# Searching for Contrast in the Beige Papers: Predicting Regional Unemployment Using NLP

Douglas Callahan | CS224n | Stanford University SCPD

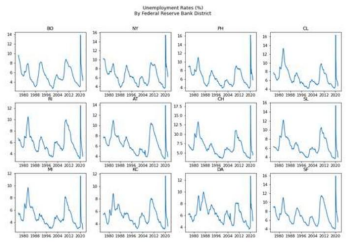
## Problem / Question

Can the "Beige Books" issued by the Federal Reserve district banks provide useful information for forecasting changes in regional unemployment rates?

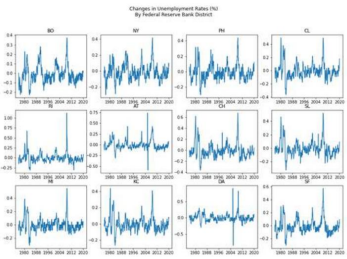
## Hypothesis

- Although imprecise and largely non-quantitative, the information in the Beige Books (if correctly analyzed and distilled) contains sufficient information to meaningfully supplement a model that predicts whether the regional unemployment rate will increase or decrease the next month.

## U.S. Regional Unemployment



## Response Variable: Month-over-Month Changes in Regional Unemployment



## Text Sample

- **Employment and Wages -** Employment has increased slightly since the previous report. Contacts throughout the District continued to note a tight labor market in a variety of industries, including construction, health care, manufacturing, and information technology. Companies have used a myriad of strategies to attract and retain workers, such as signing bonuses and paid time off. A contact in the trucking industry reported that insurance policies have prevented firms from hiring less-experienced drivers. Contacts in the agriculture industry near Memphis reported filling vacancies with temporary workers through the H-2A visa program.
- - excerpted from Beige Book Report of the St. Louis Fed (Apr. 17, 2019)

## Transformers

- To process the text samples from the Beige Books, three different transformer models were used:

**distilBERT**
- leaner version of BERT, with 40% of BERT's parameters but (supposedly) 95% of BERT's performance

**BERT**
- canonical transformer model with ~11 million parameters

**finBERT**
- BERT transformer further trained on Reuters RCV1 corpus of financial media and fine-tuned on the Financial PhraseBank

## Dimensionality Reduction

Originally, I intended to include the transformer models' final hidden states as an input to the time series. The results of this approach were disappointing, however. I suspected the problem lay with the high dimensionality of that hidden state (768 dimensions). I therefore engaged in two attempts at dimensionality reduction:

**PCA**
- First 30 PCA components of the transformer model's hidden state were added to the time-series data

**Fine-Tuning**
- Transformer model was fine-tuned on the classification task in the absence of any time-series data. The model's predictions on the classification task were then added to the time-series data.

## Classifier Heads

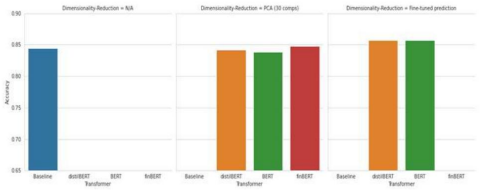
Three different classifier heads were used in the classification task:
- Feed-forward Neural Net
- Random Forests
- Gradient Boosted Trees

Ultimately, the Gradient Boosted Trees showed the best results on the test set and also (unlike the neural net) proved robust to overfitting.

Accordingly, all results displayed in this poster are based upon Gradient Boosted Trees.

## Results

- Below are the Accuracy statistics for the several models, all run on Boosted Trees:




## Conclusion

- The baseline classifier boasted an impressive ~84% accuracy rate without any supplemental NLP data
- Failure has many fathers: the somewhat disappointing improvements seen in the supplemented classifiers could have several causes
  - Poor preprocessing of data – reducing the size of the text excerpts would likely increase the transformers' ability to extract meaningful representations from them
  - Imbalanced dataset (far more decreases than increases in the unemployment rate)
  - Insufficiently large dataset, which prevented the direct use of the 768-dimension hidden states and led to generally unsuccessful dimensionality reduction techniques

## Relevant Works


- Araci, Dogu. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models, Thesis Submitted in Support of Masters of Data Science at the University of Amsterdam (2019), https://arxiv.org/pdf/1908.10063.pdf/