**COMPARATIVE ANALYSIS OF TIME SERIES MODELS FOR PRICE PREDICTION**

*Submitted in partial fulfilment of the requirements for the award of the*

*degree of*

***MASTERS OF COMPUTER APPLICATIONS***

*2022-2024*

**SUBMITTED BY**

**AJAY K 38222008**

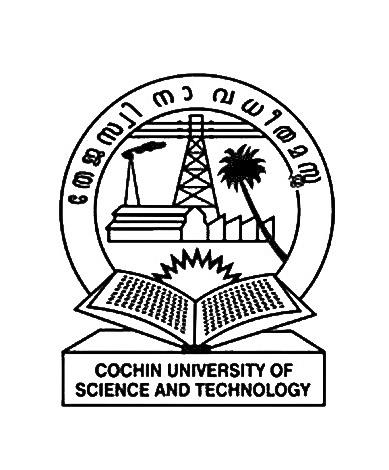
**ARAVIND E S 38222018**

**ARULDAS JAYAKRISHNAN 38222019**

**ASHKABSHA C 38222022**

**DHYAN MANOHAR 38222032**

**DEPARTMENT OF COMPUTER APPLICATIONS**



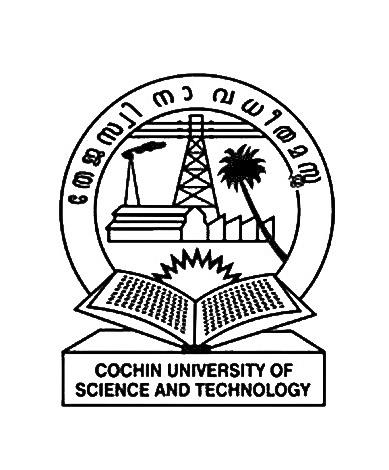
**COCHIN UNIVERSITY OF SCIENCE AND**

**TECHNOLOGY,COCHIN-22**

APRIL 2024

**DEPARTMENT OF COMPUTER APPLICATIONS**

**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**COCHIN-22**

**BONAFIDE CERTIFICATE**

This is to certify that the project report **“COMPARATIVE ANALYSIS OF TIME SERIES MODELS FOR PRICE PREDICTION”** submitted in partial fulfilment of the requirements for the award of the Degree of **Master of Computer Applications** is a record of bonafide work done by **AJAY K (38222008) , ARAVIND E S (38222018) , ARULDAS JAYAKRISHNAN (38222019) , ASHKABSHA C (38222022) and DHYAN MANOHAR (38222032)** during the period from June 2023 to November 2023 of their study in the Department Of Computer Applications at Cochin University Of Science And Technology, under my supervision and guidance.

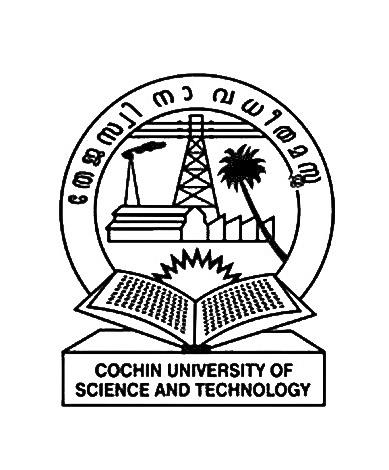
**Head of the Department Signature of the Examiner**

Department Of Computer Applications Department Of Computer Applications

CUSAT CUSAT

**DEPARTMENT OF COMPUTER APPLICATIONS**

**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**

 **COCHIN-22**

**CERTIFICATE**

This is to certify that the report entitled **“COMPARATIVE ANALYSIS OF TIME SERIES MODELS FOR PRICE PREDICTION”** submitted in partial fulfilment of the requirements for the award of the Degree of **Master of Computer Applications** is a record of bonafide work done by **AJAY K (38222008) , ARAVIND E S (38222018) , ARULDAS JAYAKRISHNAN (38222019) , ASHKABSHA C (38222022) and DHYAN MANOHAR (38222032)** during the period from June 2023 to November 2023 of their study in the Department Of Computer Applications at Cochin University Of Science And Technology, under my supervision and guidance.

**Signature of the Guide**

Department Of Computer Application

CUSAT

**DECLARATION**

We hereby declare that the report entitled **“COMPARATIVE ANALYSIS OF TIME SERIES MODELS FOR PRICE PREDICTION”** submitted in partial fulfilment of the requirements for the award of the Degree of **Master of Computer Applications** is a record of bonafide work done by usduring the period from June 2023 to November 2023 under the supervision and guidance of Dr M K SABU , Department of Computer Applications, Cochin University of Science and Technology , Cochin -22

PLACE : CUSAT **AJAY K 38222008**

**ARAVIND E S 38222018**

**ARULDAS JAYAKRISHNAN 38222019**

**ASHKABSHA C 38222022**

**DHYAN MANOHAR 38222032**

**ACKNOWLEDGEMENT**

We thank God almighty for bestowing his blessings upon us to proceed and complete the work. We have great pleasure in acknowledging the help given by various individuals throughout the project work. This project itself is an acknowledgement to the inspiration, drive and technical assistance contributed by many individuals.

We express our sincere and heartfelt gratitude to Dr. M V Judy, Head of the Department and faculty members for being helpful and cooperative during the period of the project.

We also express our deep gratitude to our guide Dr. M K Sabu, Assistant Professor in the Department of Computer Applications for her valuable guidance and timely suggestions that helped in the completion of this project in time.

We extend our sincere thanks to all the teaching and non-teaching staff of the Department of Computer Applications for providing the necessary facilities and help. Without the support of any of them this project would not have been a reality.

PLACE : CUSAT **AJAY K 38222008**

**Date :27/11/2023 ARAVIND E S 38222018**

**ARULDAS JAYAKRISHNAN 38222019**

**ASHKABSHA C 38222022**

**DHYAN MANOHAR 38222032**

**TABLE OF CONENTS**

Abstract…………………………………………………………………………………08……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….……………………………………………………………………………………………………………………………………………………………………………………………….………………………………………………………………

**LIST OF FIGURES**

**ABSTRACT**

Stock trading and investing are crucial for individuals and institutions as they provide opportunities to grow wealth, achieve financial goals, and contribute to economic development. Stock price forecasting helps investors make informed decisions, manage risk and optimize their investment portfolios based on predicted future price movements. In this project we are trying to implement a comparative study of stock price forecasting using Time Series Analysis.

Time series analysis is a statistical method used to analyze and interpret patterns, trends, and relationships within a sequence of data points collected over time. It can be used to identify complex patterns and make unbiased decision free from emotional and external influences. There are different machine learning methods to implement time series analysis. Which includes LSTM, ARIMA, RNN, CNN and CNN-LSTM. We trained models by each of these methods using historical stock prices available in yfinance and predicted the price for a given time period and compared it with the actual stock price to find accuracy of each model. Then we can use the best of these models to predict the future stock price movement.

**1.INTRODUCTION**

Stock trading and investment activities play a pivotal role in shaping individual financial landscapes and contributing significantly to the economic fabric of both individuals and institutions. The ability to forecast stock prices serves as a critical tool for investors, enabling informed decision-making, risk management, and the optimization of investment portfolios. In this project, we embark on a comparative study of stock price forecasting employing Time Series Analysis, aiming to unveil the most effective predictive models.

Time series analysis, a statistical methodology, stands as a powerful means to unravel patterns, correlations, and trends within sequential data points collected over time. Its application in the realm of stock price prediction offers a shield against emotional biases and external influences, fostering the creation of informed strategies. Various machine learning methodologies such as LSTM, ARIMA, RNN, CNN, and CNN-LSTM represent the toolkit for implementing time series analysis.

Through this research endeavor, we have trained models utilizing historical stock price data extracted from yfinance, employing each of these distinct methods to predict stock prices within a specified time frame. The crux of our study lies in comparing these predictive models against actual stock prices, seeking to determine the accuracy and efficacy of each. Subsequently, our aim is to identify and employ the most robust and reliable model for predicting future stock price movements, ultimately contributing to more informed and data-driven investment strategies.

**2.THEORETICAL BACKGROUNDS**

## Machine learning

Using statistical and computational techniques, machine learning is a subset of artificial intelligence that enables computers to learn from data without being explicitly programmed. The aim of machine learning is to create algorithms that can learn from data patterns and make predictions or choices.

Algorithms for machine learning can be classified into three categories: supervised learning, unsupervised learning, and reinforcement learning.

In supervised learning, a machine learning model is trained using labelled data. By identifying patterns in the labelled data, the model has the ability to predict outcomes. Image identification, language translation, and sentiment analysis are a few examples of supervised learning.

Using unlabeled data to train a machine learning model is known as unsupervised learning. Without knowing what those patterns could be beforehand, the model learns to spot them in the data. Anomaly detection and clustering are two examples of unsupervised learning.

Through the process of reinforcement learning, a computer learning model is taught to make decisions in a setting where it receives feedback in the form of rewards or penalties based on its actions. Over time, the model gains the ability to behave in a way that maximises its reward. Playing video games and using robots are two examples of reinforcement learning.

There are several uses for machine learning, such as in natural language processing, audio and picture identification, fraud detection, recommendation systems, and predictive analytics. There are many of fascinating new breakthroughs and innovative prospects in this quickly expanding area.

## Preprocessing

Preprocessing, which involves getting the data ready for modelling and analysis, is a crucial stage in machine learning. Preprocessing is the act of converting raw data into a format that machine learning algorithms can quickly analyse and comprehend.

Data cleaning is the process of eliminating useless, erroneous, or inconsistent data from a dataset. The performance of the machine learning model is improved by removing any noise or mistakes from the data in this stage.

Data transformation is the process of transforming data into a format that is acceptable for analysis. Among these include controlling outliers, scaling the numerical features, and translating categorical data into numerical values.

Feature extraction is the process of choosing and removing the dataset's most pertinent characteristics. By reducing the number of dimensions in the data, this step can make it simpler to read and analyse.

Feature encoding is the process of assigning numerical values to category characteristics. The majority of machine learning algorithms want numerical input, thus this is significant.

Feature scaling is the process of adjusting the numerical features to a comparable range in order to remove bias from the model. Standardisation and normalisation are two common scaling methods.

Splitting the dataset into training and testing sets is known as data splitting. The testing set is used to assess the performance of the machine learning model, whereas the training set is used to train the model.

## LSTM

For processing sequential data, such as audio, video, and text, recurrent neural networks (RNNs) of the LSTM (Long Short-Term Memory) type are employed in machine learning. To address the shortcomings of conventional RNNs in modelling long-term dependencies, LSTM networks were created.

Memory cells, gates, and activation mechanisms all function in concert within the LSTM architecture to enable long-term learning and memory. Information is stored in the memory cells and sent from one time step to the next using them. Information entering and leaving memory cells is managed by the gates. The output of the gates and memory cells are controlled by the activation processes.

In an LSTM, there are three primary types of gates:

* + - The forget gate regulates which data is removed from memory cells. Which information is no longer pertinent to the current time step is determined.
    - The input gate determines what data is added to the memory cells. It chooses the data that is crucial for the current time step.
    - The output from the memory cells is managed by the output gate. It chooses which data should be output to the following time step.

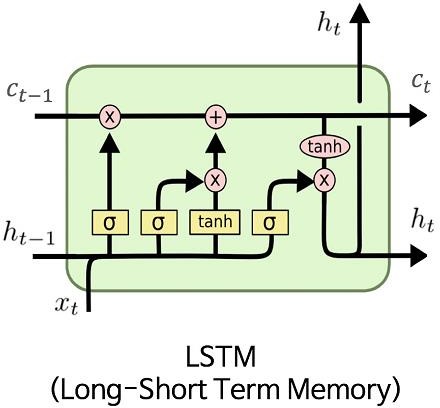
Natural language processing (NLP) activities including language translation, text categorization, and sentiment analysis frequently make use of LSTM networks. They are also utilised in applications that handle sequential data, such as voice recognition, video analysis, and others.

Time series forecasting may be done using LSTM. By using historical data from observations of a variable across time, time series forecasting makes predictions about future values.

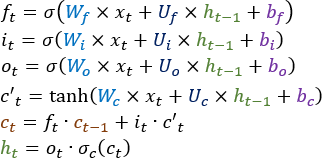
Since LSTM networks can recognise long-term relationships and patterns in the data, which are frequently necessary for producing reliable predictions, they are particularly helpful in time series forecasting jobs. LSTMs may also deal with data that has changing times between observations, missing values, and other abnormalities that are frequently present in real-world time series data.

The input to the network in time series forecasting using LSTM is a list of prior observations, and the network's output is the projected value(s) for one or more subsequent time steps. A historical dataset of the time series is used to train the LSTM, which then learns to spot trends and patterns in the data. Once trained, the model may be used to generate predictions for brand-new, unforeseen data.

* + - Recurrence is implemented similarly in LSTMs, however there are four layers working together in a very precise way rather of just one tanh layer.
    - The transformations that are used in the hidden state at time step t are shown in the following figure.
    - The cell state c, which represents the unit's internal memory, is represented by the line running across the top of the diagram.
    - The LSTM uses the i, f, o, and g gates to get around the vanishing gradient problem. The line across the bottom represents the hidden state h. The LSTM gains knowledge of these gates' parameters during training:



* + - Examining the cell's equations is another method to understand about how these gates operate inside an LSTM cell.
    - These equations explain how the value of the hidden state at time t is determined from the value of the hidden state at time t-1.



* + - These are the input, forget, and output gates: i, f, and o. The same equations are used to calculate them, but different parameter matrices—Wi, Ui, Wf, Uf, and Wo, Uo—are used.
    - The output of these gates is modulated by the sigmoid function between 0 and 1, which allows the output vectors to be multiplied element-by-element with another vector to determine how much of the second vector can flow through the first.

**2.22 TIME SERIES ANALYSIS**

Time Series Analysis stands as a fundamental statistical technique employed to unravel patterns and glean insights from sequential data points amassed over time. In the context of predicting stock prices, this method serves as a cornerstone in financial analysis, offering a profound understanding of historical market behaviors. By meticulously scrutinizing past stock price data, Time Series Analysis enables the identification of intricate trends, cyclic behaviors, and correlations that underlie market dynamics. This approach empowers analysts and investors to decipher the nuances of market fluctuations, allowing for informed predictions about potential future stock price movements. The application of various machine learning algorithms within Time Series Analysis, encompassing methodologies like LSTM, ARIMA, RNN, CNN, and CNN-LSTM, equips analysts with robust predictive models. These models utilize historical stock data to forecast potential market trajectories, offering invaluable insights to guide investment decisions, optimize portfolios, and mitigate risks in the ever-evolving financial landscape.

The predictive capabilities embedded within Time Series Analysis hold immense significance for investors seeking to navigate the complexities of financial markets. Its utilization transcends mere analysis of historical stock prices, extending to the strategic forecasting of future market behaviors. By leveraging Time Series Analysis, investors gain a vantage point into the probable trends and movements within stock prices, enabling proactive decision-making and risk management strategies. The amalgamation of statistical methodologies and machine learning algorithms allows for the construction of predictive models that encapsulate the intricate dynamics of stock markets. Ultimately, the prowess of Time Series Analysis in predicting stock prices not only aids in deciphering historical trends but also provides a compass for investors, empowering them to navigate the uncertainty of financial markets with a greater sense of foresight and informed decision-making.

**Top of Form**