Applied Data Science Project 5

Jason LI   Dejian WANG   Yuxin ZHU   Ke HAN
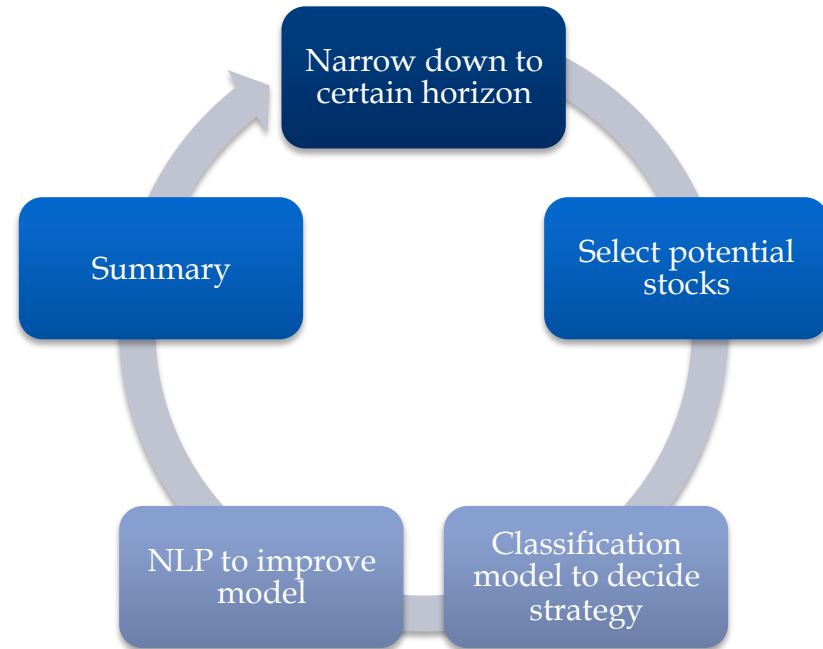
Introduction

Scope

- Using the financial data set
- Focused on Earning release date related days
- Using the US market most liquidity stock
- Across more than 10 years investment horizon

Data Description

- Using more than 1000 stocks
- Using more than 39000+ entries
- Including financial fundamental information

Data Science related

- Predictive modeling
- Classification models
- NLP to generated features
- Data visualization

Narrow down to certain horizon

Select potential stocks

Classification model to decide strategy

NLP to improve model

Summary

Step 0: Related Papers and Projects

Adaptive Asset Allocation

- Adaptive asset allocation is a process of constantly rotating into asset with
  - Adaptive momentum
  - Volatility
  - Correlation factors

Momentum

- Defined as the an asset with the greatest relative performance over the past 1 month to 1 year is more likely to exhibit stronger performance over the next few days or weeks
- Some reasons for Momentum:
  - perceptions of risk
  - human cognitive biases
  - the operation of markets

Earning announcement related

- There are two main indicators:
  - SUE (Standardized unexpected earnings)
  - EAR (Earning announcement return)

$$SUE_{i,q} = \frac{X_{i,q} - E(X_{i,q})}{\sigma_{i,q}}$$

$$EAR_{i,q} = \prod_{j=t-1}^{t+1}(1 + R_{i,j}) - \prod_{j=t-1}^{t+1}(1 + FF_j)$$

Here, FF stands for the benchmark size &book t- market FF portfolio

- And the related trading strategy are:
  - Long the stocks with positive announcement
  - Short the portfolio with negative announcement

Step 1: Data collection and Description

Load data

- Data source: Using Bloomberg terminal, Scraping from Yahoo and Wall Street Journal
- Size: including financial data 39034 entries (based on the volatility, choose the most volatility ones), news related txt files

Horizon

- Based on the Earning released projects on the previous page, we focused on the earning released date -6 to 6 horizon
- In our data set, the earning release is covering 2006 to 2016 earning release (can refer to earning calendar)

Data processing

- Calculated and processed the dividend yield ratio, enterprise value to EBITDA ratio, market to book ratio, Momentum, price to sales ratio, days(based on the data to earning release) etc.
- For example here,
  - Momentum is based on the price
    $(CLOSE(J) / CLOSE(J - N) * 100)$
  - Fama French Book to market ratio

$$\text{Book to Market} = \frac{\text{Book Value of Firm}}{\text{Market Value of Firm}}$$

*Full code can be found in the R markdown file in github
*http://www.nasdaq.com/earnings/earnings-calendar.aspx 4

Step 2: Select the stocks based on their Momentum

| Select Stocks | Statistical Description of the data |

Step 3: Classification model

Final model

- After feature selection and running linear regression,

Features using
stepwise to
choose
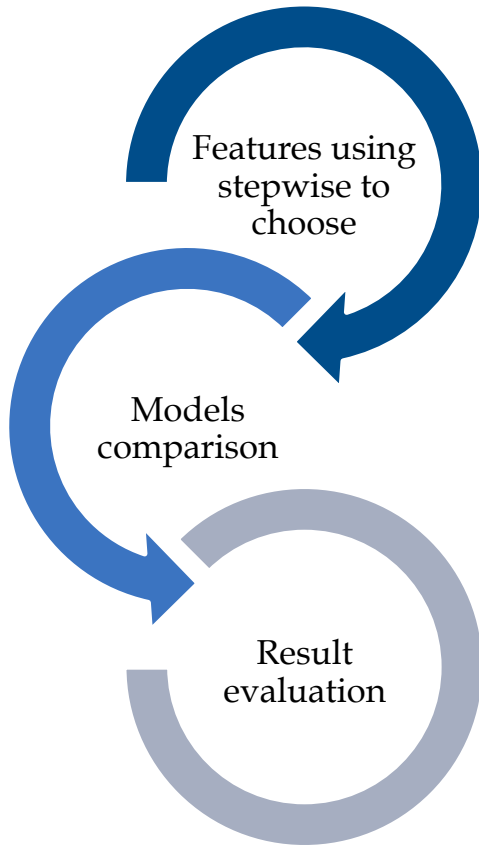
Models
comparison
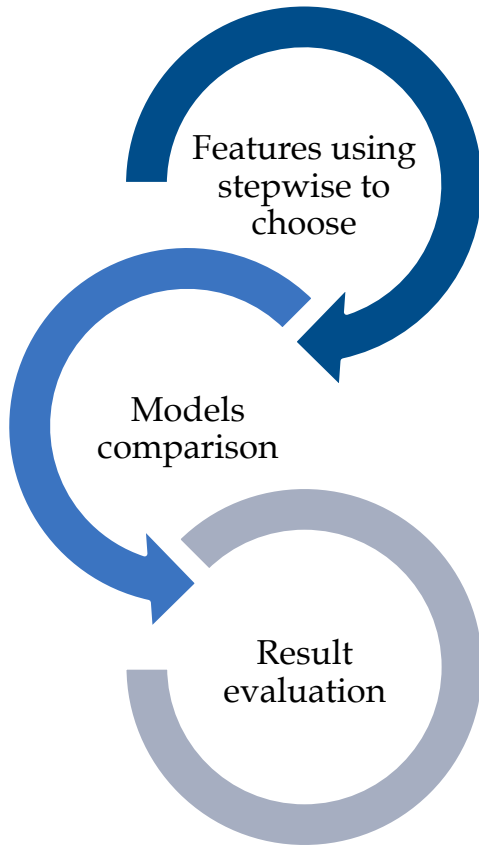
Result
evaluation

*Full code can be found in the R file attached

Step 4: Including news features

Final model

- After feature selection and running linear regression,

Features using stepwise to choose

Models comparison

Result evaluation

*Full code can be found in the R file attached

Step 5: Strategy and Back test

- (what results did the analysis produce and how do you interpret those results)

*Full code can be found in the R file attached

Summary and Improvement

Summary

- Machine learning algorithms application, potential monetization

More data

- Use data not only for
- More types,

More Feature

- Using the weather information
- To analyze characteristics, including

Algorithm optimization

- It could be improved with other distributed algorithms
- Some data preparation technique/ subgroup analysis could be done

Thank You !

Contact Information: Jason LI
Dejian WANG
Yuxin ZHU
Ke HAN