International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 20, Issue 4 (April. 2024), PP. 101-103

Deciphering Market Trends: Leveraging Machine Learning and Twitter Analysis for Stock Price Prediction

Dr Karanam Ramesh Rao
Professor , Dept of CSE-Data Science,
Sphoorthy Engineering College, Hyderabad

Abstract: *This paper explores the relationship between public sentiment and market dynamics by employing sentiment analysis and AI methodologies. Utilizing Twitter data, we gauge public mood and integrate it with historical DJIA values to forecast market trends. Validation of our approach reveals a 75% accuracy using Self-Organizing Fuzzy Neural Networks (SOFNN) on Twitter feeds and DJIA data from June 2009 to December 2009. Additionally, we implement a simplistic portfolio management strategy based on our predicted trends. Our study builds upon the work of Bollen et al., achieving comparable accuracy in forecasting stock market movements.*

Keywords: *Sentiment Analysis, Stock Market Forecast, Twitter Sentiment, Natural Language Processing (NLP)*

Date of Submission: 11-04-2024 Date of Acceptance: 23-04-2024

I. Introduction

Predicting financial market fluctuations has long been a subject of interest. The Efficient Market Hypothesis (EMH) suggests that market prices primarily reflect new information and follow a random walk pattern. Despite its widespread acceptance, some researchers have sought patterns in market behavior driven by human emotions and decision-making processes, positing a direct correlation between "public opinion" and "market sentiment". In this paper, we conduct sentiment analysis on Twitter data to discern public mood, categorizing it into four classes - Calm, Happy, Alert, and Kind, akin to fuzzy membership. We leverage these mood states along with past Dow Jones Industrial Average (DJIA) values to forecast future stock movements and devise a portfolio management strategy.

An approach builds on the methodology introduced by Bollen et al. [1], garnering significant media attention recently. Their study attempted to predict market behavior by analyzing Twitter users' sentiment. By examining tweet data from 2008 and employing the OpinionFinder and Google Profile of Mood States (GPOMS) algorithms, they categorized public sentiment into six classes: Calm, Alert, Confident, Vital, Negative, and Happy. Subsequently, they validated these sentiment time series against real events such as the U.S. elections and Thanksgiving Day in 2008. Causality analysis was also conducted to assess whether public mood states, as inferred by the OpinionFinder and GPOMS mood time series, could predict changes in DJIA closing values. Employing Self-Organizing Fuzzy Neural Networks, their results demonstrated an impressive accuracy of nearly 87% in predicting DJIA fluctuations. The remainder of this paper is structured as follows: Section 2 outlines our overall approach to addressing the problem, followed by detailed discussions of individual components in subsequent sections.

Algorithm: Our methodology closely follows the one employed by Bollen et al. [1]. Raw DJIA values are initially processed to obtain cleaned features, while tweets are fed into the sentiment analysis algorithm to generate mood values for each day. These mood states and processed DJIA values are then inputted into our machine learning framework, utilizing SOFNN to train a model for predicting future DJIA values. The learned model, in conjunction with past DJIA and mood values, informs the portfolio management system, which leverages predicted trends to make informed buy/sell decisions. Figure 1 illustrates a concise flowchart of our approach, with subsequent sections elaborating on each part of our methodology in greater detail.

Dataset: We utilize two primary datasets: Dow Jones Industrial Average (DJIA) values from June 2009 to December 2009, obtained from Yahoo! Finance, and publicly available Twitter data comprising over 476 million tweets from more than 17 million users during the same period. Given our continuous prediction and analysis, we partition tweets by day using timestamp information.

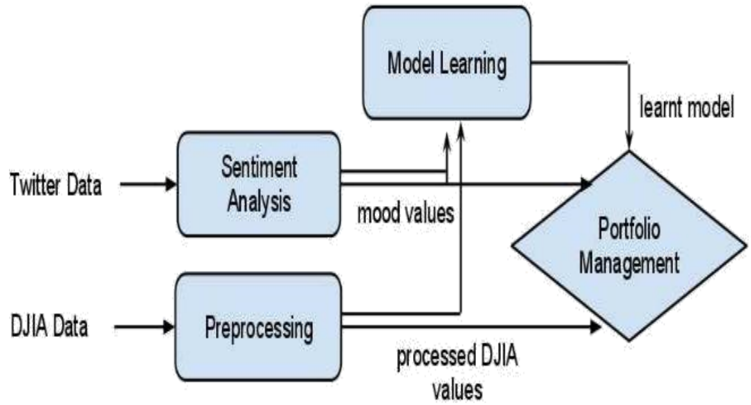

Figure 1: Our approach

Data Preprocessing: Data obtained from the aforementioned sources undergo preprocessing to facilitate meaningful analysis. We preprocess DJIA data by filling missing values for weekends and market closure days using interpolation. Additionally, to ensure data consistency and stability, we normalize values using z-scores. Opinion Analysis: Sentiment analysis forms a crucial component of our solution, as its outcome is utilized for training our predictive model. While various techniques exist for classifying text sentiment, we adopt a four-mood classification: Calm, Happy, Alert, and Kind. Despite the availability of established tools like OpinionFinder and SentiWordNet, we found them inadequate and opted to develop our sentiment analysis code.

Our sentiment analysis approach entails:

Word List Generation: We create a word list based on the Profile of Mood States (POMS) survey, a psychometric questionnaire assessing an individual's mood. This list is expanded using SentiWordNet and a standard thesaurus to ensure broad coverage. Tweet Filtering: Given the vastness of tweet data, we filter and consider only those tweets expressing sentiment, such as those containing phrases like "feel," "makes me," or "I'm." Daily Score Calculation: We employ a simple word counting algorithm to calculate scores for each POMS word for a given day. These scores are then mapped to four mood states using predetermined mapping techniques, with validation conducted around significant events to ensure accuracy.

$$\text{score card} = \frac{\text{\#of times the word matches tweets in a day}}{\text{\#of total matches of all words}}$$

Results: Our analysis reveals that SVMs and Linear Regression did not perform well, possibly due to classification issues.

Table.1 Accuracy direction

Sentiment Analysis Model	Mean Absolute %Error(MAPE)	Direction Accuracy
Linear Regression	7.05%-7.78%	64.44%-
Logistic Regression	N/A	60%
SVM	N/A	59.75%
SOFNN	9.22%-11.78%	64.44%-

However, Straight Regression exhibited significant performance, aligning with Granger Causality results. SOFNN outperformed other algorithms, particularly with a Direction Accuracy of 75.56% when considering Calm and Happy moods.

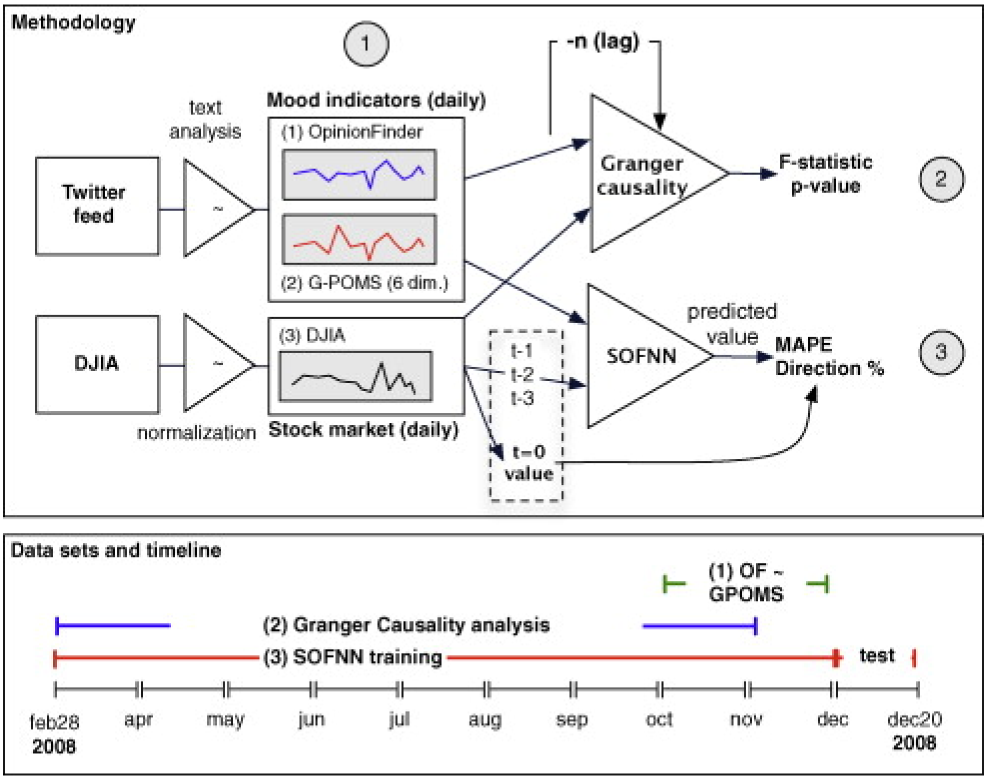

Figure 4: Anticipated versus Genuine StockQualities involving SOFNN onCalm+Happy+DJIAfor40SuccessiveDays

II. Conclusion:

Inferences drawn from our findings suggest that public mood, as inferred from Twitter data, can be effectively captured using straightforward NLP approaches. These deductions were validated against responses to significant socio-cultural events in 2009. Among the mood factors studied, only calmness and happiness exhibited Granger causation with DJIA values, indicating an impact within a three-day window. A Self-Organizing Fuzzy Neural Network (SOFNN) performed well in predicting DJIA values when trained on a feature set comprising DJIA values, calm mood, and happiness data from the past three days. Our study also demonstrated successful portfolio management using the system, yielding a profitable return over a 40-day period.

References

[1]. J. Bollen and H. Mao. Twitter thoughts- set as a economic change indicator. IEEE PC, 44(10):91-94.
[2]. C.- C. Chang and C.- J. Lin. LIBSVM:A library for help vector machines. ACM Exchanges on Wise Frameworks and Innovation, 2:27:1-27:27, 2011.
[3]. Three. G. P. Pack Leng and T. M. McGinnity. An on-line calculation formaking self-putting collectively fluffy brain agencies. Brain Organizations, 17(10):1477-1493.
[4]. Lapedes and R. Farber. Nonlinear signaldealing with making use of brain company: Forecast and framework demonstrating. In Los Alamos Public Lab Specialized Report.
[5]. E. Stefano Baccianella and F. Sebastiani. Sentiwordnet 3. Zero: An upgraded lexical asset for feeling exam and evaluation mining. In LREC. LREC.