Deriving Value From Labor Market Data

Team Gannett Peak: Congda Xu, Binqi Shen, Matthew Ko, Isaac Choi

Agenda

01 **Data Manipulation**

02 **Model Implementation**

03 **Trading Strategy Development**

04 **End Analysis**

Data Manipulation

Data Preparation

Greenwich's Job Market Data

- Filtered out jobs that were posted after Jan. 31, 2020: avoid COVID impact
- Treated each company's data in each month as a single observation
- Extracted & Aggregated all features **in the monthly interval for each company ticker**

Stock Data

- Scraped from online API: Alpha Vantage
- **Monthly Closing Price/Return** as our response variable

Feature Engineering

Salary
Average salary of job listings for each company in each month

Job Posting
Number of active job postings for each company in each month

Average Posting Duration
Average cumulative posting days for all active jobs for each company in each month

New Posting
Number of new job postings for each company in each month

New Cbsa / Max & Min Cbsa
Total number of postings to new geographic areas (core-based statistical area);
Extract if company is expanding to more urban or rural areas.

Int Count / Weight
Total number of entry-level roles postings;
Percent change in entry-level roles posting

Imp Count / Weight
Total number of important roles postings;
Percent Change in important role postings

Model Implementation

Model Implementation

1. Model Selection

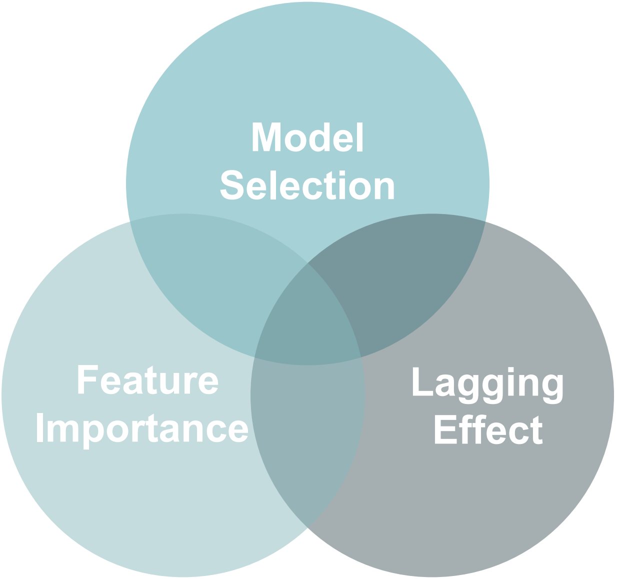
- Classification
- Regression

2. Feature Importance

- Value of Greenwich’s proprietary data

3. Lagging Effect Possibilities

- Test 6 Possibilities



Model Selection

1

Stock Trend Prediction: Random Forest Classifier

Predict Two Classes:
positive return & negative return

Performance Metric:
Accuracy: 57% ~ 65%

Identified feature importance

2

Stock Return Prediction: Vector Auto Regression Model

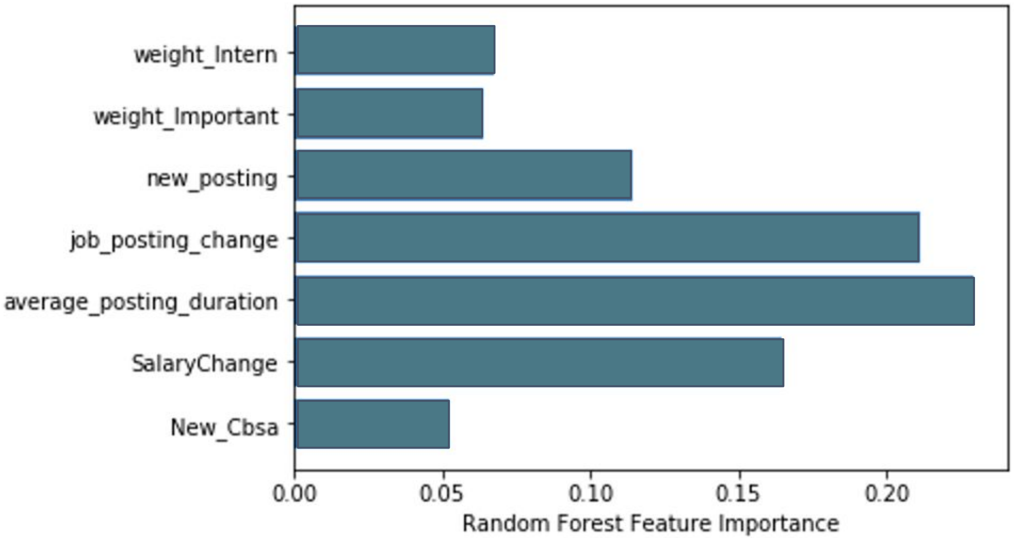
Predict Stock Returns:
stock price → then translate into return

Performance Metric:
Mean Absolute Percentage Error (MAPE): ~ 18%

Individual time series model for individual stocks

Feature Importance

- Relative Feature Importance from Random Forest Model



Lagging Effect Possibilities

Tested 6 different lagging relationships: 0 month to 5 months

- Optimal lagging period for both models: **Two-Month Lagging**
- Metrics: **Mean Absolute Error** & **Mean Absolute Percentage Error**

Lagging	MAE	MAPE
0	6.1109	18.8887
1	5.9221	18.9751
2	5.8275	18.6476
3	6.0673	18.9849
4	7.3849	24.3747
5	6.9415	22.2253

Trading Strategy Development

Stock Picking Strategy

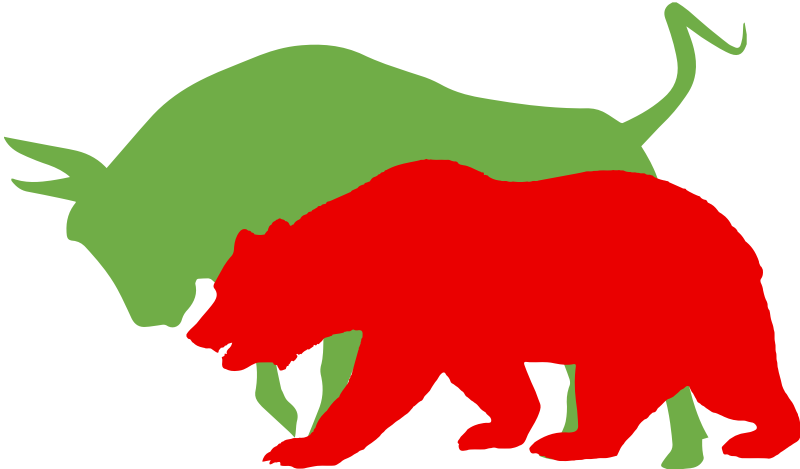
1. **For each month: Pick the top 10 stocks**
   - How to select
     - According to **classification prediction**, positive return for at least 1 of the next 2 months
     - According to **regression prediction**, predicted return in range **2% ~ 20%**
       - Why less than 20%: too much model risk for prediction over 20%; needs fundamental research
       - Why over 2%: Average monthly return for S&P is less than 2%
     - Rank stocks in this range by their **Sharpe ratio**
     - Exclude stocks in **Dow Jones**
     - Select top 10

Stock Picking Strategy

2. For each month, add from top 10 list to original portfolio

3. Drop those in the portfolio which are predicted to have negative return in next two consecutive months

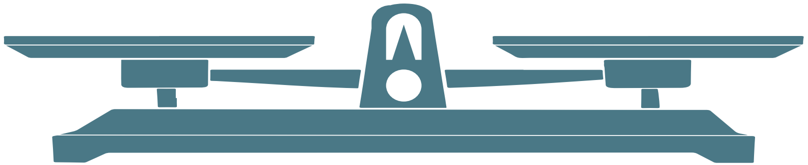
4. Drop those in the portfolio which have lost money in the past two consecutive months



Weight Allocation Strategy

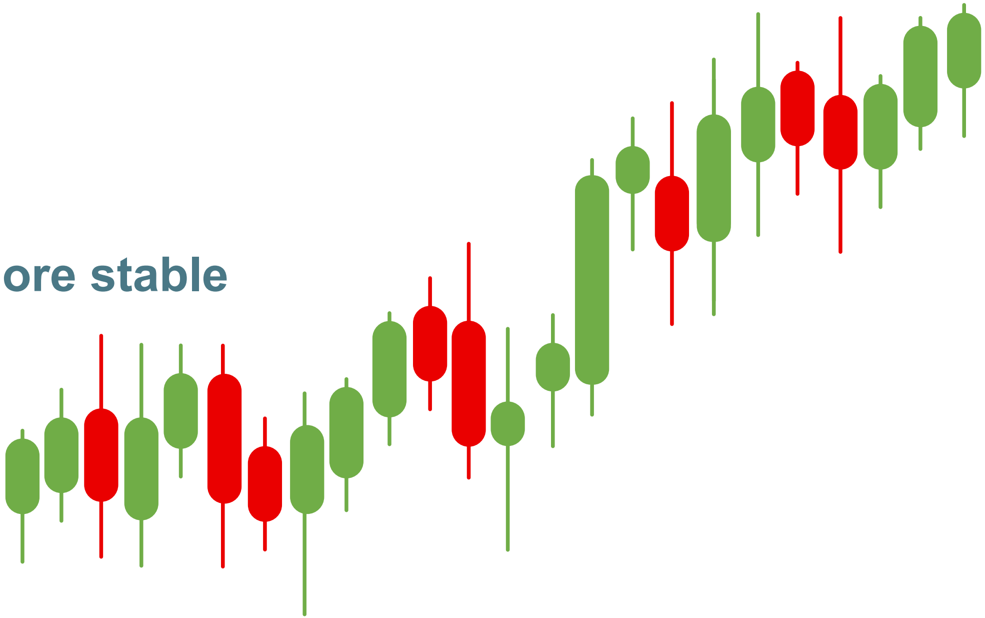
- **Initially, tried CAPM tangency portfolio weight optimization**
  - Problem 1: negative weights for stocks, we don't want shorting here
  - Problem 2: too much zero weight assigned to stocks in the portfolio
- **Two solutions**
  - Plan A: Each stock just buy one (can be scaled)
  - Plan B: Assign weight according to Sharpe ratio of stocks:

$$w_i = \frac{SR_i}{\Sigma SR_i}$$



Strategy Simulation

- **Testing period**
  - 2019-08 ~ 2020-01
  - S&P performance: 8.2%
- **Strategy performance**
  - Plan A: 16.6%
  - Plan B: 11.5%
- **In longer term, Plan B might be more stable**



Portfolio Tracking

2019-08

HEI
PLNT
MDB
CABO
INTU
MA
OKTA
RCM
PGR
CTAS

10

2019-09

HEI
PLNT
MDB
CABO
INTU
MA
OKTA
RCM
PGR
CTAS
CWST
BAND
WM
TDY
REXR

15

2019-10

HEI
MDB
CABO
INTU
MA
RCM
PGR
CTAS
CWST
BAND
WM
TDY
REXR
NEE
WCN
AYX
NOVT

17

2019-11

MDB
CABO
MA
RCM
PGR
CTAS
CWST
WM
TDY
REXR
NEE
WCN
AYX
NOVT
RSG
SUI
ELS

17

2019-12

MDB
CABO
MA
RCM
PGR
CTAS
CWST
WM
TDY
REXR
NEE
WCN
AYX
NOVT
RSG
SUI
ELS
APPF
MSA
PLNT

20

2020-01

MDB
CABO
MA
RCM
PGR
CTAS
CWST
WM
TDY
NEE
WCN
AYX
NOVT
RSG
SUI
ELS
APPF
MSA
PLNT
ADBE
AMD
SPGI
CDNS
OKTA

24

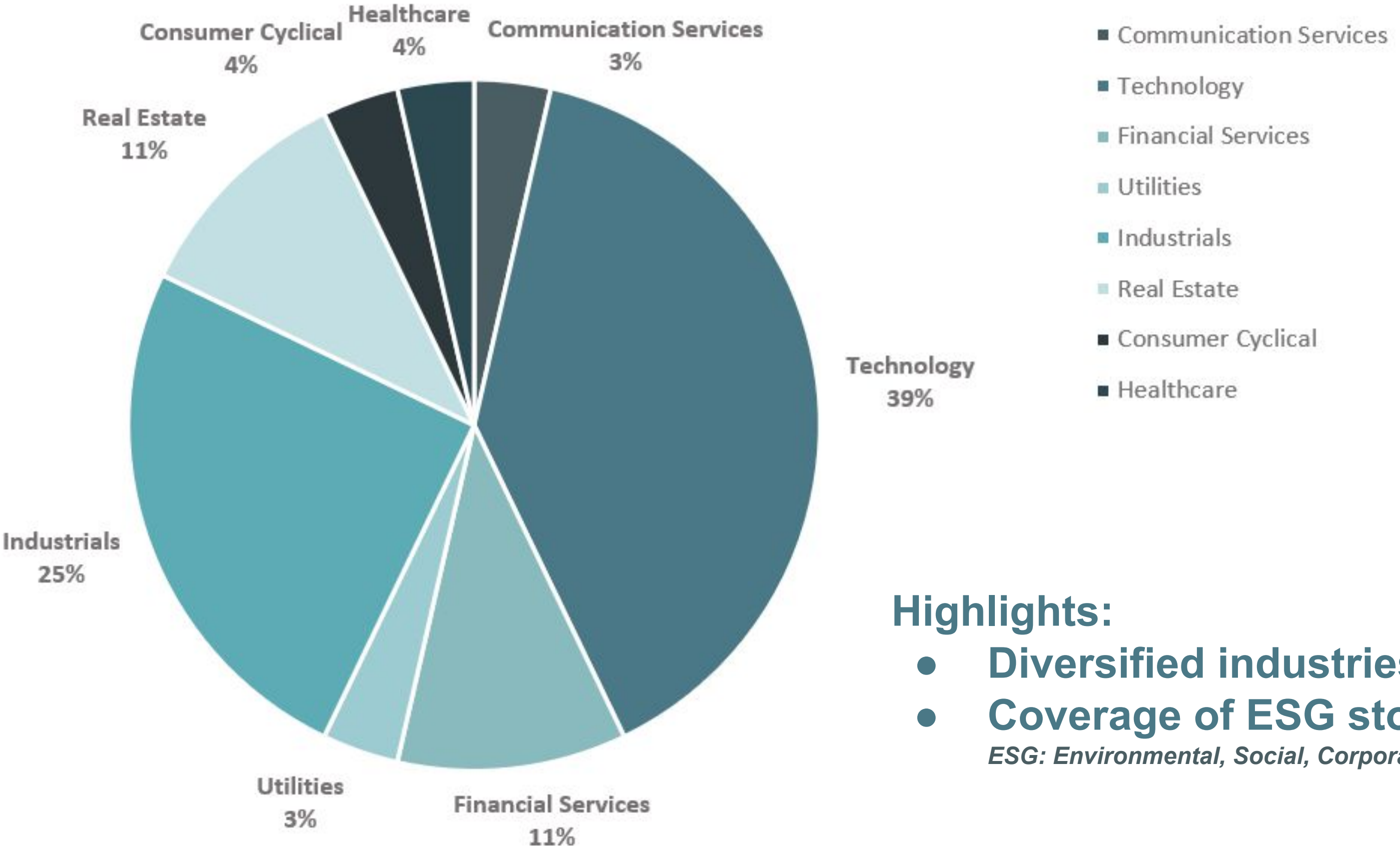
	2019-08	2019-09	2019-10	2019-11	2019-12	2020-01
Plan B	1.4%	-10.3%	0.8%	3.4%	-1.5%	19.6%
S&P	-1.8%	1.7%	2.0%	3.4%	2.9%	-0.2%

Portfolio Performance

CABO	41.52%	NOVT	15.15%	WCN	6.68%	PLNT	3.51%
AYX	33.65%	CWST	13.34%	AMD	6.37%	OKTA	1.61%
TDY	21.19%	MSA	11.60%	CD	5.48%	RCM	0.40%
MA	19.47%	RSG	9.89%	ELS	5.01%	SUI	-0.16%
APPF	17.10%	SPGI	9.82%	PGR	4.35%	INTU	-6.95%
MDB	16.90%	CTAS	9.80%	REXR	4.20%	HEI	-9.81%
NEE	16.47%	ADBE	8.17%	WM	3.77%	BAND	-35.60%

Winner: 85.7% **Loser: 14.3%** **Outperformed S&P: 46.4%**

Portfolio Decomposition



Highlights:

- **Diversified industries**
- **Coverage of ESG stocks**

ESG: Environmental, Social, Corporate Governance

End Analysis

End Analysis

Risk Analysis

- Time Limit in the data used for the project
  - Longer time: More stable variation & Pick up potential seasonal trends
- The Models chosen for this project are representative of their performance on this dataset, their performance on new data is unknown
- Weight allocation strategy we used are experimental
  - Further research might provide space for improvement

Research Potentials

- More complex time series models: Long Short Term Memory networks
- Further feature engineering and information retrieval
- Advanced weight allocation strategies
- Implementing multiple lagging times simultaneously in strategy

Any Questions?


Thank You