Emotion Based Content Credibility Prediction Model For Twitter Social Network

Faraz Ahmad, S A M Rizvi

Abstract: Twitter and other microblogging platforms are reportedly being used for propagating spams and other malicious content, which stand contrary to the vision of these platforms. This situation raises several ethical challenges in terms of deceiving people with pseudo information. Many a time, users of microblogging sites are held responsible for spreading spiteful information on social media, and they remain unknown in most of the cases. Tweet content may contain either genuine information or uncredible content. The uncredible content contains an abusive or absurd language for any caste, culture, religion or political party. This content needs to be filtered out so that it does not disturb the tranquility of nation in the long run. Therefore, filtering such content from Online Social Networks (OSNs) is the utmost requirement, and so, needs to be addressed. Earlier studies have focused on mainly feature-set belonging to content-based, topic-based and network-based categories. However, the potential of emotion-based features still remains to be explored in the domain. In this paper, IBM Watson and Meaning Cloud platforms have been used for evaluating emotions, sentiment and polarity scores in order to develop a classification model that will filter out all the uncredible content from the OSNs like Twitter. Initially, 35K tweets and associated features provided by Twitter were crawled. These tweets were further preprocessed and forwarded to six human experts for annotating it on a five-pointer credible class scores. The Multilayer perceptron, Naive Bayes, Random Forest and Support Vector Machine algorithms were used for developing a machine learning classification model for categorizing tweets into one of the given credibility classes. The acceptable level of accuracy, precision, recall, and f1 score is observed for all given credibility classes.

Index Terms: Social Media, Twitter, Credibility, Emotions, Sentiment, Machine Learning.

1 INTRODUCTION

Social networks are always dynamic in nature where people get all kinds of information, either having high impacted nature or a casual one. Twitter is one of the most influential and reputed platforms where people can post information through text, pictures, and videos, which provides benefits in several possible ways. People across the world can get rapid information associated with any event. Information sharing is the process where users can post their own thoughts or feelings and can also retweet the content of others. While sharing content, it is very essential that a user should not post anything which is detrimental to the emotions and sentiments of people belonging to different caste, culture, religion or political background. Tweets containing hate speeches, abusive or fake content can create chaos and even end into a disastrous condition. The main objective of this paper is to develop a machine learning model that helps in classifying the fake or uncredible content from the OSNs. Various researchers tried to develop a machine learning model to filter out uncredible or rumored content from Twitter social networks. The feature sets incorporated by researchers for developing classification models are user-based features, content-based features, topic-based features, network features, external resource features and other metadata features for finding the credibility score of the respective tweet. However, the potential of emotion-based features still needs to be explored in the domain. The emotion-based model will facilitate in filtering uncredible content not only posted by random Twitter users but also by the reputed/affluent users.

- Faraz Ahmad is currently pursuing Ph.D. degree in Computer Science and Applications from Department, of Computer Science, Jamia Millia Islamia University, New Delhi, India, PH-+ 91-9911096801. E-mail: faraz159020@st.jmi.ac.in
- S A M Rizvi is currently working as a Professor in Department of Computer Science, Jamia Millia Islamia University, New Delhi, India, PH-+91- 9999915251. E-mail: sarizvi@jmi.ac.in

The users who actively participated in almost every real-time event trending on social networking sites and having an average number of followers/friends count are considered as reputed/affluent users. These users updated their profile consistently and people recognized them as an influential citizen. The users who get the verification tag by Twitter mostly fall under this category. These verified users are politicians, actors, cricketers, journalists and other renowned personalities. The aim of this paper is to develop an emotion-based model that can effectively filter out uncredible content from credible ones not only posted by the general public but also by these reputed/affluent users, which provide benefits to the society in long run. The presented work is divided into five major steps: In the first step, 35K tweets were crawled using Twitter Rest API, all the extracted tweets belong to controversial hashtags and high impact occasions happened in India. The next step is data preprocessing where tweets were preprocessed based on extracted features set (user's followers count, tweet favorite count, user verified by Twitter, user-created at, tweets created at, user's description is present or not and tweet is possibly sensitive) provided by twitter. Furthermore, tweets having a length of fewer than 10 words, or tweets posted by those users who sign up on twitter for less than 3 months were discarded. Tweets from users whose user's information like profile photo, bio, user description were missing from the profile, were also discarded. After preprocessing, the next step is to annotate the data with the help of human experts (i.e. research scholars and bloggers) who have in-depth knowledge of all the occasions/events from which tweets belong. The tweets were assigned to six different experts to annotate it using a five-pointer credibility class scale. Moreover, for analyzing the topics and events associated with the tweets, a word cloud has been generated. This word cloud contains words that have occurred more than 200 times in the extracted dataset. The next step is the most significant step in which emotions, sentiment and polarity features of the tweets were evaluated using API provided by IBM Watson and Meaning-Cloud platforms. Sentiment scale ranges from -1 to +1, whereas emotions are categorized into five major classes like Joy, Anger, Sadness, Fear, and Disgust and measured on a scale of 0 to +1. The tweets polarity was measured over the six-pointer scale as P+ (Positive Polarity), P (Somewhat Positive), NONE, NEU, N (Somewhat Negative), and N+ (Negative Polarity). NONE and NEU belong to the same credibility class, so these two classes are merged together. In the last step, a machine learning classification model has been developed to classify the annotated tweets with the help of extracted features set. For tweets classification into their respective credibility classes, a Multilayer perceptron (MLP), Naive Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM) were used. The rest of the paper is organized in the following order. Section 2 discusses background study related to credibility analysis in social networks; section 3 discusses data extraction, preprocessing, annotation and features mining. Section 4 discusses model generation using machine learning techniques that are used for classification and section V is a conclusion with future directions.

2 LITERATURE REVIEW

In the current scenario, social media has been the major platform of accessing information and collecting news from all over the world however, their credibility is still questionable. Some research articles have been reviewed which helps us in understanding the cause of spreading fake content and also suitable measures to overcome these issues. The presented paper [1] measured the user influence based on three distinct parameters like in-degree (number of followers), mentions and retweets. Six million user's data were extracted, which revealed a significant amount of users hold influence on a variety of topics; however ordinary users cannot gain influence unexpectedly overnight. It can only be gained by posting relevant and insightful tweets that are acceptable by other users. In conclusion, authors demonstrated that having millions of followers on Twitter is not helpful in gaining influence, instead it is gained by active participation with the audience through mentions and retweets. Authors' [2] focused on the automatic evaluation of the credibility of the tweets posted on Twitter Social Networks. Twenty-five hundred varied events were extracted which was identified by the twitter monitor model. Based on such events 10,000 tweets were crawled and further processed by AMT workers and were labeled as either news-worthy or casual-chat. Moreover, for credibility assessment tweets were further labeled in one of the four given credible classes. For automatic credibility assessment topic-based, user-based, propagation-based, and message-based features were extracted, which helped in classifying the tweets using the J48 decision tree algorithm. Twitter is an important platform in today's world where users can update information in real-time and participate in any event and uploads the update. Authors' [3] suggested a technique that could retrieve the credibility of such events automatically. The graph-based optimization approach was implemented in order to get a solution to a problem. Furthermore, a PageRank like credibility dissemination consisting of users, tweets, and events on different social networking platforms. This paper also claimed that their methods outperform the decision tree classifier. The paper [4] presented a survey for user's awareness of tweet credibility. The differences were found between the features that the user considered appropriate for assessing the credibility and the one which is provided by the search engines. Moreover, two experiments were conducted in which numerous features were transformed systematically and found that the users are poor in judging the credibility and were often influenced by the heuristics like user name. Based on these outcomes, the strategies were discussed that how users can improve their credibility on online social networks. The presented work [5] suggested a monitor model based on unsupervised ML algorithms for monitoring emergency condition on Twitter. The paper focused on detecting credibility associated with the tweets posted by users on Twitter in emergency conditions. The datasets of such typical events were extracted and outsourced to the human experts who manually labeled the tweets into two classes viz. credible or uncredible. Furthermore, several features were extracted that facilitates in classifying tweets in one of the given two categories. Various machine learning algorithms were applied and got maximum classification accuracy of 66.7% by SVM. The presented paper [6] examined how users associated cues in online social networks changed the perception of source credibility. The six mock Twitter pages were generated and asked 289 different participants from two different universities to report their credibility. The variation in followers count and the ratio of followers to follows had been done and found that users having too many or too few followers were judged with lower trustworthiness and expertise. However, narrower the gap between the ratios of followers to follows higher could be the decree of credibility. Authors' [7] developed and deployed a browser extension named TweetCred for evaluating real-time credibility scores of the posted content on Twitter Social Networks. Credibility scores were calculated on seven pointer rating scale which was based on extensive sets of forty-five different features. The extensive datasets were crawled from twitter using Twitter streaming API based on six major events happened all over the world. Lastly, the SVM ranking algorithm was applied for assessing the credibility of tweets posted in real-time. The paper [8] suggested a technique that can be automatically used to detect fake or spiteful text from online social networks. The author extracted two different types of features; those features were associated with the user and the content which they had posted, namely "shallow features" and "implicit feature". The shallow features contain basic characteristics of the user or the content which he posts on social media whereas implicit features are attained by analyzing the social influence, detailed information regarding the user, sentiment score of posted content, etc. The experiment illustrated that the rumor detection process obtained significant enhancement as compared to the leading edge methods. Social media platforms like Twitter contain a large number of fake content relating to several events and topics. Authors' [9] built a corpus namely CredRank containing sixty million tweets that were collected from one thousand and forty-nine real-world events. These events were further tagged by thirty different AMT workers on a given credibility scale. The Latent Dirichlet Allocation algorithm is incorporated in order to identify real-world events and then manually filtered the topics specific to any particular event. The main contribution of this paper is to generate a well-labeled corpus annotated at the credibility scale by AMT workers. It furthermore enabled the researchers to explore the participation of media in facilitating the spread of the rumors. The presented paper [10] proposed a system for content credibility analysis on Twitter to avoid the proliferation of malicious content. This system contains four major components namely user reputation based component, credibility engine which classifies authentic and unauthentic content, component defining expertise of users for any particular topic and lastly feature ranking algorithm which ranks the features based on its importance for evaluating credibility. System effectiveness was evaluated on two different datasets and an acceptable level of precision and recall were found for different machine learning algorithms. The proposed work [11] used the twitter dataset consisted of real-time tweets which were crawled using river plugin provided by Twitter. These tweets were gone through manual credibility-classes labeling process. The tweets classified into one of the given credibility classes based on three broad categories of features. Initially, classification was performed using the Random Forest algorithm on "Twitter only features" and this result was compared with classification based on "reconcile only features." Reconcile steps helped in improving the manual tagging process and also to decrease the disparity between the results. At last, the combination of both the features was used for classification and found the highest classification accuracy for credibility classes with an acceptable level of precision and recall.

3 DATA PREPARATION & FEATURE GENERATION FOR MACHINE LEARNING (ML) TECHNIQUES

Initially, 35,000 tweets and their associated features were extracted using twitter rest API provided by Twitter. The tweets were further gone through preprocessing steps where all the partially updated profiles, tweets containing less than 10 words and tweets written in some other language except English were filtered out. Tweets containing only hashtags and mentions were also discarded. Moreover, only those tweets were taken into consideration whose user signup on Twitter for at least 3 months before the tweet was posted and contain at least 200 followers and friends count. This provides information about the authenticity and popularity of users among other users. After the filtration process, only 3,500 finest tweets were taken under consideration for model development. Data annotation is the next step in which we randomly selected six scholars who are having different backgrounds and opinions for the things happening within the society. All of them are frequent Twitter users and have in-depth knowledge about the events that happened within the last decade. Table 1 illustrates the examples of tweets labeled by the annotators for all four credibility classes.

TABLE 1
TWEETS LABELED BY ANNOTATORS

Acceptable	#WhyTheyHateModi Easy n simple bcz @narendramodi ji dislikes terrorism corruption scams and his(modiji) love for the nation and army persons!!!!
Slightly Acceptable	#WhyTheyHateModi First time in the history of India, unorganized sector workers like vegetable sellers, foot path shop keepers etc are given monthly □15000 pension. #PMSYM #PMShramYogiManDhan #vandematram
Neutral	RT @htTweets: An exhibition of Nizam's jewels will be inaugurated at the National Museum in Delhi on Monday after a gap of 11 years.
Slightly Unacceptable	RT @MousamiSingh1: Narendra Modi is working under BJP whereas Congress is working under Rahul Gandhi. You can see the bottom up and top down.#India
Unacceptable	RT @rishibagree: The progression Of Indian Intellectuals ⬇ ⬇ has been Hate for Modi Hate for Gujarat Hate for

Data was annotated on a five-pointer credibility scale (Credible, Somewhat-Credible, Neutral, Somewhat-Uncredible and Uncredible) for evaluating the credibility of the tweets. The Credibility-Class distributions are mentioned in Table 2.

TABLE 2
TWEETS DISTRIBUTION INTO CREDIBILITY CLASSES

Credibility Level	Distribution
Uncredible	780
Somewhat-Uncredible	356
Neutral	843
Somewhat-Credible	395
Credible	1124

The preprocessed data were further analyzed using IBM Watson and Meaning cloud tool to generate sentiment, emotions, and polarity associated with the tweets. These are the extended features that were calculated for developing a classification model. The sentiment is calculated on a scale of -1 to +1 where, 0 to 1 as +ve sentiment and 0 to -1 as –ve sentiment. The value closer is to -1 or +1 higher will be the sentiment. Emotions were categorized and calculated into 5 major categories like sadness, anger, joy, disgust, and fear on a scale of 0 to +1. And using Meaning cloud platform polarity associated with the tweets were evaluated, which categorized the results into P+, P, N, N+, None, Neutral classes.

TABLE 3
FEATURES SET AND DESCRIPTION

Features	Description
Sentiment Score	The sentiment of the tweet 0 to 1 as +ve sentiment and 0 to -1 as –ve sentiment
JOY	Joy Score calculated on 0 to 1 scale
FEAR	FEAR Score calculated on 0 to 1 scale
ANGER	ANGER Score calculated on 0 to 1 scale
DISGUST	DISGUST Score calculated on 0 to 1 scale
SADNESS	SADNESS Score calculated on 0 to 1 scale
Tweet Polarity	Polarity Scale (P+, P, N+, N, NONE, NEUTRAL)
Tweet Favorite Count	People count who liked the tweet
User Follower Count	People count who are following the user
User-Created At	First time when user signup on Twitter
Tweet Created At	Tweets posted time
User Description	User Description is present or not
Tweet Possibly Sensitive	The warning is given by Twitter that tweets are possibly sensitive.

For further analyzing the context and topics associated with the tweets, a word cloud was generated for all the words which occurred more than 200 times within the processed dataset. It elaborates that all the tweets strictly belong to the political system and other major incidences happened in India. People all over the nation showed their views for some major occasions like Demonetization, GST, Pulwama attack, Surgical Strike, etc. Some tweets are also strictly against and in favor of different active political parties in India like BJP, AAP, Congress, AIMIM, etc.

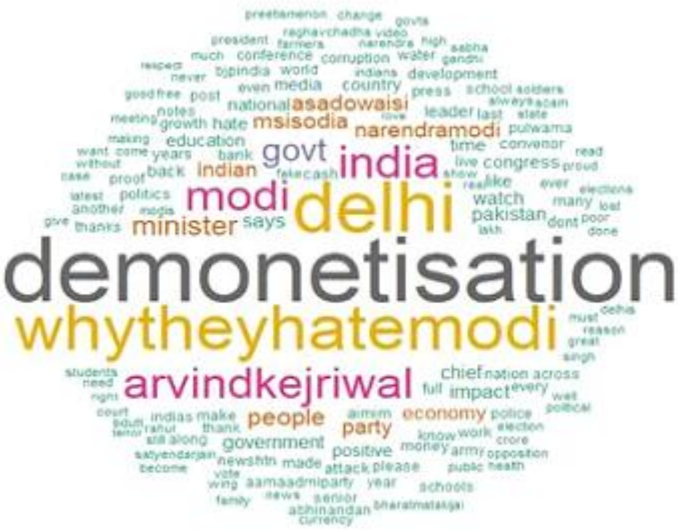

Fig 1. *Word Cloud*

4 MACHINE LEARNING MODEL GENERATION AND RESULTS DISCUSSION

Support Vector Machine, Random Forest, Naive Bayes, and Multilayer Perceptron classification model were applied for classifying tweets into its respective credibility class. Support Vector Machine (SVM) is a classifier that is used to classify both linearly and non-linearly separable data. SVM is used to draw a hyperplane in order to maximize the distances between two or more given separable classes. The number of hyperplanes depends upon the number of incorporated features and the given number of separable classes. Supports vector are the data points that provide the information regarding the exact location and orientation of the hyperplanes. These support vectors are helpful in maximizing the margins of the hyperplanes. SVM kernels are the mathematical function that takes data as an input and transforms it into the required form. There are various types of SVM kernels like linear, radial basis function, polynomial, sigmoidal etc. The incorporation of these kernels is based on the distribution of the data. In this study, an SVM classification model has been developed using a radial basis function. Altogether model generated 1469 support vector which provides an accuracy of 94.8621%. Random forest (RF) consists of a large number of distinct and uncorrelated decision trees. These uncorrelated models (tress) produce an ensemble prediction that outperforms any individual prediction model. RF can be used for both regression and classification models. The classification RF was developed in this study, which incorporates 300 trees. Furthermore, for evaluating the error rate a plot has been drawn which is shown in figure 2. As the number of trees increases, out of bag error initially drops and then it becomes relatively constant after 50 trees.

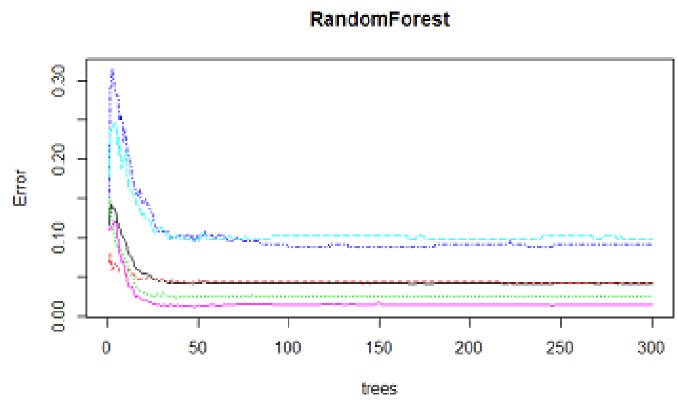

Fig 2. *Random Forest Plot*

Naive Bayes is a supervised machine learning algorithm based on Bayes theorem with an assumption that any particular predictor feature is not related to the other features within a class. Naive Bayes model is easy to build, convenient for multiclass prediction with large datasets, and also outperform various other sophisticated classifiers. Bayes Theorem helps in evaluating the posterior probability P(A|X) from (P(X|A)* P(A)) / P(X), where P(A|X) is the posterior

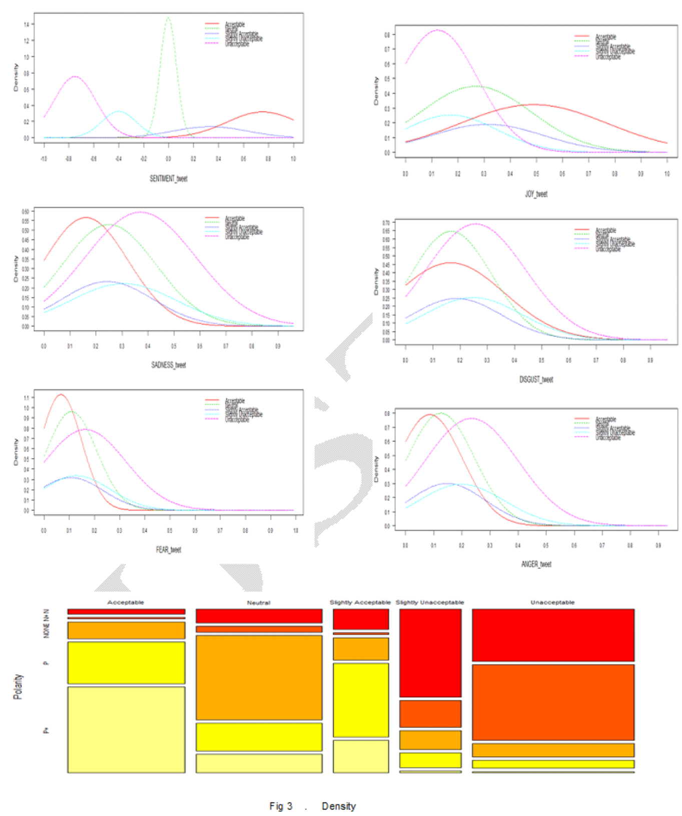

Fig 3*. Density Plots*

probability, P(A) is the predictor's prior probability, P (X|A) is the probability of predictor given class and P(X) is the class prior probability. In this study, a Naive Bayes model has been developed which provides an accuracy of 79.88%. While developing a model we drew a density plot for all seven predictor variables using Naive Bayes function which is shown in figure 3. A Multilayer Perceptron model is used for building a classification model using Keras and Tensorflow packages. This model contains multiple layers, whose computation units were connected in a feed-forward way. The model splits the data into a 70/30 ratio, 70 percent data for training, while 30 percent of the data for validation. The altogether model trained for 200 epochs, however after 150 epochs model's accuracy remains constant at 76.75%. The loss and accuracy of the model are shown in figure 4.

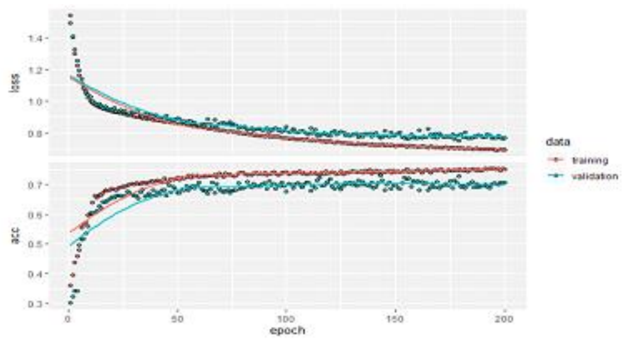

Fig 4. *Accuracy and loss plot for training and validation data*

The accuracy, precision, recall and f1 score were calculated that how efficiently feature set helped in filtering out the non-credible content from the credible one. Initially, the data set were partitioned into 70-30 training and testing sets. The overall statistics for machine learning classifiers are mentioned in Table 4. All of the ML algorithms are giving high accuracy with an acceptable level of f1 score; however, Random Forest gives the best results with 95.79% accuracy with 96.15% area under the curve followed by SVM, Naive Bayes, and MLP.

TABLE 4
OVERALL STATISTICS FOR ML CLASSIFIERS

ML Techniques	Accuracy	AUROC
Random Forest	0.9579	0.9615
SVM	0.9151	0.9346
Naive Bayes	0.7988	0.8515
Multi-Layer Perceptron	0.7675	0.8014

Furthermore, performance measures of the ML Techniques were evaluated by calculating Precision, Recall, F1 score, and AUROC. Table 5 discussed the performance measures of the implemented classifiers. Acceptable level of precision, recall and f1 score for all the credibility classes have been found. However, for slightly-acceptable class, these three performance measures are exceptionally low with Naive Bayes and MLP classifiers.

RANDOM FOREST

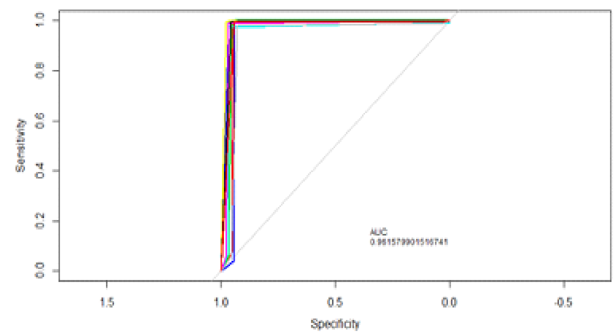

Fig 5*. ROC Curve Random Forest*

MULTILAYER PERCEPTRON

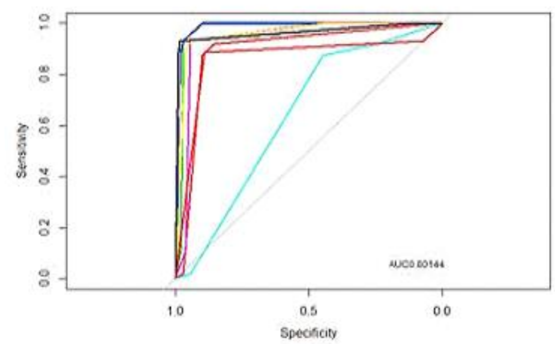

Fig 6*. ROC Curve Multi-layer perceptron*

SUPPORT VECTOR MACHINE

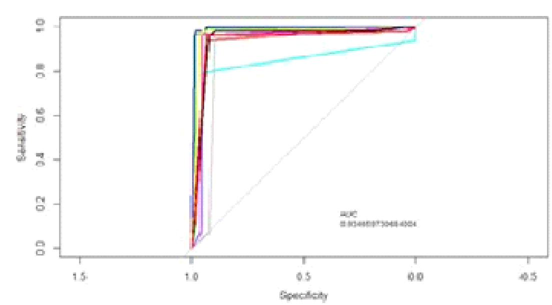

Fig 7*. ROC Curve Support Vector Machine*

Naive Bayes

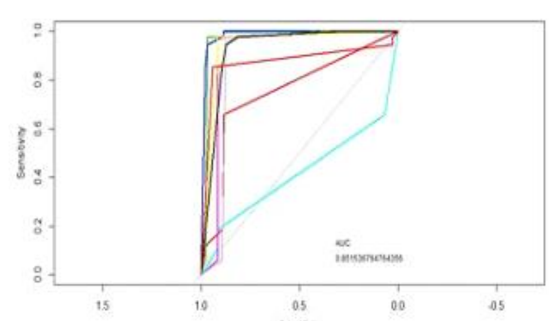

Fig 8*. ROC Curve Naive Bayes*

The figures (5, 6, 7 & 8) represent the Multiclass ROC curve which has been drawn for finding the performance of the classifiers which were classifying the tweets into five credibility classes. Whereas AUC is the area under the ROC curve, value ranges from 0 to 1, closer the value is to 1 higher will be the prediction performance.

TABLE 5

PERFORMANCE MEASURE OF MACHINE LEARNING CLASSIFIER

		Credibility Classes				
ML Techniques	Performance Measures	Credible	Somewhat Credible	Neutral	Somewhat Uncredible	Uncredible
Random Forest	Precision	0.9914	0.9807	0.9795	0.8981	0.9323
	Recall	0.9392	0.9026	0.9677	0.9238	0.9939
	F1 Score	0.9646	0.94	0.9736	0.9107	0.9622
Support Vector Machine	Precision	0.9333	0.952	0.6923	0.8974	0.9598
	Recall	0.9613	0.8981	0.8901	0.8898	0.912
	F1 Score	0.9471	0.9242	0.7788	0.8936	0.9353
Naive Bayes	Precision	0.8199	0.147	0.8106	0.7642	0.86
	Recall	0.8842	0.0588	0.8458	0.9145	0.8543
	F1 Score	0.8508	0.084	0.8278	0.8326	0.8571
Multilayer Perceptron	Precision	0.8547	0.2032	0.8544	0.8203	0.8857
	Recall	0.7936	0.5319	0.6925	0.84	0.9117
	F1 Score	0.823	0.2941	0.765	0.83	0.8985

For evaluating the accuracy of the dataset, the 10 fold cross-validations were performed for all four ML classifiers. The largest accuracy value should be considered to select the optimal model. The final evaluated results for the given models are mentioned in Table 6.

TABLE 6

10- FOLD CROSS VALIDATION

ML Classifiers	Accuracy
Random Forest	0.954352
SVM	0.948621
Naive Bayes	0.956804
Multi-layer Perceptron	0.817598

For analyzing features, box plots have been drawn, that tells about which feature will fall under which credibility classes. It is clearly shown in figure 9 that all the tweets which fall into the uncredible or somewhat uncredible classification category have a higher percentage of negative emotions like anger, sadness, fear, and disgust whereas, the tweets fall into credible or somewhat credible have a higher percentage of joy emotion. Tweets having negative sentiment and polarity will fall into the category of uncredible or somewhat uncredible credibility classes whereas, tweets having positive sentiment and polarity fall into the credible or somewhat- credible class categories.

5 CONCLUSION AND FUTURE SCOPE

The flexibility and easiness that Twitter provides for their users are incomparable, however, while sharing effective and relevant information, sometimes fake and uncredible content is also disseminated with the real content. These contents are a big threat to the society, as it contains abusive words or hate speech towards any caste, culture, society, religion or political party. The objective of this paper is to filter out all these content from online social networks that can cause big damage and create societal division. In this paper, 35k tweets were crawled with the help of trending hashtags and handles, from various high impact events, happened in India. These tweets were further preprocessed and narrowed down to the finest 3.5K tweets. The aim of this paper is to develop an emotion-based model that can effectively filter out uncredible content from credible ones not only posted by general public but also by these reputed/affluent users, which provide benefits to the society in long run. A machine learning classification model has been developed which is based on the feature set to classify the tweets into different credibility classes. A feature set is divided into two major categories, features which were extracted from the Twitter like followers count, favorite count, user verified, user-created at, tweets created at, user's description is present or not and tweet is possibly sensitive. These features helped in preprocessing the tweets at the initial level. Whereas, emotions, sentiment, and polarity features are fall under the

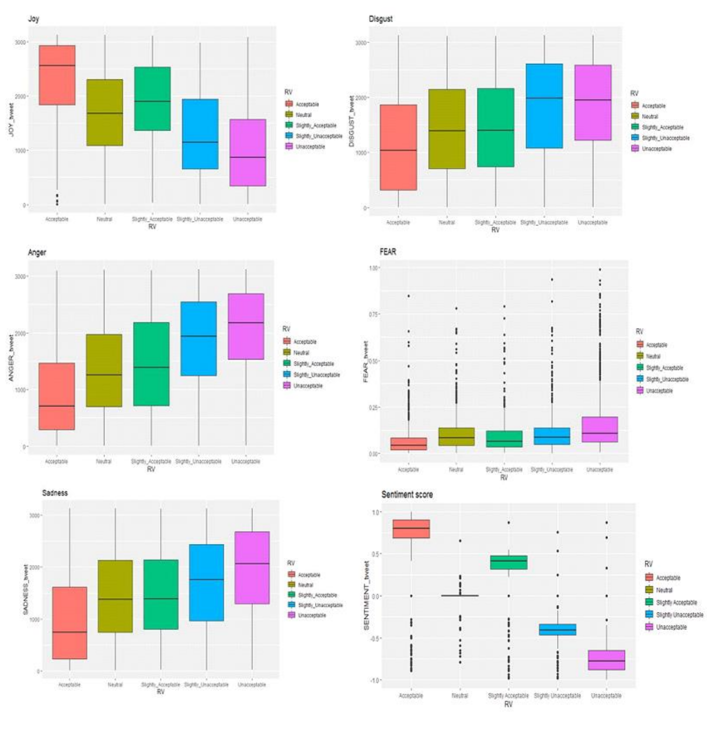

Fig 9: *Box Plot Analysis*

extended features set category that helps in evaluating the content credibility through machine learning models. Further, tweets were outsourced to six human annotators who have in-depth knowledge about every event for which tweets belong.

Only those tweet labels were taken into consideration that was annotated by three on more experts at same credibility scale. Finally, for classification purpose, four different classification algorithms have been used, namely Multilayer perceptron, Naive Bayes, Random Forest and Support Vector Machine. The best classification result of 95.79% accuracy is given by the Random Forest, followed by SVM with 91.51%, Naive Bayes with 79.88%, and Multilayer Perceptron with 76.75% respectively. The acceptable level of accuracy, precision, recall, and f1 score was observed for all given credibility classes w.r.t all four classification algorithms. For future perspectives, the work on finding the credibility of images, memes, and videos will be taken into consideration. Moreover author will also try to find the sarcasm associated with such uncredible media contents.

6 REFERENCES

[1] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring user influence in twitter: The million follower fallacy. Icwsm, 10(10-17), 30. (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)(pp. 91-100). IEEE.

[2] Castillo, C., Mendoza, M., & Poblete, B. (2011, March). Information Credibility on Twitter. In Proceedings of the 20th International Conference on World Wide Web (pp. 675-684). ACM.

[3] Gupta, M., Zhao, P., & Han, J. (2012, April). Evaluating event credibility on twitter. In Proceedings of the 2012 SIAM International Conference on Data Mining (pp. 153-164). Society for Industrial and Applied Mathematics.

[4] Morris, M. R., Counts, S., Roseway, A., Hoff, A., & Schwarz, J. (2012, February). Tweeting is believing?: understanding microblog credibility perceptions. In Proceedings of the ACM 2012 conference on computer supported cooperative work (pp. 441-450). ACM.

[5] Xia, X., Yang, X., Wu, C., Li, S., & Bao, L. (2012, May). Information credibility on twitter in emergency situation. In Pacific-Asia Workshop on Intelligence and Security Informatics (pp. 45-59). Springer, Berlin, Heidelberg.

[6] Westerman, D., Spence, P. R., & Van Der Heide, B. (2012). A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. Computers in Human Behavior, 28(1), 199-206.

[7] Gupta, A., Kumaraguru, P., Castillo, C., & Meier, P. (2014, November). Tweetcred: Real-time credibility assessment of content on twitter. In International Conference on Social Informatics (pp. 228-243). Springer, Cham.

[8] Zhang, Q., Zhang, S., Dong, J., Xiong, J., & Cheng, X. (2015, October). Automatic Detection of Rumor on Social Network. In Proceedings of the 4th National CCF Conference on Natural Language Processing and Chinese Computing (NLPCC) (pp. 113-122). Springer International Publishing.

[9] Mitra, T., & Gilbert, E. (2015, May). CREDBANK: A Large-Scale Social Media Corpus with Associated Credibility Annotations. In ICWSM (pp. 258-267).

[10] Alrubaian, M., Al-Qurishi, M., Hassan, M. M., & Alamri, A. (2016). A credibility analysis system for assessing information on twitter. IEEE Transactions on Dependable and Secure Computing, 15(4), 661-674.

[11] Lorek, K., Suehiro-Wiciński, J., Jankowski-Lorek, M., & Gupta, A. (2015). Automated credibility assessment on Twitter. Computer Science, 16(2)), 157-168.