iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55894

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

Stock Price Prediction using LSTM and KNN Algorithms

Vineet G Kowti

Abstract: ***The aim of this study is to enhance and evaluate the performance of Long Short-Term Memory (LSTM) and K Nearest Neighbor (KNN) algorithms when used to forecast stock prices. In this study, previously recorded data on price as well as other crucial factors like transaction volume and market sentiment must be collected from reliable sources. The information gathered is then cleaned, preprocessed and refined in order to make sure it is compatible with the training model. Then, we developed two forecasting models, one using (LSTM) and (KNN). LSTM models are trained using historical data, and the performance of these models is evaluated using a variety of metrics, such as mean squared error and mean error. The performance evaluation and comparison of the above two models are conducted saying LSTM model outperforms the KNN model in accuracy.***

I. INTRODUCTION

The stock market is impacted by a wide range of factors, such as political developments, economic information, and public opinion. In order to accurately estimate stock values, one must engage in a complex process that involves years of research. Construction of models that can accurately estimate stock values is now feasible because of recent developments in machine learning and deep learning approaches. The goal of this study is to construct and evaluate the predictive capability of the two models namely LSTM and KNN. The LSTM is variant of Recurrent Neural Network (RNN) architecture that is used for prediction task, whereas the KNN is straightforward yet effective machine learning algorithm used for regression problems.

The historical data is collected such as prices, trade volume and market mood. The initial step involves processing of data to check its reliability for the training model which is also called as preprocessing and followed by data analysis, scaling, train/test data split, evaluation metrices and prediction.

II. BASIC CONCEPTS

A. Pandas

Pandas is a Python package furnishing fast and suggestive data structures designed to make working with “labelled” or “relational” data. It aims to be the abecedarian high- position structure block for doing practical, real- world data analysis in Python. also, it has the thing of getting the most flexible and important open- source data analysis tool available in any language. It's formerly well on its way toward this thing.

B. NumPy

NumPy is a general- purpose array- processing package. It provides a high- performance multidimensional array object and tools for working with these arrays. It's the abecedarian package for scientific computing with Python. It's open- source software.

Besides its egregious scientific uses, NumPy in Python can also be used as an effective multi-dimensional vessel of general data.

C. Matplotlib

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library erected on NumPy arrays and designed to work with the broader SciPy mound. It was introduced by John Hunter in the time 2002. One of the topmost benefits of visualization is that it allows us visual access to huge quantities of data in fluently digestible illustrations. It consists of several plots like bar, line, histogram etc. Installation Windows, Linux and macOS distributions have matplotlib and utmost of its dependences as wheel packages.

D. Scikit

There are several Python libraries which give solid executions of a range of machine literacy algorithms. One of the best known is Scikit- learn, a package that provides effective performances of a large number of common algorithms. Scikit- Learn is characterized by a clean, invariant, and streamlined API, as well as by veritably useful and complete online attestation.

©IJRASET: All Rights are Reserved

E. TensorFlow

TensorFlow is an end- to- end open- source platform for machine literacy. It has a comprehensive, flexible ecosystem of tools, libraries, and community coffers that lets experimenters push the state- of- the- art in ML, and gives inventors the capability to fluently make and emplace ML- powered operations. It provides a group of workflows with intuitive, high- position APIs for both newcomers and experts to produce machine literacy models in multitudinous languages.

III. REQUIREMENTS SPECIFICATIONS

A. Analysis

This project helps in forecasting the stock values using LSTM and KNN algorithms. The process consists of Data collection, planning, design, and performance assessment. Data quality, model validity, and forecast accuracy affect project success. In order to reach project goals, we must thoroughly analyse data, models, and projections.

B. Design Constraints

They are the parameters that are to be considered while developing a forecasting model in order to maintain high efficiency.
Some of the parameters are:

1) Data Availability
2) Model complexity
3) Regulatory restrictions
4) Scalability

C. System Architecture

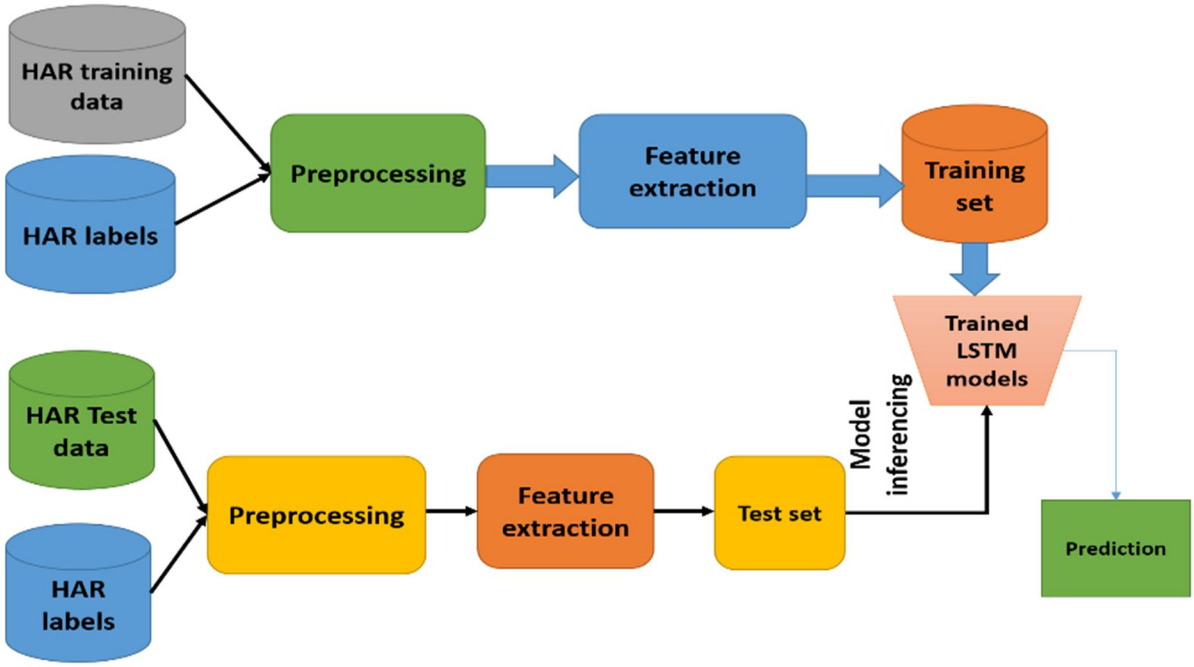


Fig.1: System Architecture

IV. IMPLEMENTATION

The first step consists of data collection, where set of data is collected from various sources and platforms. This model used python and Jupyter Notebook. This project relied on NumPy, Pandas, Scikit-Learn, Keras, and TensorFlow. Yahoo Finance used Pandas Data Reader to collect historical market prices for chosen companies. Pandas is used to extract the data to the notebook. In preprocessing, the data is collected, cleaned before it is used for implementation.
We use 40% of data to train and remaining 60% to test the algorithm. I have used TATAGLOBAL stock in this model. LSTM and KNN two model were implemented showing that LSTM is more accurate than KNN algorithm.

©IJRASET: All Rights are Reserved

V. METHODOLOGY

To build KNN and LSTM models, it generally involves multiple steps to guarantee the accuracy and reliability of the model. An example that can be taken for this model is:

1) *Data Collection:* Collecting historical data from different platforms, such as Yahoo Finance, Alpha Vantage.
2) *Data Preparation:* Formatting, cleaning, and converting raw data into a format that can be used for training and assessment models It includes techniques such as data scaling and data normalization.
3) *Model Training:* Train LSTM and KNN on a 40% dataset and assess its production using metrics such as MAE, RMSE and MSE
4) *Model Tuning:* Adjust the parameters of the LSTM model based on the reports of the model rating to improve the efficiency of the prediction.
5) *Model Testing:* Test the reliability and efficiency of the LSTM model using a set of data not used for training.
6) *Visualization:* Visualize event patterns and use charts, line charts, and histograms to analyse stock prices.
7) *Deployment:* Deploy the LSTM and KNN model to forecast future stock price.

VI. RESULTS AND ANALYSIS



Fig.2: Stock close price chart



Fig.3: Linear Regression

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



Fig.4: KNN Implementation



Fig.5: LSTM Implementation

VII. CONCLUSION

In brief, the objective of this model is to estimate its efficiency of two models i.e., LSTM and KNN, by the above research we can confidently state that LSTM model demonstrates superior performance compared to the KNN model in terms of accuracy and predictive capability. A relative examination of the LSTM and KNN models offers valuable insights into the respective pros and cons of both models when deployed for prediction of stock prices.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

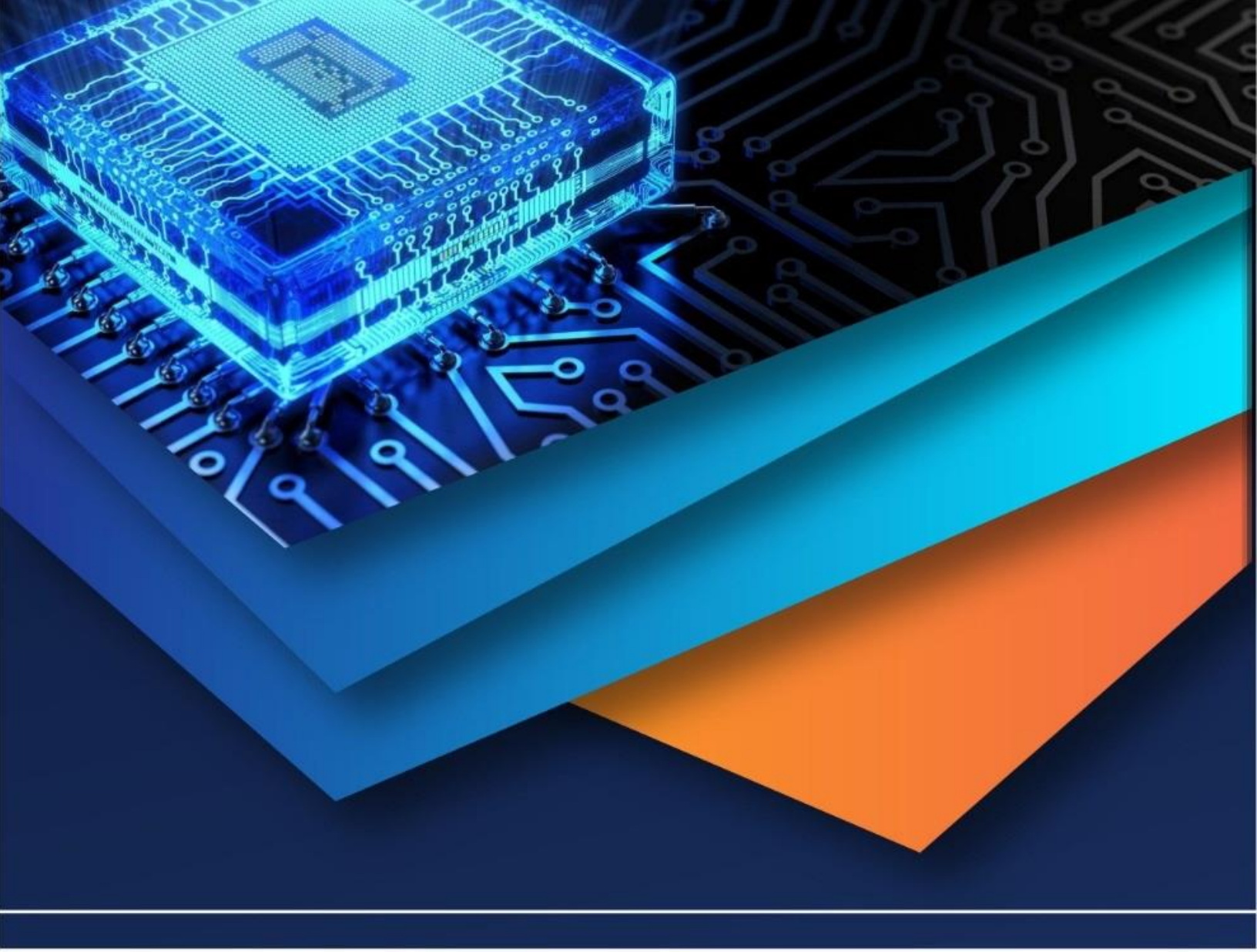
VIII. FUTURE SCOPE

There are many future directions for this project. A promising development to explore more machine learning and deep learning algorithms use for stock price prediction, such as Conventional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Furthermore, this model can be scaled continuously, the automobiles industry is one such example for using excellent innovation like machine learning and deep learning. It can be used by stock brokers also for selling and buying stocks, even helpful for investment bankers who guide their clients when to buy and sell the stocks to make huge profit.

REFERNECES

[1] https://www.sciencedirect.com/science/article/pii/S2772662222000261
[2] YouTube
[3] Simplilearn and Great Learning

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)