**Predicting Positions Outlined in a Comment Letter**

1. **Introduction and Background**

**Rulemaking and Comment Process:**

When considering changes to existing rules, many federal agencies undergo what is known as a rulemaking process. As part of the rulemaking process, agencies will ask for comments from interested parties. Depending on the issue, interested parties take the form of but not be limited to individuals, businesses, corporations, membership organizations. A typical rule making can generate thousands upon thousands of comments.

The policy work I have been involved in often involves reading through pages of comments and identifying positions on issues and arguments raised during the comment process. For a typical rulemaking, I often triage the comments and only focus on several commenters deemed to be high priority targets based on their previous advocacy on an issue.

However, I believe having to ability to analyze the entire record of a proceeding and identifying the arguments and issued raised can be useful as an advocacy tool.

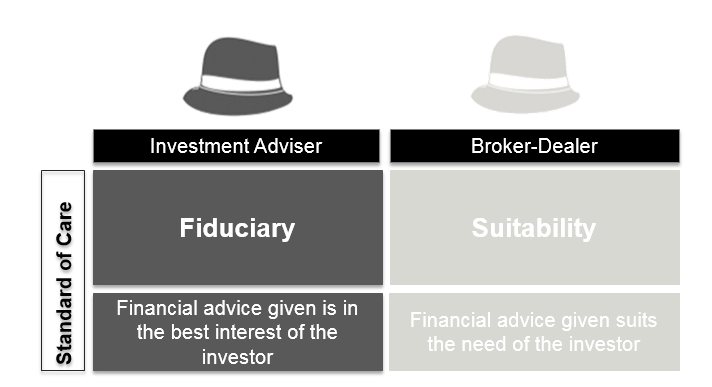
I am hoping to be able to analyze records of entire proceedings and produce guided analyses of comments using data science techniques.

**1.1 Department of Labor Conflict of Interest Rule**

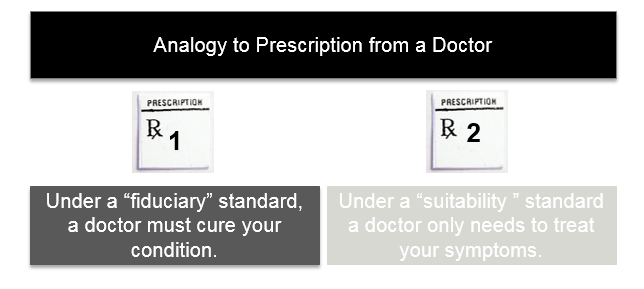
The DOL is currently engaged in a rulemaking proceeding to revise the conflict of interest rules around retirement advice for the first time in 40 years. As of August 10, 2015 there are over 2600 comments filed in this proceeding.

**Conflicted Advice:** Why does conflicted advice matter? Unless a financial services professional adheres to a fiduciary standard of conduct, the advice they give may not necessarily be in an investor’s best interest. This means that products or other advice offered could be driven by the potential for a commission or other compensation.

Generally speaking, RIAs are legally obligated to deliver financial advice in the best interest of investors, or under a “fiduciary” standard of care. However, broker-dealers (BD) only need to meet a suitability standard when delivering financial advice. There also exist many cases where an RIA is also registered as a BD and can “switch hats” when giving financial advice.



What does this mean for investors? Advice delivered under a fiduciary standard is the highest standard an investor can receive, whereas advice under a suitability standard is generally considered a lower standard. This becomes very problematic when an advisor (RIA or BD or dually registered) offers advice in the form of financial products.



Financial products can have certain commissions or fees for the advisor. Illustrated above is an analogy making the connection between fiduciary and suitability advice and a prescription from a doctor. Considering a simplistic hypothetical scenario where a medical condition can both be cured or the symptoms managed, a doctor giving a prescription under a fiduciary standard will cure the existing condition while doctor giving a prescription under a suitability standard may offer a prescription to manage symptoms rather than treat the underlying condition. To further illustrate that point, if a doctor opted for the “suitable” prescription option over the “fiduciary” prescription option because they are receiving commission from the pharmaceutical company, they are clearly giving conflicted advice.

1. **Data Sources**

Using Natural Language Processing and Naïve Baysian modeling I am attempting to determine whether comment letters filed with the Department of Labor on an the “Conflict of Interest Rule” falls into one of three categories: Support, Oppose, and Neutral.

Data set will be pdf of letters available at: <http://www.dol.gov/ebsa/regs/cmt-1210-AB32-2.html>

My training data will consist of hand-labeled comment letters Support:  31 examples, Oppose: 36 samples, and Neutral: 16 samples.

When I began the project there were 825 letters on the DOL website and as of today, there are nearly 2600.

The primary goal is to see whether or not I can create code that will be at least as good as me or my colleague in tagging what position a letter/ author has on the issue.

1. **Methodology - Creating a Dataframe from Hand-labeled data**

To create my initial dataframe for analysis, I manually created a csv file with the names of the authors from my training data set. I then downloaded the matching letters, converted them to a text file. Using glob I created a list from the letter text and appended it to my original csv file to create a dataframe for use in analysis

1. **Descriptive and Exploratory Analysis –DRAFT Current Status**

<Placeholder>

<ADD Detailed Data Dictionary of FINAL Features identified>

1. **Predictions from Training Data and Machine Learning**

<Place holder Add>

1. **Model Selection**

<Place holder Add>

1. **Predictions**

<Place holder Add>

1. **Conclusion**

<Place holder Add>