

Directional Prediction of Stock Prices using Breaking News on Twitter

Hana Alostad

School of Computing, Informatics and
Decision Systems Engineering
Arizona State University
Tempe, Arizona 85287-8809
Email: hana.alostad@asu.edu

Hasan Davulcu

School of Computing, Informatics and
Decision Systems Engineering
Arizona State University
Tempe, Arizona 85287-8809
Email: HasanDavulcu@asu.edu

Abstract—Stock market news and investing tips are popular topics in Twitter. In this paper, first we utilize a 5-year financial news corpus comprising over 50,000 articles collected from the NASDAQ website for the 30 stock symbols in Dow Jones Index (DJI) to train a directional stock price prediction system based on news content. Then we proceed to prove that information in articles indicated by breaking Tweet volumes leads to a statistically significant boost in the hourly directional prediction accuracies for the prices of DJI stocks mentioned in these articles. Secondly, we show that using document-level sentiment extraction does not yield to a statistically significant boost in the directional predictive accuracies in the presence of other 1-gram keyword features.

Keywords—stock prediction; text mining; breaking news; twitter analysis; twitter volume spike; stock trading

I. INTRODUCTION

Online social networks, like Twitter, are enabling people who are passionate about trading and investing to break critical financial news faster and they also go deep into relevant areas of research and sources leading to interesting insights. Recently Twitter has been used to detect and forecast civil unrest [1], criminal incidents [2], box-office revenues of movies [3], and seasonal influenza [4].

Stock market news and investing tips are popular topics in Twitter. In this paper, first we utilize a 5-year financial news corpus comprising over 50,000 articles collected from the NASDAQ website for the 30 stock symbols in Dow Jones Index (DJI) to train a directional stock price prediction system based on news content. Next we collect over 750,000 tweets during a 6 month period in 2014 that mention at least one of the 30 DJI stock symbols. We utilize the 68-95-99.7 rule, also known as the three-sigma rule or empirical rule [5], to define a simple method for detecting hourly stock symbol related tweet volume breakouts. Then we proceed to test our hypothesis to determine if “information in articles indicated by breaking Tweet volumes will lead to a statistically significant boost in the hourly directional prediction accuracies for the prices of DJI stocks mentioned in these articles”.

The contributions of the paper can be summarized as follows: Firstly, we show that sparse logistic regression [6] for this text based classification task with 1-gram keyword features filtered by a Chi2 [7] feature selection algorithm leads to the best overall directional prediction accuracy. Secondly, we show that using document-level sentiment extraction does not yield to a statistically significant boost in the predictive accuracies in the presence of other 1-gram keyword features. Thirdly, we

show that the breaking news based system indeed yields a statistically significant boost in directional prediction accuracy compared to the one using all news.

The rest of the paper is organized as follows. Section II presents related work. Section III presents formal problem definition. The system architecture is presented in Section IV. Section V presents the experimental data we used, and quantitative 10-fold cross validation results for the two experiments that we performed. The last section concludes the paper and discusses future work.

II. RELATED WORK

Stock price prediction problem has been studied for several years, several research papers have been published with a goal to increase the accuracy of prediction. Table I contains a summary of previous research results related to stock price or direction prediction, the input data sets used, the time-frames for prediction, the length of the period of collected data, prediction algorithms used, and the overall accuracies.

These systems have different prediction time-frames and goals. Some of them predict stock price for the intended time-frame like [10], [14] and [19]. Time frames vary as daily or next 20 minutes. Others such as [16], [12], [13], [15], [18], [8], and [17] predict stock price direction for the next day. [11] aims to predict the price direction every 2-hours, and [9] aims to predict monthly direction.

Related systems collected their input data from various sources and exchanges: [14], [18], and [19] collected stock news, tweets and price charts related to S&P 500 companies. [17] collected tweets and stock price data related to Nasdaq stocks, [16] collected tweets and stock price charts related to Dow Jones Industrial Average (DJIA), [13] collected one year of data related to Microsoft company. [9] collected stock price charts from Shenzhen Development Stock A (SDSA) exchange. [11] collected currency price and news data related to foreign exchange market (Forex). [8] collected stock price data from CNX Nifty, S&P BSE Sensex exchanges and finally [10] collected thirteen years of stock price charts data related to Goldman Sachs Group Inc.

[9], [8], and [10] used only stock price as input to predict stock price or direction with accuracies varying between 83% and 90%. [12], [13], [15], and [11] are examples of papers which utilize news as well as stock prices to predict price direction with varying accuracies between 51% and 83%.

[18] made correlation analysis between the stock price and the tweet volume, and used it to predict stock market

TABLE I. SUMMARY OF PREVIOUS RESEARCH RESULTS

| Reference | Data set | | | Time-frame | Period | Prediction | Algorithm | Accuracy |
|-----------|-------------|-------------|--------|------------|--------|------------|---------------------|--------------|
| | Stock Price | Online News | Tweets | | | | | |
| [8] | ✓ | ✗ | ✗ | Daily | 9 Yrs | Direction | Naive Bayes | 90% |
| [9] | ✓ | ✗ | ✗ | Monthly | 2 Yrs | Direction | Logistic Regression | 83% |
| [10] | ✓ | ✗ | ✗ | Daily | 13 Yrs | Price | Linear Regression | 2.54 (RMSE) |
| [11] | ✓ | ✓ | ✗ | 2 Hrs | 4 Yrs | Direction | SVM | 83% |
| [12] | ✓ | ✓ | ✗ | Daily | 14 Yrs | Direction | SVM | 79% |
| [13] | ✓ | ✓ | ✗ | Daily | 1 Year | Direction | SVM | 61% |
| [14] | ✓ | ✓ | ✗ | 20 Min | 1 Mo | Price | SVR | 51%, |
| [15] | ✓ | ✓ | ✗ | Daily | 1 Year | Direction | Neural Network | 3.70 (RMSE) |
| [16] | ✓ | ✗ | ✓ | Daily | 10 Mos | Direction | Neural Network | 88% |
| [17] | ✓ | ✗ | ✓ | Daily | 2 Mos | Direction | Decision Tree | 77% |
| [18] | ✓ | ✗ | ✓ | Daily | 3 Mos | Direction | Liner regression | 68% |
| [19] | ✓ | ✗ | ✓ | Daily | 1 Year | Price | Bayesian | 0.3% (daily) |

direction with 68% accuracy. Following work by [19] analyzed tweet spikes in combination with price action based technical indicators such as price breakout direction as an input to a Bayesian classifier for stock price prediction, yielding a daily average gain of approximately 0.3% during a period of 55 days generating a total gain of 15%. [16] used extracted sentiment information from Twitter data and a neural network classifier to predict Dow Jones Industrial average (DJIA) daily price direction with 88% accuracy. [17] also used sentiment information extracted from Twitter as input to a decision tree classifier to predict price direction for four companies in NASDAQ stock exchange with average accuracy of 77% distributed as APPL at 77%, GOOG at 77%, MSFT at 69% and AMZN at 85% during a two months period of evaluations.

III. PROBLEM DEFINITION

The correction effect of online news articles covering company related events, announcements and technical analyst reports on the stock price may take some time to show. Depending on the severity and impact of the news announcement this period may vary between few minutes to an hour, and the effect may sometimes determine the trend direction of the financial instrument for upcoming weeks or months.

One way to measure the impact of news on a stock price is to analyze the trading volume following the news announcement. Another indicator of news impact is the diffusion rates and volumes of messages on social media containing the stock symbol and news links of interest.

Twitter provides a suitable platform to investigate properties of such information diffusion. Diffusion analysis can harness social media to investigate “viral tweets” to create early-warning indicators that can signal if a breakout started to emerge in its nascent stages. In this paper, we utilize the 68-95-99.7 rule to define a simple method of tweet volume breakouts. In statistics, the 68-95-99.7 rule, also known as the three-sigma rule or empirical rule [5], states that nearly all values lie within three standard deviations (σ) of the mean (μ) in a normal distribution. We utilize a fixed sized sliding window (of length 20 hour intervals that was determined experimentally), to compute a running average and standard deviation for the hourly volumes of tweets that mention a stock symbol. Then, we identify *breakout* signals within a time-series of hourly tweet volumes for each stock symbol whenever its hourly volume exceeds $(\mu(20) + 2\sigma)$ of the

previous 20 hour periods. We consider a breakout as an indication that traders or technical analysts are sharing some exciting or important new information regarding the company or a group of companies. Next, we collect the URL links mentioned within the breaking-news hour of tweets and we designed a pair of experiments to test our hypothesis whether “information in news indicated by breaking tweet volumes will lead to statistically significant boost in the directional prediction accuracy for the prices of the related stock symbols mentioned in these articles”.

Our system has the following characteristics:

- 1) Input data: hourly stock price charts of the 30 stocks comprising the Dow Jones Index (DJI), online stock news articles for a 5 year period spanning 2010 and 2014 from NASDAQ¹ news website, the tweets related the 30 stock symbols collected from Twitter Streaming API² spanning a 6 months period between March 2014 and September 2014, and online news articles mentioned in tweets during breaking news hours.
- 2) Prediction time-frame: The collected data is analyzed and predictions are made on hourly bases.
- 3) Prediction goal: To predict the hourly price direction for the stocks mentioned in tweets during breaking news hours.

The distinguishing features of our system compared to systems mentioned in our related work are: (1) [19] used Tweeter volume spikes alongside stock price-based technical indicators for stock price turning point prediction where as our system utilizes textual content of the news mentioned in tweets during breaking Twitter volume hours to predict the hourly direction of the stock price following a breakout period. (2) [16] and [17] used extracted sentiment information alongside stock price-based technical indicators to see if sentiment information leads to a boost in the predicted direction accuracy. Our system primarily relies on textual content of the news from breaking tweet volume hours to predict price direction. We also experimented with extracted sentiment as an additional feature to see if it leads to a boost in overall accuracy. Unlike [16] and [17], our system did not experience a statistically significant boost in predictive accuracies as a result of including sentiment information alongside other textual content features. [16]’s

¹<http://www.nasdaq.com/symbol/ibm/news-headlines>

²<https://dev.twitter.com/streaming/overview>

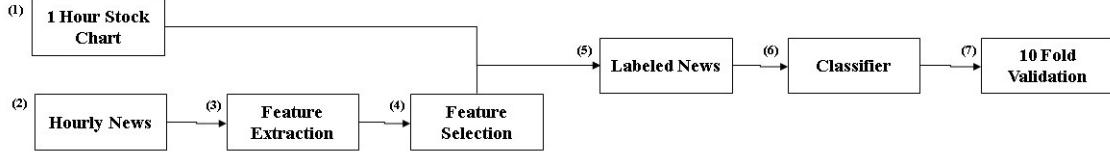


Fig. 1. System Architecture

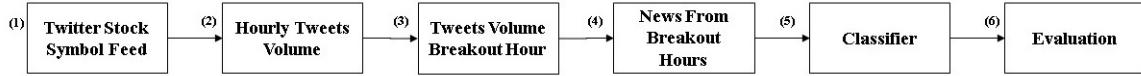


Fig. 2. Illustration of System Architecture of Experiment-2

accuracy is not comparable to ours since they are reporting the daily directional prediction accuracy for the Dow Jones Index Average (DJIA). Compared to predictive accuracies for four companies listed in [17], we have only one stock in common with their experiments, i.e. MSFT, where their system reported a daily directional predictive accuracy of 69% and our system reports an hourly directional accuracy of 82%.

IV. SYSTEM ARCHITECTURE

In order to test the hypothesis that “information in news indicated by *breaking* tweet volumes will lead to statistically significant boost in the directional prediction accuracy for the prices of the relevant stock symbols mentioned in such articles”, we designed two experiments. In the first experiment we trained a classifier using *all* stock news articles for a 5 year period spanning 2010 and 2014 from NASDAQ news website. Figure 1 illustrates the system architecture used for the first experiment. For comparison purposes we experimented with three different types of features extracted from text: 1-gram keywords, 2-gram phrases, and bi-polar sentiment (i.e. positive and negative) extracted from text. We grouped news hourly, and categorized each hourly collection as one of two categories: (1) those that led to an increased stock price or (2) those that led to a price reduction during the next hour. Next, we applied a feature selection method to reduce the number of features to only relevant ones. The details of these steps are presented in the Section IV.A. Finally we experimented with two types of text classifiers and evaluated their directional predictive accuracy using 10-fold cross validation. The results of the first experiment utilizing all stock news for all 30 company stocks are reported in Section V.B. In our second experiment, we tested the directional predictive accuracy of our classifier (i.e. trained in the first experiment above) using *only* online articles collected during hourly breaking tweet volume periods. Figure 2 illustrates the system architecture used for our second experiment. Steps involved in the second experiment were hourly profiling of the tweets mentioning a stock symbol, detection of tweet volume breakout periods, collection of online news mentioned in tweets during the breaking hours, feature extraction from news, and running of the classifier to predict the stock price direction of the next hour following a breaking hour. We compared the accuracies of the classifiers in both first and second experiments to test the validity of our hypothesis. The details of the steps involved in the second experiment are explained in Section IV.B, and the

experimental results and evaluations are presented in Section V.C.

A. Experiment-1: Hourly Price Direction Prediction using Online News

The following is a detailed description of each step used in Experiment-1:

- 1) One Hour Stock Chart: We collected hourly stock financial price charts for all the companies comprising the Dow Jones Index (DJI) using an API from ActiveTick³. For each trading hour the price direction was calculated based on the hourly Open and Close prices according to the Formula 1 below, where d represents the trading date and h represents the trading hour:

$$Dir(d, h) = \begin{cases} 1 & \text{if } Open(d, h) \leq Close(d, h) \\ -1 & \text{otherwise} \end{cases} \quad (1)$$

- 2) Hourly News: We used Web Content Extractor⁴ to collect online news articles from NASDAQ website. We stored all metadata information related to the articles like their title, url, date, time, and source in a database table. After that we fetched the news content using their urls and performed content extraction using Boilerpipe⁵.

3) Feature Extraction:

- N-gram Features from News: R for text mining tm⁶ package was used to extract keyword features from the news corpus. First all whitespaces, stop words, numbers, punctuation were removed from the documents, then all the terms were converted to lowercase and stemmed into their root words. Next features were recorded in a document-term matrix. For each stock symbol we created a pair of document-term matrices: one with 1-gram features and another with 2-gram features represented in a binary form. We used R.matlab⁷ package to create Matlab format files that corresponds to these matrices.
- Sentiment Features: To detect sentiment in news content we used a Java version of SentiStrength library⁸,

³<http://www.activetick.com/>

⁴<http://www.newprosoft.com/>

⁵<http://code.google.com/p/boilerpipe/>

⁶<http://cran.r-project.org/web/packages/tm/index.html>

⁷<http://cran.r-project.org/web/packages/R.matlab/index.html>

⁸<http://sentistrength.wlv.ac.uk/>

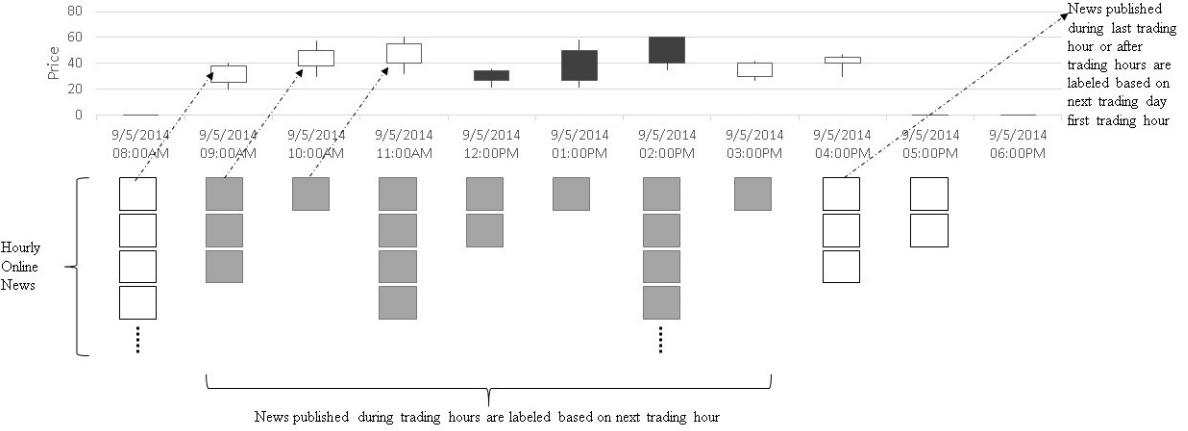


Fig. 3. An Illustration of News Labeling

SentiStrength is a classifier that uses predefined sentiment word list with human polarity and strength judgments, then it applies some rules to detect sentiment in short text [20]. [21] showed that using general word lists for sentiment analysis of large financial text leads into mis-classifying of common words in financial domain, so based on our SentiStrength initial testing results and the findings of [21] we decided to use Loughran and McDonald Financial Sentiment Dictionaries⁹ instead of the general sentiment word list supplied with SentiStrength. Besides using different sentiment word list we needed to get the sentiment for each document. We used OpenNLP¹⁰ Sentence Detector to extract sentences from each document, and then we applied the SentiStrength classifier on each sentence. We determined the majority polarity for the sentences contained in a document and used the majority polarity (i.e. positive or negative) as the sentiment for the document.

- 4) Feature Selection: Feature selection in text mining reduces the number of features to only relevant and discriminative set of features. We used Chi2 [7] feature selection algorithm from a feature selection package¹¹. Chi2 is a two phase general algorithm that selects automatically a proper critical value for statistical χ^2 test and then it removes all irrelevant and redundant features [7].
- 5) News Labeling: Figure 3 is an illustration of the news labeling step. In this phase we used the stock price direction of the following hour to categorize the directionality of the hourly collections of news articles. In order to align the news article hours with the stock chart hours we had to standardize and adjust their time zones. Formula 2 is used to label the news articles where d represents the publishing date, and h represents the publishing hour.

$$\text{Label}(d, h) = \text{Dir}(d, \text{Next}(h)) \quad (2)$$

In this paper we are assuming that the effect of published news articles will be reflected on the stock price direction

during the next hour. Formula 2 applies to all the news articles published during official trading hours which starts at 9AM and ends on 3PM in EST time zone. For articles that are published during the last trading hour, or after trading hours, or during holidays and weekends we assumed that their effect will be seen on the direction of the first trading hour of the next trading day. For this case Formula 3 is used to label those news articles.

$$\text{Label}(d, h) = \text{Dir}(\text{Next}(d), \text{First}(h)) \quad (3)$$

- 6) Classifier: We formulate price direction prediction problem as a classification problem in a general structured sparse learning framework [6]. In particular, the logistical regression formulation presented below fits this application, since it is a dichotomous classification problem (e.g. upwards vs. downwards price correction). In the formula 4, a_i is the vector representation of the news during the i^{th} hour, w_i is the weight assigned to the i^{th} document ($w_i=1/m$ by default), and $A=[a_1, a_2, \dots, a_m]$ is the document n-gram matrix, y_i is the directionality of each hour based upon the stock price action of the next hour, and the unknown x_j , the j-th element of x , is the weight for each n-gram feature, $\lambda > 0$ is a regularization parameter that controls the sparsity of the solution, $|x|_1 = \sum|x_i|$ is 1-norm of the x vector. We used the SLEP [6] sparse learning package that utilizes gradient descent approach to solve the above convex and non-smooth optimization problem. The n-grams with non-zero values on the sparse x vector yield the discriminant factors for classifying a news collection as leading to upwards or downwards price correction. n-grams with positive polarity correspond to upward direction indicators, and those with negative polarity correspond to downward direction indicators.

$$\min_x \sum_{i=1}^n w_i \log(1 + \exp(1 + y_i(x^t a_i + c))) + \lambda|x| \quad (4)$$

We also utilized an SVM classifier during our experiments using LIBSVM¹².

⁹http://www3.nd.edu/~mcdonald/Word_Lists.html

¹⁰<https://opennlp.apache.org/>

¹¹<http://featureselection.asu.edu/software.php>

¹²<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

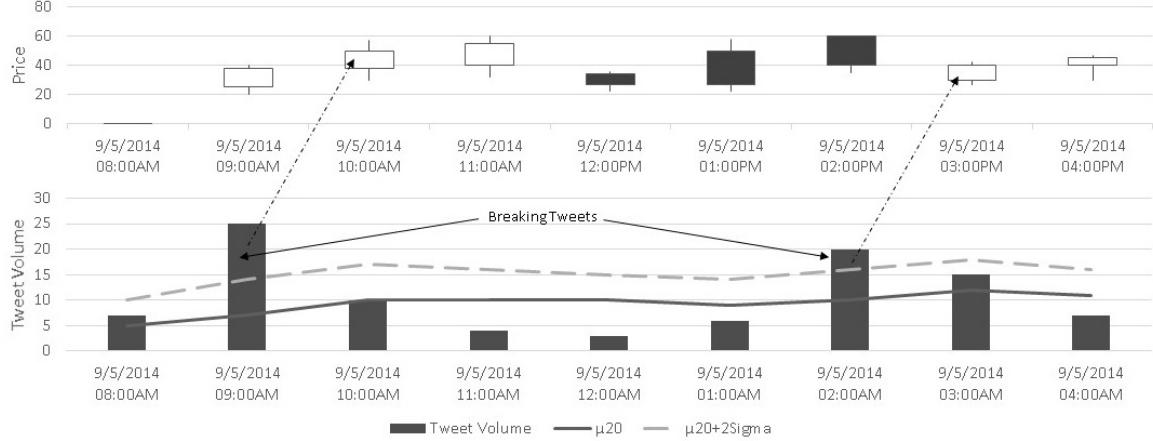


Fig. 4. An Illustration of Breaking Tweets

- 7) 10-fold cross validation: We run a total of 8 experiments for each stock symbol where we experimented: (1) with SVM and sparse logistic regression classifiers, (2) with 1-gram and 2-gram features, and (3) with and without extracted sentiment features. After the training phase of the classifier, we validated the accuracies using 10-fold cross validation. The evaluation results for the first experiment are presented in Section V.B.

B. Experiment-2: Hourly Price Direction Prediction using Breaking News

We selected the classifier with the best performance emerging from Experiment-1 to use in Experiment-2. Experiment-2 was designed to test if the online news indicated by *breaking* tweet volumes would lead to a statistically significant boost in the directional prediction accuracy for the prices of the relevant stock symbols mentioned in such news. The system architecture figure in Fig. 2 shows the steps used in this experiment. The following is a detailed description of each step:

- 1) Twitter Stock Symbol Feed: Twitter streaming API was used to collect tweets related to companies in the Dow Jones Index (DJI). In order to collect relevant tweets we used a keyword filter made from the stock symbols, either prefixed by a dollar sign (\$) or prefixed by “NYSE:” or “NASDAQ:”. For example, the keyword filter for Microsoft Corp. is \$MSFT or NYSE:MSFT. For each matching tweet we stored the stock symbol, tweet text, date, time, and the set of URLs mentioned in the tweet. If the tweet text contained more than one stock symbol then we stored the same tweet information for each mentioned stock symbol.
- 2) Hourly Tweets Volume Profiling: We utilize a fixed sized sliding window (of length 20 hour intervals) where the 20 hour intervals was determined by conducting several experiments with different intervals, to compute a running average $\mu[20]$ and standard deviation σ for the hourly volumes of tweets that mention a stock symbol.
- 3) Tweets Volume Breakout Hour: We identify *breakout* signals within a time-series of hourly tweet volumes for each stock symbol using Formula 5.

$$\text{Breakout} = \begin{cases} \text{True} & \text{if } N(d, h) \geq \mu[20](d, h) + 2\sigma(d, h) \\ \text{False} & \text{otherwise} \end{cases} \quad (5)$$

In Formula 5 N represents tweet volume on specific date d , and hour h , $\mu[20]$ is 20-hour simple moving average applied on tweets' volume, $\mu[20](d, h) + 2\sigma(d, h)$ represents the upper band for simple moving average - a 20-hour moving average plus 2-times standard deviation. If the volume of hourly tweets N exceeds the upper band value, this would indicate a volume breakout. Otherwise the tweet volume is non-breaking. In Fig. 4, the pair of dotted arrows shows two instances of tweet volume breakouts at 9/5/2014 at 9AM and 9/5/2014 at 2PM, where the corresponding articles from these hours will be used to predict the price directions of the mentioned stocks at following hours.

- 4) News From Breaking Tweets: In this step the news content of URLs found in the tweets during the breaking hours are downloaded and their textual content is extracted using the following steps:
 - a) For each breaking hour of a specific stock symbol we fetch the URLs found in tweets during the breaking hour, i.e. $\text{Breakout} = \text{True}$. In some cases the URLs were mentioned in their short URL forms, so before fetching them, they were converted to their long forms.
 - b) Fetch the URL links' content and perform content extraction from the HTML documents using the jsoup HTML parser¹³.
- 5) Classifier: After extracting the hourly breaking news and their 1-gram features we utilized the logistic regression classifier to predict the price direction for the next hour.
- 6) Evaluation: The predictive accuracies of the classifier for news following breaking hours are presented in Section V.C.

¹³<http://jsoup.org/>

TABLE II. COUNTS OF COLLECTED NEWS ARTICLES, TWEETS, AND CALCULATED BREAKING TWEETS FOR 30 DOW JONES STOCK SYMBOLS

| Stock Symbol | News | Tweets | Breaking Tweets | Avg URLs per Breaking Tweets |
|--------------|------|--------|-----------------|------------------------------|
| \$AXP | 1614 | 15251 | 155 | 10 |
| \$BA | 2006 | 19041 | 207 | 15 |
| \$CAT | 1842 | 19303 | 188 | 18 |
| \$CSCO | 1984 | 26611 | 285 | 19 |
| \$CVX | 2168 | 15897 | 202 | 13 |
| \$DD | 1553 | 11218 | 63 | 11 |
| \$DIS | 1870 | 25014 | 199 | 17 |
| \$GE | 2260 | 31336 | 274 | 21 |
| \$GS | 1878 | 52888 | 246 | 40 |
| \$HD | 1743 | 18459 | 164 | 13 |
| \$IBM | 2188 | 80412 | 248 | 50 |
| \$INTC | 2157 | 30724 | 273 | 23 |
| \$JNJ | 2232 | 18236 | 224 | 16 |
| \$JPM | 1543 | 33658 | 274 | 27 |
| \$KO | 1899 | 22688 | 208 | 21 |

| Stock Symbol | News | Tweets | Breaking Tweets | Avg URLs per Breaking Tweets |
|--------------|------|--------|-----------------|------------------------------|
| \$MCD | 1879 | 21419 | 228 | 17 |
| \$MMM | 1183 | 11438 | 122 | 9 |
| \$MRK | 1573 | 28882 | 199 | 39 |
| \$MSFT | 1733 | 59469 | 271 | 54 |
| \$NKE | 1080 | 18206 | 196 | 18 |
| \$PFE | 1841 | 50859 | 292 | 57 |
| \$PG | 1781 | 15097 | 192 | 12 |
| \$T | 1784 | 30128 | 263 | 24 |
| \$TRV | 968 | 8858 | 52 | 7 |
| \$UNH | 1133 | 11224 | 144 | 10 |
| \$UTX | 1278 | 10872 | 68 | 9 |
| \$V | 1683 | 19174 | 190 | 13 |
| \$VZ | 2194 | 20896 | 252 | 16 |
| \$WMT | 2216 | 30448 | 236 | 27 |
| \$XOM | 2378 | 22433 | 248 | 15 |

TABLE III. ACCURACY RESULTS OF EXPERIMENT 2: HOURLY PRICE DIRECTION PREDICTION USING BREAKING NEWS

| Stock Symbol | Experiment-1 Accuracy | Experiment-2 Accuracy |
|--------------|-----------------------|-----------------------|
| \$AXP | 67.84% | 69.66% |
| \$BA | 68.20% | 79.00% |
| \$CAT | 65.37% | 68.39% |
| \$CSCO | 67.70% | 69.55% |
| \$CVX | 66.60% | 65.22% |
| \$DD | 67.93% | 66.67% |
| \$DIS | 65.60% | 64.75% |
| \$GE | 66.60% | 66.67% |
| \$GS | 68.30% | 71.60% |
| \$HD | 68.70% | 75.00% |
| \$IBM | 66.10% | 80.80% |
| \$INTC | 69.60% | 67.30% |
| \$JNJ | 64.40% | 69.20% |
| \$JPM | 67.50% | 74.00% |
| \$KO | 67.90% | 68.60% |

| Stock Symbol | Experiment-1 Accuracy | Experiment-2 Accuracy |
|--------------|-----------------------|-----------------------|
| \$MCD | 67.80% | 65.30% |
| \$MMM | 70.60% | 76.50% |
| \$MRK | 69.30% | 64.80% |
| \$MSFT | 72.60% | 81.80% |
| \$NKE | 70.30% | 71.30% |
| \$PFE | 70.70% | 59.50% |
| \$PG | 70.80% | 72.70% |
| \$T | 75.80% | 82.90% |
| \$TRV | 65.40% | 75.00% |
| \$UNH | 68.40% | 75.20% |
| \$UTX | 67.80% | 58.60% |
| \$V | 71.20% | 71.00% |
| \$VZ | 65.30% | 72.80% |
| \$WMT | 66.60% | 71.00% |
| \$XOM | 65.20% | 67.30% |

V. EVALUATION OF EXPERIMENTS RESULTS

A. Collected Data

We collected online news articles and stock price charts related to 30 stock symbols in Dow Jones Index for the period between October, 2009 and September, 2014. We also collected tweets matching stock symbols for the period between March, 2014 and September, 2014. The total number of news articles collected from the NASDAQ website for the 30 stock symbols in Dow Jones Index is 53,641. The total number of collected tweets matching stock symbols is 780,139. Table II shows the counts of news articles, total number of collected tweets, and the number of hourly breakout periods for each symbol.

B. Experiment-1: Hourly Price Direction Prediction using Online News

We executed the steps described in Figure 1 on data sets collected for the 30 Dow Jones Index companies. In order to identify the best set of text features and the best classifier we had to perform several experiments. We run a total of 8 experiments for each stock symbol where we experimented: (1) with SVM and sparse logistic regression classifiers, (2) with 1-gram and 2-gram keyword features, and (3) with and without extracted sentiment for documents. After the training phase of the classifier, we validated the accuracies using 10-fold cross validation.

The evaluations show that 1-gram features led to higher overall accuracies compared to 2-gram features for both SVM

and LogisticR experiments. Also, the experimental setup with the LogisticR classifier using 1-Gram features where the sentiment features were excluded led to the maximal accuracies in 19 out of 30 cases. The second best experimental setup that achieved the maximal accuracies was also with the LogisticR classifier with 1-gram features integrated with the sentiment feature.

Hence, in order to determine the utility of extracted sentiment features we formulated the following hypotheses and applied Non-parametric sign test [22] at confidence level 95% to test if sentiment features would yield a statistically significant boost in overall prediction accuracies:

- 1) Null Hypothesis (h0): 1-gram LogisticR classifier without sentiment features accuracies' mean = 1-gram LogisticR classifier with sentiment accuracies' mean, indicating that they are at the same level of performance.
- 2) Alternative Hypothesis (h1): 1-gram LogisticR classifier without sentiment features accuracies' mean \neq 1-gram LogisticR classifier with sentiment features accuracies' mean, indicating that they are not at the same level of performance.

The p-value of the sign test to compare 1-gram LogisticR classifier without sentiment features with 1-gram LogisticR classifier with sentiment features at significance level 0.05 equals to 0.1221, which leads to the acceptance of the null hypothesis h0 and the rejection of the alternative hypothesis h1, concluding that using sentiment would *not* yield a statistically

TABLE IV. EXPERIMENT 2 FINANCIAL EVALUATION

| Stock Symbol | Gross Profit Points | Gross Loss Points | Total Profit Points | (ROI%) | Total Trades | Long Position (won%) | Short Position (won%) |
|--------------|---------------------|-------------------|---------------------|------------|--------------|----------------------|-----------------------|
| \$AXP | \$14.53 | -\$7.99 | \$6.543 | 7% | 155 | 71% | 68% |
| SBA | \$67.77 | -\$18.79 | \$48.99 | 47% | 207 | 74% | 87% |
| SCAT | \$30.67 | -\$15.51 | \$15.16 | 16% | 188 | 66% | 81% |
| SCSCO | \$13.00 | -\$4.96 | \$8.03 | 39% | 285 | 69% | 71% |
| SCVX | \$22.96 | -\$14.74 | \$8.22 | 7% | 202 | 76% | 58% |
| SDD | \$2.62 | -\$1.66 | \$0.96 | 1% | 63 | 67% | 67% |
| \$DIS | \$21.80 | -\$9.65 | \$12.15 | 16% | 199 | 59% | 83% |
| SGE | \$9.47 | -\$3.26 | \$6.21 | 27% | 274 | 76% | 64% |
| SGS | \$39.45 | -\$13.30 | \$26.15 | 18% | 246 | 74% | 70% |
| \$HD | \$21.79 | -\$5.82 | \$15.97 | 22% | 164 | 73% | 85% |
| SIBM | \$75.25 | -\$14.28 | \$60.97 | 37% | 248 | 76% | 93% |
| SINTC | \$14.54 | -\$5.96 | \$8.58 | 35% | 273 | 73% | 58% |
| \$JNJ | \$27.08 | -\$12.63 | \$14.45 | 15% | 224 | 65% | 84% |
| SJPM | \$24.12 | -\$9.69 | \$14.43 | 29% | 274 | 67% | 94% |
| SKO | \$12.63 | -\$6.37 | \$6.26 | 17% | 208 | 67% | 75% |
| \$MCD | \$23.98 | -\$14.28 | \$9.70 | 10% | 228 | 72% | 62% |
| \$MMMK | \$23.96 | -\$9.36 | \$14.60 | 11% | 122 | 71% | 96% |
| \$MRK | \$24.67 | -\$14.13 | \$10.54 | 20% | 199 | 67% | 63% |
| \$MSFT | \$26.87 | -\$6.21 | \$20.66 | 65% | 271 | 88% | 77% |
| SNKE | \$25.80 | -\$8.95 | \$16.85 | 25% | 196 | 62% | 87% |
| SPFE | \$19.17 | -\$10.24 | \$8.93 | 35% | 292 | 60% | 59% |
| SPG | \$24.97 | -\$8.80 | \$16.18 | 22% | 192 | 67% | 94% |
| ST | \$14.81 | -\$2.76 | \$12.05 | 41% | 263 | 85% | 82% |
| STRV | \$11.26 | -\$1.73 | \$9.53 | 21% | 52 | 71% | 86% |
| SUNH | \$22.34 | -\$9.23 | \$13.11 | 19% | 144 | 68% | 91% |
| SUTX | \$5.05 | -\$3.34 | \$1.71 | 1% | 68 | 50% | 86% |
| \$V | \$84.70 | -\$28.25 | \$56.45 | 31% | 190 | 75% | 68% |
| SVZ | \$16.01 | -\$4.97 | \$11.03 | 26% | 252 | 76% | 71% |
| SWMT | \$19.38 | -\$8.63 | \$10.75 | 15% | 236 | 75% | 68% |
| \$XOM | \$34.84 | -\$14.86 | \$19.98 | 22% | 248 | 72% | 62% |
| Average | \$25.85 | -\$9.68 | \$16.17 | 23% | 205 | 70% | 76% |

significant boost in the overall prediction accuracy.

C. Experiment-2: Hourly Price Direction Prediction using Breaking News

In Experiment-2 we applied steps outlined in Figure 2 to 30 stock symbols in Dow Jones Index using breaking news periods only as trade triggers. Table III shows that Experiment-2 led to a boost in predictive accuracies for 70% of the stock symbols (i.e. 21 out of 30 cases). In order to prove that Experiment-2 yields a statistically significant boost in prediction accuracy compared to Experiment-1 we applied Sign test at confidence level 95%. We formulated the following hypothesis:

- 1) Null Hypothesis (h0): Experiment-1 accuracies mean = Experiment-2 accuracies mean, indicating that they are at the same level of performance.
- 2) Alternative Hypothesis (h1): Experiment-1 accuracies mean \neq Experiment-2 accuracies mean, indicating that they are not at the same level of performance.

The p-value of the sign test to compare Experiment-1 with Experiment-2 at significance level 0.05 equals to 0.0357, which leads to the rejection of the null hypothesis h0 and the accepting of the alternative hypothesis h1 thus confirming that using 1-gram based LogisticR classifier with *breaking news* yields a *statistically significant boost* in directional prediction accuracy for 30 DJI stocks compared to using the same classifier with *all* of the stock news every hour.

Beside evaluating directional accuracies, we also performed a simulated financial evaluation of the proposed trading system by back-testing its trades and accounting for its return on investment (ROI%) for a period of 6 months, between March 2014 and September 2014. In this simulation it assumed that no commissions or fees are charged in each trade. The system was

tested on the hourly stock chart for its trade entries and exists whenever: (1) there was a hourly Tweet volume breakout for Tweets matching a stock's symbol, with (2) a trade fired in the direction (e.g. a long/buy or short/sell trade) indicated by our classifier at the beginning of the following hour for a duration of 1 hour. For each company, we calculated the amounts for the total number of trades, gross profit points, gross loss points, and the total profit points alongside percentage of long positions (won%), and percentage of short positions (won%).

Table IV shows the results of the simulated back-test evaluations. The results shows that during this period our system was profitable overall on its recommended hourly trades with each stock symbol. For stocks, one point loss or gain equals one dollar loss or gain. In Table IV, the total value of gross profit points is \$775.49 and the average value of gross profit points is \$25.85. Similarly, the total value of gross loss points is -\$290.35, and the average value of gross loss points is -\$9.68. The total value of profit points is \$485.14 and the average value of total profit points is \$16.17.

Since each stock has a different stock price, we also performed simulated trading using a diversified portfolio based on equal exposure to risk or gains from each stock in order to calculate the total and monthly average return on investment (ROI%) using this trading system. The simulated trades show that, this system would trigger a total of 6,163 trades during these 6 months period with a winning ratio of 70% for its long/buy directional trades and 76% winning ratio for its short/sell directional trades. On the average this system triggers 35 monthly trades per stock. Trading with an equally diversified portfolio yields a total (ROI%) of 23% for 6 months, indicating an average monthly (ROI%) of 3.83%. The highest total (ROI%) achieved was 65% by executing 271 trades with

Microsoft (\$MSFT), and the lowest total (ROI%) was 1% by executing 63 trades with DuPont (\$DD) and 68 trades with United Technologies (\$UTX).

VI. CONCLUSION AND FUTURE WORK

In this paper we proposed a system to predict the hourly stock price direction based on the textual analysis of news articles' content mentioning a stock symbol. Firstly, we showed that using LogisticR classifier with 1-gram keyword features leads to the best overall directional prediction accuracy. Secondly, we showed that using extracted document-level sentiment features does not yield to a statistically significant boost in directional predictive accuracies in the presence of 1-gram keyword features selected by a *Chi2* [7] feature selection algorithm. Thirdly, we show that the breaking news based system indeed yields a statistically significant boost in prediction accuracy compared to the one using all news indiscriminately. As future work, we will perform several experiments by varying time frames (i.e. 1 minute, 5 minutes, 15 minutes, 30 minutes, 1 hours, 4 hours and daily) for both grouping tweets and detecting their breakouts, and for determining the best time-frame on stock price charts that leads to the highest predictive accuracies. Furthermore, inspired by related work which utilize technical stock price indicators, we also propose to utilize indicators such as price-level breakouts, moving averages and candle-stick patterns to test if they would lead to a statistically significant boost in predictive accuracies. We also plan to experiment with different categories of stocks (i.e. large cap, mid cap and small cap) to determine their sensitivities to online news and social-media breakouts.

REFERENCES

- [1] R. Compton, C. Lee, J. Xu, L. Artieda-Moncada, T.-C. Lu, L. Silva, and M. Macy, "Using publicly visible social media to build detailed forecasts of civil unrest," *Security Informatics*, vol. 3, no. 1, 2014.
- [2] X. Wang, D. Brown, and M. Gerber, "Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information," in *International Conference on Intelligence and Security Informatics*, ser. Lecture Notes in Computer Science, IEEE Press. IEEE Press, 2012.
- [3] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, ser. WI-IAT '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 492–499.
- [4] H. Achrekar, A. Gandhe, R. Lazarus, S. Yu, and B. Liu, "Twitter improves seasonal influenza prediction," in *HEALTHINF 2012 - Proceedings of the International Conference on Health Informatics, Vilamoura, Algarve, Portugal, 1 - 4 February, 2012.*, 2012, pp. 61–70.
- [5] F. Pukelsheim, "The three sigma rule," *The American Statistician*, vol. 48, no. 2, pp. 88–91, 1994.
- [6] J. Liu, S. Ji, and J. Ye, "Slep: Sparse learning with efficient projections," *Arizona State University*, vol. 6, p. 491, 2009.
- [7] H. Liu and R. Setiono, "Feature selection via discretization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 9, no. 4, pp. 642–645, 1997.
- [8] J. Patel, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162–2162–2172; 2172, 2015, doi: 10.1016/j.eswa.2014.10.031 pmid:
- [9] J. Gong and S. Sun, "A new approach of stock price prediction based on logistic regression model," in *New Trends in Information and Service Science, 2009. NISS '09. International Conference on*, 2009, pp. 1366–1371, iD: 1.
- [10] S. S. Roy, D. Mittal, A. Basu, and A. Abraham, "Stock market forecasting using lasso linear regression model," in *Afro-European Conference for Industrial Advancement*. Springer, 2015, pp. 371–381.
- [11] A. K. Nassiroussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, "Text mining of news-headlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment," *Expert Systems with Applications*, vol. 42, no. 1, pp. 306–324, 1 2015.
- [12] M. Hagenau, L. Michael, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," vol. 55, no. 3, /2013 06, pp. 685; 685–697; 697, doi: 10.1016/j.dss.2013.02.006 pmid:
- [13] M. Y. Kaya and M. E. Karsligil, "Stock price prediction using financial news articles," in *Information and Financial Engineering (ICIFE), 2010 2nd IEEE International Conference on*. IEEE, 2010, pp. 478–482.
- [14] R. Schumaker and H. Chen, "Textual analysis of stock market prediction using financial news articles," *AMCIS 2006 Proceedings*, p. 185, 2006.
- [15] S. Lauren and S. D. Harlili, "Stock trend prediction using simple moving average supported by news classification," in *Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 International Conference of*. IEEE, 2014, pp. 135–139.
- [16] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
- [17] T.-T. Vu, S. Chang, Q. T. Ha, and N. Collier, "An experiment in integrating sentiment features for tech stock prediction in twitter," 2012.
- [18] Y. Mao, W. Wei, B. Wang, and B. Liu, "Correlating s&p 500 stocks with twitter data," in *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*. ACM, 2012, pp. 69–72.
- [19] Y. Mao, W. Wei, and B. Wang, "Twitter volume spikes: analysis and application in stock trading," in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. ACM, 2013, p. 4.
- [20] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 1, pp. 163–173, 2012.
- [21] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
- [22] E. L. Lehmann, *Nonparametrics :statistical methods based on ranks*. San Francisco: Holden-Day, e. L. Lehmann, with the special assistance of H. J. M. D'Abreu.; ;24 cm; Includes bibliographical references and index.