Columbia University

IEOR4742 – Deep Learning for OR & FE

Project Summary (Group 14)

EPS Prediction

Authors contributing to this work:

Abhishek Mukherjee, Xueru Wang, Sie Hendrata

1. **Project Objective**

We want to use price and fundamentals information to predict future earnings surprise which is the difference between analyst forecast and real EPS. There are two tasks we plan to do:

1) Predict an earnings surprise above certain percentage thresholds;

2) Predict the level of EPS on a quarterly basis;

Finally, we can use the above information to long/short the specific stocks, thus setting up a portfolio in the financial sector of S&P 500. And, also we can optimize the Portfolio using certain Portfolio optimization metric like Sharpe Ratio or Sortino Ratio.

1. **Previous groups’ work**

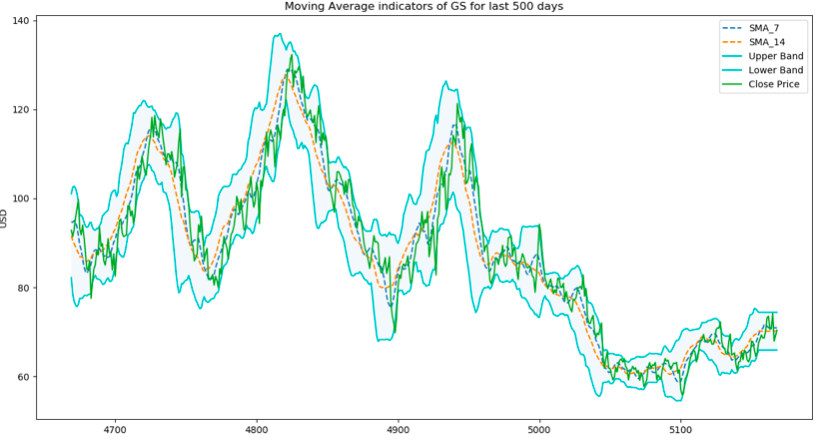
None

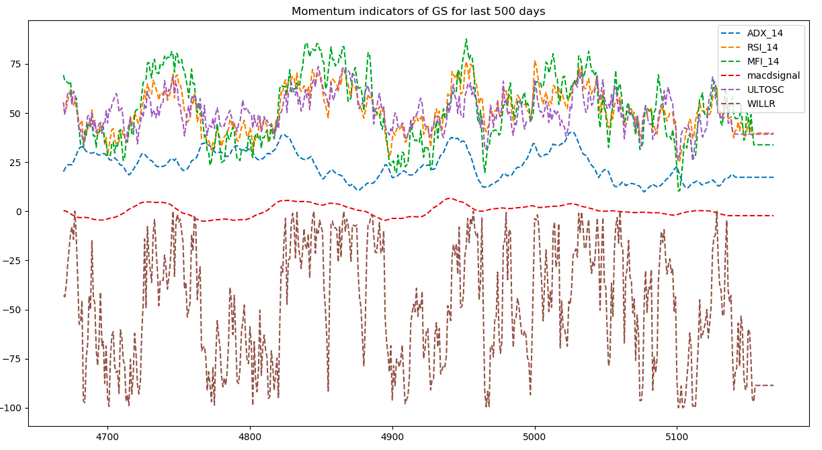
1. **What we have done**

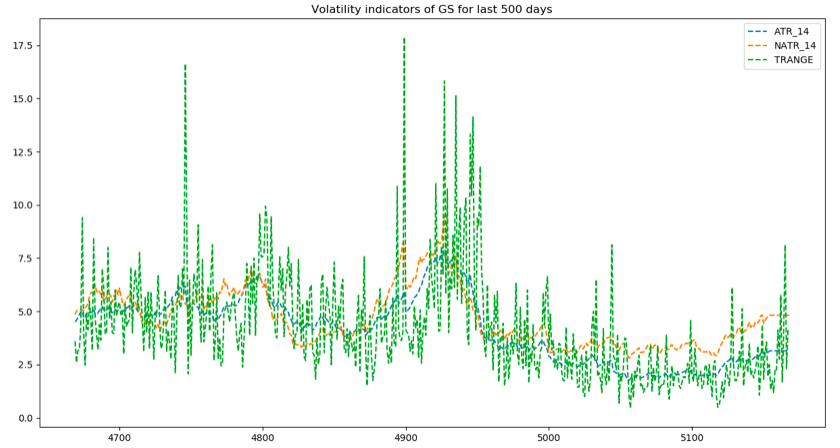
We have collected daily price data, quarterly fundamental data, and quarterly analyst forecast data of S&P 500 form FactSet. Data collected from 1997 to 2019.

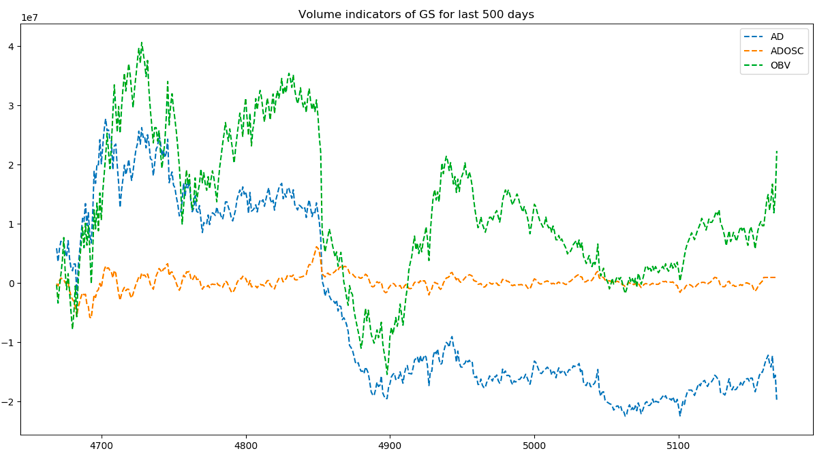
Up to now, we choose GS (Goldman Sachs) as a single stock example and what we have done are as follows:

1. **Technical Indicators**: we created 178 price Technical Indicators of GS for last 500 days, including Moving Average indicators, Momentum indicators, Volatility indicators, Volume indicators and MACD, the plots are drawn below:



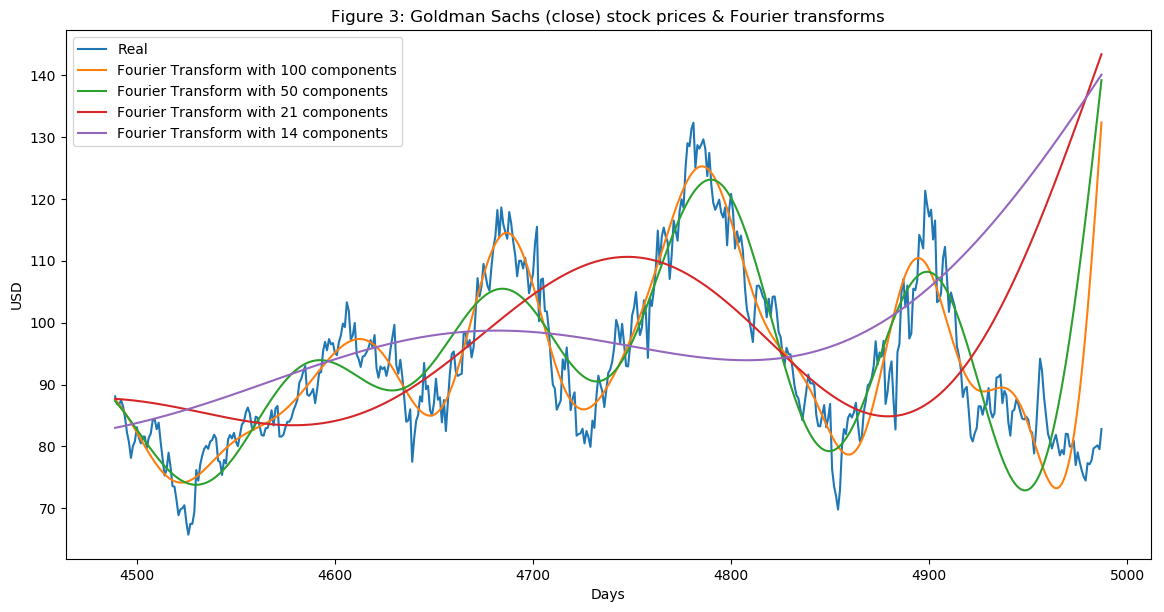




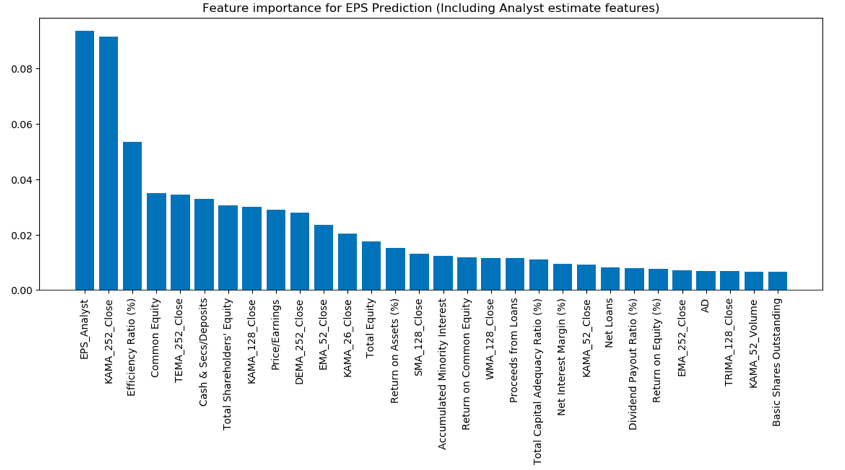


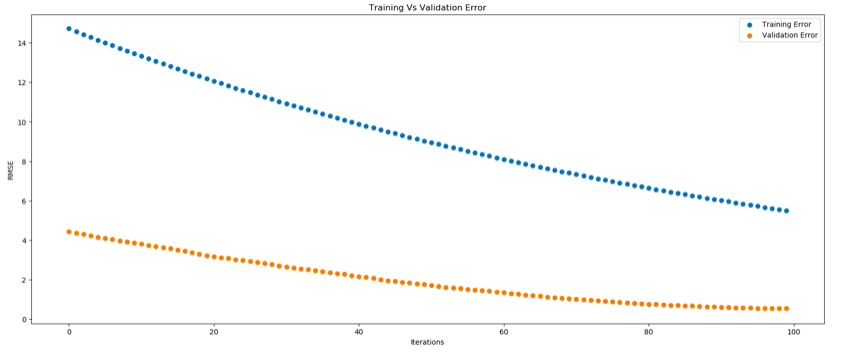


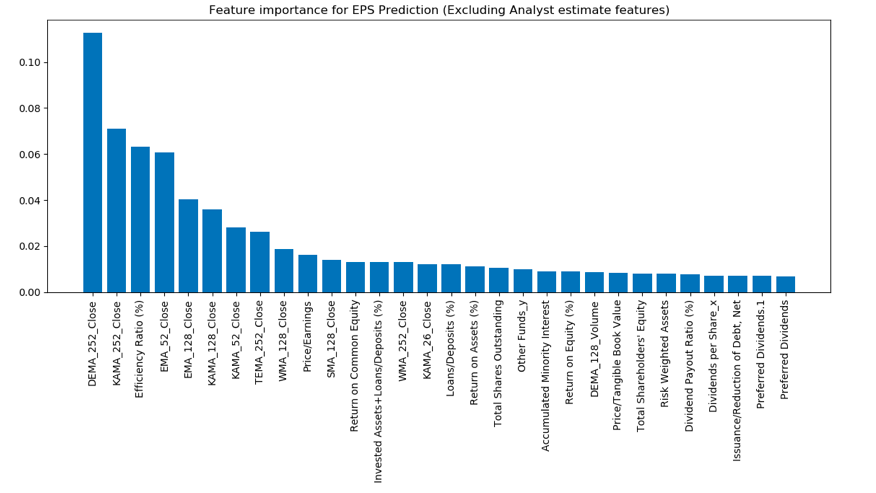
1. **Fourier transforms**: along with the daily closing price, we created Fourier transforms in order to generalize several long- and short-term trends. Using these transforms we will eliminate a lot of noise (random walks) and create approximations of the real stock movement, the transforms look like this:

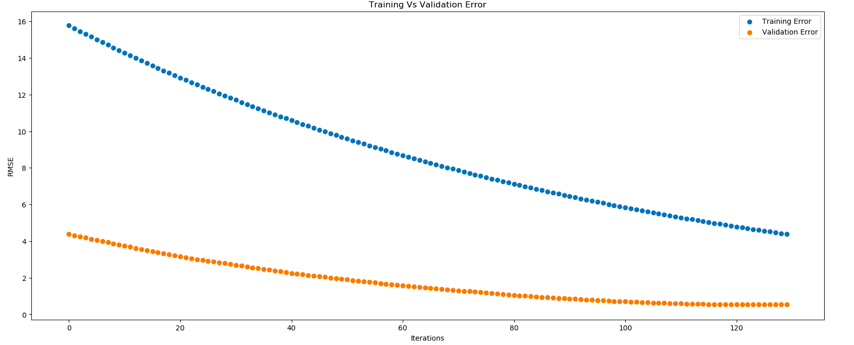


1. **Feature importance with XGBoost**: having so many features we have to test feature importance for predicting EPS. Here are our results for including analyst estimate features and excluding analysit features, both of them don’t have overfitting problem.



****

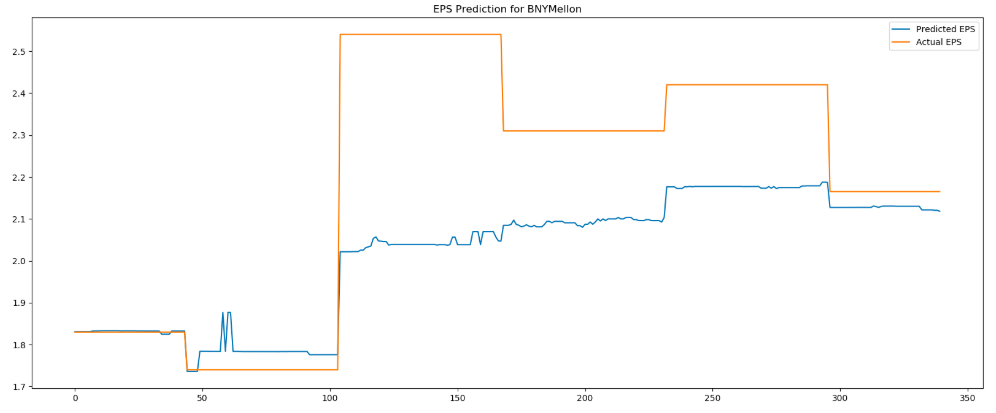
****

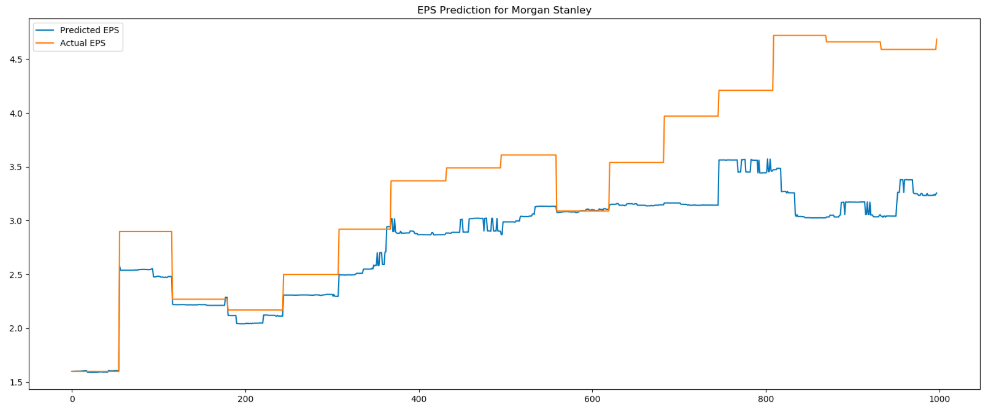
****

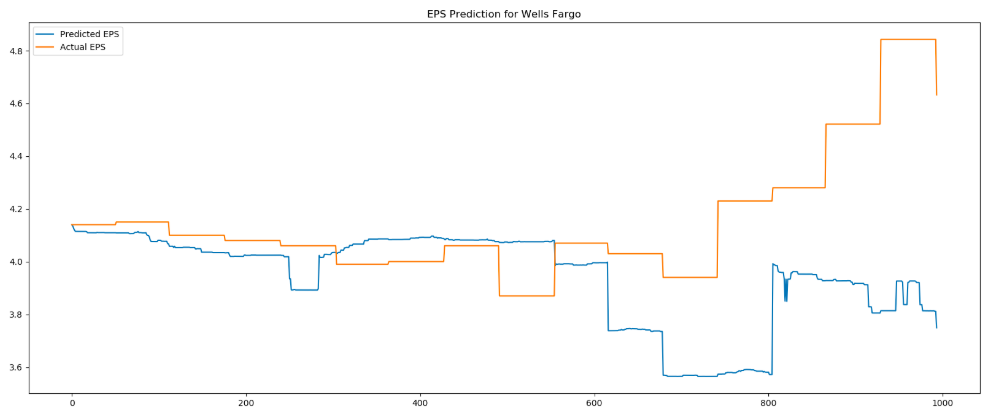
**Some Important conclusions regarding EPS prediction:**

* Also, keeping the **top 30 - 50 features** selected from the XGBoost’s Feature Importance list improves the model’s accuracy by lowering the mean root square error. For example, keeping 33 most predictive features for Goldman Sach’s EPS prediction improves Model's prediction accuracy by reducing Root Mean Squared Error from 3.33 to 2.91.
* **Keeping EPS\_Analyst feature** improves our Model's predictive accuracy for all the tested stocks which matches with out intuition in the sense that Analyst does a good job.
* Using **percentage changes in fundamental data** from past quarter and past year (for example, **((pretax\_income(current) – pretax\_income(prev\_quarter))/pretax\_income(prev\_quarter))** improves our model’s prediction accuracy drastically and its intuitive in the sense that change in fundamental dataset is a **good comparative indication** of how well the stock is performing and can be used to predict the future eps. We feature engineering all the fundamentals by incorporating their change and improved our model’s accuracy.

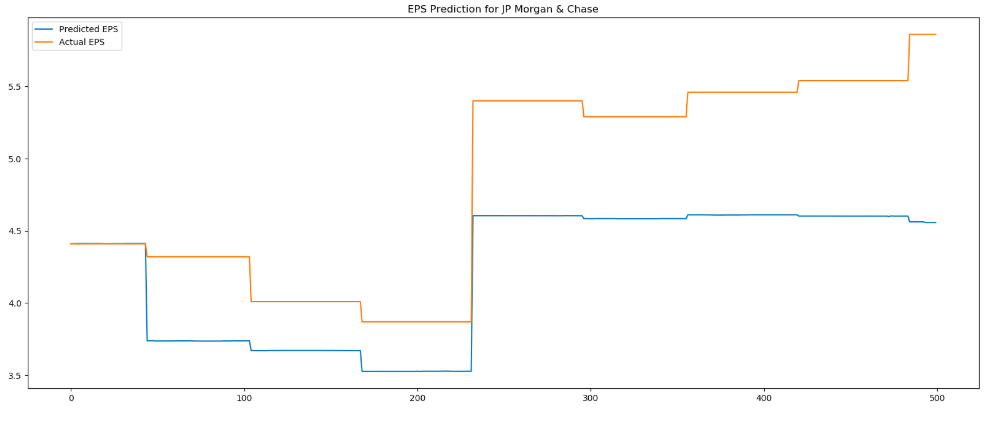
**EPS Prediction Results for few Stocks:**

**BNY MELLON:**

**Morgan Stanley:**

**Wells Fargo:** 

**JP Morgan & Chase:**



**As you can see, our model is successfully predicting if the eps is going to go up or go down and this is because of our clever feature engineering and model selection. Although, it is very difficult to predict the actual value of the eps but still our model doesn’t deviate from the actual value by a large margin and hence our model can be used for EPS prediction.**

**Also, it is important to notice that the model’s EPS prediction for a longer future is not that accurate and it is intuitive in the sense that currently it is difficult to predict the EPS for the next 3 or 4 years.**