

# Sentiment Analysis for Effective Stock Market Prediction

Aditya More

<sup>1</sup>*Department of Computer Science and Engineering, Lovely Professional University,  
Punjab, India*

---

**Abstract:** We analyze the impact of sentiment and attention variables on the stock market volatility by using a novel and extensive dataset that combines social media, news articles, information consumption, and search engine data. We apply a state-of-the-art sentiment classification technique in order to investigate the question of whether sentiment and attention measures contain additional predictive power for realized volatility when controlling for a wide range of economic and financial predictors. Using a penalized regression framework, we identify the most relevant variables to be investors' attention, as measured by the number of Google searches on financial keywords (e.g. "financial market" and "stock market"), and the daily volume of company-specific short messages posted on Stock Twits. In addition, our study shows that attention and sentiment variables are able to improve volatility forecasts significantly, although the magnitudes of the improvements are relatively small from an economic point of view.

**Keywords:** Machine learning, Sentiments Analysis, Stock market.

---

## Introduction

All Stock prediction is a challenging problem in the field of finance as well as engineering and mathematics. Due to its financial gain, it has attracted much attention both from academic side and business side. Stock price prediction has always been a subject of interest for most investors and financial analysts. Nevertheless, finding the best time to buy or sell has remained a very difficult task for investors because there are other numerous factors that may influence stock prices. In the prediction of stock market analysis a new novel approach is proposed to predict the buy or sell signal to the investors. Based on the combined result of opinion of sentences

Sentiment classification has become a very popular task in Machine Learning processing area, which tries to predict sentiment (opinion, emotion, etc.) from texts.

Today people are placing their comments and opinions on social media which can be shared by others also. Sentiment classification could be done in word/phrase level, sentence level and document level. Sentiment analysis has now become the dominant approach used for extracting sentiment and appraisals from online sources. Subjectivity

analysis focuses on dividing language units into two categories: objective and subjective, whereas sentiment analysis attempts to divide the language units into three categories; negative, positive, and neutral.

## Methodology

### Data Collection

The Dataset used for news sentiment was sourced from Kaggle. The datasets consists of Twenty Seven columns and 4100 rows. It having Twenty Five columns for top news of that day and other 2 column are date, labels(My graph is going upward or downward) and Stock data is of Dow Jones Industrial.

### Preprocessing

We used here some techniques to remove punctuation marks, to change text to lower, combining 25 columns text together (so that can perform bag of word techniques or Feature Extraction ),we split dataset into Train, Test after this all we applied feature extraction technique called **Count Vectorizer** use to convert text to vectors

## Model Training

For training Model here we used many Machine Learning Algorithms like RandomForest, Decision Tree SVM(Support Vector Machine), Xgboost, Voting Ensemble, Naïve Bayes. Each algorithm was trained on the training subset of the dataset using the extracted features, Hyperparameter tuning.

## Model Evaluation

The Trained models were evaluated on the test test to assess their performance in accurately classifying stock position. Evaluation metrics such as accuracy, confusion matrix, classification report.

### Accuracy:

accuracy = (number of correct predictions) / (total number of predictions)

### Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

### Precession

$$\text{Precision} = \frac{TP}{TP + FP}$$

### F1 Score:

$$F1 = \frac{2(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

# Result and Discussion

## Dataset

The Dataset used for news sentiment was sourced from Kaggle. The datasets consists of Twenty Seven columns and 4100 rows. It is having Twenty Five columns for top news of that day and other 2 column are date, labels(My graph is going upward or downward) and Stock data is of Dow Jones Industrial.

## Quantitative Results

The performance of Machine Learning algorithms namely RandomForest, Decision Tree SVM(Support Vector Machine), Xgboost, Voting Ensemble, Naïve Bayes was evaluate by stock market trend prediction from news sentiment tasks using a dataset sourced from Kaggle. After training and Training and testing the models on the datasets

The following metrics obtained are:

## Classification report

For 0	MNB	RFC	XGB	DT	LR	VEC
Precision	0.74	0.72	0.74	0.84	0.85	0.85
Recall	0.93	0.96	0.93	0.84	0.79	0.79
F1-Score	0.83	0.82	0.83	0.84	0.82	0.82

For 1	MNB	RFC	XGB	DT	LR	VEC
Precision	0.95	0.97	0.95	0.84	0.78	0.79
Recall	0.79	0.78	0.79	0.85	0.85	0.84
F1-Score	0.86	0.87	0.86	0.84	0.83	0.81

## Accuracy

	MNB	RFC	XGB	DT	LR	VEC
Accuracy	84.65	84.92	85.44	84.12	84.02	84.74

## Confusion Matrix:

	MNB	RFC	XGB	DT	LR	VEC
TT	138	134	138	157	159	158
TF	10	5	10	31	40	41
FT	48	52	48	29	28	28
FF	182	187	182	161	151	151

## Conclusion :

Unlike the conventional stock market prediction systems our novel approach combines the sentiments of common people through the news feeds and Sensex data to predict the behavior of stock market. The news feed

---

Aditya More

[adityamore4989@gmail.com](mailto:adityamore4989@gmail.com)

Source Code: [Code and Dataset](#)

of the stock related news are obtained for company and sentiment polarity of the news sentences are calculated for the prediction of stock news, whether it is positive, negative or neutral. The moving average stock level indicator is used to calculate the Sensex points for the stock market. Finally combination of sentiment polarity news and the Sensex points provide an efficient result to the stock market forecasters when to buy or sell their stocks.

## References :

1. P.Robert, Schumaker, and H.Chen, "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZF in Text System", *ACM Transactions on Information Systems*, Vol. 27, No. 2, pp. 1- 20, 2009.
2. W. Huang, K. K. Lai, Y.Nakamori, S. Wang, L. Yu, "Neural Networks in Finance and Economics Forecasting", *International Journal of Information Technology & Decision Making*, Vol. 6, No. 1, pp.113–140, 2007.
3. T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity: An Exploration of Features for Phrase- Level Sentiment Analysis", *Computational Linguistics*", Vol. 35, No. 3, pp. 399-433, 2009.
4. Kaggle :-  
<https://www.kaggle.com/datasets/aaron7sun/stocknews>
5. The Journal of finance and data science :-  
<https://www.sciencedirect.com/journal/the-journal-of-finance-and-data-science>

