Approach to solve Fake news classification

* Read training file and check for null values. After checking for null in each text, title and author find out their null count
* Put available data in text column where title is null and available data in title where text is null. In this way we can still have useful info in those cols.
* Put author as Unknown where author is null rather than dropping those rows to have more information.
* Convert all data into lowercase and then remove English stop words otherwise stop words will have highest percentage in text and which won’t be useful
* Stem the text to group words with similar context thus helping in reduction of vocab
* Put author and title as total which will be used as input for classifier
* Check for words which are repeating most via word cloud to get understanding of vocab
* Use glove embedding to vectorize vocab. The reason to use glove over one-hot or tfidf is that by using glove we can preserve semantic meaning of words and have better contextual representation
* Choosing glove over word2vec because word2vec will take much time for training and for this problem that much deep training is not required.
* Use LSTM as layer in model architecture over GRU and RNN for better understanding of text.
* Single direction LSTM is good because in news data fake or real can be identified by reading from one end only.
* Train for 10 epochs to get final model and get evaluation metrics.