

The background of the slide is a dark blue collage of financial data. It includes various line graphs, bar charts, and numerical values such as '0.06%', '0.05%', '1344.83', '-6.89', and '1795.6'. Some text is partially obscured or faded, creating a sense of depth and complexity. The overall aesthetic is high-tech and professional, typical of a financial or data science presentation.

ML-based Investment Analytical Tool

Meeting the needs of a changed Investment Clientele

Thong Bui, Sarah Kelley, Natarajan Shankar,
Zhongqiao Jin

New Markets, Money Management, Data Science

- New investment segment has been recognized, with huge potential
 - Millennials have invested in lower numbers - the largest potential segment
 - Millennials' investment behavior different from that of prior generations
 - Robin Hood: Raised \$176M, recently \$100 Million at a \$1.3B valuation
- To support the new market, financial Industry disruption is underway
 - Zero-fee trades
 - No storefront offices, research reports, analytical tools
- The new market is this project's business driver
 - We offer a financial application that complements this new market
 - Why us?
- Data forms the basis of investment decisions
 - Quantitative and Natural
 - But is voluminous and is in raw form
 - Needs tools and expertise, for insight
- Leading to the birth of:
 - "Investment Analytical Tool": pick the best stocks from S&P500 given preferences of long-term period and risk



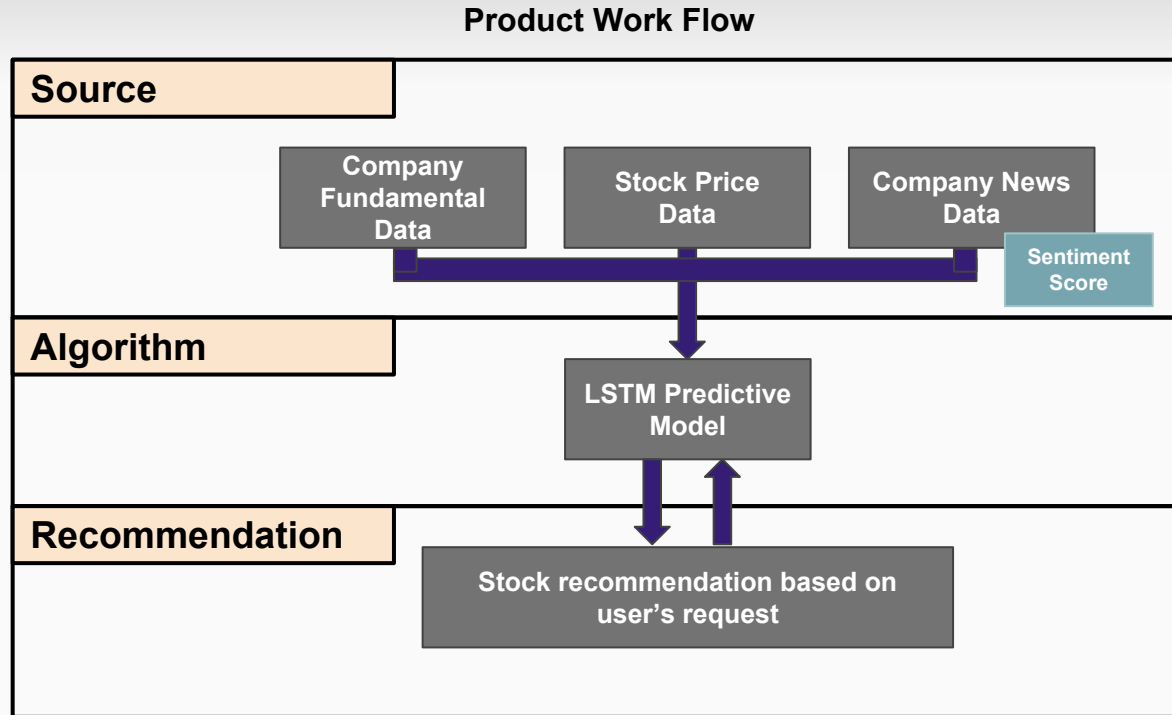
Approaches

Approaches:

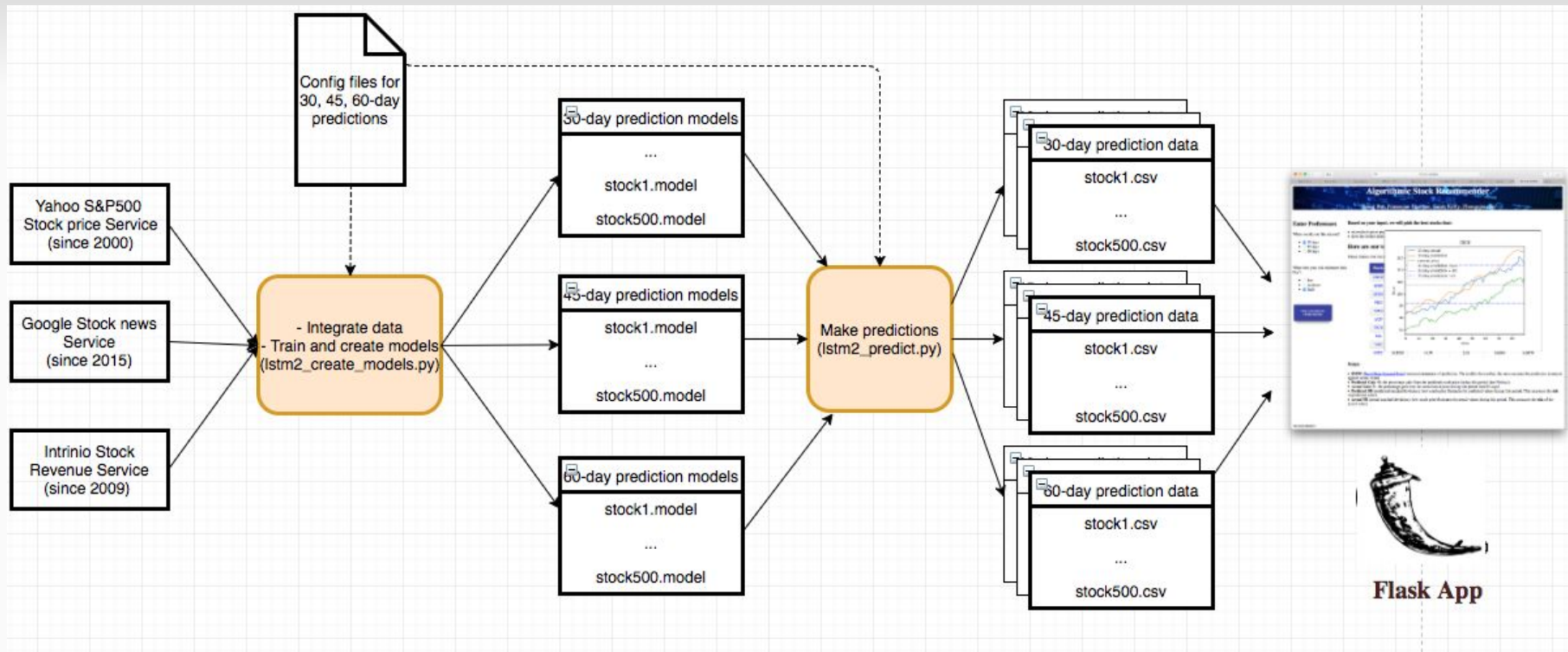
We are providing a web-based solution to recommend individual stocks for the investors. The solution is leveraging both time series/static numeric data as well as NLP data to predict the future stock price and recommend it to the user based on their risk aversion

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



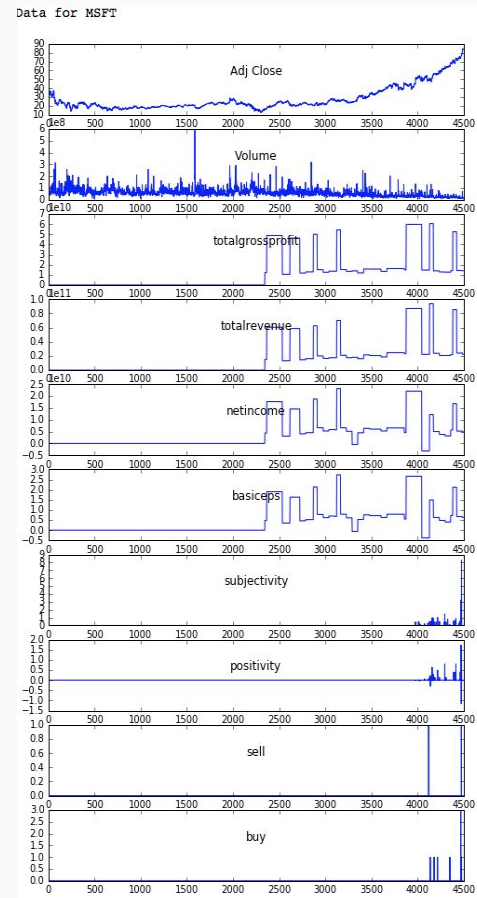
Detailed System Architecture



Data Analysis

Data Analysis:

1. Stock price data from 2000: Adj Close, Volume
2. Revenue data from 2009:
 - Use totalgrossprofit, totalrevenue, netincome, basiceps
 - Other features (operatingrevenue) are mostly 0s so we decided to drop them
3. News data starts from 2015:
 - Sentiment analysis to generate subjectivity, positivity
 - Generate buy, sell.
- Data quality concerns with free data services
- We decided to use the latest start date 2015.
- Miss the downturn (crash in 2008)

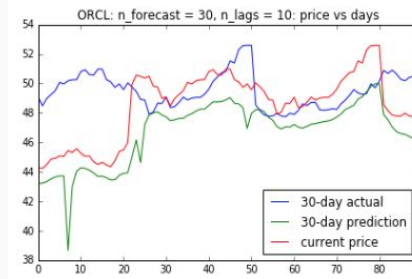
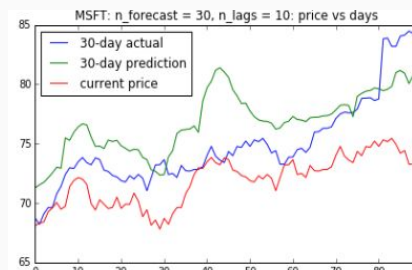
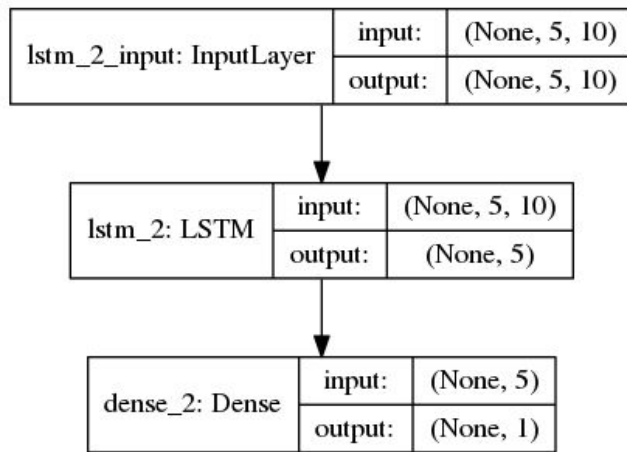


Models using time-series and LSTM

For each stock:

- Create time-series dataset and label data for the future prices of 30, 45, 60 days
- Train data: before the last 90 days. Test data: last 90 days
- LSTM: useful and successful when dealing with a time series.
 - Long term dependencies in the network is done by gating mechanisms
 - Very good at holding long term memories
- Minimize overfitting:
 - cross-validation against test data
 - early stopping (see [details](#))

Models & Predictions

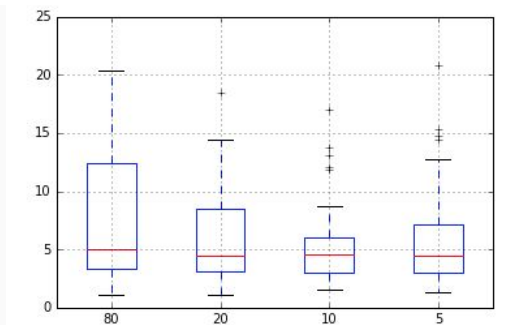
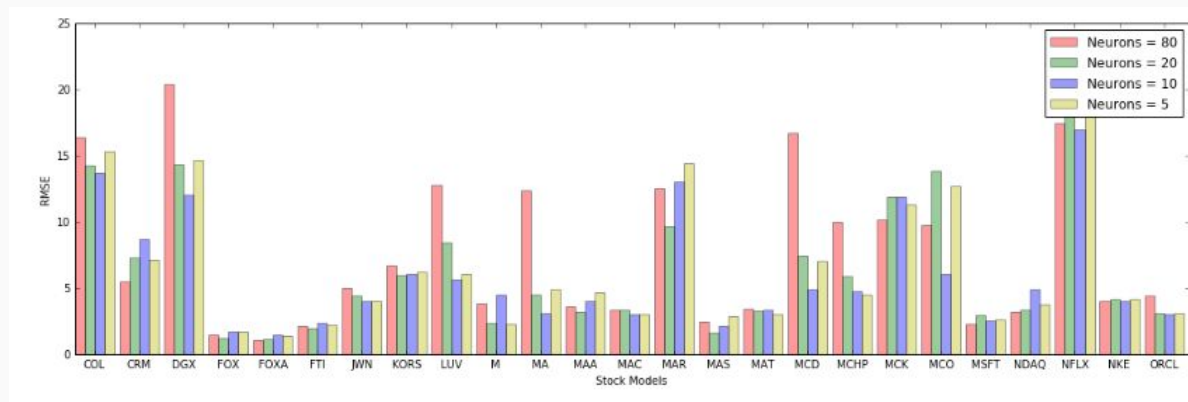


Parameter Tuning

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)

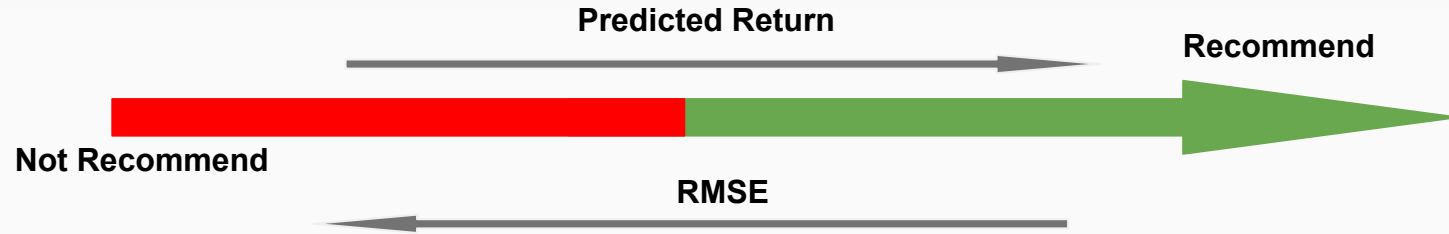
Models & Predictions (cont'd)



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 < 33.33% quantile
 - Medium risk: predicted SD is between [33.33%, 66.66%]
 - High risk: predicted SD > 66.66%
- 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

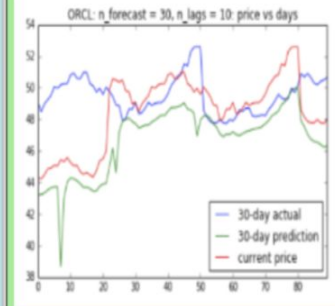
Primary Recommendation

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****

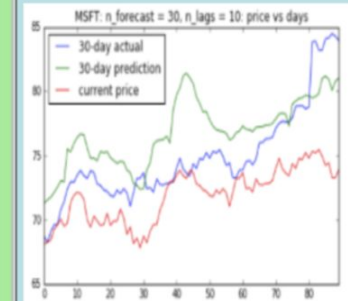
This tool is only an investment guide not an expert advisor

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:

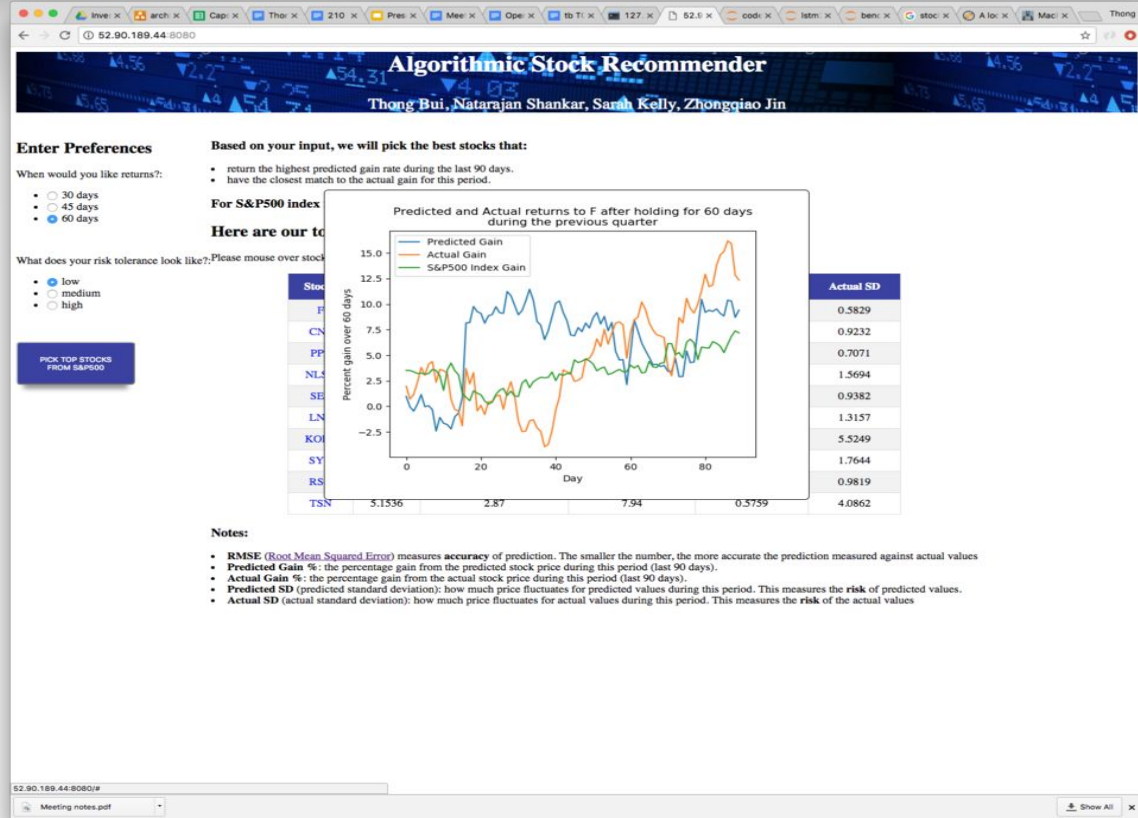
- return the highest predicted gain rate during the last 90 days.
- have the closest match to the actual gain for this period.

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



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)

- Our presented product is a prototype but it is has large potential
- It is data driven, incorporates analytical intelligence, presents a simple user interface and is intended to be open for use
- Product enhancements for functionality and robustness
 - Conduct market tests to understand usage patterns and results
 - Product internals:
 - Extend the natural language and sentiment interpretation capability
 - Extend the training data to enhance the predictive capability
 - Provide more user options
 - Provide flexible predictive time periods



Next steps



Demo

We have webapp ready: <http://52.90.189.44:8080/background>

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

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/>