Kunal V,2021, 9:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Fake News Detection Using Machine Learning

Kunal V, Akarsh M, Sameera MS Harishta, Mansi, Asst. Prof. Prathima Mabel
Department of Information Science Engineering,
Dayananda Sagar College of Engineering Bangalore,
India.
kunalrao8055@gmail.com, akarsh4200@gmail.com, sameerams19.harishta@gmail.com, mansisinghpm@gmail.com,
prathimamabel-ise@dayanandasagar.edu

Abstract-Social media interaction especially the news spreading around the network is a great source of information nowadays and such fake news are used to gain financially, our social media will many times project false facts as completely true without checking the background of the information and also during the elections due to political reasons there will be a lot of fake news generated, so detecting the fake news has become a great challenge and importance. Now a days machine learning will play a key role in classification of the information using natural language processing techniques. The easy access and exponential growth of the information available in Social media network has made it inticate to distinguish between false and true information .

Keywords:- Machine Learning, Natural Language Processing and Fake News.

I. INTRODUCTION

As we know in today's world almost internet is ruling everyone's life, from small scale business to corporate sector, schools, colleges everything is somehow reliable on internet.

As internet is given so much of priority with news items, blogs etc. there will be a lot of fake data will be created and this has brought researchers into great interest on this and also many studies have been done by researchers to find effect of this false news, after the launch of jio with low internet package price even the rural area people have also started accessing news through their digital devices only, so as a result there will lot of fake news generated.

II. KEYWORDS

1. Machine Learning:
Machine learning is an Artificial Intelligence technology where we give privilege to our computers to access our data and let them use the same data tolearn for themselves.It is basically getting a computer to perform a specific task without beingprogrammed or instructed specifically to do so. Machine learning is "learning from a collection of examples on how to perform a particular task" interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural data. In other words, it helps the machine to understand our language. The language which we speak and write. NLP enables machines to read, understand and react. AI is booming and NLP is a key feature of that.

III. PROPOSED SYSTEM

The Proposed system focuses on making an application which can be able to detect whether the provided source of information is fake or real.

1. Approach:
The approach we use in this project to implement fake news detection which mainly focuses on detecting whether the provided source of information is fake or real is by using Long Short Term Memory (LSTM) network . We will train a Long Short Term Memory(LSTM) Network to detect fake news from a given source of information. This project can be used to automatically predict whether the circulating news is fake or not. The process could be

International Journal of Science, Engineering and Technology

An Open Access Journal

done automatically without having any human involvement.

2. Long Short Term Memory(LSTM) Network:

Long short term memory are special type of recurrent neural network. They solve the vanishing gradient problem and are capable of retaining memory for a longer duration of time.

2.1 Recurrent Neural Network:

- Normally a feedforward neural network will map a fixed size input to a fixed size output. One main disadvantage of this is they do not have any time dependency or memory effect.
- Recurrent neural network is a type of ANN which is developed to take temporal dimension into consideration by having a memory(internal state).

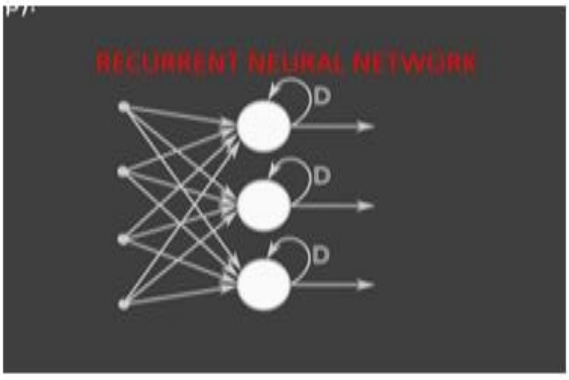


Fig 1. Recurrent Neural Network.

3. Recurrent Neural Network Architecture:

Recurrent neural network has a temporal loop in which the hidden layer not only gives an output but it feeds itself as well . Time will be considered as an extra dimension, Recurrent neural network has the ability to recall what happened in the previous time stamp so it works great with sequence of text.

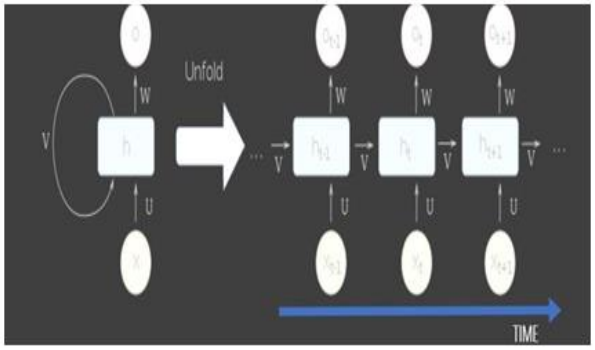


Fig 2. Recurrent Neural Network Architecture.

3.1 Advantages:

- Recurrent neural networks offer huge advantage over feedforward ANN and they are much more effiecient.
- RNN Allows us to work with sequence of vectors

Fig 3. Recurrent Neural Network Architecture.

- Sequence in inputs
- Sequence in outputs
- Sequence in both

4. Gradient descent Algorithm with Equations

To put in very simple terms, Gradient Descent is a helper algorithm that aims to achieve the required optimal solution through trial and error method. It will provide the required optimal solutions mainly based on the bias values.

It works by calculating the gradient of the cost function and moving in the negative direction until the local/global minimum is achieved, By taking the positive of the gradient local/global maximum is achieved and the learning rate will define the size of the steps taken.

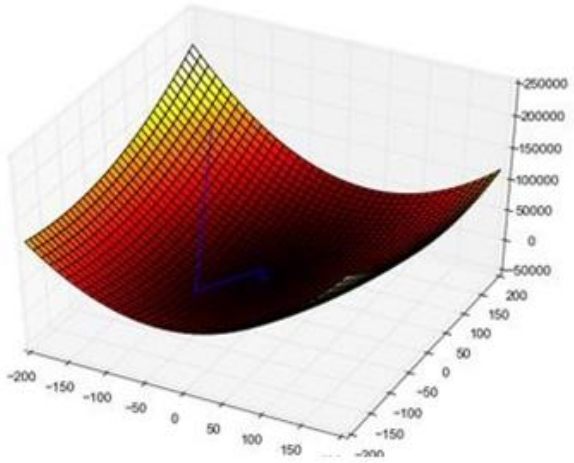


Fig 4. Gradient descent Graph.

International Journal of Science, Engineering and Technology

An Open Access Journal

The area covered in the search space will increase if learning rate increases and they are directly proportional to each other and will inturn help in reaching the global minimum faster.Training will take a very long period of time to reach optimized weight values for small learning rates.

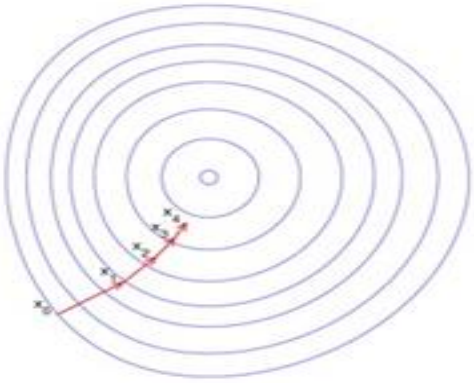


Fig 5. Positive Training graph

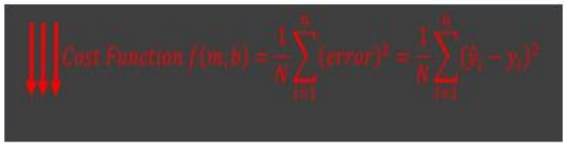
4.1 Equations:

Let's assume that we want to obtain optimal values for parameters m and b.

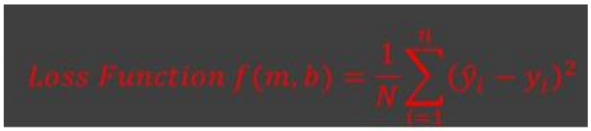
Y=B+M*X

The goal is to find best paraments.

We need to first formulate a loss function as follows:



$$Cost\ Function\ f(m,b) = \frac{1}{N}\sum_{i=1}^{n}(error)^2 = \frac{1}{N}\sum_{i=1}^{n}(\vartheta_i - y_i)^2$$



$$Loss\ Function\ f(m,b) = \frac{1}{N}\sum_{i=1}^{n}(\vartheta_i - y_i)^2$$

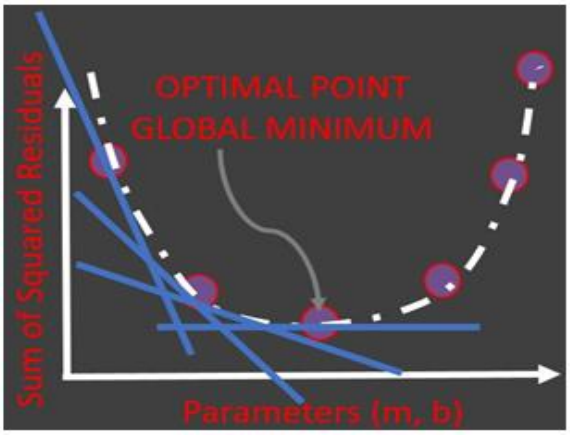
Gradient Descent works as follows:

- Calculate the gradient derivative of the loss function that is loss/w.
- Pick random values for weights (m,b) and substitute.
- Calculate the step size

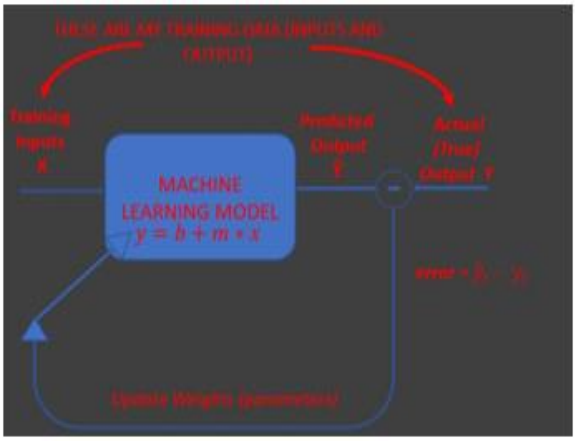
Step Size=Learning Rate * Gradient

- Update the parameters and repeat

New Weight=Old Weight-Step Size



(a)



(b)

Fig 6. Graph.

IV. DESIGN OBJECTIVES

- The design of the project is done using nine modules/ steps which has separate individual objectives of its own
- Understand the problem statmentand business case
- Import libraries and datasets
- Perform exploratory data analysis
- Perform data cleaning
- Visualize cleaned up dataset
- Prepare the data by performing tokenization and padding
- Understand the theory and intuition behind recurrent neural networks and lstm
- Understand the intuition behind long short term memory (lstm) networks
- Assess trained model performance

These are the main 9 objectives involved in the project. Collection of datasets will be done based on the category of true or false.

International Journal of Science, Engineering and Technology

An Open Access Journal

1. Data Flow Diagram:

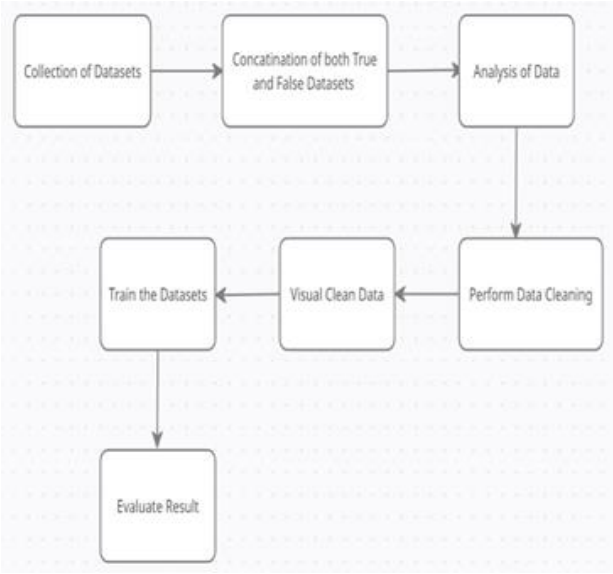


Fig 7. Data Flow Diagram.

1.1 Initially source of information /datasets will be collected based on the category of true or false and both the data will be collected in the initial stage and basically understanding the problem statement will be done.

1.2 The second objective is to implement all the main libraries and datasets .

1.3 Import libraries/datasets and perform preliminary data processing:

- Pandas,
- Numpy,
- Tensorflow,
- Seaborn,
- Matplotlib.

1.4 Then the Exploratory analysis of the data will be done will be done which involves two key challenges.

1.5 Indicate how many data samples do we have per class (i.e.: Fake and True)

1.6 List how many Null element are present and the memory usage for each dataframe.

1.7 Once the data analysis is done then the next important objective is data cleaning where the source of information or data sets which will be containing punctuations, commas etc will be totally removed out from the dataset and pure data without any punctuations will be obtained.

1.8 Next important objective is visualize the cleaned data where we will plot the number of counts for fake vs true news.

1.9 Next objective is to prepare the data by performing Tokenization and Padding.

2. Tokenization:

Tokenization allows us to vectorize text corpus by turning each text into a sequence of integers.

3. Padding:

Padding is defined as the amount of pixels added to an image.

3.1 Next objective is to understand the intuition behind Long short term Memory systems.

- LSTM Networks work better compared to vanilla RNN since they overcome vanishing gradient problem.
- In practice RNN failed to establish a long term dependencies.
- LSTM Networks are type of RNN that are designed to remember long term dependencies by default.
- LSTM can remember and recall information for a prolonged period of time.
- Recall that each line represents a full vector.

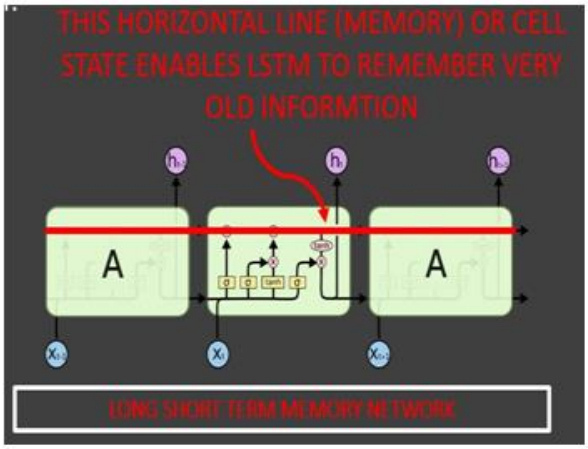


Fig 8. Graph.

3.2 Once all the modules are implemented we will evaluate the final result by checking how much percent the overall system is efficient.

4. Vanishing Gradient Problem:

- Vanishing gradient problem is a difficulty found in training certain Artificial Neural Networks with gradient based methods.
- LSTM networks work much better compared to vanilla RNN since they overcome the vanishing gradient problem.
- The error has to propogate through all the previous layers resulting in a vanishing gradient, The network weights are no longer updated as the

International Journal of Science, Engineering and Technology

An Open Access Journal

gradient goes smaller. ANN gradients are calculated during back propogation.

- The gradients keep diminishing exponentially and therefore the weights and biases are no longer being updated.

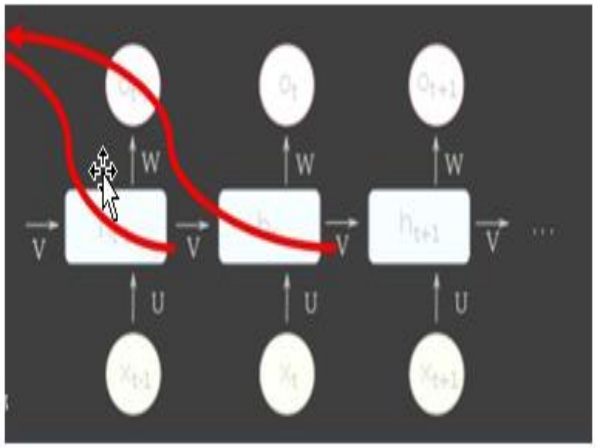


Fig 9. Block Diagram of LSTM.

V. CONCLUSION

As the spread of fake news is increasing rapidly in todays world due to the impact of everything depending on online through digital world our project would definitely help in differentiating between fake and real news.

REFERENCES

[1] Parikh, S. B., & Atrey, P. K. (2018, April). Media-Rich Fake News Detection: A Survey. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 436-441). IEEE.

[2] Conroy, N. J., Rubin, V. L., & Chen, Y. (2015, November). Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact.

[3] Helmstetter, S., & Paulheim, H. (2018, August). Weakly supervised learning for fake news detection on Twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 274-277). IEEE.

[4] Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection.

[5] Stahl, K. (2018). Fake News Detection in Social media.

[6] Nikhil Sharma Department of Computer Engineering, DCE, Gurugram, Haryana, India International Journal of Trend in Scientific Research and Development (IJTSRD).