Author classification project

Using NLP and techniques of supervised learning (including Deep Learning) and unsupervised learning (emphasizing on unsupervised for this project), and collect thousand texts from Gutenberg project (and 7 novels) for at least 10 authors, build a project to classify text-author. The project should follow the guideline as:

1. Pre-process data using Spacy and other methods.
2. Perform data exploration
3. Using Bag of Word, apply supervised models such as Naive Bayes, Logistic Regression, Decision Tree, Random Forest, KNN, SVM and Gradient Boosting, including GridSearchCV.
4. Similar to 3., but using TF-IDF.
5. Similar to 3., but using word2vec.
6. Appy RNN to do classification.
7. Using unsupervised technique, visualize bar graphs for clusters containing 10 author documents. Adjust by silhouette scores.
8. Using LSA, LDA and NNMF, print out top ten words (with their highest loading) for each topic modeling. Analyze and compare among three methods.
9. Write up analysis and conclusions.