Analyzing Stock Market Movements Using Twitter Sentiment Analysis

Tushar Rao
NSIT, Delhi, India
Email: rao.tushar@nsitonline.in

Saket Srivastava
IIIT-Delhi, India
Email: saket@iiitd.ac.in

***Abstract*—In this paper we investigate the complex relationship between tweet board literature (like bullishness, volume, agreement etc) with the financial market instruments (like volatility, trading volume and stock prices). We have analyzed sentiments for more than 4 million tweets between June 2010 to July 2011 for DJIA, NASDAQ-100 and 13 other big cap technological stocks. Our results show high correlation (upto 0.88 for returns) between stock prices and twitter sentiments. Further, using Granger's Causality Analysis, we have validated that the movement of stock prices and indices are greatly affected in the short term by Twitter discussions. Finally, we have implemented Expert Model Mining System (EMMS) to demonstrate that our forecasted returns give a high value of R-square (0.952) with low Maximum Absolute Percentage Error (MaxAPE) of 1.76% for Dow Jones Industrial Average (DJIA).**

***Keywords*-Stock market ; sentiment analysis ; Twitter ; microblogging ; social network analysis**

I. INTRODUCTION

Before the emergence of internet, information regarding company's stock price, direction and general sentiments took a long time to disseminate among people. Also, the companies and markets took a long time (weeks or months) to calm market rumors, news or false information (memes in Twitter context). This era of web technology is marked with fast pace information dissemination as well as retrieval [1]. Spreading good or bad information regarding a particular company, product, person etc. can be done at the click of a mouse [2] or even using micro-blogging services such as Twitter. In this age of fast paced information dissemination [3], *short term sentiments* play a very important role in *short term performance* of financial market instruments such as indexes, stocks and bonds.

It is well accepted that *news drive macro-economic movement in the markets*, while researches suggests that *social media buzz is highly influential at micro-economic level*, specially in the financial markets [4], [5], [6], [7]. In this work we have applied simplistic message board approach by defining bullishness and agreement terminologies derived from positive and negative vector ends of public sentiment w.r.t. each market security or index terms (such as returns, trading volume and volatility). This method is not only scalable but also gives more accurate measure of large scale investor sentiment that can be potentially used for short term hedging strategies as discussed ahead. This gives clear distinctive way for modeling sentiments for service based companies such as Google in contrast to product based companies such as Ebay, Amazon and Netflix. *The aim of this work*, is to quantitatively evaluate the *effects of twitter sentiment dynamics* around a stocks indices/stock prices and use it in conjunction with the *standard* model to improve the accuracy of prediction.

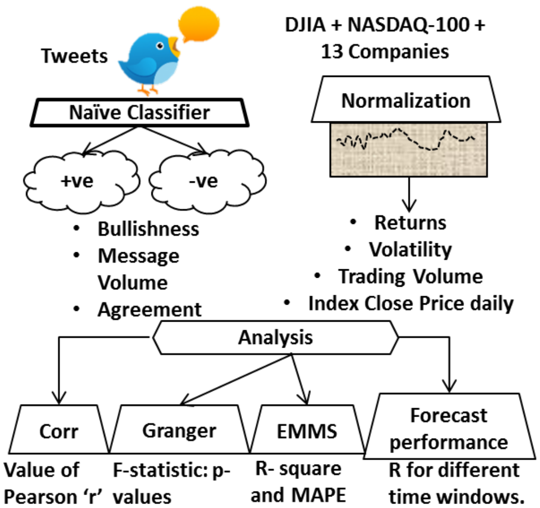

Figure 1. Flowchart of the proposed methodology. Four set of results obtained (1) Correlation results for twitter sentiments and stock prices for different companies (2) Granger's casuality analysis to causation (3) Using EMMS for quantitative comparison (4) Performance of forecasting method over different time windows

II. RELATED WORK

Gilbert et al. and Zhang et al. have used corpus from livejournal blogposts in assessing the bloggers sentiment in dimensions of fear , anxiety and worry making use of Monte Carlo simulation to reflect market movements in S&P 500 index [6], [8]. Similar and significantly accurate work is done by Bollen et al who used dimensions of GPOMS to reflect changes in closing price of DJIA [4]. Sprengers et al. analyzed individual stocks for S&P 100 companies and tried correlating tweet features about discussions of the stock discussions about the particular companies containing the Ticker symbol [7]. However this work brings new insights to exploit public sentiment to make successful hedging strategies.

III. WEB MINING AND DATA PROCESSING

In this section we describe our method of Twitter and financial data collection as summarized in Figure 1.

A. Tweets Collection and Processing

Tweets are accessible through a simple search of requisite terms through an application programming interface (API)[1]. Currently more than 250 million messages are posted on

[1]Twitter API is easily accessible at- https://dev.twitter.com/docs.

Twitter everyday (Techcrunch October 2011[2]). This study was conducted over a period of 14 months period between June 2nd 2010 to 29th July 2011. During this period, we collected 4,025,595 (by around 1.08M users) English language tweets Each tweet record contains (a) tweet identifier, (b) date/time of submission(in GMT), (c) language and (d) text. We have directed our focus DJIA, NASDAQ-100 and 13 major companies listed in Table I. These companies are some of the highly traded and discussed technology stocks having very high tweet volumes.

B. Sentiment Classification

In order to compute sentiment for any tweet we had to classify each incoming tweet everyday into *positive* or *negative* using nave classifier. For each day total number of positive tweets is aggregated as ${M_t}^{Positive}$ while total number of negative tweets as ${M_t}^{Negative}$. We have made use of JSON API from Twittersentiment [3], a service provided by Stanford NLP research group [9]. Online classifier has made use of Naive Bayesian classification method, which is one of the successful and highly researched algorithms for classification giving superior performance to other methods in context of tweets. Their classification training is done over a dataset of 1,600,000 tweets and achieved an accuracy of about 82.7% [9]. In our tweet dataset roughly 61.68% of the tweets are positive, while 38.32% of the tweets are negative for the company stocks under study. The ratio of 3:2 indicates stock discussions to be much more balanced in terms of bullishness than internet board messages where the ratio of positive to negative ranges from 7:1 [10] to 5:1 [11]; provides us with more confidence to study information content of discussions about the stock prices on microblogs.

C. Tweet Feature Extraction

One of the research questions this study explores is how investment decisions for technological stocks are affected by entropy of information spread about companies under study in the virtual space. We have only aggregated the tweet parameters (extracted from tweet features) over a day. In order to calculate parameters weekly, bi-weekly, tri-weekly, monthly, 5 weekly and 6 weekly we have taken average of daily twitter feeds over the specific period of time.

Twitter literature in perspective of stock investment is summarized in Figure 1. We have carried forward work of Antweiler et al. for defining bullishness (B_t) for each day (or time window) given as:

$$B_t = \ln\left(\frac{1+{M_t}^{Positive}}{1+{M_t}^{Negative}}\right) \tag{1}$$

Where ${M_t}^{Positive}$ and ${M_t}^{Negative}$ represent number of positive or negative tweets on a particular day t. Logarithm of bullishness measures the share of surplus positive signals and also gives more weight to larger number of messages in a specific sentiment (positive or negative). Message volume for a time interval t is simply defined as natural logarithm of total number of tweets for a specific stock/index which is $\ln({M_t}^{Positive}+{M_t}^{Negative})$. The agreement among positive and negative tweet messages is defined by:

$$A_t = 1 - \sqrt{1 - \frac{M_t^{Positive} - M_t^{Negative}}{M_t^{Positive} + M_t^{Negative}}} \tag{2}$$

If all tweet messages about a particular company are bullish or bearish, agreement would be 1 in that case. Influence of silent tweets days in our study (trading days when no tweeting happens about particular company) is less than 0.1% which is significantly less than previous research [11], [7]. Carried terminologies for all the tweet features{Positive, Negative, Bullishness, Message Volume, Agreement} remain same for each day with the lag of one day. For example, carried bullishness for day d is given by $CarriedBullishness_{d-1}$.

D. Financial Data Collection

We have downloaded financial stock prices at daily intervals from Yahoo Finance API[4] for DJIA, NASDAQ-100 and the companies under study given in Table I. The financial

Table I
LIST OF COMPANIES

Company Name	Ticker Symbol
Amazon	AMZN
Apple	AAPL
AT&T	T
Dell	DELL
EBay	EBAY
Google	GOOG
IBM	IBM
Intel	INTC
Microsoft	MSFT
Oracle	ORCL
Samsung Electronics	SSNLF
SAP	SAP
Yahoo	YHOO

features (parameters) under study are closing (C_t) value of the stock/index and the returns. Returns are calculated as the difference of logarithm to the base e between the closing values of the stock price of a particular day and the previous day.

$$R_t = \{\ln Close_{(t)} - \ln Close_{(t-1)}\} \times 100 \tag{3}$$

Trading volume is the logarithm of number of traded shares. We estimate daily volatility based on intra-day highs and lows using Garman and Klass volatility measures [12] given by the formula:

$$\sigma = \sqrt{\frac{1}{n}\sum \frac{1}{2}[\ln\frac{H_t}{L_t}]^2 - [2\ln 2 - 1][\ln\frac{C_t}{O_t}]^2} \tag{4}$$

IV. STATISTICAL ANALYSIS AND RESULTS

We begin our study by identifying the correlation between the Twitter feed features and stock/index parameters which give the encouraging values of statistically significant relationships with respect to individual stocks(indices).

[2]http://techcrunch.com/2011/10/17/twitter-is-at-250-million-tweets-per-day/

[3]https://sites.google.com/site/twittersentimenthelp/

[4]http://finance.yahoo.com/

Table II
CORRELATION (LINEAR: PEARSON-'R') MATRIX FOR STOCK/INDEX FEATURES VS TWITTER SENTIMENT FEATURES (STATISTICALLY SIGNIFICANT VALUES ARE BOLDED)

Stock Index →		NASDAQ	DJIA	AMZN	AAPL	T	DELL	EBAY	GOOG	IBM	INTC	MSFT	ORCL	SSNLF	SAP	YHOO
Close (C_t) Vs	**Positive**	0.01	0.22	-0.14	**-0.73**	**-0.67**	**-0.65**	**-0.81**	0.48	0.10	0.20	-0.27	**-0.79**	**-0.62**	0.04	**-0.70**
	Negative	-0.26	0.27	**-0.50**	**-0.64**	**-0.71**	**-0.81**	**-0.80**	0.49	-0.24	0.36	-0.43	**-0.95**	**-0.50**	**0.82**	-0.36
	Bullishness	-0.04	-0.37	**0.54**	0.18	0.21	0.21	**-0.64**	**-0.52**	**0.51**	-0.01	0.05	0.46	**-0.59**	-0.40	**-0.61**
	Carried Positive	0.06	0.27	-0.14	**-0.72**	**-0.61**	**-0.58**	**-0.78**	**-0.61**	0.13	0.20	-0.31	**-0.73**	**-0.58**	0.06	**-0.72**
	Carried Negative	-0.26	0.27	-0.48	**-0.64**	**-0.65**	**-0.77**	**-0.74**	-0.43	-0.08	0.37	-0.49	**-0.91**	**-0.37**	**0.84**	-0.49
	Carried Bullishness	-0.05	-0.31	**0.54**	-0.40	0.11	0.22	**-0.68**	**-0.52**	0.19	0.03	0.02	0.35	**-0.58**	-0.41	**-0.64**
Return Vs	**Bullishness**	**0.61**	**0.88**	0.10	0.04	-0.08	**0.65**	-0.01	0.47	-0.05	0.25	0.45	0.14	0.39	0.14	-0.41
	Agreement	0.45	**0.79**	0.22	-0.02	0.11	**0.52**	0.02	**0.50**	-0.06	0.18	0.46	0.11	0.38	0.10	-0.40
	Carried Bullishness	**0.60**	**0.84**	0.06	0.02	-0.16	**0.56**	0.05	0.44	0.06	0.22	0.42	0.23	0.31	0.07	**-0.55**
	Carried Agreement	-0.36	**0.76**	0.18	0.17	0.20	0.42	0.05	0.34	0.06	0.16	0.46	0.17	0.28	0.06	**-0.51**
Volatility Vs	**Bullishness**	-0.33	**-0.63**	**0.73**	-0.28	0.27	0.38	0.46	-0.14	-0.39	0.12	**0.70**	0.10	**-0.57**	-0.07	-0.17
	Agreement	-0.45	**-0.65**	**0.71**	-0.33	-0.34	**0.54**	0.48	-0.11	-0.42	0.14	**0.75**	0.17	**-0.57**	-0.07	-0.48
	Message Volume	-0.26	**0.77**	**-0.52**	**0.78**	**0.59**	0.45	0.40	**0.51**	0.12	0.30	**0.84**	0.20	-0.46	0.31	**-0.65**
	Carried Bullishness	-0.32	**-0.50**	**0.74**	-0.03	0.38	0.46	0.39	-0.18	-0.46	0.13	**0.70**	0.21	**-0.57**	-0.05	-0.35
	Carried Agreement	0.38	**-0.53**	**0.73**	-0.30	-0.45	**0.54**	0.40	-0.01	-0.47	0.17	**0.76**	0.26	**-0.57**	-0.05	**-0.58**
	Carried Message Volume	-0.37	**0.74**	**-0.65**	**0.77**	**0.63**	**0.50**	0.44	**0.53**	0.12	0.31	**0.81**	0.21	-0.47	0.34	**-0.69**

A. Correlation Matrix

For the stock indices DJIA and NASDAQ and 13 tech companies under study we have come up with the correlation matrix given in Table II between the financial market and Twitter sentiment features as explained in earlier section. The time period under study is monthly average as it the most accurate time window that gives competently significant values as compared to other time windows which is discussed later section IV-D.

Our approach shows strong correlation values between various features (upto 0.88 for returns from DJIA index etc.) and the average value of correlation between various features is around 0.5. Comparatively highest correlation values from earlier work has been around 0.41 [7]. As the relationships between the stock(index) parameters and Twitter features show different behavior in magnitude and sign for different stocks(indices), a uniform standardized model would not applicable to all the stocks(indices). Therefore, building an individual model for each stock(index) is the correct approach for finding appreciable insight into the prediction techniques. Returns are mostly correlated to same day bullishness by 0.61 and by lesser magnitude 0.6 for the carried bullishness for DJIA. Volatility is again dependent on most of the Twitter features, as high as 0.77 for same day message volume for NASDAQ-100. One of the *anomalies* that we have observed is that EBay gives negative correlation between the all the features due to heavy product based marketing on Twitter which turns out as not a correct indicator of average growth returns of the company itself.

B. Bivariate Granger Causality Analysis

Granger Causality Analysis (GCA) is not used to establish causality, but as an economist tool to investigate a statistical pattern of lagged correlation. A similar observation that the clouds precede rain is widely accepted. GCA rests on the assumption that if a variable X causes Y then changes in X will be systematically occur before the changes in Y. We realize lagged values of X shall bear significant correlation with Y. However correlation is not necessarily behind causation. We have made use of GCA in similar fashion as [4], [6] This is to test if one time series is significant in predicting another time series. Let returns R_t be reflective of fast movements in the stock market. To verify the change in returns with the change in Twitter features we compare the variance given by following linear models-

$$R_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \epsilon_t \tag{5}$$

$$R_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \sum_{i=1}^{n} \gamma_i X_{t-i} + \varepsilon_t \tag{6}$$

Equation 5 uses only 'n' lagged values of R_t , i.e. ($R_{t-1}, ..., R_{t-n}$) for prediction, while Equation 6 uses the n lagged values of both R_t and the tweet features time series given by $X_{t-1}, ..., X_{t-n}$. We have taken weekly time window to validate the casuality performance, hence the lag values [5]. will be calculated over the weekly intervals $1, 2, ..., 7$. From the Table III, we can reject the null hypothesis (H_o) that *the Twitter features do not affect returns in the financial markets* i.e. $\beta_{1,2,....,n} \neq 0$ with a high level of confidence (high p-values). However as we see the result applies to only specific negative and positive tweets. Other features like agreement and message volume do not have significant casual relationship with the returns of a stock index (low p-values).

[5] *lag at k* for any parameter M at x_t week is the value of the parameter prior to x_{t-k} week. For example, value of returns for the month of April, at the lag of one month will be $return_{april-1}$ which will be $return_{march}$

Table III
GRANGER'S CASUALITY ANALYSIS OF DJIA AND NASDAQ FOR 7 WEEK LAG TWITTER SENTIMENTS(** FOR P-VALUE < 0.05 AND * FOR P-VALUE < 0.1 WHICH IS 95% AND 90% CONFIDENCE INTERVAL RESPECTIVELY)

Stock Indices ↓	Lag	Positive	Negative	Bullishness	Agreement	Message Volume	Carried Positive	Carried Negative	Carried Bullishness	Carried Agreement	Carried Message Volume
DJIA	1 Week	0.614	0.122	0.891	0.316	0.765	0.69	0.103	0.785	0.759	0.934
	2 Week	0.033**	0.307	0.037**	0.094*	0.086**	0.032**	0.301**	0.047**	0.265	0.045**
	3 Week	0.219	0.909	0.718	0.508	0.237	0.016**	0.845	0.635	0.357	0.219
	4 Week	0.353	0.551	0.657	0.743	0.743	0.116	0.221	0.357	0.999	0.272
	5 Week	0.732	0.066	0.651	0.553	0.562	0.334	0.045**	0.394	0.987	0.607
	6 Week	0.825	0.705	0.928	0.554	0.732	0.961	0.432	0.764	0.261	0.832
	7 Week	0.759	0.581	0.809	0.687	0.807	0.867	0.631	0.987	0.865	0.969
NASDAQ- 100	1 Week	0.106	0.12	0.044**	0.827	0.064*	0.02**	0.04**	0.043**	0.704	0.071*
	2 Week	0.048**	0.219	0.893	0.642	0.022**	0.001**	0.108	0.828	0.255	0.001**
	3 Week	0.06*	0.685	0.367	0.357	0.135	0.01**	0.123	0.401	0.008**	0.131
	4 Week	0.104	0.545	0.572	0.764	0.092*	0.194	0.778	0.649	0.464	0.343
	5 Week	0.413	0.997	0.645	0.861	0.18	0.157	0.762	0.485	0.945	0.028
	6 Week	0.587	0.321	0.421	0.954	0.613	0.795	0.512	0.898	0.834	0.591
	7 Week	0.119	0.645	0.089	0.551	0.096	0.382	0.788	0.196	0.648	0.544

C. EMMS Model for Forecasting

We have used Expert Model Mining System (EMMS) which incorporates a set of competing methods such as Exponential Smoothing (ES), Auto Regressive Integrated Moving Average (ARIMA) and seasonal ARIMA models. In this work, selection criterion for the EMMS is coefficient of determination (R squared) which is square of the value of pearson-'r' of fit values (from the EMMS model) and actual observed values. Mean absolute percentage error (MAPE) and maximum absolute percentage error (MaxPAE) are mean and maximum values of error (difference between fit value and observed value in percentage). To show the performance of tweet features in prediction model, we have applied the EMMS twice - first with tweets features as independent predictor events and second time without them. This provides us with a quantitative comparison of improvement in the prediction using tweet features.

Table IV
EMMS MODEL FIT CHARACTERISTICS FOR DJIA AND NASDAQ-100

Index	Predictors	Model Fit statistics			Ljung-Box Q(18)		
		R-squared	MaxAPE	Direction	Statistics	DF	Sig.
Dow-30	Yes	0.95	1.76	90.8	11.36	18	0.88
	No	0.92	2.37	60	9.9	18	0.94
NSDQ-100	Yes	0.68	2.69	82.8	23.33	18	0.18
	No	0.65	2.94	55.8	16.93	17	0.46

In the dataset we have time series for a total of approximately 60 weeks (422 days), out of which we use approximately 75% i.e. 45 weeks for the training both the models with and without the predictors for the time period June 2nd 2010 to April 14th 2011. Further we verify the model performance as one step ahead forecast over the testing period of 15 weeks from April 15th to 29th July 2011 which count for wide and robust range of market conditions. Forecasting accuracy in the testing period is compared for both the models in each case in terms of maximum absolute percentage error (MaxAPE), mean absolute percentage error (MAPE) and the direction accuracy. MAPE is given by the equation 7, where $\hat{y}_i$ is the predicted value and y_i is the actual value.

$$MAPE = \frac{\Sigma^{n}{}_{i} \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \times 100 \tag{7}$$

While direction accuracy is measure of how accurately market or commodity up/ down movement is predicted by the model, which is technically defined as logical values for $(y_{i,\hat{t}+1} - y_{i,t}) \times (y_{i,t+1} - y_{i,t}) > 0$ respectively.

As we can see in the Table IV, there is significant reduction in MaxAPE for DJIA(2.37 to 1.76) and NASDAQ-100 (2.96 to 2.69) when EMMS model is used with predictors as events which in our case our all the Tweet features (positive, negative, bullishness, message volume and agreement). There is significant decrease in the value of MAPE for DJIA which is 0.8 in our case than 1.79 for earlier approaches [4]. As we can from the values of R-square, MAPE and MaxAPE in Table IV for both DJIA and NASDAQ 100, our proposed model uses Twitter sentiment analysis for a superior performance over traditional methods. Figures 2 shows the EMMS model fit for weekly closing values for DJIA and NASDAQ 100. In the figure fit are model fit values, observed are values of actual index and UCL & LCL are upper and lower confidence limits of the prediction model.

D. Prediction Accuracy using OLS Regression

Our results in the previous section showed that forecasting performance of stocks/indices using Twitter sentiments varies for different time windows. Hence it is important to quantitatively deduce a suitable time window that will give us most accurate prediction. Figure 3 shows the plot of R-square metric for OLS regression for returns from stock indexes NASDAQ-100 and DJIA from tweet board features (like number of positive, negative, bullishness, agreement and message volume) both for carried (at 1-day lag) and same week. From the figure 3 it can be inferred as we increase the time window the accuracy in prediction increases but only till a certain point that is monthly in our case beyond which value of R-square starts decreasing again. *Thus, for monthly predictions we have highest accuracy in predicting anomalies in the returns from the tweet board features.*

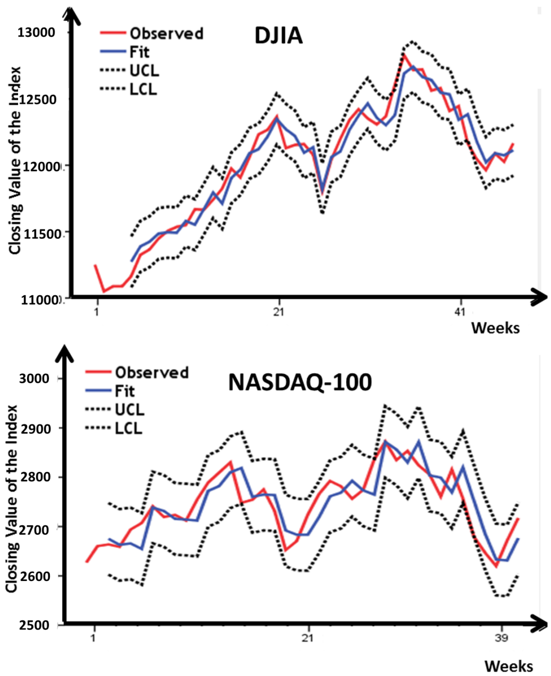

Figure 2. Plot of Fit values (from the EMMS model) and actual observed closing values for DJIA and NASDAQ-100

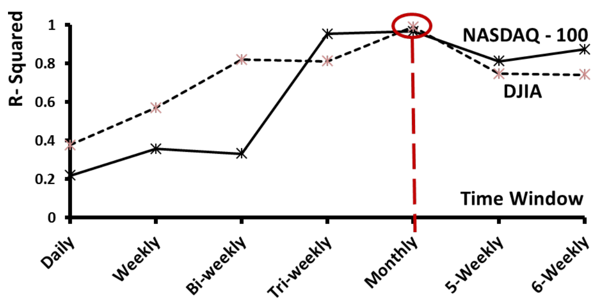

Figure 3. Plot of R-square values over different time windows for DJIA and NASDAQ-100. Higher values denote greater prediction accuracy.

V. Conclusion and Future Work

In this paper, we have worked upon identifying relationships between Twitter based sentiment analysis of a particular company/index and its short-term market performance using large scale collection of tweet data. Our results show that negative and positive dimensions of public mood carry strong cause-effect relationship with price movements of individual stocks/indices. We have also investigated various other features like how previous week sentiment features control the next week's opening, closing value of stock indexes for various tech companies and major index like DJIA and NASDAQ-100. Table V shows as compared to earlier approaches in the area which have been limited to wholesome public mood and stock ticker constricted discussions, we verify strong performance of our alternate model that captures mass public sentiment towards a particular index or company in scalable fashion and hence empower a singular investor to ideate coherent relative comparisons.

Table V
PRIOR RESEARCH IN SENTIMENT ANALYSIS FOR PREDICTING SENTIMENT ANALYSIS

Previous Approaches →	Bollen et al. [4] and Gilbert et al. [6]	Sprenger et al. [7]	Our Approach
Approach	Mood of complete Twitter feed	Stock Discussion with ticker $ on Twitter	Discussion based tracking of Twitter sentiments
Results	86.7% directional accuracy for DJIA	Corr values uptil 0.41 for S&P 100 stocks	Corr uptil 0.9, MAXPE 1.76 for DJIA and 2.69 for NASDAQ-100
Feedback/ Drawbacks	Individual modeling for stocks not feasible	News not taken into account, very less tweet volumes	Comprehensive and customizable approach. Can be used for hedging in F&O markets

Our analysis of individual company stocks gave strong correlation values (upto 0.88 for returns) with twitter sentiment features of that company. It is no surprise that this approach is far more robust and gives far better results (upto 91% directional accuracy) than any previous work.

References

[1] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[2] J. S. Brown and P. Duguid, *The Social Life of Information*. Boston, MA, USA: Harvard Business School Press, 2002.

[3] H. R. Liangfei Qiu and A. Whinston, "A twitter-based prediction market: Social network approach," *ICIS 2011 Proceedings. Paper 5*, 2011.

[4] J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," *Computer*, vol. 1010, no. 3003v1, pp. 1–8, 2010.

[5] H. Mao, S. Counts, and J. Bollen, "Predicting financial markets: Comparing survey,news, twitter and search engine data," arXiv.org, Quantitative Finance Papers 1112.1051, Dec. 2011.

[6] E. Gilbert and K. Karahalios, "Widespread worry and the stock market," *Artificial Intelligence*, pp. 58–65, 2010.

[7] T. O. Sprenger and I. M. Welpe, "Tweets and Trades: The Information Content of Stock Microblogs," *SSRN eLibrary*, 2010.

[8] X. Zhang, H. Fuehres, and P. A. Gloor, "Predicting stock market indicators through twitter i hope it is not as bad as i fear," *Anxiety*, pp. 1–8, 2009.

[9] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision."

[10] M. Dewally, "Internet investment advice: Investing with a rock of salt," *Financial Analysts Journal*, vol. 59, no. 4, pp. 65–77, 2003.

[11] M. Z. Frank and W. Antweiler, "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *SSRN eLibrary*, 2001.

[12] M. B. Garman and M. J. Klass, "On the estimation of security price volatilities from historical data," *The Journal of Business*, vol. 53, no. 1, pp. 67–78, 1980.