International Journal in IT & Engineering (IJITE)
Volume 12 Issue 11, November 2024 ISSN: 2321-1776 Impact Factor: 8.341
Journal Homepage: http://ijmr.net.in, Email: irjmss@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal

DEEP LEARNING BASED TECHNIQUES FOR FAKE NEWS DETECTION

Dr.R.Jamuna, Associate Professor in Computer Science, S.R.College, Trichy-2

J.S.Subha Shree, PG 2022-2024, S.R.College, Trichy-2

Abstract

Today in the digital era online Social Media networks have created a powerful platform for people to access, send and share any news globally. But the popularity and easy access leads to many disadvantages pertaining to the authenticity of the news. One such problem is fake news propagation. This paper identifies such misleading news using Lexicon LSTM automated model on a small scale. Due to its increasing popularity, low cost, and easy-to-access nature, online Social Media (OSM) expand networks have evolved as a powerful platform for people to access, consume, and share news.However, this has led to the large-scale distribution of fake news, i.e., deliberate, false, or misleading information. Spreading fake news is roughly as dangerous as spreading the virus or spoiling the image of person. Fake news detection attracts many researchers' attention due to the negative impacts on the society. In this work, we propose a lexicon-enhanced Long-Short Term Memory (LSTM)an automated model that is able to take into account both the news content and the social context for the identification of fake news. The model first uses sentiment lexicon as an extra information pre-training a word sentiment classifier and then gets the sentiment embeddings of words including the words not in the lexicon. Combining the sentiment embedding and its word embedding can make word representation more accurate and to detect fake news and better prediction of fake user accounts and posts. The suggested method can be further improvised with the ability to bring benefits to a variety of new activities, including preventing the spreading of fake news during elections, terrorism, natural disasters, and criminal activities for the wellbeing of humanity.

Keywords: Deep learning, Artificial intelligence (AI), Lexicon-Enhanced LSTM, Fake News, Tokenization, Lexicon.

INTRODUCTION

Fake news is false or misleading information that is a rumour portrayed as news. “Fake” here means unauthenticated or false information or news which has the aim of damaging the reputation of a person or entity, or making money through advertising revenue. With the development of the Mobile Internet and Mobile Communication technologies, social media has become an order of the day and more extensive and deeply integrated into our daily life. With easy accessibility, people tend to acquire and share information as well as express and exchange opinions and thoughts through social multimedia. Unfortunately, due to the openness of social multimedia, a large number of users, and the complexity of sources, various fake news has been fostered on websites. This widespread fake news is utilized by anti social elements to mislead readers, which could do serious harm to the society and may cause great socio-economic losses. Generally users do not have the time and skill to check the genuineness of every piece of information on media. Thence, it is necessary and pressing to discover fake news on social multimedia and guarantee users receive authentic information. Fake information can be spread in the form of text, video, pictures, and audio via social media networks such as face book and Twitter. The fake news problem has existed for a long time. Nowadays, there are different kinds of methods proposed for fake news detection task, including traditional machine learning-based methods.

1. METHODS APPLIED FOR DETECTION

Traditional methods such as Support Vector Machine (SVM), Random Forest, and Decision Tree heavily depend on hand-craft features to debunk fake news, which is labour-intensive and time-consuming. For instance, SVM-TS uses a linear classifier based on SVM along with heuristic rules to classify the posts as fake or real and employs a time-series structure to model the social feature variations. Therefore, detecting fake news can be difficult, especially with no supervising body on the internet. The growth of concern regarding the detection of unreliable news is recent. It is difficult for a human to manually detect news, even with the existence of all topics shown on social media. Therefore, there is a need for an efficient way to help us distinguish false information from true ones posted on social media. One of the efficient ways is to classify the news using Deep Learning (DL) algorithms.

Natural Language Processing (NLP)is a branch of artificial intelligence (AI) that enables computers to comprehend, generate, and manipulate human language.

NLP involves several key tasks:

- Text pre-processing: In the lexical analysis called tokenization[1] refers to the process of converting a sequence of text into smaller parts, known as tokens.
- Stemming: Stemming refers to the process of reducing a word to its word stem that affixes to suffixes and prefixes or the roots.
- Parts of speech: **Parts of Speech (PoS) tagging**, which is giving each word in a text a grammatical tag to each word such as nouns, verbs, adjectives, and adverbs.

3. EXISTING SYSTEM AND ITS LIMITATIONS

Types of Fake News

There are differing opinions when it comes to identifying types of false information. However, when it comes to evaluating the content online there are various types of fake or misleading news[3] we need to be aware of.

These include:

- **Clickbait**

These are stories that are deliberately fabricated to gain more website visitors and increase advertising revenue for websites. Clickbait stories use sensational headlines to grab attention and drive click-through to the publisher website, normally at the expense of truth or accuracy.

- **Propaganda**

Stories that are created to deliberately mislead audiences, promote a biased point of view or particular political cause or agenda.

- **Satire/Parody**

Lots of websites and social media accounts publish fake news stories for entertainment and parody. For example; The Onion, Waterford Whispers, The Daily Mash, etc.

- **Sloppy Journalism**

Sometimes reporters or journalists may publish a story with unreliable information or without checking all of the facts which can mislead audiences.

Misleading Headings

Stories that are not completely false and can be distorted using misleading or sensational headlines. These types of news can spread quickly on social media sites where only headlines and small snippets of the full article are displayed on audience newsfeeds.

- **Biased/Slanted News**

Many people are drawn to news or stories that confirm their own beliefs or biases and fake news can prey on these biases. Social media news feeds tend to display news and articles that they think we will like based on our personalised searches.

In this paper existing machine learning classifiers are utilized for fake news classification. Several supervised machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes (NB), Random Forest (RF), DecisionTree (DT), Gradient Boosting model (GBM), Logistic Regression (LR) and Voting Classifier (Logistic Regression C Stochastic Gradient Descent classifier) can be used for such classification problems.

Disadvantages of existing system

- Long Process of reporting such cases, tedious job as it is difficult to track.
- Most of the fake news cases go unreported and has low accuracy.
- Time consuming process and usually data is limited only to a small percentage.
- Response time is slow and common classifier accuracy is low.

4. PROPOSED SYSTEM AND ITS ADVANTAGES

Here a Lexicon-Enhanced LSTM model (LE-LSTM) to integrate sentiment lexicon into LSTM to capture more sentiment information of words is applied on a small scale. First, we use sentiment lexicon as the extra information to pre-train a word sentiment classifier. And then each word can get its sentiment embedding including the words not in sentiment lexicon. During the main training process, we concatenate the word embedding and its sentiment embedding as the input of LSTM and fine-tune the word sentiment classifier network. 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. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-

term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory.

Advantages of proposed system

- Outperforms in terms of accuracy in all four confusion matrix performance measures as well as execution duration.
- Reduces the number of words in the corpus and decreases the computation time.
- Gives the words strength when fed to the classification model and reduce the possibility of overfitting.This LSTM-based prediction model is more accurate and informative in predicting fake news that outperforms efficient and accurate classifier to detect fake news

5. ALGORITHM USED– LE-LSTM

- The following algorithm can be used for such classification problems .In our paper small sample data is only used. The algorithm can actually be applied when large datasets are to be analyzed. The objective of the paper is to propose a lexicon-enhanced LSTM [2] which is an automated model that will be able to take into account both the news content and the social context for the identification of fake new

```
Input: Training Fake News Dataset 'tnRSX', Training:PFake/Real News Values 'tnINY',
Testing
Output: Accuracy ('Acc_total')
1. procedure LE-LSTM MODEL ( tnRSX , tnINY )
# Hyperparameters with Arguments
2. batchsize = 32; epochs = 1; filters = 2; pool_size = 2; verbose = 2; N_EPOCH=7
3. max_features = 2000; embed_dim = 128; classes=2; input_length=3382; units=100
# Build a Deep Learning Model
4. model = Sequential ()
# Embedding Layer
5. model.add (Embedding (max_features, embed_dim, input_length))
# Convolutional Layer
6. model.add (Conv1D (filters, kernel_size=3, padding='same', activation='relu')
# Maxpooling Layer
7. model.add(MaxPooling1D(pool size))
# BiLSTM Layer
8. model.add(LSTM(units))
# Softmax Layer
9. model.add(Dense (classes, activation='softmax'))
# Compile Function
10. model.compile (loss = 'binary_crossentropy', optimizer = 'adamax', metrics= [accuracy])
# Model Summary
11. print (model.summary( ))
# Fit a Model
12. for all epochs in (1: N_EPOCH ) do
13. model.fit(tnRSX , tnINY , epochs, validation_data ( ttRSX , ttINY,
batch_size=batchsize))
#Model Evaluation
14. Acc = model.evaluate( ttRSX , ttINY, verbose, batch_size = batchsize)
15. Acc_total.append(Acc)
16. End for
17. returnAcc_total
18. End Procedure
```

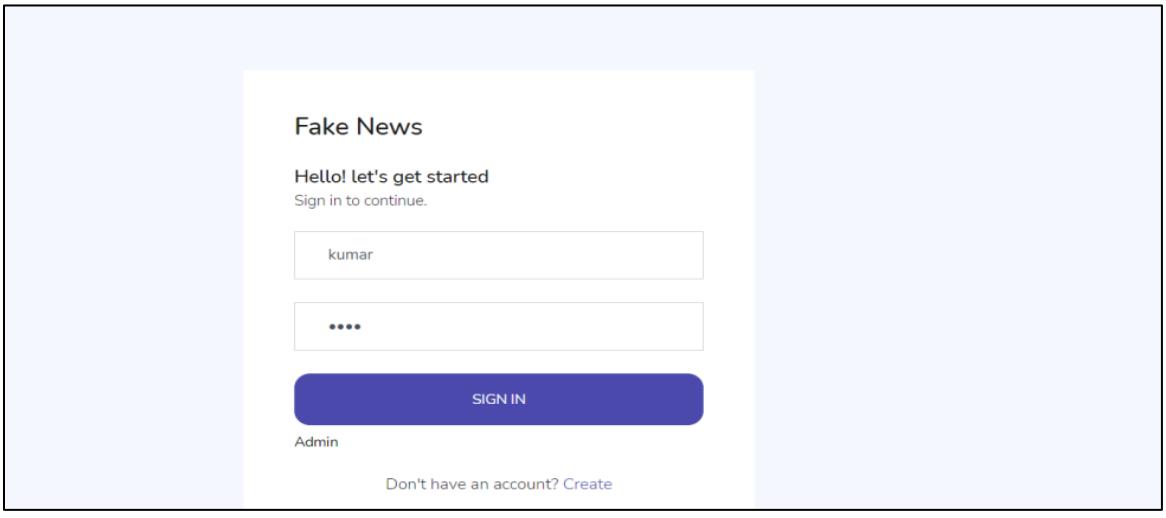
5. IMPLEMENTATION OF THE PROBLEM-SAMPLE CODING

```
import os

import base64

import io

import math

from flask import Flask, flash, render_template, Response, redirect, request, session, abort, url_for

warnings.filterwarnings("ignore")

fromsklearn.metrics import classification_report, accuracy_score, confusion_matrix

mydb = mysql.connector.connect(

host="localhost",

user="root",

  password="",

charset="utf8",

database="fake_news_classify"

)

app = Flask(__name__)

app.secret_key = 'abcdef'

UPLOAD_FOLDER = 'static'

ALLOWED_EXTENSIONS = { 'csv'}

app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER

@app.route('/', methods=['GET', 'POST'])

def index():

msg=""
```

```
ifrequest.method=='POST':
uname=request.form['uname']
pwd=request.form['pass']
cursor = mydb.cursor()
returnrender_template('index.html',msg=msg)
@app.route('/login', methods=['GET', 'POST'])
def login():
msg=""
ifrequest.method=='POST':
uname=request.form['uname']
pwd=request.form['pass']
cursor = mydb.cursor()
cursor.execute('SELECT * FROM admin WHERE username = %s AND password = %s',
(uname, pwd))
account = cursor.fetchone()
if x>0:
act="1"
st="1"
mess="Fake News!!"
elif y>0:
act="1"
st="3"
mess="True News!!"
else:
```

```
act="1"

st="2"

mess="Normal News"

mycursor.execute("SELECT count(*) FROM user_post where uname=%s &&
status=1",(uname,))

cnt = mycursor.fetchone()[0]

ifcnt==2:

msg1="warn"

elifcnt==3:

msg1="block"

mycursor.execute("update register set dstatus=1 where uname=%s",(uname,))

mydb.commit()
```

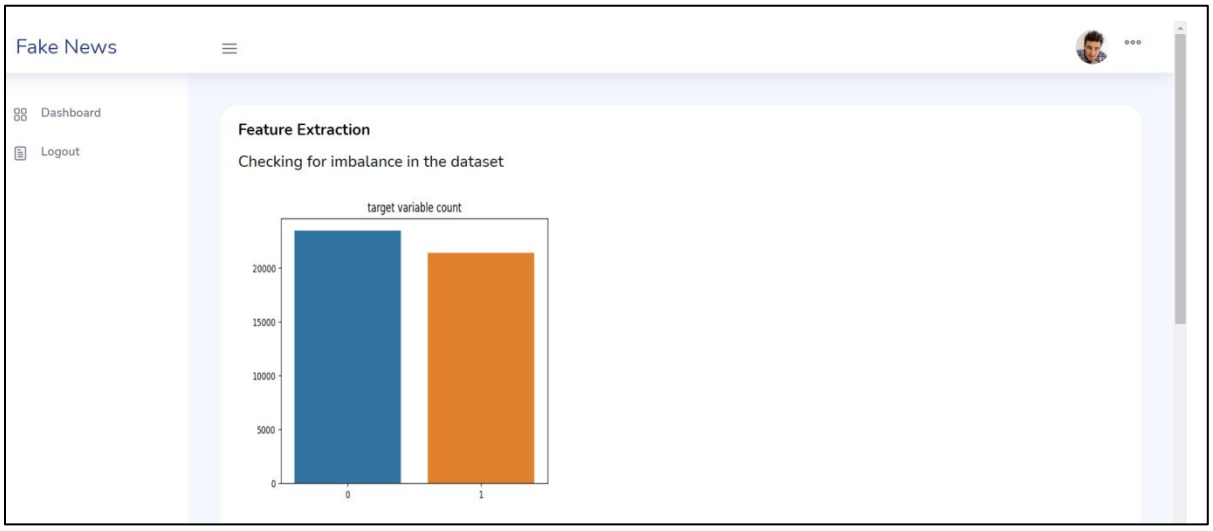
6.RESULTS AND SAMPLE OUTPUT




Fake News
Dashboard
Logout
Fake News
News Post
news
Trump travel ban should not apply to people with strong U.S. ties: court
POST
Normal News


Fake News
Dashboard
Logout
Fake News
News Post
news
That Story Trump Just Told The Press About Obama And Air Force One NEVER Happened
POST
Fake News!!


Fake News
Logout
Preprocessing
Data Cleaning
u.s. budget fight looms, republicans flip fiscal script washington (reuters) - head conservative republican faction u.s. congress, voted month huge expansion national debt pay tax cuts, called "fiscal conservative" sunday urged budget restraint 2018. keeping sharp pivot way among republicans, u.s. representative mark meadows, speaking cbs' "face nation," drew hard line federal spending, lawmakers bracing battle january. return holidays wednesday, lawmakers begin trying pass federal budget fight likely linked issues, immigration policy, even november congressional election campaigns approach republicans seek keep control congress. president donald trump republicans want big budget increase military spending, democrats also want proportional increases non-defense "discretionary" spending programs support education, scientific research, infrastructure, public health environmental protection. "the (trump) administration already willing say: 'we're going increase non-defense discretionary spending ... 7 percent,'" meadows, chairman small influential house freedom caucus, said program. "now, democrats saying that's enough, need give government pay raise 10 11 percent. fiscal conservative, don't see rationale is. ... eventually run people's money," said. meadows among republicans voted late december party's debt-financed tax overhaul, expected balloon federal budget deficit add $1.5 trillion 10 years $20 trillion national debt. "it's interesting hear mark talk fiscal responsibility," democratic u.s. representative joseph crowley said cbs. crowley said republican tax bill would require united states borrow $1.5 trillion, paid future generations, finance tax cuts corporations rich. "this one least ... fiscally responsible bills we've ever seen passed history house representatives. think we're going paying many, many years come," crowley said. republicans insist tax package, biggest u.s. tax overhaul 30 years, boost economy job growth. house speaker paul ryan, also supported tax bill, recently went meadows, making clear radio interview welfare "entitlement reform," party often calls it, would top republican priority 2018. republican parlance, "entitlement" programs mean food stamps, housing assistance, medicare medicaid health insurance elderly, poor disabled, well programs created washington assist needy. democrats seized ryan's early december remarks, saying showed republicans would try pay tax overhaul seeking spending cuts social programs. goals house republicans



7. CONCLUSIONS

The fake news problem is not new, as disinformation haslong been circulated in newspapers. Because of the internet, false news spread quickly through social media and blogs. This type of information might be harmful. Thus we must be able to distinguish between fake and actual news.This paper has introduced a fake news detection systemtaking few samples for testing by using LSTM automated algorithm as basis . In Deep learning , we have proposed the Lexicon Enhanced – Long Short Term Memory(LE-LSTM) model that has achieved a good performance.In future , the approach can be applied for the fake news detection using a real-time approach on large scale data can be used in various social media like Facebook, Twitter, Instagram, and other platforms. The suggested method has the ability to bring benefits to a variety of new activities, including preventing the spread of fake news during elections, terrorism, natural disasters, and criminal activities for the wellbeing of humanity.

8. REFERENCES

[1] Pisa P. S., de Oliveira N. R., Lopez M. A., de Medeiros D. S. V., and Mattos D. M. F., (2021) ``Identifying fake news on social networks based on natural language processing: Trends and challenges," Information, vol. 12, no. 1, p. 38.

[2] Vijay J. A., Basha H. A., and Nehru J. A., (2021) "A dynamic approach for detecting the fake news using random forest classifier and NLP in Computational Methods and Data Engineering" pp. 331-341.

[3] Nikiforos M. N., Vergis S., Stylidou A., Augoustis N., Kermanidis K. L., and Maragoudakis M., (2020) ``Fake news detection regarding the Hong Kong events from tweets," in Proc. Int. Conf. Artif. Intell. Appl. Innov. Greece: Springer pp. 177-186.