Advances in Parallel Computing Technologies and Applications
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC210144

Covid-19 Sentiment Analysis Using Deep Learning and Machine Learning

Anitha R [a,1], Ashok Kumar P M [a] and Ravi kumar T [a]
[a] *Associate Professor, Department of Computer Science & Engineering, Koneru Lakashmaiah Education Foundation, Vaddeswaram*

Abstract. As we see in later days the covid-19 is spreading broadly. In this the individuals from diverse places within the world discussing about this widespread in numerous ways. So In this respect here we are employing a estimation examination to classify each and each survey of the individuals. What estimation examination will do is it'll qualify and measure the comes about agreeing to the voice of client so for this individuals reaction is taken from online platform such as twitter stack the information and store this information in shape of datasets after that we got to do investigation on information utilizing conclusion mining. In this venture we got to take the covid-19 tweets information from twitter, based on the information we need to made the investigation of the information by utilizing the opinion analysis. In this venture we have long shortterm memory(LSTM), Naïve Bayes and Calculated Regression. Then we discover the exactness's of each calculations based on that we'll conclude that which calculation will be great in arrange to execute assumption investigation.

Keywords. LSTM, naïve Bayes, COVID-19.

1. Introduction

Opinion mining is one of a strategy to decide and made an investigation on the information in terms of positive, negative and impartial classifications. Assumption investigation makes a difference information examiners inside expansive ventures gage open conclusion, conduct nuanced advertise investigate, screen brand and item notoriety, and get it client encounters [1].

Machine learning may be a method that mimics intelligence as a part of homo sapiens. It helps in an exceedingly heap of the way like doing analysis, designing, and call making. In machine learning, there are several hidden tracks that are useful in doing many tasks. Here in each track, they are a group of neurons that will mimic the homo sapiens brain [2–4]. These characteristics of machine learning are making the output more accurate. Machine learning will help us to improve our output in the best ways. Even though it takes more time data for processing the out is accurate when compared to other models. for every track, we will perform some functions which will give activations [5, 6].

[1] Anitha R, Department of CSE, Koneru Lakashmaiah Education Foundation, Vaddeswaram.
E-mail anitharaju15@gmail.com.

Long short-term memory (LSTM) is an fake repetitive neural organize (RNN) design utilized within the field of machine learning. It was proposed in 1997 by Sepp Hochreitr and Jurgen schmidhuber. Not at all like standard feed-forward neural systems, LSTM has criticism associations. It can handle not as it were single information focuses (such as pictures) but too whole arrangements of information (such as discourse or video) [7]. A common LSTM unit is composed of a cell, an input entryway, an yield door, and a disregard entryway. The cell recollects values over subjective time interims, and three doors direct the stream of data into and out of the cell. LSTM is well-suited to classify, handle, and anticipate the time arrangement given of obscure duration. The LSTM primarily comprises 3 gates: i)Input gate (ii)Yield gate(iii)Disregard gate [8].

Credulous Bayes classification works on Baye's hypothesis. Basically naïve bayes classification calculation tends to be a standard arrangement for assumption examination task. The fundamental thought of this bayes is to calculate the probabilities of classes relegated to writings by utilizing Joint probabilities of words and classes. Navie Baye's Classification can be utilized for restricted measure Preparing information to appraise vital parameters and are very productive to actualize. Based on over-simplified suspicions of conditional likelihood and shape of the information dissemination. There are four models in NBC, Here we utilized Gaussian NB for assumption investigation which works highlights are spoken to with frequency. Y=argmaxyp(y)(pi)n=1np(xi/y) Here arg implies returns the most elevated likelihood of taken features. Calculated relapse could be a factual show that in its fundamental frame employments a calculated work to demonstrate a double subordinate variable, in spite of the fact that numerous more complex expansions exist. In relapse examination, calculated relapse or log it relapse is evaluating the parameters of a calculated show a shape of double regression [9]. Since the speculation work for logistic regression is sigmoid in nature consequently [10], The Primary important step is finding the slope of the sigmoid work. Ready to see from the induction underneath that angle of the sigmoid work takes after a certain pattern.

Machine learning may be a method that mimics intelligence as a part of homo sapiens. It helps in an exceedingly heap of the way like doing analysis, designing, and call making. In machine learning, there are several hidden tracks that are useful in doing many tasks. Here in each track, they are a group of neurons that will mimic the homo sapiens brain. These characteristics of machine learning are making the output more accurate. machine learning will help us to improve our output in the best ways. Even though it takes more time data for processing the out is accurate when compared to other models. for every track, we will perform some functions which will give activations, it is just like the process work of homo sapiens. An LSTM could be a long short term memory which can take the entire input at once and process it in like manner to the likelihood work. An LSTM is the progressed form of the repeat neural arrange which is within the field of machine learning. This demonstrate was proposed by S. Hochriter and J. schmidhuber in year 1978. Not at all like the fake neural systems (ANN) it takes after nourish forward arrange where as in LSTM it has criticism association in each and each cell. At a time LSTM will prepare a single unit of information as well as bulk data.

The applications of Long brief term memory is finding and extricating designs in information discourse or voice recognition in it. In for the most part an LSTM may be a composition of a cell input entryway, disregard door and yield door into it. And these three entryways will direct a stream of input and yield information from one put to another in specific interim of time and anticipate and extricate design in specific interim

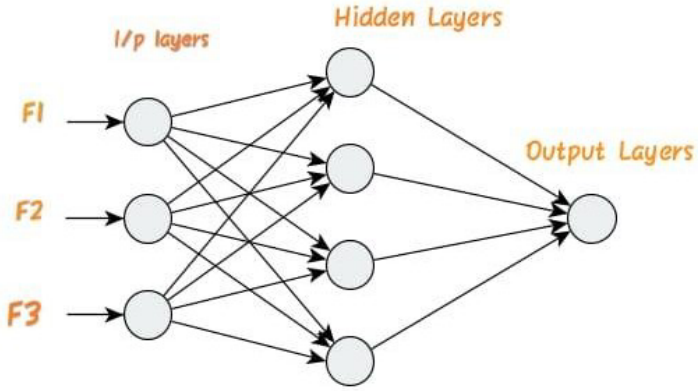

Figure 1. Neural Network Architecture

of time.Input gate: An input door could be a entryway which it'll take input and alter the input within the memory unit of cell. In this respect a sigmoid work will choose the esteem of input in or 1.A tanh esteem will dole out a weights to each and each cell and adjust esteem from -1 to 1. Forget gate: A disregard entryway could be a door which a sigmoid work will choose the esteem and get the subtle elements of the occupied from square it takes the past state and execute the another state based on past state (xt-1) and streams the yield to each cell agreeing to it. Fgt=∑(Wf.[hnt-1,xt]+bf). Output gate: An Yield entryway could be a entryway where the esteem in memory unit will choose the yield of whole piece. A sigmod work will decide that which esteem to stream within the square a tanh esteem will choose esteem of the square climate or 1 in it. In this way the the esteem of the yield will depend upon the these two handle and this handle rehashes until will get way better precision. Opt=∑(Wo.[hnt-1,xt]+bo) Hnt=opt*tanh(cnt).

Credulous Bayes classification works on Baye's hypothesis. Primarily naïve bayes classification calculation tends to be a pattern arrangement for assumption investigation task. The essential thought of this bayes is to calculate the probabilities of classes alloted to writings by utilizing Joint probabilities of words and classes. Navie Baye's Classification can be utilized for constrained estimate Preparing information to assess essential parameters and are very productive to execute. Based on over-simplified suspicions of conditional likelihood and shape of the information conveyance. There are four models in NBC, Here we utilized Gaussian NB for opinion investigation which works highlights are spoken to with frequency. This generally classified and utilized in content classification have higher victory rate to differ to the programs. Bayes Theorem $P(m/hj) = p(hj/m)p(m)p(hj)Y = argmaxyp(y)(pi)n = 1np(xi/y)$.

Calculated relapse could be a factual demonstrate that in its essential frame employments a calculated work to demonstrate a double subordinate variable, in spite of the fact that numerous more bigger expansions there. In relapse examination, calculate relapse or logit relapsing is getting the argument of a calculated show a frame of 2 side. Since the theory work for logistic regression is sigmoid in nature subsequently, The Primary important step is finding the slope of the sigmoid work. We will see from the induction underneath that slope of the sigmoid work takes after a certain design [11].

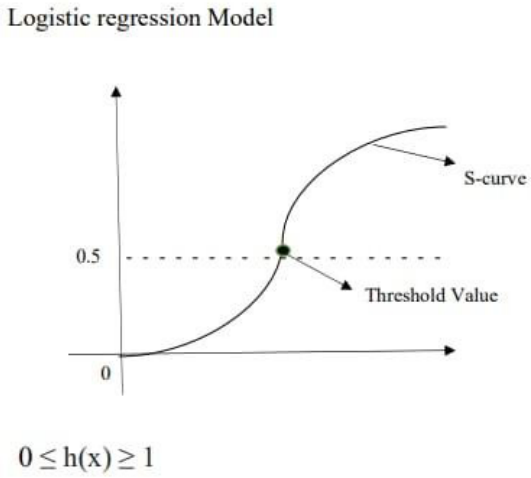

Figure 2. Logistic Regression Model

2. Methodology

Based on the data available in dataset is consists of three fields namely ID, Text, Sentiment. Opinion Investigation is the method of deciding whether a chunk of composing is positive, negative or impartial. Assumption investigation makes a difference information examiners inside expansive ventures gage open conclusion, conduct nuanced advertise investigate, screen brand and item notoriety, and get it client encounters. Based on this data available in dataset we made an analysis on the data by implementing four different types of algorithms. The four algorithms are: Deep learning [12], LSTM using RNN, Navie Bayes, Logistic Regression With the help of all these algorithms we made analysis on the data finally find the accuracy of each model.

Algorithm 1 Implementation of sentiment analysis using Artificial Neural Networks

1: import nlk, numpy, pandas, keras packages and load the dataset.
2: After importing, clean the tweets and remove the stop words from the tweets.
3: Do the text processing on tweets.
4: Convert the words into vectors using tokenizers.
5: Derive the dataset into training and test datasets.
6: Apply the sequential model as a base model with 64 neurons in hidden layers.
7: In order to remove the overfitting from the model, implement reduced, regularized and dropout model.
8: Find and accuracy and loss function for each model.

Implementation of sentiment analysis [13–15] using Naïve Bayes and Lexicons is as follows Load the covid-19 dataset into idle and Remove all outliers and noise of data to make convenient for analysis. In the second phase convert the target classes to positive to negative with help of lexicons and Divide the data into train and test data for making analysis. In the third phase Create the model for the naïve bayes for training the dataset

Algorithm 2 Implementation of sentiment analysis using LSTM

1: First we have to install all predefined libraries in Io
2: For Implementing LSTM we have to import a KERAS package in it we have to implement a Tensor flow in backend.
3: Then we have to load the dataset into it for preprocessing data.
4: Remove an unwanted noise, outliers in the data in order to make data fit for the analysis.
5: Then create basic LSTM model to fit the data in it.
6: Split the data into train and test with 70% and 30% .
7: Create model of LSTM and evaluate model.
8: Predict outcome of result by using LSTM model.
9: Run for each and every approach of model. For getting trained to model.
10: Display accuracy of the model

and Fit the train data to module. Finally Predict the target variables (attributes in dataset) and Find the accuracy of module and display result. Implementation of sentiment analysis using Logistic regression Load the covid 19 dataset into python idle. Remove all outliers and noise of data in order to make convenient for analysis. Divide the data into train and test data for making analysis.

- Fit the train data to module.
- Create the model for the logistic regression for training purpose.
- Predict the target variables with set of all independent variables i.e., attributes in dataset

3. Results and Discussion

```
print(classification_report(predictions,y_test))
```

	precision	recall	f1-score	support
anger	0.54	0.68	0.61	145
fear	0.65	0.58	0.61	158
joy	0.82	0.67	0.74	192
sad	0.72	0.77	0.74	177
micro avg	0.68	0.68	0.68	672
macro avg	0.68	0.68	0.67	672
weighted avg	0.69	0.68	0.68	672

```
accuracy = lr.score(X_test,y_test)
print(accuracy*100,'%')
```

```
67.85714285714286 %
```

Figure 3. Accuracy of regularized model

The main applications of the sentiment analysis is the sentiment analysis mainly focus on voice of the customer which means the what sentiment analysis will do is the data

collected from datasets, forums, surveys and social media platforms it will load the data and made analysis on particular data and recognize some pattern in data made easily to do analysis.

```
reg_results = test_model(reg_model, 5)
print('/n')
print('Test accuracy of regularized model: {0:.2f}%'.format(reg_results[1]*
100))
```

```
10/10 [==============================] - 0s 2ms/step - loss: 0.9854 - ac
curacy: 0.7411
/n
Test accuracy of regularized model: 74.11%
```

Figure 4. Accuracy of LSTM

```
In [15]:
y_pred=clf.predict(X_test)

In [16]:
print("Accuracy:",getAccuracy(y_test, y_pred))
```

```
Accuracy: 0.9
```

Figure 5. Accuracy of Navie Bayes and Lexicons

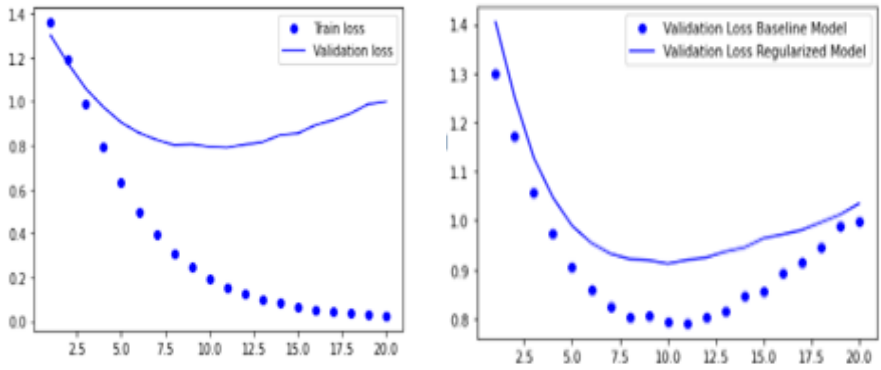

Figure 6. Loss functions of Regular model

In the field of business if a product is launched in market then business people will do survey on particular product based on voice of customer. With the help of sentiment analysis they can easily find or extract patterns into it. So then can made decisions easily. Recently we heard about covid-19 pandemic due to this so many people in the world reacted on this situation positively, negatively, and neutral and these data is taken from social media platform and made analysis on data so it is easy to take decision to the government to control pandemic based on voice of customer. There also many applications on sentiment analysis such as in medical, customer support, product analysis etc.

4. Conclusion

We conclude that implementation of sentiment analysis using lstm is better model when compared with other algorithms, even though we got 90% accuracy on implementing Navie bayes, for some cases the algorithm fails. Therefore using LSTM will give us better result.

References

[1] Wan X. A comparative study of cross-lingual sentiment classification. In 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology 2012 Dec 4 (Vol. 1, pp. 24-31). IEEE.

[2] Meng X, Wei F, Liu X, Zhou M, Xu G, Wang H. Cross-lingual mixture model for sentiment classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 2012 Jul (pp. 572-581).

[3] Shin W, Vaezi M, Lee B, Love DJ, Lee J, Poor HV. Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges. IEEE Communications Magazine. 2017 Aug 15;55(10):176-83.

[4] Shin W, Cherish MJ, Destitute. Non-orthogonal numerous get to in multi-cell systems: Hypothesis, execution, and down to earth challenges. IEEE Commun. Mag. 2017;55(10):176-83.

[5] Xu X. Machine learning-based prediction of urban soil environment and corpus translation teaching. Arabian Journal of Geosciences. 2021 Jun;14(11):1-5.

[6] Fan R, Jiang H. Optimal multi-channel agreeable detecting in cognitive radio networks. IEEE Trans. Remote Commun. 2010;9(3):1128–38.

[7] Sadhana SA, SaiRamesh L, Sabena S, Ganapathy S, Kannan A. Mining target opinions from online reviews using semi-supervised word alignment model. In 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM) 2017 Feb 3 (pp. 196-200). IEEE.

[8] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. arXiv preprint cs/0409058. 2004 Sep 29.

[9] Sulthana AR, Jaithunbi AK, Ramesh LS. Sentiment analysis in twitter data using data analytic techniques for predictive modelling. InJournal of Physics: Conference Series 2018 Apr 1 (Vol. 1000, No. 1, p. 012130). IOP Publishing.

[10] Mitola J. Cognitive radio for adaptable portable mixed media communications. Versatile Netw. Appl. 2001;6(5):435-41.

[11] Liu Y, Qin Z, Elkashlan M, Hanzo. Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access. IEICE Trans. Commun. 2009;3(6):1016-23.

[12] Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng AY, Potts C. Recursive deep models for semantic compositionality over a sentiment treebank. InProceedings of the 2013 conference on empirical methods in natural language processing 2013 Oct (pp. 1631-1642).

[13] Pan SJ, Ni X, Sun JT, Yang Q, Chen Z. Cross-domain sentiment classification via spectral feature alignment. In Proceedings of the 19th international conference on World wide web 2010 Apr 26 (pp. 751-760).

[14] Bollegala D, Weir D, Carroll J. Cross-domain sentiment classification using a sentiment sensitive thesaurus. IEEE transactions on knowledge and data engineering. 2012 May 22;25(8):1719-31.

[15] Peddinti VM, Chintalapoodi P. Domain adaptation in sentiment analysis of twitter. InWorkshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence 2011 Aug 24.