

INTA 6450 Final Exam

David Muchlinski

Instructions

Using the Russian IRA Tweet Datasets posted to Canvas, do the following:

- Develop and interpret a LDA Topic Model showing the distribution of topics over all account categories in the data. Discuss, based on the results of this topic model, the subtle ways in which the Russian Intelligence institutions attempted to psychologically divide the American electorate during the 2016 election. Pay special attention to the possible ways in which topics are distributed spatially over a latent left-right ideological dimension. Discuss the potential security implications of this kind of psychological operation on the American electorate.
- Utilize sentiment analysis from the NRC dictionary to determine which words are related to various sentiments. Discuss the results of this analysis with respect to the topics uncovered in Step 1.
- Utilizing any features you wish to engineer, plus any features that are relevant from the dataset itself, develop a supervised learning model that achieves good predictive accuracy in the following task:
 - Remove all account categories from the data except for **LeftTroll** and **RightTroll**.
 - Develop a model or series of models which show high levels of predictive accuracy when predicting **LeftTroll** or **RightTroll** behavior.
 - Interpret the model in any way you choose. You may choose to discuss which words are highly salient predictors of various trolling behavior, engage in some unsupervised clustering, or any other method of model validation you wish to use.
- **HAVE FUN!**