Dylan Saez
May 02 2022

Measuring Sentiment of Recession-Based Keywords Using News Articles and NLP –
PROPOSAL

What problem did you select and why did you select it?

In my work, I use a lot of NLP techniques such as Keyword Extraction, Topic Modeling, and
Sentiment Analysis on specific datasets for use cases. Therefore, I tried to use sentiment analysis
on a dataset I have been wanting to perform on for a while. I think that, given future extensions
of this project, this can be extremely useful in understanding how the market perceives specific
behaviors within the economy before, during, and after a recession.

What database/dataset will you use?

I will use the Factiva database. Using a method from one from the Barrett et al. (2021) paper, I
acquire the data by restricting the Factiva database. I will use the database in two batches: GFC
(2006-2009) and Post COVID-19 (20219-2022).

What NLP methods will you pick from the concept list?
1. To pre-process the data, SpaCy will be used
2. Numerical Encoding: tf-idf
3. ML algorithms: Naïve Bayes, LSTM, SVM

Will it be a classical model or will you have to customize it?
I will be using a classical algorithm with not too much customizing.

What packages are you planning to use? Why?
SpaCy
TF-IDF (Sklearn)
Sklearn: Naïve Bayes, LSTM
SVM

What NLP tasks will you work on?
Keywords, Topic Modeling, Numerical Encoding, Sentiment Tagging

How will you judge the performance of the model?
How well the model predicts the tag of sentiment for each article.

What metrics will you use?
Accuracy (F1 score)

Provide a rough schedule for completing the project.
The project will be complete by May 02, 2022