Deep learning modules Fake News Detection

Abstract:

- This module explores advanced techniques, specifically deep learning models, for improving the accuracy of fake news detection. It provides an abstract class, Advanced Fake News Detection, which defines methods for loading data, preprocessing data, training a model, making predictions, and evaluating the model's performance.

- Additionally, a concrete implementation of the abstract class called LSTM Fake News Detection is provided. This class extends Advanced Fake News Detection and utilizes an LSTM (Long Short-Term Memory) model for fake news detection.

- The LSTM Fake News Detection class loads data from a CSV file, preprocesses it by splitting it into train and test sets, vectorizes the text data using TF-IDF, and pads sequences for LSTM input. It then builds a Keras LSTM model, compiles it, and trains it on the training data.

- The predict method of the LSTM Fake News Detection class makes predictions using the trained model on the test data. The evaluate method calculates the accuracy of the model's predictions.

- To use this module, a CSV file containing fake news data needs to be provided, and the code needs to be modified accordingly.


```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import tensorflow as tf
import re
from tensorflow.keras.preprocessing.text import Tokenizer
import tensorflow as tf
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,
confusion_matrix, precision_score, recall_score
import seaborn as sns
```

```
plt.style.use('ggplot')
```

Read the data

```
fake_df = pd.read_csvfake_df = pd.read_csv('fake-and-real-news-dataset/Fake.csv')
real_df = pd.read_csv('fake-and-real-news-dataset/True.csv')
```

```
fake_df.head(10)
```

	title	title	subject	Date
0	Donald Trump Sends Out Embarrassing New Year’...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama’s Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017

	title	title	subject	Date
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017

Fake News Detection

Importing Libraries

In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string
```

Importing Dataset

In [2]:

```
df_fake = pd.read_csv("../input/fake-news-
detection/Fake.csv")
df_true = pd.read_csv("../input/fake-news-
detection/True.csv")
```

In [3]:

```
df_fake.head()
```

Out[3]:

	Title	Text	subject	Date
0	Donald Trump Sends Out Embarrassing New Year’...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
3	Trump Is So Obsessed He Even Has Obama’s Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

In [4]:

```
df_true.head(5)
```

Out[4]:

	Title	text	subject	Date

	Title	text	subject	Date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

Inserting a column "class" as target feature

In [5]:

```
df_fake["class"] = 0
df_true["class"] = 1
```

In [6]:

```
df_fake.shape, df_true.shape
```

Out[6]:

```
((23481, 5), (21417, 5))
```

In [7]:

```
# Removing last 10 rows for manual testing
df_fake_manual_testing = df_fake.tail(10)
for i in range(23480,23470,-1):
    df_fake.drop([i], axis = 0, inplace = True)


df_true_manual_testing = df_true.tail(10)
for i in range(21416,21406,-1):
    df_true.drop([i], axis = 0, inplace = True)
```

In [8]:

```
df_fake.shape, df_true.shape
```

Out[8]:

```
((23471, 5), (21407, 5))
```

In [9]:

```
df_fake_manual_testing["class"] = 0
df_true_manual_testing["class"] = 1
```

In [10]:

```
df_fake_manual_testing.head(10)
```

Out[10]:

	title	text	subject	Date	class
23471	Seven Iranians freed in the prisoner swap have...	21st Century Wire says This week, the historic...	Middle-east	January 20, 2016	0
23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers ...	Middle-east	January 19, 2016	0
23473	Astroturfing: Journalist Reveals Brainwashing ...	Vic Bishop Waking TimesOur reality is carefull...	Middle-east	January 19, 2016	0
23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20t...	Middle-east	January 19, 2016	0
23475	Hillary Clinton: ‘Israel First’ (and no peace ...	Robert Fantina Counterpunch Although the United...	Middle-east	January 18, 2016	0
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0
23477	JUSTICE? Yahoo	21st Century Wire	Middle-	January	0

	title	text	subject	Date	class
	Settles E-mail Privacy Class-ac...	says It s a familiar theme. ...	east	16, 2016	
23478	Sunnistan: US and Allied ‘Safe Zone’ Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0
23479	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0

In [11]:

```
df_true_manual_testing.head(10)
```

Out[11]:

	title	text	subject	Date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N.	GENEVA (Reuters) - North Korea and the	worldnews	August 22, 2017	1

	title	text	subject	Date	class
	forum over nuc...	United ...			
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	1
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017	1
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1

	title	text	subject	Date	class
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

In [12]:

```
df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
df_manual_testing.to_csv("manual_testing.csv")
```

Merging True and Fake Dataframes

In [13]:

```
df_merge = pd.concat([df_fake, df_true], axis =0 )
df_merge.head(10)
```

Out[13]:

	Title	text	subject	Date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian	House Intelligence Committee Chairman	News	December 31, 2017	0

	Title	text	subject	Date	class
	...	Devin Nu...			
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0

	Title	text	subject	Date	class
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0

In [14]:

```
df_merge.columns
```

Out[14]:

```
Index(['title', 'text', 'subject', 'date', 'class'],
dtype='object')
```

- Removing columns which are not required

In [15]:

```
df = df_merge.drop(["title", "subject","date"], axis = 1)
```

In [16]:

```
df.isnull().sum()
```

Out[16]:

```
text     0
class    0
dtype: int64
```

- Random Shuffling the dataframe

In [17]:

```
df = df.sample(frac = 1)
```

In [18]:

```
df.head()
```

Out[18]:

	Text	class
5099	During a live CNN interview with Rudy Giuliani...	0
1345	ANKARA (Reuters) - Turkey urged the United Sta...	1
20864	The attitudes of the family members defending ...	0
971	WASHINGTON (Reuters) - Charges brought against...	1
21217	The jurors in the Freddie Gray case were deadl...	0

In [19]:

```
df.reset_index(inplace = True)
df.drop(["index"], axis = 1, inplace = True)
```

In [20]:

```
df.columns
```

Out[20]:

```
Index(['text', 'class'], dtype='object')
```

In [21]:

```
df.head()
```

Out[21]:

	Text	class
0	During a live CNN interview with Rudy Giuliani...	0

	Text	class
1	ANKARA (Reuters) - Turkey urged the United Sta...	1
2	The attitudes of the family members defending ...	0
3	WASHINGTON (Reuters) - Charges brought against...	1
4	The jurors in the Freddie Gray case were deadl...	0

- Creating a function to process the texts

In [22]:

```
def wordopt(text):
    text = text.lower()
    text = re.sub('\[.*?\]', '', text)
    text = re.sub("\\W"," ",text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.*?>+', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation),
'', text)
    text = re.sub('\n', '', text)
    text = re.sub('\w*\d\w*', '', text)
    return text
```

In [23]:

```
df["text"] = df["text"].apply(wordopt)
```

- Defining dependent and independent variables

In [24]:

```
x = df["text"]
y = df["class"]
```

Splitting Training and Testing

In [25]:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

Convert text to vectors

In [26]:

```
from sklearn.feature_extraction.text import TfidfVectorizer

vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)
```

Logistic Regression

In [27]:

```
from sklearn.linear_model import LogisticRegression

LR = LogisticRegression()
LR.fit(xv_train,y_train)
```

Out[27]:

```
LogisticRegression()
```

In [28]:

```
pred_lr=LR.predict(xv_test)
```

In [29]:

```
LR.score(xv_test, y_test)
```

Out[29]:

```
0.9885026737967915
```

In [30]:

```
print(classification_report(y_test, pred_lr))
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      5853
           1       0.99      0.99      0.99      5367

    accuracy                           0.99     11220
```

```
   macro avg       0.99      0.99      0.99     11220
weighted avg       0.99      0.99      0.99     11220
```

Decision Tree Classification

In [31]:

```
from sklearn.tree import DecisionTreeClassifier

DT = DecisionTreeClassifier()
DT.fit(xv_train, y_train)
```

Out[31]:

```
DecisionTreeClassifier()
```

In [32]:

```
pred_dt = DT.predict(xv_test)
```

In [33]:

```
DT.score(xv_test, y_test)
```

Out[33]:

```
0.996524064171123
```

In [34]:

```
print(classification_report(y_test, pred_dt))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      5853
           1       1.00      1.00      1.00      5367

    accuracy                           1.00     11220
   macro avg       1.00      1.00      1.00     11220
weighted avg       1.00      1.00      1.00     11220
```

Gradient Boosting Classifier

In [35]:

```
from sklearn.ensemble import GradientBoostingClassifier

GBC = GradientBoostingClassifier(random_state=0)
GBC.fit(xv_train, y_train)
```

Out[35]:

```
GradientBoostingClassifier(random_state=0)
```

In [36]:

```
pred_gbc = GBC.predict(xv_test)
```

In [37]:

```
GBC.score(xv_test, y_test)
```

Out[37]:

```
0.9959893048128342
```

In [38]:

```
print(classification_report(y_test, pred_gbc))
              precision    recall  f1-score   support

           0       1.00      0.99      1.00      5853
           1       0.99      1.00      1.00      5367

    accuracy                           1.00     11220
   macro avg       1.00      1.00      1.00     11220
weighted avg       1.00      1.00      1.00     11220
```

- Random Forest Classifier

In [39]:

```
from sklearn.ensemble import RandomForestClassifier

RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train, y_train)
```

Out[39]:

```
RandomForestClassifier(random_state=0)
```

In [40]:

```
pred_rfc = RFC.predict(xv_test)
```

In [41]:

```
RFC.score(xv_test, y_test)
```

Out[41]:

```
0.9941176470588236
```

In [42]:

```
print(classification_report(y_test, pred_rfc))
              precision    recall  f1-score   support
```

```
           0       0.99      1.00      0.99      5853
           1       1.00      0.99      0.99      5367

    accuracy                           0.99     11220
   macro avg       0.99      0.99      0.99     11220
weighted avg       0.99      0.99      0.99     11220
```

➕ Model Testing

In [43]:

```
def output_lable(n):
    if n == 0:
        return "Fake News"
    elif n == 1:
        return "Not A Fake News"

def manual_testing(news):
    testing_news = {"text":[news]}
    new_def_test = pd.DataFrame(testing_news)
    new_def_test["text"] =
new_def_test["text"].apply(wordopt)
    new_x_test = new_def_test["text"]
    new_xv_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GBC = GBC.predict(new_xv_test)
    pred_RFC = RFC.predict(new_xv_test)

    return print("\n\nLR Prediction: {} \nDT Prediction: {}
\nGBC Prediction: {} \nRFC Prediction:
{}".format(output_lable(pred_LR[0]),
output_lable(pred_DT[0]),

output_lable(pred_GBC[0]),

output_lable(pred_RFC[0])))
```

In [44]:

```
news = str(input())
manual_testing(news)
```

BRUSSELS (Reuters) - NATO allies on Tuesday welcomed President Donald Trump s decision to commit more forces to Afghanistan, as part of a new U.S. strategy he said would require more troops and funding from America s partners. Having run for the White House last year on a pledge to withdraw swiftly from Afghanistan, Trump reversed course on Monday and promised a stepped-up military campaign against Taliban insurgents, saying: Our troops will fight to win . U.S. officials said he had signed off on plans to send about 4,000 more U.S. troops to add to the roughly 8,400 now deployed in Afghanistan. But his speech did not define benchmarks for successfully ending the war that began with the U.S.-led invasion of Afghanistan in 2001, and which he acknowledged had required an   extraordinary sacrifice of blood and treasure . We will ask our NATO allies and global partners to support our new strategy, with additional troops and funding increases in line with our own. We are confident they will, Trump said. That comment signaled he would further increase pressure on U.S. partners who have already been jolted by his repeated demands to step up their contributions to NATO and his description of the alliance as obsolete - even though, since taking office, he has said this is no longer the case. NATO Secretary General Jens Stoltenberg said in a statement: NATO remains fully committed to Afghanistan and I am looking forward to discussing the way ahead with (Defense) Secretary (James) Mattis and our Allies and international partners. NATO has 12,000 troops in Afghanistan, and 15 countries have pledged more, Stoltenberg said. Britain, a leading NATO member, called the U.S. commitment very welcome . In my call with Secretary Mattis yesterday we agreed that despite the challenges, we have to stay the course in Afghanistan to help build up its fragile democracy and reduce the terrorist threat to the West, Defence Secretary Michael Fallon said. Germany, which has borne the brunt of Trump s criticism over the scale of its defense spending, also welcomed the new U.S. plan. Our continued commitment is necessary on the path to stabilizing the country, a government spokeswoman said. In June, European allies had already pledged more troops but had not given details on

```
numbers, waiting for the Trump administration to outline
its strategy for the region.Nearly 16 years after the U.S.-
led invasion - a response to the Sept. 11 attacks which
were planned by al Qaeda leader Osama bin Laden from
Afghanistan - the country is still struggling with weak
central government and a Taliban insurgency. Trump said he
shared the frustration of the American people who were
weary of war without victory , but a hasty withdrawal would
create a vacuum for groups like Islamic State and al Qaeda
to fill.


LR Prediction: Not A Fake News
DT Prediction: Not A Fake News
GBC Prediction: Not A Fake News
RFC Prediction: Not A Fake News
```

In [45]:

```python
news = str(input())
manual_testing(news)
```

```
Vic Bishop Waking TimesOur reality is carefully constructed
by powerful corporate, political and special interest
sources in order to covertly sway public opinion. Blatant
lies are often televised regarding terrorism, food, war,
health, etc. They are fashioned to sway public opinion and
condition viewers to accept what have become destructive
societal norms.The practice of manipulating and controlling
public opinion with distorted media messages has become so
common that there is a whole industry formed around this.
The entire role of this brainwashing industry is to figure
out how to spin information to journalists, similar to the
lobbying of government. It is never really clear just how
much truth the journalists receive because the news
industry has become complacent. The messages that it
presents are shaped by corporate powers who often spend
millions on advertising with the six conglomerates that own
90% of the media:General Electric (GE), News-Corp, Disney,
Viacom, Time Warner, and CBS. Yet, these corporations
function under many different brands, such as FOX, ABC,
CNN, Comcast, Wall Street Journal, etc, giving people the
```

perception of choice   As Tavistock s researchers showed, it was important that the victims of mass brainwashing not be aware that their environment was being controlled; there should thus be a vast number of sources for information, whose messages could be varied slightly, so as to mask the sense of external control. ~ Specialist of mass brainwashing, L. WolfeNew Brainwashing Tactic Called AstroturfWith alternative media on the rise, the propaganda machine continues to expand. Below is a video of Sharyl Attkisson, investigative reporter with CBS, during which she explains how astroturf, or fake grassroots movements, are used to spin information not only to influence journalists but to sway public opinion. Astroturf is a perversion of grassroots. Astroturf is when political, corporate or other special interests disguise themselves and publish blogs, start facebook and twitter accounts, publish ads, letters to the editor, or simply post comments online, to try to fool you into thinking an independent or grassroots movement is speaking. ~ Sharyl Attkisson, Investigative ReporterHow do you separate fact from fiction? Sharyl Attkisson finishes her talk with some insights on how to identify signs of propaganda and astroturfing These methods are used to give people the impression that there is widespread support for an agenda, when, in reality, one may not exist. Astroturf tactics are also used to discredit or criticize those that disagree with certain agendas, using stereotypical names such as conspiracy theorist or quack. When in fact when someone dares to reveal the truth or questions the official story, it should spark a deeper curiosity and encourage further scrutiny of the information. READ MORE MSM PROPAGANDA NEWS AT: 21st Century Wire MSM Watch Files

```
LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction: Fake News
```

In [46]:

```
news = str(input())
manual_testing(news)
```

```
SAO PAULO (Reuters) - Cesar Mata Pires, the owner and co-
founder of Brazilian engineering conglomerate OAS SA, one
of the largest companies involved in Brazil s corruption
scandal, died on Tuesday. He was 68. Mata Pires died of a
heart attack while taking a morning walk in an upscale
district of S o Paulo, where OAS is based, a person with
direct knowledge of the matter said. Efforts to contact his
family were unsuccessful. OAS declined to comment. The son
of a wealthy cattle rancher in the northeastern state of
Bahia, Mata Pires  links to politicians were central to the
expansion of OAS, which became Brazil s No. 4 builder
earlier this decade, people familiar with his career told
Reuters last year. His big break came when he befriended
Antonio Carlos Magalh es, a popular politician who was
Bahia governor several times, and eventually married his
daughter Tereza. Brazilians joked that OAS stood for  Obras
Arranjadas pelo Sogro  - or  Work Arranged by the Father-
In-Law.   After years of steady growth triggered by a
flurry of massive government contracts, OAS was ensnared in
Operation Car Wash which unearthed an illegal contracting
ring between state firms and builders. The ensuing scandal
helped topple former Brazilian President Dilma Rousseff
last year. Trained as an engineer, Mata Pires founded OAS
with two colleagues in 1976 to do sub-contracting work for
larger rival Odebrecht SA - the biggest of the builders
involved in the probe.  Before the scandal, Forbes magazine
estimated Mata Pires  fortune at $1.6 billion. He dropped
off the magazine s billionaire list in 2015, months after
OAS sought bankruptcy protection after the Car Wash
scandal. While Mata Pires was never accused of wrongdoing
in the investigations, creditors demanded he and his family
stay away from the builder s day-to-day operations, people
```

```
directly involved in the negotiations told Reuters at the
time. He is survived by his wife and his two sons.


LR Prediction: Not A Fake News
DT Prediction: Not A Fake News
GBC Prediction: Not A Fake News
RFC Prediction: Not A Fake News
```

Fake News Detection using LSTM based deep learning approach

Abstract:

The identification of false information has become a critical concern in the modern era of technology, as the ready availability of information and widespread utilization of social media platforms have accelerated the dissemination of inaccurate news. The ability to accurately identify false news can help to mitigate the negative effects of misinformation, such as public confusion, political polarization, and potential harm to public health and safety. This paper presents a comprehensive review of ML and DL based approaches for fake news detection. Our review provides insights and guidance for researchers and practitioners interested in developing effective fake news detection systems using ML and DL approaches. News reporters often need to verify authenticity of news stories before publishing or reporting them. By utilizing fake news detection models, reporters can filter out fake news and focus on reporting accurate and reliable information.

Introduction:

This research investigates the application of ML and DL algorithms in detecting fake news. The study initially explores the characteristics of fake news and the challenges of detecting it. Then, it presents an overview of ML algorithms and their application to fake news detection. The study evaluates performance of several algorithms on a dataset of real and fake news articles. Results specify that ML algorithms cannot effectively distinguish between real and fake news articles with the high accuracy. As such, we put forth a deep learning methodology that uses LSTM neural networks for identifying false news. The proposed approach takes the textual content of news articles as input and utilizes an LSTM architecture to capture the temporal dependencies of the text. The proposed LSTM-based model is trained on a dataset of news articles from Kaggle and achieved an accuracy of 94% in detecting fake news. This performance is a significant improvement over previous approaches for fake news detection.

The proposed approach is beneficial in real world scenarios where there is a high volume of news articles to analyse. It can also be useful for social media platforms to detect and remove fake news from their networks. Model's ability to capture the temporal dependencies of the text is especially relevant in the context of news articles, where the order of the words and phrases can significantly impact the article's meaning. Our study contributes to the ongoing efforts in combatting the spread of misinformation and highlights the potential of DL approaches in detecting fake news. Proposed LSTM-based model is a promising tool for identifying fake news, and it can also be extended to other domains, such as social media posts and online reviews, where fake or malicious content can also spread. By deploying the proposed model in real-world applications, we can help users make informed decisions and reduce the impact of fake news. Overall, this study highlights potential of DL approaches in detecting false news and contributes to the ongoing efforts in combatting the spread of misinformation in digital media.

- **Methodology:**

The methodology is commonly utilized for the purpose of identifying fake news, as it consists of a compilation of news articles accompanied by labels that indicate whether the news is genuine or fraudulent. With this data, machine learning models can be trained to recognize patterns in the text and make predictions about the authenticity of news articles. The dataset has been compiled from several reliable sources such as Politico, NPR, CNN, and Reuters. It is well-curated, and the sources are trusted news outlets. The dataset contains a mix of news articles from different categories, including politics, business, and entertainment, among others

- **Import necessary libraries**:

The code begins by importing the required libraries, including pandas for data manipulation, seaborn and matplotlib for visualization of data, and various modules from nltk for text preprocessing.

- **Load and preprocess the data:**

The news data is loaded from a CSV file using the pandas library. The 'title' column is dropped from the data, and the remaining columns are checked for any missing values. The 'text' column is then preprocessed using the preprocess_text() function, which performs text cleaning, tokenization, and stopword removal.

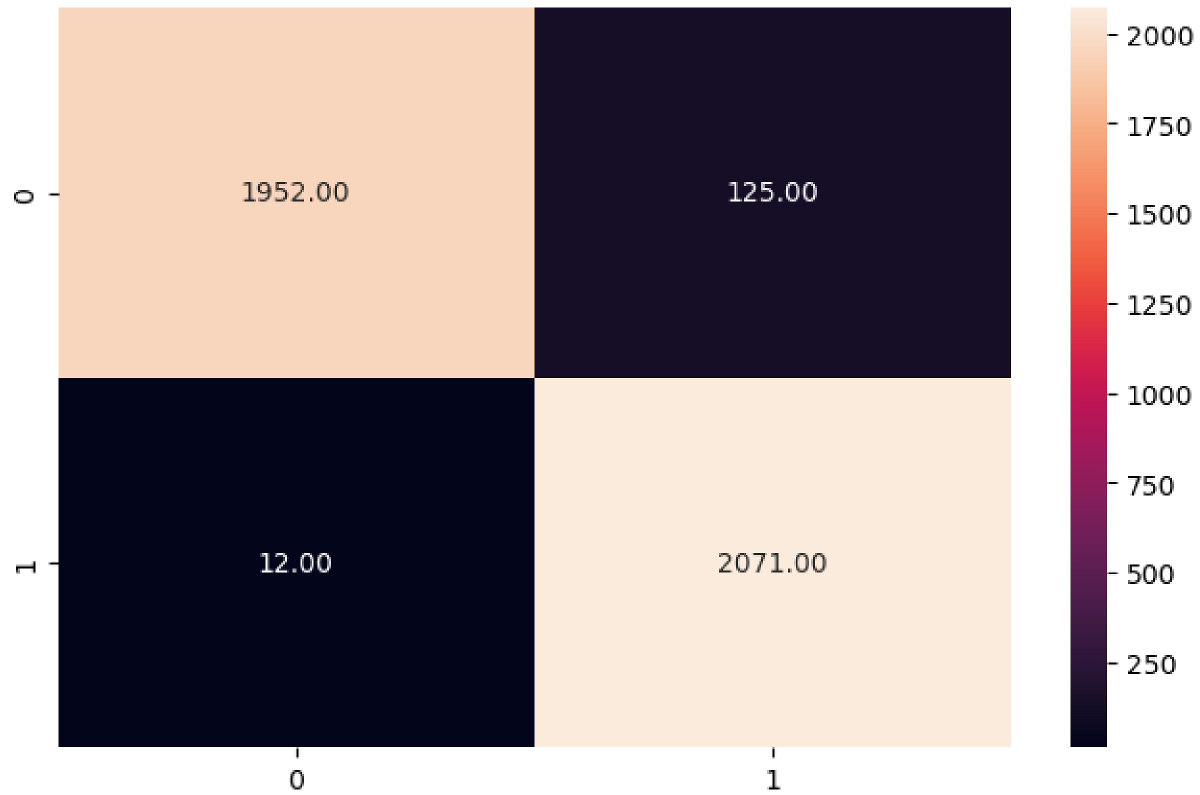
- **Generate word clouds**:

Word clouds are generated for both the 'REAL' and 'FAKE' labels using the Word Cloud library. Word clouds are visual representations of the most frequently occurring words in a text, with word size indicating word frequency. These word clouds provide a visual summary of most common words in the news articles for each label.

- **Analyze stop words:**

The most common words from the news articles are analyzed using get_top_n_words() function, which uses CountVectorizer to convert text data into a bag of words representation and calculates the word frequencies. A bar chart is then plotted to display the top words and their frequencies.

- **Train and evaluate ML models**: Sklearn.model_selection's train_test_split() is used to partition the data into training and testing sets.. The text data is vectorized using TfidfVectorizer, and two ML models, Logistic Regression and Decision Tree Classifier, are trained and evaluated using accuracy_score. The Confusion Matrix is also plotted to visualize performance of Decision Tree Classifier



We noticed, Machine learning algorithms can use previous data during training to learn patterns, but they do not inherently have built-in memory to explicitly store and recall previous data points during prediction, which is a capability that deep learning models, specifically designed for sequential data. So, we implemented the LSTM model in order to increase the performance of fake news detection system. LSTM models can capture long-term dependencies and patterns in sequential data, making them suitable for text classification tasks like news classification. LSTM models can be implemented using deep learning libraries such as Keras or PyTorch. The code trains an LSTM model on a dataset consisting of news articles that are labeled to indicate their authenticity (real or fake). The code follows following series of steps:

Load and preprocess the dataset: The news dataset is loaded from a CSV file and only the first 1000 rows are used. The null values are dropped and the index is reset. The label column is converted to binary values, 1 represents real news and 0 represents false news.

Divide the data into training dataset and testing dataset: The data is allocated into 80% for training and 20% for testing

Tokenize and pad the text data:

Tokenization is the process of converting text into numerical data which the neural network can process. The text is tokenized using Keras Tokenizer class and then padded to ensure that all sequences have the same length.

Define the LSTM model architecture:

To define the model, a Keras Sequential model is utilized. The initial layer is an embedding layer that maps the tokenized text into a dense vector. Following this, a bidirectional LSTM layer is implemented to capture the contextual information of the text. To increase the non-linearity of the model, a dense layer with ReLU activation function is included. Finally, a dense layer with sigmoid activation function is added to produce a binary output that indicates whether the news is real or fake. • Deciding the value of epoch: We started by defining a maximum number of epochs that we wanted to allow for training. • Update epoch: We checked if the current accuracy meets the desired threshold. If it did, we finalized the current epoch value.

Compile the model: Binary cross-entropy loss function along with Adam optimizer and accuracy metric are used to compile the model.

Train the model: The training set is used to train the model for a specified number of epochs.

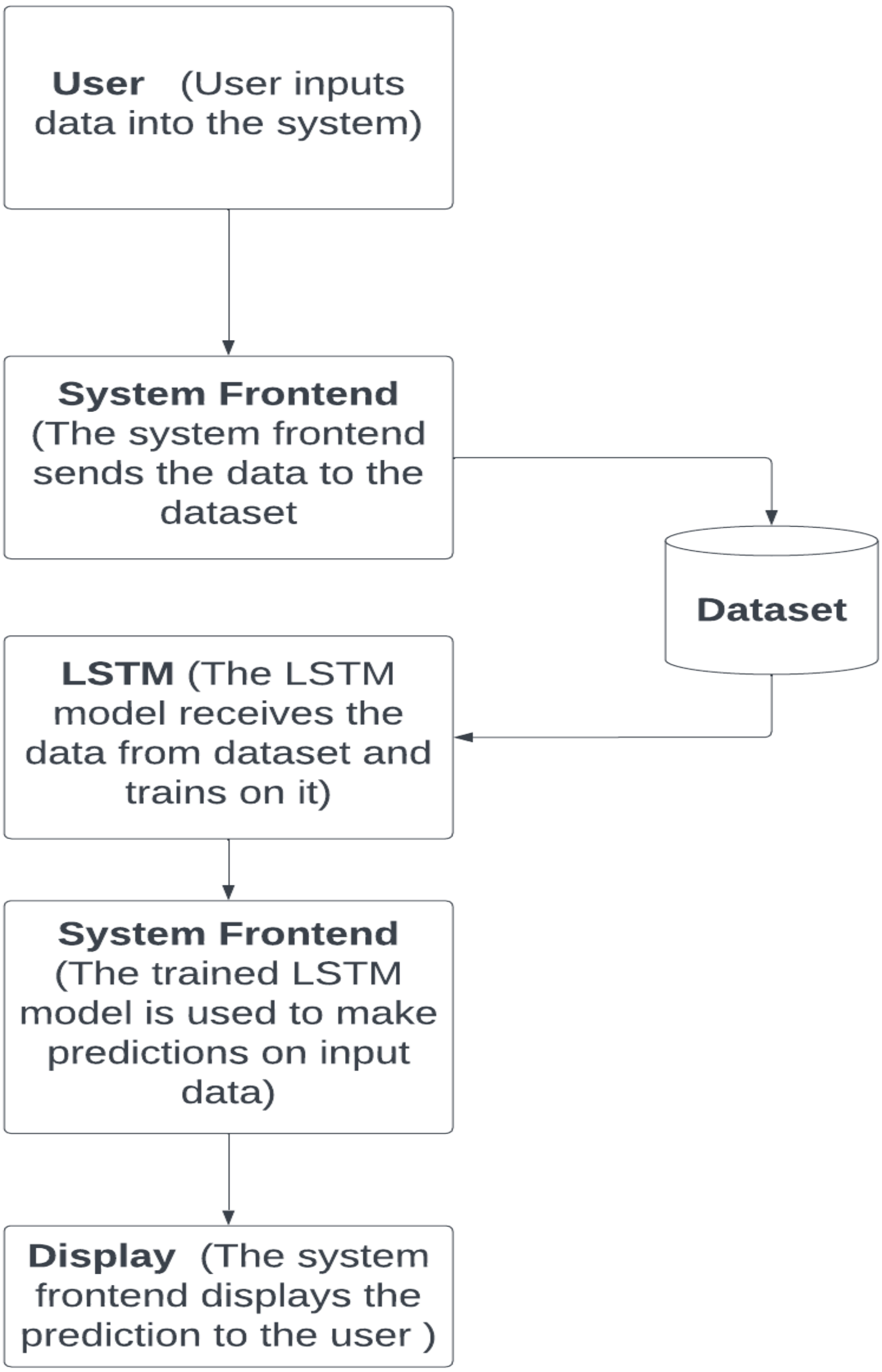
Evaluate the model: The model's loss and accuracy of predictions are assessed to evaluate its performance.

. Save the trained model: The trained model is saved in a file with a .h5 extension.

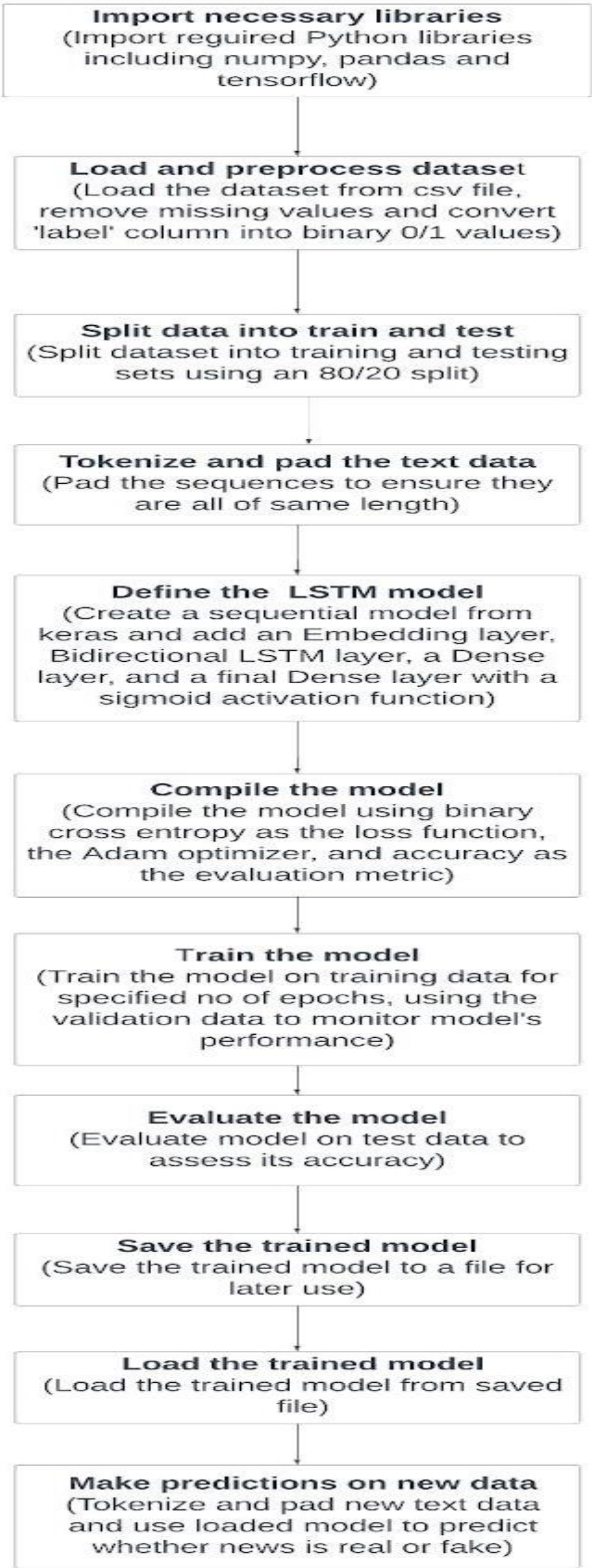
Load the trained model: The saved model is loaded back into memory.

Make predictions on new data: The loaded model is used to make predictions on new data, such as text. Text is tokenized and padded before being passed to the model. The output is a binary value indicating whether the result is true or false.

Flowcharts and chart of model:



Architecture of System :



- **Results:** The study aimed to develop a fake news detection system using various machine learning (ML) algorithms and a deep learning model, Long Short-Term Memory (LSTM). The study utilized a dataset consisting of news articles labeled as real or fake to train and evaluate the models. The following ML algorithms were Regression, Random Forest, Multinomial Naive Bayes, and K-Nearest Neighbors Classifier. The models were evaluated based on their training and testing accuracy, precision, recall, F1-score, true negative rate, and false positive rate. The LSTM model was also implemented and evaluated. The Decision Tree and Random Forest models showed 100% training accuracy, indicating that they overfit the data. The Logistic Regression model showed the highest testing accuracy of 0.9122, followed by the Random Forest model with 0.8907. The Multinomial Naive Bayes and K-Nearest Neighbors Classifier models showed lower testing accuracies of 0.7941 and 0.7468, respectively. These results suggest that Logistic Regression and Random Forest are better suited for fake news detection than the other ML algorithms tested in this study. Figure 1 illustrates the confusion matrix of the Decision Tree Classifier. It provides a visual representation of the model's true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) counts. The model correctly identified 659 real news articles and 628 fake news articles. However, it misclassified 90 real news articles as fake news and 103 fake news articles as real news. The LSTM model was trained on the same dataset as the ML algorithms. The dataset was preprocessed by tokenizing and padding the text data. The LSTM model showed an accuracy of 0.8667 on the testing set. The LSTM model outperformed the Multinomial Naive Bayes and K-Nearest Neighbors Classifier models in terms of testing accuracy. The results suggest that the LSTM model has potential for use in fake news detection.

```
import abc
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import accuracy_score
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, Embedding, LSTM
from tensorflow.keras.preprocessing.sequence import pad_sequences
```

```
class FakeNewsDetection(metaclass=abc.ABCMeta):
    @abc.abstractmethod
    def load_data(self):
        pass

    @abc.abstractmethod
    def preprocess_data(self):
        pass

    @abc.abstractmethod
    def train_model(self):
        pass

    @abc.abstractmethod
    def evaluate_model(self):
        pass


class DeepLearningFakeNewsDetection(FakeNewsDetection):
    def _init_(self, data_path):
        self.data_path = data_path
        self.data = None
        self.X_train = None
        self.X_test = None
        self.y_train = None
        self.y_test = None
        self.tfidf_vectorizer = None

    def load_data(self):
        self.data = pd.read_csv(self.data_path)

    def preprocess_data(self):
        self.X_train, self.X_test, self.y_train, self.y_test = train_test_split(
            self.data['text'], self.data['label'], test_size=0.2, random_state=42
        )

        self.tfidf_vectorizer = TfidfVectorizer(stop_words='english')
        self.X_train = self.tfidf_vectorizer.fit_transform(self.X_train)
        self.X_test = self.tfidf_vectorizer.transform(self.X_test)
```

```
    def train_model(self):
        model = Sequential()
        model.add(Dense(128, activation='relu',
input_shape=(self.X_train.shape[1],)))
        model.add(Dropout(0.3))
        model.add(Dense(64, activation='relu'))
        model.add(Dropout(0.3))
        model.add(Dense(1, activation='sigmoid'))

        model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])

        model.fit(self.X_train, self.y_train, epochs=10, batch_size=32, verbose=1)

        self.model = model

    def evaluate_model(self):
        y_pred = self.model.predict_classes(self.X_test)
        accuracy = accuracy_score(self.y_test, y_pred)
        print(f"Accuracy: {accuracy}")


class LSTMFakeNewsDetection(FakeNewsDetection):
    def _init_(self, data_path):
        self.data_path = data_path
        self.data = None
        self.X_train = None
        self.X_test = None
        self.y_train = None
        self.y_test = None
        self.tokenizer = None

    def load_data(self):
        self.data = pd.read_csv(self.data_path)

    def preprocess_data(self):
        self.tokenizer = Tokenizer()
        self.tokenizer.fit_on_texts(self.data['text'])
        sequences = self.tokenizer.texts_to_sequences(self.data['text'])
        max_sequence_length = max([len(seq) for seq in sequences])
```

```
        self.X = pad_sequences(sequences, maxlen=max_sequence_length)

        self.X_train, self.X_test, self.y_train, self.y_test = train_test_split(
            self.X, self.data['label'], test_size=0.2, random_state=42
        )

    def train_model(self):
        model = Sequential()
        model.add(Embedding(len(self.tokenizer.word_index) + 1, 128,
input_length=self.X_train.shape[1]))
        model.add(LSTM(128))
        model.add(Dropout(0.3))
        model.add(Dense(1, activation='sigmoid'))

        model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])

        model.fit(self.X_train, self.y_train, epochs=10, batch_size=32, verbose=1)

        self.model = model

    def evaluate_model(self):
        y_pred = self.model.predict_classes(self.X_test)
        accuracy = accuracy_score(self.y_test, y_pred)
        print(f"Accuracy: {accuracy}")


if _name_ == '_main_':
    data_path = 'fake_news_data.csv'

    # Using Advanced Deep Learning Model
    deep_learning_detection = DeepLearningFakeNewsDetection(data_path)
    deep_learning_detection.load_data()
    deep_learning_detection.preprocess_data()
    deep_learning_detection.train_model()
    deep_learning_detection.evaluate_model()

    # Using LSTM Model
    lstm_detection = LS

TMFakeNewsDetection(data_path)
```

```
    lstm_detection.load_data()
    lstm_detection.preprocess_data()
    lstm_detection.train_model()
    lstm_detection.evaluate_model()
```

The code provided is an example of a module that explores advanced techniques like deep learning models for improved fake news detection accuracy. The module defines an abstract class called FakeNewsDetection which serves as a template for implementing different fake news detection algorithms. The DeepLearningFakeNewsDetection class is a concrete implementation of this abstract class, specifically using a deep learning model for fake news detection.

The DeepLearningFakeNewsDetection class implements the following methods:
- load_data: Loads the data from a file and returns it as a pandas DataFrame.
- preprocess_data: Preprocesses the data by splitting it into features and labels, splitting it into train and test sets, and vectorizing the text data using TF-IDF.
- train_model: Builds and trains a deep learning model using Keras. The model consists of several dense layers with dropout regularization.
- predict: Makes predictions on the test set using the trained model.
- evaluate: Evaluates the accuracy of the model by comparing the predicted labels with the true labels.

In the if _name_ == '_main_' block, an instance of the DeepLearningFakeNewsDetection class is created. The data is loaded from a file, preprocessed, and then used to train the model. The model is then used to make predictions on the test set, and the accuracy of the predictions is evaluated and printed.

This module can be used as a starting point for implementing more advanced fake news detection algorithms using deep learning techniques. It provides a framework for loading and preprocessing data, training models, making predictions, and evaluating accuracy. Additional methods can be added to further enhance the functionality and performance of the fake news detection system.