Data Analytics Project 3

American Politics Sentiment Analysis

Text analytics has become an important tool for assessing sentiment in a variety of situations, including customer service, social media, and written works. Sentiment is defined as an immediate emotional response to an issue that may require immediate assistance. You may or may not have watched major American and European news channels like CNN or BBC and heard sentiments about American politics, i.e., views on President Trump. However, the news may be biased. Consequently, you will examine tweets from the United States and Europe with the subject of American politics.

You goal to perform sentiment analysis on a corpus of tweets regarding American politics and determine the actual sentiments. The data USTweets is available in my VIT GitHub directory:

Requirements:

Analysis: You may use Rstudio or Python (including Jupyter Notebook) to perform your analysis.

A **white-paper** report (Google what this consist of) with a maximum of five pages. You may add an appendix containing your detailed analysis results (no more than 10 additional pages). An appendix with your model code must be attached. This will give you the foundation of a conference paper.

A **PowerPoint slide deck** consisting of a maximum of about 10 slides:

1. Title slide
2. Business problem statement
3. Modeling objective
4. Model functional form and reasoning for its selection
5. Final model (variables with coefficients and explanation) summary
6. 1 deep-dive slide per model variable (a good model may only have three or four factors, even if all variables are statistically significant—modeling is an “art”).

Each slide in a PowerPoint slide deck needs to stand alone yet have few words and contain one graphic (chart, graph, table, picture, etc.) with words, but not verbose. Each slide should tell a story that an executive will understand. If you cannot effectively communicate your analysis, it is worthless—a model is only good if it is used.

Stages of Model Development

1. Business problem development and model requirement definition
2. Model design plan, including modeling objective and model functional from, plus potential data sources
3. Model development
    a. Data preprocessing

   b. Exploratory analysis
   c. Model training
   d. Model testing
4. Model post processing analysis
5. Model validation
6. Model Implementation

Things to consider while performing exploratory analysis:

1. Is a high LTI score better or worse than a low LTI score?
2. Is there a direct relationship between current home market value and income? If so, what is it? If not, why not?
3. Does the qualitative risk factor generally agree with the credit scores?
4. What is the relationship between marital status and the other variables if any?
5. What is the relationship between number of children and the other variables if any?
6. Continue to ask the data these kinds of questions.

Citations: Use ISO 690 with Numerical Reference.