JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

STOCK PRICE PREDICTION USING MACHINE LEARNING

Prediction using Long Short Term Memory Algorithm

[1]Junaid Mandviwala, [2]Suraj Jadhav, [3]Chinmay Joshi, [4]Amey Nagotkar, [5]Shrawan Vishwakarma

[1]Head of Electronics and Telecommunication Engineering Department, [2]Student, [3]Student, [4]Student, [5]Student

[1]BE Electronics and Telecommunication Engineering,
[1]Rizvi College of Engineering, Mumbai, India

Abstract: For many business analysts and academics, predicting stock market values has always been a difficult assignment. Indeed, stock market price forecasting is a fascinating field of study for investors. Many investors want to know how the market will perform in the future in order to make successful investment decisions. Indirectly effective prediction algorithms assist traders by providing supportive information such as the market direction in the future. By applying diverse algorithms to data, data mining techniques are useful for projecting the future. This project tries to increase the quality of output by anticipating stock market movements using financial news, analyst opinions, and quotes. It presents a revolutionary method for predicting the closing price of the stock market. Many researchers have contributed in some way to the field of chaotic forecasting. So far, the typical methodologies have been fundamental and technical studies. For share market prediction, another prominent method for identifying unknown and hidden patterns in data is recurrent neural networks (RNN).

IndexTerms **- Stock Price. Stock Market, Prediction, Machine Learning, RNN**

I. INTRODUCTION

The stock market refers to a collection of exchanges and other venues where shares of publicly traded firms can be bought, sold, and invested. Such financial transactions are carried out through institutionalised formal exchanges (whether real or virtual) that function under a set of rules.

A stock exchange is a highly regulated and controlled environment. The National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE) are India's two main stock exchange regulators (BSE). The stock market promotes fair pricing and transaction transparency by bringing together hundreds of thousands of market participants who want to purchase and sell shares. Unlike earlier stock markets, which issued and traded paper-based physical share certificates, today's computerised stock markets are fully electronic.

The stock market, as a primary market, allows companies to issue and sell shares to the general public for the first time via an initial public offering (IPO). This practice assists firms in receiving the necessary cash from investors. It basically means that a company divides itself into numerous shares and sells a portion of those shares to the public for a cost.

II. PROBLEM STATEMENT AND OBJECTIVE

Every day, the stock market makes headlines. Every time it hits a new high or a new low, you hear about it. If an effective algorithm could be established to anticipate the short-term price of a single stock, the rate of investment and business opportunities in the stock market may increase. Artificial Neural Networks and Convolutional Neural Networks have been used in the past to predict stock prices, with a 20 percent error loss on average. In this project, we'll look at whether a Recurrent Neural Network-based model can be used to predict stock price with a lower percentage of inaccuracy. If the answer is YES, we'll see how trustworthy and dependable it is.

III. LITERATURE SURVEY

- Paper 1: LSTM FULLY CONVOLUTIONAL NETWORKS FOR TIME SERIES CLASSIFICATION
  - Fazle Karim, Somshubra Majmudar, Houshang Darabi and Shun chen

We make a significant improvement in the existing state-of-the-art for time series classification using deep neural networks with the proposed models. Our baseline models, with and without fine-tuning, can be trained end-to-end with minimal preprocessing

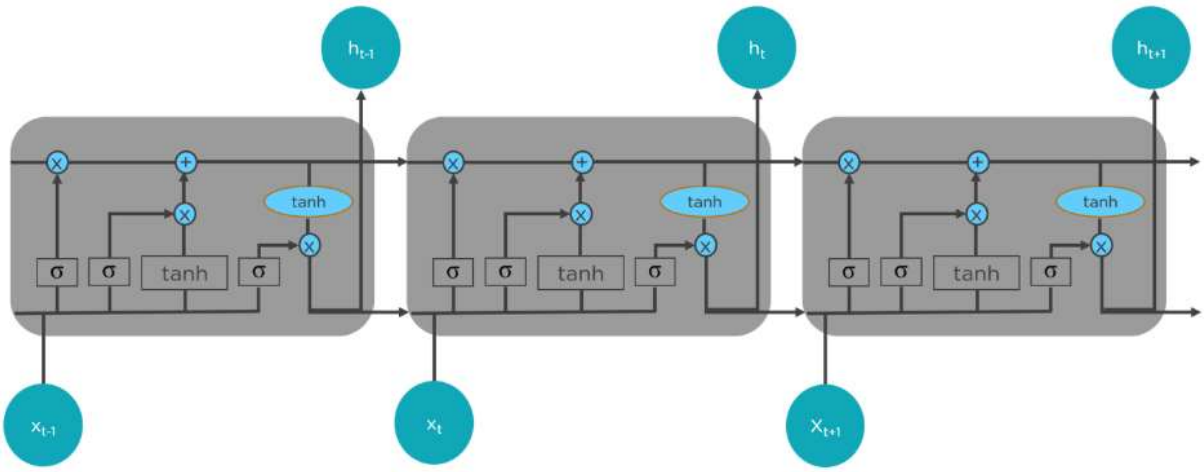
and generate much better results. With a modest increase in the number of parameters, LSTM-FCNs can significantly improve the performance of FCN models. The capacity to visually evaluate the decision process of the LSTM RNN
and offer a robust baseline on their own is provided by an LSTM-FCN. To improve the performance of a model, fine-tuning can be used as a general approach. The significant improvement in performance over FCN models demonstrates that LSTM RNNs can enhance the performance of FCN modules for time series classification. An overall study of our model's performance is presented, as well as comparisons to other methodologies. More study is needed to figure out why the attention LSTM cell fails to match the performance of the general LSTM cell on some datasets. In addition, an extension of the suggested models to multivariate time series is simple but not studied in this paper.

- Paper 2: STOCK MARKET TRENDS PREDICTION AFTER EARNING RELEASE
  - Chen Qian, Wenjie Zheng

The stock market is known to the general public as a chaotic system, and it has been proven that even models designed with empirical important elements can still have low accuracy. We tried to confine our scope to earnings release day in our research, and it turned out that we could create models with a prediction accuracy of roughly 70%. We use two sets of input features to build the model: financial statistics from a company's quarterly earnings report, market surprise due to consensus expectations in terms of digital data, and sentiment analysis of relevant articles from mainstream media of financial professionals to predict stock market movements in the after-hour period and trend the day after the release day. Experiments indicate that SVM and LWLR models outperform other models because they control data correlation, as explained in Section VI. However, we have a tiny data size (300 samples) due to the limited number of firm options, which could lead to high bias and overfitting. Company orientation and future business guidance, which are difficult to quantify, affect the stock price in addition to specific financial elements and consensus news.

IV. USED ALGORITHM

We will create our model to predict the stock values of Indian stocks using a Long Short Term Memory Network (LSTM). Long-term dependency models (LTSMs) are a sort of recurrent neural network that is used to learn long-term dependencies. It's frequently used to process and forecast time-series data.



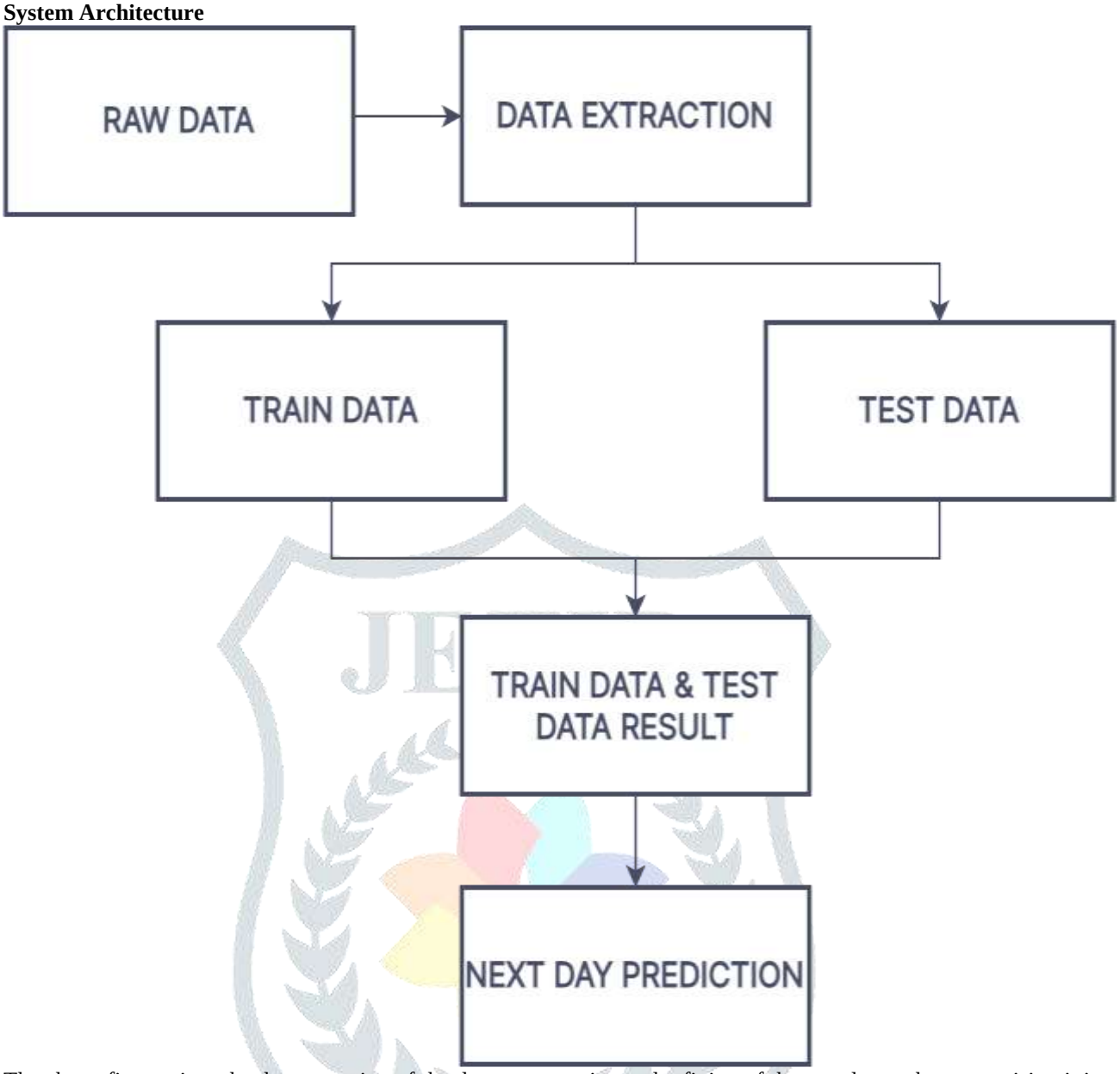
As we can see from the graphic at the top, LSTMs have a chain-like structure. A single neural network layer exists in general RNNs. LSTMs, on the other hand, are made up of four interconnected layers that communicate extremely well.

LSTMs work in a three-step process.

- The first step in LSTM is to decide which information to be omitted from the cell in that particular time step. It is decided with the help of a sigmoid function. It looks at the previous state (ht-1) and the current input xt and computes the function
- There are two functions in the second layer. The first is the sigmoid function, and the second is the tanh function. The sigmoid function decides which values to let through (0 or 1). The tanh function gives the weightage to the values passed, deciding their level of importance from -1 to 1.
- The third step is to decide what will be the final output. First, you need to run a sigmoid layer which determines what parts of the cell state make it to the output. Then, you must put the cell state through the tanh function to push the values between -1 and 1 and multiply it by the output of the sigmoid gate.

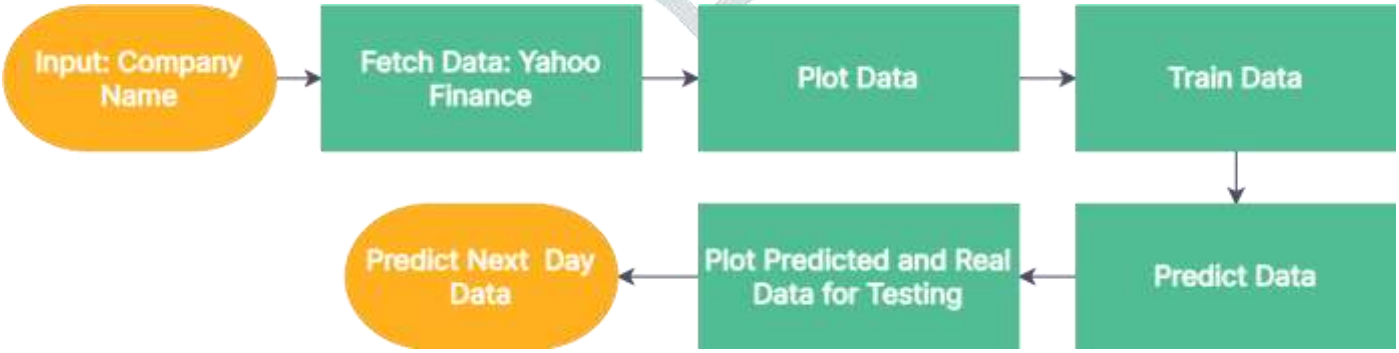
5. Methodology

- **System Architecture**



The above figure gives the demonstration of the dataset extraction and refining of the raw dataset by categorising it into two-phase of training and testing data. From the given dataset a well-modified categorization is extracted and a graph set is plotted to gain the required output which gives the stock prediction range.

- Data Flow Diagram



Data collection or data fetching is done from yahoo finance public api. Yahoo finance has all attributes related to stock such as stock price history, ownership details, Low, High, Close, Open etc. After the data is fetched from the yahoo finance data cleaning is done, for prediction we require only the close price of the stock beside that all the other columns were dropped. Once data cleaning is done visualisation of the data is done by plotting data on the graph. Now we have cleaned data, we train data to predict the next day's closing price and use that to model future values as well as to predict the trend of the stock price and then visualisation for the same is done on the graph. All these steps were discussed in detail below.

- **Data and Sources of Data**

For machine learning we require raw data, without raw data training of machine learning algorithms is not possible. For this study secondary data has been collected from the website of Yahoo Finance using 'yfinance' api. For training of model data collected

from Jan 2008 to Jan 2022. For visualisation and prediction of close data collection period is ranging from January 2008 to present day (if today is 22 april 2022) then data range will be Jan 2008 - April 2022.

Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
2008-01-01	192.565269	193.778917	190.639620	192.848465	4502158	0.0	0.0
2008-01-02	191.918028	198.552624	189.490731	196.072739	10764075	0.0	0.0
2008-01-03	195.801702	197.201444	192.860627	193.799179	5993773	0.0	0.0
2008-01-04	208.261801	208.261801	192.322575	193.277313	5185640	0.0	0.0
2008-01-07	194.078309	196.448962	191.028015	194.418137	6612148	0.0	0.0
...	...	...	...	...	...	...	...
2022-04-19	513.250000	520.950012	505.100006	511.799988	20362934	0.0	0.0
2022-04-20	512.000000	513.700012	507.000000	509.299988	11881305	0.0	0.0
2022-04-21	513.400024	518.000000	512.200012	516.299988	10035638	0.0	0.0
2022-04-22	512.900024	513.500000	499.000000	500.600006	16760456	0.0	0.0
2022-04-25	495.000000	499.200012	490.049988	494.750000	14057444	0.0	0.0

3526 rows × 7 columns

- **Data Processing**

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always the case that we come across clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put it in a formatted way. So, for this, we use data pre-processing task

- **Data Splitting**

The train-test split procedure is used to estimate the performance of machine learning algorithms when they are used to make predictions on data not used to train the model. In this step data is split for training and testing almost 80% of data is for training and 20% for testing is a basic rule in machine learning.

- **Training**

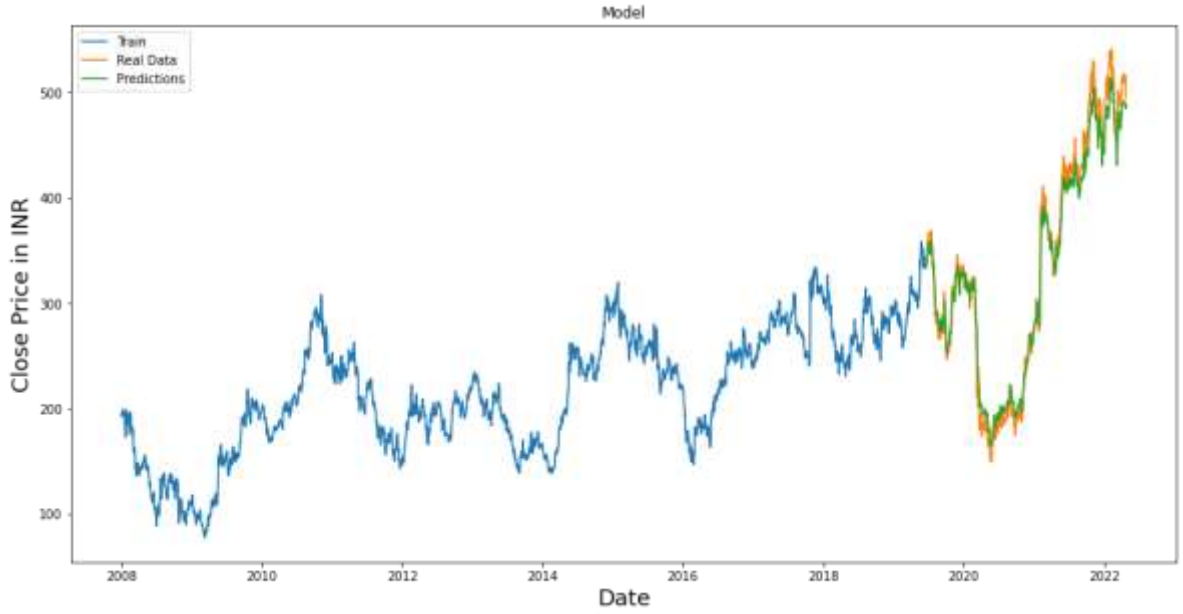
In this step, we do training data for machine analysis itself and we do another step is to validate training data because training data set will produce either overfitting or under the fitting problem that means false positive output or true negative output that means overfitting means when you go new area and 1st person give disrespect and you considering all people are same this is.

- **Testing and Evaluation:**

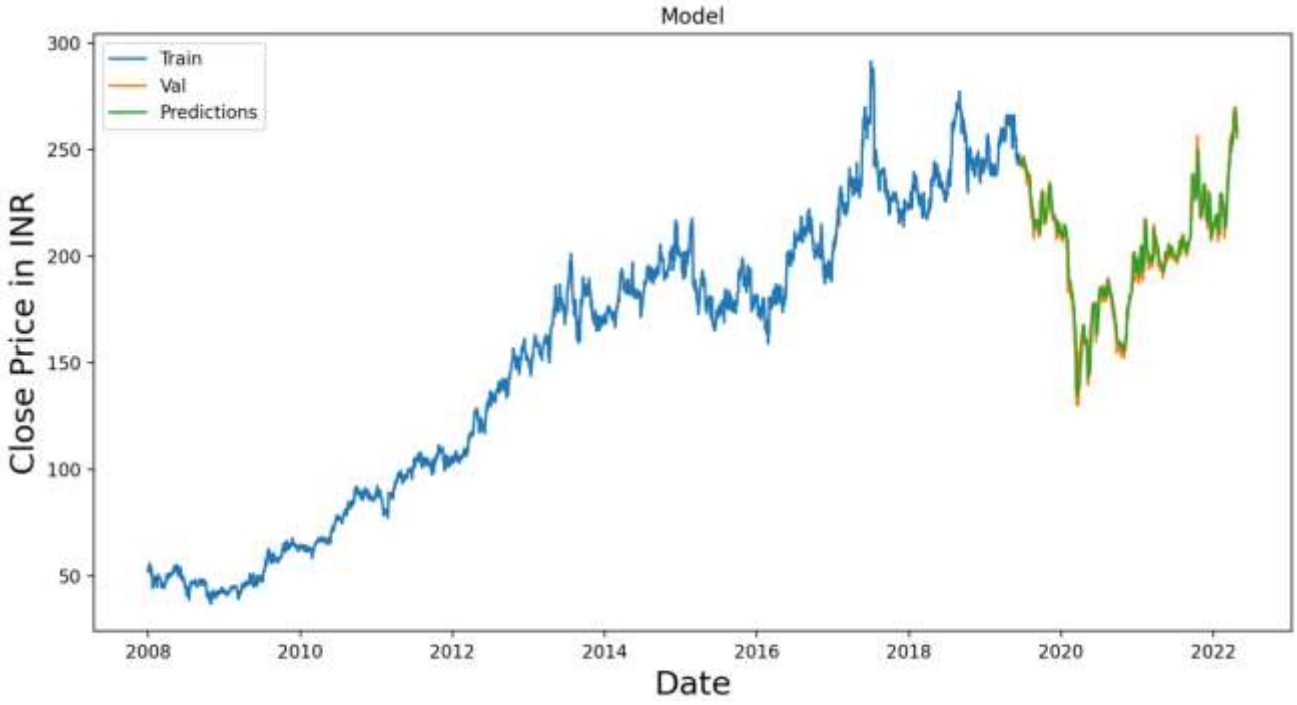
In this testing phase, we tested the model using cross-validation, we checked the stocks original price value and predicted value of the model. Matplotlib is a data visualisation library in python that provides an advanced interface for drawing attractive and informative mathematical graphics.

Here are some examples of prediction models.

1. Ticker Name: SBIN.NS (State Bank of India)
   Date: 26 April 2022
   Predicted Price for 27 April 2022: 487.26 rupees
   Closing Price for 27 April 2022: 497.25 rupees



2. Ticker Name: ITC.NS
   Date: 27 April 2022
   Predicted Price for 28 April 2022: 267.43 rupees
   Closing Price for 28 April 2022: 260.05 rupees



6. Tools Used

Hardware:
Minimum requirements needed to install and run are as follows:

1. Operating System: Windows 7 or upper, Linux, Mac
2. Processor: Dual core 2.4 GHz
3. RAM: 4 GB

Software:

1. Any Web Browser

.

7. Conclusion

In this paper, a model is constructed that uses a Recurrent Neural Network with an LSTM cell to predict correct future stock values for time series data. The stock price predicted is quite close to the actual price. In this experiment, the stock prices of five different companies were sampled. Based on this approach, the mean absolute value of the percentage of error is less than 2%. Also shown and fitted with a beta distribution was the absolute value of the error. The absolute value of the percentage of error, the coefficient of determination (R-squared), Pearson's correlation coefficient, Spearman's rank correlation coefficient, and explained variance

error were determined to validate the model's performance. The estimated findings revealed a fairly significant link between the two variables. between the current stock price and the expected stock price.

REFERENCES

[1] Fazle Karim, Somshubra Majmudar, Houshang Darabhi, and Shun Chen, 'LSTM fully convolutional networks for time series classification, *IEEE Access (Volume 6), 2017*

[2] Chen Qian, Ran an, Wenjie Zheng, "Stock market trends prediction after earning release"

[3] Sunil, "Commonly Used Machine Learning Algorithms", - *09 september 2017*