Fake news Detection

Raghul Raj Manogeran

***Abstract*- Large amounts of false information are created online for a variety of reasons, including financial and political benefits, due to its ease of spread online. The widespread dissemination of false information can have a serious negative impact on people's lives and society as a whole by (i) upsetting the ecosystem's authenticity balance, (ii) purposefully influencing consumers to adopt biased or false beliefs, and (iii) altering how people perceive and react to actual news and information. This fake news detection exercise attempts to categorize text information regarding spreading rumors into four separate labels (FALSE, MISLEADING, TRUE, UNPROVEN). The claim and the articles must be preprocessed before being compared for classification.**

I. INTRODUCTION

Newspapers, tabloids, and magazines have given way to online news sources, blogs, social media feeds, and other digital media formats as the news medium changed. Consumers now have more access to the most recent news at their fingertips. In their current form, these social media platforms are very effective and helpful for enabling users to debate, share, and discuss topics like democracy, education, and health. However, some organizations also utilize these platforms negatively, frequently to obtain financial advantage, and occasionally to sway public opinion, influence people's attitudes, or propagate satire or absurdity.

What percentage of the news we read on social media and on purportedly "reliable" news websites can we trust? It is very simple for anyone to post whatever they want, and while that may be acceptable, there is the possibility of going too far. Examples of this include posting false information online to incite panic, telling lies to influence another person's decision, or essentially anything else that could have long-lasting effects. The amount of information available online makes it difficult to distinguish between true and untrue. Due to the aforementioned reasons, it is important to detect fake news.

II. RELATED WORKS

1) Fake News Detection Using Machine Learning Ensemble Methods -

This study investigates many textual characteristics that can be utilized to identify between true and false contents. These characteristics are used to train a variety of machine learning algorithms using various ensemble approaches, and four real-world datasets are used to assess their performance. The suggested ensemble learner strategy outperforms the individual learner approach, according to experimental evaluation.

2) Fake news detection based on news content and social contexts: a transformer-based approach -

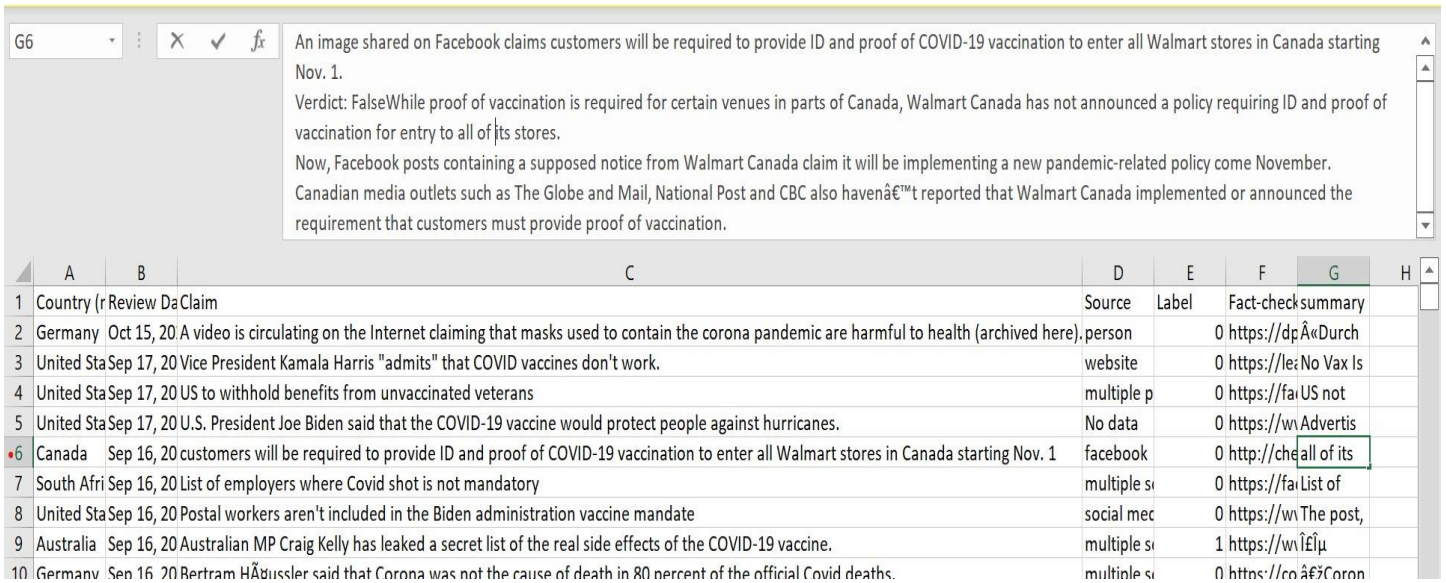
In order to identify false news, the suggested approach uses data from news articles and social situations. The model is built on a Transformer architecture, which consists of two parts: an encoder to extract meaningful representations from the fake news data and a decoder to forecast behavior based on historical data. To further aid in the classification of the news, a number of characteristics from the social contexts and news content in our model are used. In addition, a successful labeling method to solve the label shortage issue is proposed. The algorithm can detect fake news more accurately and quickly than baselines within a few minutes of it spreading (early detection), according to experimental results using real-world data.

III. MODEL DESCRIPTION

1) Data Scraping using Newspaper3k

Newspaper3k is a Python package for extracting content from the web. While parsing for lxml, it uses the requests library and depends on BeautifulSoup. Newspaper3k package web scrapes websites using advanced algorithms to extract all of the helpful text. On websites for online newspapers, it works incredibly well. Newspaper3k can scrape the complete article text for you as well as other types of information like the publish date, author(s), URL, photos, and video, to mention a few.

Image 1: Summary Generated for label False (0)

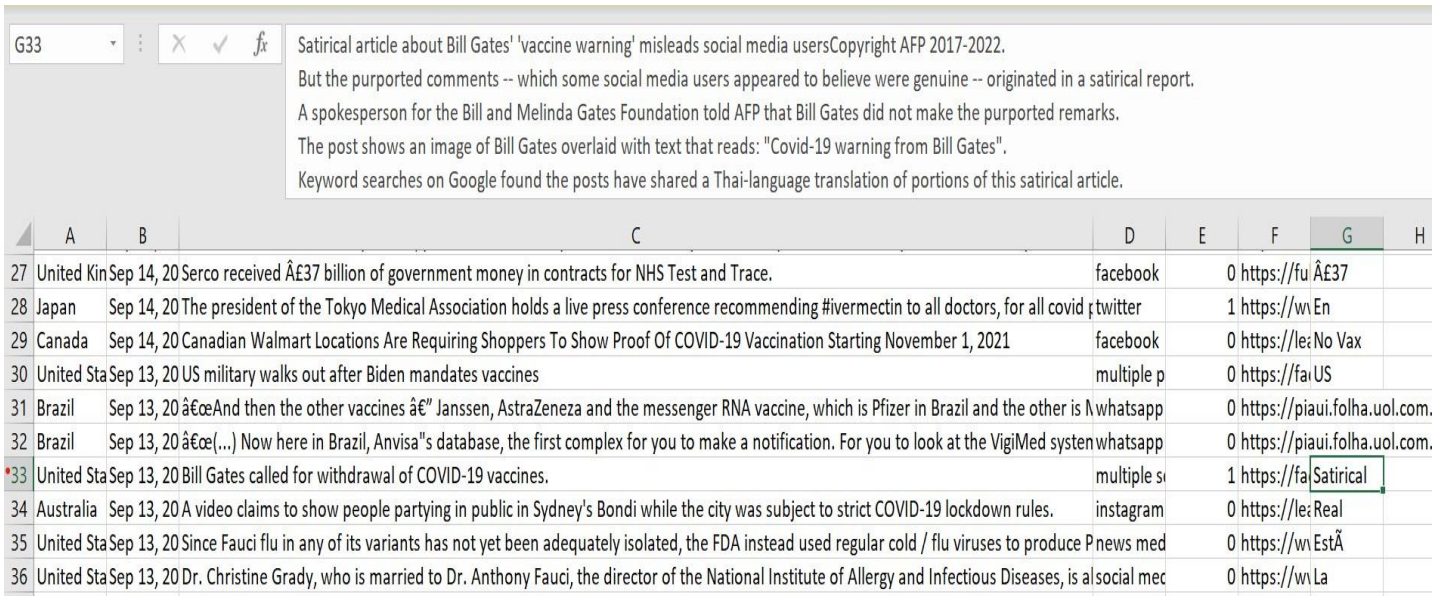


G6

An image shared on Facebook claims customers will be required to provide ID and proof of COVID-19 vaccination to enter all Walmart stores in Canada starting Nov. 1.
Verdict: FalseWhile proof of vaccination is required for certain venues in parts of Canada, Walmart Canada has not announced a policy requiring ID and proof of vaccination for entry to all of its stores.
Now, Facebook posts containing a supposed notice from Walmart Canada claim it will be implementing a new pandemic-related policy come November.
Canadian media outlets such as The Globe and Mail, National Post and CBC also havenâ€™t reported that Walmart Canada implemented or announced the requirement that customers must provide proof of vaccination.

	A	B	C	D	E	F	G	H
1	Country (r	Review Da	Claim	Source	Label	Fact-check	summary	
2	Germany	Oct 15, 20	A video is circulating on the Internet claiming that masks used to contain the corona pandemic are harmful to health (archived here).	person	0	https://dp	Â«Durch	
3	United Sta	Sep 17, 20	Vice President Kamala Harris "admits" that COVID vaccines don't work.	website	0	https://le	No Vax Is	
4	United Sta	Sep 17, 20	US to withhold benefits from unvaccinated veterans	multiple p	0	https://fa	US not	
5	United Sta	Sep 17, 20	U.S. President Joe Biden said that the COVID-19 vaccine would protect people against hurricanes.	No data	0	https://w	Advertis	
6	Canada	Sep 16, 20	customers will be required to provide ID and proof of COVID-19 vaccination to enter all Walmart stores in Canada starting Nov. 1	facebook	0	http://che	all of its	
7	South Afri	Sep 16, 20	List of employers where Covid shot is not mandatory	multiple s	0	https://fa	List of	
8	United Sta	Sep 16, 20	Postal workers aren't included in the Biden administration vaccine mandate	social mec	0	https://w	The post,	
9	Australia	Sep 16, 20	Australian MP Craig Kelly has leaked a secret list of the real side effects of the COVID-19 vaccine.	multiple s	1	https://w	Î£Îµ	
10	Germany	Sep 16, 20	Bertram HÃ¤ussler said that Corona was not the cause of death in 80 percent of the official Covid deaths.	multiple s	0	https://co	â€žCoron	

Image 2: Summary Generated for label Misleading (1)

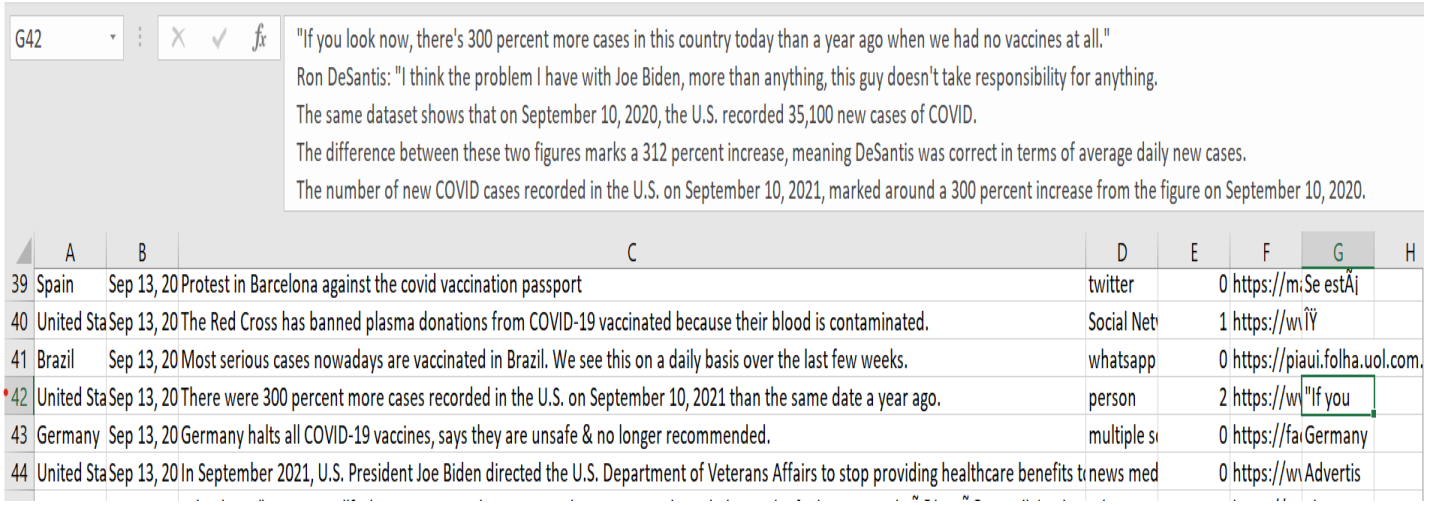


G33

Satirical article about Bill Gates' 'vaccine warning' misleads social media usersCopyright AFP 2017-2022.
But the purported comments -- which some social media users appeared to believe were genuine -- originated in a satirical report.
A spokesperson for the Bill and Melinda Gates Foundation told AFP that Bill Gates did not make the purported remarks.
The post shows an image of Bill Gates overlaid with text that reads: "Covid-19 warning from Bill Gates".
Keyword searches on Google found the posts have shared a Thai-language translation of portions of this satirical article.

	A	B	C	D	E	F	G	H
27	United Kin	Sep 14, 20	Serco received Â£37 billion of government money in contracts for NHS Test and Trace.	facebook	0	https://fu	Â£37	
28	Japan	Sep 14, 20	The president of the Tokyo Medical Association holds a live press conference recommending #ivermectin to all doctors, for all covid p	twitter	1	https://w	En	
29	Canada	Sep 14, 20	Canadian Walmart Locations Are Requiring Shoppers To Show Proof Of COVID-19 Vaccination Starting November 1, 2021	facebook	0	https://le	No Vax	
30	United Sta	Sep 13, 20	US military walks out after Biden mandates vaccines	multiple p	0	https://fa	US	
31	Brazil	Sep 13, 20	â€œAnd then the other vaccines â€" Janssen, AstraZeneza and the messenger RNA vaccine, which is Pfizer in Brazil and the other is N	whatsapp	0	https://piaui.folha.uol.com.		
32	Brazil	Sep 13, 20	â€œ(...) Now here in Brazil, Anvisa"s database, the first complex for you to make a notification. For you to look at the VigiMed systen	whatsapp	0	https://piaui.folha.uol.com.		
33	United Sta	Sep 13, 20	Bill Gates called for withdrawal of COVID-19 vaccines.	multiple s	1	https://fa	Satirical	
34	Australia	Sep 13, 20	A video claims to show people partying in public in Sydney's Bondi while the city was subject to strict COVID-19 lockdown rules.	instagram	0	https://le	Real	
35	United Sta	Sep 13, 20	Since Fauci flu in any of its variants has not yet been adequately isolated, the FDA instead used regular cold / flu viruses to produce P	news med	0	https://w	EstÃ	
36	United Sta	Sep 13, 20	Dr. Christine Grady, who is married to Dr. Anthony Fauci, the director of the National Institute of Allergy and Infectious Diseases, is al	social mec	0	https://w	La	

Image 3: Summary Generated for label True (2)

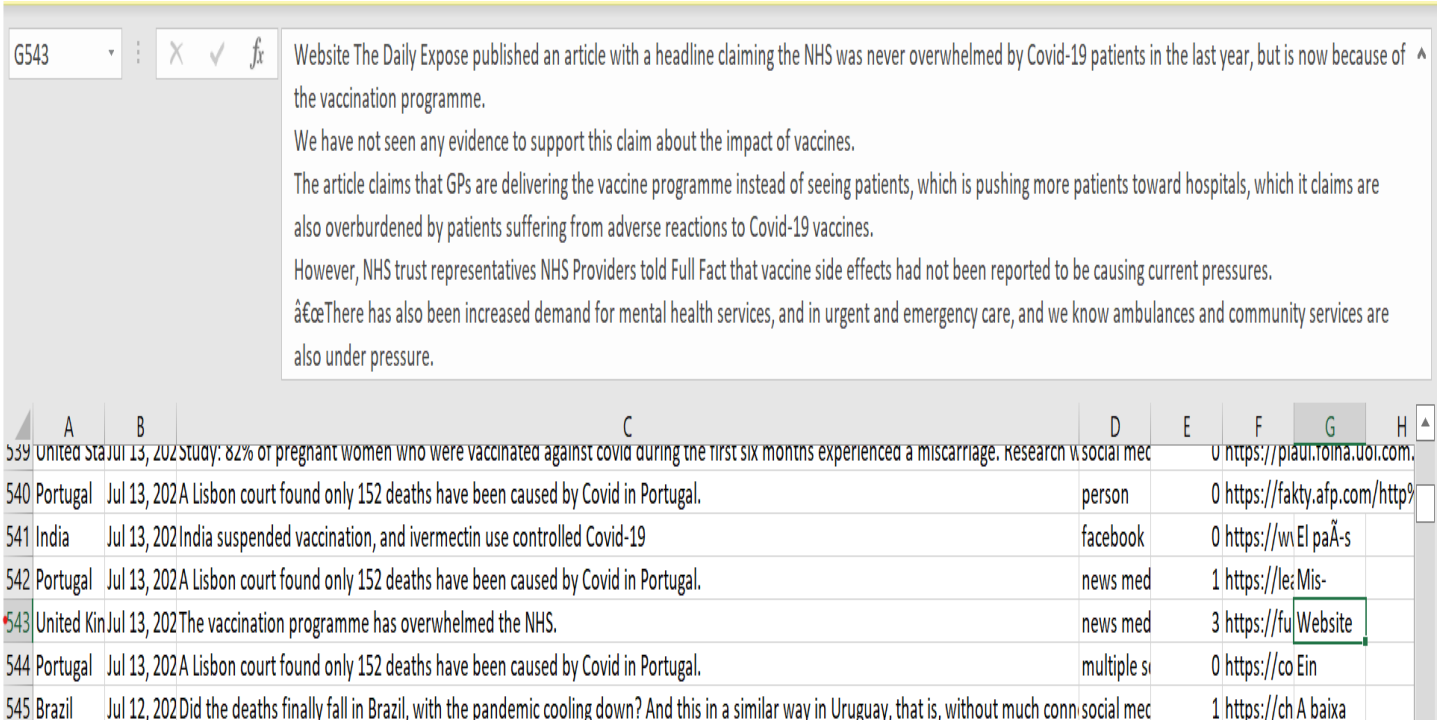


G42

"If you look now, there's 300 percent more cases in this country today than a year ago when we had no vaccines at all."
Ron DeSantis: "I think the problem I have with Joe Biden, more than anything, this guy doesn't take responsibility for anything.
The same dataset shows that on September 10, 2020, the U.S. recorded 35,100 new cases of COVID.
The difference between these two figures marks a 312 percent increase, meaning DeSantis was correct in terms of average daily new cases.
The number of new COVID cases recorded in the U.S. on September 10, 2021, marked around a 300 percent increase from the figure on September 10, 2020.

	A	B	C	D	E	F	G	H
39	Spain	Sep 13, 20	Protest in Barcelona against the covid vaccination passport	twitter	0	https://m	Se estÃ¡	
40	United Sta	Sep 13, 20	The Red Cross has banned plasma donations from COVID-19 vaccinated because their blood is contaminated.	Social Net	1	https://w	ÎŸ	
41	Brazil	Sep 13, 20	Most serious cases nowadays are vaccinated in Brazil. We see this on a daily basis over the last few weeks.	whatsapp	0	https://piaui.folha.uol.com.		
42	United Sta	Sep 13, 20	There were 300 percent more cases recorded in the U.S. on September 10, 2021 than the same date a year ago.	person	2	https://w	"If you	
43	Germany	Sep 13, 20	Germany halts all COVID-19 vaccines, says they are unsafe & no longer recommended.	multiple s	0	https://fa	Germany	
44	United Sta	Sep 13, 20	In September 2021, U.S. President Joe Biden directed the U.S. Department of Veterans Affairs to stop providing healthcare benefits t	news med	0	https://w	Advertis	

Image 4: Summary Generated for label Unproven (3)



Newspaper3k also generates a summary of the article that gives us the main points without reading the complete thing. After the data has been extracted, it can be merged and saved in a variety of forms, including CSV, JSON, and even pandas. It is available in almost 30 languages in Newspaper3k.

We have primarily used the Newspaper3k package to generate a summary of the news article given as URLs in the training dataset. We added a column named “summary” and stored the output(concise short paragraph) of the entire news article in the respective rows.

The examples of summaries generated for each of the unique labels are as follows:

Summary generated when the label is 0 - In the Image1 we can see the summary for the Row Number 6 along with the claim that is given. Since it is not true, the label is 0

Summary generated when the label is 1 - In the Image2 we can see the summary for the Row Number 33 along with the claim that is given. The label here is 1 as the information is misleading.

Summary generated when the label is 2 - In the Image3 we can see the summary for the Row Number 42 along with the claim that is given. The label here is 2 and we can see that the summary supports the claim and thus the label is True.

Summary generated when the label is 3 - In the Image4 we can see the summary for the Row Number 543 along with the claim that is given. We can clearly see in the summary that there is no evidence to support the claim and thus the label is 3 that is unproven.

2) **Data Preprocessing**

Mapping of Source names

a) The source column had many redundant values which were wrongly spelled. We have corrected them by mapping those values to their respective source. For example- the misspelled sources were “perseon”, “facbook”, etc. These values were correctly mapped under as person and social media.

b) The source column had a lot of entries which were names of different persons. All those sources were mapped to the source as- person. Example- Names like anthony fauci, bernie sanders, beverley turner, boris johnson, etc were all mapped under the source person.

c) All the social media websites like youtube, facebook, instagram, twitter, tiktok etc were all mapped to the source - social media.

d) Values like unknown, video, photo, study, meme etc were all mapped to a source as Others.

```
data['Source'].value_counts()
```

```
social media          3675
multiple sources      1089
person                 930
others                 266
news media             266
websites               158
Name: Source, dtype: int64
```

Mapping of Country names-

Considering only Top 10 countries which account for 65% of the data and binning the others countries together as "Others" -

```
#Top 10 contributing to 65 percent of the data..So using them
data['Country (mentioned)'].value_counts()[:9]*100/len(data)
```

```
United States     21.788847
India             14.379699
Brazil            10.197368
United Kingdom     6.876566
Italy              3.947368
China              3.493108
France             2.662907
Spain              2.474937
Indonesia          2.427945
Name: Country (mentioned), dtype: float64
```

3) **One-hot encoding of Countries and Sources columns**

One hot encoding is a method that involves transforming categorical information into a format that is given to the Machine Learning algorithms to help them perform better at prediction.

Data that don't relate to one another can benefit from one hot encoding. The arrangement of numbers is treated as a significant characteristic by machine learning algorithms. This technique of encoding makes our training data more useful and expressive, and it can be rescaled easily. By using numeric values, we more easily determine a probability for our values. In particular, one hot encoding is used for our output values, since it provides more nuanced predictions than single labels. We have used the sklearn library in python to perform one-hot encoding.

4) Train-Test Split

Perform train-test split of our data- The train_test_split() method is used to split the data into train and test sets. We must first separate our data into features (X) and labels (y). The dataframe is split into the X train, X test, Y train, and Y test sections. The model is trained and fitted using the X train and y train sets. The model is tested to see if it correctly predicts the outputs and labels using the X test and y test sets.

The size of the train and test sets can be explicitly tested. In our split, the training data is 80% and the remaining 20% is testing data.

5) Similarity Calculation

a) spaCy's Similarity score:

After doing the splitting, we use en_core_web_sm - spaCy's similarity calculator.spaCy is a free and open-source library for Natural Language Processing (NLP) in Python, We are able to calculate the similarity between sentences using this and we can generate a score as well. The value ranges from 0 to 1, with 1 meaning both sentences are the same and 0 showing no similarity between both sentences

b) Cross Encoder Semantic Similarity Score:

Since cosine similarity will not be able to capture the semantics or context of the sentences, we've also used a Cross-Encoder approach. Here, we pass both sentences simultaneously to the Transformer network. It gives an output value between 0 and 1 indicating the similarity of the input sentence pair. Cross-Encoders can be used whenever you have a predefined set of sentence pairs you want to score. For example, you have 100 sentence pairs and you want to get similarity scores for these 100 pairs.

6) DecisionTreeClassifier

A decision tree is a flowchart-like tree structure in which an internal node represents an attribute or feature, the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value.

We are using the following features one-hot encoded values of country column and source columns along with the similarity score. Next, we try to predict the labels for the test split from the train data by applying our trained model.

We got an accuracy of 77.36% for the test split using the spaCy's similarity score as feature

```
from sklearn import metrics
metrics.accuracy_score(y_test,dtree_predictions)
```

```
0.7736883320281911
```

We got an accuracy of 77.05% for the test split using the Cross encoder semantic similarity score as feature

```
from sklearn import metrics
metrics.accuracy_score(y_test,dtree_predictions)
```

```
0.7705559906029757
```

7) Making Predictions

Similarly, we repeat the same steps for the testing data. Repeating the same preprocessing steps for test data - Summary generation, country and source binning and calculating the similarity scores. We're then using the trained Decision Tree Classifier model to predict for the test data points.

We got a test accuracy of 86.19% using the Semantic similarity score as a feature and 85.63 using the spaCy's similarity score as a feature.

IV. EXPERIMENT

We tried out quite a few approaches to perform fake news detection before finalizing on our approach. Here is the list of approaches we tried.

1) Tf-IDF approach: We scraped the URLs provided in the Fact-checked Article column and performed a Tf-IDF vectorization to identify the critical keywords and extracted relevant sentences containing these keywords to form a summary and calculated a cosine similarity score between the claim and the extracted summary. Used the score along with Country (mentioned), Source as features for a multinomial logistic regression model. The cosine similarity scores could not capture the semantic similarity between the sentences and the Tf-IDF way of generating summary was not sufficient to decide on the labels.
2) BERT Question Answering Approach: We used the fine-tuned BERT model from the Hugging Face Transformers library to answer questions. It was trained using the CoQA dataset (Conversational Question Answering dataset). The model reads the user-provided text context and attempts to respond to any questions posed by that text context. We framed the problem in the following way: The scraped data being the user-provided text and questions being the rephrased versions of the claim asking if it is fake,true,misleading or unproven. The answer would be the label indicating the category of the news.
3) BERT Text Entailment Approach: We used the fine-tuned BERT model from the Hugging Face Transformers library to answer questions. It was trained using the MultiNLI dataset (Multi-Genre Natural Language Inference dataset). If a premise is true, then there is entailment. Simply put, a sentence Y is said to entail a sentence X if X is true and Y can be deduced logically from it. A pair of sentences in the dataset we used can either entail each other, be neutral, or contradict each other. Here, the premise is the generated summary (which is always true) while the conclusion being the claim (which can entail, contradict, or be neutral to the premise).

V. FUTURE WORKS

1. Translation: Some of the webpages had non-English content in them. We performed similarity checks between claims that were in English versus the summary that was generated in other languages. So, the scores were not accurate. So, In the future, we'd include language translation of the summary into English to get accurate measures of similarities.
2. Text preprocessing: Preprocessing of the scraped content would be something we might want to take up in the future for trying out other approaches. Examples being tokenization, stemming or lemmatization, removing stop words and punctuations, etc. Since, our current models don't require us to do that for similarity score calculations, we didn't do that for the final version.
3. Cross Validation: Different cross validation techniques like Holdout Method, K-Fold Cross-Validation can be used to identify the best possible parameters for the Decision Tree Classifier model.

REFERENCES

[1] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, Muhammad Ovais Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods", Complexity, vol. 2020, Article ID 8885861, 11 pages, 2020. https://doi.org/10.1155/2020/8885861

[2] Raza, S., Ding, C. Fake news detection based on news content and social contexts: a transformer-based approach. Int J Data Sci Anal 13, 335–362 (2022). https://doi.org/10.1007/s41060-021-00302-z

[3] https://www.sbert.net/index.html

[4] https://www.sbert.net/docs/pretrained_models.html

[5] https://www.whatismybrowser.com/detect/what-is-my-user-agent/

[6]https://www.aitude.com/multiclass-classification-on-highly-imbalanced-dataset/

[7] https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/#:~:text=We%20can%20set%20the%20maximum,get%20a%20very%20bad%20accuracy.

[8] https://www.geeksforgeeks.org/multiclass-classification-using-scikit-learn/

[9] https://medium.com/analytics-vidhya/question-answering-system-with-bert-ebe1130f8def

[10] https://towardsdatascience.com/fine-tuning-pre-trained-transformer-models-for-sentence-entailment-d87caf9ec9db
[11] https://towardsdatascience.com/question-answering-with-a-fine-tuned-bert-bc4dafd45626
[12] https://www.geeksforgeeks.org/python-measure-similarity-between-two-sentences-using-cosine-similarity/#:~:text=Python%20%7C%20Measure%20similarity%20between%20two%20sentences%20using%20cosine%20similarity,-Improve%20Article&text=Cosine%20similarity%20is%20a%20measure,A%20and%20B%20are%20vectors.
[13] https://stackoverflow.com/questions/53453559/similarity-in-spacy
[14] https://github.com/dh1105/Sentence-Entailment
[15]https://www.geeksforgeeks.org/newspaper-article-scraping-curation-python/
[16]https://www.cfoselections.com/perspective/when-to-use-a-decision-tree-for-business-planning