***CA22 Assignment 3 By group LetsArgue***

**Requirement**: ‘glove/glove.6B.100d.txt’ file which we have included in zip folder under glove folder.

1. **get\_sentimentdiff**(x) calculates the vector of polarity difference between consecutive claims as there are high chances of confirmation bias present if adjacent claim sentiment changes largely.

2.Used Vader to get sentiment score (from nltk.sentiment.vader import SentimentIntensityAnalyzer)

3. Used word2vec from spacy ('en\_core\_web\_lg').

4. Calculated ['no\_of\_major\_claims','no\_of\_claims',’ no\_of\_premises’] to check if number of major claims or claims or premises affect confirmation bias or not. Calculated [‘nooftokens\_majorclaim’,’ nooftokens\_claims’,’ nooftokens\_premises’] to check if number of tokens in major claims or claims or premises affect confirmation bias or not.

(**Thought**: More the number of major\_claims, claims, premises, more the chances that author might change or switch to opposite side of his initial stance (to brief other side of argument but end up having confirmation bias)).

5. ['confirmation\_bias'] converted as 0/1 for quicker results.

6. Used Glove vec ‘glove/glove.6B.100d.txt’ of 100 dimentions from 6Billion word vocabulary.

7. used TFIDF on text with max\_features = 1000, min\_df = 0.02, norm='l1', ngram\_range=(1,2), analyzer='word'.

8. passed [‘text\_sentiment\_scr’,’vec’,’glove\_vec’,’ claims\_padded\_polaritydiff\_vec’,’text’] as main features. Calculated corelation matrix between features & tried adding them as well but accuracy remained less than 70% so didn’t used as **main features**.

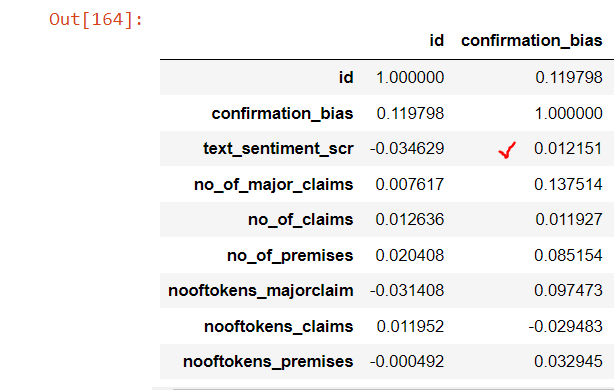
9. Used SVC & RandomForestClassifier(66%) but finalized SVC as was giving better accuracy(71%).

10. Done Hyperparameter tuning using GridSearch for both SVC & RandomForest.

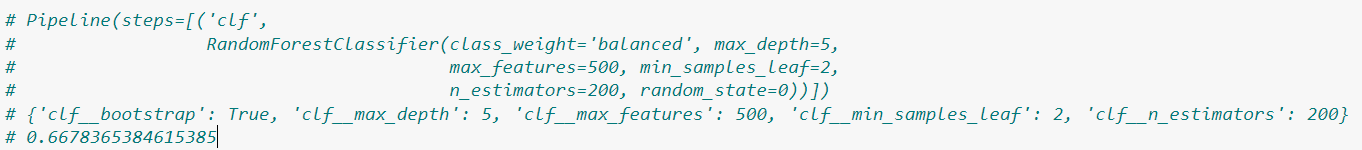
11. Tried creating ANN model as well but it was giving train accuracy to (92%) but not test accuracy each time(was getting 69% most of the time so removed in final code but pasted screen shot below).

Corelation matrix between features: There was a strong corelation between confirmation bias & no\_of\_major\_cliams(see pic below) & nooftoekns\_majorclaim. Tried using them as features but was giving accuracy (approx 64%) so removed in final features.

# Corelation matrix between features & confirmation\_bias



# Random Forest giving 66% accuracy with best parameters as below.(commented In final code)



# Trained ANN model also(with main features) but removed from final code as test accuracy was 69%.

