Stock Price Data Processing & Prediction System

Team #1

Yuhang Zhou	yz853
Jiachen Ding	jd1287
Lichuan Ren	lr629
Haofan Zhang	hz332

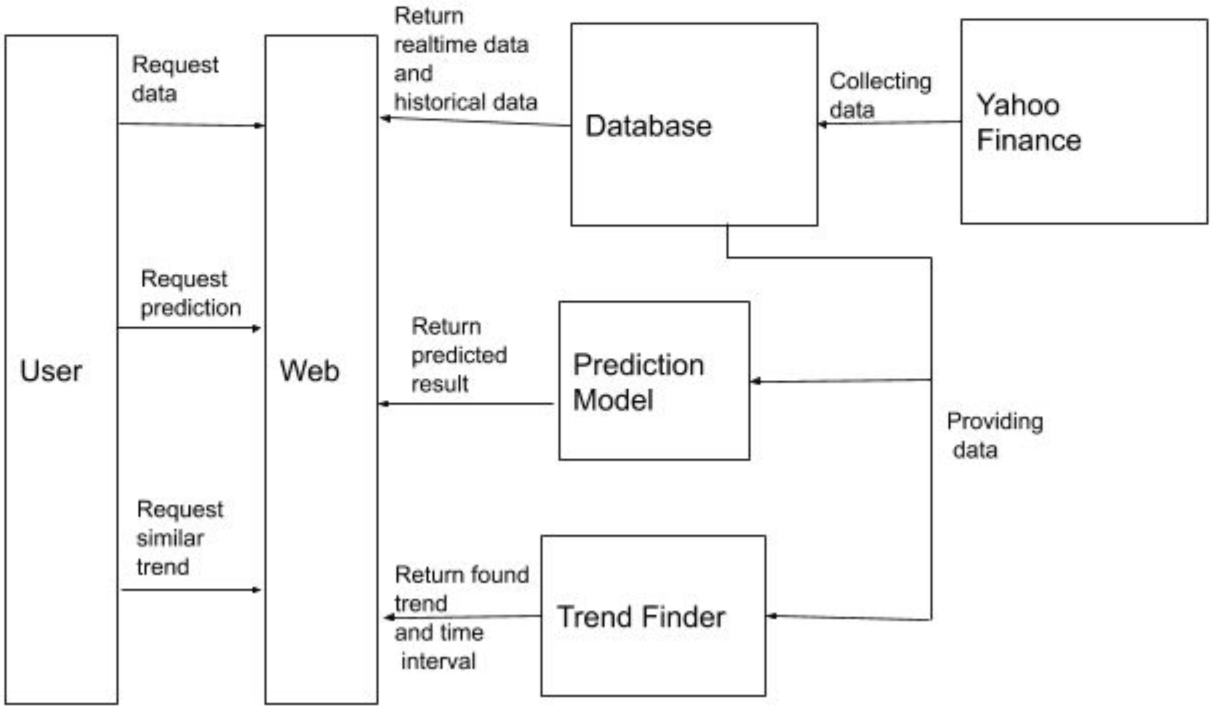
Outline

- Contribution Breakdown
- Background & Motivation
- Architecture
- Use Cases
- Programming Technology
- Web Service
- Special Feature
- Web Sources
- Prediction

Contribution breakdown

Yuhang Zhou	Web service and web front-end development
Jiachen Ding	Prediction algorithm and back-end development
Lichuan Ren	Data processing and web front-end development
Haofan Zhang	Special feature and back-end development

Architecture



Use Cases

1. Access and display the historical data or real-time data of the selected stock
2. Predict the stock price of a given stock
3. Find time intervals in the historical data with a similar trend of a given time period

Programming Technology

Language: python 3, HTML

Database: MySQL

Datatype: Numpy, Pandas,Tensor

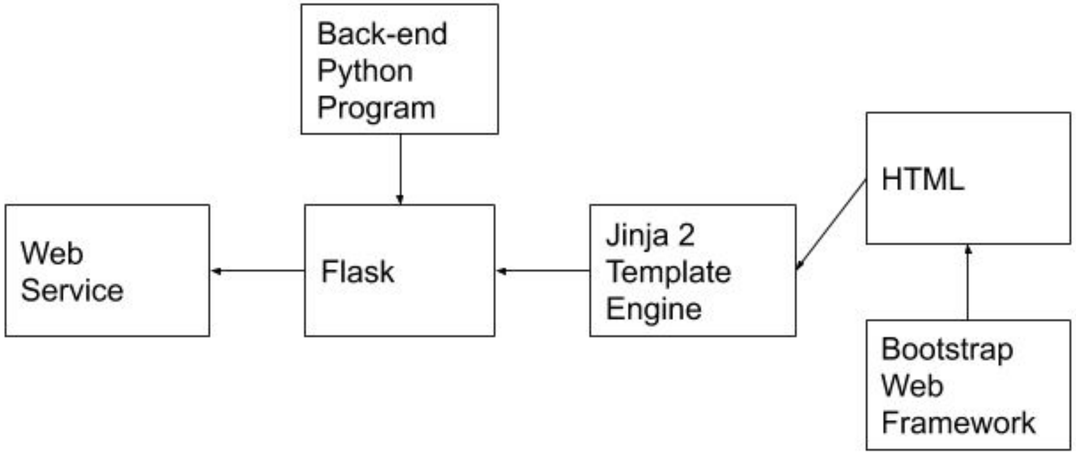
Machine learning framework: PyTorch, Keras

Data source: Yahoo Finance

Web Service

Web API: RESTful

Framework: Flask of python

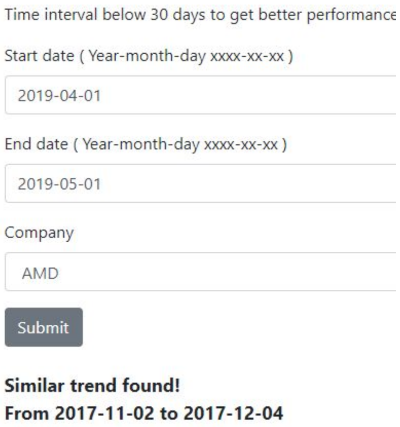


Special Feature - Trend Finder

Using Pearson correlation coefficient

to find similar trend of a stock

$$\rho_{X,Y} = \frac{\mathrm{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$



Time interval below 30 days to get better performance

Start date (Year-month-day xxxx-xx-xx)

2019-04-01

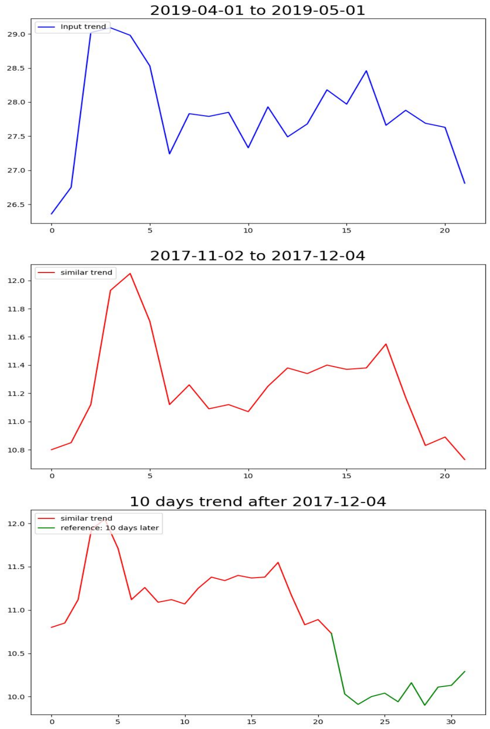
End date (Year-month-day xxxx-xx-xx)

2019-05-01

Company

AMD

Similar trend found!
From 2017-11-02 to 2017-12-04



2019-04-01 to 2019-05-01

2017-11-02 to 2017-12-04

10 days trend after 2017-12-04

0.8-1.0 Extremely strong correlation

0.6-0.8 Strong correlation

0.4-0.6 Moderate correlation

0.2-0.4 Weak correlation

0.0-0.2 Very weakly Very weakly or unrelated

Web Sources

Stock price data source from Yahoo Finance

Some prediction and trend finding methods

Picking up stock indicators based on finance researches

Prediction

SVM: Using all historical data as training data

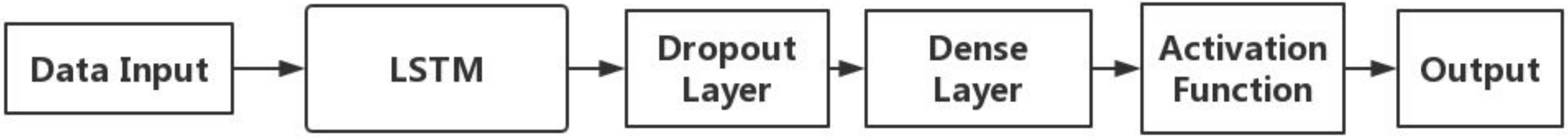
Bayesian Curve Fitting: Using the recent data as training data

LSTM: Using all historical data to train the model

Key features of LSTM: able to remember **ancient information**, the neurons can learn the temporary association between neighbor data chip, can keep long-term trend

Indicators: Close Price, Volume and Moving Average

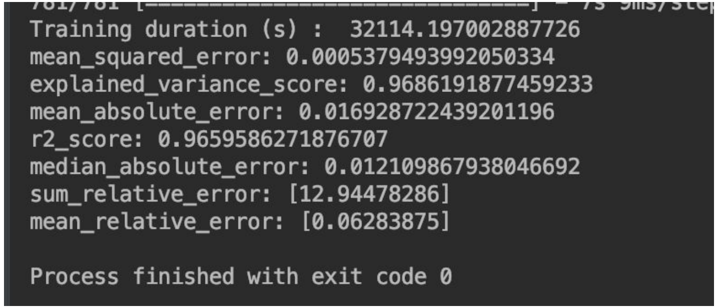
LSTM Model



- 2 LSTM-RNN
- 2 dropout layers in case of overfitting
- A dense layer to make desicion
- Data Normalization for data preprocessing

LSTM Model

- Activation function: Adam
- Loss function: Mean Squared Error
- Training/testing split: daily price of past 3 years; 4/1
- 500 epochs of training
- Specific weights for specific stocks
- Metrics:



```
781/781 [==============================] - 7s 9ms/step
Training duration (s) :  32114.197002887726
mean_squared_error: 0.0005379493992050334
explained_variance_score: 0.9686191877459233
mean_absolute_error: 0.016928722439201196
r2_score: 0.9659586271876707
median_absolute_error: 0.012109867938046692
sum_relative_error: [12.94478286]
mean_relative_error: [0.06283875]

Process finished with exit code 0
```

Demo

Background & Motivation

Background:

Nowadays, more than 50% of the US households own stocks. There are over 600 securities exchanges around the world.

Large trading organizations can employ sophisticated computer systems and armies of analysts.

Motivation:

To help the individual investor make better investment decisions.