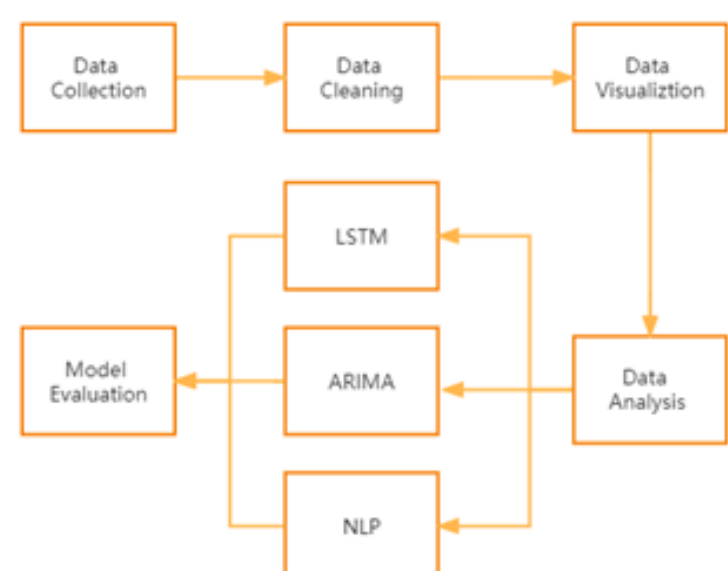# Analysis and Improvements of Stock Price Prediction Algorithm

He Xianrui, Liu Hongxin, Luo Jiehui, Xue Qianyi, Zhou Yubo

Introduction

Comments from the stock forum can reflect the price movement of stocks. Based on this idea, we propose to combine NLP and classical algorithms to find out whether NLP makes predictions more precisely.

Expriment

Data Colletion&Preparation

We use API and Web Mining to obtain historical stock quotes and relevant comments about Apple Corporation, respectively.

Data Exploration&Visualization

We need to observe the intrinsic nature of the data and visualize its changing trends

Data Analysis (LSTM)

Fig1. Prediction Results of LSTM

Data Analysis (ARIMA)

Fig2. Prediction Results of ARIMA

Sentiment analysis and stock price prediction

Apple Stock Comments Preprocessing
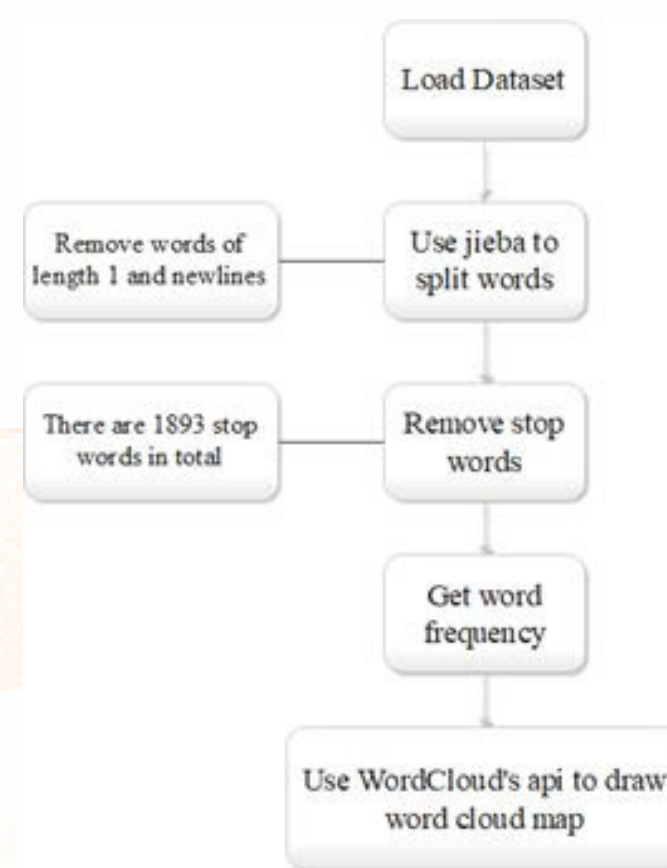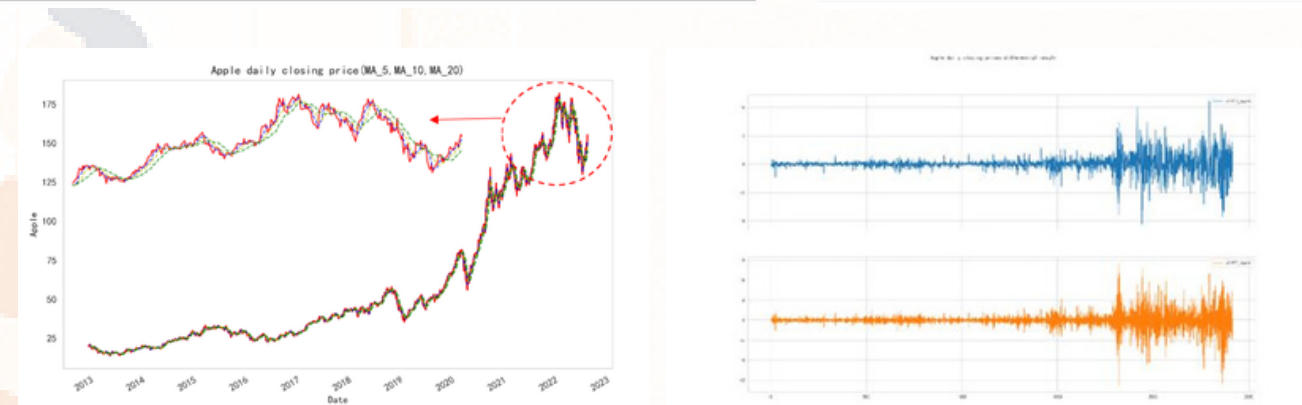

Fig3. Flow Chart of Text Preprocessing

Fig4. Apple style Word Cloud illustration

There are 907 days of relevant news or comments, among which the words that appear more often are "股票", "投资", "下跌", "上涨" and so on.

Sentiment Analysis

Fig5. Flow Chart for Calculating the Score

Table1. Input variables and interpretation of the ARIMAX model.

Symbol	Definition
Close	Daily closing price of Apple stock.
Adj_Close	Adjusted daily closing price of Apple stock.
Volume	Apple Stock Daily Volume.
Emotion	Daily Apple stock sentiment index.
Change	The difference between the closing prices of the stock for the previous 2 days.
Close_MA_5	Average of closing prices for the previous 5 days.
Close_MA_10	Average of closing prices for the previous 10 days.
Close_MA_20	Average of closing prices for the previous 20 days.

Table2. ADF test results

Test Statistic	0.494763
p-value	0.984713
Lags Used	27
Number of Observations Used	2387
Critical Value (1%)	-3.433092
Critical Value (5%)	-2.862752
Critical Value (10%)	-2.567415

Fig6. Unstationary series becomes stationary series (difference operation)

MSE	3.607
RMSE	1.899
R^2	0.996

Performance of ARIMAX(1,1,1) on the test dataset.

Conclusion

The text information of comments can also be used as an important factor in the prediction algorithm, which helps predict more accurately.