Research Paper ICI Value : 61.33| SJIF Impact Factor(2019) : 8.045| ISI Value:1.433

Volume - 7, Issue- 8, August 2019 | e-ISSN : 2347 - 9671| p- ISSN : 2349 - 0187

EPRA International Journal of Economic and Business Review -Peer Reviewed Journal

SENTIMENT INTENSITY OF MARKET NEWS: IS IT A POSITIVE SIGNAL FOR PREDICTIVE ANALYSIS OF STOCK PRICES?

Leben Johnson Mannariat

PhD Scholar, Associate Professor & Coordinator GITAM FinTech Academy, GITAM (Deemed to be) University, Rushikonda, Vizag-530017, India

ABSTRACT

Technical Analysis: the prediction of stock prices using patterns in historical stock charts is nothing short of art. Quantitative analysts spend most of their time pouring over patterns to predict the future direction and price, to maximize profits or mitigate risk. Some analysts use these signals along with Fundamental Analysis and market news, to make trading decisions for short and long term trading strategies. Algorithm Trading is the process of using computerized mathematical models, based on certain patterns, rules and factors. With the exponential growth in Algorithm trading, currently capturing ~80% of market volume and the growth in computing power, sophisticated mathematical models is the order of the day. These strategies have grown over the period from simple Excel rule-based programs to Artificial Intelligence using deep neural networks.

Market news pertaining to the stock or macro economic factors, while having an impact on the stock price and important for trading decisions was always elusive to factor in quickly, due to it being voluminous and its inherent subjectivity in nature. Natural Language Processing (NLP) of news can provide a powerful signal that can augment trading decisions. This paper aims to explore the sensitivity intensity of relevant historical news through VADER NLP and use it as an additional signal to assess the excess predictive power over and above historical time-series stock price prediction, using neural network algorithms. This model is developed using Python and Tensorflow Keras neural network libraries.

KEYWORDS: *Natural Language Processing, Sentiment Analysis, Algorithm Trading, AI/ML, Stock Trading.*

1.1 INTRODUCTION

Tasseography is the art of reading tea leaves that interprets symbols, shapes and hidden messages in tea leaves, to read the past and predict the future, which is a dwindling art form. Technical Analysis is the art of looking at historical stock price patterns, to predict the future direction and price movement of the stock. There are various patterns that are well documented and established in enabling such predictions. To name a few; Bollinger Bands, Fibonacci Series, Chaikin Money Flow, Elliott Waves etc. Fundamental Analysis is the process of reviewing the annual statements (Balance Sheet, Income Statement, etc) of a company to develop ratios that can be compared with other peer companies to assess the relative performance of one company over the other or industry. Financial Analysts use Technical and Fundamental analysis to ascertain the future price movements of stocks and provide recommendations, which are used by Financial Traders to take appropriate positions in the stocks with an aim to mitigate risk and/or increase profitability.

Algorithmic (Algo) Trading is the process of using computing power to automatically trade a portfolio of stocks, based on the mathematical algorithms modelled on certain patterns, which have proved to be very profitable. As per CNBC[1] the daily volume of Algo trading can change according to volatility and can account for almost 80% of the daily moves on U.S. stocks. Most Algo trading models are mathematical models and hence interpret numbers to model patters and make suitable trading decisions.

Market news on the other had are subjective in nature, which needs to be interpreted by analysts/traders and digested before appropriate intervention is taken to mitigate risk or increase profitability. Such decisions move the stock price according to the impact of the news, which then translates into change in numbers that the models can interpret perfectly and appropriate trades initiated. However there is much to be desired in the speed of this human interpretation, digestion and intervention in the market. If a system can quickly interpret the news automatically, it would provide a much needed edge over competition, where everyone is Algo trading.

Artificial Intelligence / Machine Learning (AI/ML), which was introduced in the late 1980s, had little appetite then, due to the dearth and complexity of computing power. These days with data becoming the life-blood of the economy and computing power growing exponentially, AI/ML has taken centre stage with enormous amounts of open-source libraries being developed to address every need. Neural Networks (NN) is a computing program that optimally combines algorithms to mimic human neurons in the brain and make appropriate decisions based on the input factors. Natural Language Processing is a form of AI that interprets written verbose text and identifies the sentiment and intensity of the text, in terms of positivity or negativity of the content. This sensitivity analysis can be an additional powerful signal in predicting the direction of the stock price movement and the price itself.

This paper aims to capture the sensitivity analysis of relevant news of stocks and use that as a additional signal to determine the appropriate close price of stocks and assess the additional predictive power obtained by harnessing this NLP signal.

2.1 LITERATURE REVIEW

Moazzam Khoja[2] paper attempts to find the relationship between news informativeness and High Frequency Trading (HFT)s' long term bets. He confirms the behaviour in small cap stocks and also showed that, in short term trading they do not front run in small caps, unlike in large cap stocks. With this information, news when captured and when used at the earliest has an advantage in the short term bets; however the news interpretation is still manual. O'Hara[3] went into details to study the effects of HFT on the market microstructure. However the research does not delve into the impact of news in HFT decision making structure. Philip Tralveaven, et al[4] highlighted the importance of live data feeds scraped from various social news data sources, however their model does not seem to elude to the fact that news sentimental intensity is used in predicting the price of the stock or portfolio. Terrence Hendershott, et al[5] analyzed the impact of Algorithmic Trading (AT) on market liquidity and concluded that AT improves liquidity and suggested that it lowers the cost of trading. It however does not dwell in studying the impact of market news on discover of stock trading price. Adam Atkins, et al[6] dwelled into the process of modelling the semantics of news text using Latent Dirichlet Allocation (LDA) and used the DJI index to study the up and down movement of the index.

Hiransha M, et al[7] compared various Neural Networks such as Artificial Neural Networks (ANN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Multi Layer Perceptron (MLP) and Convolutional Neural Network (CNN) to measure and compare the Mean Absolute Percentage Error (MAPE) of prediction across various window sizes and prediction days. They concluded that Neural Networks had lower MAPE when compared to linear models like ARIMA. Adity Bhardwaj, et al[8] reviewed the various classification methods and used Beautiful Soup libraries in Python to perform sentiment analysis on live server data values at different intervals of time. Li Guo, et al[9] analyzed the process of analysing lexicon text using various classification methods and compared their performance. Based on their findings Neural Networks outperformed other ML techniques in classifying text into categories. Bruno Miranda Henrique, et al[10] used Support Vector Regression (SVR) models to measure performance of various Brazilian, Chinese and American stocks using Technical Analysis indicators like Simple Moving Average (SVA), Weighed Moving Average (WMA), Relative Strength Indicator (RSI) and Accumulated Distribution Oscillator (ADO) on various kernels. They were able to obtain smaller training errors when using linear kernel.

Xiandong Li, et al[11] used the Harvard psychological dictionary and Loughran–McDonald financial sentiment dictionary to build a sentiment dictionary. This was then used to label the news under positive, neutral and negative categories. Five years historical news was then used with some Hong Kong Stock Exchange listed stocks to predict their prices. They conclude that sentiment analysis improves the predictive accuracy of stocks and outperforms the bag-of-words model. Axel Groß-Klußmann, et al[12] measured the impact of intraday news on stock volatility employing a high-frequency Vector Autoregressive (VAR) model. This study established that intraday company-specific news has an impact on the high-frequency trading activity, even when the earnings announcements are discarded. Michael Hagenau, et al[13] have shown that using semantically relevant news features can reduce over-fitting, leading to highly profitable trading models.

3.1 RESEARCH METHODOLOGY

The following research methodology is used to validate the hypothesis, detailed explanation is provided in the subsequent sections.

- Python programming language was chosen for this analysis, due to its ease and the availability of vast open-source libraries that can help to perform machine learning and neural network analysis.
- The dataset was split between training and validation sets at a ratio of 0.80 to 0.20, to ensure there is adequate amount of data for the model to learn.
- As the various factors being considered are of different scales, they were then normalized to ensure uniformity between the independent variables.
- News was collected and sensitivity intensity collated day-wise, using VADER sensitivity analyzer and a news sensitivity intensity vector generated.
- To baseline the prediction, Python's Statsmodel OLS library was used for each of the stocks independently.
- The program was then executed in Tensorflow's keras.sequential neural network model with each stock along with the news vector and the Mean Square Error (MSE) was calculated.
- The program was then also executed in Tensorflow's keras.sequential neural network with other macro-economic factors such as DJIA market indices, Unemployment, Market Interest Rate, with and without news vector, for each of the stocks.
- Correlation between each of the macro-economic factors was checked to ensure multicollinearity is minimized.
- The overall R-squared factor for each of the runs was collated and improvement compared, which directly attributes to the improvement in the predictive power of the model.

4.1 DATA CONNECTIVITY & NORMALIZATION

Market closes prices were collected from Quandl[14], whose API interface to Python, makes it easy to collect data from their free datasets for individual research purpose. Along with close prices macro economic factors such as GDP, Interest Rates, Consumer Price Index (CPI), Dow Jones Industrial Average (DJIA) was collected to use as independent variables where appropriate to validate the strength of the predictive signals.

These collected data was subjected through a rigorous cleanup process to ensure the time-series were uniform and contiguous. Especially for the macroeconomic factors like GDP, unemployment and interest rate, which are available on a quarterly basis, CPI was available monthly; linear progression fills were used to align the data to the daily close price data of stocks. Missing Not-a-Number (NaN) values were appropriately filled up either with forward or backup numbers or using linear progression.

The data was then normalized using the following metrics.

Normalized Data = {(RawData – MeanOfRawData) / StanDevOfRawData}

This helps to ensure the normalized data from the various sources have a common scale. Each of the independent variable was checked for multi-collinearity, and it was noticed that GDP was highly correlated to CPI and hence was dropped, as can be seen in left plot of Figure-1.

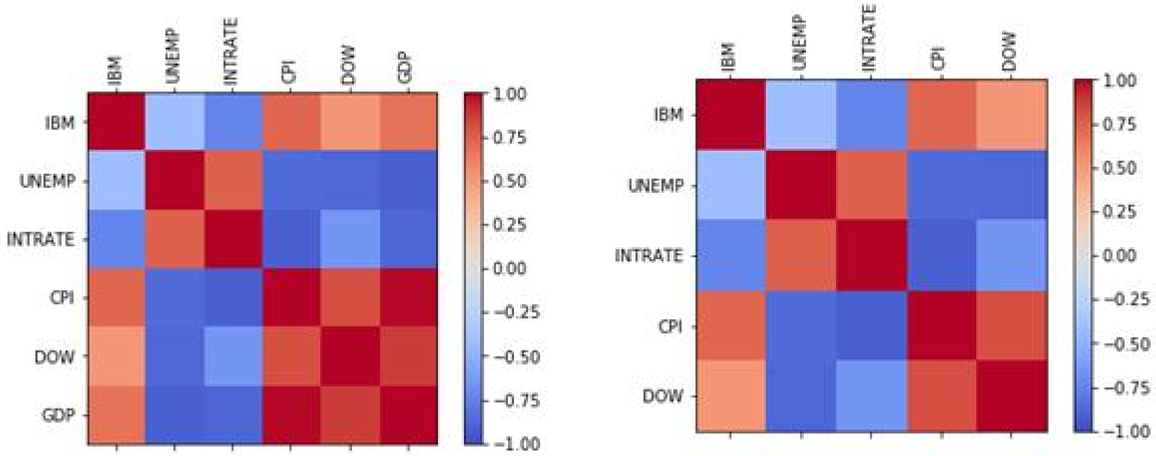

Figure-1: Correlation heatmap between dependent & independent variables

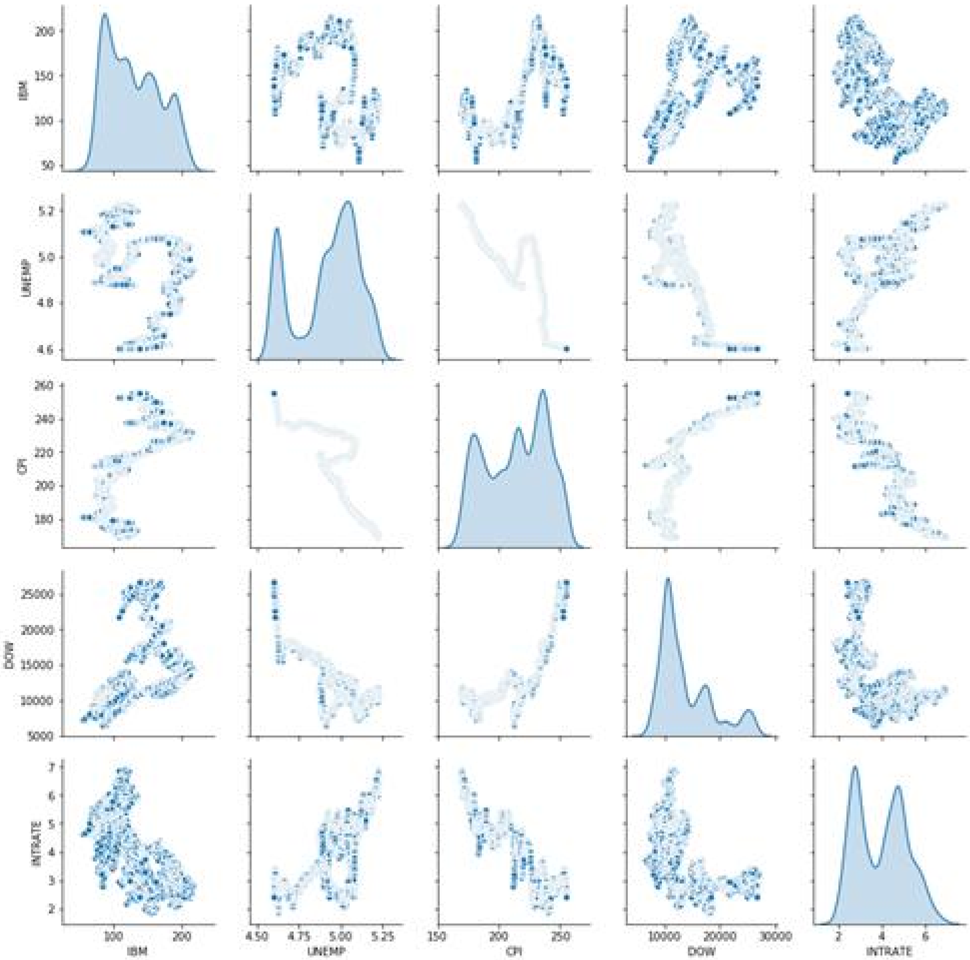

Figure-2: Correlation pairplots between dependent & independent variables

5.1 STOCK SELECTION

One of the important inputs that any machine learning algorithm needs is vast amounts of past historical data. As data was easily accessible for US market stocks, it was decided to pick stocks from the US market. With liquidity being another factor, it was decided to pick stocks from the constituents of Dow Jones Industrial Average (DJIA) index, spread across various sectors. Below are the set of 4 stocks that was picked to validate the model.

- International Business Machines (IBM)
- The Home Depot Inc (HD)
- The Procter & Gamble Company (PG)
- JP Morgan Chase & Co (JPM)

These stocks have been listed on the exchange prior to 1st Jan 2000 and hence covers the time horizon for testing and validation, which was decided to be from from 1st Jan 2000 until 30th June 2019, almost 20-years covering two recession periods of 2000 and 2008.

6.1 NATURAL LANGUAGE PROCESSING

Natural Language ToolKit (NLTK) is a set of libraries provided in Python for natural language processing. Its objects and methods can be used for classification, tokenization, parsing and semantic reasoning. Based on this it is able to provide sensitivity analysis of any textual data. For sensitivity analysis a valence aware module, ntlk.sentiment.vader[15] was used. VADER(Valence Aware Dictionary and sEntiment Reasoner) provides the sentiment of words according to semantic obtained by comparing with its word semantic library, as positive, negative, neutral and compound decimal score. This compound decimal score is the valence or intensity and varies between -1 (highly negative) to +1 (highly positive).

Beautifulsoup is a Python library that takes HTML or XML files, parses the content and provides data in just text. Also if needed any of the HTML tags can be used, for example: if you prefer to extract just the hyperlinks, it can be done. For the model the hypertext doc was converted into plain text format.

News for the stock being analyzed was collected using Thomson Reuters' Eikon API[16]. Each extraction is limited to a maximum of 150 days. In some cases where past news was available, looped the dates to get historical news, however now much could be obtained. The news extraction used the following filters to ensure the relevant news was extracted.

Topic:CEN AND (Topic:FRX OR Topic:ECI OR Topic:INT) AND R:IBM.N Language:LEN

Where; CEN: Central Bank Events, FRX: Currencies Foreign Exchange Market, ECI: Economic Indicators, INT: Interest Rate and the stock symbol in LEN:English. Reuters provides the capability to perform AND/OR/NOT operations for the filters, which was helpful to extract relevant news.

The news was then converted into a polarity score using the Sentiment Intensity Analyzer of VADER. The polarity scores are rated between positive, negative and neutral, depending on the proportion of texts that fall under each of the category, all summing up to 1. It also provides a compound score based on the overall normalized rating, which is a decimal number varying between -1 to +1. On a given day there could be multiple stories on the stock in question for which the cumulative compound sensitivity score is considered, or there could be none in which case it is taken as 0.

7.1 NEURAL NETWORKS

The neural network model was built using keras.sequential model with one input layer, one hidden layer and one output layer. The model was compiled to store Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared values, with the cost/loss function being MSE that was optimised to be minimized.

The model was then fit to the normalized train dataset and was iterated through 1,000 epoch for training the model with time-series historical data. The trained model was then evaluated against the test dataset and the predictions of the stock price captured and compared against the actual stocks prices during the test time period. The difference between the actual and predicted values is the error component, whose Mean and Standard Deviation were computed.

R-squared; which captures the explained deviation of the data points from the mean, increased significantly and reaches above 0.97 (Figure-3) with 1,000 epocs learning iterations, the rest 0.03 is the unexplained error. This R-squared plot captures all other independent variables like Unemp, IntRate, CPI, DOW alongwith relevant market News. Figure-4 the Statsmodel.api[17] regression summary sheet, shows the data collated without neural network learning.

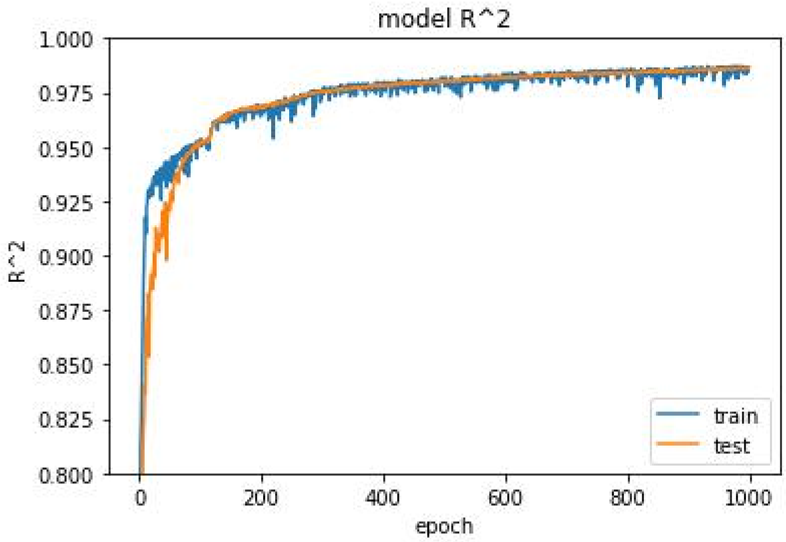

Figure-3: R-Squared values for IBM, with Unemp, IntRate, CPI, DOW & News

```
Model Summary For multi regression
                            OLS Regression Results
==============================================================================
Dep. Variable:                    IBM   R-squared:                       0.057
Model:                            OLS   Adj. R-squared:                  0.054
Method:                 Least Squares   F-statistic:                     17.26
Date:                Wed, 07 Aug 2019   Prob (F-statistic):           1.31e-16
Time:                        11:05:16   Log-Likelihood:                -8954.5
No. Observations:                1423   AIC:                         1.792e+04
Df Residuals:                    1418   BIC:                         1.795e+04
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
UNEMP         27.2988      8.220      3.321      0.001      11.174      43.424
INTRATE      -10.2052      8.879     -1.149      0.251     -27.622       7.212
CPI           37.7838     11.920      3.170      0.002      14.401      61.167
DOW            7.6517      7.701      0.994      0.321      -7.455      22.759
compound      -1.2353      4.118     -0.300      0.764      -9.313       6.842
==============================================================================
Omnibus:                       63.141   Durbin-Watson:                   0.001
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               26.797
Skew:                           0.033   Prob(JB):                     1.52e-06
Kurtosis:                       2.331   Cond. No.                         7.81
==============================================================================
```

Figure-4: OLS Summary for IBM, with Unemp, IntRate, CPI, DOW & compound News

So what can be inferred from the summary and the plot from two different sources measuring R-squared and P-values? While the P-values of IntRate, DOW and compound News are greater than the significance of 0.05, implying that these independent variables are inconsistent to predict the price, which is reaffirmed with a low Adj. R-squared value of 0.054. However they have proved different in the neural network plot, which has a very high R-squared of 0.97. Which helps to infer that multiple iterations of learning using neural networks has improved the predictive price of the stock, even with variables having a high P-value.

To further validate the findings, the same neural network code was executed multiple times by dropping the independent variables that have high P-values and the R-squared plots captured, alongwith the Statsmodel OLS regression model summary. A set of results are presented in the attached Figures-5 & 6. As can been seen that the R-squared values captured through Tensorflow's neural networks was still significant at above 0.95, while that captured through regular OLS regression was 0.056. This reaffirms that the multiple iterative learning process improves the predictive power of the model.

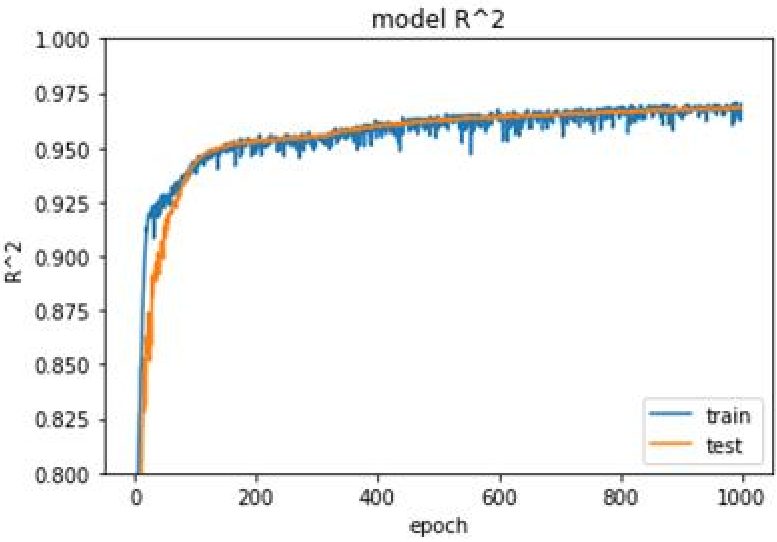

Figure-5: R-Squared values for IBM, with Unemp & CPI.

OLS Regression Results

Dep. Variable:	IBM	R-squared:	0.056
Model:	OLS	Adj. R-squared:	0.055
Method:	Least Squares	F-statistic:	42.27
Date:	Wed, 07 Aug 2019	Prob (F-statistic):	1.47e-18
Time:	10:24:01	Log-Likelihood:	-8955.4
No. Observations:	1423	AIC:	1.791e+04
Df Residuals:	1421	BIC:	1.793e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
UNEMP	23.4577	6.738	3.481	0.001	10.240	36.675
CPI	49.9874	6.782	7.370	0.000	36.683	63.292

Omnibus:	168.224	Durbin-Watson:	0.001
Prob(Omnibus):	0.000	Jarque-Bera (JB):	45.931
Skew:	0.042	Prob(JB):	1.06e-10
Kurtosis:	2.124	Cond. No.	3.60

Figure-6: OLS Model Summary values for IBM, with Unemp and CPI

8.1 NLP POTENTIAL PREDICTIVE POWER

Comparing Figures-1 & 3, it can be seen that the R-squared values have significantly improved with news (>0.975), when compared to without news (~0.95). It is important that news should be captured all through the period for which the backtest was conducted. However most unsubscribed sources did not have news archived dating back 20-years in history, hence its true potential is not fully harnessed.

The Predicted Error: difference between the predicted price and the true values during testing was computed and the errors plotted in Figure-7, showing the Scatter plot and it's Histogram.

The Predicted Error has a standard deviation of: 5.41 and a mean of: -0.78, which are very low, contributing to the high R-squared values shown earlier.

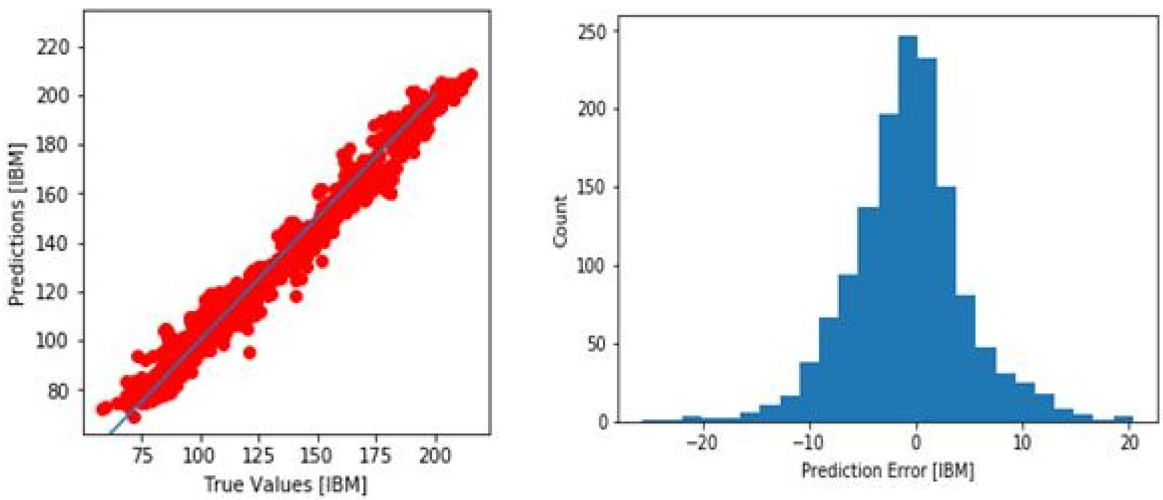

Figure-7: Predicted Error: - Predicted Price Vs True Values

9.1 CONCLUSION

Neural Networks has significantly improved the predictive power through multiple iterations that constantly improves on the cost function, when compared to standard regression. Natural Language Processing libraries have significantly enhanced the capability to perform sentimental analysis of any news that is published in the market. Historically there is a correlation established between stock volatility and price due to relevant news releases in the market.

Using the power of NLP sensitivity intensity of relevant news can be established, which was then harnessed using neural networks to improve the prediction of stock prices, as can be seen with the high R-squared value of > 0.975 (with news) and ~0.95 (without news). So going back to our title question of: "*Sentiment Intensity of Market News: Is it a Positive Signal for Predictive Analysis of Stock Prices?*" the answer is an absolute *"Yes".*

10.1 FUTURE RESEARCH

While this exercise was done based on historical stock prices, it would be very relevant to carry out research to establish the sensitivity intensity of real-time news and integrate it with algorithmic trading to quickly harness the power of news to monetize profits or mitigate risk, by executing instantaneous trades. This can further be extended to manage a portfolio of stocks, derivatives and/or commodities that can be integrated with real-time sentimental analysis of relevant market news of the instruments in the portfolio and take appropriate trading decisions.

REFERENCES

1. *Silivia Amaro, "Sell-offs could be down to machine that controls 80% of the US stock market, fund manager says", www.cnbc.com, 5th Dec 2018.*
2. *Moazzam Khoja, "Are High Frequency Traders independently informed? The study of the role of HFTs around Informational Events", March 6, 2019.*
3. *O'Hara, M., "High frequency market microstructure",.Journal of Financial Economics (2015), http: //dx.doi.org/10.10*
4. *Philip Treleaven, Michal Galas and Vidhi Lalchand, "Algorithmic Trading Review", Communications of the ACM, November 2013, Vol.56, No.11, Page:79.*
5. *Terrence Hendershott, Charled M. Jones and Albert J. Menkveld, "Does Algorithmic Trading Improve Liquidity?" The Journal of Finance, Vol. LXVI, No.1, February 2011.*
6. *Adam Atkings, Mahesan Niranjan, Enrico Gerding, "Financial news predicts stock market volatility better than close price", The Journal of Finance and Data Science, 4(2018) p120-137.*
7. *Hiransha M, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P, "NSE Stock Market Prediction Using Deep-Learning Models", Procedia Computer Science 132 (2018) 1351–1362.*
8. *Aditya Bhartwaj, Yogendra Narayan, Vanraj, Pawan, Maitreyee Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty", Procedia Computer Science 132 (2018) 1351–1362.*
9. *Li Guo, Feng Shi, Jun Tu, "Textual analysis and machine leaning: Crack unstructured data in finance and accounting", The Journal of Finance and Data Science 2 (2016) p153 170.*
10. *Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices", The Journal of Finance and Data Science 4 (2018) p183-201.*
11. *Xiandong Li, Haoran Xie, Li Chen, Jianping Wang, Xiaotie Deng, "News impact on stock price return via sentiment analysis", Knowledge-Based Systems 69 (2014) 14–23*
12. *Axel Groß-Klußmann, Nikolaus Hautsch, "When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions", Journal of Empirical Finance 18 (2011) 321–340*
13. *Michael Hagenau, Michael Liebmann, Dirk Neumann, " Automated news reading: Stock price prediction based on financial news using context-capturing features", Decision Support Systems, Volume 55, Issue 3, June 2013, Pages 685-697*
14. *Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.*
15. *Seabold, Skipper, and Josef Perktold. "Statsmodels: Econometric and statistical modeling with python." Proceedings of the 9th Python in Science Conference. 2010.*

END NOTES

1. *Silivia Amaro, "Sell-offs could be down to machine that controls 80% of the US stock market, fund manager says", www.cnbc.com, 5th Dec 2018.*
2. *Moazzam Khoja, "Are High Frequency Traders independently informed? The study of the role of HFTs around Informational Events", March 6, 2019*
3. *O'Hara, M., "Highfrequencymarketmicrostructure",.Journal of Financial Economics(2015), http: //dx.doi.org/10.10 There are no sources in the current document.16/ j.jfineco.2015.01.003*
4. *Philip Treleaven, Michal Galas and Vidhi Lalchand, "Algorithmic Trading Review", Communications of the ACM, November 2013, Vol.56, No.11, Page:79.*
5. *Terrence Hendershott, Charled M. Jones and Albert J. Menkveld, "Does Algorithmic Trading Improve Liquidity?"The Journal of Finance, Vol. LXVI, No.1, February 2011.*
6. *Adam Atkings, Mahesan Niranjan, Enrico Gerding, "Financial news predicts stock market volatility better than close price", The Journal of Finance and Data Science, 4(2018) p120-137.*
7. *Hiransha M, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P, NSE Stock Market Prediction Using Deep-Learning Models", Procedia Computer Science 132 (2018) 1351–1362.*
8. *Aditya Bhartwaj, Yogendra Narayan, Vanraj, Pawan, Maitreyee Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty", Procedia Computer Science 132 (2018) 1351–1362.*
9. *Li Guo, Feng Shi, Jun Tu, "Textual analysis and machine leaning: Crack unstructured data in finance and accounting", The Journal of Finance and Data Science 2 (2016) p153170.*
10. *Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices", The Journal of Finance and Data Science 4 (2018) p183-201.*
11. *Xiandong Li, Haoran Xie, Li Chen, Jianping Wang, Xiaotie Deng, "News impact on stock price return via sentiment analysis", Knowledge-Based Systems 69 (2014) 14–23*
12. *Axel Groß-Klußmann, Nikolaus Hautsch, "When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions", Journal of Empirical Finance 18 (2011) 321–340*
13. *Michael Hagenau, Michael Liebmann, Dirk Neumann, " Automated news reading: Stock price prediction based on financial news using context-capturing features", Decision Support Systems, Volume 55, Issue 3, June 2013, Pages 685-697*
14. *Courtesy: www. quandl.com, provides access to free datasets*
15. *Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.*
16. *Courtesy: Thomas Reuters, Eikon login received via GITAM University MoU agreement.*
17. *Seabold, Skipper, and Josef Perktold. "Statsmodels: Econometric and statistical modeling with python." Proceedings of the 9th Python in Science Conference. 2010.*