Multiclass Classification of Flood Tweets using CNN and LSTM

Sachi Verma
B.E. Student
Department of Computer Engineering
Fr. Conceicao Rodrigues College of Engineering
University of Mumbai
Mumbai, India

Prerna Sharma
B.E. Student
Department of Computer Engineering
Fr. Conceicao Rodrigues College of Engineering
University of Mumbai
Mumbai, India

Joel Ayappa
B.E. Student
Department of Computer Engineering
Fr. Conceicao Rodrigues College of Engineering
University of Mumbai
Mumbai, India

Supriya Kamoji
Assistant Professor
Department of Computer Engineering
Fr. Conceicao Rodrigues College of Engineering
University of Mumbai
Mumbai, India

Abstract— **During natural disasters, bystanders and affected people post situational updates including reports of injured or dead people, infrastructure damage, requests for urgent needs and so on. The role of social media, in particular microblogging platforms such as Twitter, as a conduit for actionable and tactical information during disasters is increasingly acknowledged. Twitter plays a major role in the rapid propagation of information during disasters. It allows accessing or dispersing crucial information or breaking news directly from the affected areas. Tweets can be published from multiple platforms and devices and are delivered to users in real time. In order to connect to a general topic, users can add hash tag as keywords to their posts. Hash tag is a meta character which is expressed as #keyword or #hash tag. Hash tag helps people to pursue their interested topics very easily and quickly. Hash tag will provide the tweets related to a common or particular topic.**

Keywords— **Flood, Twitter, CNN, LSTM, hashtag, Machine Learning, Deep Learning**

I. INTRODUCTION

Floods are the most frequent and devastating type of natural disaster causing a significant loss in terms of human lives and infrastructure worldwide, every year. The damage of flood can be significantly mitigated if timely and accurate information about the location, scale, and most affected areas is available. However, several challenges, such as the availability of reporters and other resources, etc., are associated with the gathering of such information during floods. On the other hand, social media has been proved very effective in information dissemination in such events. Hence, we have come up with a system that retrieves information from social media, namely Twitter, and processing the textual and visual information by classification and filtering to get most accurate details about floods specific to India. The development objective of the proposed Project is to improve urban flood management and to counteract natural calamities protect life and property and improve safety for the people and ensure social stability.

II. ANALYSIS AND METHODOLOGY

A. Problem Statement Analysis

An online study says that most people in the severely damaged areas of a flood end up with no means to communicate (like internet connections) and put forward their necessities. However, people in the further areas with less damage tweet dramatically greater than usual to help convey the severity of damage. [2] According to reports, as floods cause major damage to life and property every year, it is time the central and the state governments prepare a long-term plan that goes beyond piecemeal measures like building embankments and dredging to control floods. [10] This motivates us to develop our Flood Severity prediction project that will classify floods based on their severity to help the government and other organizations supply necessities accordingly, thereby enhancing flood management. The development objective of the proposed Project is to improve urban flood management and to counteract natural calamities protect life and property and improve safety for the people and ensure social stability.

B. Methodology Flow

- Tweets are collected using different available libraries to create datasets of relevant keywords in the form of hash tags

- The collected data is preprocessed to clean it

- Tweets are vectorized and then classified as relevant and non-relevant where relevant tweets are filtered out for further processing

- The relevant Tweets are then filtered into severity categories and then the location is extracted from there.
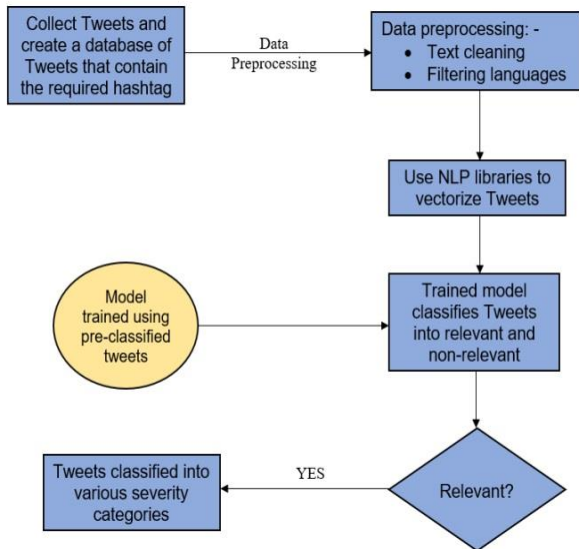
Fig. 1 – Methodology of Implementation

III.   IMPLEMENTATION

There does not exist a dataset containing of flood tweets. So, we first created a dataset from scratch that contained by scraping tweets with the required hashtags.

A.   Data Preprocessing

The collected data is cleaned and preprocessed to keep it ready for training and testing. This includes removing urls, punctuations, emojis and html tags as well as separating words which are commonly smashed together in social media text situations. Next, we conduct a spell check and correction.

We will also have to convert our text data to tokens and ultimately 'embeddings' to train our model on. A popular pre-trained model for this task is the GloVe embedding model, so we load GloVe Text Embeddings. Word embeddings are vector representations of words. The embedding corpus is simply a hash table which consists of a number of words (keys) and their corresponding vector representation (values). We then create text corpus and clean stop words. Lastly, we need to confirm our embedding matrix is of the dimensions equal to number of unique words and each words vector should be of size of the dimensional array.

B.   Classification into relevant and irrelevant tweets

For this, we used a custom RNN which includes several LSTM, GRU and Bi-directional layers, each with larger number of inputs to start and then decreasing as the model goes deeper. Used dropout at regular so that when we freeze a portion of the weights during training, we force the model to optimize with fewer parameters which typically leads to better generalization and robustness on unseen test data. Also freezing some weights means we do not have to perform gradient descent on that proportion of the weights meaning training time will be reduced. This allows for more experiments in less time. Using a large pre-trained embedding model also helped with performance. The GloVe 300d increases the models ability to learn by a large amount. Using a learning rate optimizer (ADAM in this case) helped

to converge faster and helped complete more experiments in shorter amount of time as well.

We use F1 score as the evaluation metric as:-

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

Where

$$precision = \frac{TP}{TP + FP}$$

And

$$recall = \frac{TP}{TP + FN}$$

True Positive [TP] = your prediction is 1, and the ground truth is also 1 - you predicted a positive and that's true.

False Positive [FP] = your prediction is 1, and the ground truth is 0 - you predicted a positive, and that's false.

False Negative [FN] = your prediction is 0, and the ground truth is 1 - you predicted a negative, and that's false.
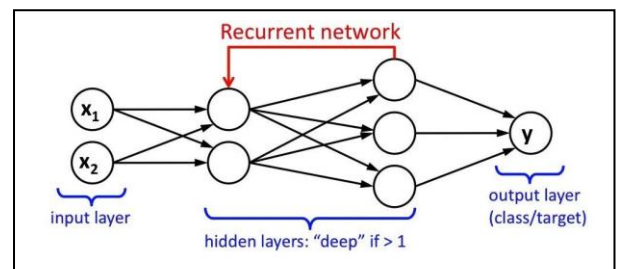

Fig. 2 – RNN Diagram

Accuracy	F1 Score	Loss
0.8228	0.6569	0.3729

Table 1 – Scores for the relevancy classification model

C.   Multiclass Tweet Classification

Text classification is a modelling technique in which the class for a given sequence is predicted using a series of sequences as input. The difficulty with this predictive modelling technique is that the input sequences don't always have the same length. Because of the sequences' constant length, which contributes to a relatively vast vocabulary, learning long-term contexts typically requires a model. For a comparative analysis, we tested the dataset on a number of Machine Learning and Deep Learning models.

1) Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.
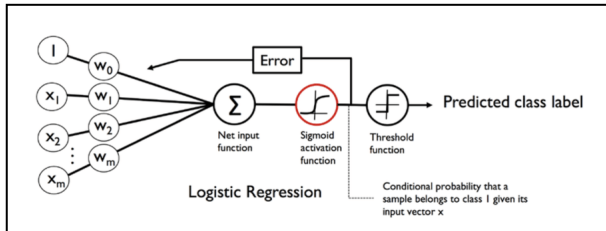

Fig. 3 – Logistic Regression Diagram

2) Linear SVC

The objective of a Linear SVC (Support Vector Classifier) is to fit to the data you provide, returning a "best fit" hyperplane that divides, or categorizes, your data. From there, after getting the hyperplane, you can then feed some features to your classifier to see what the "predicted" class is. This makes this specific algorithm rather suitable for our uses, though you can use this for many situations. LinearSVC is implemented in terms of liblinear which is the reason it has flexibility in the choice of penalties and loss functions
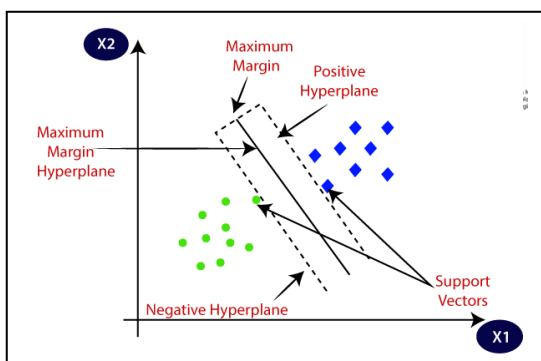

Fig. 4 – SVM Diagram

3) Multinomial Naïve Bayes

The Multinomial Naive Bayes algorithm is a Bayesian learning approach popular in Natural Language Processing (NLP). The program guesses the tag of a text, such as an email or a newspaper story, using the Bayes theorem. It calculates each tag's likelihood for a given sample and outputs the tag with the greatest chance. The Naive Bayes classifier is made up of a number of algorithms that

all have one thing in common: each feature being classed is unrelated to any other feature. A feature's existence or absence has no bearing on the inclusion or exclusion of another feature.
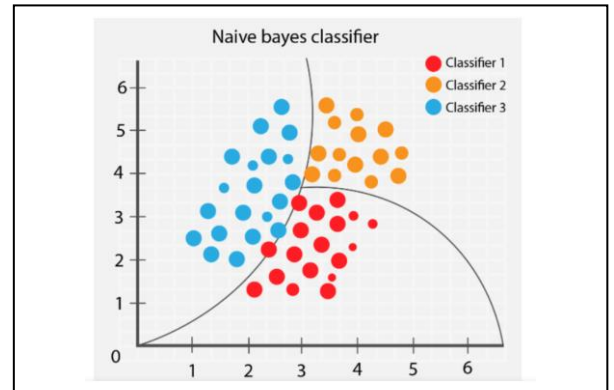

Fig. 5 – Multinomial Naïve Bayes Diagram

4) Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
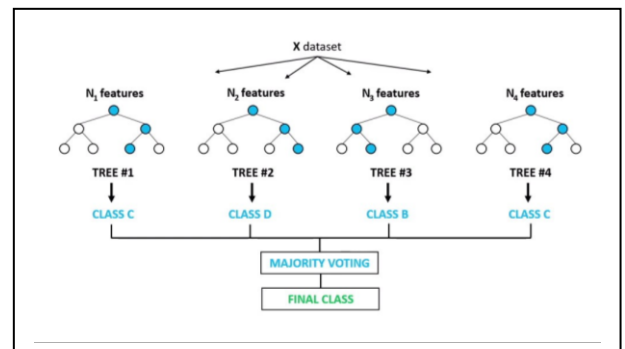

Fig. 6 – Random Forest Classifier

Comparative Analysis of the above mentioned as well as a few more Machine Learning models:-

	Logistic Regression	SVM	Naive Bayes	Random Forest
Accuracy	66.746126	70.202622	53.873659	55.423123
F1_score	35.993173	48.279275	8.027717	42.745904
Recall	66.746126	70.202622	53.873659	55.423123
Precision	66.746126	70.202622	53.873659	55.423123

Table 2 – Comparative Analysis of ML models for text classification

5) Deep Learning Models

Text analysis method based on Convolutional neural network (CNN) can obtain important features of text through pooling but it is difficult to obtain contextual information which can be leverage using LSTM.

	Accuracy	Loss
LSTM 1-Layer	0.7974	0.4029
LSTM 2-Layer	0.7625	0.4287
CNN + LSTM	0.8226	0.3594

Table 3 – Comparative Analysis of DL models for text classification

IV. CONCLUSION

These days, social media's impact during a crisis is almost exactly the same as that of traditional media. In addition, compared to other traditional media, social media has a far larger percentage of irrelevant material. A good categorization technique for social media data might be incorporated to significantly lessen the influence of irrelevant data. Collecting specific tweets pertaining to the disaster management required helps give more meaningful insights as opposed to overall disaster classification. The comparative analysis of Machine Learning as well as Deep Learning models help us understand the metrics on which they are based and hence the respective performance. When we consider Machine Learning, we can see that the SVM model gives us the most efficient output with the highest accuracy. Similarly, when it comes to Deep Learning, the combination of using CNN and LSTM together gave a better result than just using 1,2-layer LSTM models. The future scope for this would be location extraction and mapping of tweets to better assist people in need.

REFERENCES

[1] Olteanu, A., Castillo, C., Diakopoulos, N., & Aberer, K. (2014). CrisisLex: A lexicon for collecting and filtering microblogged communications in crises. In Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM) (pp. 168-177). https://doi.org/10.1609/icwsm.v8i1.14538

[2] Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013). Extracting information nuggets from disaster-related messages in social media. In Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM) (pp. 708-717).

[3] Zhang, Y., & Liu, L. (2016). A survey of opinion mining and sentiment analysis. In Mining the Social Web (pp. 169-222). Springer. DOI 10.1007/978-1-4614-3223-4_13

[4] Ritter, A., Clark, S., Mausam, & Etzioni, O. (2011). Named entity recognition in tweets: An experimental study. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1524-1534).

[5] Akter, S., Wamba, S.F. Big data and disaster management: a systematic review and agenda for future research. Ann Oper Res 283, 939–959 (2019). https://doi.org/10.1007/s10479-017-2584-2

[6] Singh, J.P., Dwivedi, Y.K., Rana, N.P. et al. Event classification and location prediction from tweets during disasters. Ann Oper Res 283, 737–757 (2019). https://doi.org/10.1007/s10479-017-2522-3

[7] Wang, Z & Ye, X 2018, 'Social media analytics for natural disaster management', International Journal of Geographical Information Science, vol. 32, no. 1, pp. 49-72. https://doi.org/10.1080/13658816.2017.1367003

[8] Hernandez-Suarez, A.; Sanchez-Perez, G.; Toscano-Medina, K.; Perez-Meana, H.; Portillo-Portillo, J.; Sanchez, V.; García Villalba, L.J. Using Twitter Data to Monitor Natural Disaster Social Dynamics: A Recurrent Neural Network Approach with Word Embeddings and Kernel Density Estimation. Sensors 2019, 19, 1746. https://doi.org/10.3390/s19071746

[9] Xiao Huang, Cuizhen Wang, Zhenlong Li & Huan Ning (2019) A visual–textual fused approach to automated tagging of flood-related tweets during a flood event, International Journal of Digital Earth, 12:11, 1248-1264, DOI: 10.1080/17538947.2018.1523956

[10] de Bruijn, J.A., de Moel, H., Jongman, B. et al. A global database of historic and real-time flood events based on social media. Sci Data 6, 311 (2019). https://doi.org/10.1038/s41597-019-0326-9

[11] Ahmouda, A.; Hochmair, H.H.; Cvetojevic, S. Using Twitter to Analyze the Effect of Hurricanes on Human Mobility Patterns. Urban Sci. 2019, 3, 87. https://doi.org/10.3390/urbansci3030087

[12] Volodymyr V. Mihunov, Nina S. N. Lam, Lei Zou, Zheye Wang & Kejin Wang (2020) Use of Twitter in disaster rescue: lessons learned from Hurricane Harvey, International Journal of Digital Earth, 13:12, 1454-1466, DOI: 10.1080/17538947.2020.1729879

[13] Karami, A., Shah, V., Vaezi, R., & Bansal, A. (2020). Twitter speaks: A case of national disaster situational awareness. Journal of Information Science, 46(3), 313-324. https://doi.org/10.1177/0165551519828620

[14] Mendon, S., Dutta, P., Behl, A. et al. A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. Inf Syst Front 23, 1145–1168 (2021). https://doi.org/10.1007/s10796-021-10107-x

[15] J. Du, C. -M. Vong and C. L. P. Chen, "Novel Efficient RNN and LSTM-Like Architectures: Recurrent and Gated Broad Learning Systems and Their Applications for Text Classification," in IEEE Transactions on Cybernetics, vol. 51, no. 3, pp. 1586-1597, March 2021, doi: 10.1109/TCYB.2020.2969705.

[16] Bodapati, S., Bandarupally, H., Shaw, R.N., Ghosh, A. (2021). Comparison and Analysis of RNN-LSTMs and CNNs for Social Reviews Classification. In: Bansal, J.C., Fung, L.C.C., Simic, M., Ghosh, A. (eds) Advances in Applications of Data-Driven Computing. Advances in Intelligent Systems and Computing, vol 1319. Springer, Singapore. https://doi.org/10.1007/978-981-33-6919-1_4

[17] Rupapara V, Rustam F, Amaar A, Washington PB, Lee E, Ashraf I. 2021. Deepfake tweets classification using stacked Bi-LSTM and words embedding. PeerJ Computer Science 7:e745 https://doi.org/10.7717/peerj-cs.745

[18] Md. Yasin Kabir and Sanjay Madria. 2019. A Deep Learning Approach for Tweet Classification and Rescue Scheduling for Effective Disaster Management. In Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '19). Association for Computing Machinery, New York, NY, USA, 269–278. https://doi.org/10.1145/3347146.3359097

[19] Moreo, A., Esuli, A. & Sebastiani, F. Word-class embeddings for multiclass text classification. Data Min Knowl Disc 35, 911–963 (2021). https://doi.org/10.1007/s10618-020-00735-3

[20] Z. Tan, J. Chen, Q. Kang, M. Zhou, A. Abusorrah and K. Sedraoui, "Dynamic Embedding Projection-Gated Convolutional Neural Networks for Text Classification," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 3, pp. 973-982, March 2022, doi: 10.1109/TNNLS.2020.3036192.

[21] Daler Ali, Malik Muhammad Saad Missen, Mujtaba Husnain, "Multiclass Event Classification from Text", Scientific Programming, vol. 2021, Article ID 6660651, 15 pages, 2021. https://doi.org/10.1155/2021/6660651

[22] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on twitter: Classification performance and challenges," in Big Data Mining and Analytics, vol. 2, no. 3, pp. 181-194, September 2019, doi: 10.26599/BDMA.2019.9020002.