Stance Detection in Fake News Challenge (Team 27)

- Meet Maheshwari
- Priya Mehta
- Sriza Sen
- Shaney Waris

Dataset Links:

- Train Dataset Link:
  - https://drive.google.com/drive/folders/1Rs1mJqJzRxAUrCN4JUkh16YaBXT4n4j-?usp=sharing

- Test Dataset Link:
  - https://drive.google.com/drive/folders/1Rs1mJqJzRxAUrCN4JUkh16YaBXT4n4j-?usp=sharing

Code Files Structure:

Link: https://drive.google.com/drive/folders/1Rs1mJqJzRxAUrCN4JUkh16YaBXT4n4j-?usp=sharing
Our project contains only two code files,

- *FS_Creation.ipynb*
  - This file creates the following features:
    - Negated words.
    - Avg per sentence sentiment (+, -, 0).
    - Number of URLs present in a tweet.
    - Lexicon sentiment of hashtags.
    - length of a tweet without stopwords.
    - No of sentences in a tweet.
    - Average sentence length.
    - Number of positive words.
    - Number of neg words.
    - Number of total hashtags in a tweet.
    - The total number of user tags present in a tweet.
    - The total number of emoji's present in a tweet.
    - Number sentiment avg.
    - Language-Agnostic BERT Sentence Embedding.
    - Universal Sentence Encoder.

- *Model Application.ipynb*
  - Create various models and compare their results.
- *sh_fs.py*
  - Features created by Shaney
- *sr_fs.py*
  - Features created by Sriza

- *stance_detect.py*
    - Module to import to find the stance of any given headline and body with 4 models and 4 different feature sets.
- *app.py*
    - Flask webapp to detect stance for any article given Headline and Body with 4 models and 4 different feature sets.
- *templates/*.html* and *static/**
    - HTML and other supporting files for Flask app.

How to run the code :

Open Terminal or Command Prompt
Run the following command

- set FLASK_APP=<path/to/app.py>
- flask run

Trained Models Link:

- https://drive.google.com/drive/folders/1Rs1mJqJzRxAUrCN4JUkh16YaBXT4n4j-?usp=sharing

Output:

We have used various models and then compared their results.

MLP with only BERT feature:

	precision	recall	f1-score	support
0.0	0.16	0.31	0.21	295
1.0	0.05	0.08	0.06	88
2.0	0.34	0.47	0.39	701
3.0	0.93	0.83	0.87	5849
accuracy			0.76	6933
macro avg	0.37	0.42	0.38	6933
weighted avg	0.82	0.76	0.79	6933

MLP with (BERT feature + Other features):

	precision	recall	f1-score	support
0.0	0.16	0.30	0.21	295
1.0	0.33	0.10	0.16	88
2.0	0.35	0.47	0.40	701
3.0	0.92	0.85	0.88	5849
accuracy			0.78	6933
macro avg	0.44	0.43	0.41	6933
weighted avg	0.82	0.78	0.80	6933

MLP with Universal Encoding

	precision	recall	f1-score	support
0.0	0.16	0.40	0.23	295
1.0	0.02	0.03	0.03	88
2.0	0.29	0.39	0.33	701
3.0	0.92	0.80	0.86	5849
accuracy			0.73	6933
macro avg	0.35	0.41	0.36	6933
weighted avg	0.81	0.73	0.77	6933

MLP with Universal Encoding + Other features

	precision	recall	f1-score	support
0.0	0.11	0.35	0.17	295
1.0	0.06	0.02	0.03	88
2.0	0.27	0.57	0.36	701
3.0	0.93	0.72	0.81	5849
accuracy			0.68	6933
macro avg	0.34	0.41	0.34	6933
weighted avg	0.82	0.68	0.73	6933

Random Forest with only BERT Feature:

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.47	0.09	0.15	701
3.0	0.85	0.99	0.91	5849
accuracy			0.84	6933
macro avg	0.33	0.27	0.27	6933
weighted avg	0.76	0.84	0.79	6933

Random Forest with (BERT feature + Other Feature):

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.50	0.08	0.14	701
3.0	0.85	0.99	0.92	5849
accuracy			0.84	6933
macro avg	0.34	0.27	0.26	6933
weighted avg	0.77	0.84	0.79	6933

Random Forest with Universal Encoding:

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.76	0.03	0.06	701
3.0	0.85	1.00	0.92	5849
accuracy			0.85	6933
macro avg	0.40	0.26	0.24	6933
weighted avg	0.79	0.85	0.78	6933

Random Forest with Universal Encoding+Other features:

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.71	0.02	0.03	701
3.0	0.84	1.00	0.92	5849
accuracy			0.84	6933
macro avg	0.39	0.25	0.24	6933
weighted avg	0.78	0.84	0.78	6933

SVC with only BERT Feature:

	precision	recall	f1-score	support
0.0	0.16	0.15	0.15	295
1.0	0.00	0.00	0.00	88
2.0	0.48	0.38	0.42	701
3.0	0.89	0.93	0.91	5849
accuracy			0.83	6933
macro avg	0.38	0.36	0.37	6933
weighted avg	0.81	0.83	0.82	6933

SVC with (BERT feature + Other features):

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.02	0.00	0.00	701
3.0	0.84	0.99	0.91	5849
accuracy			0.84	6933
macro avg	0.22	0.25	0.23	6933
weighted avg	0.71	0.84	0.77	6933

SVC with Universal Encoding:

	precision	recall	f1-score	support
0.0	0.23	0.25	0.24	295
1.0	0.00	0.00	0.00	88
2.0	0.49	0.38	0.43	701
3.0	0.90	0.94	0.92	5849
accuracy			0.84	6933
macro avg	0.41	0.39	0.40	6933
weighted avg	0.82	0.84	0.83	6933

SVC with Universal Encoding + Other features

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	295
1.0	0.00	0.00	0.00	88
2.0	0.02	0.00	0.00	701
3.0	0.84	0.99	0.91	5849
accuracy			0.84	6933
macro avg	0.22	0.25	0.23	6933
weighted avg	0.71	0.84	0.77	6933

XGBoost with only BERT feature:

	precision	recall	f1-score	support
0.0	0.20	0.08	0.11	295
1.0	0.00	0.00	0.00	88
2.0	0.42	0.23	0.30	701
3.0	0.87	0.95	0.91	5849
accuracy			0.83	6933
macro avg	0.37	0.32	0.33	6933
weighted avg	0.78	0.83	0.80	6933

XGBoost with (BERT feature + Other features):

	precision	recall	f1-score	support
0.0	0.16	0.30	0.21	295
1.0	0.33	0.10	0.16	88
2.0	0.35	0.47	0.40	701
3.0	0.92	0.85	0.88	5849
accuracy			0.78	6933
macro avg	0.44	0.43	0.41	6933
weighted avg	0.82	0.78	0.80	6933

XGBoost with Universal encoding:

	precision	recall	f1-score	support
0.0	0.26	0.06	0.10	295
1.0	0.00	0.00	0.00	88
2.0	0.57	0.20	0.29	701
3.0	0.87	0.98	0.92	5849
accuracy			0.85	6933
macro avg	0.43	0.31	0.33	6933
weighted avg	0.80	0.85	0.81	6933

XGBoost with Universal encoding + Other features

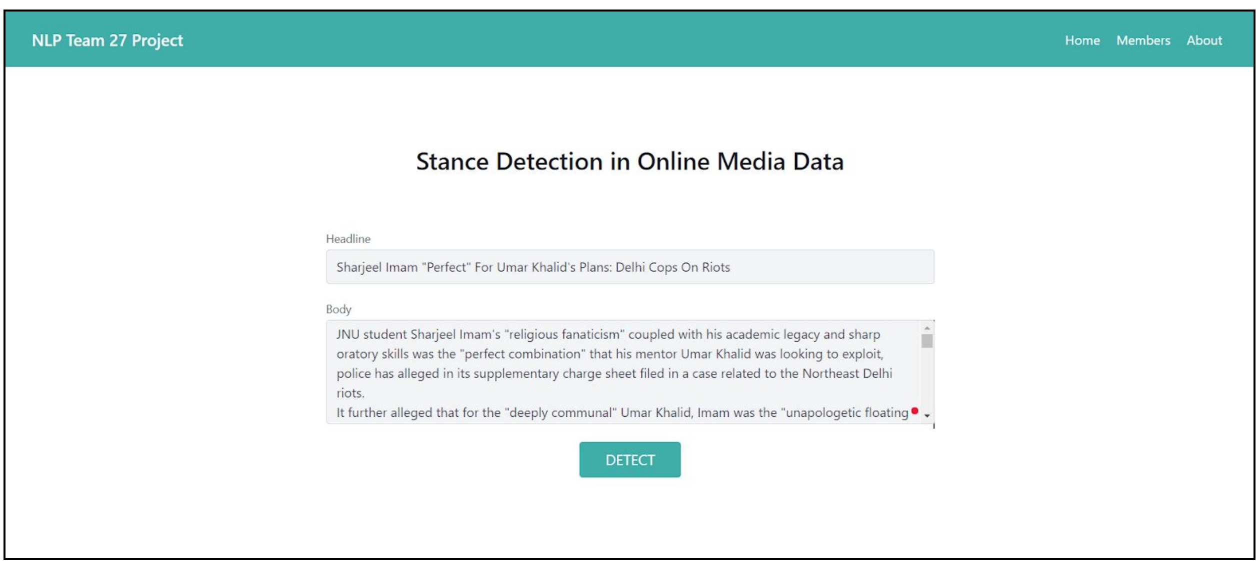
	precision	recall	f1-score	support
0.0	0.22	0.06	0.10	295
1.0	0.00	0.00	0.00	88
2.0	0.63	0.22	0.32	701
3.0	0.87	0.98	0.92	5849
accuracy			0.85	6933
macro avg	0.43	0.32	0.34	6933
weighted avg	0.81	0.85	0.81	6933

Note: Here, “Other Features” are the features other than BERT and Universal sentence encoder.

Accuracy Comparison For All models(%)

Feature Set Model	LaBSE	LaBSE+	Universal Encoder	Universal Encoder+
Random Forest	**84**	**84**	**85**	**84**
XGBoost	**83**	**78**	**85**	**85**
MLP	**76**	**78**	**73**	**68**
SVC	**83**	**84**	**84**	**84**

Screenshot of our Webapp:



Results we got:

Results

We have trained 4 models with 4 different set of features and then comapred their results.

Model	Features	Stance
Random Forest	LaBSE	unrelated
Random Forest	LaBSE + Other features	unrelated
Random Forest	Universal Encoder	unrelated
Random Forest	Universal Encoder + Other Features	unrelated
XGBoost	LaBSE	unrelated
XGBoost	LaBSE + Other features	unrelated
XGBoost	Universal Encoder	unrelated
XGBoost	Universal Encoder + Other Features	unrelated
MLP	LaBSE	discuss
MLP	LaBSE + Other features	unrelated
MLP	Universal Encoder	discuss
MLP	Universal Encoder + Other Features	unrelated
SVM	LaBSE	unrelated
SVM	LaBSE + Other features	unrelated
SVM	Universal Encoder	unrelated
SVM	Universal Encoder + Other Features	unrelated

Note: Here Others features are the features other than LaBSE and Universal Encoder. You can refer our report for the same.

End of ReadMe