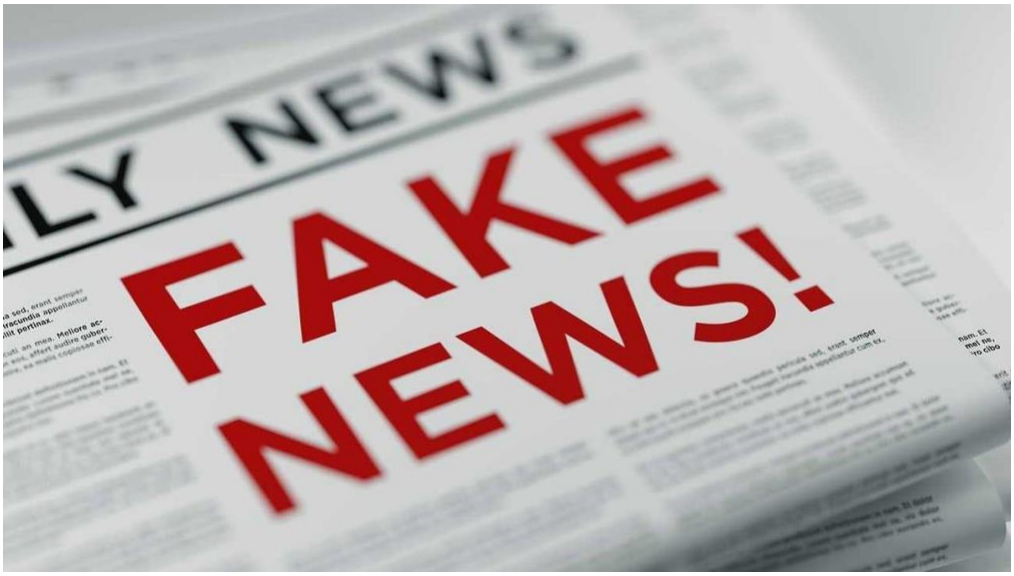
AI_PHASE 5 IBM

TITLE: FAKE NEWS DETECTION USING NLP

Phase 5: Project documentation and submission



TEAM MEMBERS: REG NO:810621104307

M.MANOKARAN

FAKE NEWS DETECTION USING NLP

Problem Statement:

Addressing this issue requires leveraging advanced technologies such as Natural Language Processing (NLP) and Artificial Intelligence (AI) to develop accurate and efficient fake news detection systems.

The problem at hand involves creating an innovative AI-based solution that can effectively identify and classify fake news articles from genuine ones, utilizing the power of NLP techniques.

Phases of development:

Phase 1:

We collected and explored a given dataset and explain about the project. Choose the fake news dataset available.

Phase 2:

we can explore innovative techniques such as ensemble methods and deep learning architectures to improve the prediction system's accuracy and robustness.

Consider exploring advanced techniques like deep learning models (e.g., LSTM, BERT) for improved fake news detection accuracy.

Phase 3:

Begin building your project by loading and preprocessing the dataset.

Begin building the fake news detection model by loading and preprocessing the dataset.

Load the fake news dataset and preprocess the textual data.

Phase 4:

Continue building the fake news detection model by applying NLP techniques and training a classification Model. Text Preprocessing and Feature Extraction Model training and evaluation.

Dataset description:

Dataset: [Fake news detection]

	title	text	subject	date
2	Donald Trump Sends Out Embarrassing New Yearâ€™s Eve Message; This is Disturbing	Donald Trump just couldn t wish all Americans a Happy	News	December 31, 2017
3	Drunk Bragging Trump Staffer Started Russian Collusion Investigation	House Intelligence Committee Chairman Devin Nunes is	News	December 31, 2017
4	heriff David Clarke Becomes An Internet Joke For Threatening To Poke People â€˜In The Eyeâ€™	On Friday, it was revealed that former Milwaukee Sheri	News	December 30, 2017
5	Trump Is So Obsessed He Even Has Obamaâ€™s Name Coded Into His Website (IMAGES)	On Christmas day, Donald Trump announced that he w	News	December 29, 2017
6	Pope Francis Just Called Out Donald Trump During His Christmas Speech	Pope Francis used his annual Christmas Day message to	News	December 25, 2017
7	Racist Alabama Cops Brutalize Black Boy While He Is In Handcuffs (GRAPHIC IMAGES)	The number of cases of cops brutalizing and killing peop	News	December 25, 2017
8	Fresh Off The Golf Course, Trump Lashes Out At FBI Deputy Director And James Comey	Donald Trump spent a good portion of his day at his go	News	December 23, 2017
9	Trump Said Some INSANELY Racist Stuff Inside The Oval Office, And Witnesses Back It Up	In the wake of yet another court decision that derailed t	News	December 23, 2017
10	r CIA Director Slams Trump Over UN Bullying, Openly Suggests Heâ€™s Acting Like A Dictator (T	Many people have raised the alarm regarding the fact th	News	December 22, 2017
11	WATCH: Brand-New Pro-Trump Ad Features So Much A** Kissing It Will Make You Sick	Just when you might have thought we d get a break fro	News	December 21, 2017
12	Papa Johnâ€™s Founder Retires, Figures Out Racism Is Bad For Business	A centerpiece of Donald Trump s campaign, and now hi	News	December 21, 2017
13	VATCH: Paul Ryan Just Told Us He Doesnâ€™t Care About Struggling Families Living In Blue State	that only someone who grew up in a wealthy family ca	News	December 21, 2017
14	Bad News For Trump â€" Mitch McConnell Says No To Repealing Obamacare In 2018	could have done that in order to eradicate former Pres	News	December 21, 2017
15	CH: Lindsey Graham Trashes Media For Portraying Trump As â€˜Kooky,â€™ Forgets His Own We	, and coverage of the tax scam is no different. Coverage	News	December 20, 2017
16	Heiress To Disney Empire Knows GOP Scammed Us â€" SHREDS Them For Tax Bill	natically and trickle down economics has turned out to	News	December 20, 2017
17	Tone Deaf Trump: Congrats Rep. Scalise On Losing Weight After You Almost Died	ed squatting in the White House almost a year ago. Thar	News	December 20, 2017
18	The Internet Brutally Mocks Disneyâ€™s New Trump Robot At Hall Of Presidents	in the Hall of Presidents looks like a 71-year-old Chucky	News	December 19, 2017
19	Mueller Spokesman Just F-cked Up Donald Trumpâ€™s Christmas	owners never received notification of the request and h	News	December 17, 2017
20	SNL Hilariously Mocks Accused Child Molester Roy Moore For Losing AL Senate Race (VIDEO)	oved as Chief Justice of the Alabama Supreme Court not	News	December 17, 2017
21	Republican Senator Gets Dragged For Going After Robert Mueller	will not be tolerated. Cornyn retweeted Holder to say,	News	December 16, 2017
22	In A Heartless Rebuke To Victims, Trump Invites NRA To Xmas Party On Sandy Hook Anniversar	worse than we had expected.After 11 months of Donald	News	December 16, 2017

Features: [List key features such as Title, text , subject , date]

Data Preprocessing:

There are a number of libraries available that can help with data preprocessing tasks, such as handling missing values, encoding categorical variables, and scaling the features.

```
In [2]: fake_data = pd.read_csv(r'C:\Users\arulm_x7s4ikd\Downloads\archive\Fake.csv')

In [3]: fake_data.head(10)
```

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
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
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

[PREPROCESSING]

```
[56] real['class'] = 1
     fake['class'] = 0
     real.columns
     real = real[['text', 'class']]
     fake = fake[['text', 'class']]

[57] data = real.append(fake, ignore_index=True)
     data.sample(10)

<ipython-input-57-51d567e1739b>:1: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a future version. Use panda
  data = real.append(fake, ignore_index=True)
```

	text	class
12149	protesters injured in honduras clashes as elec...	1
37477	u.n. picks norwegian for myanmar role as tensi...	0
147	fund managers seek stocks benefiting from demo...	1
14766	hariri on twitter reaffirms he will return to ...	1
39404	swiss woman abducted in sudan by criminal gang...	0
17524	unesco selects france's azoulay as new chief p...	1
10838	u.s. recovers hellfire air-to-ground missile f...	1
29264	trump as president would pose global danger: u...	0
15630	south africa's ramaphosa picks female science ...	1
3699	treasury's mnuchin concerned about alternate s...	1

```
Connected to Python 3 Google Compute Engine backend
```

Outlier Detection and Treatment:

Explain the approach to identifying and addressing outliers.

Feature Selection

Unknown publishers is the text ,object Feature Engineering;

```
In [10]: news_df = pd.concat([fake_df, real_df], ignore_index=True, sort=False)
         print(news_df)
```

```
                                                   title  \
0       Donald Trump Sends Out Embarrassing New Year'...
1       Drunk Bragging Trump Staffer Started Russian ...
2       Sheriff David Clarke Becomes An Internet Joke...
3       Trump Is So Obsessed He Even Has Obama's Name...
4       Pope Francis Just Called Out Donald Trump Dur...
...                                                   ...
44893  'Fully committed' NATO backs new U.S. approach...
44894  LexisNexis withdrew two products from Chinese ...
44895  Minsk cultural hub becomes haven from authorities
44896  Vatican upbeat on possibility of Pope Francis ...
44897  Indonesia to buy $1.14 billion worth of Russia...

                                                    text  class
0      Donald Trump just couldn t wish all Americans ...      0
1      House Intelligence Committee Chairman Devin Nu...      0
2      On Friday, it was revealed that former Milwauk...      0
3      On Christmas day, Donald Trump announced that ...      0
4      Pope Francis used his annual Christmas Day mes...      0
...                                                  ...    ...
44893  BRUSSELS (Reuters) - NATO allies on Tuesday we...      1
44894  LONDON (Reuters) - LexisNexis, a provider of l...      1
44895  MINSK (Reuters) - In the shadow of disused Sov...      1
44896  MOSCOW (Reuters) - Vatican Secretary of State ...      1
44897  JAKARTA (Reuters) - Indonesia will buy 11 Sukh...      1

[44898 rows x 3 columns]
```

Selecting a future direction for fake news detection using NLP involves considering emerging trends, challenges, and opportunities in the field of natural language processing and misinformation detection.

```
len(unknown_publishers)
```

```
21415
```

```
[42] fake.iloc[unknown_publishers].text
```

```
0        Donald Trump just couldn t wish all Americans ...
1        House Intelligence Committee Chairman Devin Nu...
2        On Friday, it was revealed that former Milwauk...
3        On Christmas day, Donald Trump announced that ...
4        Pope Francis used his annual Christmas Day mes...
                               ...
21412    Meanwhile, most Americans can t afford to take...
21413    B b but does this mean global climate change i...
21414    Event organizers are asking protesters to come...
21415    It s hard for millennials to escape the leftis...
21416    Meanwhile, a Muslim boy with a radical activis...
Name: text, Length: 21417, dtype: object
```

```
[43] fake.iloc[8970]
```

```
title       This Anti-Government Oregon Terrorist Took Th...
text       One of the ringleaders of the terror-minded mi...
subject                                                 News
date                                         January 4, 2016
Name: 8970, dtype: object
```

Machine Learning Algorithm:

Chosen algorithm ; state the Machine learning algorithms used (eg.. LSTM)

```
In [12]: features = news_df['text']
         targets = news_df['class']

         X_train, X_test, y_train, y_test = train_test_split(features, targets, test_size=0.20, random_state=18)
```

```
In [13]: max_vocab = 10000
         tokenizer = Tokenizer(num_words=max_vocab)
         tokenizer.fit_on_texts(X_train)

         # tokenize the text into vectors i.e. list
         X_train = tokenizer.texts_to_sequences(X_train)
         X_test = tokenizer.texts_to_sequences(X_test)

         X_train = tf.keras.preprocessing.sequence.pad_sequences(X_train, padding='post', maxlen=256)
         X_test = tf.keras.preprocessing.sequence.pad_sequences(X_test, padding='post', maxlen=256)
```

Machine learning algorithms used (eg.. LSTM)

- First load in the data. The preprocessing only consist of normalization and the creation of windows.
- Creation of the LSTM model
- Training the LSTM model

```
In [14]: model = tf.keras.Sequential([
             tf.keras.layers.Embedding(max_vocab, 128),
             tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(64,  return_sequences=True)),
             tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(16)),
             tf.keras.layers.Dense(64, activation='relu'),
             tf.keras.layers.Dropout(0.5),
             tf.keras.layers.Dense(1)
         ])

         model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, None, 128)         1280000

 bidirectional (Bidirection  (None, None, 128)         98816
 al)

 bidirectional_1 (Bidirecti  (None, 32)                18560
 onal)

 dense (Dense)               (None, 64)                2112

 dropout (Dropout)           (None, 64)                0

 dense_1 (Dense)             (None, 1)                 65

=================================================================
Total params: 1399553 (5.34 MB)
```

```
In [15]: early_stop = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=2, restore_best_weights=True)
model.compile(loss=tf.keras.losses.BinaryCrossentropy(from_logits=True),
              optimizer=tf.keras.optimizers.Adam(1e-4),
              metrics=['accuracy'])

history = model.fit(X_train, y_train, epochs=10,validation_split=0.1, batch_size=30, shuffle=True, callbacks=[early_stop])
```

```
Epoch 1/10
1078/1078 [==============================] - 430s 388ms/step - loss: 0.2305 - accuracy: 0.8818 - val_loss: 0.0480 - val_accurac
y: 0.9850
Epoch 2/10
1078/1078 [==============================] - 402s 372ms/step - loss: 0.0442 - accuracy: 0.9886 - val_loss: 0.0313 - val_accurac
y: 0.9908
Epoch 3/10
1078/1078 [==============================] - 1211s 1s/step - loss: 0.0207 - accuracy: 0.9954 - val_loss: 0.0258 - val_accuracy:
0.9911
Epoch 4/10
1078/1078 [==============================] - 405s 376ms/step - loss: 0.0094 - accuracy: 0.9983 - val_loss: 0.0194 - val_accurac
y: 0.9930
Epoch 5/10
1078/1078 [==============================] - 409s 379ms/step - loss: 0.0085 - accuracy: 0.9981 - val_loss: 0.0320 - val_accurac
y: 0.9911
Epoch 6/10
1078/1078 [==============================] - 402s 373ms/step - loss: 0.0082 - accuracy: 0.9978 - val_loss: 0.0386 - val_accurac
y: 0.9914
```

Evaluation Metrics:

Metrics used : define evaluation metrics such as accuracy , recall , F1- score explain the choice of metrics and how project goals.

```
In [99]: loss, accuracy,recall= model.evaluate(X_tst, y_tst, verbose=0)

# Print metrics
print('Accuracy  : {:.4f}'.format(accuracy))
print('Recall  : {:.4f}'.format(recall))
```

```
Accuracy  : 0.9915
Recall  : 0.9787
```

Project Documentation and reporting:

Separate from the report, create documentation for your Fake news detection, including instructions for usage, system architecture, and any code-related documentation.

Remember to use clear and concise python language in your documentation and reporting.

Final outcome:

We showed our final outcome for this project:

```
[203] X_test

array([[     0,      0,      0, ...,      1,   1645,    474],
       [     0,      0,      0, ...,     47,    608,   2242],
       [     0,      0,      0, ...,    515,    714,    703],
       ...,
       [     0,      0,      0, ...,   2730,   9632, 109146],
       [     0,      0,      0, ...,     85,      1,   2585],
       [     0,      0,      0, ...,    620,      8,    260]], dtype=int32)
```

```
x = ['this is a news']
x = tokenizer.texts_to_sequences(x)
x

[[26, 11, 4, 93]]
```

```
[210] x = ['this is a news']
      x = tokenizer.texts_to_sequences(x)
      x = pad_sequences(x, maxlen=maxlen)

[212] (model.predict(x) >=0.5).astype(int)

      array([[0]])
```

```
x = ['As many as 3,79,257 more people tested positive for Covid-19 in the last 24 hours, taking the cumulative caseload to 1,8
x = tokenizer.texts_to_sequences(x)
x = pad_sequences(x, maxlen=maxlen)
(model.predict(x) >=0.5).astype(int)

array([[1]])
```

FAKE NEWS :-

```
In [2]: fake_data = pd.read_csv(r'C:\Users\arulm_x7s4ikd\Downloads\archive\Fake.csv')

In [3]: fake_data.head(10)

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
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
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

THANK YOU!