Stock Sentiment Analysis

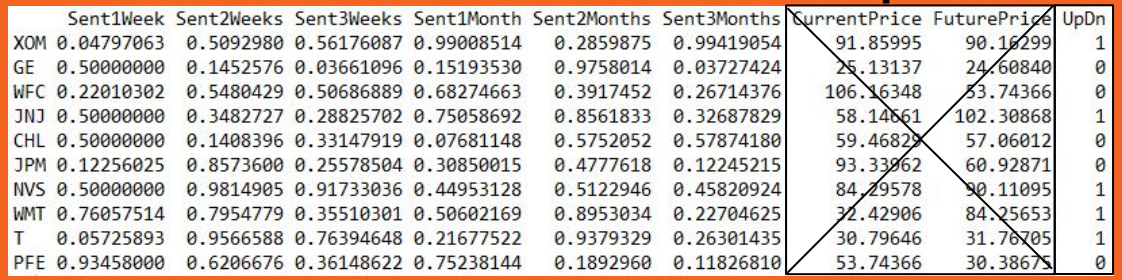
Mitchell Wade, Reazul Hoque, Curtis Wilkerson, Miller Moore

Overview and Objective

- Scrape historical tweets for 50 largest stock on the NYSE (i.e., BMY, C, XOM, CVS, etc)
- Score sentiment for tweets by stock and date
- Use historically aggregated sentiment scores to predict future price movement
- Modelling choice will be random forest in classification mode
- Response variable - price difference over next time period (e.g., monthly)
    - “1” if price is Up and “0” if price is down

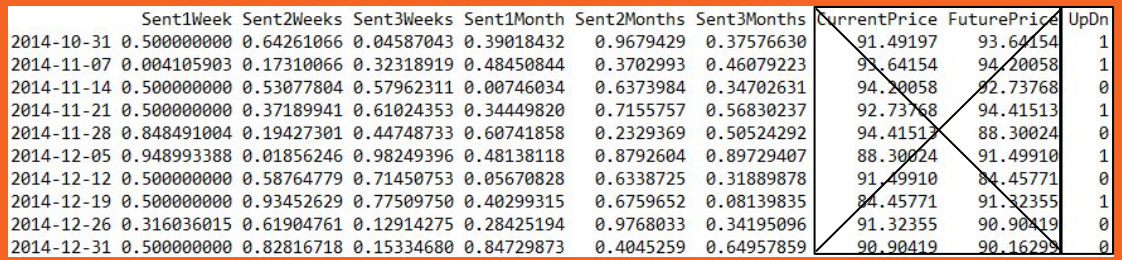
Possible Approaches

- Two approaches:
  - (1) Model all stocks at once (treat as independent observations)



	Sent1Week	Sent2Weeks	Sent3Weeks	Sent1Month	Sent2Months	Sent3Months	CurrentPrice	FuturePrice	UpDn
XOM	0.04797063	0.5092980	0.56176087	0.99008514	0.2859875	0.99419054	91.85995	90.16299	1
GE	0.50000000	0.1452576	0.03661096	0.15193530	0.9758014	0.03727424	25.13137	24.60840	0
WFC	0.22010302	0.5480429	0.50686889	0.68274663	0.3917452	0.26714376	106.16348	53.74366	0
JNJ	0.50000000	0.3482727	0.28825702	0.75058692	0.8561833	0.32687829	58.14661	102.30868	1
CHL	0.50000000	0.1408396	0.33147919	0.07681148	0.5752052	0.57874180	59.46829	57.06012	0
JPM	0.12256025	0.8573600	0.25578504	0.30850015	0.4777618	0.12245215	93.33962	60.92871	0
NVS	0.50000000	0.9814905	0.91733036	0.44953128	0.5122946	0.45820924	84.29578	90.11095	1
WMT	0.76057514	0.7954779	0.35510301	0.50602169	0.8953034	0.22704625	32.42906	84.25653	1
T	0.05725893	0.9566588	0.76394648	0.21677522	0.9379329	0.26301435	30.79646	31.76705	1
PFE	0.93458000	0.6206676	0.36148622	0.75238144	0.1892960	0.11826810	53.74366	30.38675	0

  - (2) Model each stock alone (e.g., XOM) where dates are observations



	Sent1Week	Sent2Weeks	Sent3Weeks	Sent1Month	Sent2Months	Sent3Months	CurrentPrice	FuturePrice	UpDn
2014-10-31	0.500000000	0.64261066	0.04587043	0.39018432	0.9679429	0.37576630	91.49197	93.64154	1
2014-11-07	0.004105903	0.17310066	0.32318919	0.48450844	0.3702993	0.46079223	93.64154	94.20058	1
2014-11-14	0.500000000	0.53077804	0.57962311	0.00746034	0.6373984	0.34702631	94.20058	92.73768	0
2014-11-21	0.500000000	0.37189941	0.61024353	0.34449820	0.7155757	0.56830237	92.73768	94.41513	1
2014-11-28	0.848491004	0.19427301	0.44748733	0.60741858	0.2329369	0.50524292	94.41513	88.30024	0
2014-12-05	0.948993388	0.01856246	0.98249396	0.48138118	0.8792604	0.89729407	88.30024	91.49910	1
2014-12-12	0.500000000	0.58764779	0.71450753	0.05670828	0.6338725	0.31889878	91.49910	84.45771	0
2014-12-19	0.500000000	0.93452629	0.77509750	0.40299315	0.6759652	0.08139835	84.45771	91.32355	1
2014-12-26	0.316036015	0.61904761	0.12914275	0.28425194	0.9768033	0.34195096	91.32355	90.90419	0
2014-12-31	0.500000000	0.82816718	0.15334680	0.84729873	0.4045259	0.64957859	90.90419	90.16299	0

Workplan

- Lead text analyzers - Mitchell and Reazul
- Lead modellers - Curtis and Miller

- 10/23/15 - Acquire and score as many tweets per stock as possible
- 10/30/15 - Finish data cleaning and matrix construction
- 11/6/15 - Formulate and build initial model
- 11/13/15 - Tune and validate model
- 11/20/15 - Test model and draw conclusions (i.e., answer if effective)
- > 11/20/15 - Present on results and conclusions