcuts, and slang.5 We also use the SCOWL (Spell Checker Oriented Word Lists) a spell English dictionary 6 that consists of 349,554 English words. The SCOWL dictionary is suitable for English spell checkers for most of English dialects. Although, the SCOWL dictionary contains places names (e.g., names of countries and famous cities), after testing it on Nepal Earthquake data, we found that its coverage is not complete and a large number of cities/towns of Nepal are missing.

Evaluation metrics

Classification Accuracy: The proportion of the total number of predictions that were correct.

F1 Score: F1-Score is the harmonic mean of precision and recall values for a classification problem.

Recall: The proportion of actual positive cases which are correctly identified.

Precision: The proportion of positive cases that were correctly identified

Proposed framework:

Current writing shows the utilization of internet based life, for example, Twitter, Facebook and YouTube for curating, examining what's more, abridging emergency related data so as to settle on choices and reactions (Imran, Castillo, Lucas, et al. 2014; Vieweg, Hughes, et al. 2010; Imran, Castillo, Diaz, et al. 2015; Terpstra et al. 2012; Tsou et al. 2017). Among webbased life examines, most of them center around Twitter, essentially considering its practicality and accessibility of data from an enormous client base. In (Hagen et al. 2017), the creators broke down Twitter system structure to comprehend the stream of data and how various entertainers and networks contribute towards compelling points. Avvenuti et al. research Earthquake Alerts and Report System, which endeavours tweets, to see how such frameworks can be valuable during emergency related occasions (Avvenuti et al. 2017). The framework gathers tweets during a continuous emergency occasion, channels superfluous substance identifies an occasion, evaluates harm, and for fathom ability, it gives a perception. Creators presume that such a framework is exceptionally significant for emergency related occasions. The investigation of Kim also, Hastak examine how crisis offices and associations can even more likely arrangement activity techniques for a calamity by using people's data on an informal community (Kim and Hastak 2018). For the programmed investigation of webbased life literary and media streams current writing reports a few AI and computational techniques. A large portion of these techniques for the most part utilize regulated or solo (e.g., bunching and point displaying) strategies. The directed strategies incorporate great AI calculations, for example, Random Woodland and Deep Neural Network, for example, Convolutional Neural Network (CNN), for a total study of these procedures and their applications in the emergency informatics space see (Imran, Castillo, Diaz, et al. 2015; Castillo 2016).

The best in class explore on conclusion examination is for the most part centered around grouping feeling in both of two marks (i.e., positive, or negative) or five names (i.e., positive to exceptionally negative) from printed data (Pang, Lee, et al. 2008, for example, film audits (Pang, Lee, and Vaithyanathan 2002), tweets (Paltoglou and Thelwall 2010), and paper articles and remarks (Celli et al. 2016). For supposition examination, one of the usually utilized methodologies is to utilize supposition dictionary (i.e., SentiWordNet, Sentiment Treebank, and Psycholinguistic highlights) (Cambria et al. 2016; Socher et al. 2013; Alam, Danieli, et al. 2018) as highlights for planning the opinion classifier. In (Nagy and Stamberger 2012), the creators report the utilization of emojis alongside SentiWordNet helps in improving the arrangement of opinion from microblogs dataset gathered during calamities and emergencies. Socher et al. present the utilization "Notion Treebank" can help in recognizing supposition marks with a precision of 80.7% to 85.4% (Socher et al. 2013). Other normal methodologies incorporate the usage of word embeddings along with profound neural systems. The broad relative examinations can be found in SemEval tweet order task (see Rosenthal et al. 2017). After some time, a few open-source devices have additionally been created. Among them, one of the most broadly utilized devices is the Stanford CoreNLP . The use of microblogging platforms such as Twitter during the sudden onset of a crisis situation has been increased in the last few years. Thousands of crisis-related mes- sages that are posted online contain important informa- tion that can also be useful to humanitarian organizations for disaster response efforts, if processed timely and effec- tively (Hughes and Palen, 2009; Imran et al., 2015). Many different types of processing techniques ranging from machine learning to natural language processing to computational linguistics have been developed (Corvey et al., 2010) for different purposes (Imran et al., 2016). Despite there exists some resources e.g. (Temnikova et al., 2015; Olteanu et al., 2015), however, due to the scarcity of rel- evant data, in particular human-annotated data, crisis informatics researchers still cannot fully utilize the capabilities of different computational methods. To overcome these issues, we present to research community a corpora consisting of labeled and unlabeled crisis-related Twitter messages. Moreover, we also provide normalized lexical re- sources useful for linguistic analysis of Twitter messages.

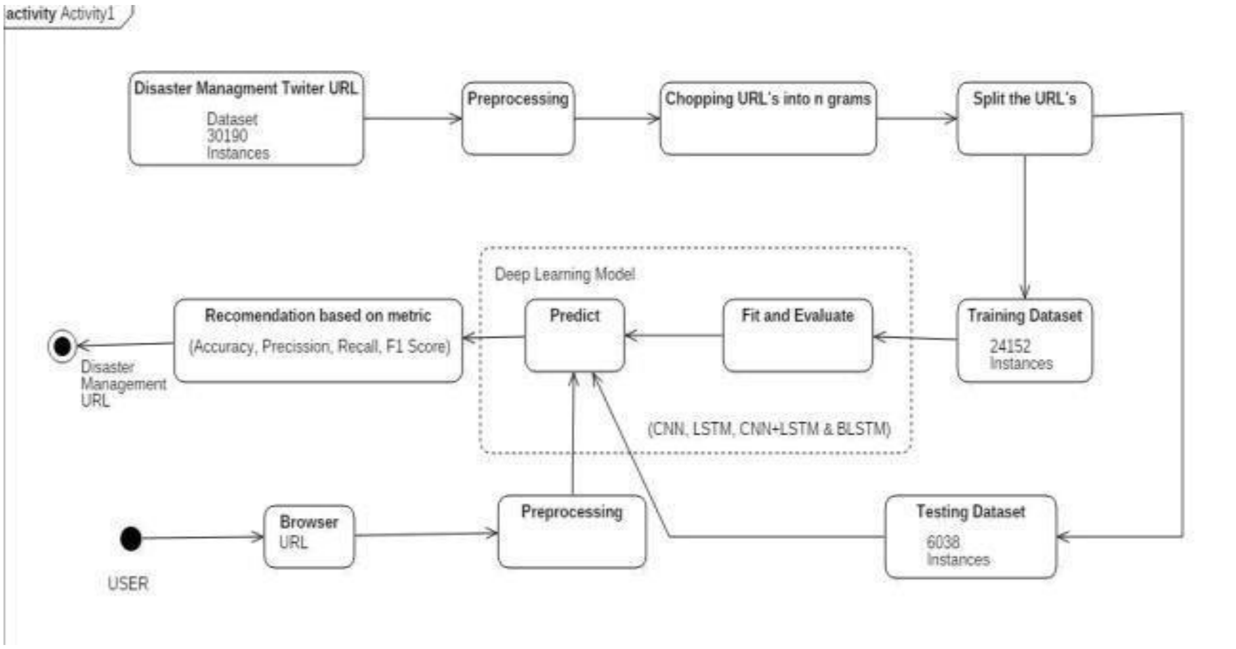
Moreover, humanitarian, and formal crisis response organizations such as government agencies, public health care NGOs, and military are tasked with responsibilities to save lives, reach people who need help, etc. (Vieweg et al., 2014). Situation-sensitive requirements arise during such events and formal disaster response agencies look for actionable and tactical information in real-time to effectively estimate early damage assessment, and to launch relief efforts accordingly. Many Natural-Language- Processing (NLP) techniques such as automatic summarization, information classification, named-entity recognition, information extraction can be used to process such social media messages (Bontcheva et al., 2013; Imran et al., 2015). However, many social media messages are very brief, informal, and often contain slangs, typographical errors, abbreviations, and incorrect grammar (Han et al., 2013).

Data Description:

We gathered emergency related messages from Twitter posted during 19 unique emergencies that occurred from 2013 to 2015. Table 1 shows the rundown of emergency occasions alongside their names, emergency type (for example seismic tremor, flood), nations where they occurred, and the quantity of tweets every emergency contains. We gathered these messages utilizing our Artificial Intelligence for Disaster Response stage (Imran et al., 2014). Artificial Intelligence for Disaster Response is an open-source stage to gather furthermore, characterize Twitter messages during the beginning of a helpful emergency. Artificial Intelligence for Disaster Response has been utilized by UN OCHA during many serious fiascos, for example, Nepal Earthquake; Typhoon Hagupit. Artificial Intelligence for Disaster Response gives distinctive helpful approaches to gather messages from Twitter utilizing the Twitter' s spilling API. One can utilize various information gathering techniques. For instance, gathering tweets that contain a few watchwords and are explicitly from a specific land territory/locale/city (for example New York).One of the basic necessities of many of these tasks is the availability of data, in particular human-annotated data. In this paper, we present human-annotated Twitter corpora collected during 19 different crises that took place between 2013 and 2015. To demonstrate the utility of the annotations, we train machine learning classifiers. Moreover, we publish first largest word2vec word embeddings trained on 52 million crisis-related tweets. To deal with tweets language issues, we present human-annotated normalized lexical resources for different lexical variations. Twitter has been extensively used as an active communication channel, especially during mass convergence events such as natural disasters like earthquakes, floods, typhoons (Imranet al., 2015; Hughes and Palen, 2009). Where the UCI repository has many datasets to work on. Among them the datasets were used by the many data community members to test the algorithms and tools. Most of the datasets were based on Classification problems. Among them tweet classification is one which plays a major role in classification problem. For instance, gathering tweets that contain a few watchwords and are explicitly from a specific land territory/locale/city (for example New York).

Analysis & Implementation Results:

In the wake of breaking down all the datasets with the recently referenced multi-name remark that I could find to use them as getting ready data. I likewise set up a gradient descent algorithm with Twitter embeddings as pattern. The rest of the undertaking of this task is to utilize LSTM and ANN to improve the presentation of classifier. The motivation to pick LSTM and ANN is that it has accomplished condition of-workmanship execution in numerous NLP undertakings, and it is publicly released. I built up a few LSTM-based models and they all outperform the gauge execution and also some ANN models with step function. The fine-tuned approach in the paper of Devlin et al. and LSTM-based gradient descent algorithm accomplished best execution, boosting up 3% regarding exactness, Matthew coef, F-1 and also we trying to check the same with ANN.



Steps for Implementation:

Identification of Data set

Text Processing

Identification of candidate OOV words

Evaluation Metrics

Results:

```
Unnamed: 0  ...                                label
0          0  ...              other_useful_information

[1 rows x 4 columns]
   Unnamed: 0  ...                             label
0           0  ...         other_useful_information
1           1  ...         other_useful_information
2           2  ...         other_useful_information
3           3  ...         other_useful_information
4           4  ...         other_useful_information
```

```
5        5  ...                                        not_related_or_irrelevant
6        6  ...                                        not_related_or_irrelevant
7        7  ...                                        not_related_or_irrelevant
8        8  ...                                         injured_or_dead_people
9        9  ...                                        donation_needs_or_offers_

[10 rows x 4 columns]
100%|██████████| 407873900/407873900 [00:15<00:00, 25793650.90B/s]
Writing example 0 of 17679 Writing
example 10000 of 17679 train 17679
16
1
16
ANNForSequenceClassification(
  (ANN): ANNModel(
    (embeddings): ANNEmbeddings(
      (word_embeddings): Embedding(30522, 768, padding_idx=0)
      (position_embeddings): Embedding(512, 768)
      (token_type_embeddings): Embedding(2, 768)
      (LayerNorm): ANNLayerNorm()
      (dropout): Dropout(p=0.1, inplace=False))

Epoch:   0%|          | 0/1 [00:00<?, ?it/s]
Fininshed: 0.00% (0/1104)
Fininshed: 19.93% (220/1104)
Fininshed: 39.86% (440/1104)
Fininshed: 59.78% (660/1104)
Fininshed: 79.71% (880/1104)
Fininshed: 99.64% (1100/1104)
```

Epoch: 100%|██████████| 1/1 [32:41<00:00, 1961.75s/it]

Writing example 0 of 4420

Output:

accuracy_score: 0.7529411764705882

precision_score macro: 0.6823439400117844

precision_score micro: 0.7529411764705882

recall_score macro: 0.6555834472227754

recall_score micro: 0.7529411764705882

f1_score macro: 0.6631546661216651

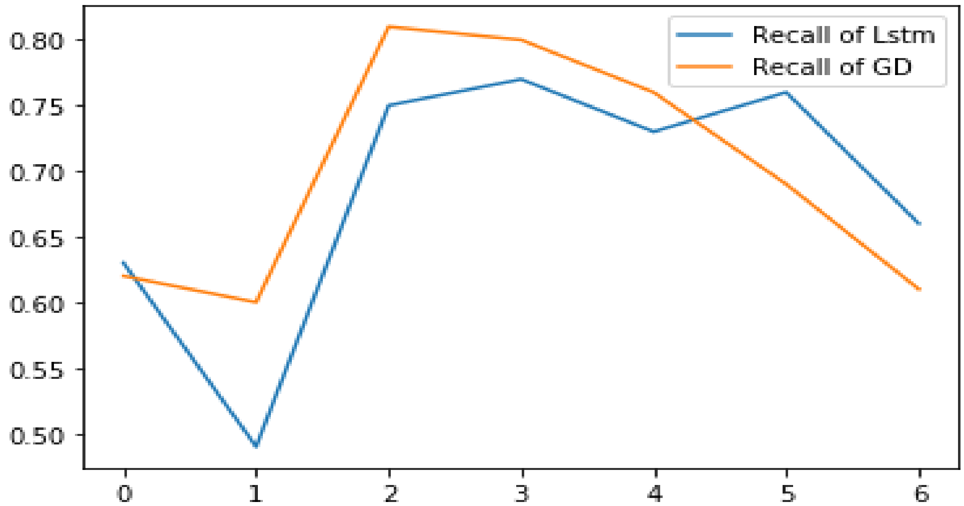
f1_score micro: 0.7529411764705881

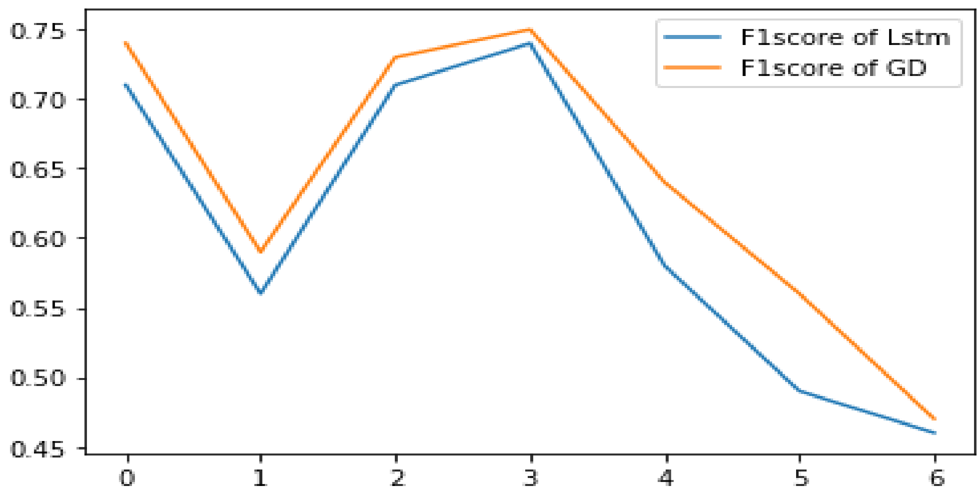
Evaluation Metrics:

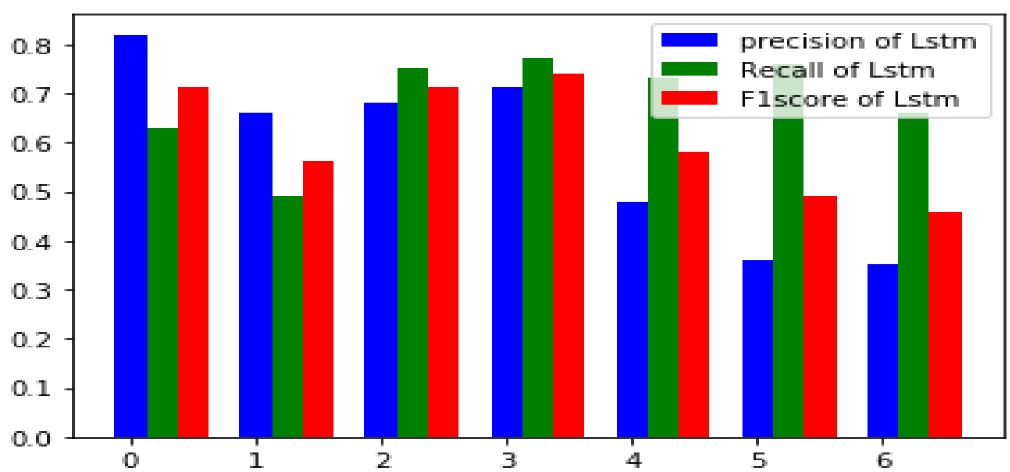
These Values below are Preliminary values for Evaluation Metrics

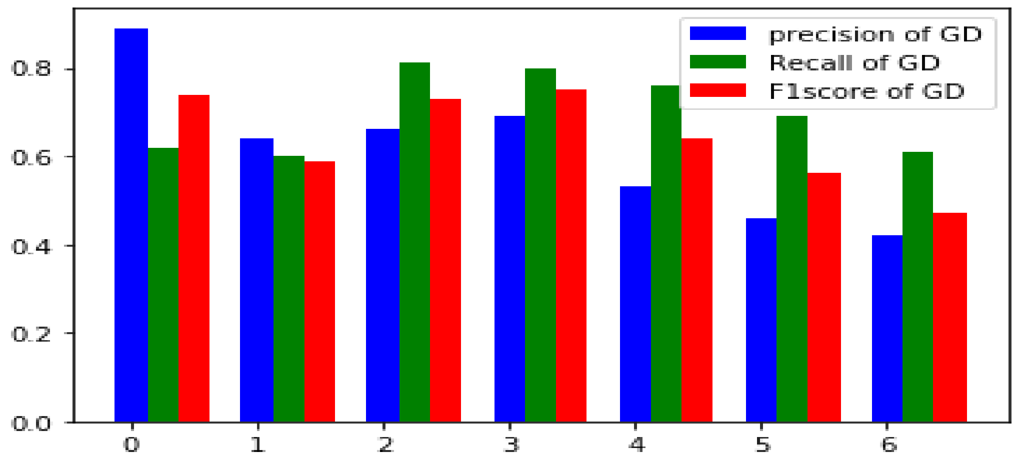
Accuracy	Precision		Recall		F1-Score	
	macro	micro	macro	micro	macro	micro
0.7554	0.6722	0.7554	0.6604	0.7554	0.6633	0.7554

Statistical Analysis:









Conclusion:

Internet based life has been attracting considerations from scientists and professionals the field of debacle the board. Exact message classification is a vital necessity to settle on choices from the bounteous however uproarious client produced information. ANN-based classifiers could achieve better execution compared with the bi-LSTM pattern model. A few marks are preferable unsurprising over others. Uncertainty and subjectivity are an incredible deterrent to lift up execution of classifier. The nature of the information needs improvement.

References/Bibliography:

1. Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. "What to Expect When the Unexpected Happens: Social Media Communications Across Crises Human Factors; Measurement". In: (). DOI: 10.1145/ 2675133.2675242. URL: http://dx.doi.org/10.1145/2675133.2675242.
2. Alexandra Olteanu et al. "CrisisLex : A Lexicon for Collecting and Filtering

Microblogged Communi- cations in Crises”. In: the Eighth International AAAI Conference on Weblogs and Social Media. 2014, pp. 376–385.

3. Jeffrey Pennington, Richard Socher, and Christopher Manning. “Glove: Global Vectors for Word Rep- resentation”. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014. ISBN: 9781937284961. DOI: 10.3115/v1/D14-1162.

4. Bernd Resch, Florian Usländer, and Clemens Havas. “Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment”. In: Cartogra- phy and Geographic Information Science 45.4 (2018), pp. 362–376. DOI: 10.1080/15230406.2017. 1356242. URL: http://www.tandfonline.com/action/journalInformation?journalCode= tcag20.

5. Christian Reuter and Marc-Andre Kaufhold. “Fifteen years of social media in emergencies: A retro- spective review and future directions for crisis Informatics”. In: Journal of Contingencies and Cri- sis Management 26.1 (Mar. 2018), pp. 41–57. DOI: 10.1111/1468- 5973.12196. URL: http: //doi.wiley.com/10.1111/1468-5973.12196.

6. Stefan Stieglitz et al. “Sense-making in social media during extreme events”. In: Journal

of Contingencies and Crisis Management 26.1 (Mar. 2018), pp. 4–15. ISSN: 14685973.

DOI: 10.1111/1468-5973.12193. URL: http://doi.wiley.com/10.1111/1468-5973.12193.

7. Kevin Stowe et al. “Improving Classification of Twitter Behavior During Hurricane Events”. In: the 6th International Workshop on Natural Language Processing for Social

Media. Melbourne, Australia, 2018, pp. 67–75. URL: http://aclweb.org/anthology/W183512.

8. Jacob Devlin et al. “ANN: Pre-training of Deep Bidirectional Transformers for

Language Understanding”. 2018.

9. Yury Kryvasheyeu et al. “Rapid assessment of disaster damage using social media activity”. In: Science Advances (2016). ISSN: 23752548. DOI: 10.1126/sciadv.1500779.

10. Benjamin Herfort et al. “Twitter Analysis of River Elbe Flood”. In: International

Conference on Information Systems for Crisis Response and Management. Pennsylvania, USA, 2014. URL: http: //www.producao.usp.br/handle/BDPI/46102.

11. Firoj Alam et al. “A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria”. In: the

15th International Conference on Information Systems for Crisis Response and Management (ISCRAM). Rochester, USA, 2018.