FINAL REPORT

by 2020a7ps0139u NEELESH NAGPAL

Submission date: 16-May-2023 10:57PM (UTC+0400)
Submission ID: 2072885389
File name: -Machine_Learning_and_Sentiment_based_Investment_analysis_1.pdf (1.33M)
Word count: 5312
Character count: 28673

REPORT

ON

MACHINE LEARNING
AND
SENTIMENT BASED INVESTMENT ANALYSIS

BY

NEELESH NAGPAL 2020A7PS0139U CS

COURSE NUMBER: CS F377

Prepared in Fulfillment of the
Design Project Course

AT

BITS Pilani, Dubai Campus
Dubai International Academic City (DIAC)
Dubai, U.A.E

Second Semester
(FEBRUARY 2023 - MAY 2023)

BITS Pilani, Dubai Campus
Dubai International Academic City (DIAC)
Dubai, U.A.E

Course Name: Design Project **Course No:** CS F377

Duration: 6-Feb-2023 to 31-May-2023 **Date of Start:** 6-Feb-2023

Title of the Project: Machine Learning and Sentiment Based Investment Analysis

Student Name: Neelesh Nagpal **Student ID:** 2020A7PS0139U

Discipline of Student: Computer Science Engineering

Name of Project Supervisor: Dr. Tamizharasan P S

Key Words: Investment strategies, Stock Exchange, Risk Management, Sentiment analysis, Machine Learning, Deep Learning

Project Areas:

- Investment Analysis
- Sentiment Analysis
- Regression analysis
- Machine learning in Finance
- ANNs and RNNs

Abstract:

Stock price prediction has always been an idea of intrigue amongst the finance industry. Traditional mathematical models have their own limitations in terms of time complexity and efficiency in making these predictions. Modern data analysis techniques aim to exploit the power of Machine Learning and Natural Language Processing to be able to make credible predictions by not only considering the stock price performance in the market, but also the sentiment of investors in the market through news feed, social media, etc. The hypothesis is that firm level sentiment affects the stock performance, too. This project aims to provide a systematic approach to analyze stock price performance, through exploratory data analysis, followed by application of regression and neural networks to solidify the predictions of the direction of the stock market. Sentiment analysis of the stocks under consideration involves classifying an opinion as positive or negative sentiment, each affecting the stock prediction differently.

ACKNOWLEDGEMENTS

I am deeply appreciative towards Dr. Srinivasan Madapusi, BPDC Director, for giving us students the opportunity to gain firsthand experience by working on projects under the guidance of experienced faculty.

Additionally, I extend my gratitude to my Project Supervisor, Dr. Tamizharasan P S, for his unwavering guidance, support, and motivation. His expertise and experience helped me understand the fundamental concepts of sentiment analysis and its application in investment analysis and portfolio management and enabled me to grasp the standards of research at a professional level. Finally, I am grateful to my family for their constant encouragement and positivity, which played a significant role in my successful completion of the project.

Student Name: **NEELESH NAGPAL**

ID Number: **2020A7PS0139U**

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1.LITERATURE REVIEW

Nan Jing et al. (2021) proposed a hybrid model that combines deep learning with investor. sentiment analysis for predicting stock prices. CNN was used to classify investor sentiments and LSTM was used to capture temporal dependencies in stock prices and a sentiment analysis algorithm to extract sentiment from financial news articles about the SSE (Shanghai Stock Exchange). The sentiment score generated was used as an input feature to the LSTM.

Arthur Emanuel de Oliveira Carosia et al. (2021) proposed the use of sentiment analysis techniques based on ANNs to form the basis of investing strategies in the Brazilian stock market. Different ANNs were studied, and their performances were compared, CNN gave the most promising results in terms of F1 score and accuracy. CNN was used to classify news sentiments as positive, negative, or neutral. These sentiment scores were then used to develop trading strategies-based sentiment momentum and sentiment reversal. The results showed that sentiment-based investment decisions deem higher returns and better risk management than conventional buy and hold strategies and technical stock analysis.

Konstadin Mishev et al. (2020) evaluated the performance of sentiment analysis techniques in financial markets from traditional lexicon-based strategy to modern transformer-based approach. The index evaluated was S&P500. Among the lexicon-based approach, it was found that the hybrid lexicon outperforms baseline lexicons like general or finance specific lexicons. The paper also evaluated machine learning models like SVMs, Random Forests and deep learning models like LSTM and Transformer model. Authors found that BERT (Bidirectional Encoder Representations from Transformers outperforms other models.

Ashish Pathak et al. (2019) proposed a hybrid approach by combining sentiment analysis and neural networks to develop a robust model that aims to predict financial markets behavior better than when these techniques are used independently. The results were experimentally verified by applying the model on some top stocks in the Indian Stock market scenario.

Yunchuan Sun et al. (2018) developed a Guba-based sentiment analysis technique to better understand the forecasting of the Chinese Stock market. Guba is an online social media platform where people can post and share their views about their favorite stocks. This novel approach directly addresses the investor sentiment by exploiting the vast volumes of data shared in Guba, by classifying sentiment into good, bad, or neutral.

Raj Parekh et al. (2022) proposed an approach to forecast cryptocurrencies price fluctuation. They developed an approach that not only considers the currency's historic price and investor sentiments, but also its interdependency on other cryptocurrencies in the market. The model developed by DL-GuesS modelled the price prediction of Dash by analyzing the tweets and price history of Dash, Bitcoin and Litecoin.

Ramkrishna Patel et al. (2021) developed a two-model approach that combined NLP for feature extraction and LSTM for prediction of stock prices. They used Text Blob and Vader sentiment packages for feature extraction and converting textual data to numerical data. Then these values were fed into the LSTM RNN for predicting the closing price after 90 days.

Kristian Bondo Hansen et al. (2022) developed an approach to analyze alternate data besides just the investor sentiment. According to them, alternative data is something that is not easily available, is expensive to gather and is outside the market. Some examples of alternative data include satellite imagery, GPS data, social media commentary. These factors may seem irrelevant to investment analysis, but they have an indirect impact on the financial market.

Shri Bharathi et al. (2017) developed an approach to make stock market predictions by combining both Sensex points and RSS feeds. RSS stands for Really Simple Syndication. Opinion mining is conducted to develop a sense of firm level sentiment of a stock.

Liang Li et al. (2020) proposed a machine learning approach to develop smart investment strategies using the K-Means algorithm. It took into consideration the data of China Merchants

Bank's Capricorn Smart investment. Efficiency of the investment was studied using machine learning algorithms and declared smart investments were the way to move forward in the world of finance.

Sahar Sohangir et al. (2018) proposed a deep learning-based approach to determine if they can improve sentiment analysis of financial social media platforms like StockTwits and SeekingAlpha to extract vital information about the mass investor sentiment that drives the financial markets. They implemented sentiment analysis using both data mining and deep learning. Data mining approaches included SVMs, Naïve Bayes and Decision Trees. Deep learning approaches included various neural network models like CNNs, LSTM and doc2vec. CNN outperformed data mining approaches as well as topped within its own category.

Yue Qiu et al. (2022) created a new measure of sentiment that considers the importance of text-based content and financial anomalies. This sentiment index can be used to effectively forecast stock trends. They incorporated the day-of-the-week effect as well while making predictions and experimentally proved significantly improved accuracy in predictions of various machine learning algorithms like SVM, NB, RF, Logistic regression, KNN, GBDT.

Wataru Souma et al. (2019) proposed an enhanced deep learning model for sentiment analysis in stock prediction and mapping investing strategies. To create the word vectors needed as inputs for the deep learning TensorFlow network, they utilized articles from both the Wikipedia and Gigaword five corpus published in 2014. They implemented the global vectors for word representation approach to generate these word vectors. They made use of a hybrid model for training based on RNNs and LSTM units.

Yuming Li et al. (2019) introduced in this paper, the main concept to be the application of deep reinforcement learning to the decision-making process of stock trading, as well as using this technique to predict stock prices. The three classic models in DRL, Dueling Double Deep Q-Network (Dueling DDQN), Deep Q-Network (DQN), and Double Deep Q-Network (DDQN) were selected. DQN displayed the most promising results. They suggested the use of Adaboost for stocks with which investors were more familiar with.

1.2.FINANCIAL MARKETS AND INVESTMENT ANALYSIS

Financial markets have been a topic of great intrigue and innovation since their inception. People get involved in investment analysis for a variety of reasons ranging from making quick profits to ensuring a safe return in the future long run.

Investment analysis is the tool through which people make these decisions. Investment analysis forms the backbone of the financial services industry. It is the process of evaluating the potential profitability and risk of an investment opportunity. It involves examining financial and other relevant data to determine the viability of an investment opportunity and to make informed investment decisions. All investments are subject to market risk and investors should be cautious in making smart and calculated decisions. An overview of the investment analysis journey includes the following steps:

1. **Objective**: People entering the investment arena must be clear about the objective of their investment so that they can assess their risk-taking capabilities appropriately. They must have a clear idea of where their finances stand and avoid taking unnecessary risk if their liabilities do not allow them to.
2. **Risk Assessment**: People should access their risk based on their objective. The higher the risk, the higher is the return of an investment. Investors should carefully try and manage a suitable trade-off between risk and return.
3. **Market Analysis**: People should analyze the market scenario well and based on the competitive landscape and economic trends they should be able to narrow down the investments which suit their interests while minimizing risk simultaneously.
4. **Financial Analysis**: Once an investment opportunity is narrowed down, people should analyze the history of that particular investment and its performance in recent years. They should only go ahead with the investment if they deem the commodity or stock to be financially healthy.
5. **Investment Strategy**: Depending upon all the analysis conducted so far, investors must carefully plan out an investment strategy which includes appropriate asset allocation and diversification. The goal is to reach the objective of the investment while minimizing risk.
6. **Monitoring**: Once an investment has been made, an investor must track the performance of their portfolio to check if the chosen investment is performing as per the expectation. If the investment fails to meet the objective, then the investment strategy must be readjusted as soon as possible.

CHAPTER 2

DATA VISUALIZATION AND STOCK PRICE ANALYSIS

2. DATA VISUALIZATION

2.1. EXPLORATORY ANALYSIS

Data visualization is a crucial step when it comes to performing any sort of analysis. It is the first time we can observe some patterns from the data. These patterns form the basis of subsequent modelling and hypothesis development. This provides a more comprehensive understanding of the market before any investor takes an investment decision.

The stock dataset consists of 8 different stocks taken in consideration from 2012 right through 2020 consisting of ~2000 entries. The dataset also tracks the SP500 index which gives an overall strength of the top 500 companies in the US. The stocks under consideration are: AAPL (Apple), BA (Boeing), T (AT&T), MGM (MGM Resorts), AMZN (Amazon), IBM, TSLA (Tesla motors) and GOOG (Google). The attempt is to cover all sectors of the economy. The graph below shows the stock fluctuations in the considered period. The prices have not been normalized and the raw data has been plotted. A steep drop in prices around March 2020 accounts for the emergence of COVID-19. However, the market bounced back soon.

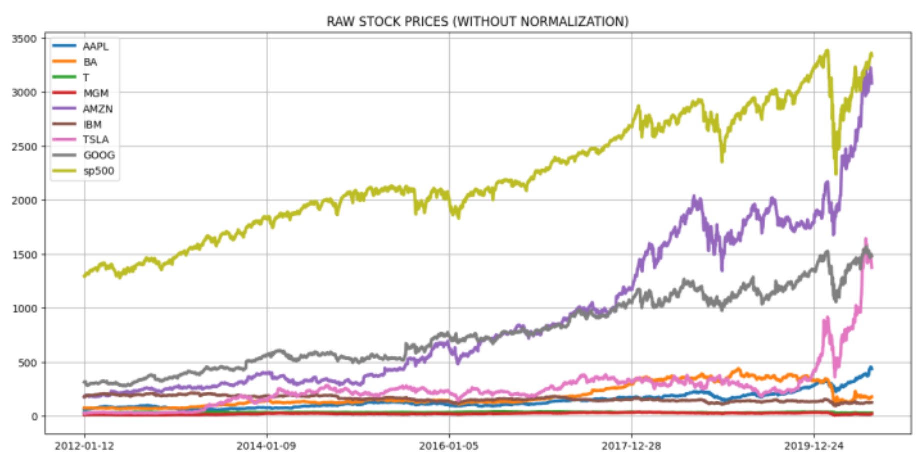


FIG 1

STOCK PRICE VS DATE

The data once normalized can be plotted again to assume the same starting point of all the stocks. Normalization is necessary before any actual comparisons can be made among the stocks.

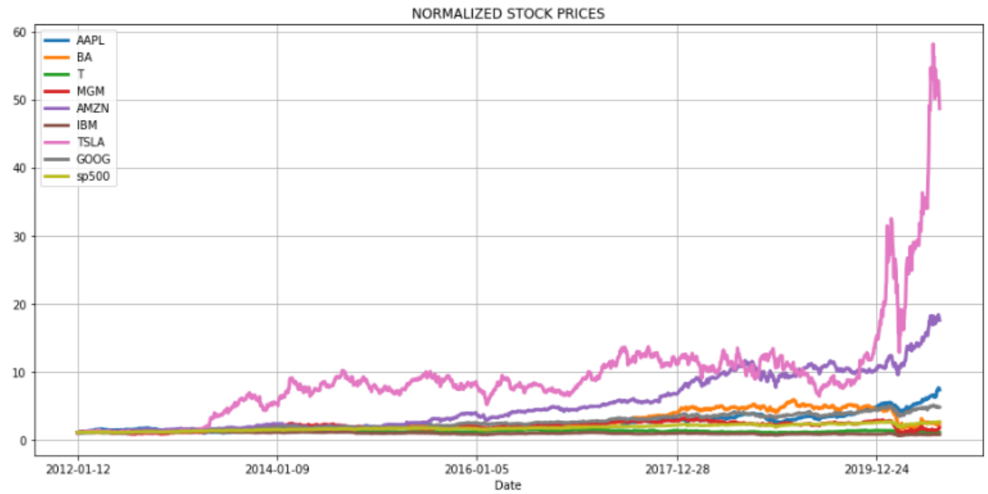


FIG 2

NORMALISED STOCK PRICE VS DATE

Daily returns of stocks are a good measure to determine the stock's performance. Positive return implies a bullish market and growing stock price. However, a negative return means the price of the stock fell as compared to the previous day.

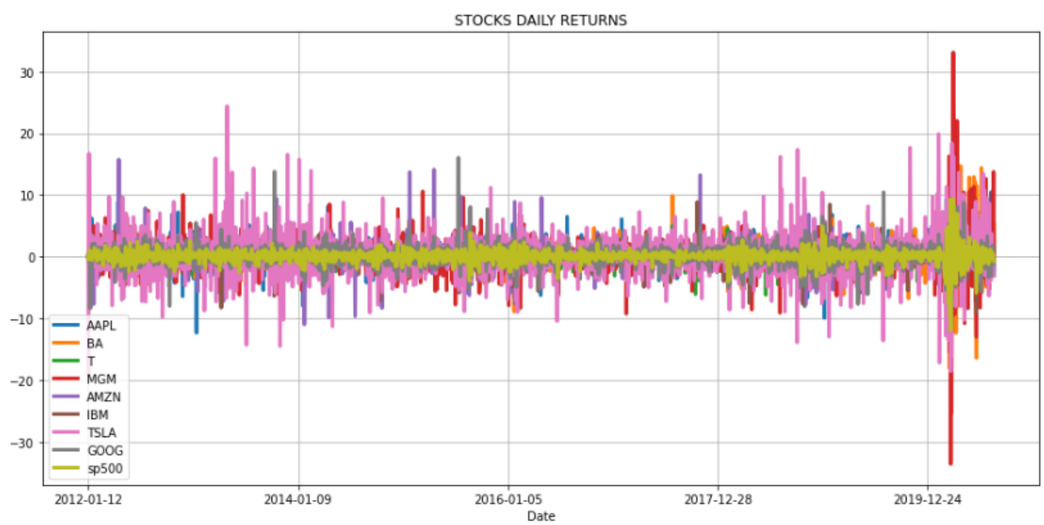


FIG 3

STOCK DAILY RETURNS

A huge drop can be seen in the returns section of MGM around March 2020 as the hotel industry took a big financial hit due to COVID-19.

Another way to interpret daily returns of stocks is using histograms. The farther the values deviate from 0 (standard deviation) the more risk is associated with the particular security in general.

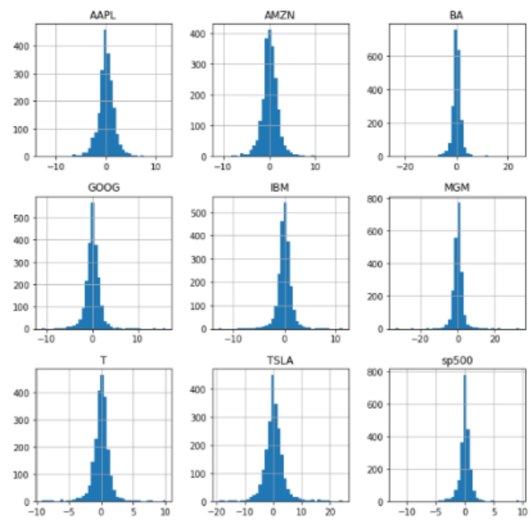
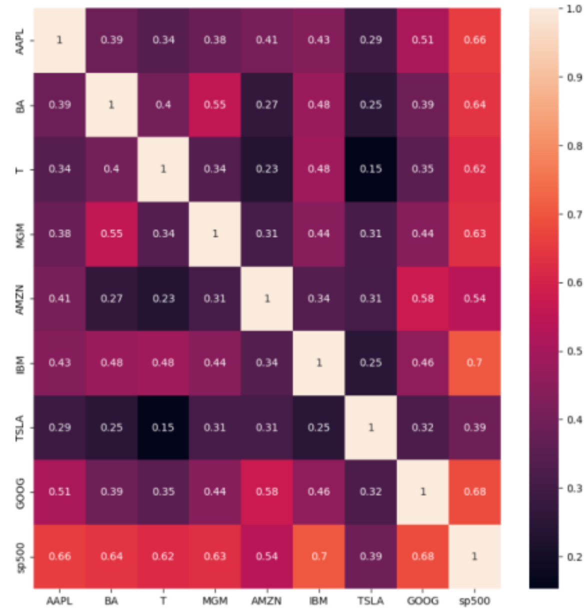


FIG 4

RISK VISUALISATION

Stocks with a wider bottom imply more fluctuations in the returns. MGM showed major deviations with returns reaching between -20% to 20%. This implies MGM was a risky stock to invest in.

A correlation matrix in the form of a heatmap can be visualized to perform some analysis on how the performance of one stock affected the other and vice versa. The heatmap obtained is as below:



<u>FIG 5</u>

<u>CORRELATION MATRIX</u>

<u>HEATMAP INFERENCES:</u>

[1] Every stock would have a 1 correlation with itself as it completely defines the variance in performance of its own.

[2] Strong positive correlation between S&P500 and IBM as well as between S&P500 and GOOG. This is because both IBM and GOOGLE are within S&P500 (Fortune 500).

[3] MGM and Boeing also show a highly positive correlation. This is because both hotel and transport industry are related to many extents.

CHAPTER 3

REGRESSION ANALYSIS IN FINANCE

3. REGRESSION ANALYSIS IN FINANCE

3.1.EXPLORING REGRESSION OPTIONS

A regression model is a basic and straightforward machine learning model that sims to establish a linear relationship between an independent variable (X) and a dependent variable (Y).

The most obvious choice of linear regression may not be the best choice while predicting stock prices since we have more than one independent variable (Price and Volume) and one dependent variable that we are trying to predict (Next day's price). Moreover, price and volume will have a high collinearity and the model may overfit the data in case of linear regression. Thus, linear regression may not be the best model to deploy.

The next option is Ridge Regression. This model works best for datasets with high multicollinearity. It uses L2 regularization. The model moves towards increasing bias to get a more generalized model that does not overfit the data and reduces variance. This is done by changing the slope of the regression line (α).

3.2. RIDGE REGRESSION MODEL

3.2.1. DATA PREPARATION

The data for the model needs 2 input features which are closing price and volume of stocks traded on that day. The target variable will be the next day's closing price. The target column is generated by adding a new column with 1 up shifted values of the closing price column. The stock taken in consideration is **TSLA**. The data has been scaled using Min-Max Scaler. The data is split in 65:35 ratio for training and testing purposes, respectively. The graph below represents the training and test data visualized. The two lines in each graph represent closing price (Blue) and stock volume (Orange) respectively.

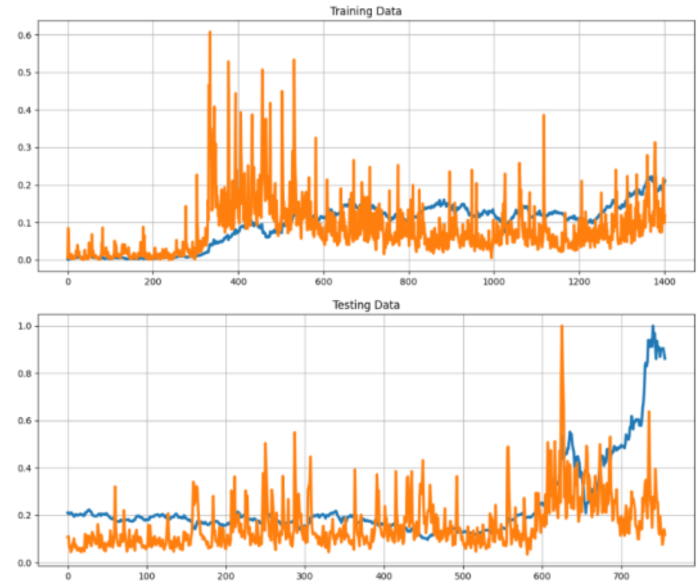


FIG. 6

TRAIN AND TEST SPLIT

3.2.2. BUILDING AND TESTING MODEL

The ridge regression model is built and trained with 65% of data as training data. The model uses this data to learn the patterns and interdependencies between X and Y. The model is then put to the test by feeding it data that it has never seen before. The achieved accuracy is ~83%.

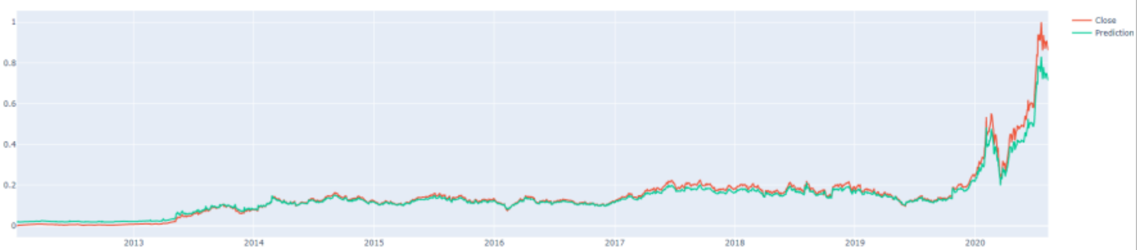


FIG. 7

ORIGINAL V/S PREDICTED TSLA STOCK (REGRESSION)

The results are plotted for original vs the predicted price of the TSLA stock. The model does a decent job in trying to generalize the relationship between the current price and predicted price. It can be seen the model deviates slightly more towards the 2018-2020 values since that was the test data and testing accuracy was bound to be less than training accuracy.

CHAPTER 4

DEEP LEARNING IN STOCK PRICE ANALYSIS

4. DEEP LEARNING IN STOCK PRICE ANALYSIS

Deep learning is at the forefront of analysis when it comes to making predictions, especially in the stock market which is so volatile. Traditional machine learning models are limited by their computational powers since it is impossible to feed all the input features to the model. This is where neural networks come into play. These neural networks can be trained to behave like actual neurons and make decisions and classifications. Deep learning models provide the potential for making much more reliable predictions by figuring out complex relationships that may be missed by conventional mathematical models. With the ever-increasing computational power and faster algorithms, deep learning models can be combined with natural language processing (NLP) techniques like sentiment analysis to make even more accurate forecasts.

4.1.NEURAL NETWORK OPTIONS

There are a number of options for neural networks that can be considered while developing a model to predict stock prices:

[1] Feedforward Neural Networks (Vanilla): They are the feedforward neural networks that train through iterations and gradient descent. These neural networks, however, have no concept of memory or temporal dependence. These factors are crucial when making time sensitive predictions like stock prices.

[2] Recurrent Neural Networks (RNN): These networks are designed to keep the temporal dimension into consideration by having a memory feedback loop but suffer from a vanishing gradient problem.

[3] Long Short-Term Memory (LSTM): These networks are designed to remember long term dependencies unlike Vanilla RNNs.

Thus, an LSTM based model is the best choice for making stock price predictions.

4.2.LSTM TIME SERIES MODEL

4.2.1. BUILD AND TRAIN THE MODEL

The dataset used for the training of ridge regression is used again. Here, too, the model will have 2 input features (Price and Volume) and 1 output feature (Target Price). The data is split 70:30 train-test split. The model is built with 3 hidden layers and 150 neurons each. The linear activation function is chosen, and minimum squared error is the cost function that the mode aims to minimize.

4.2.2. TESTING

The model is trained with 70% of data. It is put to the test by passing the rest 30% of data. The predicted values are recorded in a new data frame and the results are plotted against the ground truth values.

FIG. 8

ORIGINAL V/S PREDICTED TSLA STOCK (LSTM)

The model performs exceedingly better than the ridge regression model. The model can detect the pattern of sharp fluctuations in the test data and does not miss by a big margin. The model can further be optimized by changing and trying the output by varying the hidden layers, number of neurons, changing the activation function, etc.

CHAPTER 5

SNETIMENT BASED INVESTMENT ANALYSIS

5. SENTIMENT ANALYSIS IN INVESTMENT STRATEGY

5.1.INTRODUCTION

Artificial Intelligence, Deep Learning and Machine Learning have been transforming the world of finance and investing. AI-powered investment advisors can perform real time insight analysis of an investment and produce a recommended strategy in an incredibly brief time. This has come to be known as HFT (High Frequency Trading). Artificial Intelligence has enabled advisors to assess alternative datasets like ship cargo movements and weather forecasts to manage hedging strategies. Artificial Intelligence has also made it possible to predict a stock's behavior by considering the traffic on a particular company's website.

One of the most important and powerful applications of artificial intelligence in the world of investing is sentiment analysis or often called opinion mining. This Natural Language Processing (NLP) technique allows financial analysts to develop a sense of sentiment in the market, which is the general opinion of investors towards the company.

Sentiment based investment analysis is an approach to assessing the sentiment of people and investors towards a particular asset or the whole company. This NLP backed technique is based on the fundamental idea that people's sentiments and public relations of a company affects their investment opportunities along with their cashflows. Analyzing these sentiments provides analysts and potential investors with valuable insights into the market trends and hence makes more informed choices and alter strategies. There are various methods to gather and analyze investor sentiment data, like surveys, polling, social media analysis and news analysis.

Sentiment analysis works best when combined with other forms of analysis like technical or fundamental analysis of stocks. Sentiment analysis allows investors to alter their investing strategy. For example, an investor may look for an opportunity to invest in a stock that is being undervalued because of negative sentiments. Sentiment analysis works well in giving a general market sentiment and an idea of momentum to find the right time to invest.

However, due to its volatile nature, Sentiment analysis is not a foolproof method to judge an investment opportunity since sentiments in the market can be swayed by media coverage or political interferences which may give rise to biases and inaccuracies.

Thus, it is only advisable to use sentiment analysis along with other analysis results as well to reduce the probability of over-dependence on any one source of information.

5.2. DATA CLEANING

The goal is to build a sentiment detector model using NLP that can work as a binary classifier of text-based information into either positive or negative sentiment. The dataset consists of texts collected from twitter with the keywords of stocks or companies in question. The texts have been assigned 1 or 0 depending on the sentiment it produces. The aim will be to train a model such that it can classify any new text data as 0 or 1. Data cleaning includes removing punctuation followed by removing stop words. The punctuation simply includes the characters in the string as follows!"#$%&'() *+, -. /:;<=>?@[\]^_`{|}~. The stock sentiment data is processed to remove these characters. Next step is to remove stop words. These words are the words that are used too commonly and hold no actual meaning when it comes to analyzing sentiment. Some examples of stop words include it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs,' etc. The tokens are obtained using the gensim package. A word cloud can be plotted to visualize which keywords are most important. The aim is to find keywords which people use when they are positive or bullish about the market and when they are negative or bearish about the market.

FIG. 9

POSITIVE SENTIMENT WORD CLOUD

FIG. 10

NEGATIVE SENTIMENT WORD CLOUD

5.3. TOKENIZING AND PADDING

Tokenization is the process of dividing the vast text data into smaller segments, which are easier to analyze and train the model. It often involves converting text data sequence into integers. The task is to gather unique words in the data and divide the data into train and test split. The words are then encoded to their corresponding integer sequences using the tokenizer package. Total unique words obtained are 9268. The maximum length of any sentence in the dataset is 20. To maintain uniformity, padding is performed to the tokenized sequences. This involves adding zeroes to the sequence such that all vectors are of same length.

5.4. LSTM MODEL

The goal is to reduce the complexity of data that is fed into the network. The embedded layer contains reduced features as compared to the actual dataset. This saves resources and improves efficiency of the model drastically. The task is like PCA or autoencoders. The embedded package is used to develop an embedded layer. The model is trained with 256 neurons in the input layer with relu activation and the output layer has 2 neurons 1 depicting each sentiment (0 for negative and 1 for positive). The padded train data is passed as the input. The model shows a validation accuracy of ~76% with 4 epochs to make sure the model does not overfit the data.

5.5. MODEL ASSESSMENT

The model is evaluated with data it has not seen before. The output is a 2-dimensional vector with both arguments depicting probabilities of the sentiment being positive or negative. The actual output of the model is the value that is larger of the two probabilities. A confusion matrix can be used to visualize the accuracy. As per the matrix, the model correctly classified ~1600 samples as true positive and ~2600 as true negative. However, the model misclassified 50 entries as false positive and ~1100 entries as false negative. The final testing accuracy obtained is ~78%.

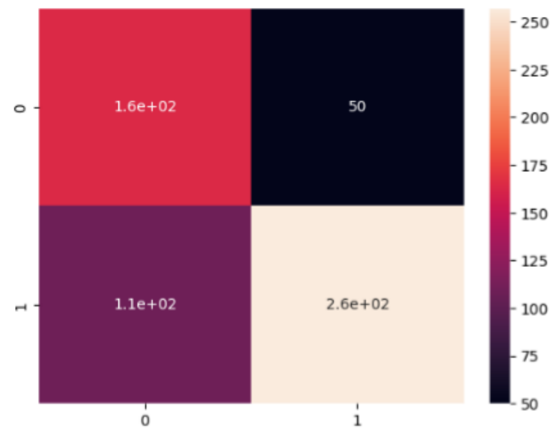


FIG. 11

CONFUSION MATRIX

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6. CONCLUSION AND FUTURE SCOPE

6.1. CONCLUSION

Stock price analysis and portfolio management is a core finance field that is continuously moving more and more towards integrating AI and Machine learning in its domain. The sheer efficiency and potential to accurately analyze the market direction is the reason why more and more security analysts want to make use of machine learning techniques to predict stock prices. Machine learning provides a huge computational advantage when it comes to analyzing market prices over traditional mathematical models. This report aims to provide a highly reusable approach to stock price analysis. Ridge regression is used along with L2 Regularization as the machine learning model to predict stock prices. The future direction of the market is analyzed by the spot price and stock volume each day. These parameters form the input features to the model. The ridge regression model performed decently well to predict the stock prices and not overfit the data with an accuracy of ~83%. However, the same stock price analysis was performed with an LSTM Model, showed even closer predictions. Deep learning is an extremely powerful technique when it comes to analyzing data and generalizing patterns, even more so than traditional machine learning models like regression that are restricted by its mathematical nature of computation. The hypothesis considered in this report is that stock prices are not only dependent on the spot price and stock volume, but also on investor sentiment. The way a particular firm is viewed in the market and how people conceive of that stock is also an integral part of analysis. The data can be gathered from any social media channel like Twitter, news feed, etc. regarding the stocks in question. The aim here was to build an NLP model and train it to predict investor sentiment using an LSTM classifier that can predict probability a particular sentence as a positive or negative sentiment towards the company. This sentiment analysis, even though quite powerful, cannot be the only factor an investor can base their decision on. The previous analysis done using machine learning and deep learning models must be combined to gather an overall view of the stock.

6.2.FUTURE SCOPE

The scope in the direction of integrating computational AI in investment analysis is vast. Robo-advisers are being developed capable of performing millions of instance calculations and processing per second. The scope involves introducing newer and faster machine learning algorithms to make predictions. This report only considered the spot price and stock volume, however there are hundreds of other features that can be considered to have an impact on the stock price like geographic location, company profile, firm domain and many more. The deep learning model deployed here is LSTM. The output of the model can be improved by manipulating hidden layers, changing activation functions, introducing other input features, etc. High Frequency Trading (HFT) is a new branch in trading that focusses on performing analysis in the form of millions of instructions per second and gaining insights on real time data and not historical data. The future of AI in investment analysis is a promising one. It must be noted that investors can improve their portfolio decision making by a mix of AI and their own instincts and judgements.

REFERENCES

[1] R. Patel, V. Choudhary, D. Saxena and A. K. Singh, "LSTM and NLP Based Forecasting Model for Stock Market Analysis," *2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT)*, Meerut, India, 2021, pp. 52-57, doi: 10.1109/ICACFCT53978.2021.9837384.

[2] Jing, N., Wu, Z., & Wang, H. (2021, September). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems With Applications*, *178*, 115019. https://doi.org/10.1016/j.eswa.2021.115019

[3] de Oliveira Carosia, A. E., Coelho, G. P., & da Silva, A. E. A. (2021, December). Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning. *Expert Systems With Applications*, *184*, 115470. https://doi.org/10.1016/j.eswa.2021.115470

[4] Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, *8*, 131662–131682. https://doi.org/10.1109/access.2020.3009626

[5] Dash, A. S., & Pathare, S. B. (2022). A Survey of Sentiment Analysis Through Machine Learning for Forecasting Indian Stock Market Trend. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4057274

[6] Sun, Y., Fang, M., & Wang, X. (2018, March 5). A novel stock recommendation system using Guba sentiment analysis. *Personal and Ubiquitous Computing*, *22*(3), 575–587. https://doi.org/10.1007/s00779-018-1121-x

[7] He, H., & Dai, S. (2022). A prediction model for stock market based on the integration of independent component analysis and Multi-LSTM. *Electronic Research Archive*, *30*(10), 3855–3871. https://doi.org/10.3934/era.2022196

[8] Hansen, K. B., & Borch, C. (2022, January). Alternative data and sentiment analysis: Prospecting non-standard data in machine learning-driven finance. *Big Data & Society*, *9*(1), 205395172110707. https://doi.org/10.1177/20539517211070701

[9] Shaikh, A., & Shinde, S. (2022). Prediction of Stock Market Prices using Prediction algorithm and Sentiment Analysis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4111874

[10] Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018, January 25). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data*, *5*(1). https://doi.org/10.1186/s40537-017-0111-6

[11] Qiu, Y., Song, Z., & Chen, Z. (2022, January 27). Short-term stock trends prediction based on sentiment analysis and machine learning. *Soft Computing*, *26*(5), 2209–2224. https://doi.org/10.1007/s00500-021-06602-7

[12] Souma, W., Vodenska, I., & Aoyama, H. (2019, January). Enhanced news sentiment analysis using deep learning methods. *Journal of Computational Social Science*, *2*(1), 33–46. https://doi.org/10.1007/s42001-019-00035-x

[13] Li, Y., Ni, P., & Chang, V. (2019, December 23). Application of deep reinforcement learning in stock trading strategies and stock forecasting. *Computing*, *102*(6), 1305–1322. https://doi.org/10.1007/s00607-019-00773-w

FINAL REPORT

ORIGINALITY REPORT

14%	12%	10%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	link.springer.com Internet Source	1%
2	www.researchgate.net Internet Source	1%
3	ouci.dntb.gov.ua Internet Source	1%
4	jfmp.sbu.ac.ir Internet Source	1%
5	Sakshi Chhabra, Deepika Saxena, Smruti Rekha. "A Review on Secure Cloud Resource Management", Institute of Electrical and Electronics Engineers (IEEE), 2023 Publication	1%
6	thescipub.com Internet Source	1%
7	hrcak.srce.hr Internet Source	1%
8	Ali Shavandi, Majid Khedmati. "A multi-agent deep reinforcement learning framework for	1%

algorithmic trading in financial markets", Expert Systems with Applications, 2022
Publication

9	eprints.pknstan.ac.id Internet Source	1 %
10	cric.ku.dk Internet Source	1 %
11	www.bu.edu Internet Source	1 %
12	Tingcheng Mo, Chi Xie, Kelong Li, Yingbo Ouyang, Zhijian Zeng. "Transmission effect of extreme risks in China's financial sectors at major emergencies: Empirical study based on the GPD-CAViaR and TVP-SV-VAR approach", Electronic Research Archive, 2022 Publication	1 %
13	www.jetir.org Internet Source	<1 %
14	www.coursehero.com Internet Source	<1 %
15	papers.ssrn.com Internet Source	<1 %
16	www.studymode.com Internet Source	<1 %
17	ethesis.nitrkl.ac.in Internet Source	<1 %

18	github.com Internet Source	<1 %
19	repository.iiitd.edu.in Internet Source	<1 %
20	Wataru Souma, Irena Vodenska, Hideaki Aoyama. "Enhanced news sentiment analysis using deep learning methods", Journal of Computational Social Science, 2019 Publication	<1 %
21	repositories.lib.utexas.edu Internet Source	<1 %
22	"Advances in Digital Marketing and eCommerce", Springer Science and Business Media LLC, 2020 Publication	<1 %
23	ebin.pub Internet Source	<1 %
24	research-api.cbs.dk Internet Source	<1 %
25	www.research.manchester.ac.uk Internet Source	<1 %
26	"Advances in Computing and Intelligent Systems", Springer Science and Business Media LLC, 2020 Publication	<1 %

27	"Neural Information Processing", Springer Science and Business Media LLC, 2023 Publication	<1%
28	Ibrahim Said Ahmad, Azuraliza Abu Bakar, Mohd Ridzwan Yaakub, Mohammad Darwich. "Sequel movie revenue prediction model based on sentiment analysis", Data Technologies and Applications, 2020 Publication	<1%
29	Piyush Vyas, Martin Reisslein, Bhaskar Prasad Rimal, Gitika Vyas, Ganga Prasad Basyal, Prathamesh Muzumdar. "Automated Classification of Societal Sentiments on Twitter With Machine Learning", IEEE Transactions on Technology and Society, 2022 Publication	<1%
30	aircconline.com Internet Source	<1%
31	doras.dcu.ie Internet Source	<1%
32	jfin-swufe.springeropen.com Internet Source	<1%
33	researchbank.rmit.edu.au Internet Source	<1%
34	scholar.utc.edu Internet Source	<1%

35	www.worldscientific.com Internet Source	<1 %
36	Sepp Hochreiter, Jürgen Schmidhuber. "Long Short-Term Memory", Neural Computation, 1997 Publication	<1 %
37	"Communication and Intelligent Systems", Springer Science and Business Media LLC, 2022 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		