FOOD PRICE PREDICTION USING DIFFERENT MODELS ON A COMMON DATASET AND ITS ANALYSIS

RATAN KUMAR PRADHAN

B.Tech, Third Year, CSE

TANMAYA SAMANTRAY

B.Tech, Third Year, CSE

ANSHUMAN MOHAPATRA

B.Tech, Third Year, CSE

Project Report

on

FOOD PRICE PREDICTION USING DIFFERENT MODELS ON A COMMON DATASET AND ITS ANALYSIS

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

Bachelor of Technology
in
Computer Science and Engineering

Submitted by
Ratan Kumar Pradhan
2001106208
Tanmaya Samantray
2001106227
Anshuman Mohapatra
2001106550

Under the guidance of

Dr. Subasish Mohapatra

Department of Computer Science and Engineering
Odisha University Of Technology and Research
Bhubaneswar, Odisha - 751003

Department of Computer Science and Engineering

Odisha University Of Technology and Research, Bhubaneswar

Supervisor's Certificate

This is to certify that the project entitled **FOOD PRICE PREDICTION USING DIFFERENT MODELS ON A COMMON DATASET AND ITS ANALYSIS** submitted by **Ratan Kumar Pradhan**,**Tanmaya Samantray** and **Anshuman Mohapatra** bearing registration number **2001106208, 2001106227**, **2001106550** respectively to the Department of Computer Science and Engineering , Odisha University of Technology and Research, formerly College of Engineering and Technology, Bhubaneswar, is a record of bonafide project work under my supervision and I consider it worthy of consideration for partial fulfilment of the requirements for the award degree of Bachelor of Technology in Computer Science and Engineering under Odisha University of Technology and Research, Bhubaneswar.

Dr. Subasish Mohapatra

(Project Guide)

Department of Computer Science and Engineering

Odisha University Of Technology and Research, Bhubaneswar

Certificate of Examination

This is to certify that the project entitled **FOOD PRICE PREDICTION USING DIFFERENT MODELS ON A COMMON DATASET AND ITS ANALYSIS** submitted by **Ratan Kumar Pradhan**,**Tanmaya Samantray** and **Anshuman Mohapatra** bearing registration number **2001106208, 2001106227**, **2001106550** respectively, to the Department of Computer Science and Engineering , Odisha University of Technology and Research, formerly College of Engineering and Technology, Bhubaneswar, is a record of bonafide project work under the guidance of **Dr. Subasish Mohapatra**. We the below signed, after checking the project mentioned above and the official record book(s) of the students, hereby state our approval of the report submitted in partial fulfilment of the requirements of the degree of Bachelor of Technology in Computer Science and Engineering at Odisha University of Technology and Research Bhubaneswar.

Dr. Subasish Mohapatra	**Dr. Subasish Mohapatra**
(Project Guide)	(Head of the Department)

Acknowledgement

It is our privilege and solemn duty to express our deepest sense of gratitude to **Dr. Subasish Mohapatra**, under whose guidance we carried out this work. We are indebted, for his valuable supervision, heart full cooperation and timely aid and advice till the completion of the project in spite of his pressing engagements. We would like to express our deep sense of gratitude to **Mr. Ashis Kumar Mishra**, Coordinator of the Department for encouragement and inspiration throughout my project work.We wish to record our sincere gratitude to our respected Head of the Department, **Dr. Subasish Mohapatra** for his constant support and encouragement in preparation of this project.

We take this opportunity to express our hearty thanks to all those who helped us in the completion of our project work. We are very grateful to the authors of various articles on the Internet, for helping us become aware of the research currently ongoing in this field.

We are very thankful to our parents for their constant support and love. Last, but not the least, we would like to thank our classmates for their valuable comments, suggestions and unconditional support.

Ratan Kumar Pradhan

Tanmaya Samantray

Anshuman Mohapatra

Declaration

We certify that

I. We certify that this work has been done by ourselves under the general supervision of our supervisor.

II. The work has not been submitted to any other Institute for any degree or diploma.

III. We have followed the guidelines provided by the Institute in writing the report.

IV. Whenever we have used materials (data, theoretical analysis, figures, text) from the other sources, we have given due credit to them by citing them in the text of the report and giving their details in the references.

V. Whenever we have quoted written materials from other sources, we have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Ratan Kumar Pradhan

Anshuman Mohapatra

Tanmaya Samantray

ABSTRACT:

Food and goods connected to food have fluctuating prices. A change in pricing has an impact on both the government's monetary policy and the consumer's purchasing habits. The Consumer Food Price Index (CFPI) captures changes in food costs over a certain time period. In India, the Central Statistical Organisation publishes the CFPI on a monthly basis. Additionally, it displays inflation and aids in prompt remedial action by the government. In this study, we anticipate the consumer food price index in India using a machine learning technique. The use of Artificial Neural Network (ANN) models with back propagation learning to forecasting future values of CFPI has been the special emphasis of this work.

In essence, the Consumer Price Index (CPI) is a gauge of the economic phenomenon of inflation that is rife in the retail sector. Technically, it is calculated by compiling the price variations seen in the items and services that make up the majority of the nation's consumption. LSTM, ARIMA, and GARCH are just a few of the machine learning models used in this study to anticipate food prices. The model with the smallest difference between actual and projected values is found after further accuracy analysis using several metrics, including Mean Absolute Error, Mean Absolute Percentage Error, etc.

KEYWORDS:

food price,price prediction, LSTM, ARIMA, GARCH,regression

Contents

INTRODUCTION:

The main aim of firms is profit maximization.As per Pradeepta et al. [1], To achieve this goal, the constant updating and forecasting of selling prices is of fundamental importance for every company. Although digital transformation is a phenomenon that involves all companies, from small to large, many of them still update prices by hand through logics that are not always clear nor objective and transparent, but rather based on the experience and expertise of those in charge of updating the price list.As stated by Doganis et al. [2], the automation of price prediction and update can provide a strong productivity boost by freeing up human resources, which can thus be allocated to more creative and less repetitive tasks. This also increases the morale and commitment of employees; it also speeds up the achievement of goals, and improves accuracy by minimizing human errors. Subjectivity is also reduced: once the operating criteria have been established, forecast algorithms will keep behaving consistently. This in turn means an improvement in compliance.

Staple food material prices can be a trending topic in the market. The fluctuation of the price is influenced by many factors. For instance, the weather, oil price, and etc are the external factors of the staple food price as stated by Rusiri Illesinghe in [3]. Indeed, the prediction of staple food fluctuation price is important for the farmers, consumers, even government. [4]

Time-series forecasting has always been a major topic in data science with plenty of applications.The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. . According to Jenkins in [5],Traditional econometric methods were proven inadequate when dealing with significant amounts of volatile time-series data. As a result, interest in machine learning methods has grown as they offer data-driven solutions that don't require prior assumptions.

There are lots and lots of models available in Machine Learning and Data Analysis. All these models have their own pro's and con's and not every model is suitable for everything. In this project we mainly used three Models: ARIMA, GARCH, LSTM.

ARIMA:

An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.A statistical model is autoregressive if it predicts future values based on past values. ARIMA makes use of lagged moving averages to smooth time series data. It implicitly assume that the future will resemble the past.They are widely used in technical analysis to forecast future security prices. According to investopedia [6] ,this model is preferred for data sets that are mostly stable and have low volatility.

GARCH:

According to investopedia [7] ,Generalised AutoRegressive Conditional Heteroskedasticity or GARCH is a statistical model used in time-series analysis where the variance error is believed to be serially auto-correlated.GARCH is appropriate for time series data where the variance of the error term is serially auto-correlated following an autoregressive moving average process.GARCH is useful to assess risk and expected returns for assets that exhibit clustered periods of volatility in returns.

GARCH models are used when the variance of the error term is varying. That is, the error term is heteroskedastic. Heteroskedasticity refers to the irregular pattern of variation of an error term, or variable, in a statistical model.

LSTM:

The excellent feature of ANNs, when applied to time series forecasting problems is their inherent capability of non-linear modelling, without any presumption about the statistical distribution followed by the observations. The optimal model emerges dynamically from the data itself.ANNs are data-driven and self-adaptive by nature .

As per Pradeepta et al [8], Long Short Term Memory is a type of recurrent neural network. LSTM networks are well-suited for classifying, processing and making models based on time series data, since there could be lags of unknown duration in a time series. The recurrent neural network uses long short-term memory blocks to provide context for the way the program receives inputs and creates outputs. The long short-term memory cell or block is a complex unit with various components such as weighted inputs, activation functions, inputs from previous blocks and eventual outputs.

LITERATURE SURVEY:

1."Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices" by Lorenzo Menculini , Andrea Marini , Massimiliano Proietti, Alberto Garinei , Alessio Bozza , Cecilia Moretti and Marcello Marconi at mdpi, 2021.

According to this paper,setting sale prices correctly is of great importance for firms, and the study and forecast of prices time series is therefore a relevant topic not only from a data science perspective but also from an economic and applicative one.
The results of this research indicates that the combination of CNNs and LSTMs yields the most accurate results for all the three products, but require the longest and computationally more expensive tuning. On the contrary, Prophet performances were not brilliant, but model tuning and data preparation were particularly quick. ARIMA and LSTM-only neural networks showed good performance both in terms of accuracy and time required for model selection and training.

2. *"An Introductory Study on Time Series Modeling and Forecasting" by Ratnadip Adhikari and R. K. Agrawal ,2013.*
A time series containing records of a single variable is termed as univariate. But if records of more than one variable are considered, it is termed as multivariate. A time series can be continuous or discrete.
According to the author ,the proper selection of the model orders (in case of ARIMA), the number of input, hidden and output neurons (in case of ANN) and the constant hyper-parameters (incase of SVM) is extremely crucial for successful forecasting.

3. *"Forecasting: theory and practice" by Fotios Petropoulos,Daniele Apiletti, Vassilios Assimakopoulos, Clara Cordeiro and Florian Ziel at International Journal of Fourcasting Volume 38 ,Issue 3,2022.*
The theory of forecasting is based on the premise that current and past knowledge can be used to make predictions about the future. In particular for time series, there is the belief that it is possible to identify patterns in the historical values and successfully implement them in the process of predicting future values. However, the exact prediction of futures values is not expected.
Accepting the advantages and limitations of systematic forecasting methods and most importantly avoiding any exaggerated expectations of what it can achieve is critical. Such methods do not possess any prophetic powers, they simply extrapolate established patterns and relationships to predict the future and assess its uncertainty. Their biggest advantage is their objectivity and ability for optimal extrapolation. Their biggest disadvantages are: (*i*) the patterns and the relationships must remain fairly constant during the forecasting phase for the forecasts to be accurate, and (*ii*) uncertainty must not be fat-tailed so it can be measured quantitatively.

4.”*Integrated Feature Selection of ARIMA with Computational Intelligence Approaches for Food Crop Price Prediction” by Yuehjen E. Shao and Jun-Ting Dai at Hindawi,2018.*

According to the paper ,the rapid increase in demand for food has contributed to a continuous rise in food prices, which directly threatens the lives of over 800 million people around the world who are reported to be chronically undernourished. Consequently, food crop price prediction has attracted considerable attention in recent years.

The experimental results reveal that the proposed integrated models of CI and ARIMA-SVR provide predictions of food crop prices without requiring extensive effort to obtain the future values of explanatory variables.

5. “An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting” by H.F.Zou,G.P.Xia,and H.Y.Wangat Neurocomputing Volume 70, Issue 16-18,2007.

It compares the predictive performance of ARIMA, artificial neural network and the linear combination models for forecasting wheat price in Chinese market. It is found that ANN model is best as well as at capturing a significant number of turning points.

They investigated the forecasting measurement of turning points of three forecasting methods mentioned above. Price forecasting is an integral part of commodity trading and price analysis. When evaluating forecasting models, turning point forecasting power is as important as quantitative accuracy with small errors.

6.”A survey of machine learning techniques for food sales prediction” by Grigorios Tsoumakas at SpringerLink,2018.

Food sales prediction is concerned with estimating future sales of companies in the food industry, such as supermarkets, groceries, restaurants, bakeries and patisseries. Accurate short-term sales prediction allows companies to minimize stocked and expired products inside stores and at the same time avoid missing sales.

Sales prediction is typically done arbitrarily by managers. However, skilled managers are hard to find and they are not always available. Therefore, sales prediction should be supported by computer systems that can play the role of a skilled manager when she is not there and/or help her take the right decision by providing estimates of future sales.

7. “A neural network model based on the multi-stage optimization approach for short-term food price forecasting in China” by Zou Haofei, Xia Guoping,Yang Han,Yang Fangting at Experts Systems with Application Volume 33,Issue 2,2007.

Studies have demonstrated that back-propagation neural network can be effectively used to uncover the nonlinearity in the financial markets. But , back-propagation algorithm suffers the problems of slow convergence, inefficiency, and lack of robustness.

This project’s main focus was to out-sample forecasting for the wheat prices of Chinese food market.The forecasting accuracy achieved beyond the training data is the ultimate and the most important measure of performance.
It also aimed at investigating the forecasting measurement of turning points of three forecasting methods mentioned above. Price forecasting is an integral part of commodity trading and price analysis. When evaluating forecasting models, turning point forecasting power is as important as quantitative accuracy with small errors.

8. “The use of ARIMA models for reliability forecasting and analysis” by S.L.Ho,M.Xie at Computer & Industrial Engineering, Volume 35, Issue 1-2,1998.
This paper investigated the approach to repairable system reliability forecasting based on the Autoregressive Integrated Moving Average (ARIMA) models. This time series technique made very few assumptions and is very flexible. It was theoretically and statistically sound in its foundation and no a priori postulation of models was required when analysing failure data. An illustrative example on mechanical system failures was presente. Comparison was also made with the traditional Duane model.

9. “Regression Based Price Prediction of Staple Food Materials Using Multivariate Models” by K. Venkateswara Rao,D. Srilatha,D. Jagan Mohan Reddy,Venkata Subbaiah Desanamukula,and Mandefro Legesse Kejela at Hindawi, 2022.
Profit margins for essential foodstuffs could be a demand rising problem. There are several variables influencing currency fluctuations. For example, the various variables of commodity food prices are climate, crude prices, and so on. Forecasting the fluctuating prices of basic foodstuffs is also relevant even for the government, producers, and customers.The positive peak of basic commodity prices in Asia, especially India, has caused everybody concerned and endanger the system’s stability. Consistent food prices offer diverse benefits for emerging regions, such as enhancing economic development and avoiding hunger traps for small producers and workers.This research used ARCH and GARCH models to estimate food prices.

10. “Predicting Staple Food Materials Price Using Multivariables Factors (Regression and Fourier Models with ARIMA)” by Said Fadlan Ashari, P.H.Gunawan, Yanti Rusmawati at IEEE,2022.
ARIMA or known as Box-Jenkins method is a prediction method to forecast time series data.there are few problems for researchers to use ARIMA, one of them is that many obstructions in ARIMA models are hard to be implemented. But recently, ARIMA is not a difficult method to use anymore. There has been much research and experiments done to automate ARIMA modelling in 25 previous years. Moreover, there are some studies to seek stationarity as well as the application of seasonal and non-seasonal data in many sectors.

Here the author used two type of regression models, multiple linear and Fourier regression models.The results of both regression models are shown satisfying with produce high accuracy.

11. "Machine Learning Approach for the Prediction of Consumer Food Price Index" by Prakash kumar Sarangi,Deepti Sinha,Sachin Sinha,Neetu Mittal at ICRITO, 2021.

This paper has applied the machine learning approach in forecasting the consumer food price index in India.It has focussed on applicability of ANN models with back propagation learning in predicting the future values of CFPI.

The experimental results show that a simple ANN model with back propagation algorithm is highly capable in forecasting the future values of CFPI.

12. "Multi-item time series prediction using autoregressive integrated moving average and long short term memory on perishable products" by Abdullah 'Azzam;Shelly Elvina Salsabila;Suci Miranda at AIP Conference ,2023.

According to the author, "Perishable products are products whose quality will decrease with the increasing age of the product which is not more than 14 days. This will affect consumer satisfaction in product selection. Forecasting errors on perishable products can cause losses due to this."

This study focused on prediction using ARIMA and LSTM. According to the author, ARIMA is the better model even though its RMSE was higher than LSTM. The LSTM method is overfitting which means that the results of the data training are good but cannot generalise to new data so it cannot be used to make predictions regularly.

OBSERVATION AND MOTIVATION:

1. Consumer Price Index (CPI) in essence is a measure of inflation rampant in the retail sector of the economy.

2. Forecasting the fluctuating prices of basic foodstuffs is also relevant even for the government, producers, and customers.Staple food material prices can be a trending topic in the market. The fluctuation of the price is influenced by many factors. For instance, the weather, oil price, and etc. are the external factors of the staple food price. Indeed, the prediction of staple food fluctuation price is important for the farmers, consumers, government and everyone.

3. In our survey we found that different authors used different models on different dataset to get their result.We have decided to compare some of those models over a common dataset and compare their performance. We have chosen 3 different models used for time series analysis. These three models, LSTM, GARCH, and ARIMA have never been used compared together in any other research we have studied till now.

MODELS USED:

LSTM:

Recurrent neural networks (RNNs) of the LSTM (Long Short-Term Memory) type are frequently employed in deep learning for time series analysis and natural language processing. In contrast to conventional RNNs, LSTM uses a memory cell that may selectively recall or forget data from earlier time steps to solve the vanishing gradient problem. The simple LSTM model and its uses will be covered in this article.

The input gate, forget gate, and output gate are the three gates that make up the LSTM model. The LSTM may selectively store or delete data from earlier time steps thanks to these gates, which regulate the flow of information into and out of the memory cell.

How much of the fresh input should be added to the memory cell is decided by the input gate. A sigmoid activation function that produces a value between 0 and 1, where 0 indicates that no new input is added and 1, which indicates that all new input is added, controls it.

How much of the prior memory should be erased is decided by the forget gate. A sigmoid activation function that generates a value between 0 and 1, where 0 indicates that all prior memory is deleted and 1, which indicates that all prior memory is maintained, controls it as well.

How much memory should be output to the following time step is decided by the output gate. A sigmoid activation function that produces a value between 0 and 1, where 0 indicates that no memory is outputted and 1, which indicates that all memory is outputted, controls it.

A memory cell in the LSTM model is also used to store data from earlier time steps. Every time step, the memory cell is updated based on the input, forget, and output gates. The current time step's predictions are then based on the updated memory cell.

The steps below are typical ones to utilise the LSTM model:

1. Prepare the data: The LSTM model should be used after preparing the data. The information must be presented as a time series, with each row representing a time step and each column representing a variable.

2. Split the data: Training and testing sets should be created from the data. While the testing set is used to gauge the model's effectiveness, the training set is used to fit the model.

3. Normalize the data: The data must be normalized so that the mean and variance are both zero. For the LSTM model to operate effectively, this is critical.

4. Specify the model architecture: The input dimensions, output dimensions, and activation functions of the LSTM model should all be specified.

5. <u>Train the model</u>: A suitable optimization approach, such as stochastic gradient descent (SGD), should be used to train the LSTM model on the training data.

6. <u>Assess the model</u>: Using the right metrics, such as mean squared error (MSE) or root mean square error (RMSE), the performance of the LSTM model should be assessed on the testing set.

The LSTM model has several uses in a variety of disciplines, including finance, economics, and natural language processing. Stock prices, exchange rates, and other financial variables may all be predicted using the LSTM model in the fields of finance and economics. The LSTM model may be used in natural language processing for sentiment analysis, text categorization, and language translation.

In conclusion, the LSTM model is an effective tool for natural language processing and time series analysis. The model can selectively store or delete data from earlier time steps thanks to three gates that regulate the flow of information into and out of a memory cell. The stages of preparing the data, dividing the data, normalizing the data, establishing the model architecture, training the model, and assessing the model are routinely followed in order to use the model. The LSTM model may be a helpful tool for forecasting and decision-making and has several applications in many different industries.

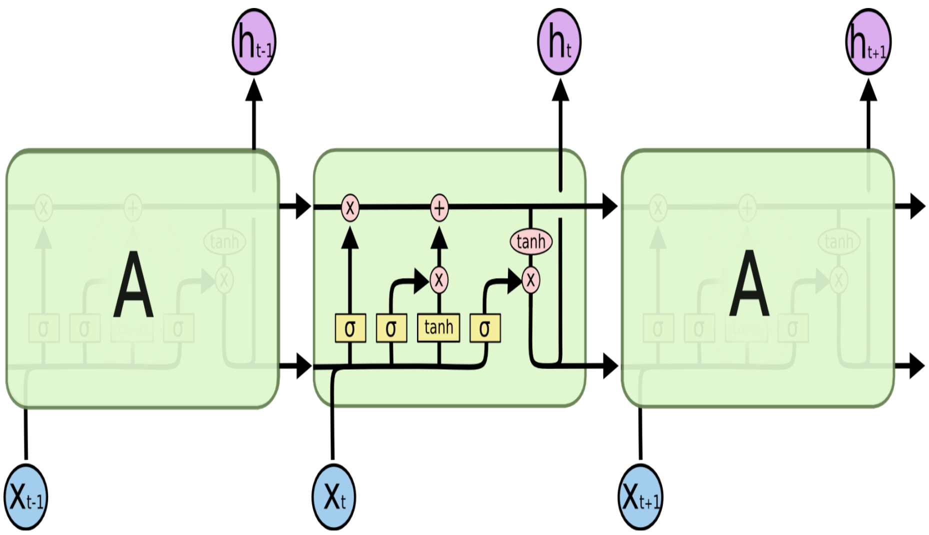

Figure 1. LSTM diagram

ARIMA:

The time series model known as ARIMA, or Autoregressive Integrated Moving Average, is often used to predict future values of a series based on its historical trends. The model is popular among academics and professionals in a variety of industries, including engineering, banking, and economics, due to its simplicity and strength.

The moving average (MA), the integrated (I), and the autoregressive (AR) components are the three parts of the ARIMA model. While the MA component simulates the dependency of the current value on previous mistakes, the AR component simulates the dependence of the current value on past values of the series. For the AR and MA components to function effectively, the series must be made immobile using the I component.

The following procedures are normally taken in order to employ the ARIMA model:

1. <u>Verify stationarity</u>: Before using the ARIMA model, it is important to determine if the time series is stationary. A stationary time series is one in which the mean, variance, and autocorrelation do not vary during the course of the data. For the AR and MA components to function effectively, stationarity is crucial. To check for stationarity, a number of statistical tests may be used, including the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.

2. <u>Differentiate the series</u>: If the series is not stationary, it can be differentiated by taking the first difference (that is, deducting the current value from the prior value). The sequence can be repeated a number of times until it stops moving.

3. <u>Establish the order of differencing</u>: The order of differencing is the quantity of differences that must be made in order for the series to become stationary. The autocorrelation and partial autocorrelation plots of the differenced series can be used to establish the order of differencing.

4. <u>Identify the AR and MA components' relative positions</u>: The autocorrelation and partial autocorrelation plots of the differenced series may be used to identify the relative positions of the AR and MA components. The partial autocorrelation plot displays the correlation between the series and its lagged values after the effects of intermediate lags have been removed from the autocorrelation plot, which displays the correlation between the series and its lagged values.

5. <u>Fit the model</u>: After establishing the order of the ARIMA model, the model may be adjusted to the data using maximum likelihood estimation or another suitable technique. The series' future values may then be predicted using the fitted model.

The ARIMA model is a straightforward yet effective time series model that may be used to predict future values of a series based on its historical behaviour. The moving average component, the integrated component, and the autoregressive component make up the model.

In order to apply the model, one normally performs the following steps: confirming stationarity; differencing the series; deciding the order of differencing; deciding the order of the AR and MA components; and finally, fitting the model to the data. The ARIMA model may be a helpful tool for forecasting and decision-making and has several applications in many different industries.

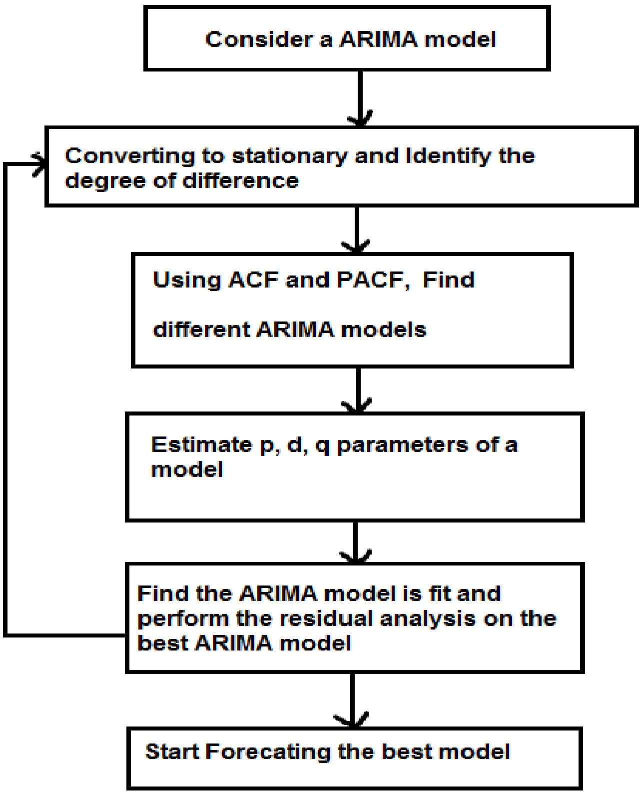

Figure 2. ARIMA flow-chart

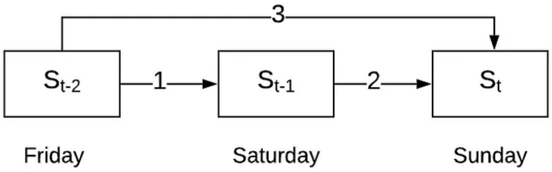

Figure 3. Auto-correlation

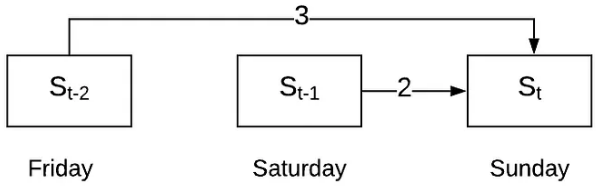

Figure 4. Partial-correleation

GARCH:

For modelling the volatility of financial time series, the statistical model known as GARCH (Generalised Autoregressive Conditional Heteroskedasticity) is frequently used in finance and economics. Bollerslev presented the model as an expansion of the ARCH (Autoregressive Conditional Heteroskedasticity) model in 1986.

The conditional variance of a time series may be modelled using the GARCH model, which is helpful for predicting risk in the financial markets. The model posits that a time series' variance is not constant but changes with time and is influenced by earlier values of the time series.

The conditional mean equation and conditional variance equation are the two equations that make up the GARCH model. Typically, an autoregressive (AR) or moving average (MA) process is used to simulate the conditional mean equation. In order to simulate the conditional variance equation, historical squared errors and conditional variance values are used.

The basic form of the GARCH model is as follows:

$$\mathbf{Y_t = \mu_t + \varepsilon_t} \qquad \text{(Equation 1)}$$

$$\mathbf{\varepsilon t = \sigma_t * e_t} \qquad \text{(Equation 2)}$$

$$\mathbf{\sigma^2 t = \hat{w} + \alpha_1 * \varepsilon^2_{\{t-1\}} + \beta_1 * \sigma^2_{\{t-1\}}} \qquad \text{(Equation 3)}$$

where $\mathbf{Y_t}$ is the observed time series, $\boldsymbol{\mu_t}$ is the conditional mean of $\mathbf{Y_t}$, $\boldsymbol{\varepsilon_t}$ is the error term, $\boldsymbol{\sigma_t}$ is the conditional standard deviation of the error term, $\mathbf{e_t}$ is a random variable with zero mean and unit variance, $\mathbf{\hat{w}}$ is the constant term, $\boldsymbol{\alpha_1}$ and $\boldsymbol{\beta_1}$ are the parameters that control the persistence of the volatility, and $\boldsymbol{\varepsilon^2_{\{t-1\}}}$ and $\boldsymbol{\sigma^2_{\{t-1\}}}$ are the squared error and conditional variance from the previous time step, respectively.

By utilizing either maximum likelihood estimation (MLE) or quasi-maximum likelihood estimation (QMLE), the GARCH model may be calculated. The QMLE technique allows for more diverse error term distributions whereas the MLE approach presupposes that the error term has a normal distribution.

In comparison to alternative models for modelling volatility, such as the historical volatility approach and the implied volatility method, the GARCH model provides a number of benefits. The GARCH model has the ability to represent the volatility's persistence and clustering, which are frequently seen in financial markets. The GARCH model also has the benefit of being able to predict future volatility, which is advantageous for risk management and portfolio optimization.

In economics and finance, the GARCH model has several uses. The GARCH model may be applied to finance to simulate the volatility of stock prices, currency rates, and other financial variables. The model may also be used to calculate a portfolio's value at risk (VaR), which is a gauge of the possible loss that could happen under unfavorable market circumstances. The GARCH model in economics may be used to simulate the volatility of macroeconomic variables like inflation and GDP.

The GARCH model is a potent tool for simulating the volatility of financial time series, to sum up. The approach enables the modelling of a time series' conditional variance, which is helpful for calculating risk in the financial markets. The conditional mean and variance equations, which make up the model, may be estimated using either maximum likelihood estimation or quasi-maximum likelihood estimation. In comparison to alternative models for modelling volatility, such as the historical volatility approach and the implied volatility method, the GARCH model provides a number of benefits. The model may be a valuable tool for forecasting and risk management and has many applications in finance and economics.

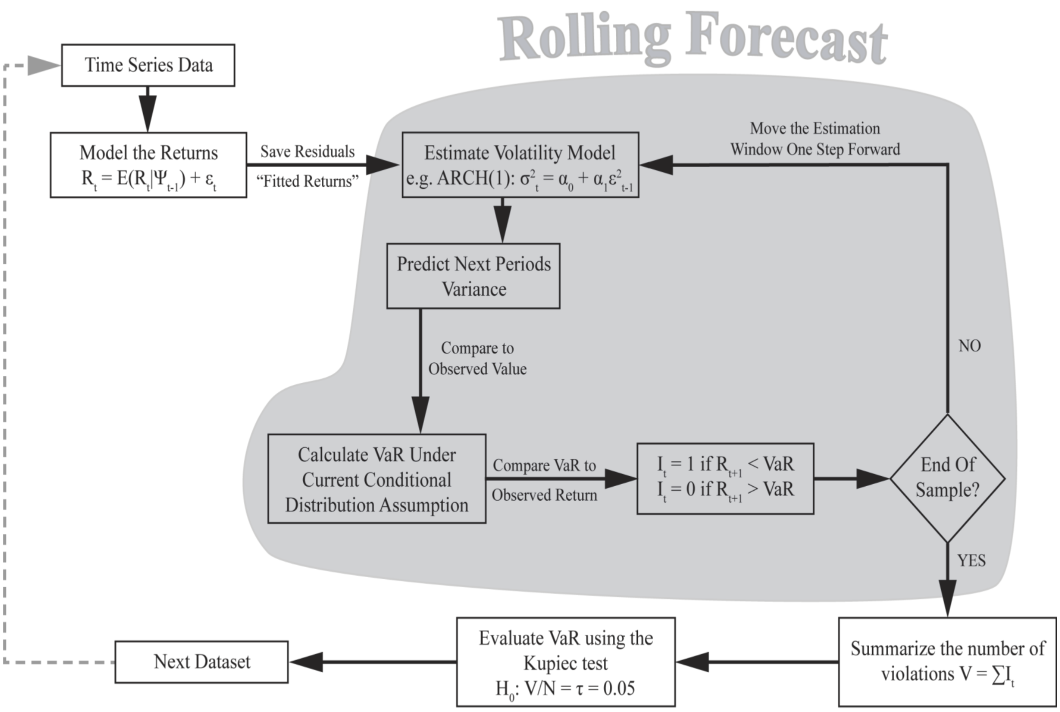

Figure 5. Garch rolling forecast

EXPERIMENT AND RESULT ANALYSIS:

Here the LSTM, GARCH, and ARIMA models have been analysed using the dataset obtained from the Central Statistical Office (CSO) under the Ministry of Statistics and Program Implementation, Government of India.

All the simulations of the various models have been done on a 64-bit Linux(Google Colab) system with AMD Ryzen 5 3550H CPU @2100 MHz with 16GB Memory and 107GB Storage. The development environment in Python 3.10 and the packages that have been used are Scikit-learn, NumPy, ARCH, TensorFlow, and Pandas.

Dataset:

Table 1: Dataset Description

SL. NO.	DATASET PARAMETERS	DESCRIPTION
1	No. of Documents	1
2	Document Size	47 kb
3	Dataset Dimensions	267 X 3
4	Data Source	CSO, MOSPI

The simulation has been done using “Vegetables” column as index out of the many columns available in the document. This has been done to minimize the effect of “seasonality” on the ARIMA model which affects its performance. The CPI of Vegetables are least affected by the seasonality factor.

Performance Analysis:

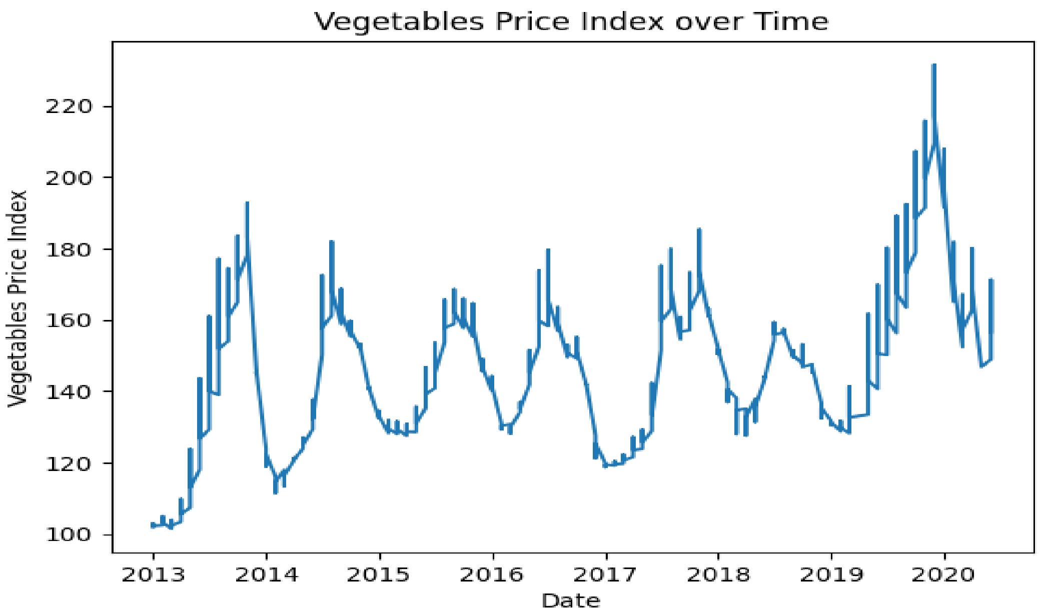

Figure 6: Graphical representation of fluctuation in the Price of Vegetables over the time

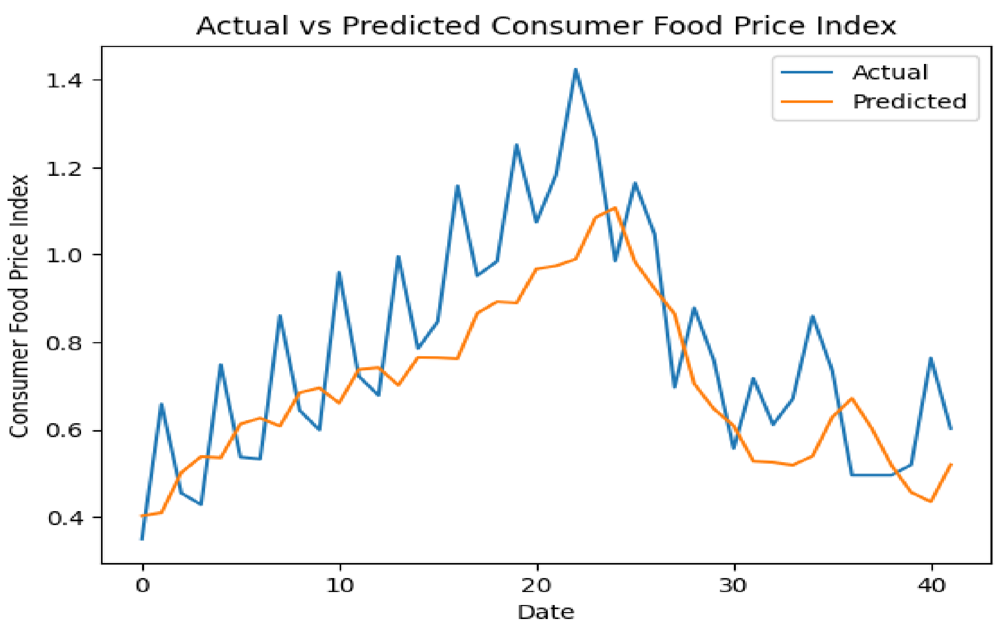

Figure 7: Graphical representation of Actual vs Predicted value for the **LSTM** model

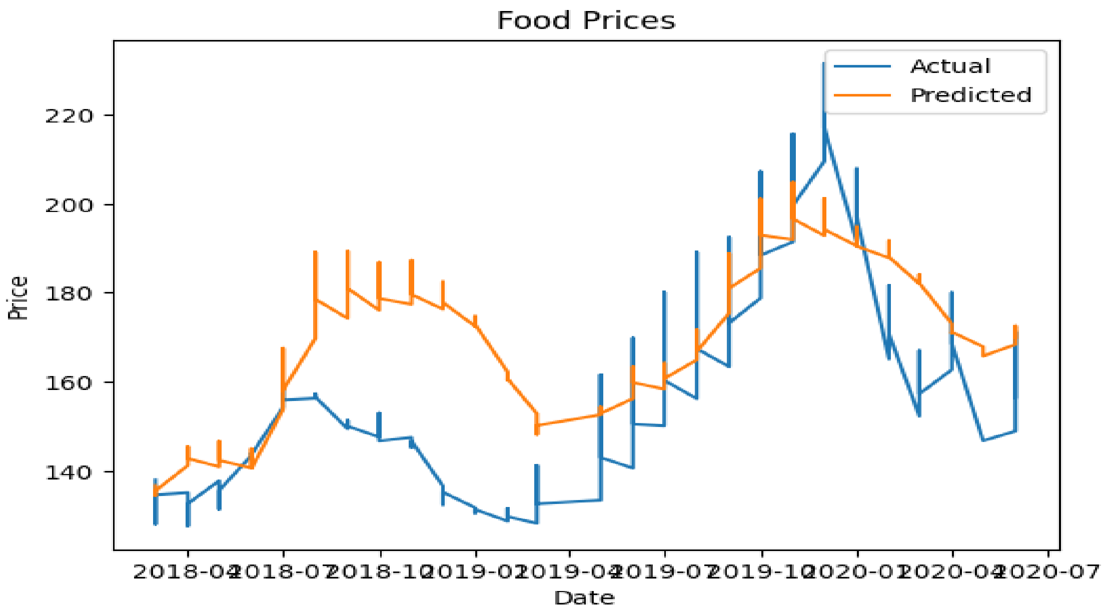

Figure 8: Graphical representation of Actual vs Predicted value for the **ARIMA** model

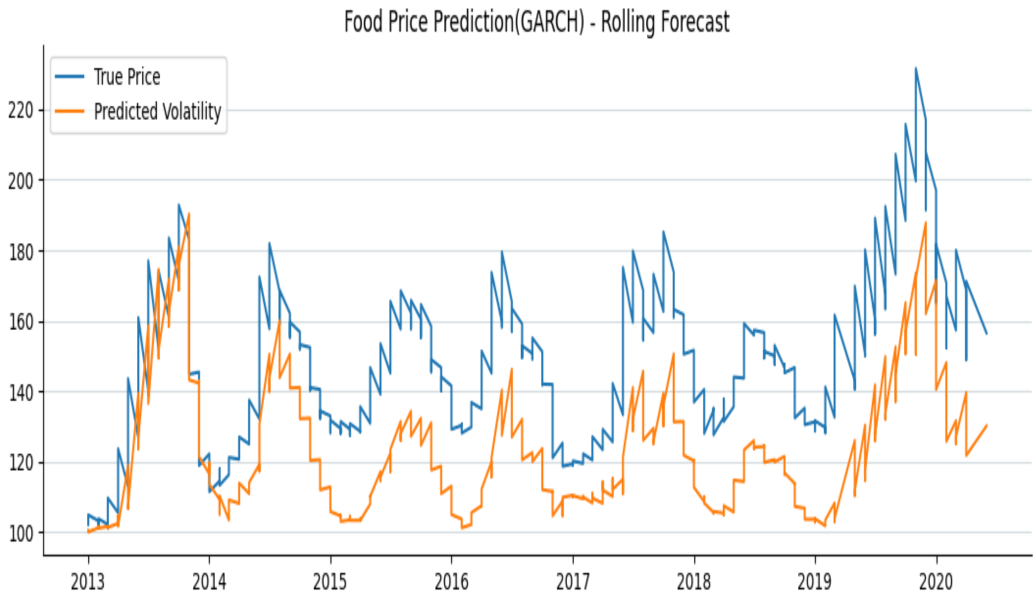

Figure 9: Graphical representation of Actual vs Predicted value for the **GARCH** model

Model Evaluation:

The Models' accuracy has been evaluated based upon the following parameters:

MAE(Mean Absolute Error) : It gives the average distance between the predicted value and the true value

MAE = (∑| actual - prediction|)/no. of observations (Equation 4)

MSE(Mean Squared Error) : It is the average of the square of the errors. The larger the number the larger the error.

MSE = (∑(| actual - prediction|)^2)/no. of observations (Equation 5)

RMSE(Root Mean Squared error) : RMSE is the square root of MSE. MSE is measured in units that are the square of the target variable, while RMSE is measured in the same units as the target variable.

RMSE = sqrt(MSE) (Equation 6)

MAPE(Mean Absolute Percentage Error) : It represents the average of the absolute percentage errors of each entry in a dataset to calculate how accurate the forecasted quantities were in comparison with the actual quantities.

MAPE = 1/n(∑|(actual - prediction)/actual)|) (Equation 7)

Table 2: Comparison of Different Models using various parameters

MODELS	LSTM	ARIMA	GARCH
MAE	12.47	16.65	24.29
MSE	385.23	444.84	765.55
RMSE	19.62	21.09	27.66
MAPE	8.84	11.18	16.04

CONCLUSION AND FUTURE WORK:

The most often used criterion for evaluating the accuracy of a time series forecast is the mean absolute percentage error (MAPE), which is employed in our data set, the consumer food price index (CFPI). The model is more accurate the lower the value of MAPE. The LSTM model is the best accurate model for a time series forecasting project, such as the prediction of food prices, according to the MAPE value seen in the above table. It also has the least absolute error.

In this experiment, three distinct time-series analysis and prediction models were examined. In the future, we intend to develop this project even more, develop a fresh hybrid model that makes use of these modules, and evaluate the accuracy and performance of the new model.

Future research will address the drawbacks of the current models, such as seasonality in the case of ARIMA. The hybrid model can use these models as a part of it or some different models in addition to these models.

REFERENCES

1. Sarangi, P.K., Chawla, M., Ghosh, P., Singh, S., Singh, P.K. (2021). FOREX trend analysis using machine learning techniques: INR vs USD currency exchange rate using ANN-GA hybrid approach, Materials Today:Proceedings.

2. Doganis P, Alexandridis A, Patrinos P, Sarimveis H (2006) Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. J Food Eng 75(2):196–204

3. https://medium.com/@rusirij/food-price-prediction-using-regression-model-training-and-predicting-638af744df1d

4. https://www.researchgate.net/publication/354863477_Forecasting_of_Inflation_Rate_Contingent_on_Consumer_Price_Index_Machine_Learning_Approach

5. G.E.P. Box, G. Jenkin (1970) Time Series Analysis, Forecasting and Control

6. https://www.investopedia.com/terms/b/box-jenkins-model.asp

7. https://www.investopedia.com/terms/g/garch.asp

8. Sarangi. P. K. & Sarangi, P. (2010). Short-Term Load Forecasting Using Neural Network Technology. IUP Journal of Computer Sciences, 4(2), pp 15-23.