Trading With Sentiment

Using Social Media & News Sentiment to Better Financial Predictions

Zinan Chen
Siyu Liu
Qiuhao Chengyong
Tyler McMurray
Senbo Zhang

Our Inspiration:
Emotions in Finance decision-making

Greed and **fear** are two main drivers of the stock market.

We believe that emotions have a big influence into the performance of stocks.

We believe bridging the gap between traditional financial data and new forms of data like NLP Techniques can improve financial predictions.

Our Goal

- Use NLP techniques on data sources to feature engineer for machine learning algorithms
  - Using Bert in Python, Using Sentiment Analysis in R
- Merge results to try to create a portfolio to showcase the effectiveness of using sentiment with more traditional predictions
  - Regression Algorithm + Sentiment
- Using Alternative Algorithms like LSTM Neural Networks & Sentiment for better Financial Predictions

Data Used

Twitter Dataset

The Twitter data provided by a Twitter Intelligence and Analytics Website: Followthehashtag.com.

Twitter information

Our dataset includes tweets with cashtag and stock code in 79 days including tweets, likes, share and number of followers

Kaggle Dataset

Our news dataset includes the 2008 to 2016 top 20 news' topics from Reddit.

Yahoo Finance

We also pulled financial information from Yahoo Finance from 2015 to 2016

Why We Believe Twitter Can Represent Public Sentiment

Facts about Twitter

Over

300 Million Users

500 Million Tweets sent per day

Tons of Financial Information

Problem: Using sentiment to predict increase/decrease of an individual stock

---

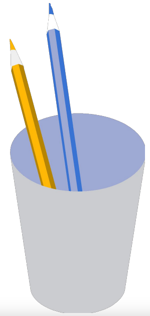
Our answer: Random Forest Model in 7-8 Days Windows Performs Best

Boosting

Accuracy

~50%

You may want to get more calculation



Random Forest

Accuracy

52%

You can try to make money on stock

Selected

Linear Regression

Accuracy

~50%

Predict stock buy or sell is hard

You need a rest

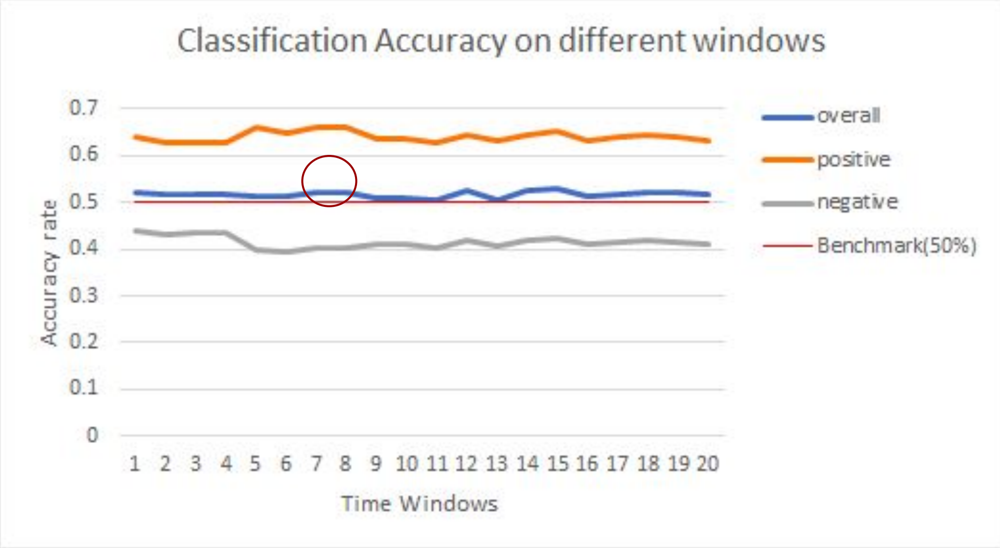
The Best Buy/Sell Classifier From Sentiment

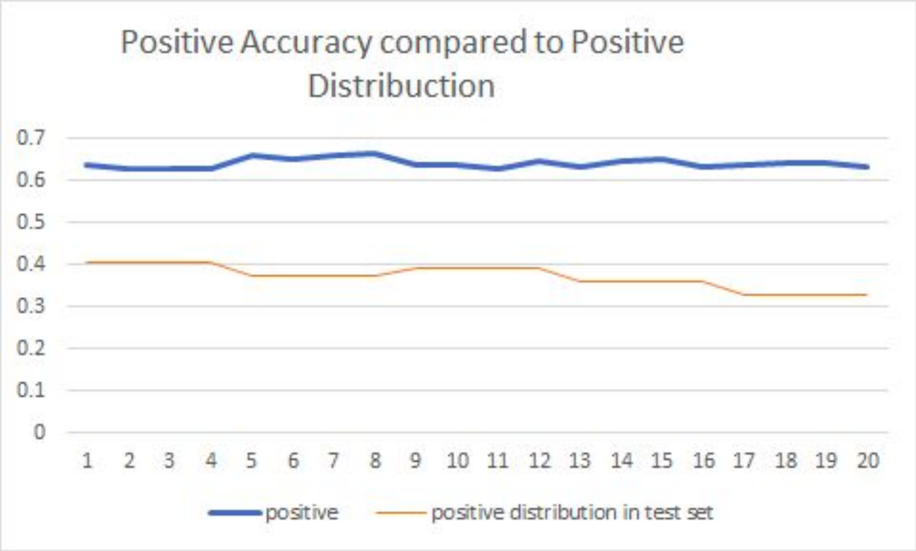
Benchmark: 50%

Unfortunately the overall accuracy is not significant from 50%

Best Windows:

7-8 Days





Perform better on 'Increase Prediction

Around 65%

Benchmark:
Original distribution of Increased stock in testset.

It performs better when predict Increase

Creating A Portfolio To Try To Validate Results

- Create a Portfolio from top predicted stocks by combing price predictions with sentiment predictions
- Try many different machine learning algorithms to predict stock prices
- Try different window sizes
  - Adjust how many days of closing stock price are included in the models
- Evaluate the portfolio by multiple aspects
  - Compare against Random Sampled Baskets
  - Compare against how all stocks did in general

Merging Price Predictions With Sentiment Classifier

Best Algorithm For Price Prediction:
- Linear Regression
- Using 8 days of data in the past

Algorithm Without Sentiment Component:
- 60% of Randomly Generated Buckets
- 30% of Random Buckets beat the average of the market

Name	Gains_Losses	Average	Beat_Winner	Beat_Aggregate
Winner	0.220	0.044	--	Yes
Basket1	-0.249	-0.050	No	No
Basket2	0.055	0.011	No	Yes
Basket3	-0.153	-0.031	No	No
Basket4	-0.018	-0.004	No	Yes
Basket5	-0.345	-0.069	No	No
Basket6	-0.472	-0.094	No	No
Basket7	-0.400	-0.080	No	No
Basket8	-0.076	-0.015	No	Yes
Basket9	-0.313	-0.063	No	No
Basket10	-0.205	-0.041	No	No
** Numbers were scaled*			0%	30%

Our Portfolio Conclusion:

- When combining traditional predictions with the sentiment aspect our winning portfolio performed very well
- 0% of Random Baskets beat our Winning Basket
- Our Winning Basket beat the Average of the market
  - The market is what we consider all other stocks used during the predictions for our winning portfolio
- NLP Techniques clearly helps better financial predictions in this scenario

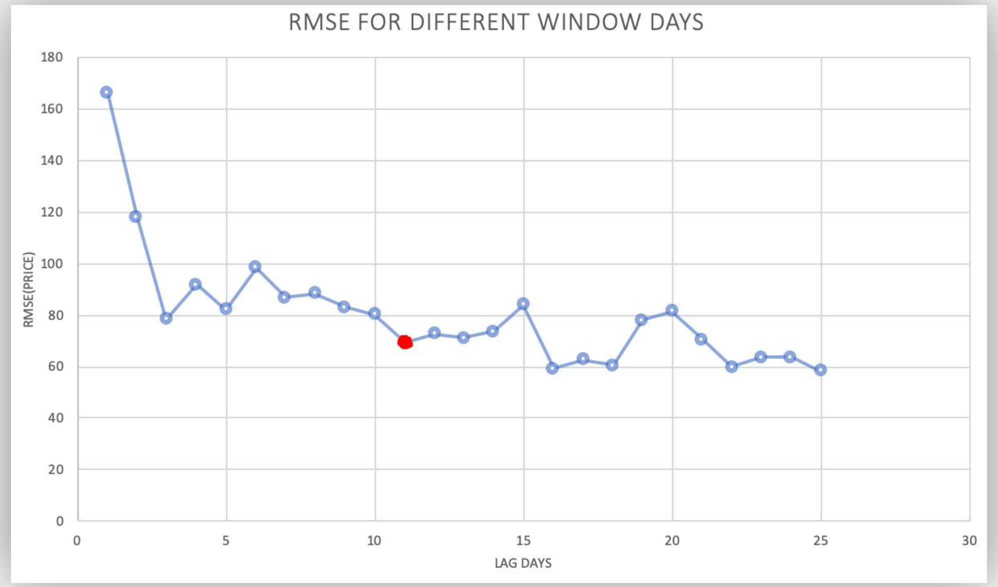
Time Series: LSTM Model

- Advance of the LSTM Model
    - ➢ Three gates help to solve vanishing gradient problem over backpropagation-through-time:
        - Forget Gate
        - Input Gate
        - Output Gate
        - Help to store important information and ignore unimportant information

LSTM Model Prediction Results

Advance Model: LSTM + Sentiment

- Two Steps Design to combine the Time series and Sentiment
    - Create basic model based on historical data
    - Add an extra feature “news sentiment” to the model
    - Check the change on accuracy - **RMSE**



- Lag between market information and its impact on stock price

  - Different Windows with Model
  - Compare the RMSE find the best lag days

Advance Model: Lower RMSE and Higher Accuracy

- Advance Model Results
  - RMSE significantly decrease from 120 to **71**
  - Predict Price fit the actual price better
  - Future direction: Add more variables to better financial prediction

The Scope of Our Project

- A ton of horizontal steps and a few big vertical steps with our project
- We kept changing data sets as more fruitful ones came about but kept trying the same concept
- Transition from trying to make a bulletproof trading strategy to a more realistic and proveable concept
- Found out about LSTM Model too late and could not scale it to the other aspect of our project
- Our Portfolio worked why aren’t we millionaires!

THANK YOU and Congratulations!!!

Q & A