

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is light green. They are positioned diagonally, with the blue one partially covering the green one.

# Stock price predictor

Yue Pan

11/17/2020



## Motivation:

Stock investment, which has become an important financial tool, is increasingly important for modern investors to pursue wealth. However, what we can't ignore is that high returns always come with high risks. Especially for the young people who just step into the stock market, their social experience and risk tolerance are relatively weak.

**So, I design a software helping investors get on the right track to some extent.**

## Goal:

The goal is to predict the trend of stock price (up or down) by collecting and analyzing key words from recent news of a user input stock symbol or company name.



## 3 parts:

### Front-end:

Take user input stock symbol or company name, check if the input is valid, translate it into stock symbol for the middle end.

### Middle-end:

Takes a stock symbol and returns a bunch of recent news that are related to the company by crawling [www.forbes.com](http://www.forbes.com), save the most relevant 20 news titles.

### Back-end:

Train a neural network model to do classification based on sentiment judgement on the 20 news titles.



# Front-end

1. Check input validity (is the input listed in NASDAQ)
2. Translate the input to a stock symbol

```
ypan@cobalt08 /data/user/ypan/bin/stockPredictor/front-end $ python front.py
Please choose enter stock symbol or company name, 1.symbol 2.company name:1
Please type a stock symbol: AMZN
ypan@cobalt08 /data/user/ypan/bin/stockPredictor/front-end $ cat frontOut.txt
AMZN    Amazon.com, Inc. Common Stock
```

frontOut.txt with key word: AMZN

# Middle-end

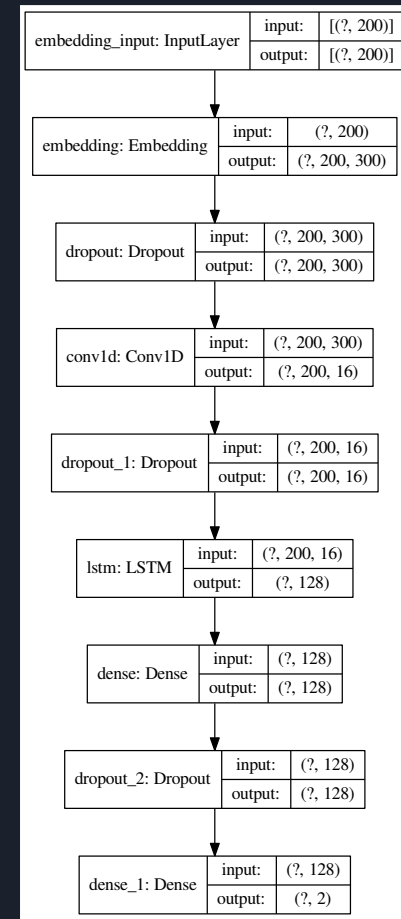
1. Use requests and beautiful soup package to build a web crawler.
2. Save the most relevant 20 news title in a temporary txt file.

```
ypan@cobalt08 /data/user/ypan/bin/stockPredictor $ cat middle-end/middleOut.txt
Magic Johnson's Latest Assist For Small Businesses: Partnering With Amazon To Spotlight Where To Shop This
Holiday Season
Wrap It Up, 2020: As Retailers Report Q3 Earnings, Eyes On The Holiday Season
They're Here! How To Avoid 5 Worst Holiday Shopping Scams
Wall Street Favoring Stay-At-Home Trade As Virus Cases Mount And Vaccine Still A Ways Away
Why Launching The K-Pop Group P1Harmony Involved Making A Film
Features Of A Fixed Index Annuity: Surrender Charges
Covid-19 Surge At Odds With Vaccine Optimism
Sears Appears To Have Given Up On The Holiday Selling Season
Even Multi-Bagger Stocks Aren't Immune To Major Drawdowns, Research Shows
Burberry Doubles Down On Digital And Localization Strategies As Growth Returns
Want Speedy Delivery? Walmart Is Revving Things Up With Self-Driving Cars
For The Stock Market, The Election Isn't Nearly Over
Payments Stocks And The Covid-19 Vaccine: A Window To 2021's Winners And Losers
1 Click To Squeeze An 8.2% Dividend From Apple
Alibaba's Singles' Day Brings In Record $74 Billion In Pandemic Year
Big Tech, Data Privacy And The Board's Role
Until The Economy Fully Recovers, Many More Companies Will Face Distress
Ford's Electric E-Transit May Hit Sweet Spot On Cost, Range, Payload For Commercial Customers
What It's Like To Open A Bookstore During The COVID-19 Pandemic
Investors Sell Tech And Buy Value On Covid-19 Vaccine Hopes
```

middleOut.txt with key word: AMZN

# Back-end

1. Replace contractions in headlines with formal terms, remove special meaningless characters and stop words.
2. Load word embeddings and map every word to its embedding and limit the length of each headline to be less than 15 and do padding if the headline is short.
3. Split the whole data set into 90% / 10% for training / testing.
4. Inspired by the architecture suggested in this paper (<https://www.aclweb.org/anthology/C16-1229.pdf>) but with some changes to fit it to the problem.
5. Train a neural network model to do classification based on sentiment judgement.



Model architecture

# Evaluation & Prediction:

```
ypan@cobalt08 /data/user/ypan/bin/stockPredictor/results $ cat evaluation.txt
Train on 1432 events, Validate on 358 events, Test on 199 events
Training loss = 1.293, Training accuracy = 0.654
Testing loss = 1.213, Testing accuracy = 0.508
      precision    recall  f1-score   support

     0       0.56       0.28       0.37        104
     1       0.49       0.76       0.60         95

 accuracy          0.51        199
 macro avg       0.52       0.52       0.48        199
 weighted avg    0.53       0.51       0.48        199

[[29 75]
 [23 72]]
```

evaluation.txt

```
ypan@cobalt08 /data/user/ypan/bin/stockPredictor/results $ cat prediction.txt
Date: 2020-11-15
The stock price will drop!
```

prediction.txt with key word: AMZN



# Future Plan

1. Combine 3 parts more naturally
2. Make the UI user-friendly
3. Generalize the model for more stock markets
4. Generalize the web crawler to more websites
5. Try different neural network architectures
6. Tune neural network parameters