

TRADING ACTION CLASSIFICATION

THU PHAM
FALL 2019


1. Introduction
2. Progress Report
3. Data Overview
4. Methodology



INTRODUCTION

- ▣ Utilize historical data, financial indicators, and sentimental analysis to apply machine learning methods.
- ▣ Forecast the movement of stock in a daily basis.
- ▣ Focus on trading action (long/short position).

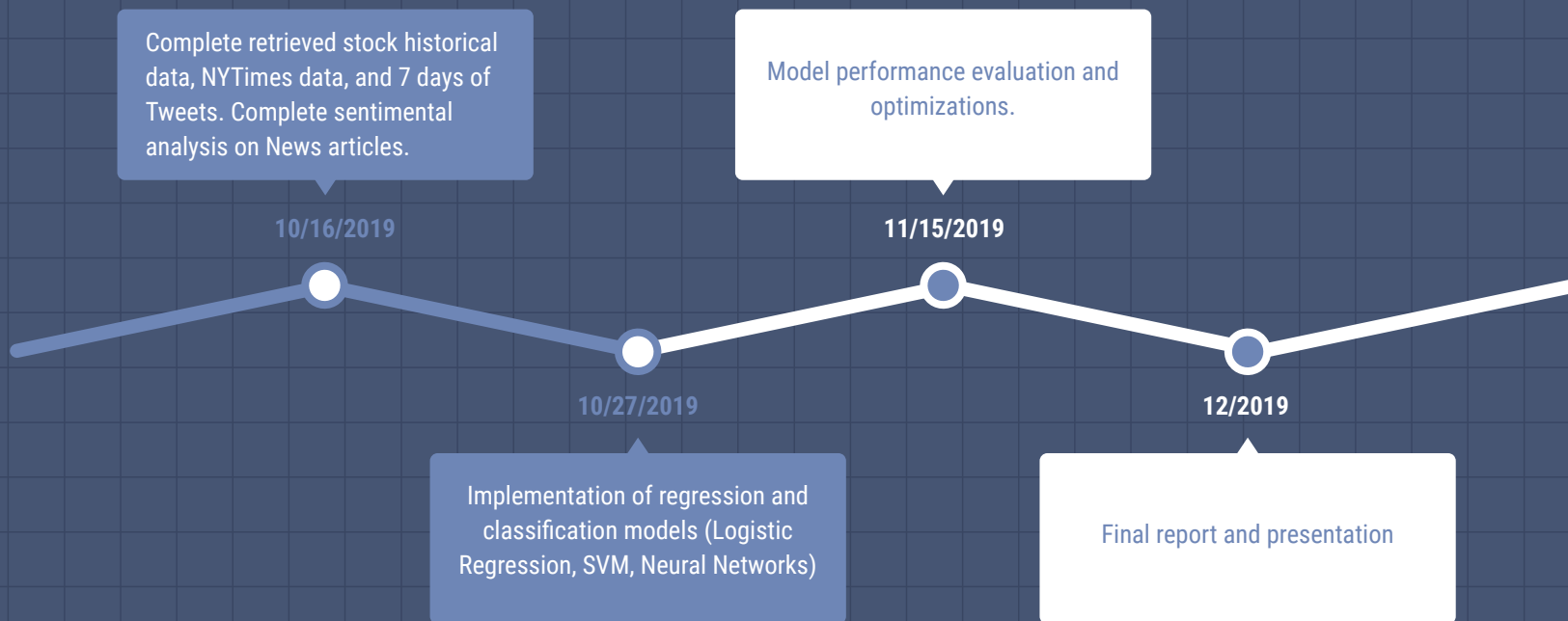




“October: this is one of the peculiarly dangerous months to speculate in stocks. The others are July, January, September, April, November, May, March, June, December, August and February.”

Mark Twain

PROGRESS REPORT



DATA OVERVIEW

Historical Stock Price

- Alpha Vantage API
- Daily
- Since 2000

	timestamp	open	high	low	close	volume
timestamp						
2019-10-15	2019-10-15	236.39	237.64	234.88	235.32	19012889
2019-10-14	2019-10-14	234.90	238.13	234.67	235.87	24106900
2019-10-11	2019-10-11	232.95	237.64	232.31	236.21	41698900
2019-10-10	2019-10-10	227.93	230.44	227.30	230.09	28253400
2019-10-09	2019-10-09	227.03	227.79	225.64	227.03	18692600
2019-10-08	2019-10-08	225.82	228.06	224.33	224.40	27955000
2019-10-07	2019-10-07	226.27	229.93	225.84	227.06	30576500
2019-10-04	2019-10-04	225.64	227.49	223.89	227.01	34619700
2019-10-03	2019-10-03	218.43	220.96	215.13	220.82	28606500
2019-10-02	2019-10-02	223.06	223.58	217.93	218.96	34612300

DATA OVERVIEW

Convert closing price to classification data

- Long stock (1) if the price increase by 2.5%.
- Short stock (0) if otherwise.

	timestamp	open	high	low	close	volume	prev_close	action
timestamp								
2019-10-15	2019-10-15	236.39	237.64	234.88	235.32	19012889	235.87	0.0
2019-10-14	2019-10-14	234.90	238.13	234.67	235.87	24106900	236.21	0.0
2019-10-11	2019-10-11	232.95	237.64	232.31	236.21	41698900	230.09	1.0
2019-10-10	2019-10-10	227.93	230.44	227.30	230.09	28253400	227.03	0.0
2019-10-09	2019-10-09	227.03	227.79	225.64	227.03	18692600	224.40	0.0
2019-10-08	2019-10-08	225.82	228.06	224.33	224.40	27955000	227.06	0.0
2019-10-07	2019-10-07	226.27	229.93	225.84	227.06	30576500	227.01	0.0
2019-10-04	2019-10-04	225.64	227.49	223.89	227.01	34619700	220.82	1.0
2019-10-03	2019-10-03	218.43	220.96	215.13	220.82	28606500	218.96	0.0
2019-10-02	2019-10-02	223.06	223.58	217.93	218.96	34612300	224.59	0.0

DATA OVERVIEW

New York Times API

- Headlines and articles archive from the past 20 years
- Merge to the stock historical data frame
- Sentimental analysis

	timestamp	open	high	low	close	volume	prev_close	action	neg	neu	pos
2000-01-03	2000-01-03	104.8750	112.5000	101.6880	111.938	133949200	102.813	1.0	0.051	0.871	0.078
2000-01-04	2000-01-04	108.2500	110.6250	101.1880	102.500	128094400	111.938	0.0	0.056	0.904	0.039
2000-01-05	2000-01-05	103.7500	110.5630	103.0000	104.000	194580400	102.500	0.0	0.093	0.828	0.079
2000-01-06	2000-01-06	106.1183	107.0000	95.0000	95.000	191993200	104.000	0.0	0.079	0.835	0.086
2000-01-07	2000-01-07	96.5000	101.0000	95.5000	99.500	115183600	95.000	1.0	0.072	0.838	0.090
2000-01-10	2000-01-10	102.0000	102.2500	94.7500	97.750	126266000	99.500	0.0	0.081	0.850	0.068
2000-01-11	2000-01-11	95.9380	99.3750	90.5000	92.750	110387200	97.750	0.0	0.086	0.846	0.069
2000-01-12	2000-01-12	95.0000	95.5012	86.5000	87.188	244017200	92.750	0.0	0.115	0.789	0.096
2000-01-13	2000-01-13	94.4840	98.7500	92.5000	96.750	258171200	87.188	1.0	0.097	0.818	0.085
2000-01-14	2000-01-14	100.0000	102.2500	99.3750	100.438	97594000	96.750	1.0	0.097	0.832	0.071

2,131,988

headlines



DATA OVERVIEW

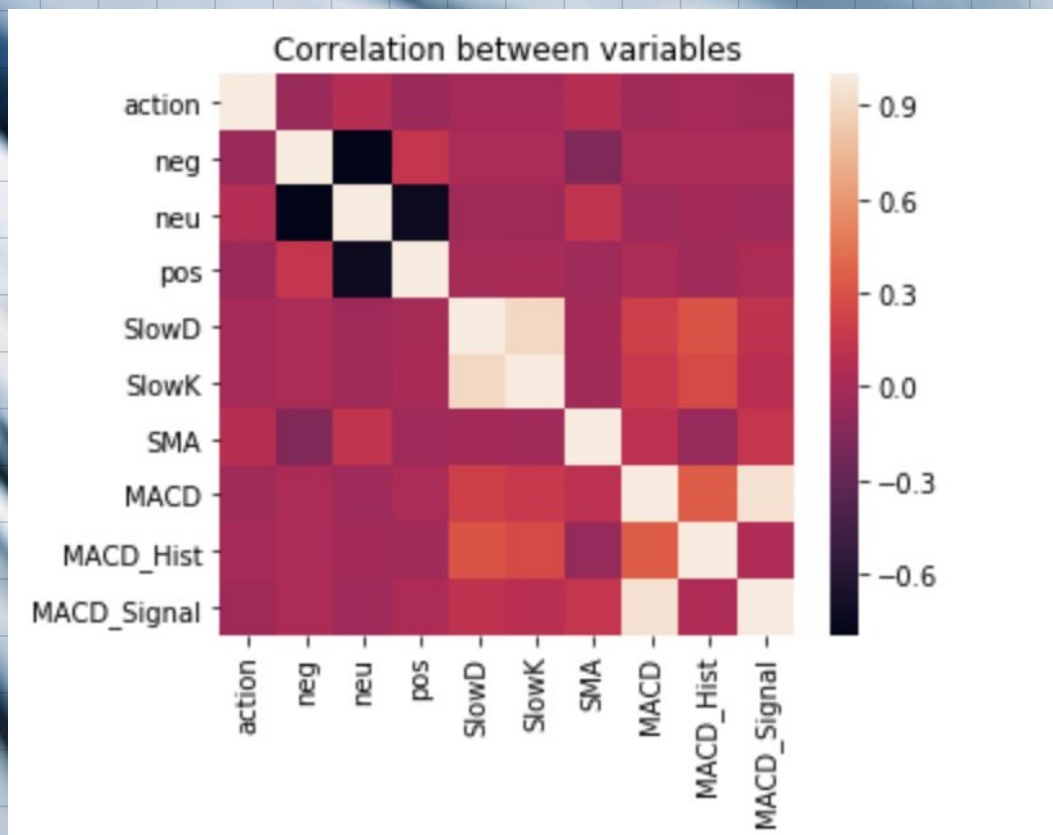
Twitter Static Sentimental Analysis from Kaggle

- AAPL
- 2016-2019
- Daily

date	ts_polarity	twitter_volume
1/1/16	0.11969256	417
1/2/16	0.14077416	495
1/3/16	0.18113164	518
1/4/16	0.07038878	1133
1/5/16	0.13363479	1430
1/6/16	0.07204194	1949
1/7/16	0.07436948	2289
1/8/16	0.05159477	2235
1/9/16	0.03234206	892
1/10/16	0.14592163	625
1/11/16	0.01944325	1222
1/12/16	0.12136357	1293

DATA OVERVIEW

	timestamp	open	high	low	close	volume	prev_close	action	neg	neu	pos	SlowD	SlowK	SMA	MACD	MACD_Hist	MA
0	2000-01-03	104.8750	112.500	101.688	111.938	133949200	102.813	1.0	0.051	0.871	0.078	86.9150	83.4605	227.842	4.9364	0.9771	
1	2000-01-04	108.2500	110.625	101.188	102.500	128094400	111.938	0.0	0.056	0.904	0.039	85.5056	89.1786	226.710	4.5060	0.7911	
2	2000-01-05	103.7500	110.563	103.000	104.000	194580400	102.500	0.0	0.093	0.828	0.079	80.6293	88.1060	225.310	4.0429	0.5257	
3	2000-01-06	106.1183	107.000	95.000	95.000	191993200	104.000	0.0	0.079	0.835	0.086	77.2606	79.2323	224.069	3.5916	0.2058	
4	2000-01-07	96.5000	101.000	95.500	99.500	115183600	95.000	1.0	0.072	0.838	0.090	74.7201	74.5496	223.276	3.4743	0.1400	



METHODOLOGY



Logistic Regression

SVM

Sequential Kernel

- Implement separately the complete data set with NY Times sentimental analysis itself vs. the dataset with added Twitter data for limiting year.
- Compare the result.

THANKS!

Any questions?

Find me at:

<https://github.com/thu2pham/tradingclassification>

