1- Preprocessing:
   Remove Arabic stop words, remove hyperlinks, remove stock market tickers like $GE, remove English and removing punctuation
   Stemming for Arabic using ISRIStemmer.
2- Feature extraction and model:
   - Tf_idf into logistic regression:
     Tf_idf is implemented from scratch tf = log10(count(token)+1) where count(token) is number of times token occurred in tweet and idf = log10(N/df_t) N is number of documents and df_t Is number of documents in which term t occurs.
     Logistic regression using sklearn library.

     For stance:

(1000,)

	precision	recall	f1-score	support
-1	0.55	0.16	0.24	70
0	0.40	0.32	0.36	126
1	0.86	0.94	0.90	804
accuracy			0.81	1000
macro avg	0.60	0.47	0.50	1000
weighted avg	0.78	0.81	0.78	1000

     For category:

(1000,)

	precision	recall	f1-score	support
advice	0.00	0.00	0.00	10
celebrity	0.86	0.79	0.82	145
info_news	0.69	0.87	0.77	545
others	0.25	0.06	0.10	17
personal	0.52	0.51	0.52	128
plan	0.24	0.09	0.13	82
requests	0.33	0.05	0.09	20
restrictions	0.00	0.00	0.00	2
rumors	0.00	0.00	0.00	15
unrelated	0.56	0.25	0.35	36
accuracy			0.67	1000
macro avg	0.35	0.26	0.28	1000
weighted avg	0.62	0.67	0.63	1000

   - Embedding into RNN
     Embedding and RNN using Pytorch library with paramters (batch size:512, layers_num:1, n_echos:5,embedding dim: 50)

Accuracy for stance is: 0.804

Accuracy for category is: 0.545

3- Use first approach (tf_idf into lr) to test because it gives better accuracy