Approach :

I saw that every sentence structure might be unique to particular author and used the author name as part of sentence. Once I clean the data by removing html tags and removing punctuations by some words and fine cleaning it , I observed the data was imbalanced. There were two minority classes and one majority class. I tried to bring the minority samples Up by using SVM SMOTE technique.

I tried to calculate the probablities group by author names for each text which I thought might help me in modelling. At last I used TF - IDF vectorizer and LIGHT GBM to train my model. I fine tuned the model using gridsearch CV and found the best parameters. Though I was getting much better score in LSTM and Bi-LSTM but due memory constraint I had to go with LIght GBM model. At last I was able to achieve max of 45 score because I feel that beyond that score it was overfitting.