

# FAKE NEWS DETECTION

USING LSTM

Group Members :  
Siddhesh Chotaliya  
Prince Kanojia  
Ishika Vaish

Guide : Prof. Rahul Jiwane

# INTRODUCTION



FAKE NEWS DETECTION

- Fake news is misinformation or manipulated news.
- The rapid development of the Internet allows a quick spread of information through social networks or websites.
- Social media is an almost unavoidable part of our society now can it be a trusted news source?
- If people lose faith in information, they will no longer be able to access even the most vital information
- NCRB data: 214% rise in cases relating to fake news, rumours



IMAGE



TEXT



AUDIO

89 ;  
1 " 3 P Y 3  
" 3 M A a .  
B d . M A a .  
B Q O K 1 G P IL G  
3 F 8 = a q ; m 8 y Y r .  
B . 10 P Y k 3 > L G M m  
K k . " 4 m G G ? ~  
P m G G :  
K P m G G

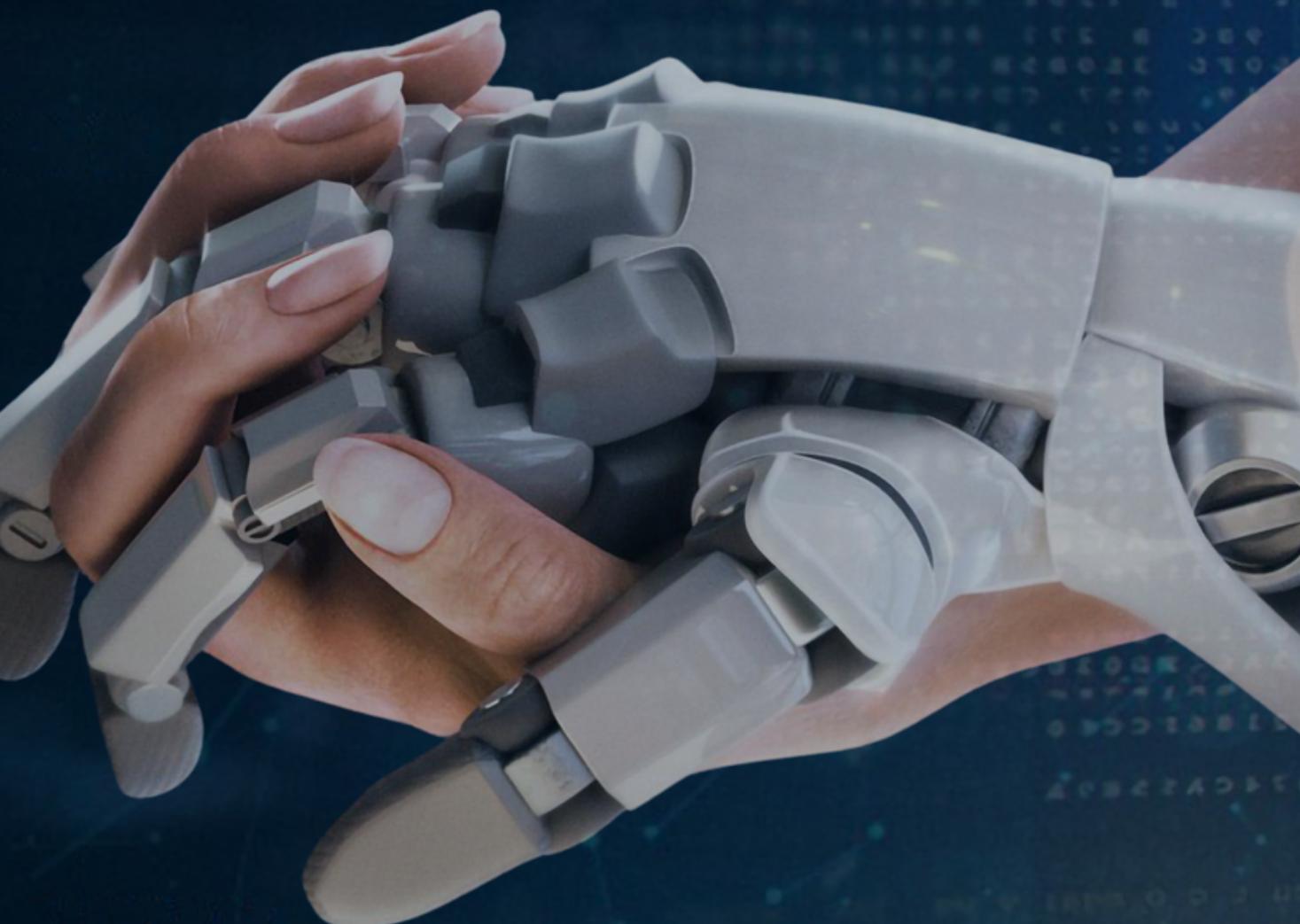


" TECHNOLOGY IS THE CAUSE, TECHNOLOGY IS THE SOLUTION "



# ABSTRACT

- Automatic credibility analysis of news articles is a current research interest.
- Deep Learning Models such as CNN, RNN, LSTM, etc are used for solution.
- Our approach is to develop a model wherein it will detect whether the given news is false or true using LSTM (long short-term memory) and other machine learning concepts such as NLP, word embedding, etc.
- First, we have introduce the datasets which contain both fake and real news and conduct various experiments to organize fake news detector.





# LITERATURE REVIEW

**01**

International Institute of Information Tech International Institute of Information Technology [2018], Bangalore, India. 3HAN A Deep Neural Network for Fake News Detection Technology [2018], Bangalore, India. 3HAN -A Deep Neural Network for Fake News Detection.

**02**

International Research Journal of Engineering and Technology (IRJET)  
Fake News Detection using LSTM Tejaswini Yesugade, Shrikant Kokate, Sarjana Patil, Ritik Varma, Sejal Pawar

**03**

M. Granik and V. Mesyura, “Fake news detection using naive Bayes classifier,” 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

**04**

Fake News detection, <https://www.kaggle.com/jruvika/fake-news-detection>, 2019/07/13

## 01. Bag Of Words

BoW is Natural Language Processing method and Information Retrieval method. NLP model are used on the numbers we cannot use text data into our model. Therefore BoW model is used to preprocess text data by converting it into bag of words.

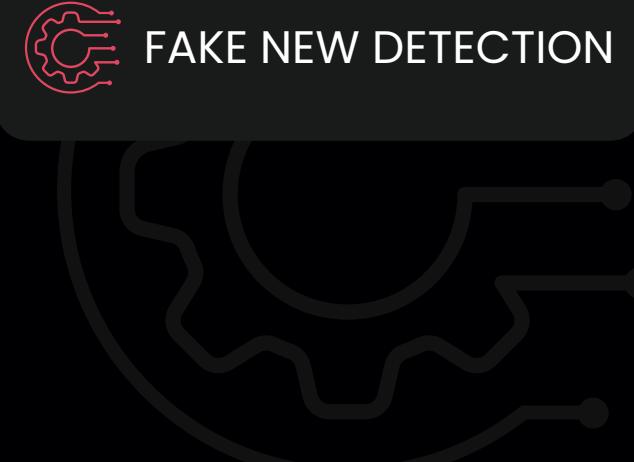
## 02. Naive Bayes

Naïve Bayes uses probabilistic approaches and are based on Bayes theorem. They deal with probability distribution of variables in the dataset and predicting the response variable or value. They are mostly used for text classification

## 03. LSTM

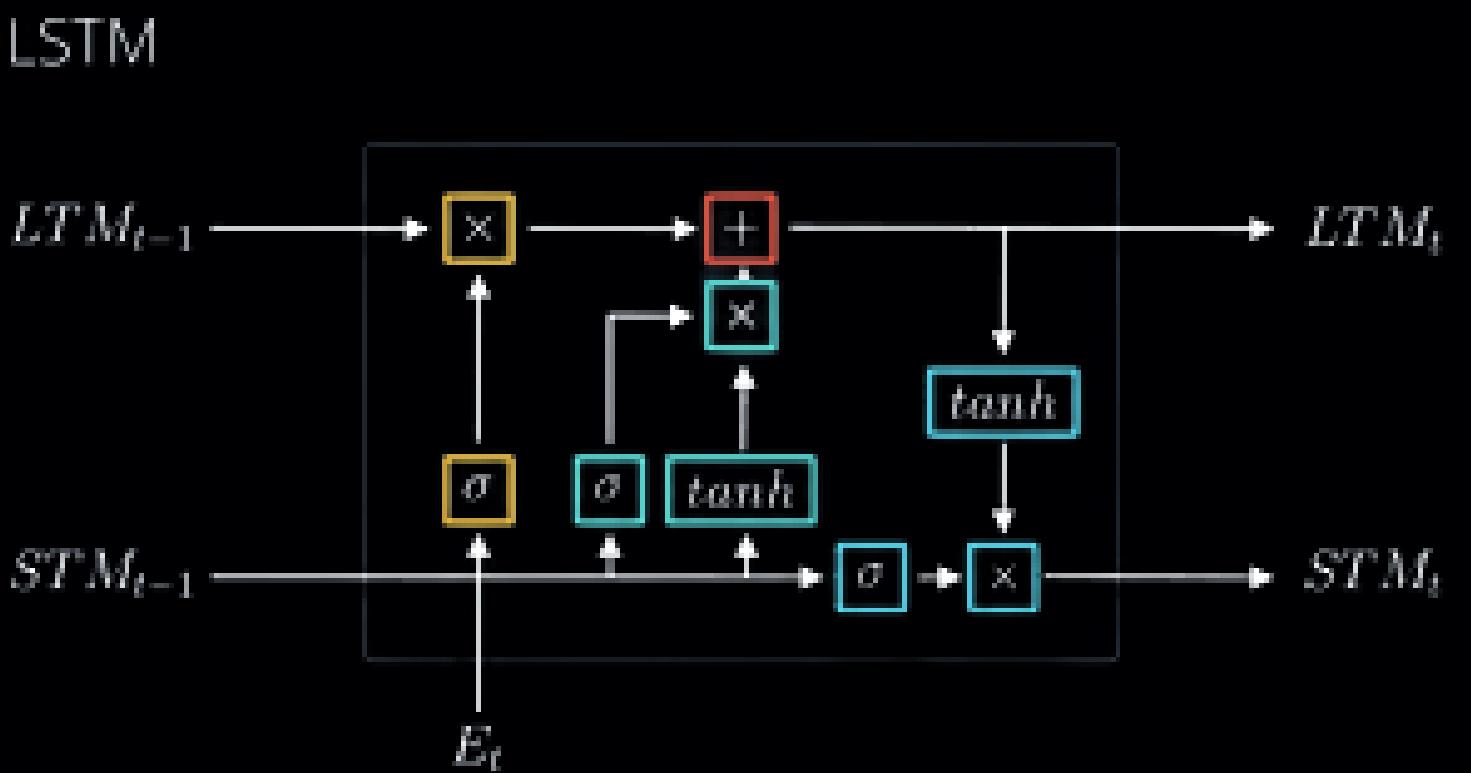
Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. It is used for processing, predicting and classifying on the basis of time series data.



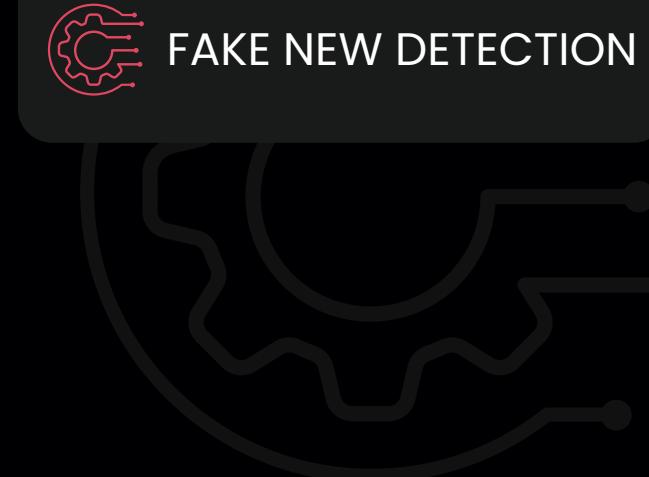


# LSTM ALGORITHM

- 01** Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture
- 02** LSTM networks are well-suited to classifying, processing and making predictions based on time series data
- 03** LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs.
- 04** Using LSTM :
  - Google - Speech recognition on smartphone
  - Amazon - Amazon Alexa
  - Apple - Quick type function on Alexa and Siri
  - Microsoft - End to End Speech Translation



# IMPLEMENTATION



01

## PREPROCESSING

- To transform data into the relevant format the data set needs preprocess.
- Remove all the stop words from the dataset.
- Converting them to lowercase and removing punctuation.
- Then, this collection of words is appended to document.

02

## WORD INDEX OF TOKENIZE DATASET

- Word tokenizing, appends text to a list and the list be named as documents.

03

## WORD EMBEDDING

- One hot Representation: We cannot give input in the form of text format to the algorithm so we have to convert them into the numeric form
- Word embedding apply feature extraction on the provided input vector. In total 40 vector features are considered.

04

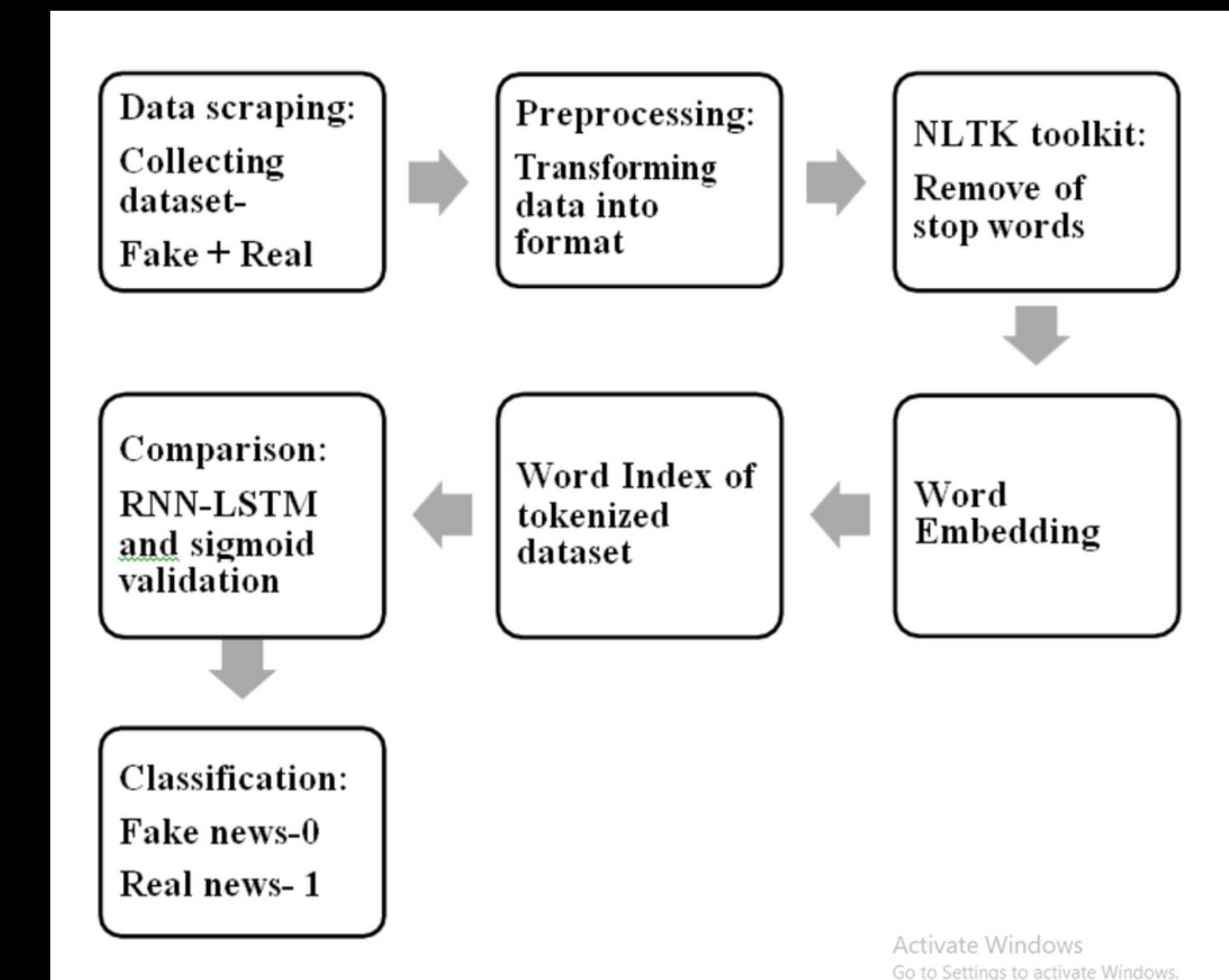
## MODEL

- Output from the word embedding is provided to the model. The machine learning model implemented here is a sequential model
- Then training and testing of model is done.

04

## CLASSIFICATION

- For both preprocessed testing data the result is predicted. If the predicted value > 0.5 Classified as 1 is real and 0 is fake.



Activate Windows  
Go to Settings to activate Windows.



FAKE NEWS DETECTION

# OUTPUT

01

Creating LSTM model

```
: model = Sequential()
model.add(Embedding(vocab_size, output_dim = DIM, weights = [embedding_vectors], input_length=maxlen, trainable = False ))
model.add(LSTM(units=128))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer = 'adam', loss= 'binary_crossentropy', metrics=['acc'])

model.summary()
Model: "sequential_3"
Layer (type)          Output Shape         Param #
=====
embedding_1 (Embedding)    (None, 1000, 100)     23187300
lstm_1 (LSTM)           (None, 128)            117248
dense_1 (Dense)         (None, 1)              129
=====
Total params: 23,304,677
Trainable params: 117,377
Non-trainable params: 23,187,300
```

02

```
: x = [' Pakistani posts of Kirpan & Pimpal destroyed by Indian Army in an immediate retaliation to beheading of soldiers ']
x = tokenizer.texts_to_sequences(x)
x= pad_sequences(x, maxlen = maxlen)

: (model.predict(x) >=0.5).astype(int)

: array([[0]])
```

**1 = True News, 0 = Fake News**

03

```
In [185]: x = ['In one of the world's biggest deals in the tech world, billionaire Elon Musk has taken control of Twitter. Musk will p'
x = tokenizer.texts_to_sequences(x)
x= pad_sequences(x, maxlen = maxlen)
<
In [186]: (model.predict(x) >=0.5).astype(int)
Out[186]: array([[1]])
```

**1 = True News, 0 = Fake News**



FAKE NEWS DETECTION

# APPLICATIONS



## Journalism

The major spread of information and trusted source is through newspapers and news channels, so this detection can be used to verify the news before broadcasting it.



## Social Media

In today's world of social media, it is easy to manipulate any information or news. Such manipulated news misguides the readers. It is important to identify that news is fake or real. This paper provides various techniques that can be used in detection and classification of information.



# CONCLUSION

- 01** In this digital age, where hoax news is present every where in digital platforms, there is an ultimate need for fake news detection and this model serves its purpose
  
- 02** LSTM can be used to get better results. This Model gives better results with accuracy of 91.05% which is very promising, we can further increase results by increasing training data.
  
- 03** LSTM can be used to get better results. This Model gives better results with accuracy of 91.05% which is very promising, we can further increase results by increasing training data.
  
- 04** We also learnt about different algorithms for classification, prediction, etc. By which we can apply these knowledge for future projects.

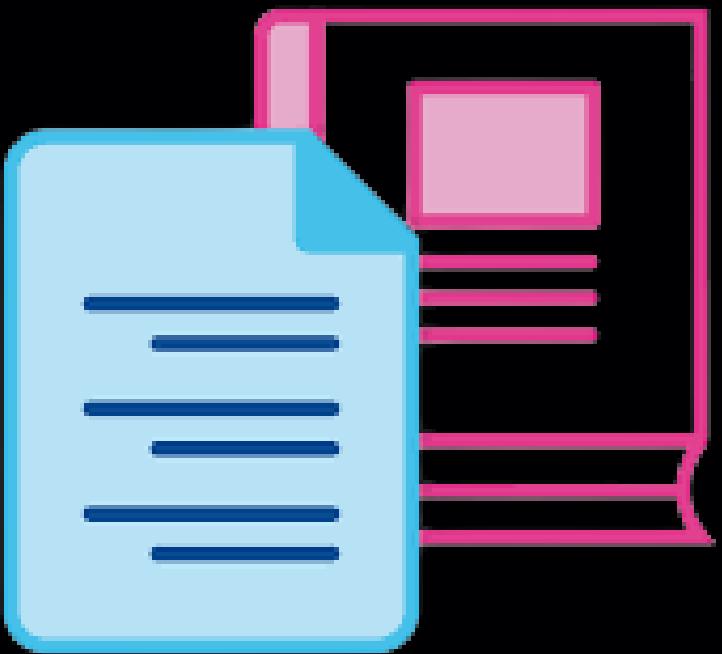




FAKE NEWS DETECTION

# REFERENCES

- <https://www.irjet.net/>
- <https://www.researchgate.net/>
- <https://en.wikipedia.org/>
- <https://ieeexplore.ieee.org/>
- <https://stackoverflow.com/>
- Bhatt, Gaurav "Combining Neural, Statistical and External Features for Fake News Stance Identification." Companion of the Web Conference 2018 on the Web Conference 2018. International World Wide Web Conferences Steering Committee, 2018.





FAKE NEWS DETECTION

# THANK YOU