Stock Analysis and Prediction Using ANN and Linear Regression

Abstract:-

For many years the following question has been a source of continuing controversy in both academic and business circles: To what extent can the past history of a common stock's price be used to make meaningful predictions concerning the future price of the stock? Although there are many different chartist theories, they all assume that the past behavior of a security's price is rich in information concerning its future behavior. History repeats itself in that "patterns" of past price behavior will tend to recur in the future. Application of artificial neural network architectures for stock market prediction has been actively studied in recent years. In recent years, there has been a lot of interest in the use of artificial neural network architectures and other machine learning algorithms like SVM, linear regression, bayesian tree for stock market prediction. The research proposes the use of Artificial Neural Network that is feed forward multi-layer perceptron with error backpropagation as well as few machine learning techniques and develops a model with 80% training data and 128 batch size. The results show that both architectures possess the ability to separate winners and losers from a sample of Edelweiss stocks, and the selected portfolios outperform the benchmark. Our study argues that ANN shows superior performance over machine learning techniques.

Introduction:-

Stock trading is the process of purchasing and selling publicly traded financial assets on a stock market portal, with hordes of buyers and sellers from around the world fully involving at any time the market is transparent. In the case of stocks, the market generally operates on a "willing buyer", "willing seller" trade, in which consumer and whole seller prices are compared for suitability. If there is no good fit, there is no deal, and the trade either wants to wait for an upcoming match or expires. So there are simply many such reasons and noises manipulating market movements, forecasting the stock market is an extremely challenging and complicated problem. A predictive tool that Stockbrokers can use to guide on exact price movements, as a basis of investment, is therefore desirable. This can be an artificial intelligence (AI) system based on neural networks. Other methods in time series prediction are linear regression, SVM, Decision Tree, Random Forest etc.

Problem Statement:-

Trading in shares is big business in many economies. Based on the information on their websites, Stockbrokers do not seem to have any intelligent tool that can help them advise clients on which stocks are suitable for any buy or sale trade. These websites provide information that points to use of fundamental, technical and time series analysis methods, as depicted. These prevalent methods show a trend on future movement and not the likely trade price for any stock in future. It is therefore desirable to have a tool that does not just point a direction of price movement, but also indicates the most likely price value of the stock itself. A Linear Regression and an ANN model that is well tuned with the appropriate parameters can be used to develop such a predictive tool.

Data Pre-processing:-

The sample data for the study were provided by Edelweiss Broking Limited. The dataset contains the following attributes: (date, open, high, low, close, adjclose, volume). These data are collected and presented on a quarterly basis. We had to clean up the raw data before we could fit it to a model.

1> **Dropping of Features:-**Features having high volume of values missing were eliminated across dataset (stock).

2> **Completing Missing Entries:-**By comparing their values of the missing items to the average of the nearby entries by mean substitution method the missing items are replaced.

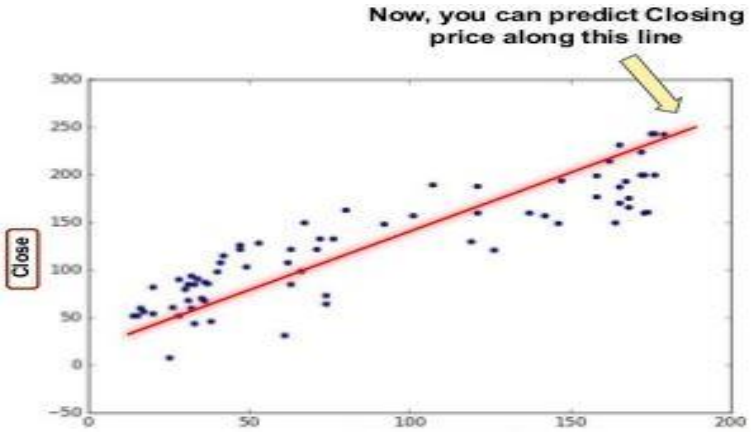
3> **Normalization:-**For better interpretation of the dataset we normalize the data.

4> **Standardization:-**Because the features scales may vary, standardization is used to improve the performance of all of our forecasting models' features.

5> **Creating a Timeline:-**We chose data spanning the first quarter of 2008 to the fourth quarter of 2018. Stocks that were discovered after the first quarter of 2008, the earliest known observation, were removed from the Stock universe.

	Open	High	Close	Volume
0	41.299999	41.685001	40.525002	0.002930
1	39.700001	40.799999	39.564999	0.002860
2	40.000000	40.599998	36.119999	0.002611
3	35.000000	44.400002	40.880001	0.002955
4	40.400002	40.400002	37.910000	0.002741

Linear Regression:- Linear regression is one of the regression methods that can be used to classify numerical data. It generates a linear function by computing weight values (w) for each feature (). The following is an example of the function:
X represents the regression value for a single instance of data. Here is an illustration of linear regression to help you understand it better.

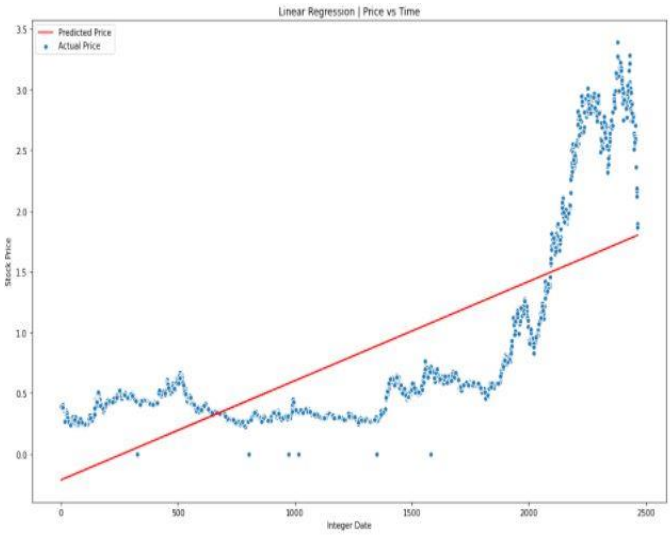


As seen in the image above, the distribution of data is represented by the linear line. The error or residual is the distance between each instance data and the linear line. In order to minimize the mean error, the best weight value of each of the feature is located while creating the linear function.

Predicted Method:-

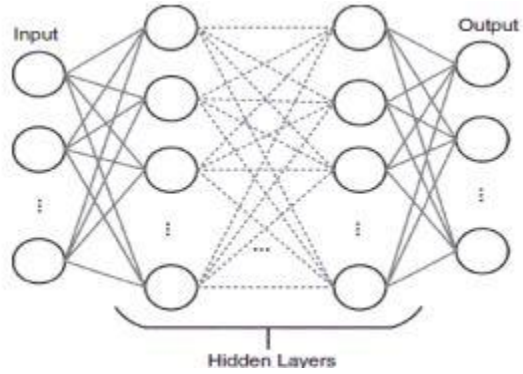
First, we separate the data into two sections: training and testing. The training section is then used to begin analysis and define the model. Data was used for training purposes 80 percent of the time and testing purposes 20 percent of the time. The transformation process's goal has been making our represent more structured. Preprocessed data includes general real valued data such as currency amounts, percentage formats. The pre-processed data is standardized into percentage-oriented data during the data Transformation process. Percentage of ordinal conversion table lists the percentage value ranges that correspond to their ordinal enumerated values. The conversion table is used to assign each enumerated value the dataset. This method produces identical categories enumerated values for different variables, even when the range of vales for 1 variable differs from the range of values for another variable. This method also clarifies the categories within the variable whose numerical values vary widely from 1 range the next. During the data training stage every regression classifier has used a predictive analytic over the dataset one after the other. Percentage split instructs regression classifier t divide the dataset proportionally into training and testing data. Training data facilitates the learning process for each classifier in order for it to develop its own regression rules. The regression rule

was applied to the testing data to forecast future stock price trends. The outcome of the tests was then evaluated.



Predicted Values of Stocks by Linear Regression

ANN-Based Stock Prediction Model:- Artificial neural networks are basically designed to fathom complicated issues that normal machine learning algorithms or easy neural networks cannot. Artificial Neural Networks simulates the working of a human brain and obviously determines accuracy of the predictions on its own, without the intervention of human brain. Artificial neural networks are capable of learning what they parse and can generalize or create patterns with the derived knowledge. . ANNs are far different from the existing traditional methods in training or precisely programming systems as they need detailed rules, which cover at the outset each possible outcome. The process of discriminating the category into which piece of data belongs to can be defined as classification task; Prominent use of this approach is in programming a neural network. Such ability in classifying live patterns or examples is generalization.



Description of an ANN

Training the Network- ANNs are much used for teaching the network. This method of teaching a network is termed as training the network. Training a network literally means, fitting a network to our training data set. This method is inspired from algebraic equations where in the process is to fit information to a mathematical model or time series line. . Training of ANNs means the model is programmed to handle well the given tasks based on unforeseen scenarios. When an artificial network is insufficiently trained, it's called under fitting that implies network has not learnt that particular training set adequately. We choose data spanning the first quarter of 2008 to the fourth quarter of 2018. Stocks that were discovered after the first quarter of 2008, the earliest known observation, were removed from the Stock universe. Before testing results are generated, for final model training the validation set is combined with the training set .

EXPERIMENT-

1. Relative Return-

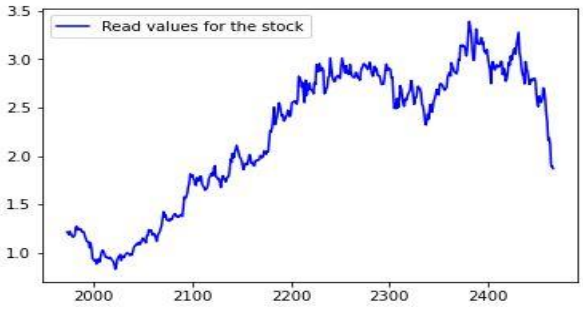
In the trial, the absolute return of a stock data is calculated. The relative return is defined as the difference in a stock's absolute return against return some benchmark. The factors affecting the broader market can be filtered out by finding the difference between the overall market performance and the performance each individual stock. Theoretically applying such a technique benefits in reducing the complexity of the prediction problem while improving model stability and prediction performance.

2. Portfolio Construction-

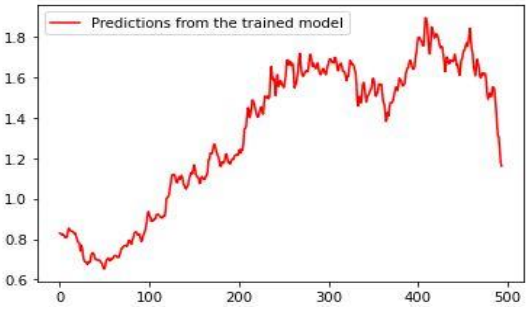
After the models generate simulated results for individual stocks, these stocks are graded based on their quarterly predicted relative returns. Stocks having highest prediction of relative returns are used to build a `Buy' portfolio, while stocks with the lowest predicted relative returns are used t build a `Sell' portfolio. As part this experiment, we build portfolios using an equal-weight strategy. This implies that the hypothetical investment should be spread evenly over the stocks in the portfolio. During the model validation stage, portfolios get built for each quarter of the validation set, and in the test set during the final testing stage.

Results and Observation:-

Our research study's goal is to assist stock brokers and investors in making stock market investments. The prediction plays an important role in stock market business, which is a complicated and challenging process due to the stock market's dynamic nature. This study demonstrates that when the data input is standardized in a common data type via the customized transform process, the results for regression techniques can be improved. Thus for ordinal information for predicting based upon ranking system adds new dimension to predicting output. On the validation set, both ANN and linear regression models are validated. Following validation, we decide on an ANN with hidden layers. The function activation of the hidden layer and output layer were rectifier and sigmoid, respectively. For the loss function, the ANN employs mean squared error, and the Adam algorithm is used for optimization.



Real Values of Stocks



Predicted Values of Stocks by ANN