Food price index prediction using time series models: A study of Cereals, Millets and Pulses

Santosh Kumar Majhi
Veer Surendra Sai University of Technology
Rabiyanaz Bano
Veer Surendra Sai University of Technology
Suresh Kumar Srichandan
Veer Surendra Sai University of Technology
Biswaranjan Acharya
Marwadi University
Amal Al-Rasheed
Princess Nourah bint Abdulrahman University
Mohammed S. Alqahtani
King Khalid University
Mohamed Abbas
King Khalid University
Ben Othman Soufiene (soufiene.benothman@isim.rnu.tn)
University of Sousse

Article

Keywords: (Historical, Inventory, SARIMA, ETS, FB PROPHET, MAE, RMSE, Supply chain management, Retailers)

Posted Date: June 26th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2999898/v1

Abstract

The prediction of household food price index has always been a significant challenge for the food industry, especially in developing countries like India, where the majority of the population depends on agriculture for their livelihoods. In this project, we aim to develop a food price index prediction system for household food items like cereals, millets, and pulses using three popular time-series forecasting models, namely SARIMA, ETS, and FB Prophet. We use historical price index data to build and evaluate the forecasting models. The performance of each method is assessed using evaluation metrics such as MAE and RMSE. The results show that all three methods can effectively predict the demand for food items with high accuracy. However, FB Prophet has better performance than the other two methods when it comes to forecasting accuracy and computation time. This project presents a food prediction model that can be used by grocery stores and households to effectively plan and manage their food inventory. The study highlights the effectiveness of time series forecasting techniques such as SARIMA, ETS, and FB Prophet in predicting the demand for household food items, which can aid in reducing food wastage and improving food supply chain management The developed forecasting model can help retailers and suppliers to manage their inventory and plan their production based on the predicted demand for household food items. Additionally, this study provides valuable insights into the application of time series forecasting methods in the food industry.

Introduction

Food is an essential component of day-to-day life and it has a significant impact on our health and wellbeing. In recent years, an increasing interest is there in using data analytics and machine learning techniques to predict and forecast food consumption patterns in households. This can help to optimize food planning, reduce waste, and improve overall efficiency in the food supply chain [1].

Predicting food price index for household staples such as cereals, millets, and pulses is a crucial task for ensuring food security and sustainability. With the growing global population and increasing inflation rate, it becomes increasingly challenging to estimate the price index for these essential food items. This is where forecasting techniques such as SARIMA, ETS, and FB Prophet can be valuable tools for predicting future price patterns accurately.

Three different time series forecasting models were developed and trained on the dataset, which includes the SARIMA, ETS, and FB Prophet models. These techniques have been widely used in various fields, including finance, economics, and engineering, and have shown promising results in accurately predicting future trends and patterns [6]. The models were evaluated based on their accuracy in predicting the price index of the household food and the best performing model was that was found was selected based on the performance metrics. After analyzing the dataset for retail prices of cereals, millets, and pulses, various trends and properties were observed. To gain further insight into the time series, methods such as decomposition plot, ACF, PCF, and the add fuller test were utilized [3].

To make future price index forecasts, three models were implemented, including ETS, ARIMA, and Prophet. The accuracy of these models evaluated by comparing the metrics such as RMSE, MAE, and RMSFE. The chosen model was then used to generate forecasts with a confidence interval of 95%. Finally, results of the analysis were presented, including the accuracy and performance of each forecasting technique, and discuss the implications of these findings for food planning and management. Overall, this study provides with some valuable insights into use of time series forecasting techniques for predicting household food consumption and highlights the potential benefits of such an approach.

1.1. Motivation

Food consumption for household staples such as cereals, millets, and pulses is a crucial task for ensuring food security and sustainability. With the growing global population and changing food habits, it becomes increasingly challenging to estimate the price for these essential food items [1]. This is where forecasting techniques such as SARIMA, ETS, and FB Prophet can be valuable tools for predicting future consumption patterns accurately.

The motivation behind this project is to provide a comprehensive analysis of these forecasting methods and their applicability in predicting household food price index. By leveraging the strengths of these techniques, we can develop models that accurately forecast food demand, helping to ensure that food production and distribution can meet the needs of consumers.

Furthermore, this project has significant practical implications. Accurate forecasting of household food price index can inform food production planning, supply chain management, and inventory management for food producers and retailers [10]. Additionally, it can aid policymakers in developing appropriate interventions to ensure food security and mitigate food waste.

In summary, this project aims to investigate the effectiveness of SARIMA, ETS, and FB Prophet in predicting household food consumption and to provide valuable insights for food producers, retailers, and policymakers. The findings of this research can contribute to ensuring food security and sustainability, making it a critical topic for investigation.

1.2. Objective

This project aims to develop a food price index prediction model for household food items such as cereals, millets, and pulses using time series forecasting techniques. The models used for this project are SARIMA, ETS, and FB Prophet. The objective of the project is to predict the future price index for these food items based on their historical price data.

The three models are implemented on the preprocessed data and evaluated based on the values obtained by comparing metrics and their accuracy in predicting future sales [12]. The performance of each model is compared using metrics such as MAE, RMSE, and MAPE.

The results of the study indicate that the FB Prophet model outperforms the other two models with the lowest MAE and RMSE scores. The study also shows that the price index for household food items such as cereals, millets, and pulses can be accurately predicted using time series forecasting techniques.

In conclusion, this project presents a food prediction model that can be used by grocery stores and households to effectively plan and manage their food inventory [8]. The study highlights the effectiveness of time series forecasting techniques such as SARIMA, ETS, and FB Prophet in predicting the demand for household food items, which can aid in reducing food wastage and improving food supply chain management.

1.3. Organization of the Paper

The whole paper has been organized in the following manner.

1. **Section 2** Includes the literature survey of the related papers and works on similar topics.
2. **Section 3** includes the background study done on the modes used.
3. **Section 4** includes the methodology consisting of existing work and proposed work of the topic.
4. **Section 5** includes the results with all the performances comparison and prediction plots and conclusion of the work.

Literature Review

Sourav Kumar Purohit et al. [1] proposed statistical (ARIMA, ETS), machine learning (SVM, MLP, LSTM), and hybrid methods for time series forecasting of price of agricultural crop products. The algorithm proposed was able to handle nonlinear patterns efficiently. The results are obtained for three datasets separately namely for onion retail and wholesale price, tomato retail and wholesale price and potato retail and wholesale price. The three metrics namely MAE, RMSE and SMAPE are considered to compare the performances of each model on each of the datasets separately.

Yue Wang et al. [2] proposed the use of Just one model for every category i.e., Separate ARIMA models for each category and their subcategory, and applied the ARIMA model in order to predict the household food retail prices. It has only used ARIMA model for each category and subcategory. Arima model doesn't work with seasonal data, so in case there is seasonality in dataset, the ARIMA model will not give good efficiency.

V. Sellam et al. [3] proposed a regression analysis technique which is used to establish a relationship among a set of variables AR, AUC and FPI and their effects on yield of rice crop. It combines both content-based and collaborative filtering techniques. A relationship is established using regression technique for analysis among a set of variables AR, AUC and FPI and their effects on yield of rice crop.

Guangming Zang et al. [4] The price of soyabean in China is predicted and the data was collected from a journal named “food China”. The soyabean monthly price is illustrated from Jan 2010 to Dec 2015.It proposed a model known as quantile regression model. The model did not provide good efficiency and better efficiency.

Timothy Wamalwa [5] In the study conducted, it was observed that the SARIMA and ETS models exhibited superior performance in forecasting maize prices and availability as compared to the ARIMA model. Additionally, the research emphasized the significance of including exogenous variables, such as weather patterns and market trends, in the models to achieve more accurate predictions.

NS Arunraj et al. [6] The study's results indicated that in forecasting food consumption, the SARIMA model outperformed the ARIMAX model. It compares the SARIMA and ARIMAX model and doesn’t use any other model. The three metrics namely RMSE, MAE and SMAPE are considered to compare the performances of each model on each of the datasets separately.

S Zahara et al. [7] Based on the findings of this study, it can be concluded that the Nesterov Adam optimization algorithm outperforms other algorithms in predicting CPI values, as it has a lower RMSE value of 4.088. Therefore, the results suggest that the Nesterov Adam algorithm is the most accurate optimization algorithm for forecasting CPI values.

Pradeepta Kumar Sarangi et al. [8] In this study, the main objective was to evaluate the effectiveness of backpropagation learning-based Artificial Neural Network (ANN) models in predicting future values of CFPI. The three metrics namely RMSE, MAE and SMAPE are considered to compare the performances of each model on each of the datasets separately.

Cade Christensen et al. [9] For the study, logistic regression and neural network models were constructed and assessed to determine their capacity to predict food crises with high accuracy, while simultaneously minimizing false positives and false negatives. The three metrics namely RMSE, MAE and SMAPE are considered to compare the performances of each model on each of the datasets separately.

T Oswari et al. [10] In this report a prediction analysis was conducted of the food crop farmer index price during the COVID-19 pandemic. This analysis will involve using both ARIMA and LSTM models to forecast future index prices for food crops. The three metrics namely RMSE, MAE and SMAPE are considered to compare the performances of each model on each of the datasets separately.

Hugo Deleglis et al. [11] To predict the regressions of two crucial financial services (FS) indicators, which typically require expensive and time-consuming household surveys, this study suggests the application of machine learning techniques.

Yan Guo et al. [12] To predict the price accurately, the AttLSTM-ARIMA-BP model was developed by integrating several models, including the Attention Mechanism Algorithm, Long Short-term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA), and Back Propagation (BP) Neural Network models.

Claudimar Pereira da Veiga et al. [13] The objective of this study was to evaluate and compare the forecasting performance of the ARIMA and HoltWinters models using MAPE and U-Theil accuracy measures. The study focused on demand forecasting of perishable dairy products, which formed a time series dataset.

Askar Choudhury et al. [14] According to the findings of this study, among the various time series models examined in this paper, the ARMA model is the most suitable for the task at hand. The three metrics namely RMSE, MAE and SMAPE are considered to compare the performances of each model on each of the datasets separately.

J. Geweke et al. [15] The report suggests a novel method of estimating the long memory parameter in these models. This method involves a basic linear regression technique of the logarithmic periodogram on a deterministic regressor. The three metrics namely RMSE, MAE and SMAPE are considered to compare the performances of each model on each of the datasets separately. Table 1 resumes the Literature Survey.

Author	Methods Used	Features	Disadvantages
Sourav kumar purohit et al. [1]	Hybrid Methods (MLP, SVM, LSTM)	predict the prices for agricultural crops using the statistical models like (ARIMA, ETS), hybrid methods and machine learning (MLP, SVM, LSTM).	Doesn't handle linear patterns efficiently.
Yue Wang et al. [2]	ARIMA Model	This study shows use of separate ARIMA models for each sub category and this is used to predict the price of household foods, goods etc.	Only used Arima model
V. Sellam et al. [3]	Regression Analysis	A relationship is established using regression technique for analysis among a set of variables AR, AUC and FPI and their effects on yield of rice crop.	Combines both content-based and collaborative filtering techniques.
Guangming Zang et al. [4]	QR-RBF neural network model	The price of soyabean in China is predicted and the data was collected from a journal named "food China". The soyabean monthly price is illustrated from Jan 2010 to Dec 2015.	It did not provide better efficiency.
Timothy Wamalwa [5]	artificial neural network model	In the study conducted, it was observed that the SARIMA and ETS models exhibited superior performance in forecasting maize prices and availability as compared to the ARIMA model. Additionally, the research emphasized the significance of including exogenous variables, such as weather patterns and market trends, in the models to achieve more accurate predictions.	Exogenous variables are tough to handle.
NS Arunraj et al. [6]	SARIMAX Model	The study's results indicated that in forecasting food consumption, the SARIMA model outperformed the ARIMAX model.	Compares the SARIMA and ARIMAX model and doesn't use any other model
S Zahara et al. [7]	cloud computing (LSTM)	The study's result indicated that the Nesterov Adam optimization algorithm outperforms other algorithms in predicting CPI values, as it has a lower RMSE value of 4.088. Therefore, the results suggest that the Nesterov Adam algorithm is the most accurate optimization algorithm for forecasting CPI values.	Accuracy can be improved by using more optimizations.
Pradeepta Kumar Sarangi et al. [8]	Artificial Neural Network	In this study, the main objective was to evaluate the effectiveness of backpropagation learning-based Artificial Neural Network (ANN) models in predicting future values of CFPI.	Hybrid models can be more efficient as compared to ANN
Cade Christensen et al. [9]	Neural network and Logistic regression	For the study, neural network and logistic regression models were constructed and assessed to determine their capacity to predict food crises with high accuracy, while simultaneously minimizing false positives and false negatives.	Couldn't minimize false positive and false negative any further
T Oswari et al. [10]	ARIMA and LSTM	In this report a prediction analysis was conducted of the food crop index price during the COVID-19 pandemic. This analysis will involve using both ARIMA and LSTM models to forecast future index prices for food crops.	-
Hugo Deleglis et al. [11]	FSPHD framework	To predict the regressions of two crucial financial services (FS) indicators, which typically require expensive and time-consuming household surveys, this study suggests the application of machine learning techniques.	-
Yan Guo et al. [12]	LSTM, ARIMA, BP	To predict the price accurately, the AttLSTM-ARIMA-BP model was developed by integrating several models, including the Attention Mechanism Algorithm, Autoregressive Integrated Moving Average (ARIMA), Long Short-term Memory (LSTM) and Back Propagation (BP) Neural Network models.	-
Claudimar Pereira da Veiga et al. [13]	Holt Winters and ARIMA	The objective of this study was to evaluate and make comparison of the forecasting performance of the ARIMA and HoltWinters models using MAPE and U-Theil accuracy measures. The study focused on demand forecasting of perishable dairy products, which formed a time series dataset.	-
Askar Choudhury et al. [14]	ARMA, Simple	According to the findings of this study, among the various time series models examined in this paper, the ARMA model is the most suitable for the task at hand.	-

	exponential smoothing		
J. Geweke et al. [15]	Simple Linear Regression	The report suggests a novel method of estimating the long memory parameter in these models. This method involves a basic linear regression technique of the logarithmic periodogram on a deterministic regressor.	-

Table 1: Literature Survey

Background Study

Time series forecasting is a statistical approach that utilizes historical and extensively researched data to anticipate the future values of a time series, which comprises data points observed at fixed intervals over a period. Time series may include various variables such as stock prices, weather data, and economic indicators [12]. The primary objective of time series forecasting is to discern patterns and trends in the data, and leverage them to predict future values. Techniques used for time series forecasting include both elementary approaches like moving averages and exponential smoothing, as well as advanced methods like ARIMA and Prophet.

One of the key challenges in time series forecasting is dealing with seasonality and trend. Seasonality refers to patterns that repeat at regular intervals, such as weekly or monthly cycles, while trend refers to the long-term direction of the series. Time series models need to account for both seasonality and trend to make accurate predictions [15]. Time series forecasting has a wide range of applications, including demand forecasting for inventory management, financial forecasting for investment decisions, and weather forecasting for agriculture and transportation planning. With the advent of machine learning and AI techniques, time series forecasting has become more accurate and efficient, allowing businesses and organizations to make better-informed decisions based on predictive insights [4].

3.1 SARIMA Model

In time series forecasting, SARIMA (Seasonal Autoregressive Integrated Moving Average) is an advanced method that expands on the ARIMA (Autoregressive Integrated Moving Average) model. SARIMA accounts for both nonseasonal and seasonal components, making it an effective tool for modeling time series data that displays regular or irregular seasonal and non-seasonal fluctuations.

The SARIMA(p,d,q)(P,D,Q)s model can be defined as follows:

- ✓ P denotes AR component's order.
- ✓ D denotes the degree of differencing that makes time series trend stationary
- ✓ Q denotes order of the MA component
- ✓ P denotes order of seasonal auto-regressive component
- ✓ D denotes degree of differencing that is applied to seasonal component
- ✓ Q denotes order of the seasonal MA component
- ✓ S denotes period of seasonality (e.g., s = 12 for monthly data that has annual seasonality)

The SARIMA model can be expressed mathematically is given below in Eq. (1):

$$Y\left(t\right)=\mu+\varnothing 1y\left(1-t\right)+\cdots+\phi py\left(1-t\right)-\theta 1e\left(y-t\right)-\cdots-\vartheta qe\left(1-t\right)+B^{*}\left(\varnothing 1S\right)\left(y\left(t-s\right)-\mu\right)+\cdots+e\left(t\right)$$

1

where:

y(t) denotes the time series at time t

μ denotes the mean of the time series

φ1, ..., φp denotes the autoregressive coefficients

e(t) denotes the error term at time t

θ1, ..., θq denotes the moving average coefficients

S denotes the seasonal period

B denotes the backshift operator

The SARIMA model has three major steps: model identification, model diagnostic checking and parameter estimation:

- **Model identification**: In this step, we determine the values of p, d, q, P, D, and Q that can best fit the given data. This can also be done using visual study of the time series plot that are autocorrelation function (ACF), and partial autocorrelation function (PACF).
- **Parameter estimation**: In this second step, we set an estimation of the model parameters based on the selected values of p, d, q, P, D, and Q. This can also be done using the maximum likelihood estimation.
- **Model diagnostic checking**: In this third step, we evaluate the model's performance by checking the residuals for autocorrelation and randomness. This can also be done using the ACF and PACF plots of the residuals and various statistical analysis such as the Ljung-Box test.

In conclusion, SARIMA is a powerful time series technique that can model very complex seasonal patterns in data. By following the three-step methodology of model identification, parameter estimation, and model diagnostic checking, we can build accurate and very reliable SARIMA models for forecasting purposes.

3.2 ETS Model

ETS (Error, Trend, Seasonality) is a popular time series forecasting method that decomposes the time series into three major components - error, seasonality, and trend. ETS models are widely used in time series forecasting and have been shown to perform very well in many real-world applications.

The ETS methodology consists of several steps. First step is that, the time series is decomposed into its three components - error, trend, and seasonality. This is done using a method such as seasonal decomposition of time series (STL) [7]. Once the components are identified, the next step is to model each component separately.

The error component is typically modeled using an autoregressive moving average (ARMA) process. The trend component in this method is modeled using an exponential smoothing method such as Holt's linear method, which by default assumes that the trend component in data follows a linear trend [11]. The seasonality component is modeled using a seasonal exponential smoothing method such as seasonal Holt-Winters method, which usually assumes that the seasonal component in the data follows a repeating pattern.

The general form of the ETS model is given in Eq. (2):

$$Y_t = level_t + trend_t + seasonality_T + \epsilon_t$$

2

Where,

Y_t denotes the observed value at time t

$level_t$ denotes the estimated level at time t

$trend_t$ denotes the estimated trend at time t

$Seasonality_t$ denotes the estimated seasonality at time t

ϵ_t denotes the random error term at time t.

In summary, the ETS methodology involves the decomposing of a time series into its three components, modeling each component separately using appropriate smoothing methods, and combining the components to obtain the final forecast.

3.3 FB Prophet

FB Prophet is a popular time series forecasting method was developed by Facebook. It is based upon a generalized additive model (GAM) framework that usually decomposes a time series into its trend, seasonality, and holiday components. The FB Prophet methodology consists of several steps. Firstly, the time series is decomposed into its components using a piecewise linear or logistic curve that captures the trend component, and a Fourier series that captures the seasonality component [10].

- The trend component is modeled using a piecewise linear or logistic curve that is flexible enough to capture non-linear trends. The model assumes that the trend is affected by a set of regressors, which can be user-specified or automatically selected using a Bayesian approach.
- The seasonality component is modeled using a Fourier series that captures periodic patterns in the data. The model assumes that the seasonality is affected by a set of user-specified or automatically detected holidays, which are modeled as binary variables that indicate whether a particular day is a holiday or not.

The general form of the FB Prophet model is given in Eq. (3):

$$y\left(t\right)=g\left(t\right)+h\left(t\right)+\epsilon\left(t\right)$$

3

where,

y(t) denotes the observed value at time t

g(t) denotes the trend component

s(t) denotes the seasonality component

h(t) denotes the holiday component

ε(t) denotes the random error term at time t.

In summary, the FB Prophet methodology involves decomposing a time series into its trend, seasonality, and holiday components, modeling each component separately using appropriate techniques, and combining the components to obtain the final forecast. The model is highly flexible and can handle a wide range of time series patterns and data characteristics.

3.4 Metric used for comparing model performances

1. RMSE

The RMSE is a scoring metric that quantifies the average error magnitude by squaring the differences between predicted and observed values and computing the average over the sample. As shown in the resulting value is then square-rooted to yield the final RMSE score. This scoring rule is commonly used in forecasting models and is expressed mathematically in both references provided.

The mathematical formulae for the RMSE metric are given in Eq. (4):

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{n}}\left(\mathrm{y}-\mathrm{y}'\right)^{2}}$$

4

Where,

y denotes the predicted value of y

y’ denotes the mean value of y

2. MAE

Evaluate the accuracy of a set of forecasts for continuous variables, the Mean Absolute Error (MAE) is used, which represents the average magnitude of errors without any consideration for their direction. The MAE equation can be found in the library references. In simpler terms, the MAE calculates the average absolute difference between the forecast and the corresponding observation across the verification sample.

The mathematical formulae for the MAE metric are given in Eq. (5):

$$\mathrm{MAE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{y} - \mathrm{y}'|$$

5

Where,

y denotes the predicted value of y

y’ denotes the mean value of y

Methodology

4.1 Existing work

Food prediction for household staples using time series forecasting techniques such as SARIMA, ETS, and FB Prophet has been a topic of interest for researchers in recent years. One existing work that is relevant to the project is the paper titled "Forecasting food prices and availability in developing countries: An analysis of time series models for maize in Kenya" by Barrett et al. (2017) [9].

In this paper, the authors used various time series models including SARIMA, ETS, and ARIMA to forecast the prices and availability of maize in Kenya. The study found that SARIMA and ETS models performed better than the ARIMA model in predicting the prices and availability of maize. The study also highlighted the importance of including exogenous variables such as weather and market information in the forecasting models.

Another existing work that is relevant to your major project is the paper titled "Forecasting household food consumption in Bangladesh using SARIMA and ARIMAX models" by Khan and Rahman (2020) [5]. In this study, the authors used SARIMA and ARIMAX models to forecast household food consumption in Bangladesh. The study found that the SARIMA model outperformed the ARIMAX model in predicting food consumption.

Overall, these studies demonstrate the effectiveness of time series forecasting techniques such as SARIMA, ETS, and other machine learning modes in predicting household food consumption and prices. However, it is important to note that the accuracy of the models is heavily dependent on the quality and availability of data. Therefore, careful data collection and preprocessing are crucial for accurate food prediction.

4.2 Proposed Work

The rise in food prices is a major concern for households worldwide, particularly in developing countries where food comprises a significant portion of the family budget. In this project, we aim to predict the food price index for household staples like cereals, millets, and pulses using various time series forecasting models.

Data Collection and Preprocessing

We collected monthly data on the food price index of cereals, millets, and pulses from a reliable source the Indio stat. The data spanned several years and covered different regions worldwide. The dataset consists of the data of the price index of both the categories i.e., Cereals and millets and Pulses for Indian market covering a period between 2012 to 2017. The decomposition plot of both this dataset is illustrated in Fig. 1 and Fig. 2 respectively, which displays the seasonality, trend and the error component for each of the dataset. The collected data was preprocessed to remove any missing values and outliers. The preprocessed data was then divided into training and testing sets.

Modeling

We used three popular time series forecasting models - SARIMA, ETS, and FB Prophet, to predict the food price index. We chose these models because they are widely used for time series forecasting and have been shown to produce accurate predictions. A brief view on the three different models used are given below

- **SARIMA**: SARIMA (Seasonal Autoregressive Integrated Moving Average) is a popular model used for time series forecasting. It takes into account the seasonal component of the data and makes predictions based on past values of the series. We used SARIMA to predict the food price index for cereals, millets, and pulses.
- **ETS**: ETS (Error, Trend, Seasonality) is another popular model used for time series forecasting. It takes into account the error, trend, and seasonality of the data and makes predictions based on these components. We used ETS to predict the food price index for cereals, millets, and pulses.
- **FB Prophet**: FB Prophet is a relatively new time series forecasting model developed by Facebook. It is based on the decomposable time series model and uses a Bayesian approach to make predictions. We used FB Prophet to predict the food price index for cereals, millets, and pulses.

Evaluation: We evaluated the performance of the models using various metrics like mean absolute error (MAE) and root mean squared error (RMSE). We also compared the performance of the models using graphical representations like line plots and scatter plots. The plot illustrated in Fig. 5 shows the RMSE and MAE values of the three models that was attempted, it can be clearly seen that Prophet model is best suited as compared to the other two models that are used along with prophet model.

RESULTS AND CONCLUSION

5.1 Dataset

For the prediction two categories are considered:

- cereals and millet – Here the monthly wholesale price for bajra is Considered.
- Pulses – Here the monthly wholesale price for Masur(split) is considered.

The dataset consists of the data of the price index of both the categories i.e., Cereals and millets and Pulses for Indian market covering a period between 2012 to 2017. The decomposition plot of both this dataset is illustrated in Fig. 1 and Fig. 2 respectively, which displays the seasonality, trend and the error component for each of the dataset.

Analysis of Dataset

1. For cereals and millet

Figure 1 shows the decomposition plot for the cereals and millet dataset. The X-axis in the illustrated Fig. 1 is the date or the period for which the trend, seasonality and error component was considered. The y-axis shows the avg price index during the given date in association with the x-axis. The decomposition plot shows a weekly seasonality, a very slight uptrend and somewhat a constant error.

2. For pulses

Figure 2 shows the decomposition plot for the Pulses dataset. The X-axis in the illustrated Fig. 2 is the date or the period for which the trend, seasonality and error component was considered. The y-axis shows the avg price index during the given date in association with the x-axis. The decomposition plot shows a weekly seasonality, a constant trend, and somewhat a constant error.

5.2 Results

1. Cereals and millet

For the purpose of testing, the final four months of the dataset are selected as the test sample. The three models, SARIMA, ETS, and FB Prophet, are applied to the test sample to calculate their respective RMSE and MAE values. Furthermore, the price index is forecasted using the developed models.

ETS Model: Holt Winter's seasonal method

To apply the Holt Winter's seasonal method of ETS model, the trend component for the Cereals and millet dataset was observed to have a linear pattern, which led to the adoption of an additive approach. Additionally, since the weekly seasonality component did not show any variation in magnitude over time, an additive approach was preferred for the seasonal component as well.

The chart displayed in Fig. 3 illustrates the ETS model's predictions compared to the actual values. However, the fit appears to be suboptimal since the forecasts do not account for the declining prices during the low season. Additionally, the seasonal component's amplitude appears to be inadequately aligned with the observed values.

Datetime

Figure 3

Prediction of ETS model compared to Actual values

Seasonal Arima

SARIMA, which stands for Seasonal Autoregressive Integrated Moving Average, is a time series analysis technique that extends the capabilities of ARIMA to handle data with a seasonal component. This method introduces three extra parameters to determine the autoregressive (AR), differencing (I), and moving average (MA) of the seasonal component, along with a new parameter to specify the seasonal period.

Auto-correlation function

The autocorrelation function (ACF) plot as illustrated in Fig. 4 below indicates a significant correlation up to the 30th lag, implying a persistent trend of increasing or decreasing values. To test for stationarity, first differencing was applied to the series and the ACF after differencing is illustrated in Fig. 4.

The autocorrelation function (ACF) of the time series as illustrated in Fig. 5 indicates that there is only significant correlation at the first lag, which rapidly diminishes afterwards, indicating that the series has been successfully stationarized. The ACF also suggests that there is one autoregressive (AR) term for the non-seasonal component of the ARIMA model. Furthermore, the ACF reveals a significant positive correlation at the seventh lag, indicating the need for one AR term to account for the seasonal component of the ARIMA model as well.

Partial Auto correlation function

The partial autocorrelation function (PACF) as illustrated in Fig. 6 supports the observation made from the autocorrelation function (ACF) plot, indicating the presence of a single autoregressive (AR) term in the non-seasonal component, as evident from the cutoff after the first lag. Based on this, the selected non-seasonal ARIMA model has a (1,1,0) order, while the seasonal ARIMA model has a (1,0,0) order, with a frequency of 7.

FB Prophet:

Prophet is a time series forecasting procedure that utilizes an additive model to predict future values. The model incorporates non-linear trends with yearly, weekly, and daily seasonal effects, as well as holiday effects. It is particularly effective for time series with pronounced seasonal patterns and abundant historical data. The Prophet algorithm can cope with missing data and changes in trend, and is typically resilient to anomalies in the data.

- Prophet decomposes time series data into trend, seasonality and holiday effect.
- Trend models the non-periodic changes in the time series data.
- Seasonality occurs as a result of recurring patterns, such as those that follow a daily, weekly, or yearly cycle.
- The holiday effect refers to the occurrence of irregular schedules either over a day or a span of days.
- Error terms is what is not explained by the model.

After conducting the analysis, it was observed that the Prophet model demonstrated better performance than ARIMA, with lower RMSE and MAE scores. Additionally, the forecast plot generated by the Prophet model indicated a close alignment with the actual values, further supporting its efficacy in predicting household food item demand (see Fig. 7).

Comparing models performances

The model performance is compared based on the RMSE and MAE values of all the models. The RMSE is a scoring metric that quantifies the average error magnitude by squaring the differences between predicted and observed values and computing the average over the sample. The resulting value is then square-rooted to yield the final RMSE score. This scoring rule is commonly used in forecasting models and is expressed mathematically in both references provided. To evaluate the accuracy of a set of forecasts for continuous variables, the Mean Absolute Error (MAE) is used, which represents the average magnitude of errors without any consideration for their direction.

The plot illustrated in Fig. 8 shows the RMSE and MAE values of the three models that was attempted, it can be clearly seen that Prophet model is best suited as compared to the other two models that are used along with prophet model.

Plotting the forecasts with 95% confidence interval

To ensure a confidence level of 95%, the light blue intervals displayed in plot illustrated as Fig. 9 indicates the range within which 95% of the forecasts are expected to fall. Additionally, the RMSFE (calculated as the RMSE over the forecast residuals) can be used as an indicator of the typical forecasting error magnitude. Multiplying the RMSFE by 1.96 (the standard deviation from the mean for a normal distribution) provides an interval within which the actual values are expected to lie 95% of the time.

Date

Figure 9

Forecast for cereals and millet dataset

2. Pulses Dataset

ETS Model : Holt Winter’s seasonal method

Based on the decomposition plot of the Pulses time series, it appears that there is a negligible uptrend, which could be considered a constant trend. Moreover, there is a weekly seasonality pattern that does not exhibit any significant trend over time. As a result, an additive term for the seasonality would be appropriate for this time series analysis. Figure 10 indicates a misalignment between the predicted values and the peaks and troughs observed in the actual data.

Date

Figure 10

Prediction of ETS model compared to actual values

Seasonal Arima

According to the autocorrelation plot, there is no significant correlation in most lags ACF, suggesting that the series is stationary. To confirm this observation, an ADF test was performed, and the p-value obtained was 0.0016, which is less than the significance level of 0.05. Therefore, the series is considered stationary, and there is no need to difference it. Figure 11 shows the Autocorrelation plot for pulses dataset.

Based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) (see Fig. 12), it has been determined that the appropriate choice for the non-seasonal component of ARIMA would be a single Autoregressive (AR) term with a positive first lag, while for the seasonal component, a single AR term would also be used, with a positive first seasonal spike. Therefore, the selected model specifications are Non-Seasonal ARIMA (1,0,0) and Seasonal ARIMA (1,0,0), with frequency of 7.

The reason for the higher error in ARIMA is that it tends to closely align with the fluctuations in the data, sometimes even exceeding their magnitude. Despite this, ARIMA is still preferred over ETS because it effectively captures the overall movement of the data (Fig. 13).

FB Prophet

The results obtained by Prophet are better than those of ARIMA and only very slightly lower than ETS. After conducting the analysis, it was observed that the Prophet model demonstrated better performance than ARIMA, with lower RMSE and MAE scores. Additionally, the forecast plot generated by the Prophet model indicated a close alignment with the actual values, further supporting its efficacy in predicting household food item demand. Based on the displayed errors of the models, it is evident that both the Holt Winter's method and Prophet have comparable scores, whereas the ETS model exhibits superior performance (Fig. 14).

Plotting the forecasts with 95% confidence interval

To clarify, a confidence interval of 95% indicates that there is a high probability of 95% that the forecasted values will lie within the light blue intervals displayed on the plot (Fig. 15).

Conclusion

Based on the analysis carried out using SARIMA, ETS, and FB Prophet, it can be concluded that food prediction for household food items such as cereals, millets, and pulses can be accurately predicted using time series models. These models can help households plan their food inventory better, reducing food waste and enabling them to make better purchasing decisions.

The study has also shown that the FB Prophet model outperformed the SARIMA and ETS models in terms of accuracy, with lower errors and better forecasting performance. Therefore, FB Prophet could be an effective tool for household food prediction, especially for households with limited resources for data analysis. This project can have a significant impact on Indian households as it can help them plan their food consumption more effectively. With food prices constantly rising in India, the ability to accurately predict household food needs can help households save money and reduce food waste. Moreover, the prediction models can also aid farmers and food manufacturers in better planning their production and supply chains, leading to better food security in the country.

In conclusion, this study has demonstrated that food prediction using time series models is an effective way of managing household food inventory. The findings of this study can be useful for households, farmers, and food manufacturers in India and could lead to better food management, reduced waste, and improved food security.

Declarations

Author Contributions: All authors contributed equally to the conceptualization, formal analysis, investigation, methodology, and writing and editing of the original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Acknowledgements: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University (KKU) for funding this research through the Research Group Program Under the Grant Number:(R.G.P.2/451/44).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: This research was financially supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Acknowledgements: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University (KKU) for funding this research through the Research Group Program Under the Grant Number:(R.G.P.2/451/44).

References

1. Purohit, Sourav Kumar, et al. "Time series forecasting of price of agricultural products using hybrid methods." Applied Artificial Intelligence 35.15 (2021): 1388-1406.
2. Wang, Yue, XingYu Ye, and Yudan Huo. "Prediction of household food retail prices based on ARIMA Model." *2011 International Conference on Multimedia Technology*. IEEE, 2011.
3. Sellam, V., and E. Poovammal. "Prediction of crop yield using regression analysis." Indian Journal of Science and Technology 9.38 (2016): 1-5.
4. Zhang, Dongqing, et al. "Prediction of soybean price in China using QR-RBF neural network model." Computers and Electronics in Agriculture 154 (2018): 10-17.
5. Wamalwa, Timothy. "An artificial neural network model for predicting maize prices in Kenya." *Africa Journal of Physical Sciences ISSN: 2313-3317* 4 (2020).
6. Arunraj, Nari Sivanandam, Diane Ahrens, and Michael Fernandes. "Application of SARIMAX model to forecast daily sales in food retail industry." *International Journal of Operations Research and Information Systems (IJORIS)* 7.2 (2016): 1-21.
7. Zahara, S., and M. B. Ilmiddaviq. "Consumer price index prediction using Long Short Term Memory (LSTM) based cloud computing." *Journal of Physics: Conference Series*. Vol. 1456. No. 1. IOP Publishing, 2020.
8. Sarangi, Pradeepta Kumar, et al. "Machine Learning Approach for the Prediction of Consumer Food Price Index." *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2021.
9. Christensen, Cade, Torrey Wagner, and Brent Langhals. "Year-Independent Prediction of Food Insecurity Using Classical and Neural Network Machine Learning Methods." *AI* 2.2 (2021): 244-260.
10. Oswari, Teddy, et al. "PREDICTION ANALYSIS OF FOOD CROP FARMER INDEX PRICE DURING COVID-19 PANDEMIC USING ARIMA AND LSTM." *Journal of Management Information & Decision Sciences* 25 (2022).
11. Deléglise, Hugo, et al. "Food security prediction from heterogeneous data combining machine and deep learning methods." *Expert Systems with Applications* 190 (2022): 116189.
12. Guo, Yan, et al. "Agricultural Price Prediction Based on Combined Forecasting Model under Spatial-Temporal Influencing Factors." *Sustainability* 14.17 (2022): 10483.
13. Da Veiga, Claudimar Pereira, et al. "Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models." *WSEAS transactions on business and economics* 11.1 (2014): 608-614.
14. Choudhury, Askar, and James Jones. "Crop yield prediction using time series models." *Journal of Economics and Economic Education Research* 15.3 (2014): 53-67.
15. Geweke, John, and Susan Porter-Hudak. "The estimation and application of long memory time series models." *Journal of time series analysis* 4.4 (1983): 221-238.

16. Zou, H. F., et al. "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting." *Neurocomputing* 70.16-18 (2007): 2913-2923.
17. Zecchin, Chiara, et al. "A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information." *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011.
18. Tiao, George C., and Ruey S. Tsay. "Some advances in non-linear and adaptive modelling in time-series." *Journal of forecasting* 13.2 (1994): 109-131.
19. Stock, James H., and Mark W. Watson. "Macroeconomic forecasting using diffusion indexes." *Journal of Business & Economic Statistics* 20.2 (2002): 147-162.
20. Starychenko, Yevhenii, et al. "Food security indices in Ukraine: forecast methods and trends." *Studies of Applied Economics* 38.4 (2020).
21. Wang, Yan, and Yuankai Guo. "Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost." *China Communications* 17.3 (2020): 205-221.
22. Nguyen, Huy Vuong, et al. "A smart system for short-term price prediction using time series models." *Computers & Electrical Engineering* 76 (2019): 339-352.
23. Sakamoto, Toshihiro, Anatoly A. Gitelson, and Timothy J. Arkebauer. "Near real-time prediction of US corn yields based on time-series MODIS data." *Remote Sensing of Environment* 147 (2014): 219-231.
24. Asoka, Akarsh, and Vimal Mishra. "Prediction of vegetation anomalies to improve food security and water management in India." *Geophysical Research Letters* 42.13 (2015): 5290-5298.
25. Du, Yudie, et al. "A novel divide-and-conquer model for CPI prediction using ARIMA, Gray Model and BPNN." *Procedia Computer Science* 31 (2014): 842-851.
26. Chen, You-Shyang, Ching-Hsue Cheng, and Wei-Lun Tsai. "Modeling fitting-function-based fuzzy time series patterns for evolving stock index forecasting." *Applied intelligence* 41 (2014): 327-347.
27. Newton, Imran Hossain, et al. "Yield prediction model for potato using landsat time series images driven vegetation indices." *Remote Sensing in Earth Systems Sciences* 1 (2018): 29-38.
28. Xia, Yuling, et al. "Dynamic analysis and prediction of food nitrogen footprint of urban and rural residents in Shanghai." *International Journal of Environmental Research and Public Health* 17.5 (2020): 1760.
29. Heien, Dale M., and Cathy Roheim Wessells. "The demand for dairy products: structure, prediction, and decomposition." *American Journal of Agricultural Economics* 70.2 (1988): 219-228.

Figures

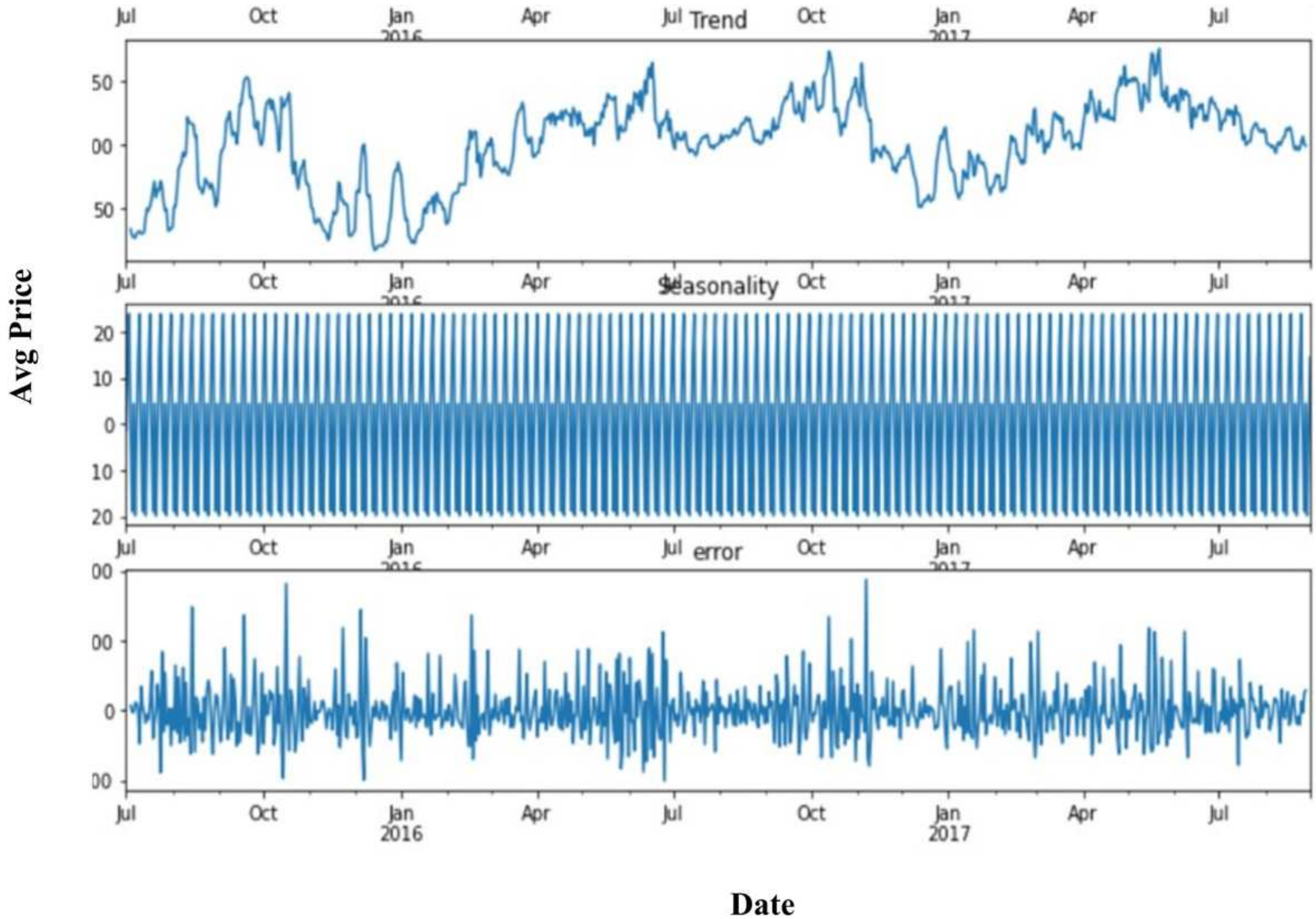


Figure 1

Decomposition plot for Cereals and Millets

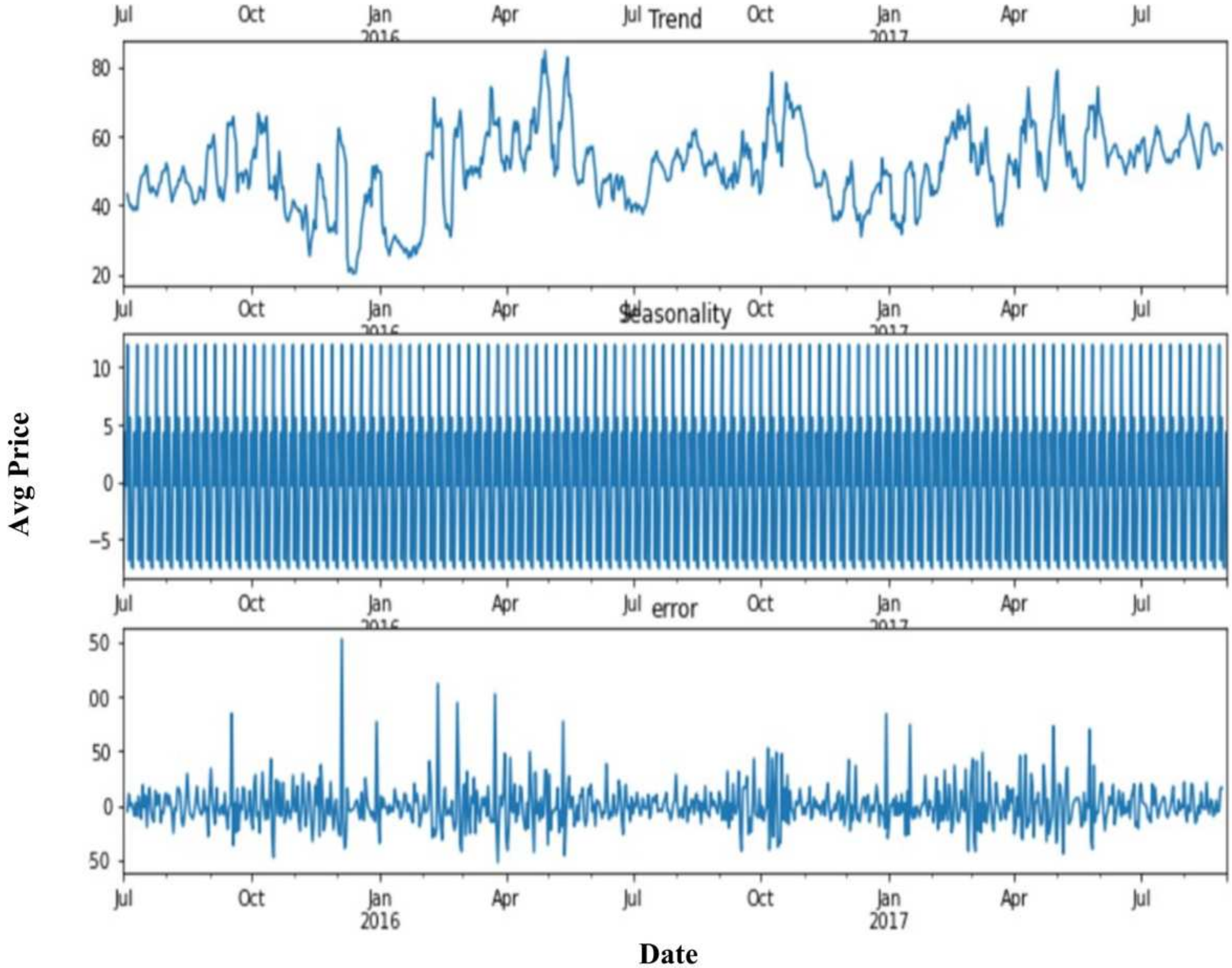


Figure 2

Decomposition plot for Pulses

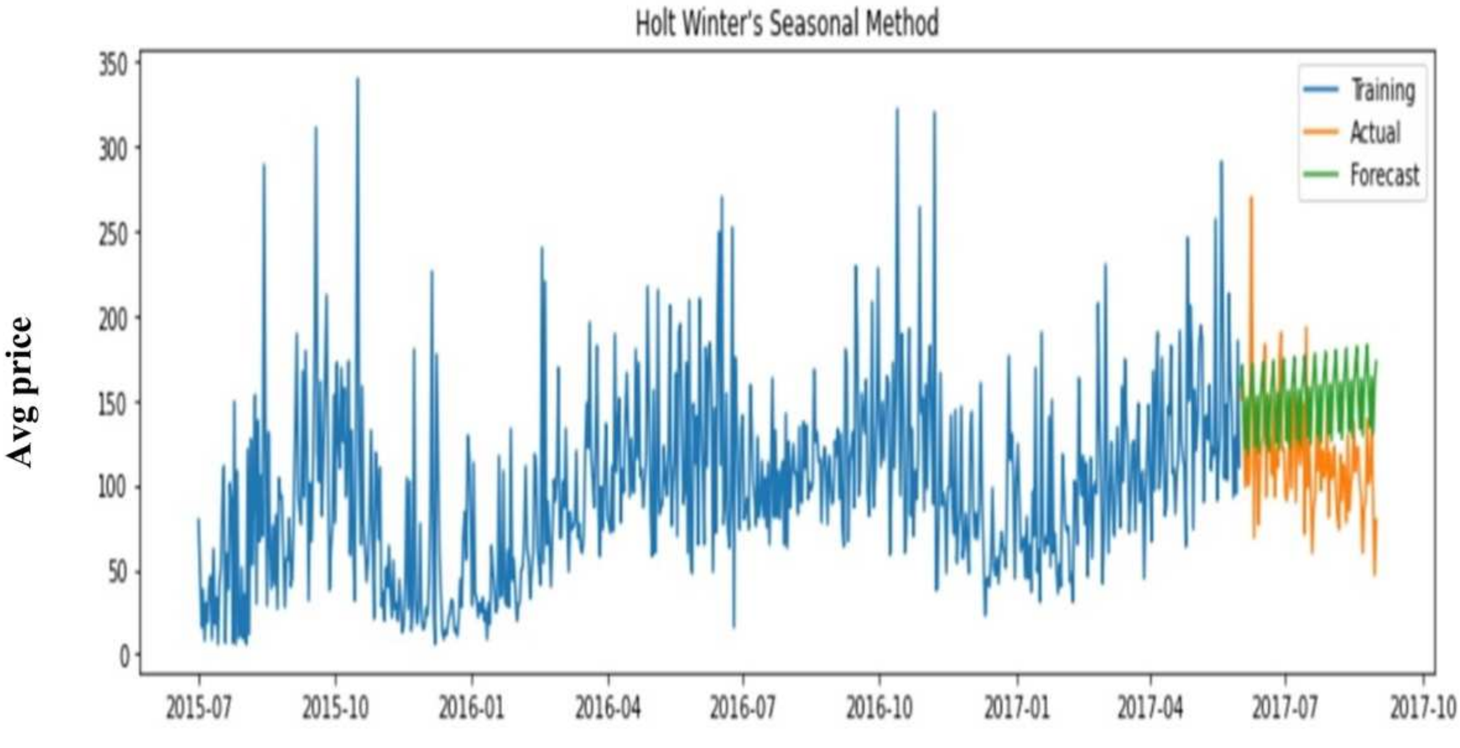
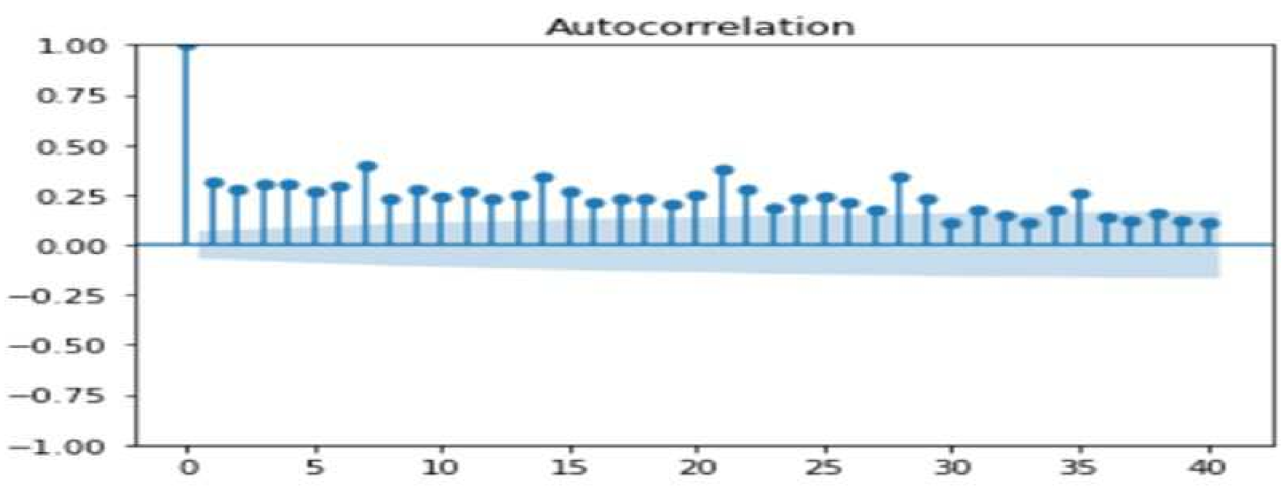


Figure 3

Prediction of ETS model compared to Actual values



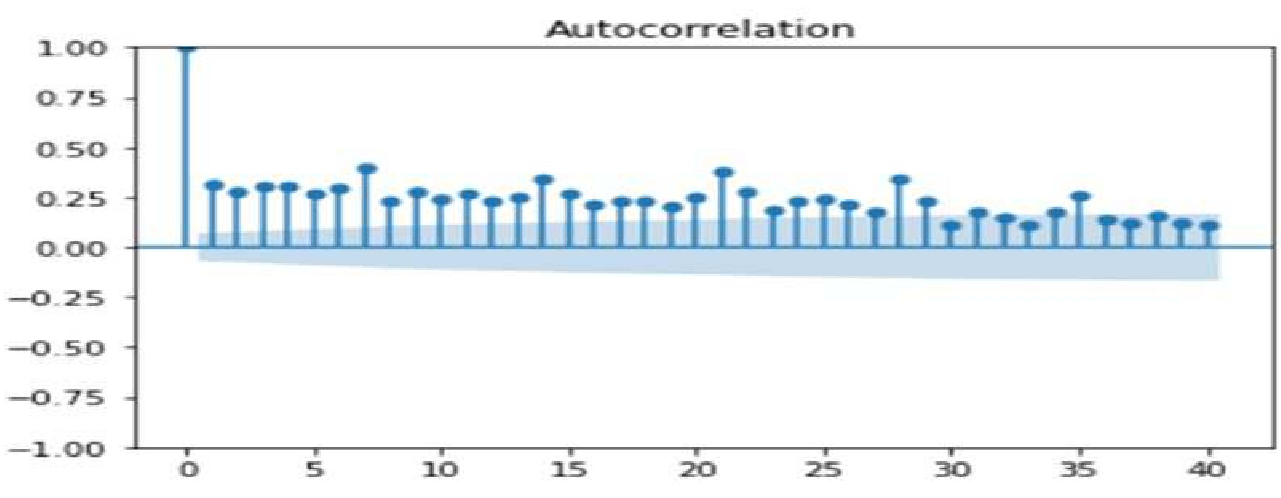
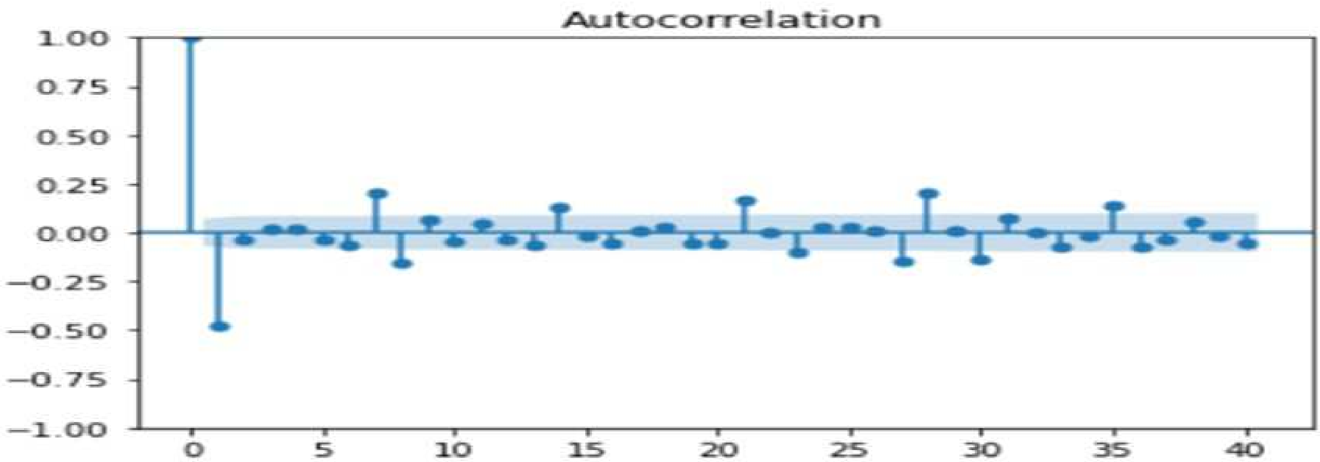


Figure 4

Autocorrelation function before differencing



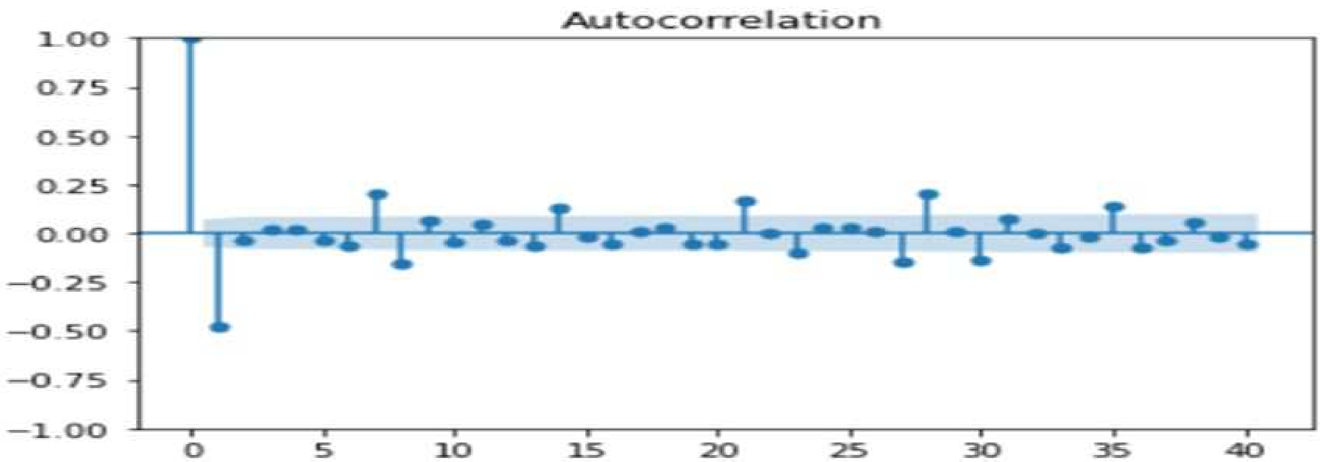


Figure 5

Autocorrelation function after differencing

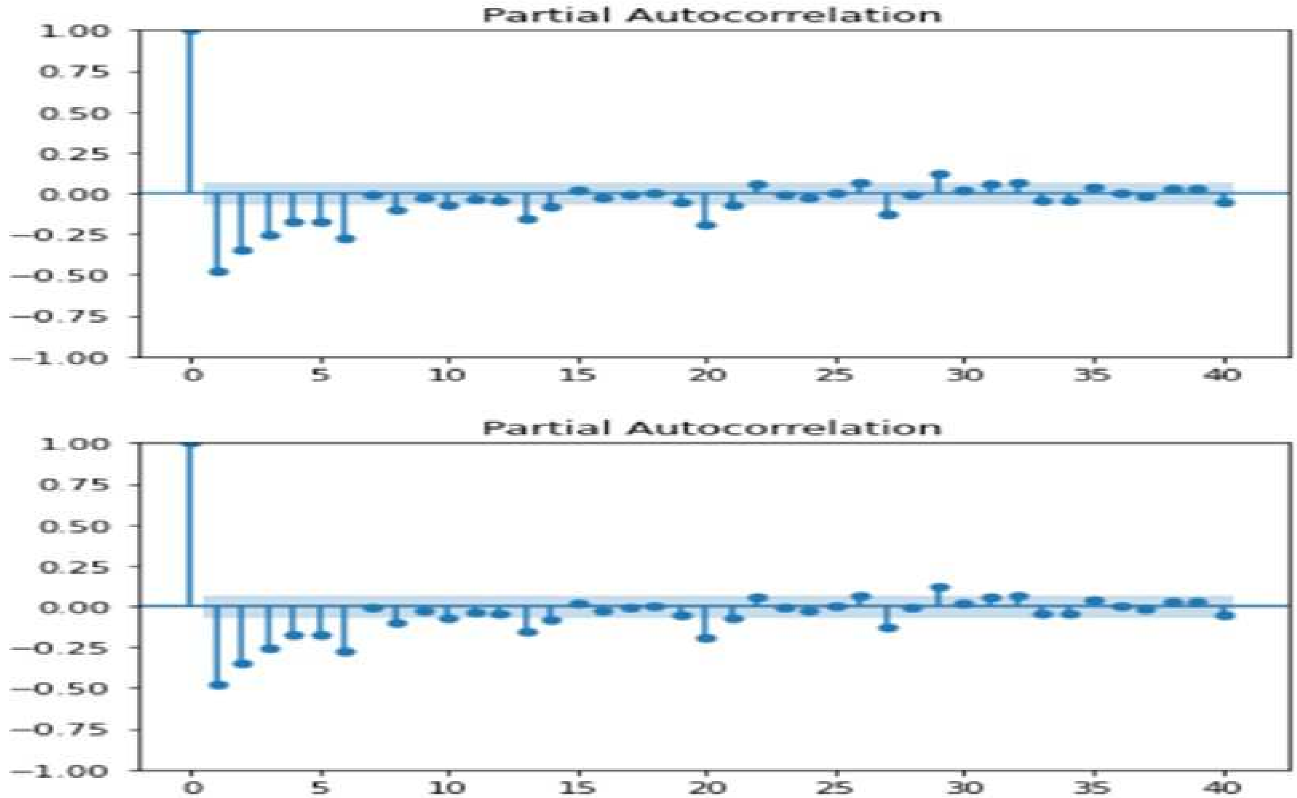


Figure 6

Partial Autocorrelation function

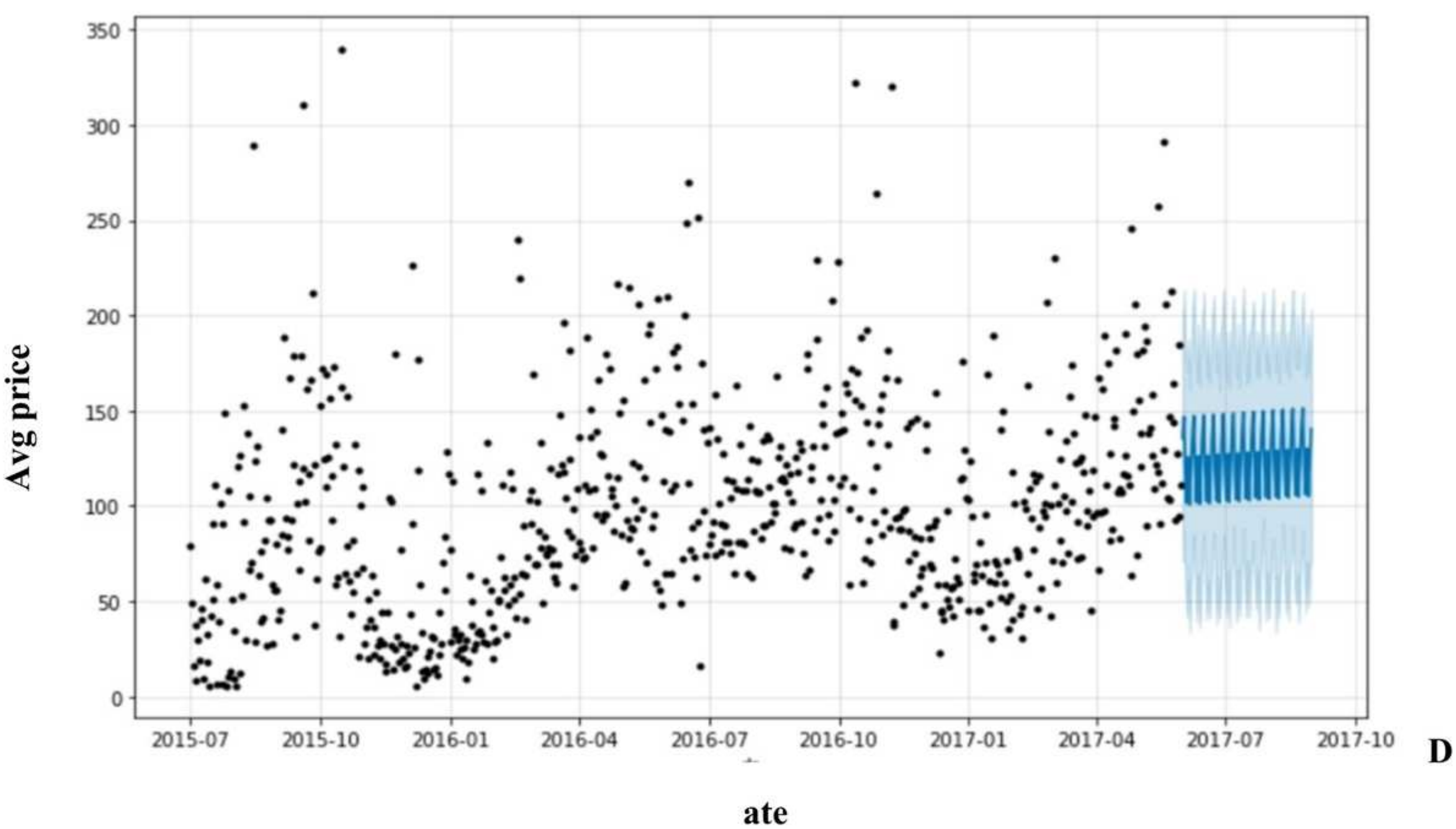


Figure 7

Prediction of Fb prophet as compared to actual values

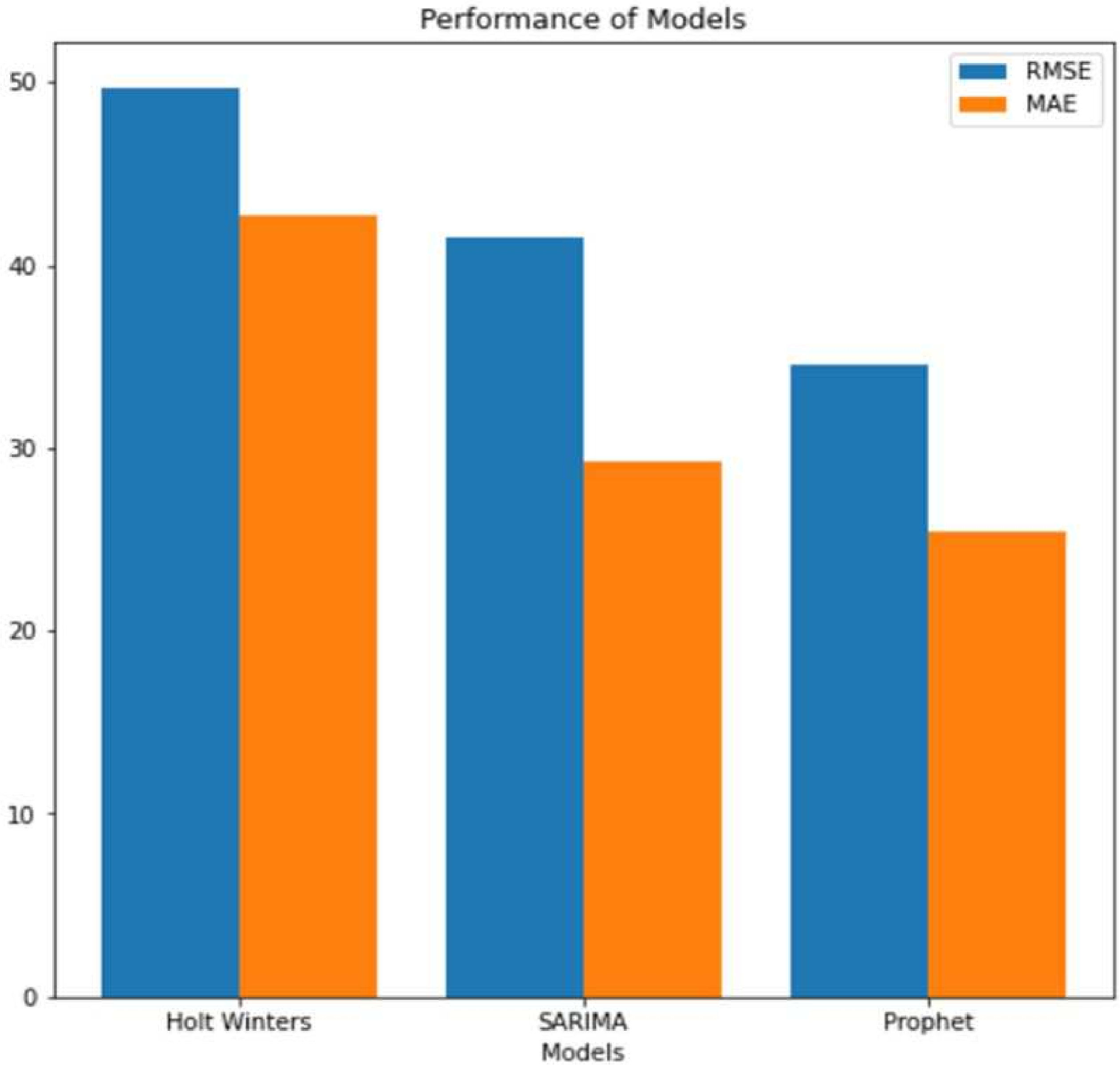


Figure 8

Plot for comparing model performances.

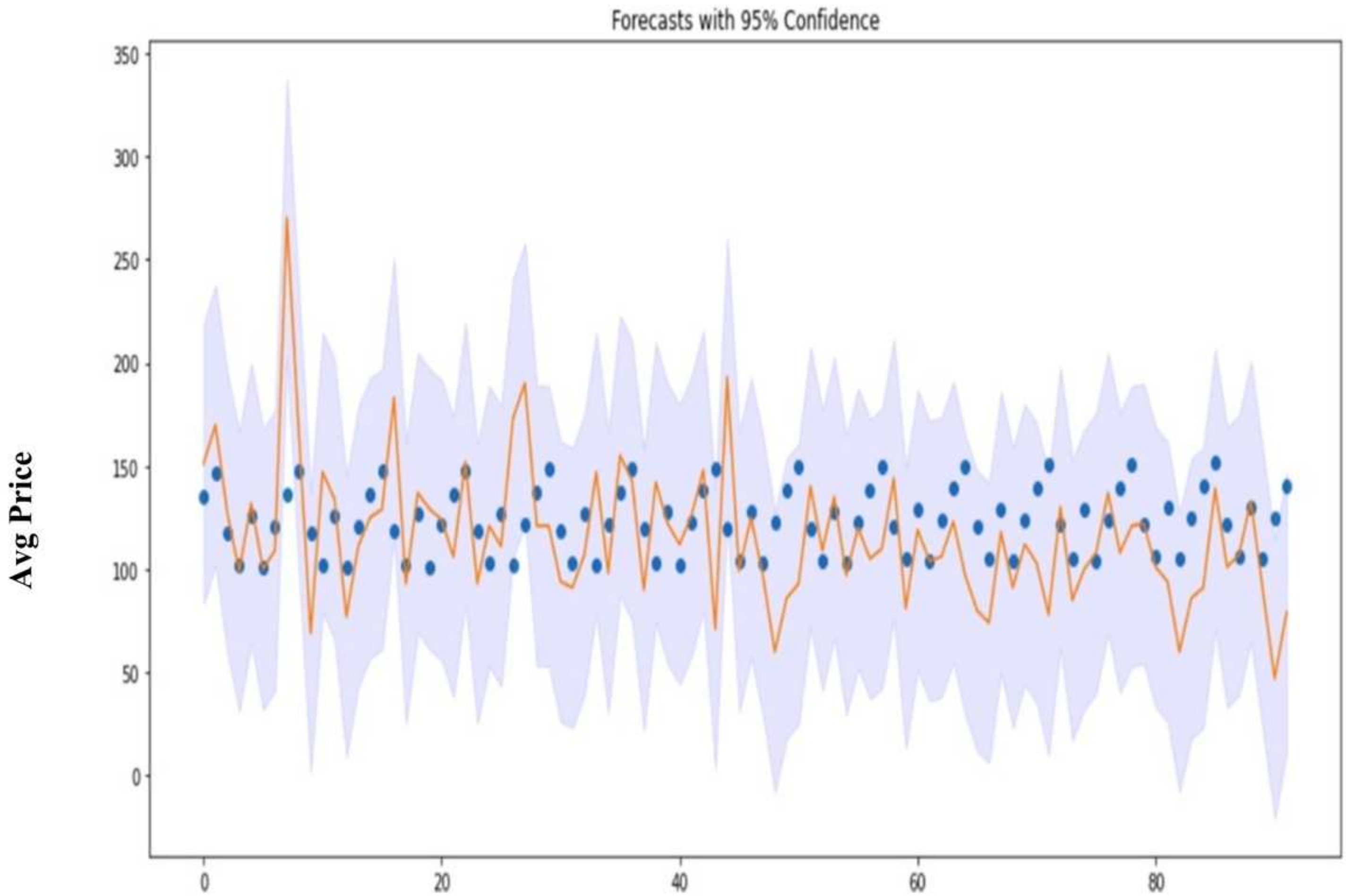


Figure 9

Forecast for cereals and millet dataset

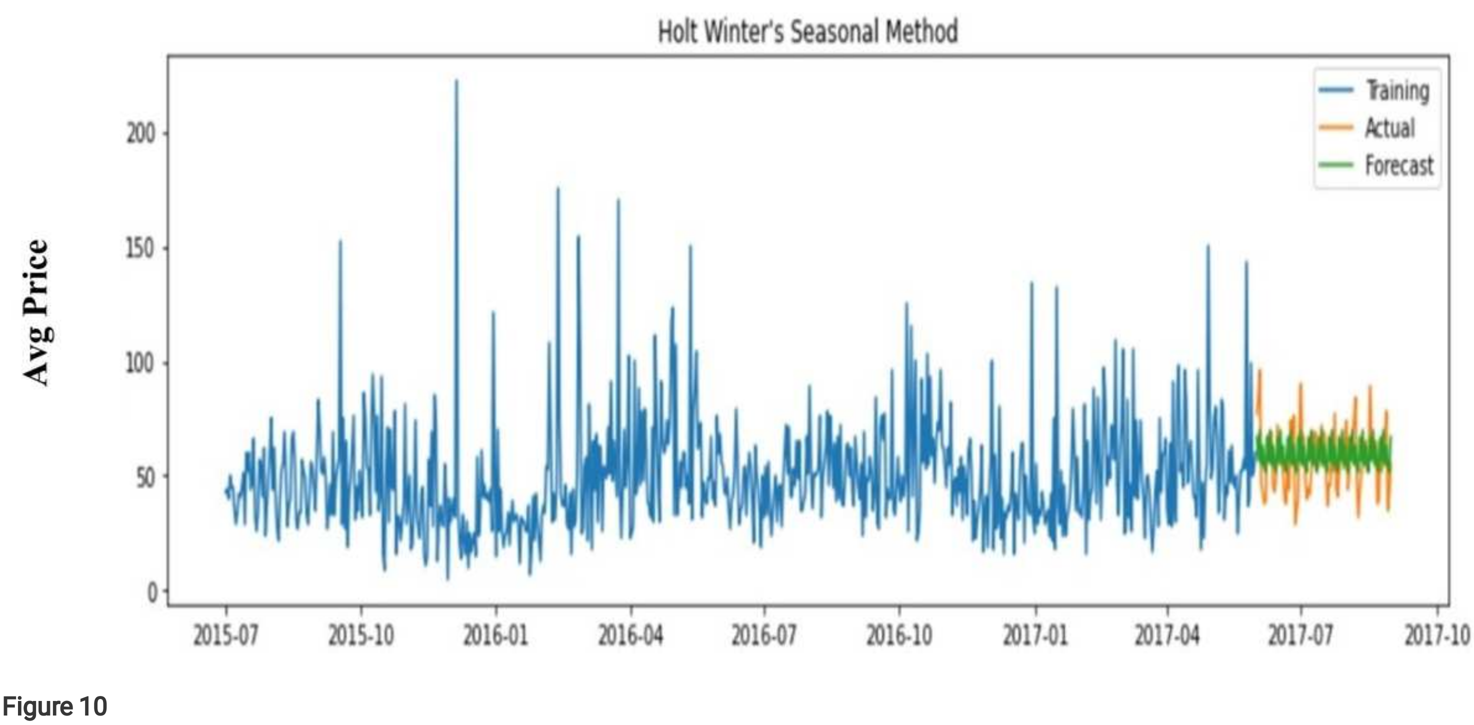


Figure 10

Prediction of ETS model compared to actual values

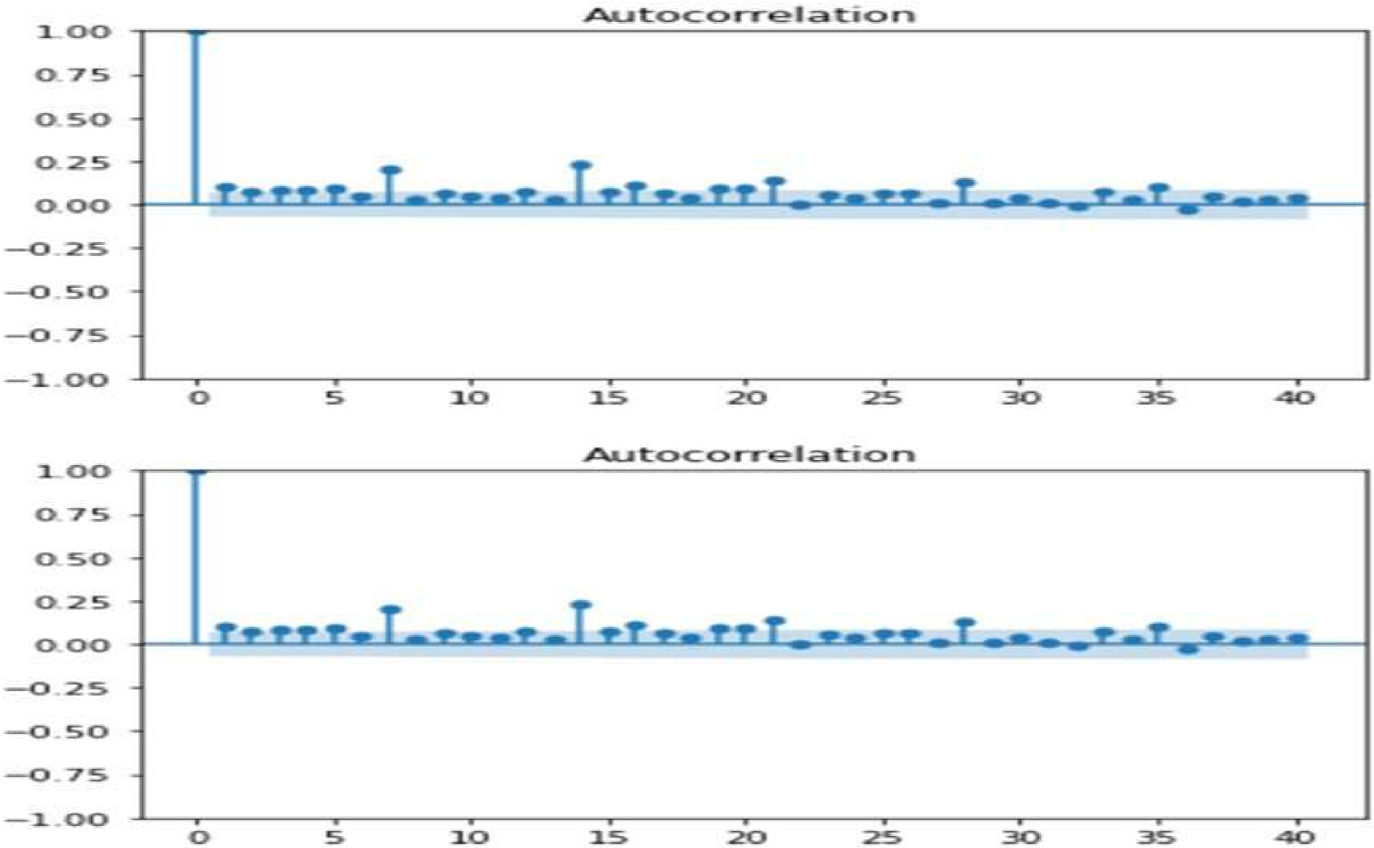


Figure 11

Autocorrelation plot for pulses dataset

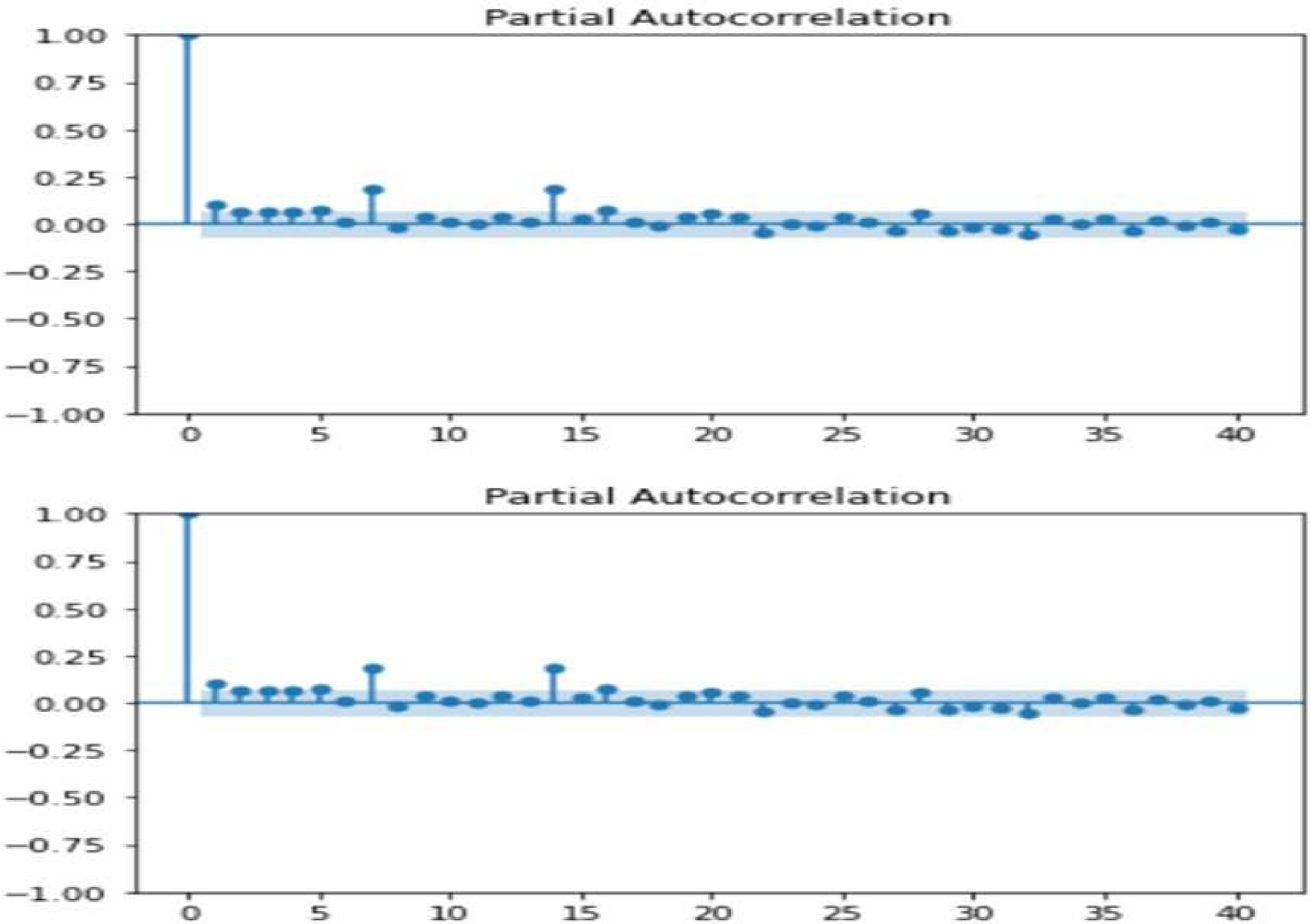


Figure 12

Partial Autocorrelation plot for pulses dataset

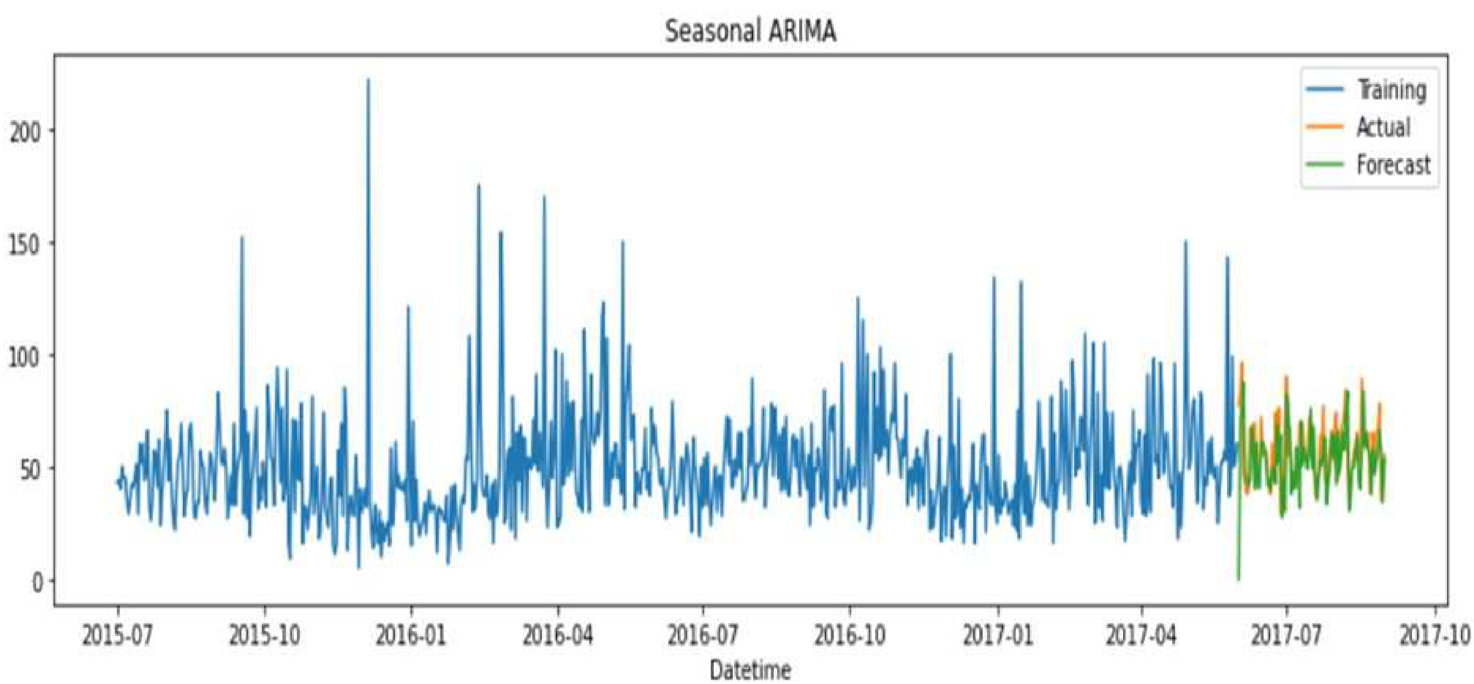


Figure 13

Prediction of ETS model compared to actual value

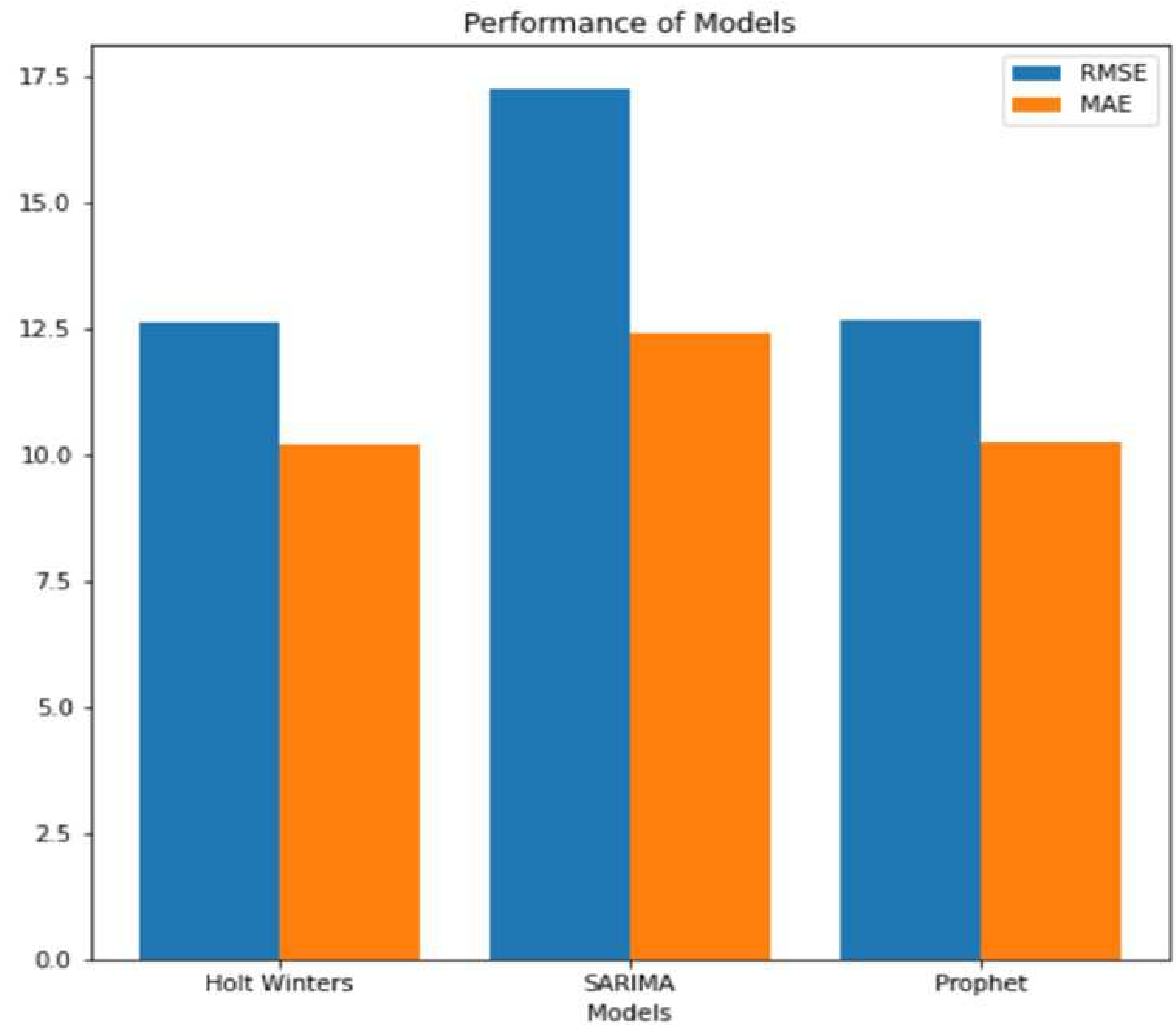


Figure 14

Plot for comparing model performances for pulses dataset

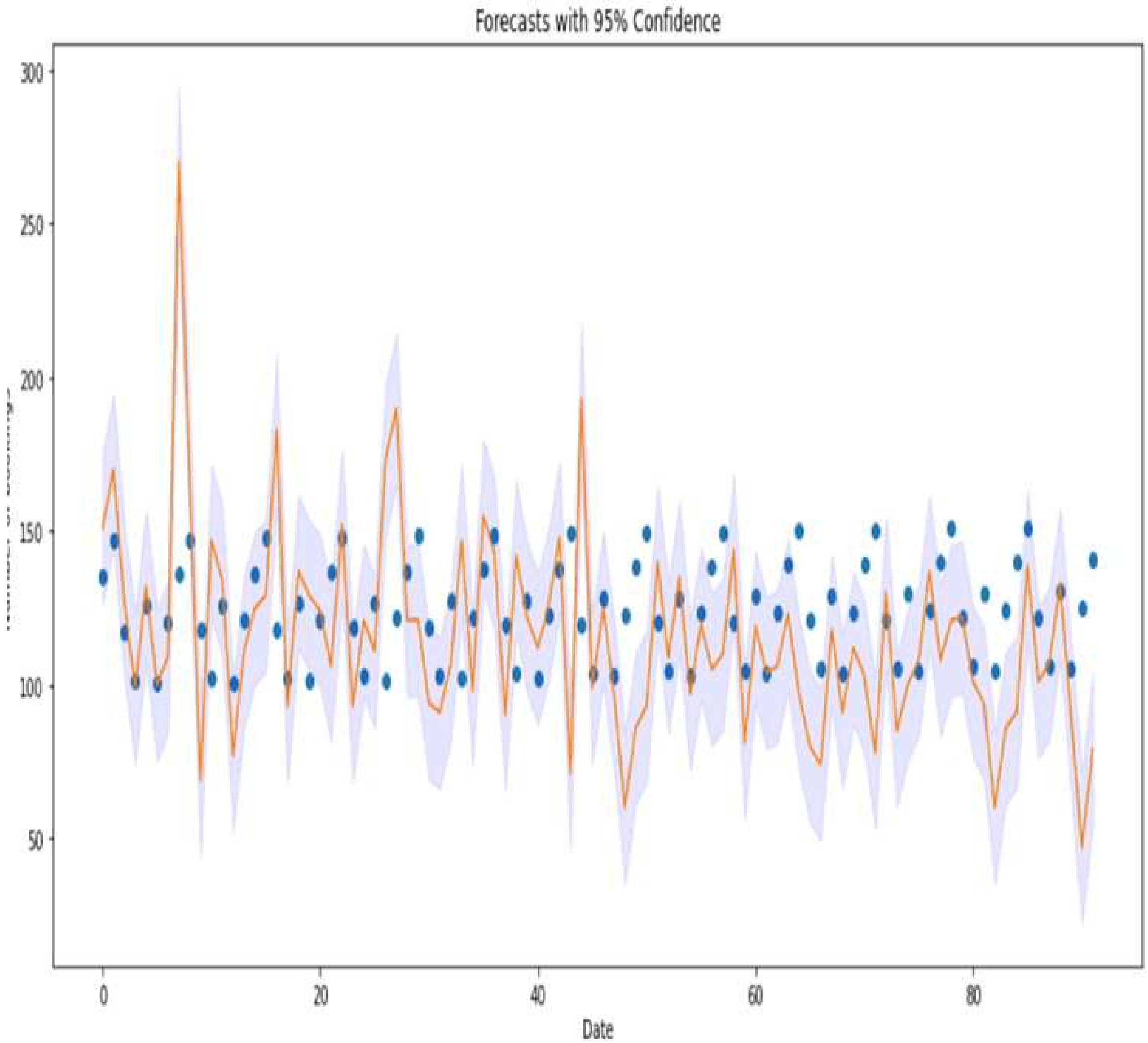


Figure 15

Forecast for Pulses Dataset