**Stock Price Prediction Using LSTM and Stacked Regression**

P.Nani Siva Rama Krishna K.Nikhil reddy N.Bhupendra sai

700731657 700739505 700734676

**Abstract**

Predicting stock market prices is a difficult task that has traditionally necessitated extensive human-computer interaction. Traditional batch processing methods cannot be used efficiently for stock market analysis because stock prices are correlated. We propose an online learning algorithm that makes use of Long Short Term Memory (LSTM), a type of recurrent neural network (RNN) in which the weights for individual data points are adjusted using stochastic gradient descent. This will produce more accurate results when compared to existing stock price prediction algorithms. With data of varying sizes, the network is trained and tested for accuracy, and the results are tabulated. The system's accuracy is then compared to that of an Artificial Neural Network. The purpose of this paper is to evaluate and compare LSTM deep learning architectures for financial time series prediction in the short and long term. One of the most difficult real-world applications of time series prediction is widely regarded as this. Unlike traditional recurrent neural networks, LSTM allows for any size time step and is not affected by the vanishing gradient problem. We consider both bidirectional and stacked Regression predictive models, as well as shallow neural networks and simple forms of LSTM networks in our experiments. The analyses are carried out with the help of a publicly available dataset of stock market closing prices.Overall, the project's ultimate goal is to forecast how the market will behave in the future by analysing sentiment from previous tweets.

**Keywords**: stock prediction, Long Short Term Memory, Recurrent Neural Network, online learning, stochastic gradient descent

**Introduction**

The stock market is a massive network of investors and traders who buy and sell stocks, causing the stock price to rise or fall. The supply and demand principles govern stock prices. The ultimate goal of purchasing stock is to profit by investing in companies with rising perceived value (i.e., share price). Share price movements can be traced back to specific Key Performance Indicators; stock markets are inextricably linked to the world of economics (KPIs). The opening stock price ('Open,' end-of-day price ('Close,' intraday low price ('Low,' intraday peak price ('High,' and total volume of stocks traded during the day ('Volume' are the five most commonly used KPIs.

One of the most successful RNN architectures is long short-term memory. The memory cell is a unit of computation invented by LSTM that replaces traditional artificial neurons in the network's hidden layer. Networks can effectively associate memories and remote in time input with these memory cells, allowing them to grasp data structure dynamically over time while maintaining high prediction capacity.

"Others' thought process" has always been an important piece of information for the great majority of us while making decisions. The Web and the Internet have made it possible to learn about the perspectives and experiences of people who are neither our personal acquaintances nor well-known expert commentators, that is, people we have never heard of. As a result, an increasing number of people are using the Internet to share their ideas with others. Individual consumers' advantage in web-based speculations about products and administrations, as well as the potential influence such conclusions have, is a driving force behind this field of inquiry. Furthermore, there are many challenges associated with this cycle that must be overcome. to be trampled to achieve the best results We focused this research on the fundamental philosophy that occurs in these systems and the procedures that must be done to address the challenges that arise.

**Motivation**

Businesses are primarily concerned with client satisfaction and product reviews. Changes in social media mood have been demonstrated to correspond with changes in stock markets. Identifying and resolving customer complaints increases customer happiness and an organization's trustworthiness. As a result, an unbiased automated method is required to classify client reviews about any concern. Companies may have mountains of consumer feedback collected in today's climate, where we are understandably suffering from data overload (albeit this does not imply better or deeper insights), but it is still hard for normal humans to analyse it manually without error or bias. Companies with the best intentions are frequently trapped in an information vacuum. You are aware that you require insights to guide your decision-making, and you are aware that you are lacking them but are unsure how to obtain them. Sentiment research provides some insight into the most important concerns, at least from the standpoint of customers. Because sentiment analysis may be automated, judgements can be based on a large amount of data rather than pure intuition.

**Main Contributions & Objectives**

"What other people think" has always been a valuable piece of information for most of us while making decisions. The Internet and the Web have suddenly made it possible to learn about the perspectives and experiences of folks who are neither our personal acquaintances nor well-known professional critics that is, people we have never heard of. In turn, an increasing number of people are making their views known to strangers over the Internet. Individual consumers' interest in online opinions regarding products and services, as well as the potential influence such opinions have, is a driving force behind this field of study. And there are other problems involved in this process that must be overcome. to be trampled all over in order to achieve the desired results We conducted this survey to basic methodology that normally occurs in this procedure and actions that are to be taken be taken to solve the difficulties encountered

**Objective**

Time Series forecasting and modelling is critical in data analysis. Time series analysis is a subset of statistics that is widely employed in subjects such as econometrics and operations research. Time series are commonly utilised in analytics and data science. Stock prices are variable, and their value is determined by a variety of factors. This project's main goal is to forecast stock values utilizing **Stacked Regression and LSTM**

**Related Work**

**Manh Ha Duong and Boriss Siliverstovs conducted research.**

The relationship between share prices and collective investment in major European countries such as France, Germany, Italy, the Netherlands, and the United Kingdom is being investigated. As European financial markets become more integrated, share prices in different European countries are likely to become even more correlated. This trend may also result in economic convergence across European countries if stock market changes influence real economic components such as investment and consumption. Indeed, our vector autoregressive models indicate that the positive connection between changes in equity prices and investment is considerable in general. As a result, 6 monetary authorities should watch share price reactions to monetary policy and their effects on the business cycle.

**Stock Market Prediction Using Machine Learning**

V Kranthi Sai Reddy, Student, ECM, Sreenidhi Institute of Science and Technology, completed the research work.

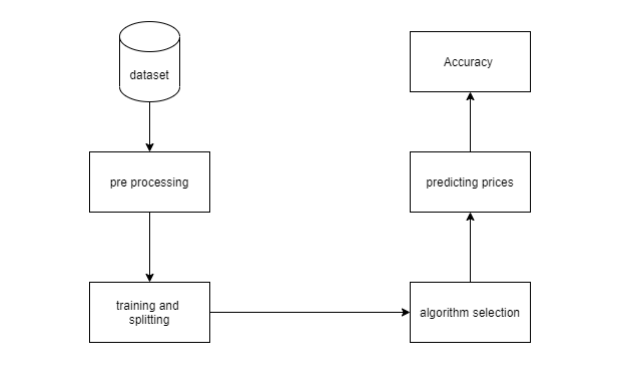
Hyderabad is in India. Global inventory buying and selling is one of the most important activity in finance. Stock market forecasting is the process of attempting to anticipate the future value of a stock or other financial instrument traded on a monetary exchange. This paper explains how to anticipate an inventory using Machine Learning. The majority of stockbrokers use technical and fundamental or time collection analysis when generating stock predictions. Python is the programming language used to anticipate the inventory marketplace's use of system mastering. We propose a Machine Learning (ML) technique in this research. The available shares information and advantage intelligence are then used to generate an accurate projection. This look at makes use of a system in this situation.Using the Support Vector Machine (SVM) method to forecast inventory costs for large and tiny capitalizations and within the three distinct markets, utilising expenses with every day and up to date frequencies.

**Forecasting the Stock Market Index Using Artificial Intelligence Techniques**

Lufuno Ronald Marwala conducted research. A dissertation submitted to the University of the Witwatersrand's Faculty of Engineering and the Built Environment in Johannesburg in fulfilment of the requirements for the Master of Science in Engineering degree. The weak form of the Efficient Market Hypothesis (EMH) asserts that forecasting the future price of an asset based on knowledge contained in historical prices of an asset is impossible. This indicates that the market performs like a random walk, making forecasting impossible. Furthermore, due to the inherent complexity of the financial system, financial forecasting is a challenging undertaking. The goal of this effort was to model and anticipate the future price of using artificial intelligence (AI) technology. an index of the stock market Based on past price data, three artificial intelligence techniques, namely neural networks (NN), support vector machines, and neuro-fuzzy systems, are used to anticipate the future price of a stock market index. Artificial intelligence approaches can account for financial system complexities and are employed as financial time series forecasting tools.

**Proposed Framework**

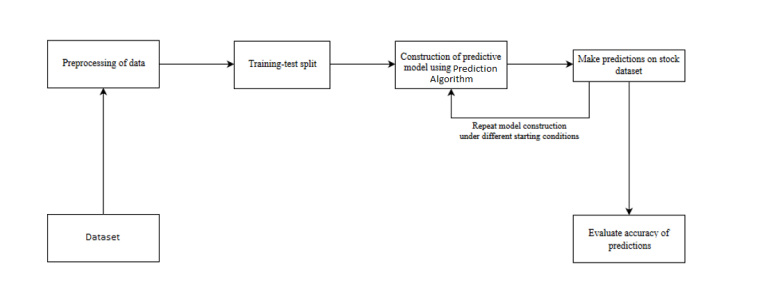
**Component Of System**



A component diagram is a type of diagram in UML. The goal is also distinct from the previous diagrams discussed. It does not define the system's functionality, but it does describe the components that enable such functions.

Component diagrams are used for visualising, describing, and documenting component-based systems, as well as developing executable systems through forward and reverse engineering. Component diagrams are essentially class diagrams that focus on a system's components and are frequently used to describe a system's static implementation perspective.

**Architecture**



**Analysis**

The financial market is a dynamic and complex system in which anyone can buy and sell currencies, stocks, shares, and derivatives through virtual platforms facilitated by brokers. The stock market allows investors to purchase shares of public firms through exchange or over-the-counter trading. This market has provided investors with the opportunity to make money and live a prosperous life by investing small quantities of money at a low risk compared to the risk of starting a new business or the necessity for a high-paying job. Many factors influence stock markets, resulting in market uncertainty and excessive volatility. Although humans can take orders and transmit them to the market, automated trading systems (ATS) run on computer software. Programs can process orders faster and more accurately than people. However, in order to evaluate and manage the performance of ATSs, risk management systems, and human-based safety measures are all required. Many factors are incorporated and considered when developing an ATS, such as the trading strategy to be used, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable future stock value prediction, and specific news related to the stock being analysed.

**System configuration**

This project may be run on standard hardware. We ran the entire project on an Intel I5 processor with 8 GB RAM and a 2 GB Nvidia Graphic Processor. It also has two cores that run at 1.7 GHz and 2.1 GHz. The first half of the process is the training phase, which takes about 10-15 minutes, and the second part is the testing phase, which just takes a few seconds to generate predictions and calculate accuracy.

**Hardware Requirements:**

• RAM: 4 GB

• Storage: 500 GB

• CPU: 2 GHz or faster

• Architecture: 32-bit or 64-bit

**Software requirements**

• Python 3.5 in Google Colab is used for data pre-processing, model training and prediction.

• Operating System: windows 7 and above or Linux based OS or MAC OS.

**Methodology**

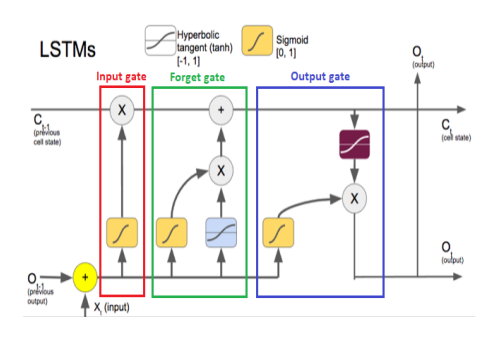
LSTMs are RNN subsets that can detect context-specific temporal dependencies over long time periods. Each LSTM neuron is a memory cell capable of storing additional information, i.e. it maintains its own cell state. Normal RNN neurons simply take in their previous hidden state and the current input to produce a new hidden state, whereas an LSTM neuron also takes in its previous cell state and produces a new cell state.

1. Sequential for initializing the neural network

2. Dense for adding a densely connected neural network layer

3. LSTM for adding the Long Short-Term Memory layer

4. Dropout for adding dropout layers that prevent over fitting



Data visualization is the discipline of attempting to understand data by displaying it in a visual context in order to highlight patterns, trends, and connections that might otherwise go undetected.

Python has a number of excellent graphing packages that are jam-packed with useful functionality. Python provides a great library for creating dynamic or highly customizable charts.

To get a little overview, here are a few popular plotting libraries:

* [**Matplotlib:**](https://matplotlib.org/)low level, provides lots of freedom
* [**Pandas Visualization:**](https://pandas.pydata.org/pandas-docs/stable/visualization.html)easy to use interface, built on Matplotlib
* [**Seaborn:**](https://seaborn.pydata.org/)high-level interface, great default styles
* [**plotnine:**](https://plotnine.readthedocs.io/en/stable/)based on R’s ggplot2, uses Grammar of Graphics
* [**Plotly:**](https://plot.ly/python/)can create interactive plots

**Type 1 Stacking Model**

Ensemble methods in statistics and machine learning combine numerous learning algorithms to achieve greater predictive performance than the learning algorithm alone. The stacking model is a group one. It use a meta-learning algorithm to find the optimum way to combine predictions from two or more base machine learning methods.

Estimator stack with a final regressor.

Stacked generalization involves stacking the output of individual estimators and calculating the final prediction with a regressor. Stacking allows you to capitalise on the strengths of each individual estimator by feeding their output into a final estimator.

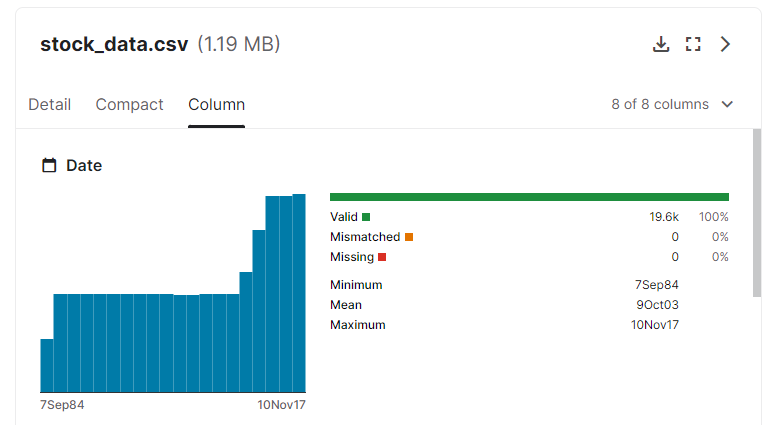
It is worth noting that estimators\_ are fitted on the entire X, whereas final estimator\_ is trained using cross-validated predictions of the base estimators via cross val predict.

*class*sklearn.ensemble.**StackingRegressor**(*estimators*, *final\_estimator=None*, *\**, *cv=None*, *n\_jobs=None*, *passthrough=False*, *verbose=0*)[[source]](https://github.com/scikit-learn/scikit-learn/blob/f3f51f9b6/sklearn/ensemble/_stacking.py#L675)

**Data Description**

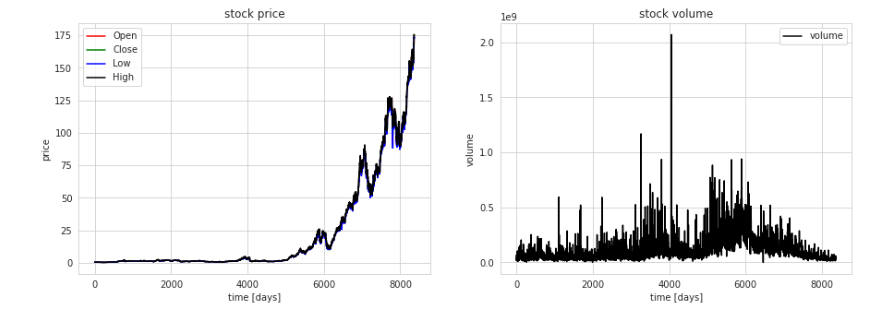
<https://www.kaggle.com/datasets/sanamps/stock-price-prediction-with-rnn>

At this point, we look for aggregated data from various sources before choosing a recent dataset from kaggle. This dataset is divided into three size-related parts (1 year, 6 months, and 3 months). Attributes in this dataset include Open, High, Low, Close, and Volume, but I chose Close as the label data and other attributes to extract features that help predict results.

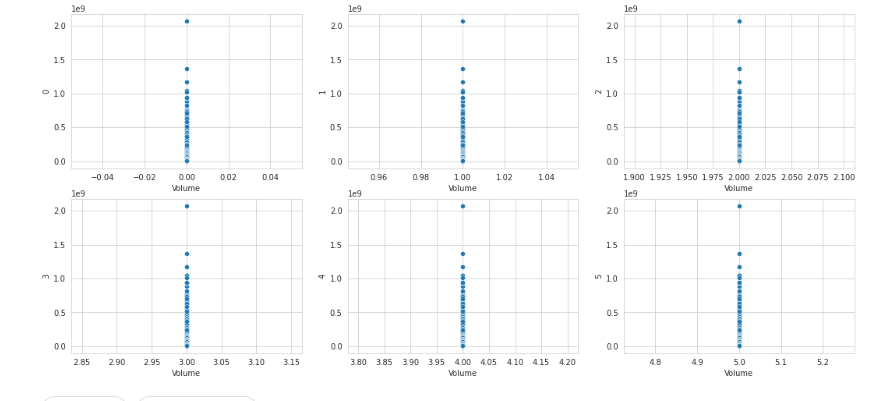


**Results**

**Data Analysis**

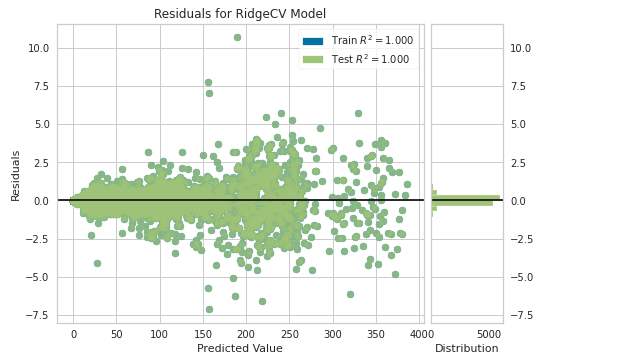


**Scatter Plot**



ridge's rmse after training: 0.5697901161442734

Computational runtime of this algo: 0.95 seconds



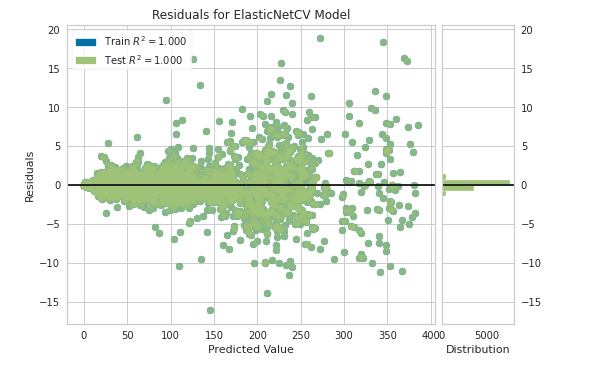
lasso's rmse after training: 1.0653253710276687

Computational runtime of this algo: 1.19 seconds



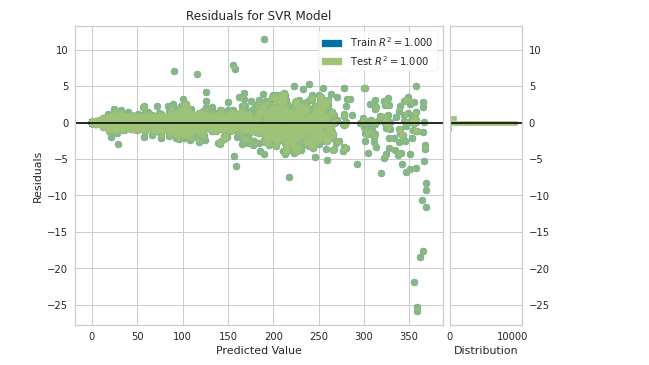
elasticnet's rmse after training: 1.0653227609999345

Computational runtime of this algo: 1.72 seconds



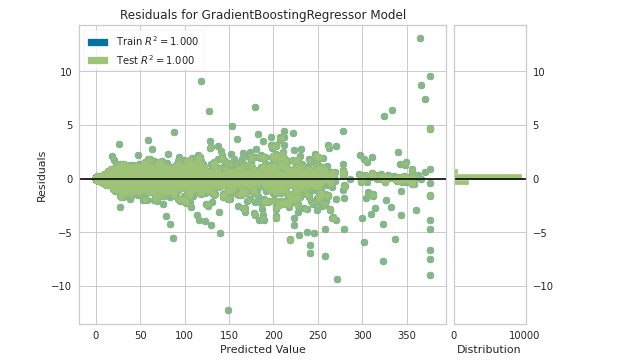
svr's rmse after training: 0.5856566427534331

Computational runtime of this algo: 58.98 seconds



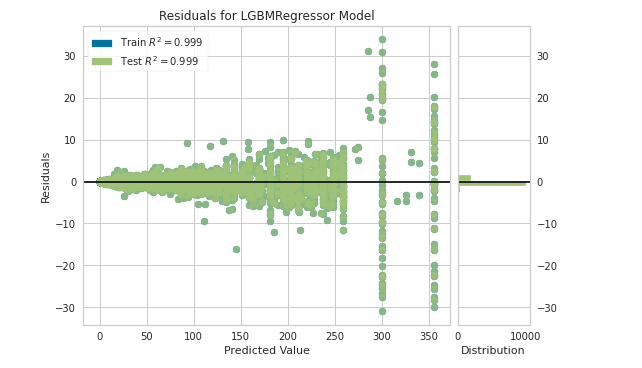
gbr's rmse after training: 0.8460081969703255

Computational runtime of this algo: 85.49 seconds



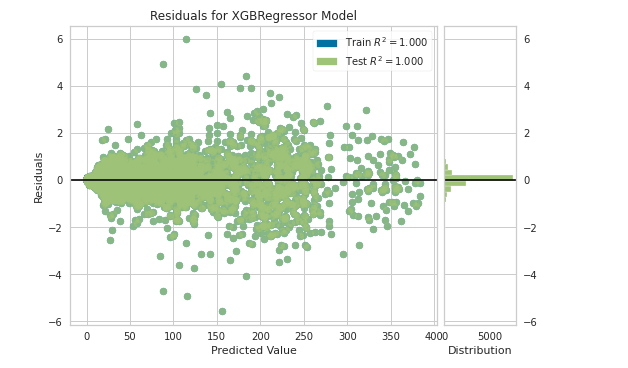
lgbm's rmse after training: 1.7597254824546376

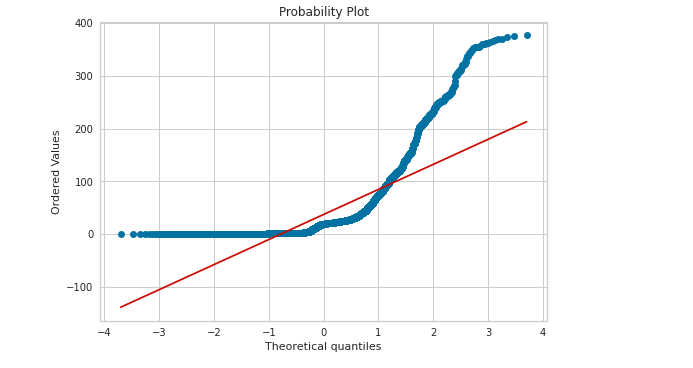
Computational runtime of this algo: 13.24 seconds



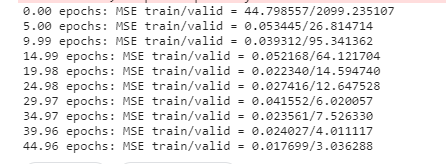
xgboost's rmse after training: 0.6893820259407684

Computational runtime of this algo: 13.36 seconds

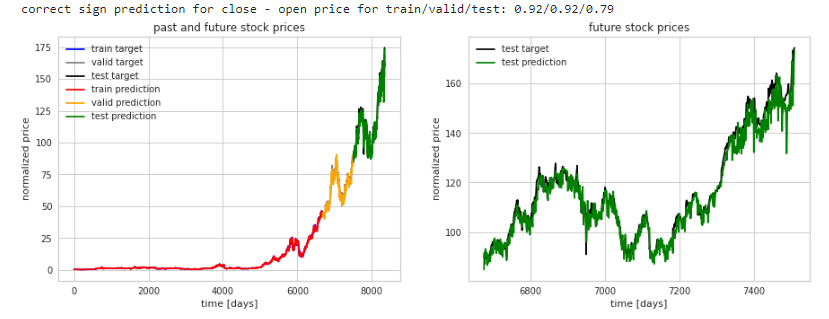




LSTM Epochs



**Final Outcome**



**References**

[1] S. Nikita, S. Abhijit, and S. Aayush, 2021. LSTM FOR STOCK MARKET PREDICTION Irjet.net (online IRJETV6I3219.pdf is available at: <https://www.irjet.net/archives/V6/i3/IRJETV6I3219.pdf> [Accessed November 30, 2021]

[2] R. Mangalampalli, P. Khetre, and V. Malviya, 2021. HYBRID ARIMA AND GRU MODELS FOR STOCK PREDICTION Ijert.org (online). Available at: <http://www.ijert.org/research/stock-prediction-using-hybrid-arima-and-gru-models> IJERTV9IS050550.pdf> [Accessed November 30, 2021]

[3] A. Nikfarjam, Emadzadeh, and S. Muthaiyah, 2021. Approaches to text mining for stock market prediction [online] https://www.academia.edu. Available at: http://www.academia.edu/10400600/Text mining approaches for stock market prediction [Accessed November 30, 2021].

[4] E. Chonga, C. Hanb, and F. Park, 2021. Methodology, Data Representations, and Case Studies for Deep Learning Networks in Stock Market Analysis and Prediction [online] durham-research-online. URL: <https://vdocument.in/durham-research> onlinedroduracukdroduracuk21533121533pdf-deeplearning.html> [Accessed November 30, 2021].

[5]. A comparison of linear regression and support vector regression is presented in the online proceedings of the 2016 International Conference on Green Engineering and Technologies (IC-GET). S. Kavitha, S. Varuna, and R. Ramya.

[6]. "Granule Based Knowledge Representation for Intra and Inter Transaction Association Mining," Queensland University of Technology, Wanzhong Yang.  July 2009 technology

[7]. R.V. Argiddi and S. S. Apte, "A Study of Association Rule Mining in Fragmented Item-Sets for Transaction Outcome Prediction in Stock Trading Systems"  IJCET-2012.