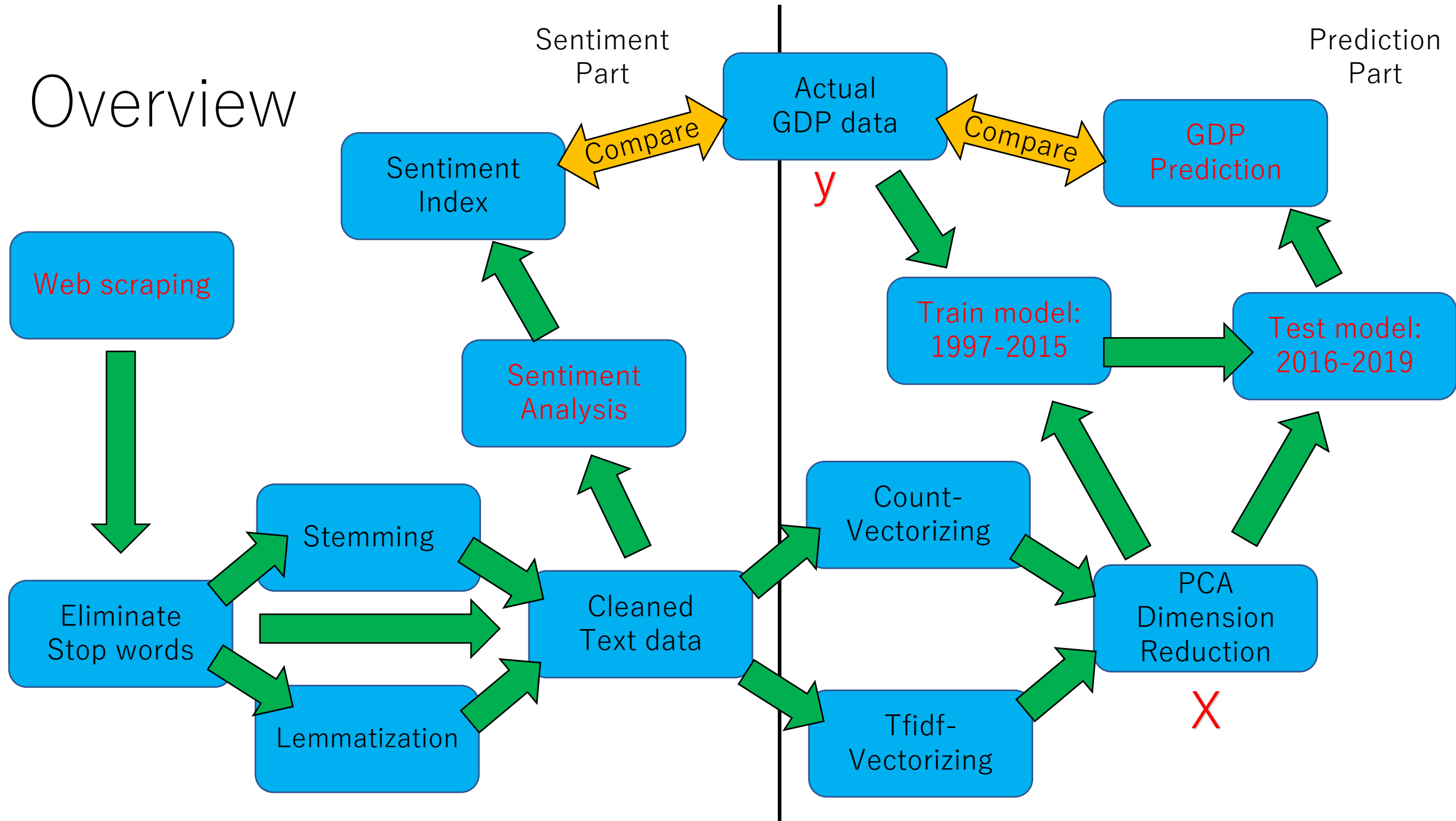# Sentiment Analysis of Beige Book & the Prediction for GDP

QMSS
Taketo Muroya

The Beige Book

- Officially, Summary of Commentary on Current Economic Conditions by Federal Reserve District
- Each Federal Reserve Bank gathers information on current economic conditions in its District
- Published eight times per year

- Word Cloud Example(June, 2019):

Web scraping

- BeautifulSoup
- Year: 1997-2019
- Get URLs and then extract Text
- Different layouts need different coding

The Federal Reserve Board

January 22, 1997

Summary

Summary

Districts

Boston
New York
Philadelphia
Cleveland
Richmond
Atlanta
Chicago
St. Louis
Minneapolis
Kansas City
Dallas
San Francisco

Full report

Prepared at the Federal Reserve Bank of Atlanta and based on information collected before October 20, 1997. This document summarizes comments received from businesses and other contacts outside the Federal Reserve and is not a commentary on the views of Federal Reserve officials.

Most District reports characterized early autumn's economic activity as moderate to strong in most areas. Retail sales in most Districts in September were weaker than anticipated, with some pickup in October. Automobile sales were mostly slower. Manufacturing activity accelerated or remained at high levels for most Districts, and the outlook was generally positive. Despite the residential real estate markets' recent uneven performance, activity remains at elevated levels overall, and most commercial real estate markets continue to perform at strong levels. The energy extraction sector continues to strengthen. Labor shortages have intensified for key sectors, such as the high-tech and energy extraction industries. Yet, reports of increasing wage pressures are infrequent. Prices remain stable, as competitive factors and resistance from buyers are making it difficult for producers to pass on any cost increases. As harvesting progresses, recent rains have had varying effects on crops nationwide. Commercial lending remained strong in many Districts, while consumer lending was more mixed.

Home > Monetary Policy > Beige Book > November 27, 2019

Beige Book

Full Report

National Summary

National Summary

This report was prepared at the Federal Reserve Bank of Dallas based on information collected on or before November 18, 2019. This document summarizes comments received from contacts outside the Federal Reserve System and is not a commentary on the views of Federal Reserve officials.

Overall Economic Activity

Economic activity expanded modestly from October through mid-November, similar to the pace of growth seen over the prior reporting period. Most Districts reported stable to moderately growing consumer spending, and increases in auto sales and tourism were seen across several Districts. In manufacturing, more Districts reported an expansion in the current period than the previous one, though the majority continued to experience no growth. The picture for nonfinancial services remained quite positive, with most Districts reporting modest to moderate growth. Transportation activity was rather mixed across Districts. Reports from the banking sector indicated continued but slightly slower growth in loan volumes. Home sales were mostly flat to up, and residential construction experienced more widespread growth compared to the prior report. Construction and leasing activity of nonresidential real estate continued to increase at a modest pace. Agricultural conditions were little changed overall, remaining strained by weather and low crop prices. Activity in the energy sector deteriorated modestly among reporting Districts. Outlooks generally remained positive, with some contacts expecting the current pace of growth to continue into next year.

Sentiment Analysis

Real GDP level over time

Real GDP rate over time

Sentiment index over time

- Simple counting Positive or Negative words using the same dictionary in class

- The index seems like the combination of GDP level and rate

- Has Upward Trend and captures the Recession periods such as 2008
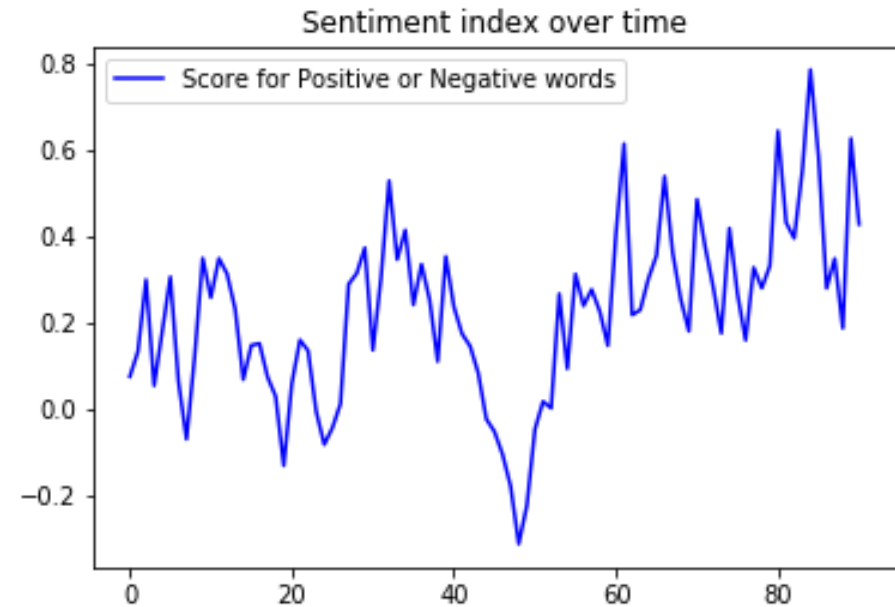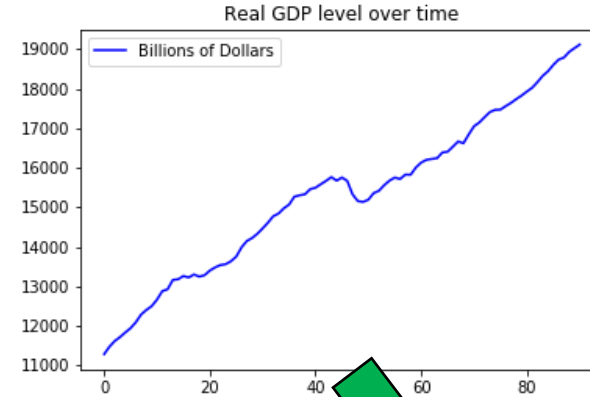
Model Training

- Random Forest Regression (n=1000, PCA_target=0.95)
- train_y = GDP growth rate (1997-2015)
- train_X = Text vectors (1997-2015)
- Tfidf is usually better than Count-Vectorizing
- Lemmatization looks better than Stemming

	n-gram=1 Tfidf-Vec	n-gram=1 Count-V	n-gram=2 Tfidf-Vec	n-gram=2 Count-V	n-gram=3 Tfidf-Vec	n-gram=3 Count-V	n-gram =1-3 Tfidf-Vec	n-gram =1-3 Count-V
Stop Word	Train:0.90 CV:-7.49 Test:-2.79	Train:0.91 CV:-6.18 Test:-3.16	Train:0.91 CV:-6.82 Test:-1.91	Train:0.89 CV:-6.56 Test:-0.88	Train:0.88 CV:-8.40 Test:-2.08	Train:0.86 CV:-8.47 Test:-1.41	Train:0.92 CV:-5.55 Test:-0.87	Train:0.91 CV:-5.85 Test:-2.03
Stemming	Train:0.93 CV:-6.43 Test:-1.04	Train:0.92 CV:-7.16 Test:-0.91	Train:0.90 CV:-7.74 Test:0.21	Train:0.89 CV:-7.26 Test:-1.81	Train:0.87 CV:-8.56 Test:-0.30	Train:0.88 CV:-7.49 Test:-0.40	Train:0.92 CV:-7.47 Test:-0.24	Train:0.91 CV:-7.23 Test:-0.63
Lemmatization	Train:0.91 CV:-6.10 Test:-0.36	Train:0.92 CV:-5.77 Test:-2.28	**Train:0.90 CV:-6.44 Test:0.25**	Train:0.89 CV:-6.73 Test:-0.96	Train:0.87 CV:-7.83 Test:-1.08	Train:0.87 CV:-7.41 Test:-0.83	Train:0.93 CV:-5.52 Test:-1.23	Train:0.89 CV:-6.36 Test:-1.29

Stop Word: Eliminate stop words,   Stemming: Eliminate stop words + Stemming,   Lemmatization: Eliminate stop words + Lemmatization

Prediction for GDP

- Prediction for GDP rate is not so accurate
- The data have Autocorrelation which might affect the accuracy of prediction

Prediction for GDP rate (Tfidf-Vec, n=2, lemmatization)

Train score: 0.90
CV score: -6.44
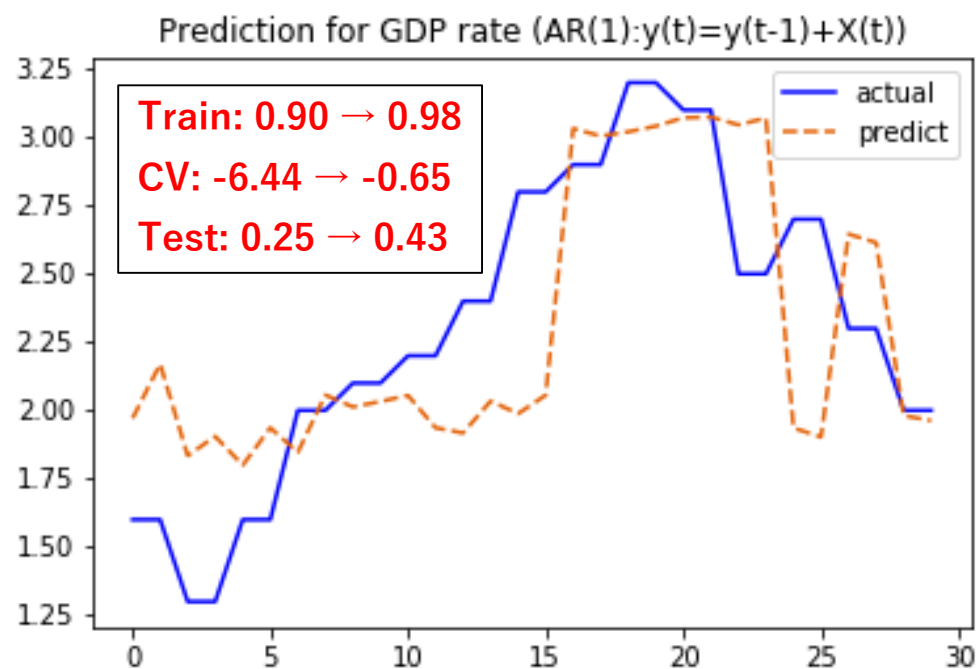Test score: 0.25

Prediction for GDP

- Prediction for GDP rate is not so accurate
- The data have Autocorrelation which might affect the accuracy of prediction

Prediction for GDP rate (AR(1):y(t)=y(t-1)+X(t))

Train: 0.90 → 0.98
CV: -6.44 → -0.65
Test: 0.25 → 0.43
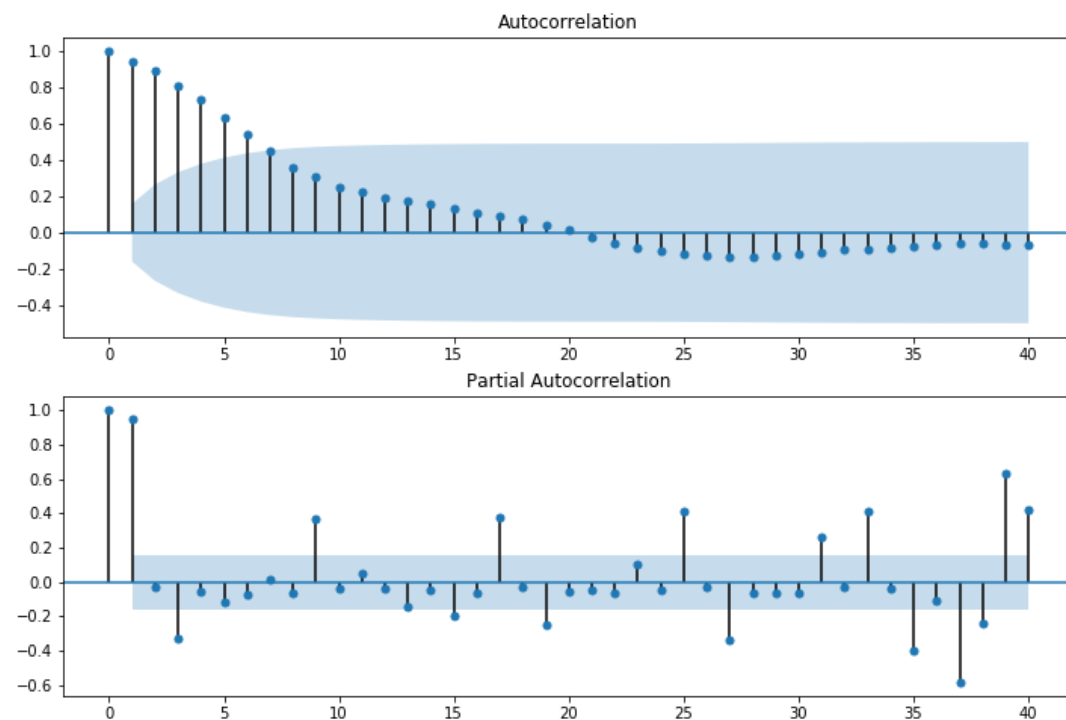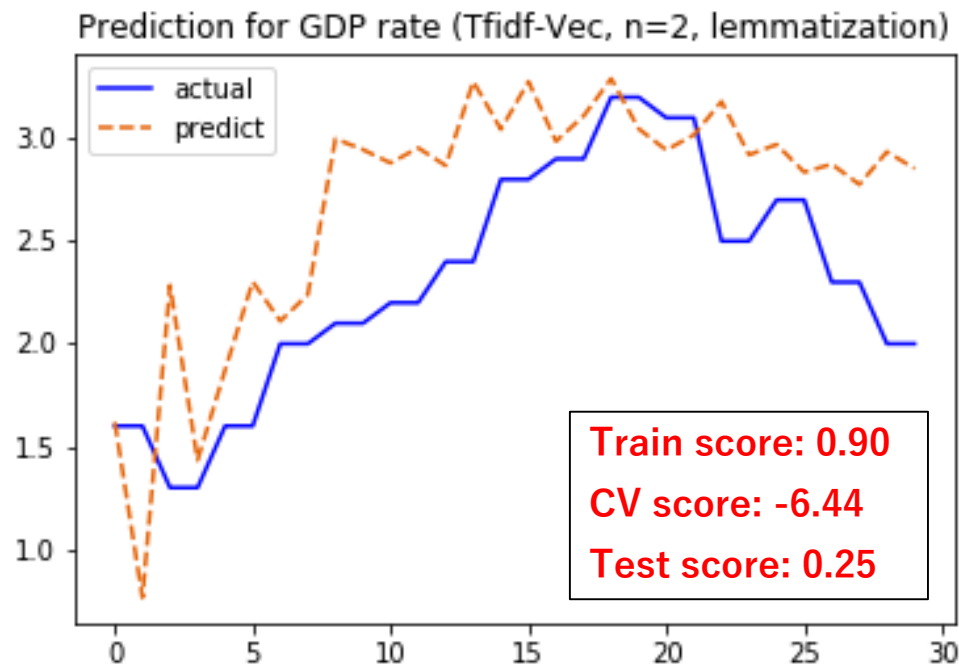
Autocorrelation

Partial Autocorrelation
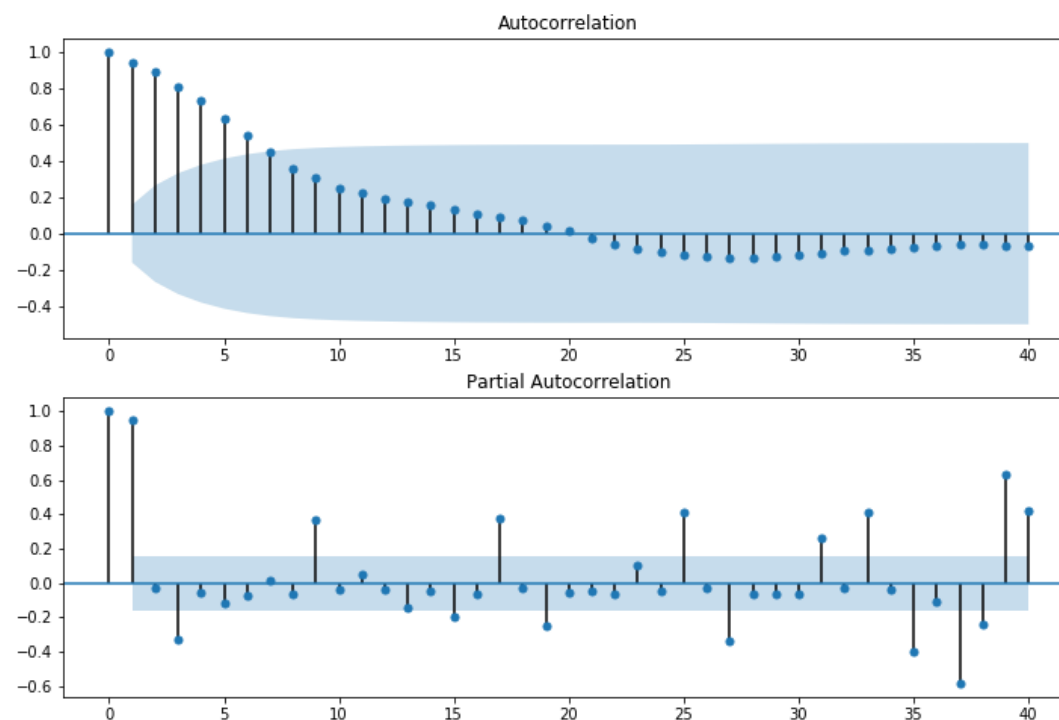
$$AR(1): GDP\_rate_t = \textbf{GDP\_rate}_{t-1} + Text\ Vector_t + e_t$$

Conclusion

Sentiment Analysis

✓Although the sentiment index just counts positive or negative words in Beige Book, it seems to <u>roughly capture the features of economic conditions</u>

✓<u>Special dictionary of economic sentiment words</u> (if it exists) could help the accuracy of sentiment index

Prediction for GDP

✓Tried to make the supervised-learning prediction <u>using the different data source (Beige Book & GDP)</u>

✓<u>Time Series model such as ARIMA</u> could improve the accuracy of prediction (Trying to find the best Time Series model)