ML-based Investment Analytical Tool

Meeting the needs of a changed Investment Clientele

Thong Bui, Sarah Kelley, Zhongqiao Jin, Natarajan Shankar

Millennials and Money Management

- Millennials - the largest potential segment of the financial market
- Millennials' investment behavior different from that of prior generations
  - Articulation of a lack of trust in standard institutions
  - Demonstrate marked gravitation towards technology
  - Seek humble fee structure, socially responsible approaches
  - Prepared to make decisions, assume risk
- To support the Millennials, financial Industry disruption is underway, driven by companies like RobinHood
  - "a stock brokerage built with the needs of a new gene mind."
  - Zero-fee trades
  - No storefront offices, research reports, analytical tools
  - Raised $176M, recently $100 Million at a $1.3B valuation
- Our business driver:
  - Offer a financial application that complements this space
  - Why us?

Data Science in the New Market

1. Data forms the basis of investment decisions
    a. Multiple forces move markets and decisions, all characterized with data:
        i. Quantitative
            1. Fundamentals: PE Ratio, Revenue Growth
            2. Economic Data: Inflation, Unemployment, Monetary
            3. Behavioral: Psychology, Social Sciences
        ii. Textual: News, Whispers, Politics
2. Data is available for consumption but is voluminous and in raw form
    a. Needs tools and expertise to leverage, for insight
        i. Apps that augment the new business model have relevance
            1. Intelligence extraction from multiple forms of data
            2. Complex models but complexity hidden from the user
            3. Provide a simple tool
3. Leading to the birth of:
    a. “Investment Analytical Tool”

Approaches

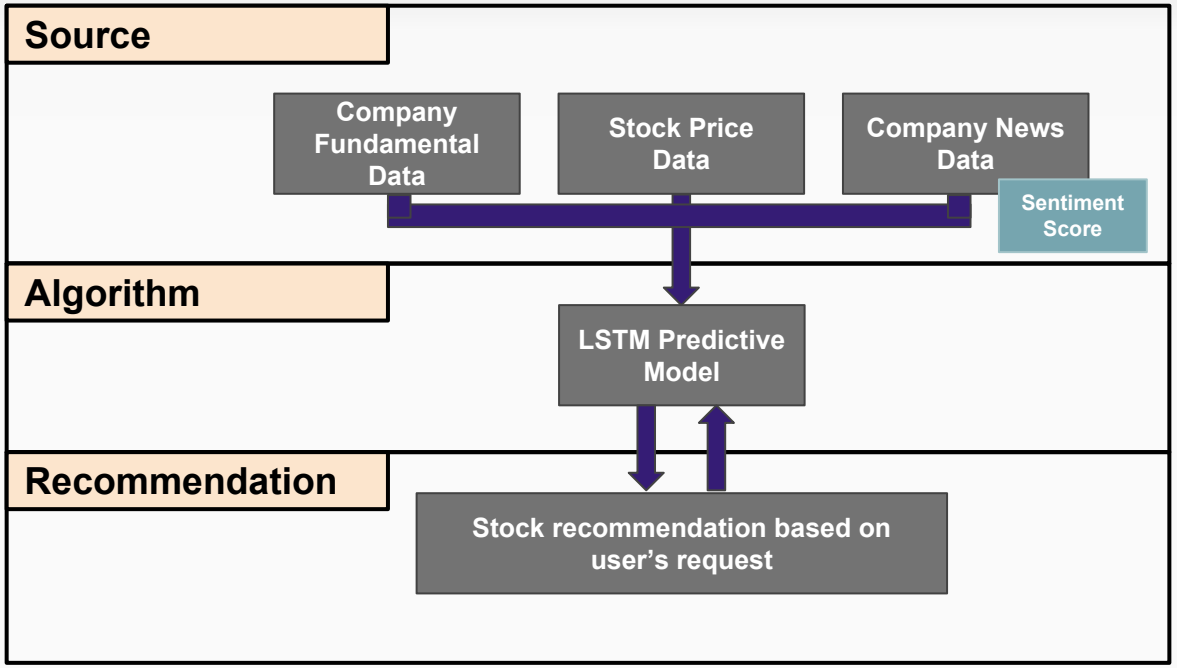
Approaches:

We are providing a web-based solution to recommend individual stocks for the investors. The solution is leveraging both time series/static numeric data as well as NLP data to predict the future stock price and recommend it to the user based on their risk aversion

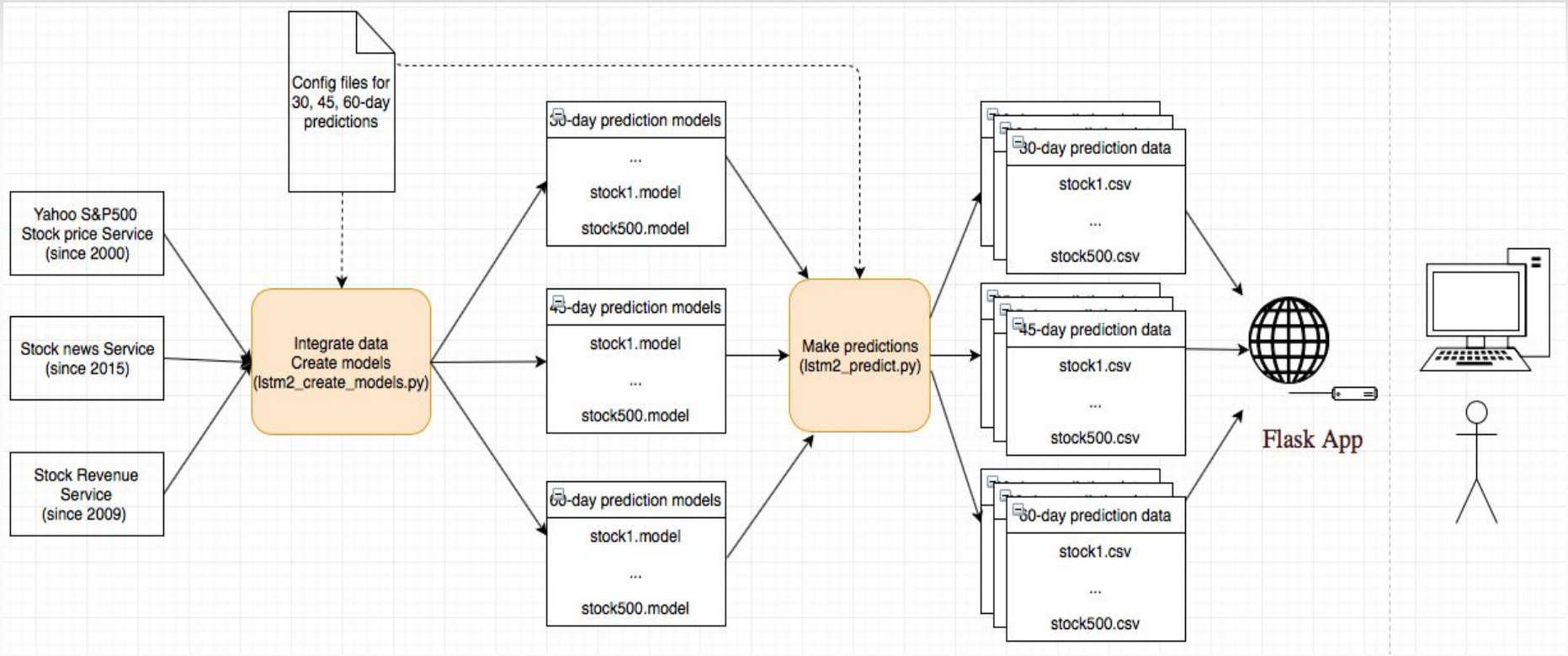
Highlights:

- Enhanced input feature coverage
- News sentiment feature inputs
- LSTM-based predictive model
- Cross validated results
- Risk exposure caveat
- Interactive UI allow user to select Risk aversion

Product Work Flow



Detailed System Architecture

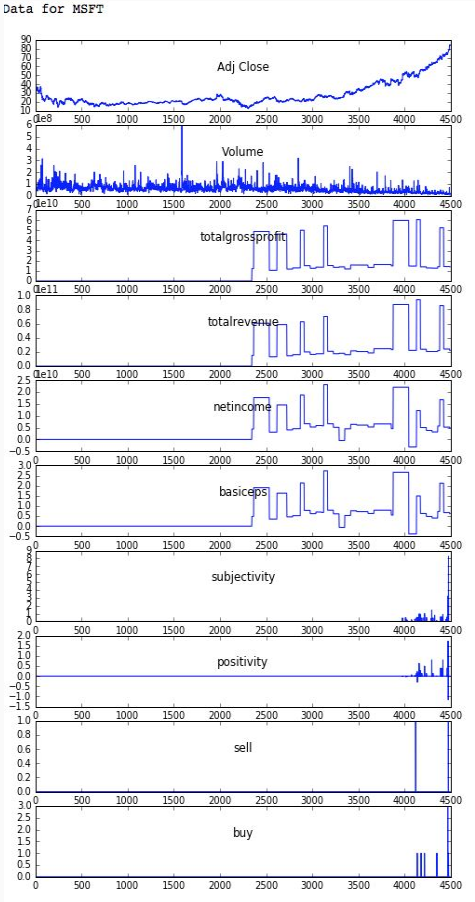


Data Analysis

Data Analysis:

- Stock data starts from 2000: Adj Close, Volume
- Revenue data starts from 2009:
  - Can we map Q1, Q2, Q3, Q4, FY to correct dates?
  - Keep totalgrossprofit, totalrevenue, netincome, basiceps
  - operatingrevenue is 0s so we decided to drop them
- News data starts from 2015
  - subjectivity, positivity, sell, buy
  - buy, sell are sparse.
- Data quality concerns

We decided to use the latest start date 2015

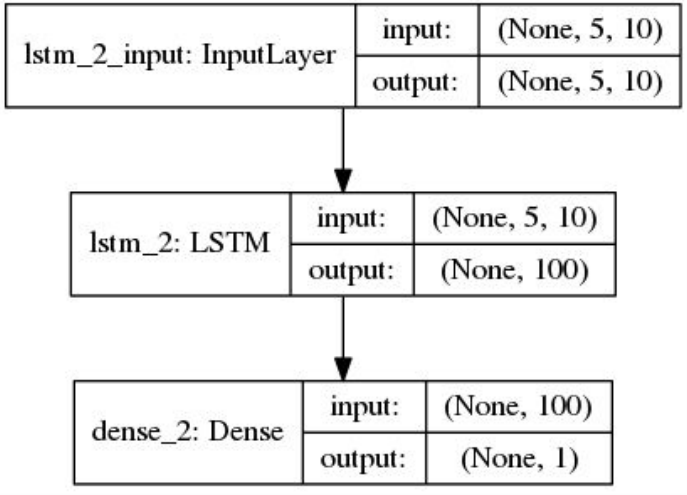


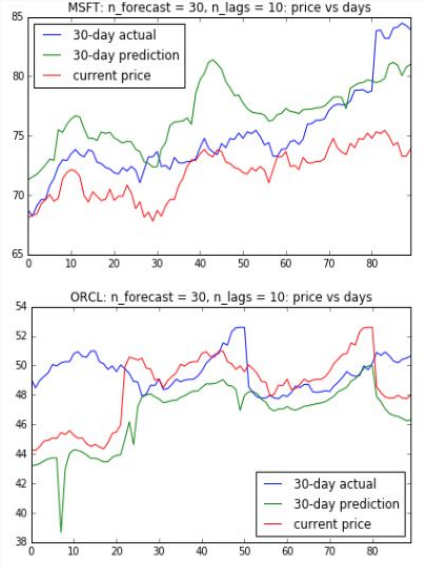
Models & Predictions

Models using time-series and LSTM

For each stock:

- Combine the 3 datasets since 2015
- Create time-series dataset to predict different long-term prices: 30, 45, 60 days
- Train data: before the last 90 days. Test data: last 90 days
- LSTM: Recurrent networks, of which LSTM is one of the most successful, are generally useful when you're dealing with a time series.
  - Very good at holding long term memories
  - Long term dependencies in the network is done by gating mechanisms
- Cross-validation against test data (last 90-day)
- Minimize overfitting: early stopping (see details)

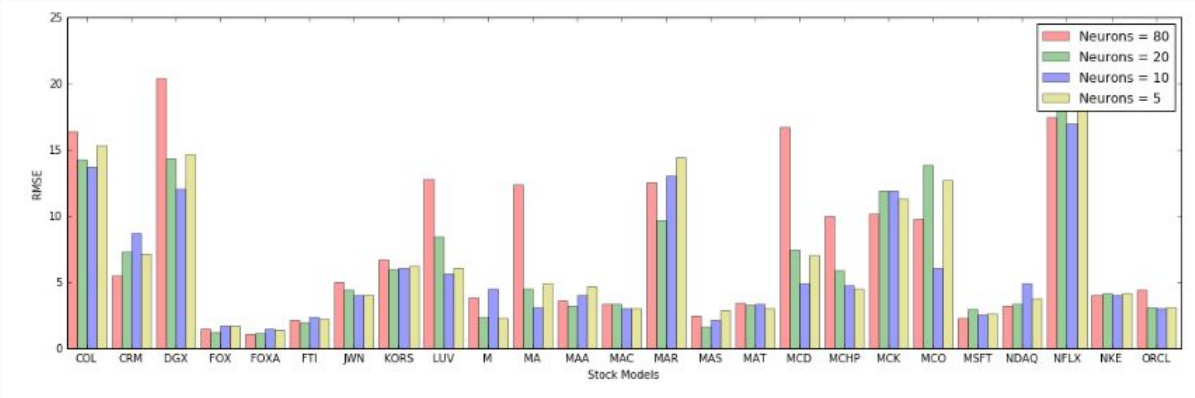


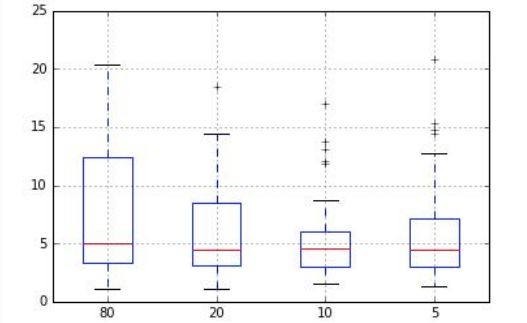


Models & Predictions (cont'd)

Parameter Tuning

We pick some parameters to tune for optimal RMSE and use them in config file for modeling
Ex: The optimal number of neurons for models to predict 30-day stock prices can be different from 45 or 60-day models (see the end of this notebook for more details)

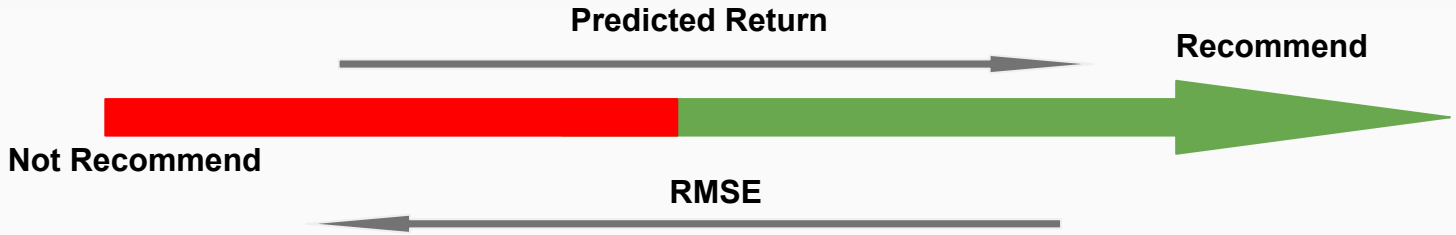




Results & Evaluation Criteria

Results:

The solution provides recommendation with a list of 10 stocks out of S&P 500 pool to users based on a combination of predicted return and the accuracy of the prediction. The solution also considered risk aversion and the expected return period.



Evaluation:

- Used RMSE (root-mean-square error) to measure the **accuracy** of prediction. RMSE is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed.
- Leveraged SD (standard deviation) to measure the **volatility** of stocks during this period - how much the prediction price is changed during this period
  - Low risk: predicted SD < 0.25 percentile
  - Medium risk: predicted SD is between [0.25, 0.75]
  - High risk: predicted SD > 0.75
- Used S&P 500 index as the benchmark to measure the historical prediction efficiency.

User Interface Architecture

- Interactive application to allow users to input risk and timeframe preferences
- Flask served over a gunicorn webserver
  - connection to Python-based backend model
  - Dynamic HTML pages

Original User Interface Design

W261: Final Project

Sarah, Shankar, Thong, Xiangjiao

Enter Preferences

How much would you like to invest?:

When would you like returns?:
- (•) 30 days
- ( ) 45 days
- ( ) 60 days

What risk profile do you want?:
- (•) Low
- ( ) Medium
- ( ) High

Get Recommendations

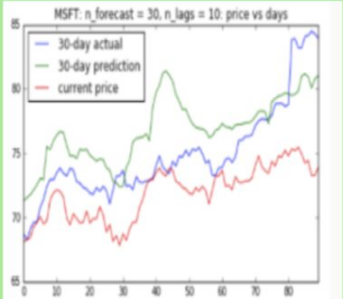
Primary Recommendation

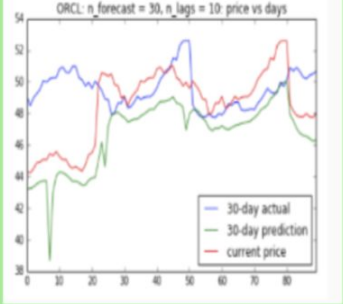
You want to invest 9000 dollars

You want returns in 60 days.

You are comfortable with a high risk level.

Predicted Performance:





Investing involves **RISKS**

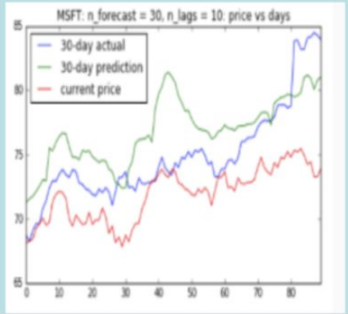
This tool is only an investment guide not an expert advisor

You could lose all your money

Invest carefully

Investment Alternative

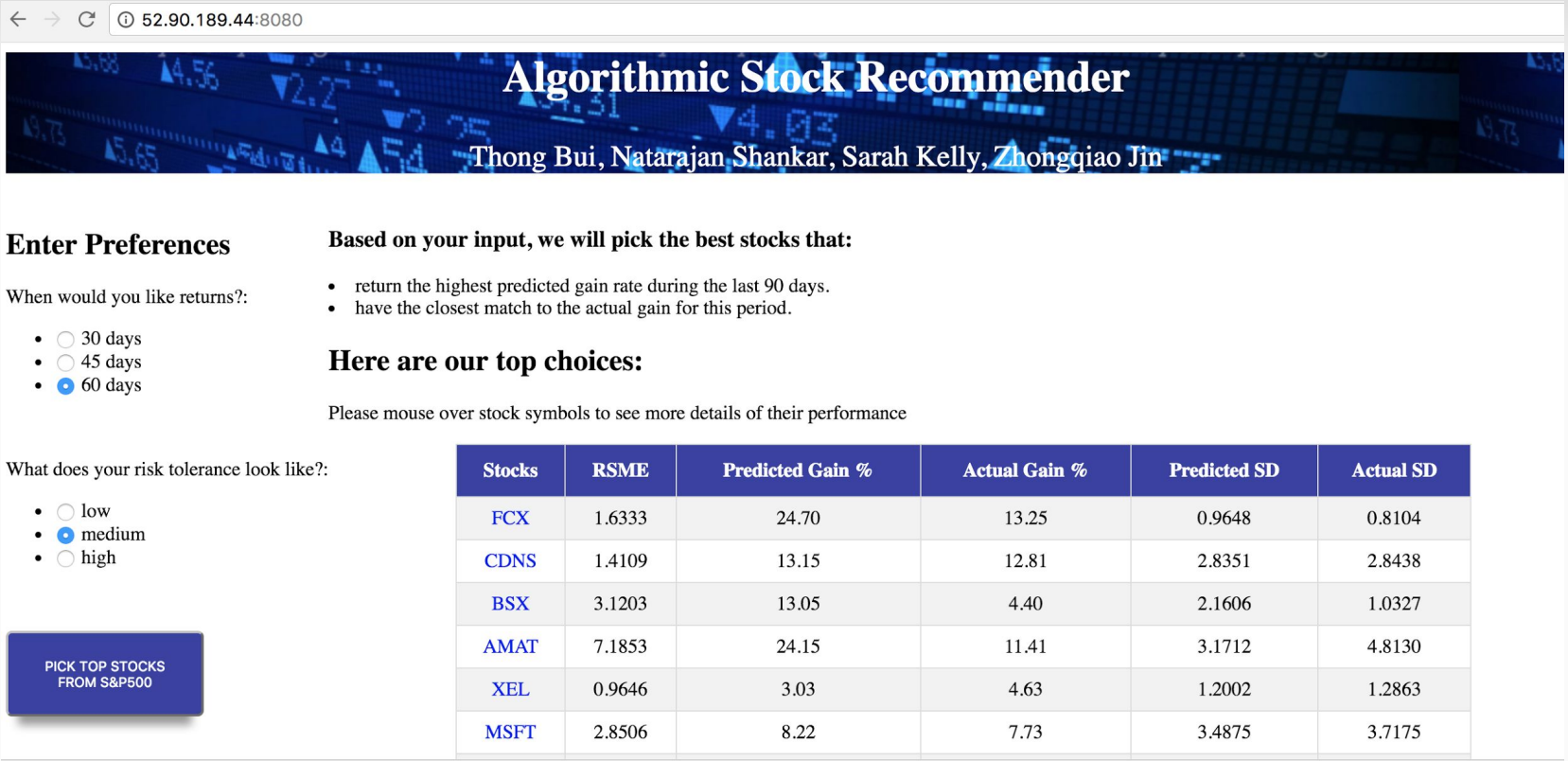
An alternative to the Primary recommendation



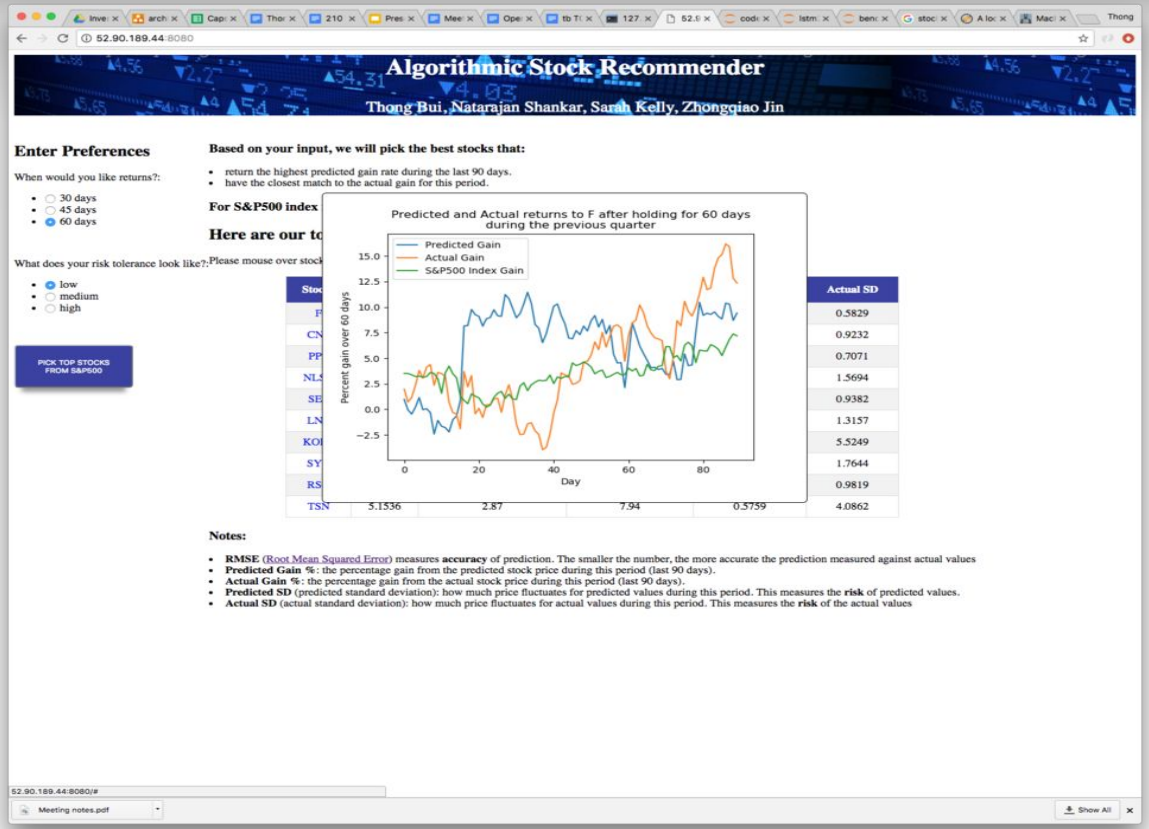
User Testing and Feedback

- User testing with 10 millennials
- Demographics:
  - 21-36; all college educated, 6 women 4 men
  - None in the financial industry or DS
- Core feedback:
  - Our interface is messy and doesn't look professional -- "honestly it looks a little like a project from a HTML 101 class"
  - Don't understand how the model works or why they should trust it
  - Found the graph extremely confusing
  - Mixed feedback on level of detail: got both people wanting more detail and feeling like there was too much data

Improved UI


52.90.189.44:8080
Algorithmic Stock Recommender
Thong Bui, Natarajan Shankar, Sarah Kelly, Zhongqiao Jin
Enter Preferences
When would you like returns?:
30 days
45 days
60 days
What does your risk tolerance look like?:
low
medium
high
PICK TOP STOCKS FROM S&P500
Based on your input, we will pick the best stocks that:
return the highest predicted gain rate during the last 90 days.
have the closest match to the actual gain for this period.
Here are our top choices:
Please mouse over stock symbols to see more details of their performance
Stocks RSME Predicted Gain % Actual Gain % Predicted SD Actual SD
FCX 1.6333 24.70 13.25 0.9648 0.8104
CDNS 1.4109 13.15 12.81 2.8351 2.8438
BSX 3.1203 13.05 4.40 2.1606 1.0327
AMAT 7.1853 24.15 11.41 3.1712 4.8130
XEL 0.9646 3.03 4.63 1.2002 1.2863
MSFT 2.8506 8.22 7.73 3.4875 3.7175

Improved UI


52.90.189.44:8080
Algorithmic Stock Recommender
Thong Bui, Natarajan Shankar, Sarah Kelly, Zhongqiao Jin
Enter Preferences
When would you like returns?:
30 days
45 days
60 days
What does your risk tolerance look like?:
low
medium
high
PICK TOP STOCKS FROM S&P500
Based on your input, we will pick the best stocks that:
return the highest predicted gain rate during the last 90 days.
have the closest match to the actual gain for this period.
For S&P500 index
Here are our to
Please mouse over stock
Predicted and Actual returns to F after holding for 60 days during the previous quarter
Predicted Gain
Actual Gain
S&P500 Index Gain
Percent gain over 60 days
Day
Actual SD
0.5829
0.9232
0.7071
1.5694
0.9382
1.3157
5.5249
1.7644
0.9819
TSN 5.1536 2.87 7.94 0.5759 4.0862
Notes:
RMSE (Root Mean Squared Error) measures accuracy of prediction. The smaller the number, the more accurate the prediction measured against actual values
Predicted Gain %: the percentage gain from the predicted stock price during this period (last 90 days).
Actual Gain %: the percentage gain from the actual stock price during this period (last 90 days).
Predicted SD (predicted standard deviation): how much price fluctuates for predicted values during this period. This measures the risk of predicted values.
Actual SD (actual standard deviation): how much price fluctuates for actual values during this period. This measures the risk of the actual values
52.90.189.44:8080/#
Meeting notes.pdf
Show All

How we communicate value

- Core question is how do we show our potential users why they should trust us
- Limitations:
  - Technical knowledge of audience
  - Balancing demonstrating value with not being overly certain
- Main tactic: written page describing at a higher level, with details on request
  - Overview of process in friendly language
  - Describe model testing/verification
  - Describe limitations
- Core feedback:
  - Reading model description test, users generally felt that they had a good basic understanding
  - “I feel like the website isn’t trying to trick me -- you are upfront about the uncertainty”
  - “Can I use this now?” (I said no)

Demo

We have webapp ready

Observations: For the last 90 days:

- Within the same time period: higher risk yields higher returns
- Longer time period yields higher returns
    - Ex: returns of 60 days > 45 days > 30 days

Next steps

What questions are still open? Where might you go from here?

- Data quality limitation:
  - From 2015
  - Pay for some financial services (expensive) to get better quality of news and quarterly reports
  - We can use data before 2008 stock crash -> better for training and predictions
- Better sentiment analysis software
- Better architecture:
  - store data in some DB
- More sophisticated UI:
  - allow user to pick a stock, input actual today’s price and other data to predict price for 30, 45, 60 days

Thank you!

Q&A

Github: https://github.com/thongnbui/MIDS_capstone

- documents: all the documents created for this project
- WebApp: flask webapp code
- code: all the back-end codes
- config: all the config files used

References:

- https://machinelearningmastery.com/