

# Fake News Detection with Machine Learning

A Project presented by **Group 15**

GUIDE- Prof. Dr. Parnika Kekare



# Group Members

## Group 15

**RISHABH JHA**

2175050

**PRASHANSA BANSAL**

2175043

**SAHIL GAVTHANKAR**

2175054

**VIVEK ROTE**

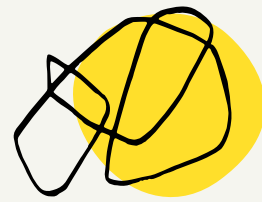
2175065

# About Content Refining

Content Refining is one of the biggest problems faced by Social Media Companies as well as a news organization. While this job was done by humans(Content Moderators) but due to the fact that there are irregularities in human behavior and it takes a huge toll on mental health therefore nowadays this work is done through AI using Machine Learning.

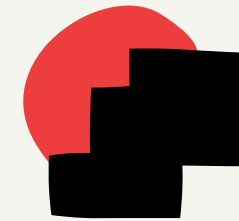


# The Problem



## **What we want to solve**

Detecting Fake news i.e to create a tool that will refine and remove fake news using AI technologies like ML and NLP.



## **Hypothesis**

Our main hypothesis was to create a model that would detect fake news as it is a primary source of spreading hate and panic among masses

# Objectives

To create a tool that will detect and remove fake news articles.

---

The primary goal was to detect fake news by scanning the string but in the further part of the project we want to our model be able to detect fake photos and video resources.

---

While getting the result is an important step but we also want to work on performance, so that our tool takes less time but do the same task. Accuracy has been the main focus

# Technology

- PYTHON
- NLP
- ML
- AI

# Example



## EXAMPLE HEADLINE

"Robert Plant Ripped up \$800M Led Zeppelin Reunion Contract"

## EXAMPLE SNIPPETS FROM BODY TEXTS AND CORRECT CLASSIFICATIONS

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."

CORRECT CLASSIFICATION: AGREE

"... No, Robert Plant did not rip up an \$800 million deal to get Led Zeppelin back together. ..."

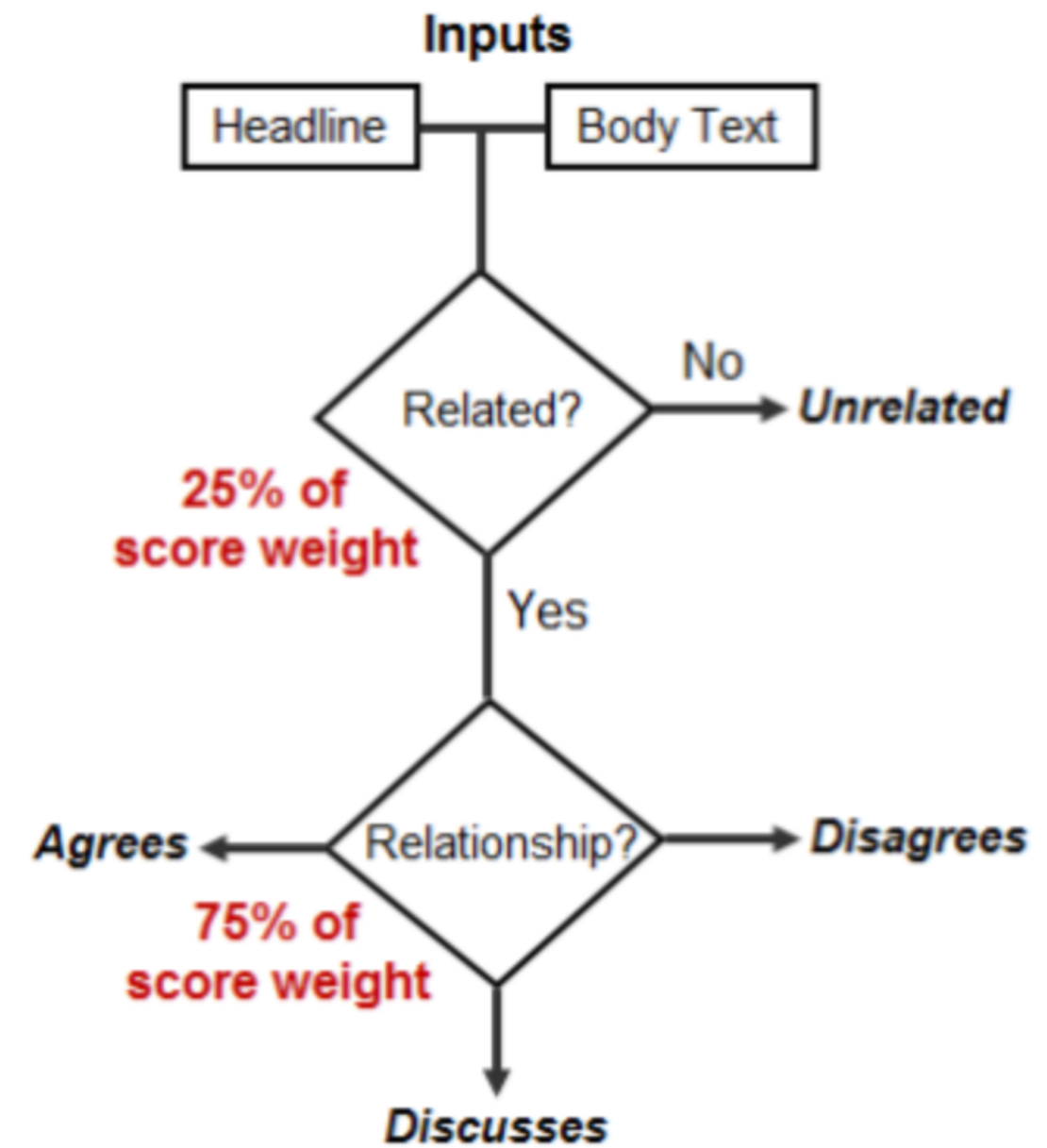
CORRECT CLASSIFICATION: DISAGREE

"... Robert Plant reportedly tore up an \$800 million Led Zeppelin reunion deal. ..."

CORRECT CLASSIFICATION: DISCUSSES

"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."

CORRECT CLASSIFICATION: UNRELATED



# MODULES







# List of Modules.

- Data cleaning
- Tokenization
- Data Visualization
- NLP
- Model Fitting

# DATA CLEANING

```
# Show cleaned up news after removing stopwords  
print(df['clean'][0])
```

```
['budget', 'fight', 'looms', 'republicans', 'flip', 'fiscal', 'script', 'washington', 'reuters', 'head', 'conservative', 'repu  
lican', 'faction', 'congress', 'voted', 'month', 'huge', 'expansion', 'national', 'debt', 'cuts', 'called', 'fiscal', 'conserv  
ative', 'sunday', 'urged', 'budget', 'restraint', 'keeping', 'sharp', 'pivot', 'republicans', 'representative', 'mark', 'meadow  
s', 'speaking', 'face', 'nation', 'drew', 'hard', 'line', 'federal', 'spending', 'lawmakers', 'bracing', 'battle', 'january',  
'return', 'holidays', 'wednesday', 'lawmakers', 'begin', 'trying', 'pass', 'federal', 'budget', 'fight', 'likely', 'linked', '  
ssues', 'immigration', 'policy', 'november', 'congressional', 'election', 'campaigns', 'approach', 'republicans', 'seek', 'con  
rol', 'congress', 'president', 'donald', 'trump', 'republicans', 'want', 'budget', 'increase', 'military', 'spending', 'democr  
ts', 'want', 'proportional', 'increases', 'defense', 'discretionary', 'spending', 'programs', 'support', 'education', 'scienti  
ic', 'research', 'infrastructure', 'public', 'health', 'environmental', 'protection', 'trump', 'administration', 'willing', 'g  
ing', 'increase', 'defense', 'discretionary', 'spending', 'percent', 'meadows', 'chairman', 'small', 'influential', 'house', '  
reedom', 'caucus', 'said', 'program', 'democrats', 'saying', 'need', 'government', 'raise', 'percent', 'fiscal', 'conservativ  
e', 'rationale', 'eventually', 'people', 'money', 'said', 'meadows', 'republicans', 'voted', 'late', 'december', 'party', 'deb  
t', 'financed', 'overhaul', 'expected', 'balloon', 'federal', 'budget', 'deficit', 'trillion', 'years', 'trillion', 'national'  
'debt', 'interesting', 'hear', 'mark', 'talk', 'fiscal', 'responsibility', 'democratic', 'representative', 'joseph', 'crowley'  
'said', 'crowley', 'said', 'republican', 'require', 'united', 'states', 'borrow', 'trillion', 'paid', 'future', 'generations',  
'finance', 'cuts', 'corporations', 'rich', 'fiscally', 'responsible', 'bills', 'seen', 'passed', 'history', 'house', 'represen  
atives', 'think', 'going', 'paying', 'years', 'come', 'crowley', 'said', 'republicans', 'insist', 'package', 'biggest', 'overh  
ul', 'years', 'boost', 'economy', 'growth', 'house', 'speaker', 'paul', 'ryan', 'supported', 'recently', 'went', 'meadows', 'm  
king', 'clear', 'radio', 'interview', 'welfare', 'entitlement', 'reform', 'party', 'calls', 'republican', 'priority', 'republi  
an', 'parlance', 'entitlement', 'programs', 'mean', 'food', 'stamps', 'housing', 'assistance', 'medicare', 'medicaid', 'healt  
h', 'insurance', 'elderly', 'poor', 'disabled', 'programs', 'created', 'washington', 'assist', 'needy', 'democrats', 'seized',  
'ryan', 'early', 'december', 'remarks', 'saying', 'showed', 'republicans', 'overhaul', 'seeking', 'spending', 'cuts', 'social'  
'programs', 'goals', 'house', 'republicans', 'seat', 'senate', 'votes', 'democrats', 'needed', 'approve', 'budget', 'prevent',  
'government', 'shutdown', 'democrats', 'leverage', 'senate', 'republicans', 'narrowly', 'control', 'defend', 'discretionary',  
'defense', 'programs', 'social', 'spending', 'tackling', 'issue', 'dreamers', 'people', 'brought', 'illegally', 'country', 'ch  
ldren', 'trump', 'september', 'march', 'expiration', 'date', 'deferred', 'action', 'childhood', 'arrivals', 'daca', 'program',  
'protects', 'young', 'immigrants', 'deportation', 'provides', 'work', 'permits', 'president', 'said', 'recent', 'twitter', 'me  
sages', 'wants', 'funding', 'proposed', 'mexican', 'border', 'wall', 'immigration', 'changes', 'exchange', 'agreeing', 'help',  
'dreamers', 'representative', 'debbie', 'dingell', 'told', 'favor', 'linking', 'issue', 'policy', 'objectives', 'wall', 'fundi  
g', 'need', 'daca', 'clean', 'said', 'wednesday', 'trump', 'aides', 'meet', 'congressional', 'leaders', 'discuss', 'issues', '  
ollowed', 'weekend', 'strategy', 'sessions', 'trump', 'republican', 'leaders', 'white', 'house', 'said', 'trump', 'scheduled',  
'meet', 'sunday', 'florida', 'republican', 'governor', 'rick', 'scott', 'wants', 'emergency', 'house', 'passed', 'billion', 'p  
ckage', 'hurricanes', 'florida', 'texas', 'puerto', 'rico', 'wildfires', 'california', 'package', 'exceeded', 'billion', 'requ  
sted', 'trump', 'administration', 'senate', 'voted']
```



```
In [20]: # join the words into a string
df['clean_joined'] = df['clean'].apply(lambda x: " ".join(x))
```

```
In [21]: df
```

```
Out[21]:
```

	title	text	subject	isfake	original	clean	clean_joined
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	1	As U.S. budget fight looms, Republicans flip t...	[budget, fight, looms, republicans, flip, fisc...	budget fight looms republicans flip fiscal scr...
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	1	U.S. military to accept transgender recruits o...	[military, accept, transgender, recruits, mond...	military accept transgender recruits monday pe...
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	1	Senior U.S. Republican senator: 'Let Mr. Muell...	[senior, republican, senator, mueller, washing...	senior republican senator mueller washington r...
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	1	FBI Russia probe helped by Australian diplomat...	[russia, probe, helped, australian, diplomat, ...	russia probe helped australian diplomat washin...
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	1	Trump wants Postal Service to charge 'much mor...	[trump, wants, postal, service, charge, amazon...	trump wants postal service charge amazon shipm...
...	...	...	...	...	...	...	...
44893	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	0	McPain: John McCain Furious That Iran Treated ...	[mcpain, john, mccain, furious, iran, treated,...	mcpain john mccain furious iran treated sailor...
44894	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	0	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	[justice, yahoo, settles, mail, privacy, class...	justice yahoo settles mail privacy class actio...
44895	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	0	Sunnistan: US and Allied 'Safe Zone' Plan to T...	[sunnistan, allied, safe, zone, plan, territor...	sunnistan allied safe zone plan territorial bo...
44896	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	0	How to Blow \$700 Million: Al Jazeera America F...	[blow, million, jazeera, america, finally, cal...	blow million jazeera america finally calls qui...
44897	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	0	10 U.S. Navy Sailors Held by Iranian Military ...	[navy, sailors, held, iranian, military, signs...	navy sailors held iranian military signs neoco...

44898 rows × 7 columns

# TOKENIZATION

```
In [32]: print("The encoding for document\n",df.clean_joined[0],"\n is : ",train_sequences[0])
```

The encoding for document

budget fight looms republicans flip fiscal script washington reuters head conservative republican faction congress voted month huge expansion national debt cuts called fiscal conservative sunday urged budget restraint keeping sharp pivot republicans representative mark meadows speaking 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 november congressional election campaigns approach republicans seek control congress president donald trump republicans want budget increase military spending democrats want proportional increases defense discretionary spending programs support education scientific research infrastructure public health environmental protection trump administration willing going increase defense discretionary spending percent meadows chairman small influential house freedom caucus said program democrats saying need government raise percent fiscal conservative rationale eventually people money said meadows republicans voted late december party debt financed overhaul expected balloon federal budget deficit trillion years trillion national debt interesting hear mark talk fiscal responsibility democratic representative joseph crowley said crowley said republican require united states borrow trillion paid future generations finance cuts corporations rich fiscally responsible bills seen passed history house representatives think going paying years come crowley said republicans insist package biggest overhaul years boost economy growth house speaker paul ryan supported recently went meadows making clear radio interview welfare entitlement reform party calls republican priority republican parlance entitlement programs mean food stamps housing assistance medicare medicaid health insurance elderly poor disabled programs created washington assist needy democrats seized ryan early december remarks saying showed republicans overhaul seeking spending cuts social programs goals house republicans seat senate votes democrats needed approve budget prevent government shutdown democrats leverage senate republicans narrowly control defend discretionary defense programs social spending tackling issue dreamers people brought illegally country children trump september march expiration date deferred action childhood arrivals daca program protects young immigrants deportation provides work permits president said recent twitter messages wants funding proposed mexican border wall immigration changes exchange agreeing help dreamers representative debbie dingell told favor linking issue policy objectives wall funding need daca clean said wednesday trump aides meet congressional leaders discuss issues followed weekend strategy sessions trump republican leaders white house said trump scheduled meet sunday florida republican governor rick scott wants emergency house passed billion package hurricanes florida texas puerto rico wildfires california package exceeded billion requested trump administration senate voted

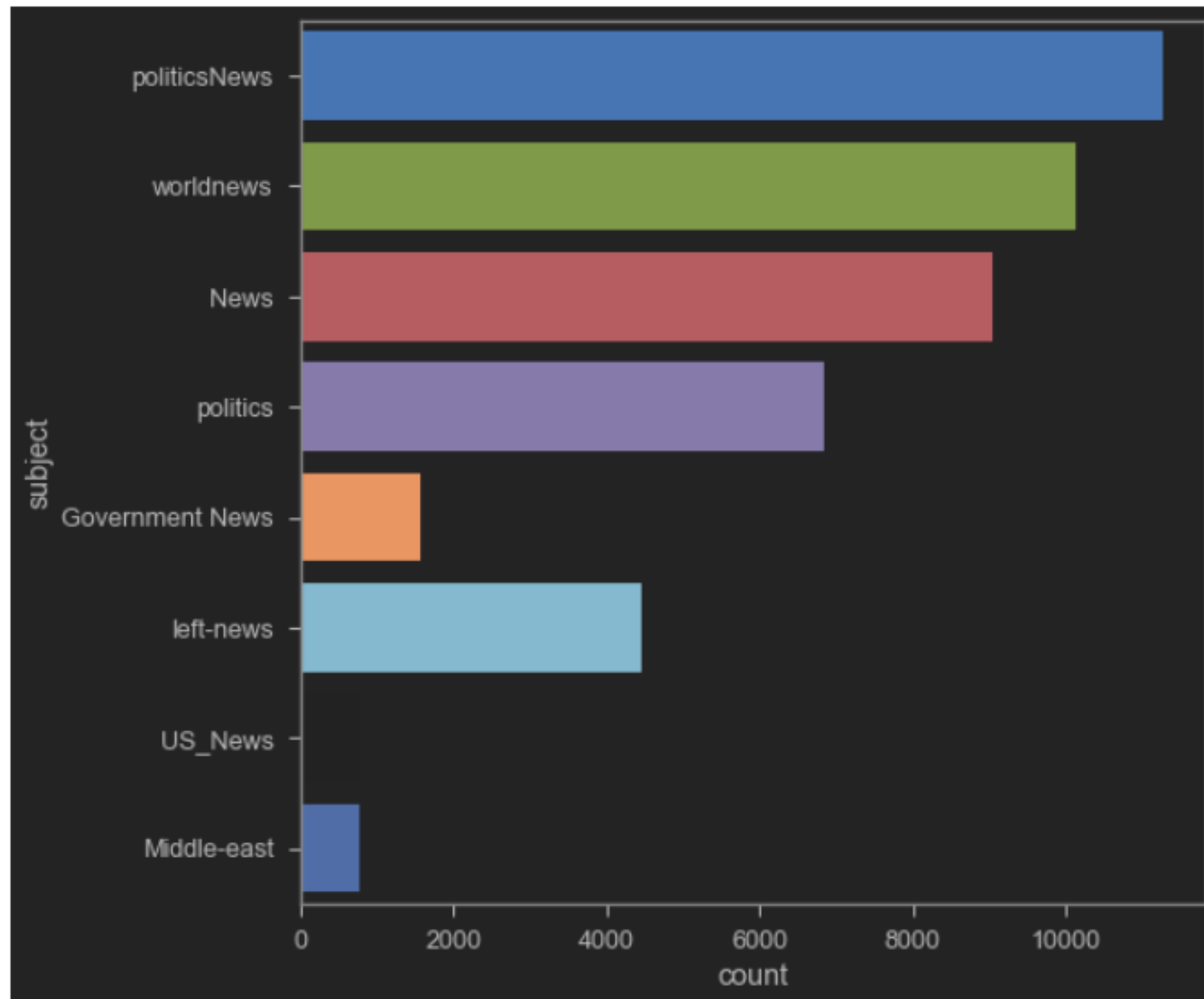
is : [1054, 51, 23, 4707, 26037, 7072, 510, 7802, 205, 1, 843, 2546, 2297, 10782, 536, 5887, 814, 6279, 264, 3443, 7305, 2538, 27, 8050, 278, 7265, 21147, 44, 20, 16, 12, 903, 24, 7, 183, 23, 470, 783, 27, 3532, 3443, 1509, 4768, 278, 7265, 21147, 5388, 2479, 3080, 4, 6452, 5777, 303, 2801, 139, 1446, 303, 4, 657, 7802, 205, 181, 21148, 29, 691, 10, 1, 19, 1564, 837, 4, 1067, 27, 1030, 2676, 27, 69, 11190, 478, 160, 2353, 695, 24, 7, 19, 490, 1016, 12131, 3773, 2801, 799, 5815, 5404, 3080, 9347, 9347, 2295, 77, 799, 2801, 2353, 373, 3589, 651, 1521, 573, 802, 7, 19, 97, 2677, 27, 120, 273, 20, 22, 506, 307, 2142]



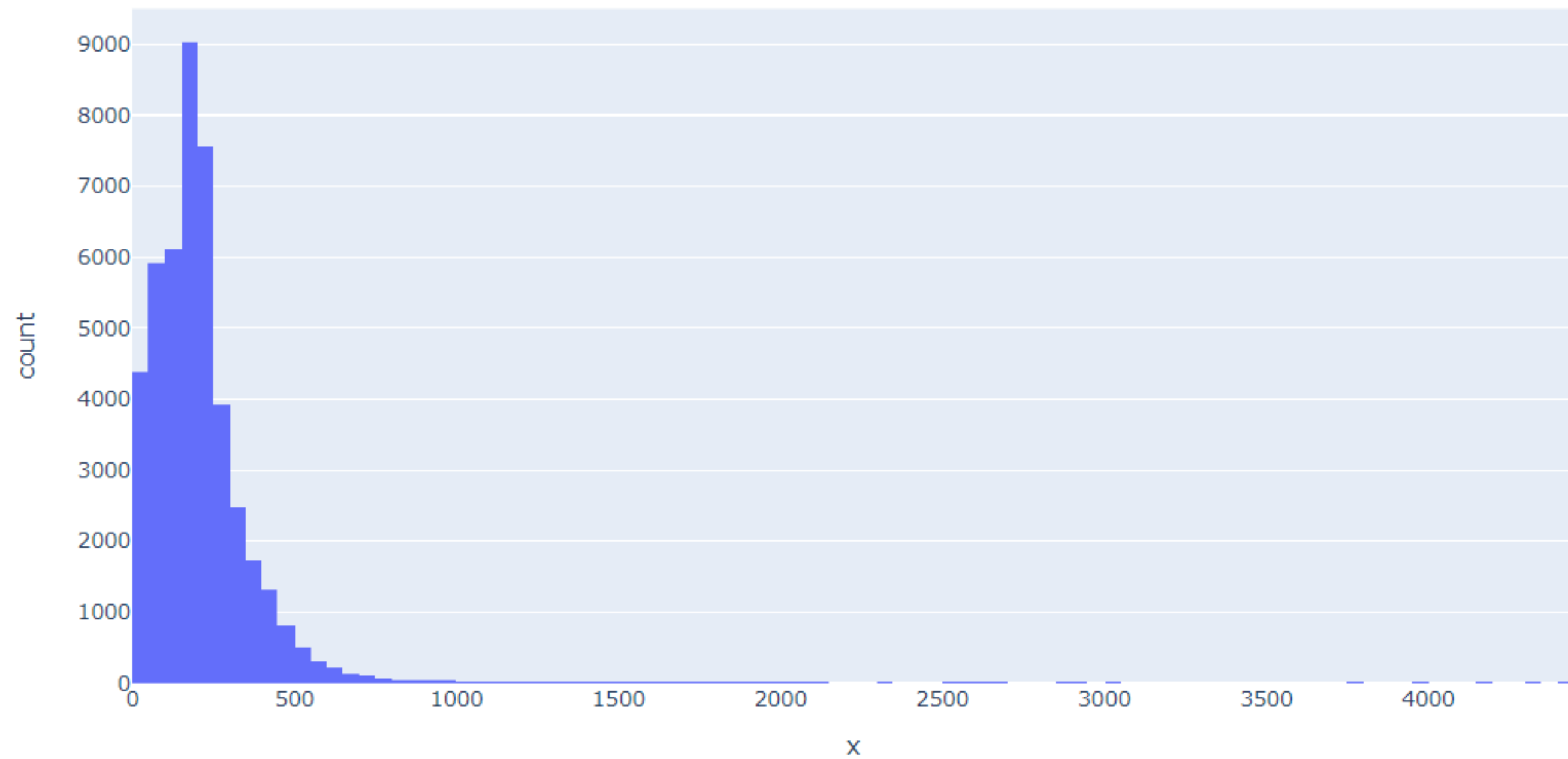
# VISUALIZATION

```
In [24]: # plot the number of samples in 'subject'  
plt.figure(figsize = (8, 8))  
sns.countplot(y = "subject", data = df)
```

```
Out[24]: <matplotlib.axes._subplots.AxesSubplot at 0x1c1bfd31400>
```



```
] : # visualize the distribution of number of words in a text
import plotly.express as px
fig = px.histogram(x = [len(nltk.word_tokenize(x)) for x in df.clean_joined], nbins = 100)
fig.show()
```



# NLP

```
In [35]: # Sequential Model
model = Sequential()

# embeddindg layer
model.add(Embedding(total_words, output_dim = 128))
# model.add(Embedding(total_words, output_dim = 240))

# Bi-Directional RNN and LSTM
model.add(Bidirectional(LSTM(128)))

# Dense layers
model.add(Dense(128, activation = 'relu'))
model.add(Dense(1, activation= 'sigmoid'))
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['acc'])
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
embedding (Embedding)	(None, None, 128)	13914112
=====		
bidirectional (Bidirectional)	(None, 256)	263168
=====		
dense (Dense)	(None, 128)	32896
=====		
dense_1 (Dense)	(None, 1)	129
=====		
Total params: 14,210,305		
Trainable params: 14,210,305		
Non-trainable params: 0		
=====		

# MODEL FITTING

```
In [33]: # Add padding can either be maxlen = 4406 or smaller number maxlen = 40 seems to work well based on results
padded_train = pad_sequences(train_sequences,maxlen = 40, padding = 'post', truncating = 'post')
padded_test = pad_sequences(test_sequences,maxlen = 40, truncating = 'post')
```

```
In [34]: for i,doc in enumerate(padded_train[:2]):
        print("The padded encoding for document",i+1," is : ",doc)
```

```
The padded encoding for document 1 is : [ 1054    51    23  4707 26037  7072    510  7802    205     1   843  2546
 2297 10782    536  5887    814  6279    264  3443  7305  2538     27  8050
   278  7265 21147     44     20    16    12   903    24     7   183    23
   470   783    27  3532]
The padded encoding for document 2 is : [  452 11308     24  6502  2408  5389 10063 15930  2222  1014     27 10063
 15930   211   159   727 27293  5352  1105    10     1  1129    57   191
     7    19  1286    10     1  1589  2004   407  1129  1425  3249   313
   498   13   106    21]
```

```
In [40]: # if the predicted value is >0.5 it is real else it is fake
prediction = []
for i in range(len(pred)):
    if pred[i].item() > 0.5:
        prediction.append(1)
    else:
        prediction.append(0)
```

```
In [41]: # getting the accuracy
from sklearn.metrics import accuracy_score

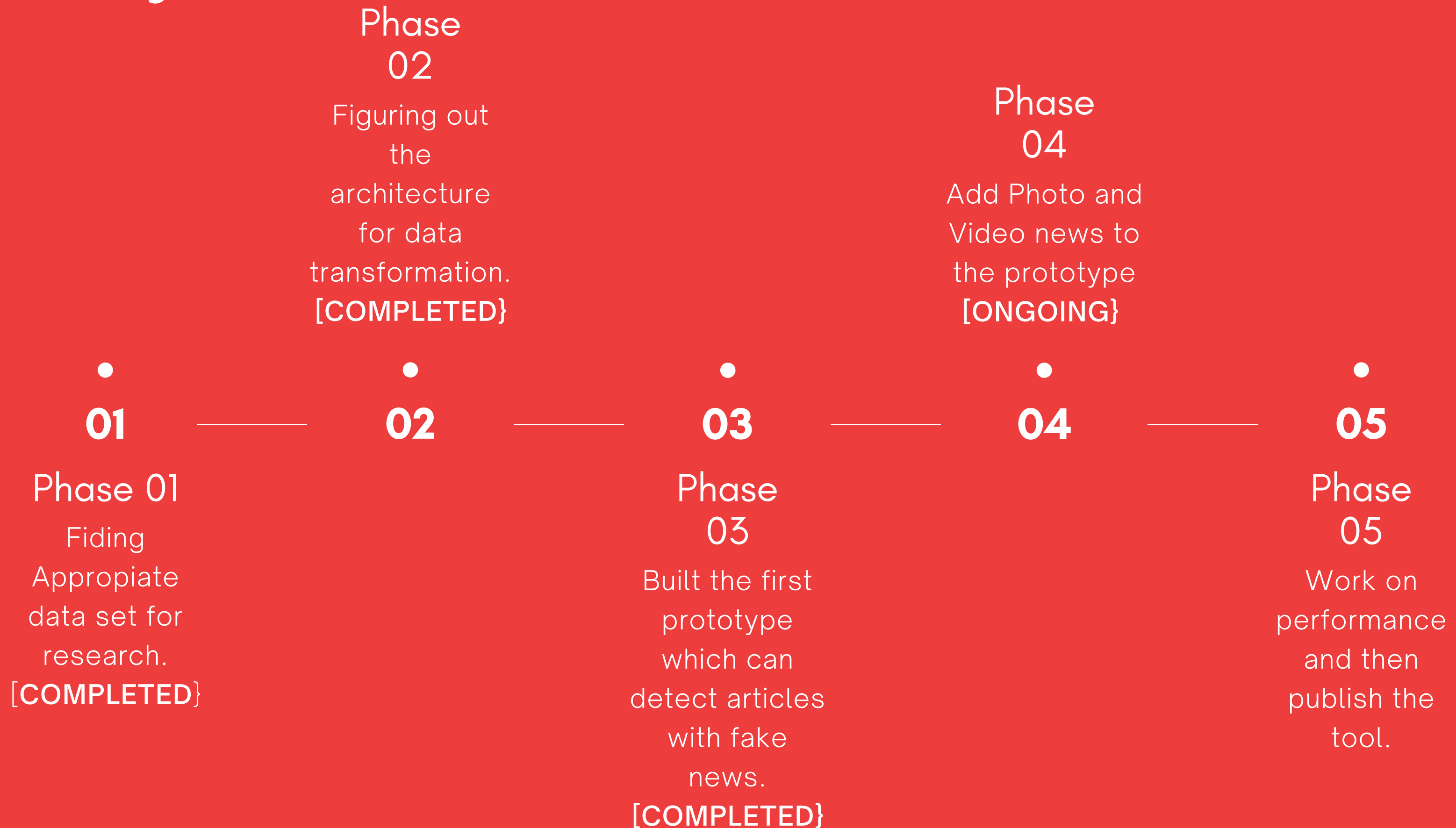
accuracy = accuracy_score(list(y_test), prediction)

print("Model Accuracy : ", accuracy)
```

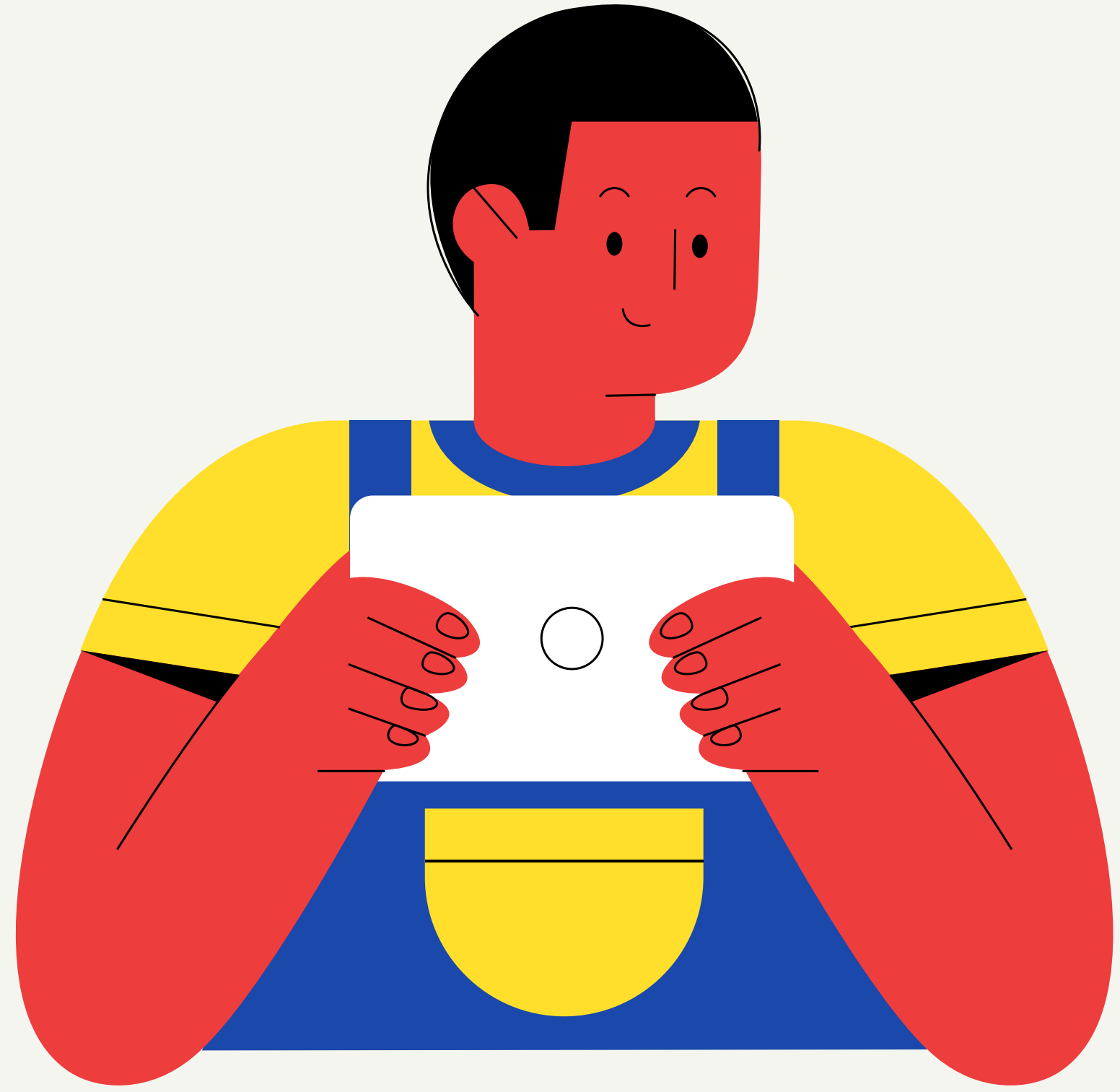
Model Accuracy : 0.9952115812917595



# Project Timeline



# Conclusion



So the fake news detection project was successfully moderated and completed. Phase 2 hopefully next sem would be the continuation of the same with Fake picture detection associated with fake news.

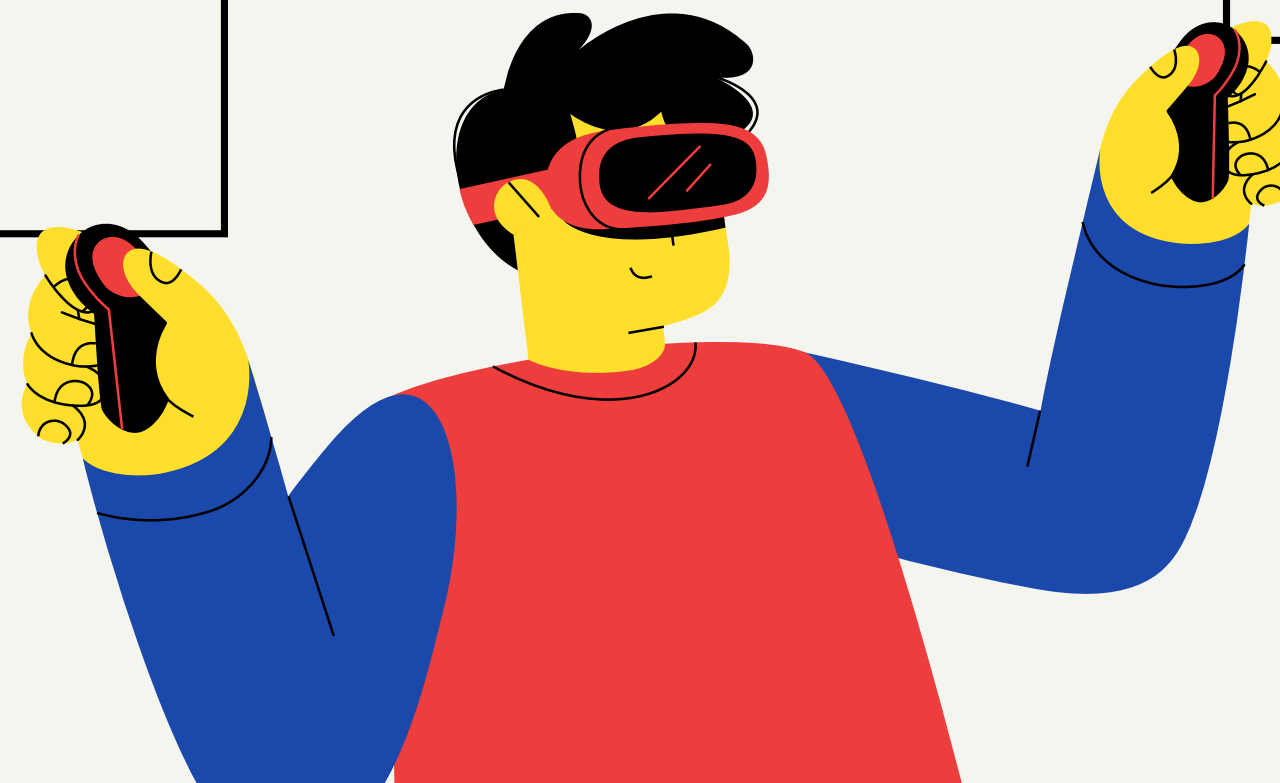
---

**Project highlights**

We would like to release this tool once image integration is done and would like to see how much efficiency it can show in real-life scenarios.

---

**Future Scope**



Thank You