Text driven stock market prediction based on deep learning methods

TechTive

Agenda

1. Introduction

1. **Mood can be extracted**: Many studies have shown that social mood can influence investment behavior (such as Theory of Social Mood [Nofsinger, 2005]). Thus sentiment from news can be helpful for financial market prediction, such as stock market [Tetlock et al, 2008].

2. **AI is powerful**: With development of AI techniques (such as deep learning), stock market prediction based on text mining from social media become more and more feasible, such as [Bollen et al, 2011].

3. **Less is More**: Sentiment extraction from news headline is a popular and well-built method [Strapparava & Mihalcea, 2008] with less noise.

Business Problem identification:

How to extract useful information from news and utilize it to predict financial market?
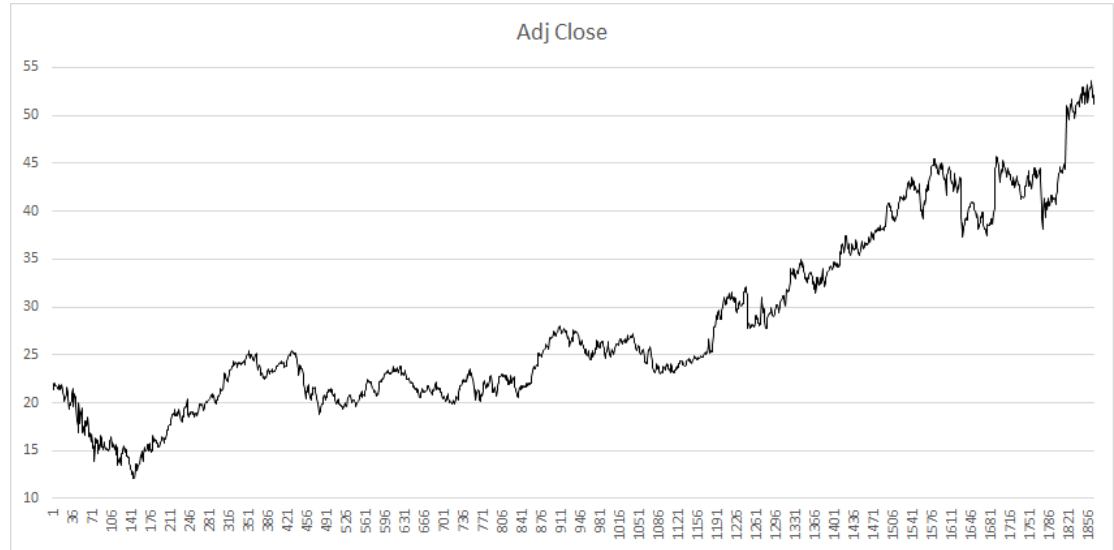
2. Data description

Target firms: Apple, Amazon, Facebook, Google, Microsoft, IBM, Tesla

Time period: 2014/03/10-2014/08/26

Stock data source: Yahoo Finance

**Domains:** Open, High, Low, Close prices, Adjusted close prices , Trading volume

Adj Close

Date	Open	High	Low	Close	Adj Close	Volume
3/10/2014	75.480003	76.190002	75.477142	75.845711	68.090271	44646000
3/11/2014	76.492859	76.96286	76.084282	76.584282	68.753311	69806100
3/12/2014	76.358574	76.764282	76	76.658569	68.820023	49831600
3/13/2014	76.777145	77.094284	75.594284	75.807144	68.055634	64435700
3/14/2014	75.541428	75.841431	74.714287	74.955711	67.291275	59299800
3/17/2014	75.385712	75.709999	75.121429	75.248573	67.554192	49886200
3/18/2014	75.128571	75.995712	75.028572	75.914284	68.151825	52411800
3/19/2014	76.03714	76.605713	75.571426	75.894287	68.133865	56189000
3/20/2014	75.69857	76.095711	75.335716	75.528572	67.80555	52099600
3/21/2014	75.989998	76.25	75.190002	76.124283	68.340355	93511600

4

2. Data description

- **News headlines data**
  - **Headline sources:** 422,419 news headlines from 11,237 websites
  - **Time period:** from 2014-03-10 to 2014-08-26
  - **News topics:** business, technology, entertainment and health
  - **News Types:** (firm) relevant news, all technical news

    (firm) relevant technical news
  - **Story :** the main subject (company) of the news
  - **Hostname:** where the article was posted

Relevant Tech News

	Biz	Tech	Ent	Health
APPL			Relevant News	
IBM		Tech News		
…				
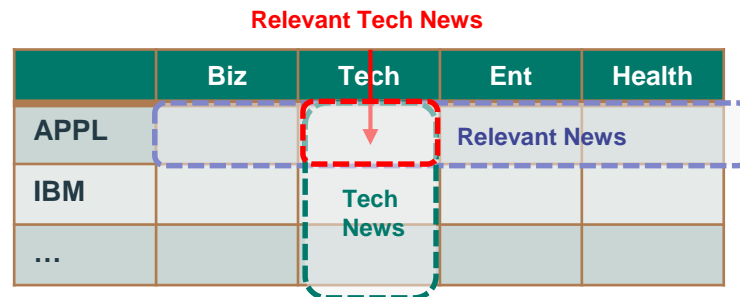
ID	TITLE	CATEGORY	STORY	HOSTNAME	TIMESTAMP
1	Fed official says weak data caused by weather, should not slow taper	b	ddUyU0VZz0	www.latime	1.39447E+12
2	Fed's Charles Plosser sees high bar for change in pace of tapering	b	ddUyU0VZz0	www.livemir	1.39447E+12
3	US open: Stocks fall after Fed official hints at accelerated tapering	b	ddUyU0VZz0	www.ifamag	1.39447E+12
4	Fed risks falling 'behind the curve', Charles Plosser says	b	ddUyU0VZz0	www.ifamag	1.39447E+12
5	Fed's Plosser: Nasty Weather Has Curbed Job Growth	b	ddUyU0VZz0	www.money	1.39447E+12
6	Plosser: Fed May Have to Accelerate Tapering Pace	b	ddUyU0VZz0	www.nasdac	1.39447E+12
7	Fed's Plosser: Taper pace may be too slow	b	ddUyU0VZz0	www.marke	1.39447E+12
8	Fed's Plosser expects US unemployment to fall to 6.2% by the end of 2014	b	ddUyU0VZz0	www.fxstree	1.39447E+12
9	US jobs growth last month hit by weather:Fed President Charles Plosser	b	ddUyU0VZz0	economictim	1.39447E+12
10	ECB unlikely to end sterilisation of SMP purchases - traders	b	dPhGU51Dcr	www.iii.co.u	1.39447E+12
11	ECB unlikely to end sterilization of SMP purchases: traders	b	dPhGU51Dcr	in.reuters.co	1.39447E+12
12	EU's half-baked bank union could work	b	dPhGU51Dcr	blogs.reuters	1.39447E+12
13	Europe reaches crunch point on banking union	b	dPhGU51Dcr	in.reuters.co	1.39447E+12

No.	Website	Freq	Density
1	in.reuters.com	2877	0.0068108
2	www.huffingtonpost.com	2603	0.0061621
3	www.businessweek.com	2420	0.0057289
4	www.contactmusic.com	2334	0.0055253
5	www.dailymail.co.uk	2258	0.0053454
6	www.nasdaq.com	2228	0.0052744
7	www.examiner.com	2085	0.0049359
8	www.globalpost.com	1975	0.0046755
9	www.latimes.com	1913	0.0045287
10	www.bizjournals.com	1882	0.0044553

5

3. Experiment: Proposed Models

- **Conventional Approach Using Stock Price:**
    - **Price-based: ARIMA**

- **Machine Learning Approaches Using Headlines:**
    - **Sentiment-Based: SVM**
    - **Non-Sentiment-Based: RNN**
    - **Non-Sentiment-Based: LSTM**

6

Description of ARIMA model prediction

- **Evaluation time period:** 10th March, 2014 – 30th June, 2014
- **Prediction time period:** 1st July, 2014 – 31st July, 2014 (in all 22 transaction days)
- **Example:** Facebook based on ARIMA (2,2,2)

Date	Stock Price	Trend	Prediction Price based on ARIMA(2,2,2)	Prediction trend	Match
6/30/2014	67.290001		67.290001		
7/1/2014	68.059998	1	67.4177	1	1
7/2/2014	66.449997	0	67.0755	0	1
7/3/2014	66.290001	0	67.1494	1	0
7/7/2014	65.290001	0	66.8485	0	1
7/8/2014	62.759998	0	66.8852	1	0
7/9/2014	64.970001	1	66.6173	0	0
7/10/2014	64.870003	0	66.6246	1	0
7/11/2014	66.339996	1	66.3828	0	0
7/14/2014	67.900002	1	66.3669	0	0
7/15/2014	67.169998	0	66.1458	0	1
7/16/2014	67.660004	1	66.1115	0	0
7/17/2014	66.410004	0	65.9068	0	1
7/18/2014	68.419998	1	65.8579	0	0
7/21/2014	69.400002	1	65.6661	0	0
7/22/2014	69.269997	0	65.6058	0	1
7/23/2014	71.290001	1	65.4241	0	0
7/24/2014	74.980003	1	65.3548	0	0
7/25/2014	75.190002	1	65.1812	0	0
7/28/2014	74.919998	0	65.1046	0	1
7/29/2014	73.709999	0	64.9375	0	1
7/30/2014	74.68	1	64.8552	0	0
7/31/2014	72.650002	0	64.6931	0	1
				Prediction accuracy	**0.40909091**

Prediction result of ARIMA model

Company	Predication correctness	Model	AIC	BIC
Apple	50.00%	ARIMA(2, 1, 2)	225.7535	237.5371
	50.00%	ARIMA(2, 1, 1)	225.9152	235.3421
	50.00%	ARIMA(1, 1, 2)	225.83	235.2568
	50.00%	ARIMA(1, 2, 2)	227.3184	236.6936
	50.00%	ARIMA(2, 2, 2)	228.5208	240.2398
	45.45%	**ARIMA(1, 1, 0)**	**222.4748**	**227.1883**
	45.45%	ARIMA(0, 1, 1)	222.475	227.1884
	45.45%	ARIMA(1, 1, 1)	224.3659	231.436
	45.45%	ARIMA(1, 2, 1)	225.4521	232.4835
	45.45%	ARIMA(2, 2, 1)	225.0521	232.0836
Amazon	45.45%	ARIMA(2, 1, 2)	523.5049	535.2884
	45.45%	ARIMA(2, 1, 1)	526.5866	536.0134
	50.00%	ARIMA(1, 1, 2)	522.2959	531.7228
	45.45%	ARIMA(1, 2, 2)	526.6151	535.9903
	45.45%	ARIMA(2, 2, 2)	539.017	550.736
	45.45%	ARIMA(1, 1, 0)	524.4863	529.1997
	45.45%	ARIMA(0, 1, 1)	524.4812	529.1946
	45.45%	**ARIMA(1, 1, 1)**	**520.9218**	**527.9919**
	45.45%	ARIMA(1, 2, 1)	539.234	546.2654
	45.45%	ARIMA(2, 2, 1)	524.2235	531.2549
Facebook	22.73%	**ARIMA(2, 1, 2)**	**288.9593**	300.7428
	50.00%	ARIMA(2, 1, 1)	295.5431	304.9699
	50.00%	ARIMA(1, 1, 2)	293.8098	303.2366
	50.00%	ARIMA(1, 2, 2)	296.232	305.6072
	40.91%	ARIMA(2, 2, 2)	302.849	314.568
	50.00%	ARIMA(1, 1, 0)	291.546	296.2595
	50.00%	**ARIMA(0, 1, 1)**	291.546	**296.2594**
	50.00%	ARIMA(1, 1, 1)	292.6344	299.7046
	50.00%	ARIMA(1, 2, 1)	295.2607	302.2921
	50.00%	ARIMA(2, 2, 1)	295.0876	302.1191

Prediction result of ARIMA model

| Company	Predication correctness	Model	AIC	BIC
Google	50.00%	ARIMA(2, 1, 2)	557.8292	569.6128
	50.00%	ARIMA(2, 1, 1)	559.1286	568.5555
	54.55%	ARIMA(1, 1, 2)	557.6681	567.0949
	54.55%	ARIMA(1, 2, 2)	559.5036	568.8788
	68.18%	ARIMA(2, 2, 2)	559.4831	571.2022
	54.54%	ARIMA(1, 1, 0)	557.4612	562.1746
	54.54%	ARIMA(0, 1, 1)	557.1963	561.9097
	45.45%	ARIMA(1, 1, 1)	557.3646	564.4347
	63.63%	**ARIMA(1, 2, 1)**	**555.2298**	**562.2612**
	63.63%	**ARIMA(2, 2, 1)**	**555.2298**	**562.2612**
Microsoft	40.91%	**ARIMA(2, 1, 2)**	**86.44131**	98.22485
	36.36%	ARIMA(2, 1, 1)	90.75954	100.1864
	36.36%	ARIMA(1, 1, 2)	90.43673	99.86356
	36.36%	ARIMA(1, 2, 2)	88.49174	97.86696
	36.36%	ARIMA(2, 2, 2)	87.0211	98.74013
	54.54%	ARIMA(1, 1, 0)	90.08427	94.79769
	54.54%	ARIMA(0, 1, 1)	90.08249	94.79591
	54.54%	**ARIMA(1, 1, 1)**	87.40701	**94.47714**
	31.82%	ARIMA(1, 2, 1)	90.11436	97.14578
	31.82%	ARIMA(2, 2, 1)	88.31618	95.3476
IBM	45.45%	ARIMA(2, 1, 2)	305.6733	317.4568
	50.00%	**ARIMA(2, 1, 1)**	**303.6116**	313.0384
	50.00%	ARIMA(1, 1, 2)	303.8555	313.2823
	68.18%	ARIMA(1, 2, 2)	307.1758	316.551
	63.64%	ARIMA(2, 2, 2)	308.0023	319.7213
	45.45%	ARIMA(1, 1, 0)	305.1655	309.8789
	45.45%	**ARIMA(0, 1, 1)**	304.986	**309.6994**
	45.45%	ARIMA(1, 1, 1)	303.7964	310.8665
	63.64%	ARIMA(1, 2, 1)	353.3993	360.4307
	63.64%	ARIMA(2, 2, 1)	305.4135	312.4449
Tesla	54.55%	ARIMA(2, 1, 2)	516.6564	528.4399
	54.55%	ARIMA(2, 1, 1)	514.6984	524.1253
	54.55%	ARIMA(1, 2, 2)	514.6628	524.0897
	50.00%	ARIMA(1, 2, 2)	514.3804	523.7556
	50.00%	ARIMA(2, 2, 2)	515.6002	527.3193
	45.45%	ARIMA(1, 1, 0)	512.5183	517.2317
	45.45%	**ARIMA(0, 1, 1)**	**512.3571**	**517.0705**
	54.54%	ARIMA(1, 1, 1)	513.7484	520.8186
	50.50%	ARIMA(1, 2, 1)	513.3618	520.3932
	50.50%	ARIMA(2, 2, 1)	513.0184	520.0498

Sentiment & Price trend: Apple, Amazon
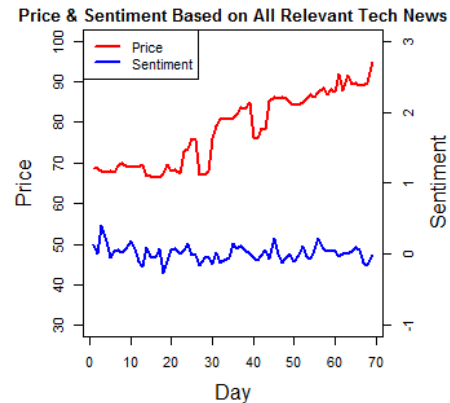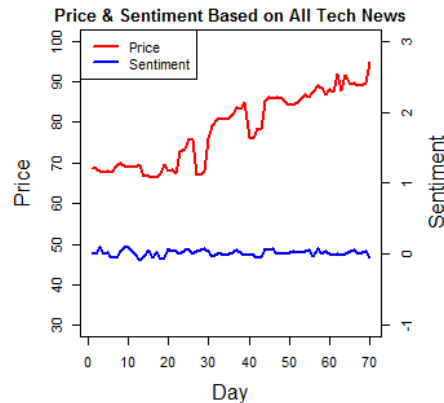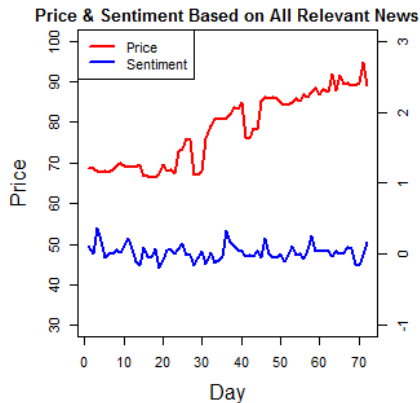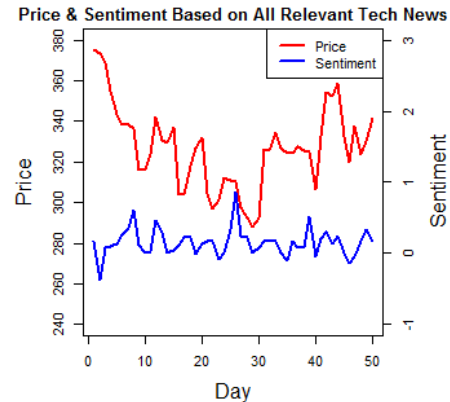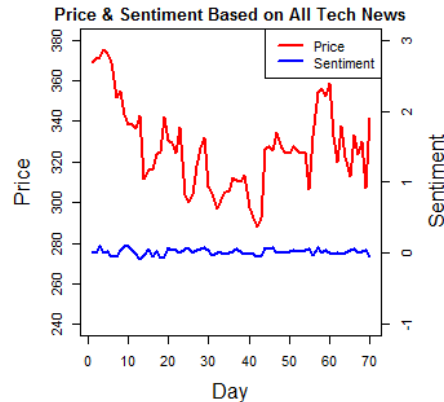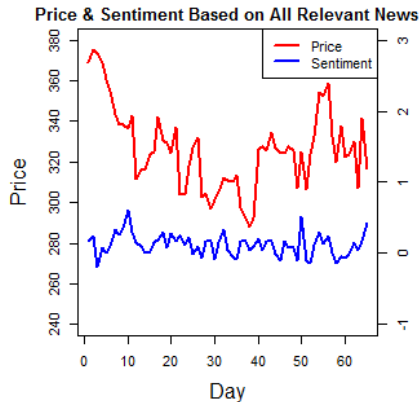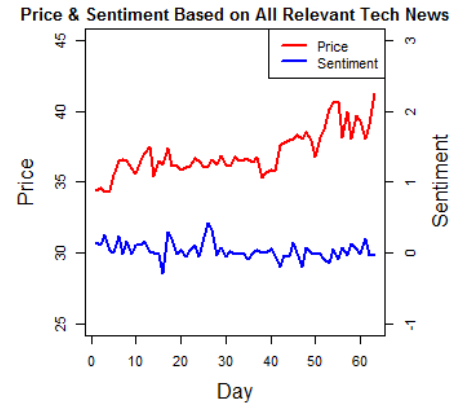
Sentiment & Price trend: Facebook, Google

Facebook
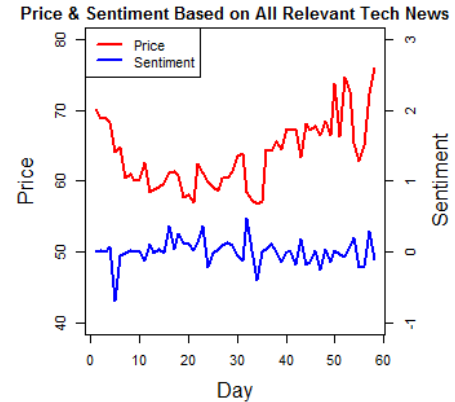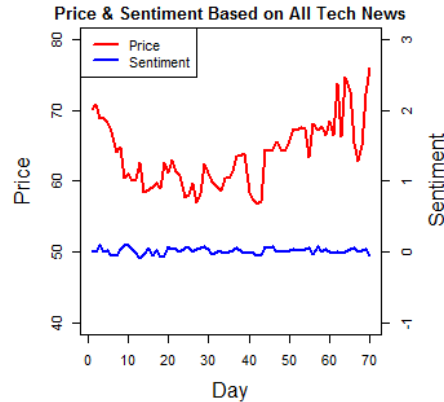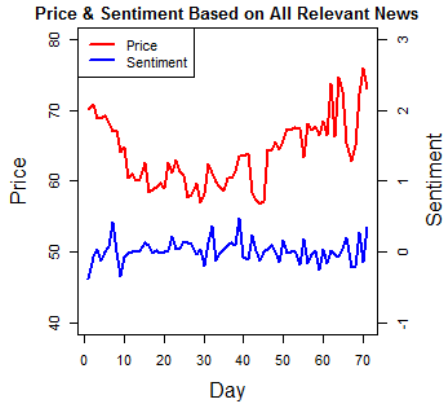
Google
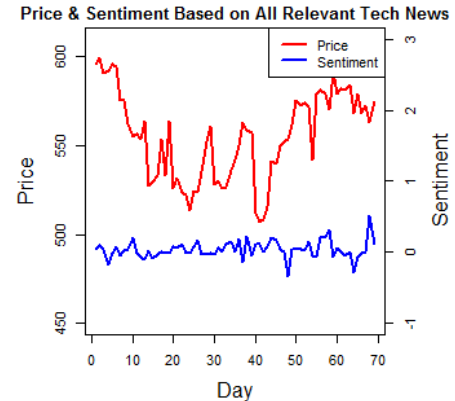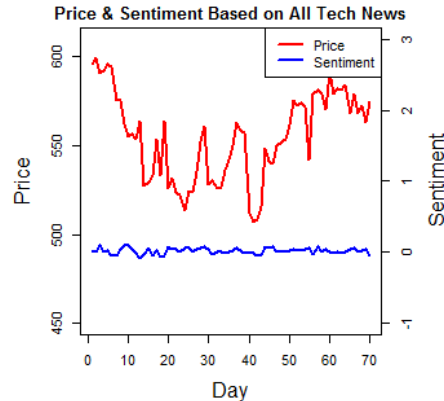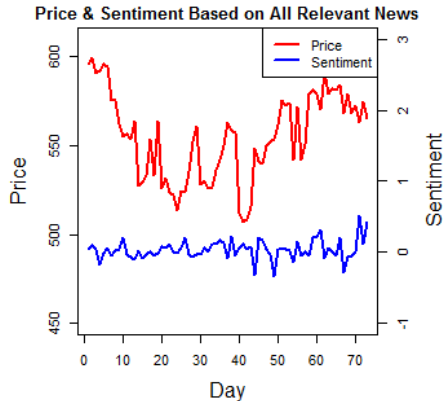
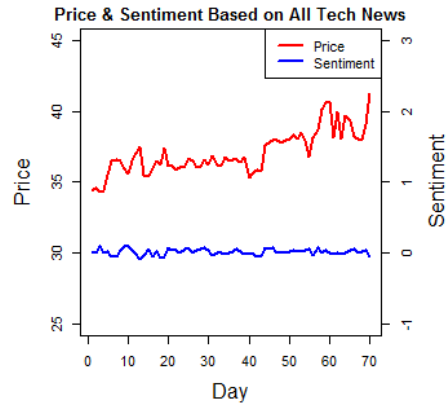# Sentiment & Price trend: Microsoft, IBM

12
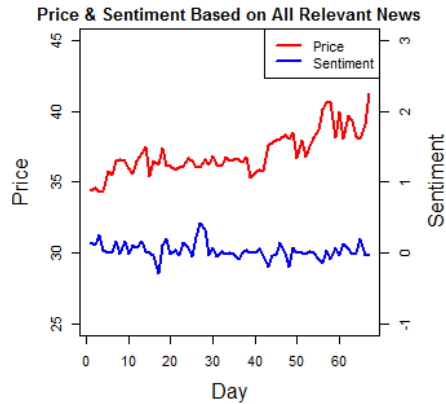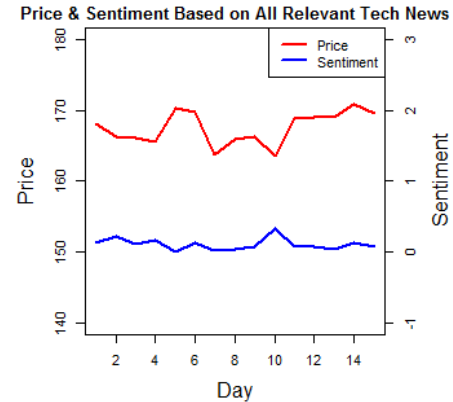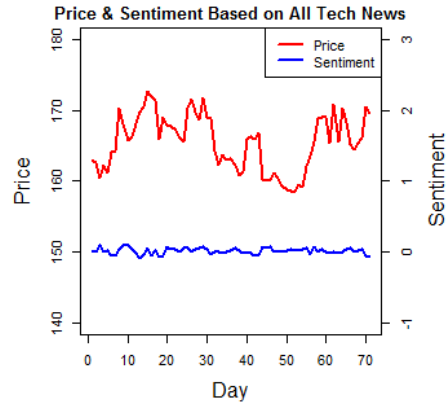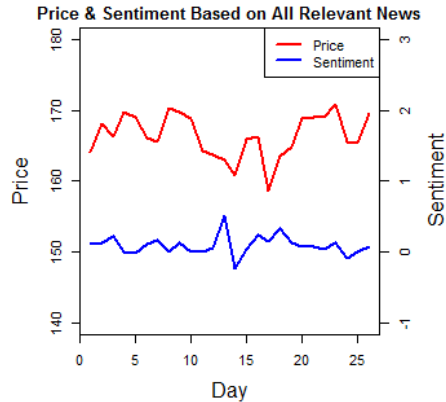
Sentiment & Price trend: Tesla

Tesla
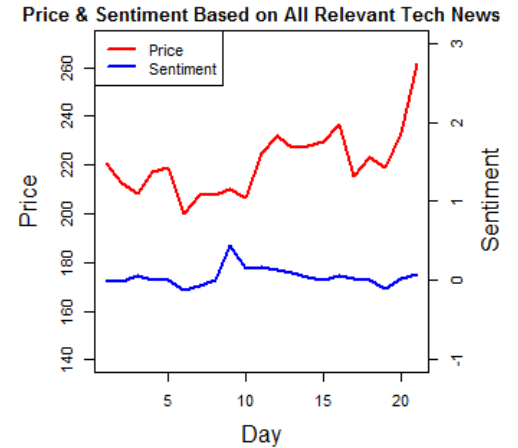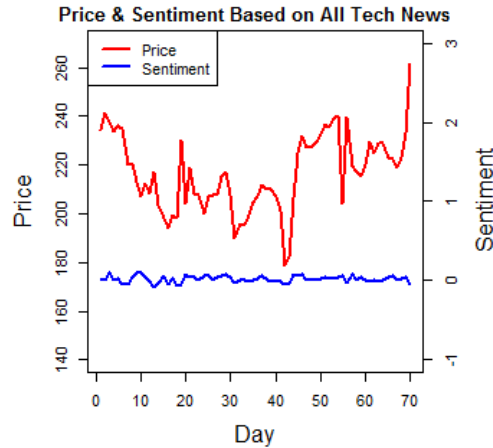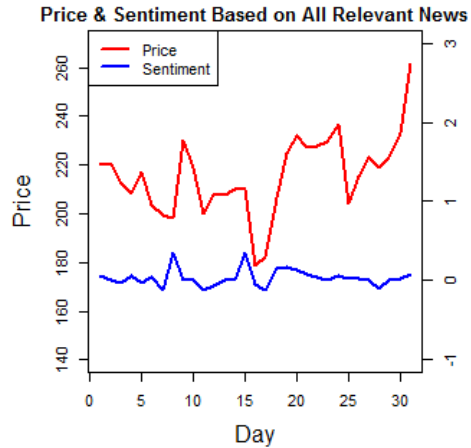
Stock response showed a time lag

Correlation between the sentiment and the price trend

Company	News type	Same day	One day lag	Two day lag	Three day lag
Apple	All relevant news	0.061	-0.093	**-0.278**	-0.191
	All tech news	-0.084	-0.010	**-0.142**	0.041
	All relevant tech news	-0.009	-0.123	**-0.247**	-0.180
Amazon	All relevant news	-0.002	-0.132	-0.078	**0.230**
	All tech news	-0.135	0.096	**-0.233**	0.062
	All relevant tech news	-0.106	-0.090	-0.023	**0.155**
Facebook	All relevant news	0.031	-0.001	0.106	**0.116**
	All tech news	0.022	**0.084**	-0.072	-0.060
	All relevant tech news	0.090	**-0.190**	0.152	0.066
Google	All relevant news	**-0.209**	0.068	0.144	-0.012
	All tech news	**-0.104**	0.009	-0.857	0.035
	All relevant tech news	**-0.303**	0.029	0.068	0.078
IBM	All relevant news	-0.008	-0.046	0.057	**0.301**
	All tech news	-0.137	-0.013	0.026	**0.207**
	All relevant tech news	-0.005	0.061	-0.091	**0.340**
Microsoft	All relevant news	0.028	**0.421**	-0.190	-0.073
	All tech news	-0.039	**0.185**	0.081	-0.025
	All relevant tech news	0.011	0.018	**-0.069**	0.013
Tesla	All relevant news	**-0.172**	0.041	0.161	0.073
	All tech news	0.083	**-0.094**	-0.034	0.014
	All relevant tech news	-0.040	0.204	**0.402**	0.223

Most companies' stock response showed time lag (two days and three days)

SVM model

Source Data

Text Data Proprecessing

Sentiment Analysis: NLTK Tool

Training Set (News before 2014-07-01)

SVM

Testing Set (News after 2014-07-01)

Trained Model

Results

- Stemming
- Remove Stop Words
- ...

All Relevant News

Tech Relevant News
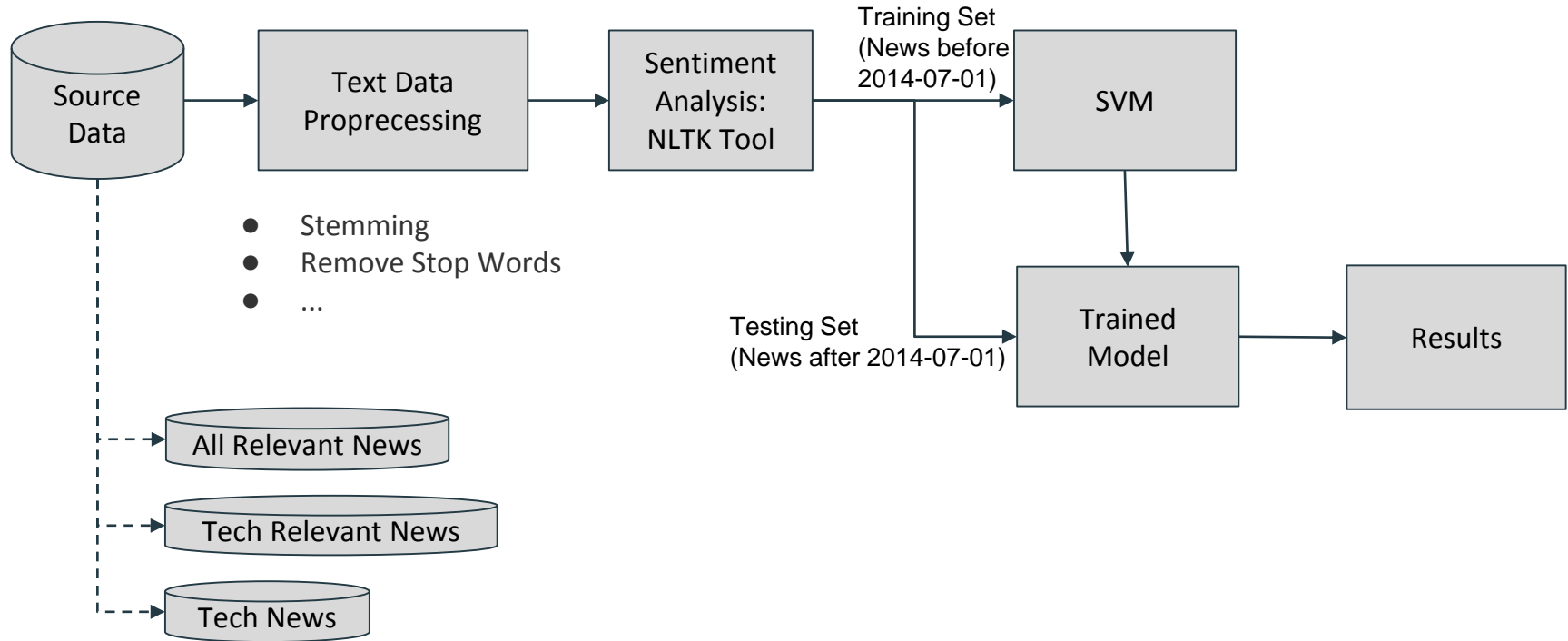
Tech News

Experiment Results - SVM

Table: Prediction accuracy on different stocks(SVM)

Stock	Inputs			Inputs	Average Accuracy	Accuracy			
	All Relevant News	Tech Relevant News	Tech News			0 lag	1 lag	2 lag	3 lag
Apple	√			AAPL Apple news	72.6%	68%	68%	64%	**90%**
		√		AAPL Apple tech news	72.6%	67%	68%	65%	**90%**
			√	AAPL tech news	**73.1%**	76%	67%	65%	**84%**
Amazon	√			AMZN Amazon news	59.6%	56%	49%	60%	**73%**
		√		AMZN Amazon tech news	**63.6%**	54%	55%	**78%**	68%
			√	AMZN tech news	61.4%	51%	65%	62%	**68%**
Facebook	√			FB Facebook news	**79.5%**	61%	**92%**	90%	76%
		√		FB Facebook tech news	78.5%	59%	**91%**	89%	75%
			√	FB tech news	69.6%	60%	73%	**74%**	71%
Google	√			GOOG google news	70.6%	62%	**81%**	69%	71%
		√		GOOG google tech news	70.2%	67%	**78%**	64%	71%
			√	GOOG tech news	**73.1%**	70%	75%	**75%**	72%
IBM	√			IBM IBM news	72.3%	75%	57%	**82%**	75%
		√		IBM IBM tech news	**75.4%**	**85%**	53%	79%	85%
			√	IBM tech news	60.7%	56%	54%	63%	**70%**
Microsoft	√			MSFT microsoft news	68.6%	55%	**82%**	72%	66%
		√		MSFT microsoft tech news	68.7%	55%	**82%**	72%	66%
			√	MSFT tech news	**70.8%**	64%	71%	71%	**77%**
Tesla	√			TSLA Tesla news	**66.6%**	56%	53%	**83%**	74%
		√		TSLA Tesla tech news	61.4%	56%	52%	**83%**	54%
			√	TSLA tech news	66.3%	70%	**76%**	65%	54%

Examples

Date	Title	Sentiment	Today		Tomorrow		Day after tomorrow	
			Predict	Real	Predict	Real	Predict	Real
2014-08-28	Apple's Smartwatch Will Debut On September 9	0.0	0	1	1	0	0	0
2014-08-28	Apple Fails Again in Bid for Samsung Smartphone Sales Ban	-0.7506	0	1	0	0	0	0
2014-08-28	12.9 inch Bigger Jumbo iPad: Apple Gets Ready To Revitalize Tablet Business	0.3612	1	1	0	0	0	0

17

RNN model

RNN: Recurrent Neural Network.

Feature:

> Detect and learn the pattern hidden in the sequence.
>
> When it read the word, it still "remember" what has been read in last steps.
>
> No manually created features

Apple's Tim Cook Attends Pride Event After Being Outed as Gay

```
RNN
```

```
Prediction of stock price
```
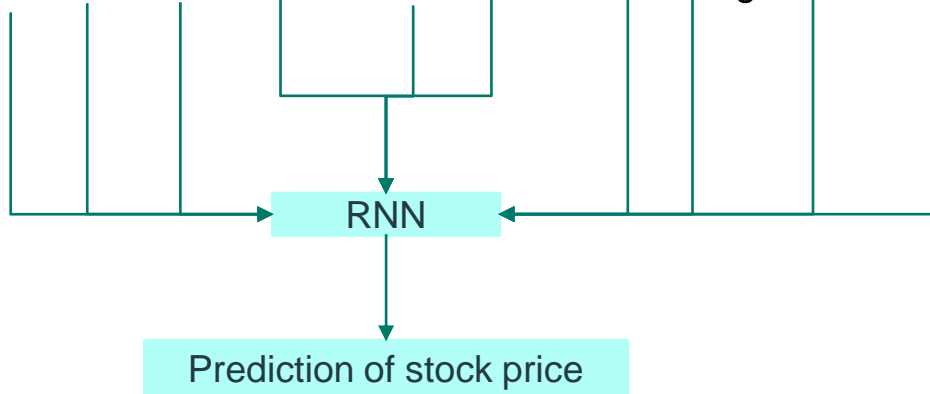
LSTM model

LSTM: Long Short Term Memory Recurrent Neural Network.

Feature:

An advanced version of RNN.

It could learn automatically which word in the previously read texts is more important compared with other words.

Apple's Tim Cook Attends Pride Event After Being Outed as Gay

Memory filtering mechanism

LSTM

Prediction of stock price

RNN

Table: Prediction accuracy on different stocks(RNN)

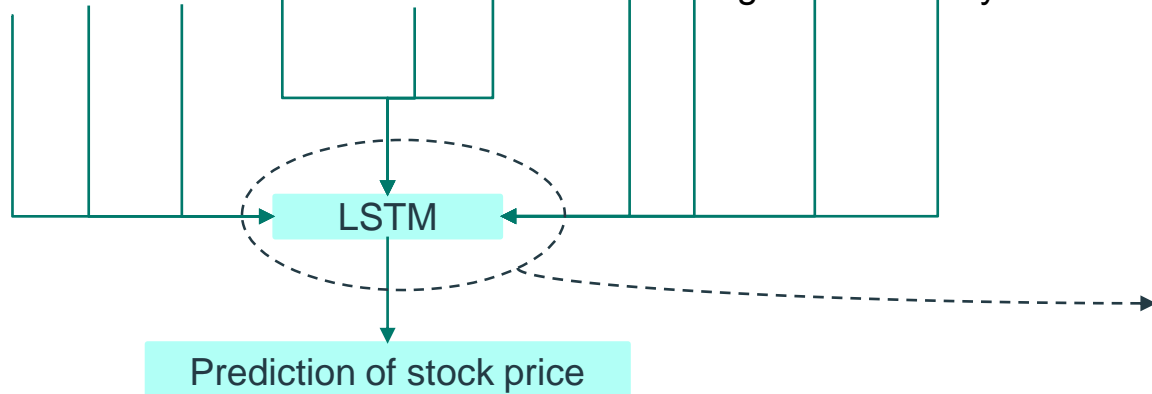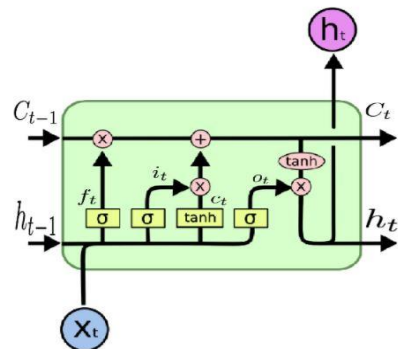| Stock	Inputs			Accuracy				
	All Relevant News	Tech Relevant News	Tech News	Average Accuracy	0 lag	1 lag	2 lags	3 lags
Apple	√			**62.0%**	55.3%	58.6%	59.5%	**74.73%**
		√		59.2%	56.5%	54.8%	57.4%	**68.09%**
			√	59.8%	59.8%	55.1%	58.4%	**65.81%**
Amazon	√			54.4%	51.0%	50.7%	43.9%	**72.03%**
		√		52.6%	53.2%	44.9%	43.3%	**68.90%**
			√	**56.6%**	50.2%	56.8%	**59.9%**	59.66%
Facebook	√			**70.2%**	57.9%	62.8%	**88.5%**	71.81%
		√		66.6%	55.8%	59.6%	**80.6%**	70.45%
			√	64.1%	56.3%	63.6%	**71.1%**	65.24%
Google	√			61.0%	56.6%	**66.1%**	63.4%	57.96%
		√		54.7%	**58.7%**	55.5%	50.9%	53.64%
			√	**64.8%**	59.3%	64.1%	68.0%	67.92%
IBM	√			47.1%	25.0%	56.9%	**81.8%**	24.62%
		√		51.8%	21.6%	53.0%	**78.5%**	53.88%
			√	**53.1%**	43.4%	50.9%	58.4%	**59.69%**
Microsoft	√			**63.8%**	54.1%	71.9%	**66.1%**	63.27%
		√		63.1%	51.8%	**69.6%**	66.9%	64.02%
			√	58.8%	52.6%	55.9%	61.3%	**65.56%**
Tesla	√			**60.7%**	43.1%	52.7%	**82.9%**	64.06%
		√		52.0%	42.7%	52.4%	**56.7%**	56.07%
			√	57.7%	56.7%	**65.9%**	62.5%	45.67%

LSTM

Table: Prediction accuracy on different stocks(LSTM)

| Stock	Inputs			Accuracy				
	All Relevant News	Tech Relevant News	Tech News	Average Accuracy	0 lag	1 lag	2 lags	3 lags
Apple	√			**61.8%**	56.6%	56.7%	59.5%	**74.33%**
		√		60.8%	61.1%	53.4%	54.2%	**74.34%**
			√	59.3%	54.9%	57.2%	58.5%	**66.56%**
Amazon	√			54.4%	54.0%	51.1%	42.0%	**70.60%**
		√		49.6%	57.4%	41.3%	41.7%	**57.95%**
			√	**56.3%**	49.3%	56.2%	59.3%	**60.43%**
Facebook	√			**68.9%**	58.0%	55.8%	**88.1%**	73.79%
		√		68.5%	57.2%	60.7%	**83.3%**	72.76%
			√	63.3%	56.4%	63.5%	**70.6%**	62.64%
Google	√			57.7%	58.6%	**64.2%**	54.3%	53.53%
		√		56.7%	**62.1%**	58.1%	49.5%	57.12%
			√	**63.3%**	59.7%	62.6%	**67.6%**	63.31%
IBM	√			47.1%	25.0%	56.9%	**81.8%**	24.62%
		√		52.7%	15.2%	53.0%	**78.5%**	64.19%
			√	**53.0%**	44.7%	51.0%	57.6%	**58.70%**
Microsoft	√			**63.5%**	54.6%	67.5%	**68.3%**	63.33%
		√		62.0%	52.0%	65.2%	**67.9%**	62.84%
			√	59.5%	54.0%	56.2%	61.7%	**66.25%**
Tesla	√			57.3%	42.1%	40.0%	**82.9%**	64.18%
		√		57.6%	42.5%	52.4%	**82.5%**	53.14%
			√	**60.7%**	58.2%	**65.4%**	61.9%	57.29%

4. Conclusions

1. **Accuracy in bold** represents the accuracy of model M_i with best performance for Firm F_j:

$$max_{NT,L} \; Acc\{M_i \, (F_j, NT, L)\}$$

2. Accuracy in red represents that accuracy of best model among models M_i ($i = 1,2,3$) for Firm F_j:

$$max_{i,NT,L} \; Acc\{M_i \, (F_j, NT, L)\}$$

F

L

NT
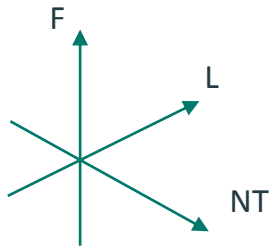
$M_i(F_j, NT_k, L_l)$

$i = 1,2,3;$
$j = 1,\ldots,7;$
$k = 1,2,3;$
$l = 1,\ldots,4;$

3 Models using headlines (M)

7 Firms (F)

3 News Types (NT)

Lag (L): 0-3 days

Firms		ARIMA(p,d,q)	All Relevant				Tech Relevant				Tech News			
			Cor(Sent, Trend)	SVM	RNN	LSTM	Cor(Sent, Trend)	SVM	RNN	LSTM	Cor(Sent, Trend)	SVM	RNN	LSTM
Apple	Lags	(1,1,0)	2	3	3	3	2	3	3	3	2	3	3	3
	Accuracy	45%	-	90%	75%	74%	-	90%	68%	74%	-	84%	66%	67%
Amazon	Lags	(1,1,1)	3	3	3	3	3	2	3	3	2	3	2	3
	Accuracy	45%	-	73%	72%	71%	-	78%	69%	58%	-	68%	60%	60%
Facebook	Lags	(0,1,1)	3	1	2	2	1	1	2	2	1	2	2	2
	Accuracy	50%	-	92%	89%	88%	-	91%	81%	83%	-	74%	71%	71%
Google	Lags	(1,2,1) or (2,2,1)	0	1	1	1	0	1	0	0	0	2	2	2
	Accuracy	64%	-	81%	66%	64%	-	78%	59%	62%	-	75%	68%	68%
Microsoft	Lags	(1,1,1)	1	1	2	2	2	1	1	2	1	3	3	3
	Accuracy	55%	-	82%	66%	68%	-	82%	70%	68%	-	77%	66%	66%
IBM	Lags	(2,1,1) or (0,1,1)	3	2	2	2	3	0	2	2	3	3	3	3
	Accuracy	50%	-	82%	82%	82%	-	85%	79%	79%	-	70%	60%	59%
Tesla	Lags	(0,1,1)	0	2	2	2	2	2	2	2	1	1	1	1
	Accuracy	45%	-	83%	83%	83%	-	83%	57%	83%	-	76%	66%	65%
Average Accuracy		-	-	83%	76%	76%	-	84%	69%	72%	-	75%	65%	65%

Table: Prediction Accuracy for different model of different Firms, different News Types and Time of Lags

4. Conclusions & Discussion

1. Different models: Sentiment-based method performs better than non-sentiment-based method;

- When using headline, almost for every firms, the best model is **SVM model** supported by sentiment;

- In future study, we can consider using sentiment analysis to improve performance of RNN/LSTM model

Sentiment-Based

Non-Sentiment-Based

Firms		ARIMA(p,d,q)	All Relevant				Tech Relevant				Tech News			
			Cor(Sent, Trend)	SVM	RNN	LSTM	Cor(Sent, Trend)	SVM	RNN	LSTM	Cor(Sent, Trend)	SVM	RNN	LSTM
Apple	Lags	(1,1,0)	2	3	3	3	2	3	3	3	2	3	3	3
	Accuracy	45%	-	90%	75%	74%	-	90%	68%	74%	-	84%	66%	67%
Amazon	Lags	(1,1,1)	3	3	3	3	3	2	3	3	2	3	2	3
	Accuracy	45%	-	73%	72%	71%	-	78%	69%	58%	-	68%	60%	60%
Facebook	Lags	(0,1,1)	3	1	2	2	1	1	2	2	1	2	2	2
	Accuracy	50%	-	92%	89%	88%	-	91%	81%	83%	-	74%	71%	71%
Google	Lags	(1,2,1) or (2,2,1)	0	1	1	1	0	1	0	0	0	2	2	2
	Accuracy	64%	-	81%	66%	64%	-	78%	59%	62%	-	75%	68%	68%
Microsoft	Lags	(1,1,1)	1	1	2	2	2	1	1	2	1	3	3	3
	Accuracy	55%	-	82%	66%	68%	-	82%	70%	68%	-	77%	66%	66%
IBM	Lags	(2,1,1) or (0,1,1)	3	2	2	2	3	0	2	2	3	3	3	3
	Accuracy	50%	-	82%	82%	82%	-	85%	79%	79%	-	70%	60%	59%
Tesla	Lags	(0,1,1)	0	2	2	2	2	2	2	2	1	1	1	1
	Accuracy	45%	-	83%	83%	83%	-	83%	57%	83%	-	76%	66%	65%
Average Accuracy		-	-	83%	76%	76%	-	84%	69%	72%	-	75%	65%	65%

Table: Prediction Accuracy for different model of different Firms, different News Types and Time of Lags

4. Conclusions & Discussion

2. **News headline can be used to predict financial market;**

- Methods using headline to predict financial market trend outperform methods using only price significantly;

- In the future study, we can consider using price as well as textual data to predict financial market;

Best model in methods using headlines

Firms		News Type	Model	Lags	Accuracy
Apple	Using price	-	ARIMA(1,1,0)	-	45%
	Using headlines	All Relevant or Tech Relevant	SVM	3	90%
Amazon	Using price	-	ARIMA(1,1,1)	-	45%
	Using headlines	All Relevant	SVM	2	78%
Facebook	Using price	-	ARIMA(0,1,1)	-	50%
	Using headlines	All Relevant	SVM	1	92%
Google	Using price	-	ARIMA(1,2,1) or ARIMA(2,2,1)	-	64%
	Using headlines	All Relevant	SVM	1	81%
Microsoft	Using price	-	ARIMA(1,1,1)	-	55%
	Using headlines	Tech Relevant	SVM	1	82%
IBM	Using price	-	ARIMA(2,1,1) or ARIMA(0,1,1)	-	50%
	Using headlines	Tech Relevant	SVM	0	85%
Tesla	Using price	-	ARIMA(0,1,1)	-	45%
	Using headlines	Tech Relevant	SVM	2	83%
Average Accuracy	Using price				51%
	Using headlines				84%

Table: Comparison between method of using price (ARIMA) and best model among the methods of using headlines

23

4. Conclusions & Discussion

3.   Different news types showed different predictive power

- The scope of news type

Relevant Tech News

	Biz	Tech	Ent	Health
APPL			Relevant News	
IBM		Tech News		
…				

- Relevance news makes significant change:

  a) **Tech News** outperforms other news type doesn't exist in our study;

  b) For Apple, Amazon, Facebook, Google, the **All Relevant news** outperforms other news type;

  c) For Apple, Microsoft, IBM and Tesla, **Tech Relevant** news outperform others;

Best model in methods using headlines

Firms		News Type	Model	Lags	Accuracy
Apple	Using price	-	ARIMA(1,1,0)	-	45%
	Using headlines	All Relevant or Tech Relevant	SVM	3	**90%**
Amazon	Using price	-	ARIMA(1,1,1)	-	45%
	Using headlines	All Relevant	SVM	2	**78%**
Facebook	Using price	-	ARIMA(0,1,1)	-	50%
	Using headlines	All Relevant	SVM	1	**92%**
Google	Using price	-	ARIMA(1,2,1) or ARIMA(2,2,1)	-	64%
	Using headlines	All Relevant	SVM	1	**81%**
Microsoft	Using price	-	ARIMA(1,1,1)	-	55%
	Using headlines	Tech Relevant	SVM	1	**82%**
IBM	Using price	-	ARIMA(2,1,1) or ARIMA(0,1,1)	-	50%
	Using headlines	Tech Relevant	SVM	0	**85%**
Tesla	Using price	-	ARIMA(0,1,1)	-	45%
	Using headlines	Tech Relevant	SVM	2	**83%**
Average Accuracy	Using price				51%
	Using headlines				**84%**

Table: Comparison between method of using price (ARIMA) and best model among the methods of using headlines

4. Conclusions & Discussion

4. **Different days of lag**: **We find that there exist time lags in financial market prediction when using news headline.**

- Different firms showed different days of time lags.

- Time lag for the firms: (Ranking from large to small):
  - Apple, Amazon;
  - Microsoft, IBM;
  - Facebook;
  - Tesla;
  - Google;

Table: Comparison of time lag between different models

Firms		All Relevant					Tech Relevant					Tech News					Ave. Lag for 3 News Type
		Cor(S,T)	SVM	RNN	LSTM	Ave. Lag	Cor(S,T)	SVM	RNN	LSTM	Ave. Lag	Cor(S,T)	SVM	RNN	LSTM	Ave. Lag	
Apple	Lags	2	3	3	3	2.75	2	3	3	3	2.75	2	3	3	3	2.75	2.75
	Accuracy	-	90%	75%	74%	-	-	90%	68%	74%	-	-	84%	66%	67%	-	-
Amazon	Lags	3	3	3	3	3.00	3	2	3	3	2.75	2	3	2	3	2.50	2.75
	Accuracy	-	73%	72%	71%	-	-	78%	69%	58%	-	-	68%	60%	60%	-	-
Facebook	Lags	3	1	2	2	2.00	1	1	2	2	1.50	1	2	2	2	1.75	1.75
	Accuracy	-	92%	89%	88%	-	-	91%	81%	83%	-	-	74%	71%	71%	-	-
Google	Lags	0	1	1	1	0.75	0	1	0	0	0.25	0	2	2	2	1.50	0.83
	Accuracy	-	81%	66%	64%	-	-	78%	59%	62%	-	-	75%	68%	68%	-	-
Microsoft	Lags	3	1	2	2	2.00	3	1	1	1	1.75	1	3	3	3	2.50	2.08
	Accuracy	-	82%	66%	68%	-	-	82%	70%	68%	-	-	77%	66%	66%	-	-
IBM	Lags	1	2	2	2	1.75	2	0	2	2	1.50	3	3	3	3	3.00	2.08
	Accuracy	-	82%	82%	82%	-	-	85%	79%	79%	-	-	70%	60%	59%	-	-
Tesla	Lags	0	2	2	2	1.50	2	2	2	2	2.00	1	1	1	1	1.00	1.50
	Accuracy	-	83%	83%	83%	-	-	83%	57%	83%	-	-	76%	66%	65%	-	-
Average Lag		1.71	1.35	1.45	1.45	1.96	1.86	1.13	1.27	1.36	1.79	1.43	1.59	1.47	1.54	2.14	-

- The possible reasons might be:
  - Different investors groups for different firms, e.g., younger investor group vs elder investor group;
  - Special news was reported during observational period for some firms, e.g., earning announcement, new product release;

Thank You!