

**Problem Statement-**

We are addressing the problem of predicting the Stock Market behavior using StockTwits Sentiment and Posting Volume [1,2,3]. It has been well known that stock related decisions are influenced by people's emotions and moods [4]. We therefore consider this factor to predict the stock market. There are also many microblogging websites that provide an interface for users to share their views about the market. We therefore define indicators based on information from these microblogging websites and attempt to predict the market. We consider two important indicators namely daily returns, volatility and via the regression model predict these factors. We use sentiment data for predicting the daily returns whereas the predicting volatility is determined by volume indicators.

**Methodology:****Data/Framework:**

StockTwits is micro-blogging platform like twitter which is used by communities of financial domain. Messages pulled using StockTwits API service has special cashtags( like hashtag ) eg. \$GOOG for google, which helps to reduce noise. Open source R tool and Python parsers can be used to perform data and regression computations.

Price/Volume/Indexes data can be pulled from Thompson Reuters Datastream and Chicago Board Options Exchange.

**Steps:**

We are going to use multiple regression model whose parameter values will be adjusted by applying least square algorithm. Due to additive nature of this model, it is widely used in financial domain [1]. For each financial target ( volatility and daily returns) we are going to test three different regression models with new variations (\*- Table1).

For each stock we look for words “bullish” ( $Bull_t$  for day  $t$ ) or “bearish” ( $Bear_t$ ) [6] and derive sentiment indexes like  $bind_t$ ,  $TIS_t$  and volatility indexes like  $MA_t$ . We are considering following models for predicting returns ( $R^t$ ) and volatility ( $\ln(\hat{\sigma}_t)$ ) [1]

Predicting Returns ( $R^t$ )	Predicting Volatility( $\ln(\hat{\sigma}_t)$ )
$R^t = f(R_{t-1})$ (baseline)	$\ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1}))$ (baseline)
$R^t = f(R_{t-1}, \ln(TIS_{t-1}))$	$\ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1}), \ln(nt-1), \ln(nt-2))$
$R^t = f(bind_{t-1}, \ln(TIS_{t-1}))$ *	$\ln(\hat{\sigma}_t) = f(\ln(\sigma_{t-1}), \ln(nt-1), \ln(MA_{t-1}))$ *

Table 1:

**Evaluation:**

To evaluate the quality of above models, we are going to use Root-Mean-Squared Error (RMSE) error metric. The lower values RMSE indicate better prediction Model. For better evaluation we can consider

fix sized rolling window evaluation scheme. To evaluate prediction results we can apply the equality of prediction statistical test to consider pair wise comparison w.r.t baseline model [1,7]

#### **Citations:**

[1] Nuno Oliveira, Paulo Cortez, and Nelson Areal. On the Predictability of Stock Market Behavior using StockTwits Sentiment and Posting Value. Springer

[2] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1{8, 2011.

[3] C. Oh and O.R.L. Sheng. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. *ICIS 2011 Proceedings*, 2011

[4] J.R. Nofsinger. Social mood and financial economics. *The Journal of Behavioral Finance*, 6(3):144{160, 2005.

[5] Shimon Kogan, Dmitry Levin, Bryan R. Routledge, Jacob S. Sagi, Noah A. Smith. Predicting Risk from Financial Reports with Regression, *ACM 2009*

[6] Mao, H., Counts, S., Bollen, J.: Predicting financial markets: Comparing survey, news, twitter and search engine data. *arXiv preprint arXiv:1112.1051* (2011)

[7] Harvey, D., Leybourne, S., Newbold, P.: Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13(2), 281–291 (1997)