International Journal of Technology and Human Interaction
Volume 19 • Issue 1

Classification of Tweets Into Facts and Opinions Using Recurrent Neural Networks

Murugan Pattusamy, University of Hyderabad, India*
Lakshmi Kanth, University of Hyderabad, India

ABSTRACT

In the last few years, the growth rate of the number of people who are active on Twitter has been consistently spiking. In India, even the government agencies have started using Twitter accounts as they feel that they can get connected to a greater number of people in a short span of time. Apart from the social media platforms, there are an enormous number of blogging applications that have popped up providing another platform for the people to share their views. With all this, the authenticity of the content that is being generated is going for a toss. On that note, the authors have the task in hand of differentiating the genuineness of the content. In this process, they have worked upon various techniques that would maximize the authenticity of the content and propose a long short-term memory (LSTM) model that will make a distinction between the tweets posted on the Twitter platform. The model in combination with the manually engineered features and the bag of words model is able to classify the tweets efficiently.

KEYWORDS

Classification of Tweets, Long Short-Term Memory Model, Recurrent Neural Networks

INTRODUCTION

Given an option, every individual wants their opinions to be heard and accepted. To accommodate this need, social networking platforms such as Facebook, Twitter, and Telegram, etc. mark their space in the online market. Every platform offers individuals the opportunity to post as much content as they wish. In order to make the post unique, there are high chances that the information shared by the individual will be biased with their opinions than the underlying facts. The need to classify the facts from opinions is therefore essential. The opinions and facts when channelized have got the potential to generate their sentiments. Hence, it is the responsibility of the platform provider to differentiate between facts and opinions to ensure that panic does not prevail in the community (Chatterjee, Deng, Liu, Shan, & Jiao, 2018).

In the past years, the number of people who are active on Twitter has been consistently spiking. Despite having many competitors, Twitter is a widely used marketing tool. In India, even the government agencies have started using the Twitter account as they can get connected to a greater number of people in a short period. Credit to the technological advancements, whatever happens at

DOI: 10.4018/IJTHI.319358 *Corresponding Author

any place on the globe, it gets cascaded to every other part of the globe. With this, there is a plethora of content that is being generated. On an average, every second, around 6000 tweets are emerging, which corresponds to over 3,50,000 tweets per minute, 500 million tweets per day and around 200 billion tweets per year (Hasan, Orgun, & Schwitter, 2019). Interesting insights can be obtained through this data. At the same time, it is desirable to eliminate data points that have opinions. It is crucial that before gaining insights from the tweets, it is beneficial to differentiate the tweets based on their authenticity by considering the person who is tweeting (Deng, Sinha, & Zhao, 2017; Wiebe & Riloff, 2005; Wright, 2009). Dealing with such a humongous volume of data needs much effort. With the advancements of big data technologies and also with the enhanced computational power, dealing with such a variety of data, growing at a rapid pace is convenient. If there is less authenticity in a particular tweet, it may comprise of personal belief or the sentiment of the person.

Understanding both the opinions of the individuals and the facts around the subject has got its business opportunities. In order to tap this potential, the initial step would be to differentiate between the opinions and the facts. The semantics of the tweets should be analyzed before understanding the sentiment of the tweets. After obtaining the sentiment of the tweets, categorize them into their respective classe (opinion or fact). In this research work, the tweets that were related to the airstrike carried out by India in retaliation to the attack on the Indian CRPF soldiers at Pulwama have been considered. This data is analyzed because the situation was panic-driven as the whole of television broadcasting was emphasizing upon this subject.

Moreover, there was an election fever that was picking up in India around the same time. Also, a solution of this sort can be applied to various other instances dealing with varied subject areas. Interestingly, the approach can be extended to other platforms (such as WhatsApp, Instagram) as well.

To address this particular problem statement, the study demonstrates a new algorithm that classifies the authentic tweets from the opinions shared through tweets. In this process, a set of features are manually generated, which enables differentiating the tweets effectively and efficiently. This serves the purpose of supervising the activity that we are performing. These manually engineered features will then be combined with the Bag of Words (BOW) model generated as part of the Natural Language Processing (NLP). After combining the features explicitly, we then use the Long Short Term Memory (LSTM) network, which is an extension of the RNN model (Goodfellow, Bengio, & Courville, 2016). We benchmark the performance of the LSTM network using a labelled dataset (test dataset) and compare its results with other popular and relevant models (Evermann, Rehse, & Fettke, 2017; Ghiassi, Zimbra, & Lee, 2017; Tumasjan, Sprenger, Sandner, & Welpe, 2010; Wang, Wang, Li, Abrahams, & Fan, 2014; Wiebe & Riloff, 2005).

This study makes contributions to (1) understand the importance of distinguishing the authentic tweets from mere opinions shared by the people on the twitter platform, (2) develop a deep learning model by combining the two different types of feature sets to classify the tweets from Twitter, (3) project the best fit model for the given dataset for the purpose of sentiment analysis of the data, and (4) contrast the significance of the finalized model in the present-day situation. It is known that LSTM is good at handling quasi data. Hence, the hypothesis as LSTM is the best model for sentiment analysis is formulated. Though research has been done earlier with LSTM in accordance with textual data. The uniqueness of this work comes from integrating the LSTM model with the BOW features and manually engineered features.

RELATED WORKS

There is a wide range of work that is presently being carried out in this particular domain. Notably, in the last few years in this decade, an enormous number of applications have popped up in this area. These applications can be categorized into the following: event detection, semantic analysis and sentiment analysis. On the whole, much work is being carried out extensively to understand the social media data (tweets in this context) and get the facts right. Interestingly, the problem statements picked up by the researchers are primarily related to either solving business problems using the technological advancements or comparing various technical approaches in handling a generic approach. In that context what we observed was that the problem of understanding the authenticity of the messages posted on Twitter has still got the scope to be enhanced with the implementation of the advanced technical approach like LSTM for addressing a bigger purpose.

Event Detection

Twitter has been fast emerging in recent years. Users are using Twitter to report real-life events. Although event detection has long been a research topic, the characteristics of Twitter make it a non-trivial task to deal with the identification of such an event. Tweets reporting such events are usually overwhelmed by high flood messages. Moreover, the event detection algorithm needs to be scalable, given the sheer number of tweets. This data is beginning to be used as a basis for detecting, monitoring and analyzing the characteristics of both natural and human-made disasters (Cheng & Wicks, 2014). By taking advantage of both the speed and coverage of Twitter, the events can be detected in a timely manner. The tweet is often associated with spatial and temporal information. Based on this t when and where an event happens. E.g., via monitoring an incoming tweet "Shooting outside the Irving mall." At 2:38 pm on July 24, the event is detected immediately, and the location and time of the crime are also extracted (Li, Lei, Khadiwala, & Chang, 2012; Luo, Zhang, & Duan, 2013; Tumasjan et al., 2010; Wright, 2009; Zhao et al., 2014). TwitterNews+, an event detection system that incorporates specialized inverted indices and an incremental clustering approach to provide a low computational cost solution to detect both significant and minor newsworthy events in real-time from the Twitter data stream (Hasan et al., 2019). Though this process is useful, it does not handle the aspect of distinguishing facts from non-facts (Aggarwal, Gopal, Gupta, & Singh, 2012; Bollen, Mao, & Zeng, 2011; Evermann et al., 2017; Ghiassi et al., 2017; Kraus & Feuerriegel, 2017).

Semantic Analysis

It is clear that from the event detection analysis, that classification of tweets into different classes is not possible. We can only understand that something has happened. In order to dig down to another level, we need to understand the semantics involved in the tweets (Landauer, Foltz, & Laham, 1998). Through this process, we will be able to understand the meaning of a given word based on the context in which it is used. It is important to understand the word and also the context in which that particular word is being used in order to make a meaningful insight. Part-of-speech (POS) tagging is essential for a variety of applications such as parsing, information extraction, and machine translation. Dialectal POS tagging is becoming increasingly important due to the ubiquity of social media, where users typically write in their dialects to match how they speak in their daily interactions. Dialectal text poses exciting challenges such as lack of spelling standards, the pervasiveness of transformative, morphological operations, such as word merging and letter substitution or deletion, in addition to lexical borrowing from foreign languages. The rationale for the separation is that different dialects have different affixes, make different lexical and word ordering choices, and are influenced by different foreign languages (Cheng & Wicks, 2014; Darwish et al., 2018; He, Wu, Yan, Akula, & Shen, 2015).

Sentiment Analysis

Understanding the meaning of the word based on the context is only halfway through in our journey (Thelwall, Buckley, & Paltoglou, 2012). In order to classify the tweets either into facts or opinions, the sentiment of every tweet should be understood. To do this, each tweet should be considered independently. Then, each word in the tweet has a separate feature. From this, we can get the Bag of Words (BoW) model (Ghiassi et al., 2017). The sentiment analysis on social media is challenging in nature as the complexity involved with the data is high. Apart from the complexity, another aspect is the availability of the data.

For the purpose of extracting features, there are many techniques. Amongst the available techniques, the predominantly used technique is BOW model. In this, various features are obtained corresponding to each document. The objective of this model will be to identify exhaustively (to the extent possible) and mutually exclusive features (Chatterjee et al., 2018; Deng et al., 2017; Sahu & Khandekar, 2020).

Aspect-Level Sentiment Classification

This is a fine-grained task in sentiment classification, aiming to extract sentiment polarity from opinions towards a specific aspect of a word that has made tremendous improvements in recent years. There are three critical factors for the aspect-level sentiment classification: contextual semantic information towards aspect words, the correlation between aspect words and their context words, and location information of context words concerning aspect words (Shuang, Ren, Yang, Li, & Loo, 2019). This area is exciting but lacks adequate data to understand how efficient it is in addressing specific real-time scenarios.

RESEARCH GAP

Though there are existing tools and applications to sort out the tweets from Twitter into facts and opinions, the precision with which they are being classified is not satisfactory. In this context, there is a scope to come up with a new technique which can be exposed as API to bring out the best of the classification possible, which can minimize errors. The objective of this research will be achieved if a new technique is proposed, and which will give higher accuracy compared to other deep learning techniques and other machine learning algorithms. In order to propose that, the data from Twitter is considered as this is the platform which has gained much significance over a period of time.

IDENTIFYING THE AUTHENTICITY OF THE TWITTER MESSAGES

In this section, the implementation aspects of the model used to identify if a particular tweet is authentic or not is elaborated. BOW model might not be ideal, because all the tweets might not be authentic. For this purpose, manually engineered features are used to validate the source. This will help to overcome drawbacks of BOW. Hence, both kinds of features (BOW & manually engineered) will be used in segregating facts and opinions. However, the way the features should be used together needs to be handled systematically. The reason is: if the manually engineered features are directly combined or merged with the BOW features, there are high chances that the model built will be biased towards the BOW features. To minimize the bias and efficiently classify the tweets, each set of features is fed into the model through different layers and then combined at the hidden layer. Nevertheless, in the proposed approach, we are using RNN. RNN can bring in the impact of multiple layers of DNN into a single layer. Therefore, it is more effective in terms of performance when compared to that of the DNN. Also, the impact of manually engineered features is not lost when they are combined directly with the BOW features. The combination of the BOW features and the manually engineered features together will give us the complete collection of features corresponding to the tweets. With RNN, there is a concept of internal memory; RNNs can remember important things about the input they receive, which enables them to be precise in making the prediction (Goodfellow et al., 2016).

The RNN-LSTM Model

In contrast to the DNN, the RNN makes the decision based on the present input and also based on the previous results and the weights accordingly. This is possible because the information is passed

Figure 1. Flow chart of proposed work

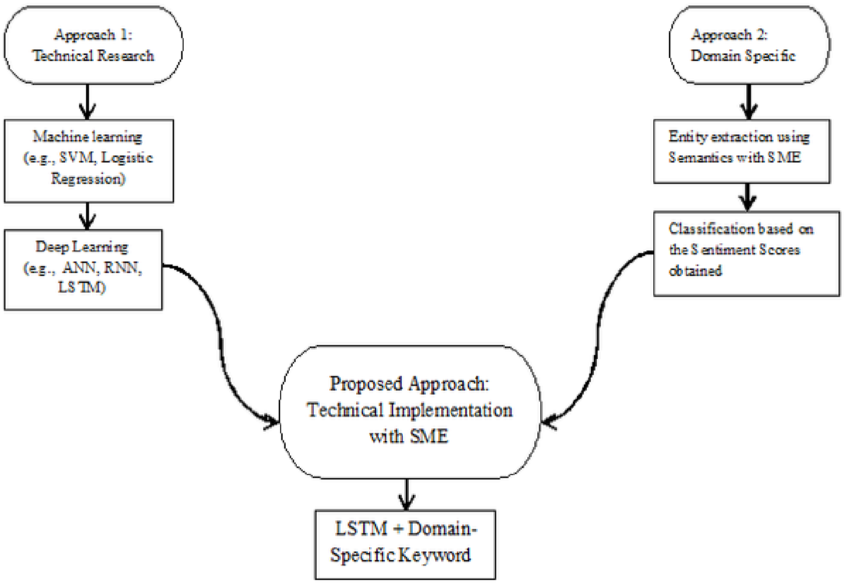
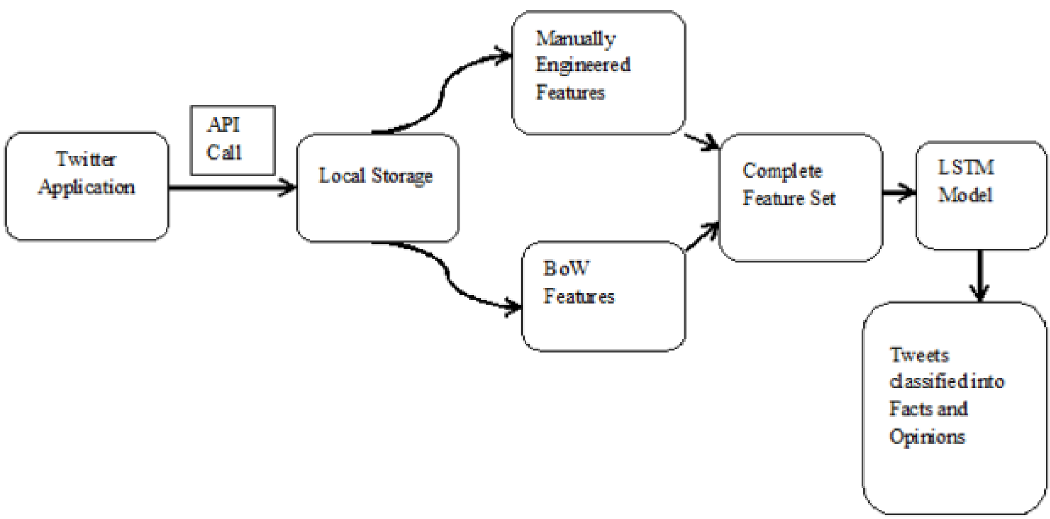


Figure 2. Analysis plan



in a loop in the RNN model. Through this, the appropriate weights are assigned to the inputs based on which the results are obtained. On the other hand, in the case of DNN, the data of the previous cycle is not considered while generating the output (Goodfellow et al., 2016).

The recurrent cycle considers both the present and past values. Here, it is the immediate past. This aspect differentiates the model from DNN because the model gets trained based on the sequential data that is passed onto it. The advantage of RNN over DNN is that the back-propagation is automatically handled in the case of RNN. However, the limitation with RNN is that if the change between weights from one iteration to another iteration is less, then the time it takes to compute will be high. To overcome this limitation, we use the LSTM model, which is an extension of the RNN. In the LSTM, there will be a memory unit which stores the results for faster computations (Abdel-Nasser & Mahmoud, 2019; Goodfellow et al., 2016; LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015). The storage in the memory is done based on the importance of the input, which can be decided based on the weights assigned to them. The significant advantage of LSTM over others is that it maintains a constant error back-propagation within the memory cells. This results in its ability to bridge long time gaps, especially in handling textual data. The LSTMs can also handle noise, distributed representations and continuous data. LSTM is also local with both time and space, i.e., it would be quick in generating the results, and also it would implement optimal utilization of the processing space, hence making the processing quicker.

In this research, 300 nodes have been used, and the dropout value of 0.5 is fixed to ensure that there is not overfitting of the model (Farquad & Bose, 2012). To get the output, a dense layer with activation function of sigmoid is defined. In this study, the accuracy rate of RNN+LSTM model is compared with Deep Neural Network, Support Vector Machine, Random Forest algorithm, Logistic Regression and Naïve Bayes classification algorithm.

The python code corresponding to it is shown in Table 1.

The libraries that are used in the processing of implementing the proposed approach are shown in Table 2.

Table 1. Python code

```
RNNB1 = Sequential()
RNNB1.add(LSTM(300,input_shape=(1,X.shape[1]),return_sequences=True))
RNNB1.add(Dropout(0.5))
RNNB1.add(LSTM(300))
RNNB1.add(Dropout(0.5))
RNNB1.add(Dense(Y.shape[1], activation = 'sigmoid'))
RNNB1.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics=['accuracy'])
```

Table 2. Libraries used to implement the proposed approach

```
import os
os.environ['KERAS_BACKEND'] = 'theano'
import pandas as pd
from keras.models import Sequential
from keras.layers import Dense, Merge, Dropout, LSTM
from keras.layers.normalization import BatchNormalization
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score
import numpy as np
from numpy.random import seed
import warnings
warnings.filterwarnings("ignore")
from joblib import load,dump
import sys
sys.setrecursionlimit(10000)
```

METHODOLOGY

Manually Engineered Features

In order to ensure robustness in the model, it is essential to explore features that are external to the text. These features can be extracted from the source of the tweets, based on the user's profile, geographical location and other parameters. For effective use of the manually engineered features, the features are binary-coded. This minimizes the task of encoding the data explicitly at a further stage.

Feature 1: Title Capitalization

In the context of tweets, behavioural aspects play a prominent role in identifying the tweets. The messages posted with facts are usually formal in their narration and free of errors. The indicator for such messages is through the use of capital letters at the start of each word that they make as part of their title. The number of words with only the first letter in uppercase are calculated. If more than three words satisfy the criterion, we assign a 1 for this binary feature (Chatterjee et al., 2018).

Feature 2: URL

Usually, facts also refer to the source from which they have obtained the information. Though this is for a referential purpose, this speaks of the integrity of the message. To count on this aspect, weightage is given to the mention of URL or any other references in the text.

Feature 3: Followers

Genuine people always have their mark irrespective of the platform on which they are. It is the same notion that works in this context. People prefer accurate information over false information. These are the kind of people who have a greater number of followers compared to people who spam the platforms. Therefore, the followers' count of the user is a significant identifier in the evaluation of a text into a fact or not a fact. The benchmark for the number of followers is considered to be 500.

Feature 4: Numbers

It is noted that most of the facts are quoted with stats or numerical values corresponding to it. Therefore, a text comprised of numbers is an interesting indicator to be considered as a fact or not a fact. Thus, we check if the given text has numbers or not.

Feature 5: Repeating Characters

When the intent is to share an opinion, there will be an informal way in which the text will be entered. There will be a repetition of characters, e.g., "….!!!". These kinds of representations usually reflect that the user is only sharing his opinion instead of a fact. For this variable, if any variable repeats more than twice, then the value will be one corresponding to this variable.

Feature 6: All Uppercase

Similar to the repetition of the characters, presenting the whole of the text in uppercase letters is also an indication that the posted text is an opinion.

Feature 7: Twitter Terms

For every platform, there are certain terms that are specific to that platform. Similarly, in twitter, there are a few terms that are used (Appendix A). If any of these words from the list appears in the text, there will be fewer chances for it to be categorized as a fact.

Once the manually engineered features are made, the Bag of Words model is built using the Count Vectorizer. In the count vectorizer, the number of features is restricted to 1000. However, the rationale behind choosing 1000 features is to minimize the sparseness in the data. The features thus obtained are combined with the labelled manually engineered features to get a holistic data set.

This data set is then divided for training the model and testing the model. The proportion of split is 80:20 respectively. Once the data segments are obtained, we train the model using the training data and validate the model using the testing data.

Evaluation of the Intended Model

This evaluation of the model is done by comparing the proposed model with other models which we build.

DATA

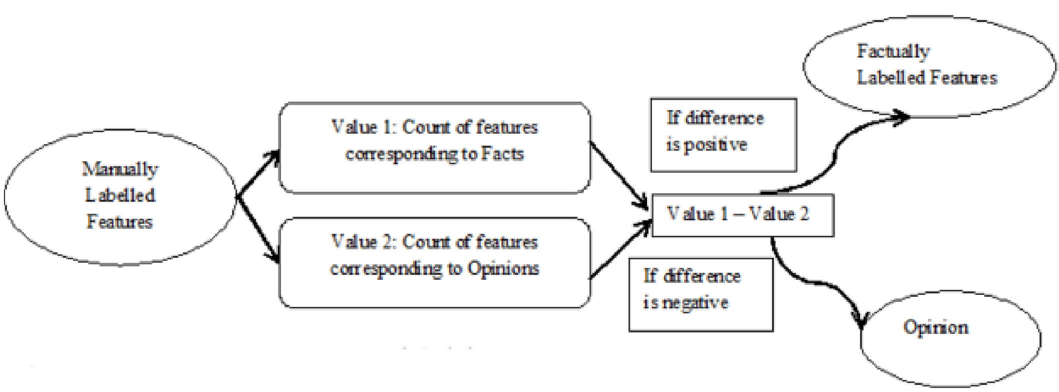
Twitter has an API through which the tweets can be extracted based on a token or keyword. In this context, the token is "#IndiaStrikesBack". After the tweets have been obtained; as part of pre-processing, only the tweets that were tweeted in the English language were considered, and the tweets tweeted in other languages were omitted. As a result, we got 12489 unique tweets.

Data Labeling

After the tweets for analysis have been obtained, it is important to label the tweets either as a fact or as an opinion. This kind of labelling is important because this is the basis upon which the sentiment analysis will be done. One way to deal with this is to do the analysis manually. However, doing it manually for all the 12489 tweets is tedious and time taking. Therefore, a technique based on manually engineered features is used (Chatterjee et al., 2018). From the manually engineered features, the attributes corresponding to a fact and opinion are obtained. The weights corresponding to a fact and an opinion for an individual tweet were calculated. Then, these weights are aggregated (for fact and opinion) as these are just binary encoded. If the count of the fact-based attributes is more, then it is segmented into the class of fact, else it is segmented into the class of an opinion. This is also done in such a way that there is a balance between the classifications and also to ensure that there is no bias in the classification process. At the end of this, there are 4534 tweets categorized as facts and 7955 tweets categorized as opinions.

This method of labelling will give results on par with the manual labelling process. However, when a person starts labelling the messages manually, there is a high chance that there will be bias in his results. However, there are many other options through which we can label the text data. One such approach is using marketing tools like Snorkel in which there will be built-in labelling functions

Figure 3. Classification of facts and opinions



(LF). These LF can also be customized as per the subject which is being considered. Nevertheless, it requires much overhead in implementing.

Baselines

After understanding various research topics and referring to other models used earlier, the following baseline models have been made:

1. **RNN+LSTM:** This is a special version of RNN which overcomes the pitfalls of the RNN. In this version, the memory functionality is enhanced to ensure that the output of the current input is retained with the help of added gates.
2. **Deep Neural Networks (DNN):** This is a neural network model in which the weights of the features are adjusted through feed forward propagation and backward propagation. The differentiating point between the LSTM and the DNN is the architecture where there will be a recurrence of the functionality in the RNN along with memory aspect which is not the case with DNN
3. **Support Vector Machines (SVM):** This is a classification technique that is used to differentiate the data points that are passed into it. In this technique, the objective will be to choose a hyperplane which can distinguish the data in such a way that the boundary formed from either side of the entities to the hyperplane is maximum.
4. **Logistic Regression (LR):** Logistic regression is also a classification technique which is used to classify the data points into different classes. This is an extension to the linear regression where the dependent variable is categorical in nature.
5. **Random Forest (RF):** In the random forest classifier, there will be multiple trees that will be built with the square root of the number of features existing. Along with this, there will be a bootstrapping mechanism, which resembles the replacement of the randomly chosen samples which are used for model training that will be implemented. This is an ensemble model that bags the decision trees to ensure that there is no overfitting.
6. **Naive Bayes (NB):** This is a supervised learning algorithm that is based on the Bayes theorem, which focuses on the posterior probability. Compared to the other techniques, this is simple to implement, especially when the context is comprised of text.

The average measures obtained by applying the mentioned algorithms (SVM, NB, LR, RF) indicate that the SVM algorithm outperformed. From the average precision results of SVM, NB LR and RF, it is observed that the SVM algorithm outperforms the other algorithms. The SVM, LR and RF classifiers did not show much difference in the performance when the BOW features are used and when the BOW features and manually engineered features are used. In contrast to these results, when the DNN approach is used, there is an increase in the accuracy (LeCun et al., 2015; Mikalef, Pappas, Krogstie, & Giannakos, 2017; Sahu & Khandekar, 2020; Schmidhuber, 2015). With this, we can say that the manually engineered features play a prominent role in performing the sentiment analysis on the text. The naïve-Bayes algorithm is used because it is more suited for text classification as it will converge more quickly than discriminative models like logistic regression. Therefore, for Naïve-Bayes, we will need less training data to build a model. The Random-Forest classifier will handle the missing values and maintain the accuracy of a large proportion of data. If there are more trees, then it will not even allow overfitting. However, on the contrary, there is a random forest with more trees and more depth that will require a lot of computation power, and the latency will be high (Sahu & Khandekar, 2020). Before exposing the data to the baseline models, it was ensured that the data is balanced, which helps in minimizing the bias in classifying the tweets (Chatterjee et al., 2018).

Figure 4. Twitter slangs

#FF	FOE	LAGO	OH	TGTSIOI
AAP	FTW	LMFO	OTB	TMB
BM	HBL	LOT	PPP	TQRT
CBOT	HWEVR	LYS	REDLYFE	TWIT
DNA	IOT	MOV	RTHX	TYFF
F4F	JGH	MVO	SML	WBOS
FMOT	LA	NEWEETER	SWEEPLE	WMBY
FTC	LLBLOG	OOYF	TCAT	WTFS
HAB	LOML	POLITWEE	TGIC	XTB
HT	LU	QOTP	TJP	YOCO
IMF	MIMI	RT	TPL	BTW
JATA	MUVA	SA	TWARS	CHK
KYR	NEJ	SMHD	TWTR	BFN
LGY	OOTT	SWAG	WATN	OMG
LOL	PMOT	TC	WMBU	OMG
LTOTD	PWAT	TGFAD	WTF	ZZZZZ
MISTWEET	RQB	TINGLE	WWTT	FCUK
MULC	S2G	TON	YOBK	DA
NEF	SHXP	TW	YTS	FAB
OOMF	STOW	TWITTERE	AYEG	F2F
PART	TBT	VMA	BTW	ETA
PTP	TEOG	WKYP	DM	EMA
RLRT	TIL	WTB	ELW	IDK
S/S	TNJ	WWD	FGF	IC
SFTW	TTYLXOX	YNI	FTAD	KK
STBH	TWITT	YOLTBTST	GSTG	NTS
T1Z	UOT	775	HR	TWART
TCOT	WGD	AWS	IGERS	
THOT	WPR	BTTP	IWTKCR	
TMC	WWBA	DETWEET	KOT	
TSS	YIF	EHT	LFL	
TWITE	YOLO	FDU	LOAL	
UAYA	TV4	FSTOW	LRT	
WDILY	AGIG	GLWS	MFBTY	
WOMW	BTTE	HMU	MT	
WTWT	CMSU	IDTS	NCBD	
YBM	EHE	ITGFT	OLLI	
YOFO	FBTL	KFB	PAP	
#LESM	FSF	LBOD	PRT	
AF	GBP	LMR	RIH	
BOL	HDM	LPC	RTZ	
CCW	IDGHP	MBF	SERPT	
DWEET	IT	MRT	SPEET	
FATC	JK	MYWB	TCIC	

RESULTS

For the SVM model built, the kernel used is a linear kernel. The reason for using this kernel is that it will give the best results for text classification. The accuracy obtained through this is 94.7%. For evaluating the logistic regression model, cross-validation of 5 was used. The reason for using five instead of 10 is that if this is increased, there are high chances of over-fitting of the data. Also, it will consume more computation time. The accuracy obtained through this model is 81.9%. With Random Forest, we can get more accuracy if we increase the number of trees and the depth of the forest but to evaluate if it suits our problem statement, we cannot have more trees and maximum depth. The maximum depth considered in this case is two for which we obtained the accuracy of 79.7%. When we tried out the traditional Naïve-Bayes technique, which is the simplest of all, the accuracy was not so high compared to the other models. It yielded only 69.2% accuracy. However, the computational time is very less compared to all the other models. Amongst all these, the DNN was earlier proposed to be the best model for a combination of BOW + manually engineered features. Similarly, in this context, it brought an accuracy of 97%.

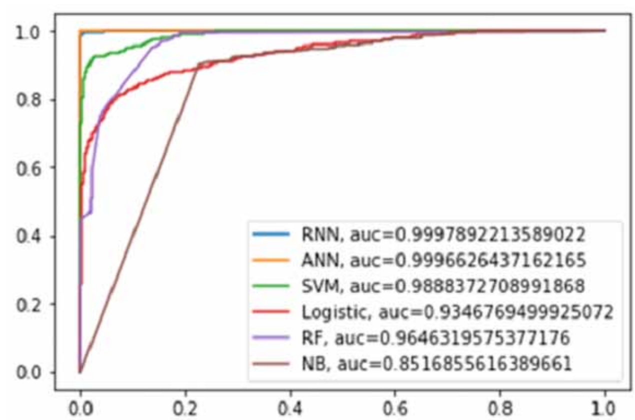
Nevertheless, DNN will not be giving consistent results as the weights keep changing for every forward and back-propagation that we do. The RNN gave a stand-out accuracy of 98%. Though the difference between DNN and RNN is not significant, RNN will be more consistent in delivering its results compared to those of DNN. Also, if the volume of data grows, then RNN will become robust in making its predictions. On the contrary, if the data size continues to grow, then DNN might need more memory consumption. RNN uses the LSTM network to overcome this problem, giving itself an edge over the DNN.

In the above graph, area Under the Curve (AUC) is the metric that is used to evaluate the performance of the model. As the AUC value for the RNN based model is more, it is the recommended

Table 3. Model accuracy rate

S.No	Model	Accuracy (in%)
1	Recurrent Neural Network + Long Short Term Memory Model	98
2	Deep Neural Network Model	97
3	Support Vector Machine Model	94.70
4	Logistic Regression Model	81.90
5	Random Forest Model	79.70
6	Naïve Bayes Model	69.20

Figure 5. ROC curve



model for sentiment analysis. The AUC quantifies the ability of the model to distinguish between the two classes that are considered.

DISCUSSION

In this study, a Long Short Term Network, which is an extension of RNN is proposed to classify the Twitter messages into facts and opinions (Chatterjee et al., 2018; Deng et al., 2017; Ghiassi et al., 2017; Hasan et al., 2019), thus, helping in getting to know the authenticity of the tweets. This kind of classification is really important because, with the increased access to social media and digital platforms, everyone is in a state of a hurry to obtain prominence. With this, they are flooding with digital content, which is leading to the spamming of the digital platforms. From that zone, in order to have a purposeful and meaningful platform, we need to filter all the unwanted or not so important stuff. This study can also be extended to other platforms like Instagram, Facebook and other blogging sites as well. Extension of this research can also be done in a review monitoring system in which we can find if a particular system is fake or genuine. Also, in India, which is happening to be the hub for digital platforms, this will be really helpful to mitigate the scenario similar to Cambridge Analytica from happening. As part of this work, it can be concluded that the LSTMs are more robust in handling the text data and also to extract the sentiment of the text. The key contribution to the literature will be the combination of manually engineered features and BOW features when fed into the LSTM model will give results with high precision.

LIMITATIONS AND SCOPE FOR FUTURE RESEARCH

The manually engineered features are not exhaustive. Future studies can build an exhaustive feature list model can increase the robustness For a particular incident, there will be many keywords, but in this study, we have considered only a single keyword (Chatterjee et al., 2018). If we can deal with the analysis comprising all the prominent keywords, then that will give a complete picture instead of a glimpse. Moreover, that will be helpful in understanding and evaluating the situation in a better way. The pre-processing of text can be done using word embedding techniques like Word2Vec or Glove. These being neural network-based techniques, they will help to implement the solution for much more complex text data. Automate the information retrieval from Twitter through Kafka. Using the Convolutional Neural Net (CNN) architecture instead of Recurrent Neural Net. As research says that CNNs are advanced compared to RNNs, but the challenge is CNN's require much computational power and are complex to implement.

CONFLICT OF INTEREST

The authors declare no competing interests.

FUNDING INFORMATION

This research work is not funded by any of the organizations.

REFERENCES

Abdel-Nasser, M., & Mahmoud, K. (2019). Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing & Applications*, *31*(7), 2727–2740. doi:10.1007/s00521-017-3225-z

Aggarwal, R., Gopal, R., Gupta, A., & Singh, H. (2012). Putting Money Where the Mouths Are: The Relation Between Venture Financing and Electronic Word-of-Mouth. Information Systems Research, 23(3-part-2), 976-992. doi:10.1287/isre.1110.040210.1287/isre.1110.0402

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. doi:10.1016/j.jocs.2010.12.007

Chatterjee, S., Deng, S., Liu, J., Shan, R., & Jiao, W. (2018). Classifying facts and opinions in Twitter messages: A deep learning-based approach. *Journal of Business Analytics*, *1*(1), 29–39. doi:10.1080/2573234X.2018.1506687

Cheng, T., & Wicks, T. (2014). Event detection using Twitter: A spatio-temporal approach. *PLoS One*, *9*(6), e97807. doi:10.1371/journal.pone.0097807 PMID:24893168

Darwish, K., Mubarak, H., Abdelali, A., Eldesouki, M., Samih, Y., Alharbi, R., & Kallmeyer, L. et al. (2018). Multi-dialect Arabic pos tagging: A CRF approach. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Deng, S., Sinha, A. P., & Zhao, H. (2017). Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Systems*, *94*, 65–76. doi:10.1016/j.dss.2016.11.001

Evermann, J., Rehse, J.-R., & Fettke, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems*, *100*, 129–140. doi:10.1016/j.dss.2017.04.003

Farquad, M., & Bose, I. (2012). Preprocessing unbalanced data using support vector machine. *Decision Support Systems*, *53*(1), 226–233. doi:10.1016/j.dss.2012.01.016

Ghiassi, M., Zimbra, D., & Lee, S. (2016). Targeted Twitter Sentiment Analysis for Brands Using Supervised Feature Engineering and the Dynamic Architecture for Artificial Neural Networks. Journal of Management Information Systems, 33(4), 1034-1058. doi:10.1080/07421222.2016.1267526 doi:10.1080/07421222.2016.1267526

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Hasan, M., Orgun, M. A., & Schwitter, R. (2019). Real-time event detection from the Twitter data stream using the TwitterNews plus Framework. *Information Processing & Management*, *56*(3), 1146–1165. doi:10.1016/j.ipm.2018.03.001

He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, *52*(7), 801–812. doi:10.1016/j.im.2015.04.006

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, *104*, 38–48. doi:10.1016/j.dss.2017.10.001

Lakshmi Kanth is a graduate student in the MBA Business analytics program and having expertise in Natural Language Processing and conversational chatbot.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259–284. doi:10.1080/01638539809545028

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. doi:10.1038/nature14539 PMID:26017442

Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C. (2012, 1-5 April 2012). *TEDAS: A Twitter-based Event Detection and Analysis System.* Paper presented at the 2012 IEEE 28th International Conference on Data Engineering. doi:10.1109/ICDE.2012.125

Luo, X., Zhang, J., & Duan, W. (2013). Social Media and Firm Equity Value. Information Systems Research, 24(1), 146-163. doi:10.1287/isre.1120.0462 doi:10.1287/isre.1120.0462

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and e-Business Management*, *16*(3), 547–578. doi:10.1007/s10257-017-0362-y

Sahu, T. P. & Khandekar, S. (2022). A Machine Learning-Based Lexicon Approach for Sentiment Analysis. In I. Management Association (Ed.), Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines (pp. 836-851). IGI Global. https://doi.org/10.4018/978-1-6684-6303-1.ch044 doi:10.4018/IJTHI.2020040102

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. doi:10.1016/j.neunet.2014.09.003 PMID:25462637

Shuang, K., Ren, X., Yang, Q., Li, R., & Loo, J. (2019). AELA-DLSTMs: Attention-Enabled and Location-Aware Double LSTMs for aspect-level sentiment classification. Neurocomputing, 334, 25-34. doi:https://doi.org/10.1016/j.neucom.2018.11.084 doi:10.1016/j.neucom.2018.11.084

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, *63*(1), 163–173. doi:10.1002/asi.21662

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. Social Science Computer Review, 29(4), 402–418. https://doi.org/10.1177/0894439310386557 doi:10.1177/0894439310386557

Wang, G. A., Wang, H. J., Li, J., Abrahams, A. S., & Fan, W. (2014). An Analytical Framework for Understanding Knowledge-Sharing Processes in Online Q&A Communities. ACM Trans. Manage. Inf. Syst., 5(4), Article 18. doi:10.1145/2629445 doi:10.1145/2629445

Wiebe, J., & Riloff, E. (2005). Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. In Computational Linguistics and Intelligent Text Processing (pp. 486-497). doi:10.1007/978-3-540-30586-6_53

Wright, A. (2009). Our sentiments, exactly. *Communications of the ACM*, *52*(4), 14–15. Advance online publication. doi:10.1145/1498765.1498772

Zhao, L., Chen, F., Dai, J., Hua, T., Lu, C. T., & Ramakrishnan, N. (2014). Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling. *PLoS One*, *9*(10), e110206. doi:10.1371/journal.pone.0110206 PMID:25350136

P. Murugan is working as an Assistant Professor in the School of Management Studies, University of Hyderabad (UoH), Hyderabad. His papers are published in SSCI and ABDC listed journals. He also has presented papers at many international and national level conferences.