A Project report on

STOCK MARKET TREND PREDICTION USING K-NEAREST NEIGHBOUR (KNN) ALGORITHM

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

K. AKSHITHA
19H51A0513

TANISHA GARG
19H51A0527

T. NIKHILA
19H51A0581

Under the esteemed guidance of

Dr S Kirubakaran
Associate Professor

Department of Computer Science and Engineering
CMR COLLEGE OF ENGINEERING & TECHNOLOGY
(An Autonomous Institution under UGC & JNTUH, Approved by AICTE, Permanently Affiliated to JNTUH, Accredited by NBA.)
KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2019- 2023

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase-1 report entitled **"STOCK MARKET TREND PREDICTION USING K-NEAREST NEIGHBOUR (KNN) ALGORITHM"** being submitted by K. Akshitha (19H51A0513), Tanisha Garg (19H51A0527), T. Nikhila (19H51A0581) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Dr S Kirubakaran
Associate Professor
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Dr S Kirubakaran, Associate Professor**, Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academic, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the Teaching & Non- teaching staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Mr. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

K.Akshitha 19H51A0513
Tanisha Garg 19H51A0527
T.Nikhila 19H51A0581

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	LIST OF FIGURES	ii
	ABSTRACT	iii
1	**INTRODUCTION**	1
	1.1 Problem Statement	2
	1.2 Research Objective	3
	1.3 Project Scope and Limitations	3
2	**BACKGROUND WORK**	4
	2.1. Methodology	5
	2.2. K-Nearest Neighbour	6
	2.2.1.Introduction	6
	2.2.2.Merits and Demerits	6
	2.2.3.Implementation Results	7
	2.3. Support Vector Regression	8
	2.3.1.Introduction	8
	2.3.2.Merits and Demerits	8
	2.3.3.Implementation Results	9
	2.4. Long Short Term Memory	10
	2.4.1.Introduction	10
	2.4.2.Merits and Demerits	10
	2.4.3.Implementation Results	11
3	**RESULTS AND DISCUSSION**	**12**
	3.1. Comparison of Existing Solutions	13
4	**CONCLUSION**	**15**
	4.1 Conclusion	16
5	**REFERENCES**	17

List of Figures

FIGURE NO.	TITLE	PAGE NO.
2.1	Methodology	5
2.2	Stock analysis using KNN	7
2.3	Stock analysis using SVR	9
2.4	Stock analysis using LSTM	11
3.1	Predicted data by KNN	13
3.2	Predicted data by SVR	13
3.3	Predicted data by LSTM	14
3.4	Comparison between the algorithms	14

ABSTRACT

Stock prediction has always been a challenging problem for statistics experts and finance. The main reason behind this prediction is buying stocks that are likely to increase in price and then selling stocks that are probably to fall. Generally, there are two ways for stock market prediction. Fundamental analysis is one of them and relies on a company's technique and fundamental information. We will be evaluating performance of KNN(K-Nearest Neighbor)supervised machine learning algorithm. In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange. The programming language that is used to predict the stock market using machine learning is Python. In this project we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called K-Nearest Neighbor to predict stock prices for the large and small capitalization's and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

CHAPTER 1
INTRODUCTION

CHAPTER 1
INTRODUCTION

Stock market movement has always been ambiguous for investors because of various influential factors. This study aims to significantly reduce the risk of trend prediction with machine learning and deep learning algorithms. Stock Market prediction remains a secretive and empirical art. Few people, if any, are willing to share what successful strategies they have. A chief goal of this project is to add to the academic understanding of stock market prediction. The hope is that with a greater understanding of how the market moves, investors will be better equipped to prevent another financial crisis. The project will evaluate some existing strategies from a rigorous scientific perspective and provide a quantitative evaluation of new strategies.

1.1 PROBLEM STATEMENT

Stock market attracts thousands of investors' hearts from all around the world. The risk and profit of it has great charm and every investor wants to book profit from that. People use various methods to predict market volatility, such as K-line diagram analysis method, Point Data Diagram, Moving Average Convergence Divergence, even coin tossing, fortune telling, and so on. Now, all the financial data is stored digitally and is easily accessible. Availability of this huge amount of financial data in digital media creates appropriate conditions for a data mining research. The important problem in this area is to make effective use of the available data.

1.2 RESEARCH OBJECTIVE

The prediction of share prices is the function of deciding the future price of a company stock or other commercial tool traded. Prediction of some movements allowed from some patterns can be found. People are always attracted to invest in share market and stock exchanges as they provide huge financial profits, which is also an important for finance research. Prediction of share price is very difficult issue it depends upon such huge numbers of factors such organization financial status and national policy and so on. Nowadays stock costs are influenced because of numerous reasons such as organization related news, political, socially efficient conditions and cataclysmic events. Many studies have been performed for the prediction of stock index value and daily direction of change in the stock index. Such huge numbers of models have been created for foreseeing the future stock costs yet everyone has their own weaknesses. This project expects to study, develop and assess different techniques so as to foresee future stock trades..

1.3 PROJECT SCOPE AND LIMITATIONS

The research helps a lot of new investors in deciding when to buy or sell a particular stock. It also helps in understanding the sentiments of experienced financial analysts and financial news data more quickly than doing the same manually. Disadvantages Current research makes use of neural networks which have the drawback of slow convergence rate and local optimum. To overcome the problem of slow convergence the author uses a pattern matching algorithm to select the input data to train the network which is an increased overhead. We will implement the system using other different machine learning techniques like Support Vector machine, decision trees, long short term memory and knn.

CHAPTER 2
BACKGROUND WORK

CHAPTER 2
BACKGROUND WORK

2.1 METHODOLOGY

The existing algorithms that are used to help predict stock market data will be implemented with the same dataset so it helps for comparison of the results. First for the dataset we are considering the stock opening and closing prices of Reliance company for day trading.

The dataset is first divided into training and test datasets and is pre-processed by sorting, removing null values and reducing noise so that we have a clean dataset. Then this is passed to the algorithms that will first be trained and then be tested each giving graph as result for easier comparision.

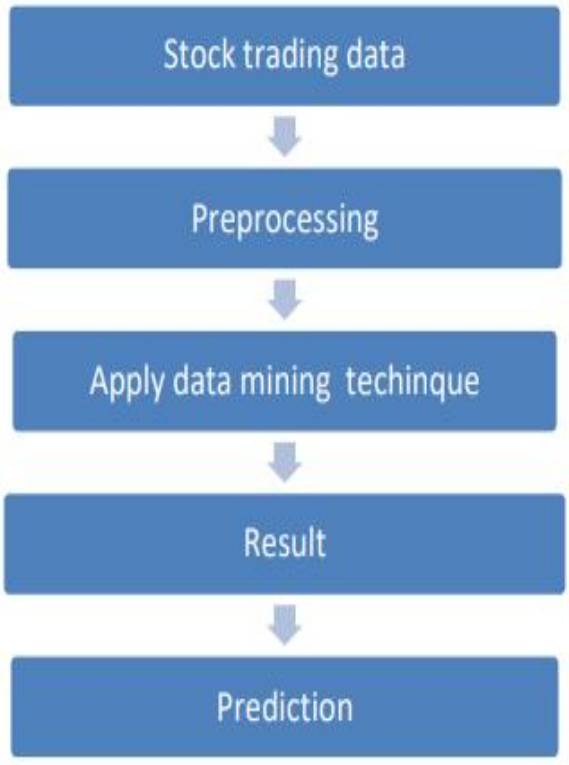

Figure 2.1 Methodology

2.2 K-NEAREST NEIGHBOUR ALGORITHM

2.2.1 INTRODUCTION

K-nearest neighbor technique is a machine learning algorithm that is considered as simple to implement. The stock prediction problem can be mapped into a similarity based classification. The historical stock data and the test data is mapped into a set of vectors. Each vector represents N dimension for each stock features. Then, a similarity metric such as Euclidean distance is computed to take a decision. In this section, a description of kNN is provided. kNN is considered a lazy learning that does not build a model or function previously, but yields the closest k records of the training data set that have the highest similarity to the test (i.e. query record). Then, a majority vote is performed among the selected k records to determine the class label and then assigned it to the query record. The prediction of stock market closing price is computed using kNN as follows:

a) Determine the number of nearest neighbors, k.

b) Compute the distance between the training samples and the query record.

c) Sort all training records according to the distance values.

d) Use a majority vote for the class labels of k nearest neighbors, and assign it as a prediction value of the query record.

2.2.2 MERITS AND DEMERITS

- No Training Period: KNN is called Lazy Learner (Instance based learning). It does not learn anything in the training period. It does not derive any discriminating function from the training data. In other words, there is no training period for it. It stores the training dataset and learns from it only at the time of making real time predictions. This makes the KNN algorithm much faster than other algorithms that require training e.g. SVM, Linear Regression etc.
- Since the KNN algorithm requires no training before making predictions, new data can be added seamlessly which will not impact the accuracy of the algorithm.

- KNN is very easy to implement. There are only two parameters required to implement KNN i.e. the value of K and the distance function (e.g. Euclidean or Manhattan etc.)
- Does not work well with large dataset: In large datasets, the cost of calculating the distance between the new point and each existing points is huge which degrades the performance of the algorithm.
- Does not work well with high dimensions: The KNN algorithm doesn't work well with high dimensional data because with large number of dimensions, it becomes difficult for the algorithm to calculate the distance in each dimension.
- Need feature scaling: We need to do feature scaling (standardization and normalization) before applying KNN algorithm to any dataset. If we don't do so, KNN may generate wrong predictions.
- Sensitive to noisy data, missing values and outliers: KNN is sensitive to noise in the dataset. We need to manually impute missing values and remove outliers.

2.2.3 IMPLEMENTATION RESULTS

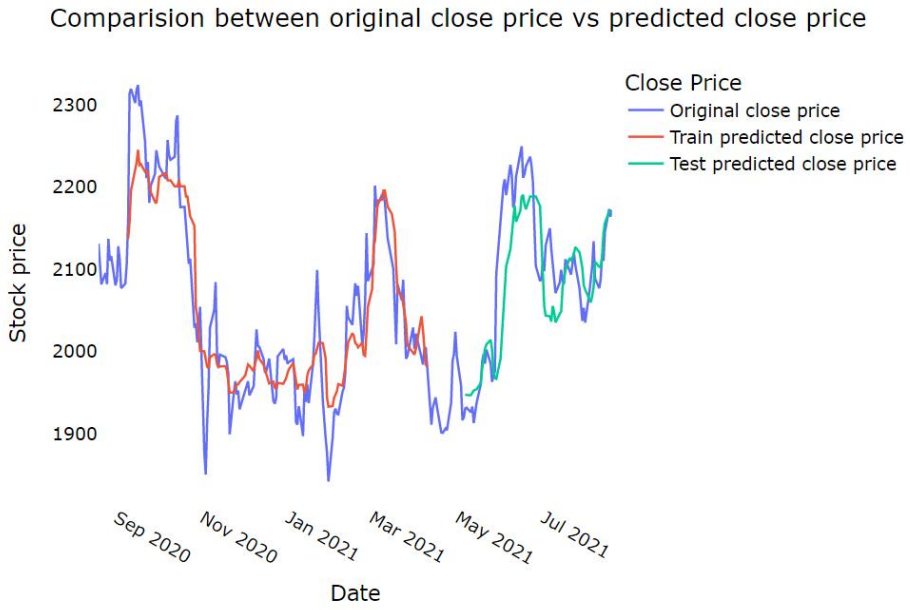

Figure 2.2 Stock analysis using KNN

2.3 SUPPORT VECTOR REGRESSION

2.3.1 INTRODUCTION

- SVM for regression is popularly and widely used for classification problems in machine learning, to predict stock market price and trend.
- It can also be used as a regression method, maintaining all the main features that characterize the algorithm.
- It is a useful and powerful technique to recognize pattern of time series dataset. It can produce good prediction result if the value of parameters can be determined properly.

2.3.2 MERITS AND DEMERITS

- Support vector regression can predict stock prices for small and large capitalization ,employing prices with both daily and up-to the minute frequencies.
- SVR have a predictive power especially when using a strategy of updating the model periodically.
- It is effective in instances where the number of dimensions is larger than the number of specimens.

- The support vector regression is missing with optimization technique further it does not have any dedicate method to deal with missing data.
- It is obtained at that the proposed model with optimized parameters take less amount of time to predict the closing price with each dataset. When compared to SVR which takes around a minute on average.

2.3.3 IMPLEMENTATION RESULTS

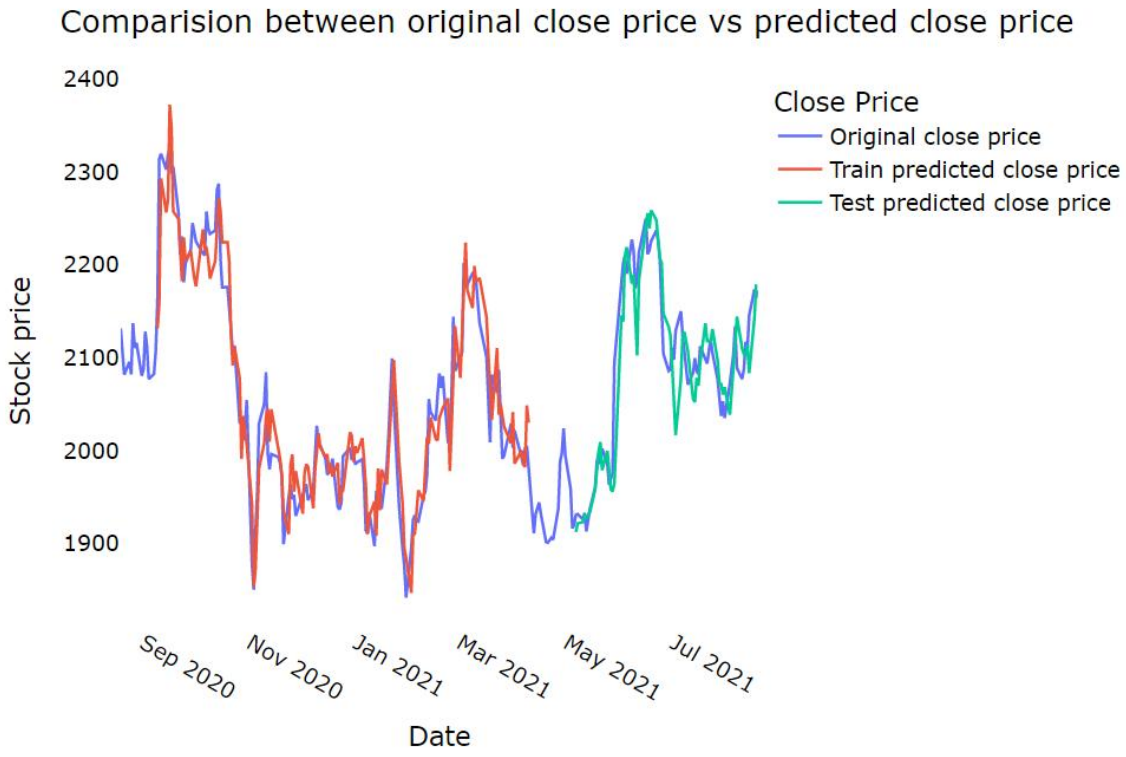

Figure 2.3 Stock analysis using SVR

2.4 LONG SHORT TERM MEMORY

2.4.1 INTRODUCTION

- LSTMs are widely used for sequence prediction problems and have proven to be extremely effective. The reason they work so well is that LSTM can store past important information and forget the information that is not.
- It has been so designed that the vanishing gradient problem is almost completely removed, while the training model is left unaltered.

2.4.2 MERITS AND DEMERITS

- Long time lags in certain problems are bridged using LSTMs where they also handle noise, distributed representations, and continuous values.
- With LSTMs, there is no need to keep a finite number of states from beforehand.
- LSTMs provide us with a large range of parameters such as learning rates, and input and output biases.

- They are in fact characterized by high noise-to-signal ratio, which makes it difficult for a machine learning model to find patterns and predict future prices.
- LSTMs take longer to train. Dropout is much harder to implement ,they are sensitive to different random weight initialization.

2.4.3 IMPLEMENTATION RESULTS

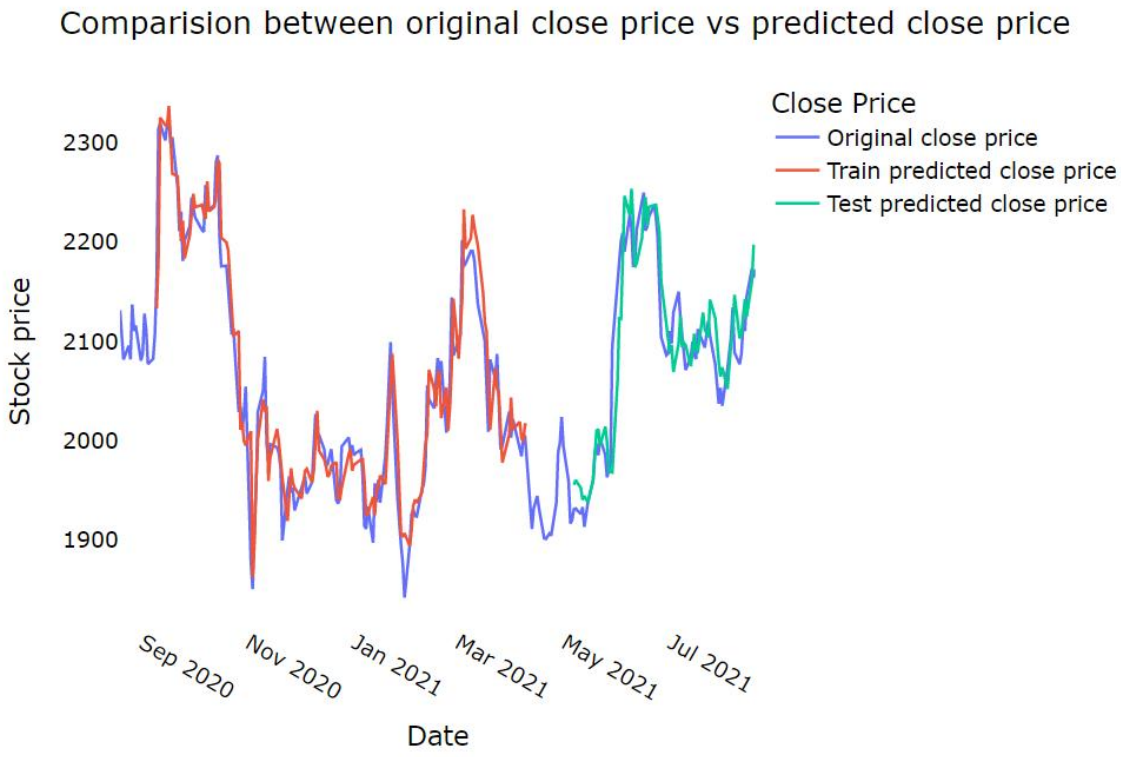

Figure 2.4 Stock analysis using LSTM

CHAPTER 3
RESULTS AND DISCUSSION

CHAPTER 3

RESULTS AND DISCUSSION

3.1 COMPARISION OF EXISTING SOLUTIONS

The below figure shows the final output using KNN algorithm.

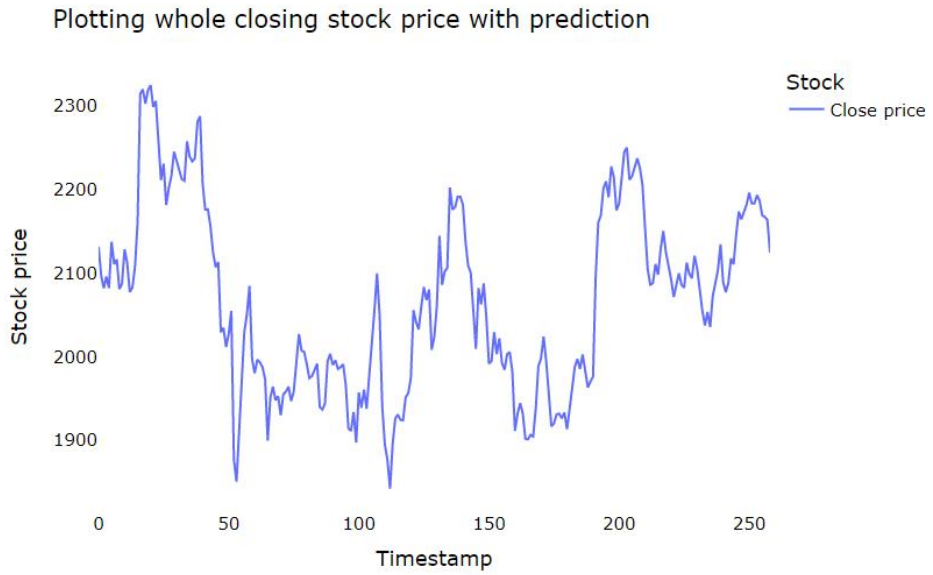

Figure 3.1 Predicted data by KNN

The below figure shows the final output using SVR algorithm.

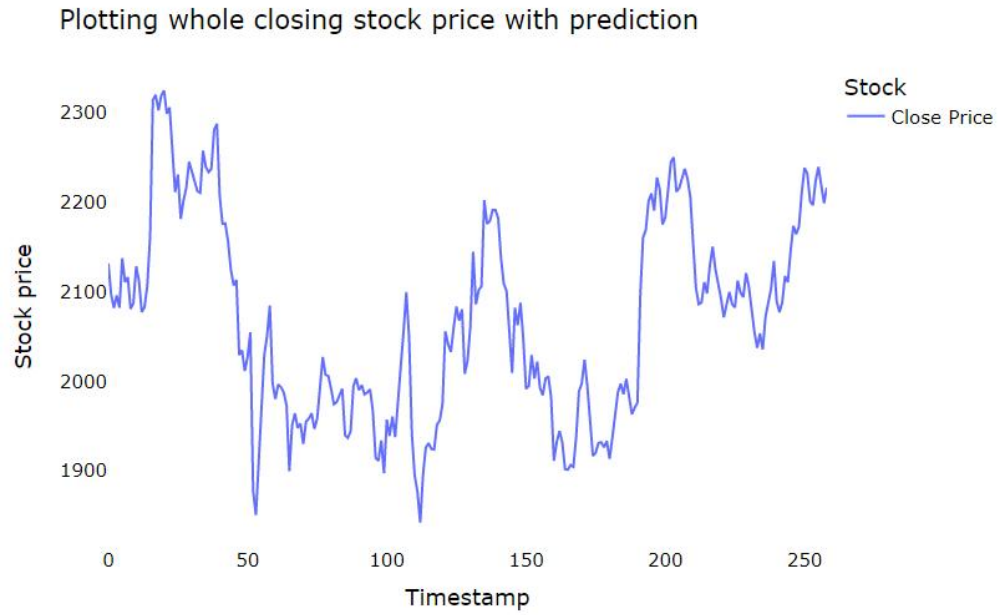

Figure 3.2 Predicted data by SVR

The below figure shows the final output using LSTM algorithm

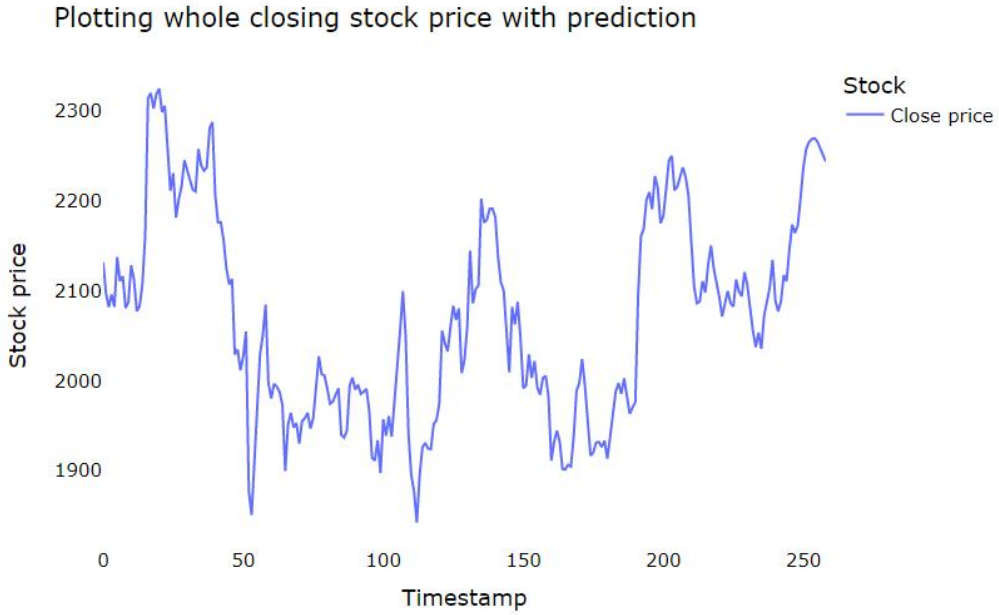

Figure 3.3 Predicted data by LSTM

The below figure shows the comparison between the above 3 algorithms and its clearly seen that LSTM is more accurate when compared with KNN and other algorithms.

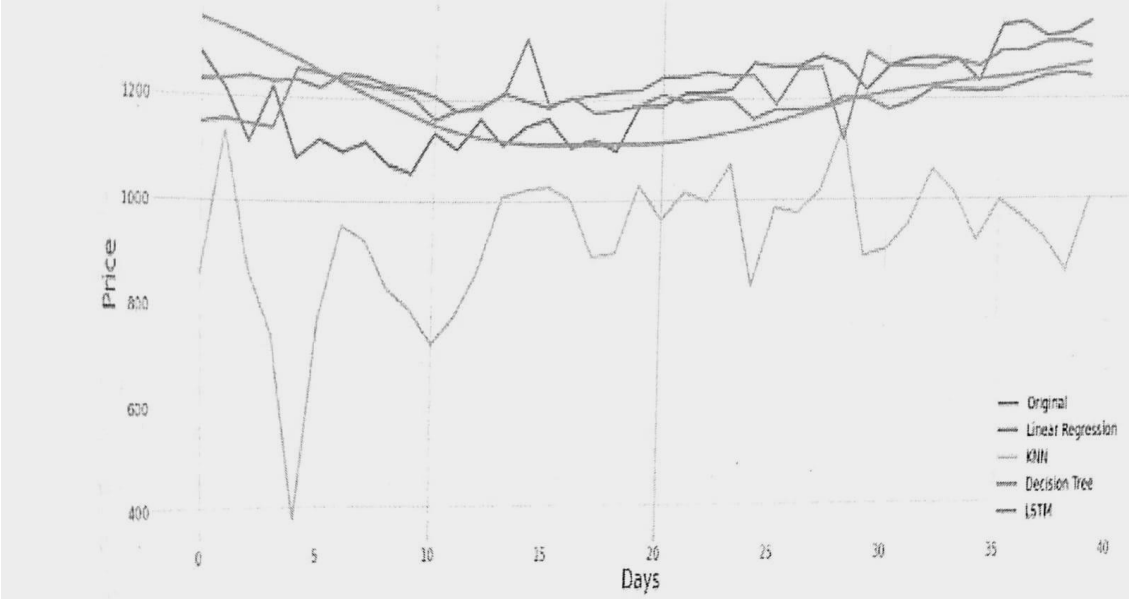

Figure 3.4 Comparison between the algorithms

CHAPTER 4
CONCLUSION

CHAPTER 4
CONCLUSION

4.1 CONCLUSION

On comparison of the algorithms KNN, SVR and LSTM it was found that all the algorithms perform well and that LSTM has the best performance but KNN algorithms only drawback is that it does not consider the non-centric data points. KNN requires lesser time to predict and is very easy to implement so efficiency can be improved by an upgraded version of KNN by combining it with Bayes theorem that considers these data points. KNN algorithm will calculate the distance from all the points in the proximity of the unknown data and will filter out the ones nearer to it. The cleaning and cleansing of data will be done after the raw data has been imported and processed. After the computational mean the output or the final layout will be displayed in the form of graph. To overcome the drawbacks of these algorithms we propose a solution which combines the KNN with Baye's theorem and which might give more efficient results.

CHAPTER 5
REFERENCES

CHAPTER 5
REFERENCES

[1]. https://www.analyticsvidhya.com/blog/2021/12/stock-price-prediction-using-lstm/#h2_2

[2]. "Stock market prediction using data mining techniques"- https://deliverypdf.ssrn.com/delivery.php?ID=068096001126025103068084119125115117122011005035010028104088106024123021074004070091031007000042033007027068118097123006121078049007088022007109085104119125089097105024020000023076021112001067095083115092072005000004104103101089083127116088104112073091&EXT=pdf&INDEX=TRUE

[3]. https://www.emerald.com/insight/content/doi/10.1108/AGJSR-05-2022-0053/full/html#sec004

[4]. https://www.kaggle.com/code/ysthehurricane/advanced-stock-pred-using-svr-rfr-knn-lstm-gru/notebook

[5]. https://www.projectpro.io/article/stock-price-prediction-using-machine-learning-project/571#mcetoc_1ftklmk3qp

[6]. https://www.analyticsvidhya.com/blog/2021/12/stock-price-prediction-using-lstm/#h2_2