Purchase and Analytics for Grace Marketing

[1] Mohammed Farooq Abdulla FM, [2] Tamilselvan V, [3] Harshini V S, [4] Deepthikka R S

[1] Assistant Professor-II, Department of Information Science and Engineering,
Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India.
[2] [3] [4] UG-Scholar, Department of Information Science and Engineering,
Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India.
Corresponding Author Email: [1] farooqabdhulla.ise@kct.ac.in, [2] tamilselvan.18is@kct.ac.in,
[3] harshini.18is@kct.ac.in, [4] deepthikka.18is@kct.ac.in

Abstract— In recent years development of computer systems were able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data is known as machine learning.In this phase sales of different lubricants were predicted using a multivariate time series forecasting algorithm.Previously it showed that the model was accurate in predicting the engine oil sales for a particular time.Using Regressions the accuracy of sales prediction was less (74%) and the models like SVM and Random forest were showing signs of over fitting.The accuracy obtained in the multivariate time series forecasting was good than other algorithms.Time series algorithms are used extensively for forecasting time-based data.In time series ARIMA,SARIMA and SARIMAX are the common methods to forecast time based data.SARIMAX are efficient in forecasting data which has seasonality trends than ARIMA which are good in forecasting data which is stationary in nature Time series methods are extensively used for forecasting time based data.In time series ARIMA,SARIMA and SARIMAX are the common methods to forecast tie based data.ARIMA is the abbreviation of Auto Regressive Integrated Moving Average a model which explains a given time series model based on its lags and other values.SARIMAX is the abbreviation of Seasonal Auto Regressive Integrated Moving Average with Xegeneous variables. ARIMA model is best for forecasting stationary time series data and SARIMAX is used for forecasting values which is seasonal in nature.

Keywords - Time series,ARIMA,SARIMA, SARIMAX,Support Vector Machine,Decision Tree,Random Forest

I. INTRODUCTION

The project Purchase and Analytics for Grace marketing is a consultancy project done for a client,the owner of Grace marketing who is an authorized dealer of TATA AUTOCOMP SYSTEMS LTD, 99 Lubricants and Champol Lubricants and sells engine oils, Gear oils, Coolants, Hydraulics, Power steering, Radiator coolants, Grease, Marine oils, Synthetic and semi synthetic oils, Diesel Exhaust Fluids, Filters and all type of industrial oils.The client is planning to move a part of his sales online and wants some strategies to increase his sales and earn a marginal profit.The project is designed to perform tasks like ordering engine oils,checking the stock periodically,recommendation of engine oils to a particular vehicle according to customers input for better engine life,predicting the volumes sold,stock to be bought every month and customer analytics to find some trends and insights based on the data.

Engine oil lubricates the engine parts for its efficient movement, by reducing the extent of metal contact, friction, and heat generation within the engine. In addition, cleaning and cooling the engine parts, sealing the gaps, and rust prevention are the other functions carried out by the engine oil. Engine oils are prepared by addition of substances, such as viscosity index improvers, antioxidants, metal detergents, and others to the base oil. Engine oil type varies depending upon the engine type.

India automotive lubricants market is estimated to reach $ 9.6 billion by 2022. Increasing demand for automotive lubricants is expected on the account of increasing sales of vehicles and automobiles.

Time series models are optimized for forecasting continuous values, such as product sales, over time. A time series model can predict trends based only on the original dataset that is used to create the model.Since the dataset consists of sales by volume in months we do time series to forecast values for a short period of time.

II. PROPOSED SOLUTION

The project has been divided into five modules.The first module is building a model to predict the price, sales,profit.of the engine oils using machine learning models.The second module website is about the web page with login authentication,and user authentication using HTML and Django.The third module is all about ordering the stocks over the state and paying for the same, and also stock checking for client so that they can have a eye on their stocks and buy the lubricants from the dealer whenever the stocks is empty.The next module is customer data analysis where data analysis is done on the details shared by the customers using analytical tools like Tableau and Python.The last module is recommendation of engine oils for the vehicles to the customers based on the user inputs like Car type,model,fuel type and give adequate addinol and engine oil recommendations.

III. LITERATURE SURVEY

Prapanna Mondal .et al"Study Of Effectiveness Of Time Series Modeling (ARIMA) In Forecasting Stock Prices" the

aim was to study about ARIMA and the stock values across different sectors have been predicted based on the historical datas and factors like GDP etc in 2014.[1]

Xin James He stated-"Time Series vs. SVR Models" compared the weekly crude oil price forecasting and to identify the best forecasting model among various forecasting models, including time series and other machine learning models like SVR had some problems like over fitting in 2018.[2]

Shengwei Wang.et al-Application of seasonal time series model in the precipitation forecast did an comparison of advantages and disadvantages of SARIMA and the regression model with seasonal latent variable in forecasting precipitation .SARIMA model ha s good model fitting degree in decision-making for irrigation.An appropriate model for the time series defined as the Box–Jenkins methods,there are problems about this model because of its failure in forecasting especially if in the past the sequence of time series has changes in 2013[3].

Musa Abubakar Alkali .et al on "The Impact Of Macroeconomic Variables On Real Estate Price Forecasting Modeling In Abuja Nigeria" stated macroeconomic variables play a significant part in developing the real estate price forecasting models. The macroeconomic variables has a strong influence on real estate price forecasting using ARIMA and ARIMAX models in 2019.[4]

Bohdan.Pavlyuchenko-"Machine-Learning Models for Sales Time Series Forecasting" Data from Rossmann Store Sales have been taken into account and models like Extra Tree,ARIMA,Random Forest, Lasso Neural Network were built and the sales of the different stores in the area was predicted in 2018.[5]

Aarati Gangshetty.et al in "Time Series Prediction of Temperature in Pune using Seasonal ARIMA Model" states that the nonlinear relationship between environmental factors and load changes can change the difference between predicted results and actual results.The forecasted values where then compared with the actual values by the ARIMA and SARIMAX model and MAE,MSE,RMSE were calculated in 2021.[6]

N Viswam.et al on"Stock market prediction using time series analysis",this paper he states, he mainly focuses on the accuracy of forecasting stock values for various sectors using ARIMA model which will allow new investors understand the market and make a expertise decision to invest in the stock market without confusions based on the historical data in the market in 2018.[7]

Soham Talukdar on "Appliance Energy prediction using Time Series Forecasting"stated that he made a comparative study of different Machine Learning Algorithms like Random Forest,SARIMAX etc.This experiment gives insights to reduce energy consumption and identify trends and appliances involved. The model is built using time series forecasting which gives a better result than other algorithms in 2018.[8]

Shambulingappa H S on"Crude Oil Price Prediction using Machine Learning".A Study to predict crude oil sales in India.The primary objective is to predict the Crude Oil Sales in India.Time Series models like AR,MA,ARIMA were built where ARIMA performs better due to it less Test MSE value than other models in 2020.[9].

Vasanthakumar .et al on "Short-term traffic flow prediction using seasonal ARIMA model with limited input data" stated that he plans to overcome the above traffic flow prediction by proposing a prediction scheme using Seasonal ARIMA (SARIMA) model for short term prediction of traffic flow using only limited input data in peak hours of Chennai for a period of 24 hrs. A comparison model with historic average and naive method was attempted. The effect of increase in sample size of input data on prediction results was inferred. Short term prediction of traffic flow during morning and evening peak periods was also attempted using both historic and real time data and SARIMA had a better performance with limited input datas in 2015. [10]

Navya Sri Kalli on "Predicting Total Business Sales using Time Series Analysis".The aim of the model was to predict the sales using machine learning models like moving average, simple exponential smoothing, Holt's model, ARIMA, and SARIMAX and the model SARIMAX gave better results in 2020.[11]

Peng Chen .et al"Forecasting crime using ARIMA model".The aim was to predict the crime cases in a city in China for a week using past 50 week crime datasets stated that using ARIMA models and comparing with Simple Exponential smoothing and double exponential smoothing and showed an outcome better than exponential smoothing in 2008.[12]

Nari Sivanandam .et al"Applications of SARIMAX models to forecast daily sales in Food Industry".The aim was to forecast the daily food sales using Machine learning algorithms SARIMA which accounts and considers all the effects due to the demand influencing factors, to forecast the daily sales of perishable foods in a retail store and how to improve traditional SARIMA model with SARIMAX model in 2016.[13]

Catherine McHugh .et al ,"Forecasting Day-ahead Electricity Prices with A SARIMAX Model".The perspective of this paper was to forecast electricity prices with the SARIMAX model considering all external factors and the RMSE was found which was low and predicted the electricity price accurately.[14].

Adhistya Erna Permanasari on "SARIMA (Seasonal ARIMA) Implementation on Time Series to Forecast The Number of Malaria Incidence" stated that SARIMA model had a better accuracy in forecasting Malaria cases in the United states,The model was also be able to represent the historical data without much deviation and had a mediocre Mean Absolute Percentage Error in forecasting in 2013.[15]

IV. METHODOLOGY

The problem statement was to build a website for a lubricant dealer to improve his sales.The project has been divided into five modules like,forecast the engine oil sales, engine oil boxes sold,customer analysis, based on the data

from January 2017 December-2019 collected from the dealer, A Web Page has been built to interact with the users which has the features like login authentication to validate the user login, a products page to view all the products sold by the dealer,a recommendation page to recommend the suitable engine oil and addinol varieties to a particular vehicle based on the users input which is done using HTML,JS,CSS and finally embedding all to a website for deployment..

Data cleaning has been done for the datas collected by the user.The Date attribute has been converted into number format for forecasting. The forecast is been done using models like Support Vector Machine,Decision tree,Random-Forest(62%),Regression(75%).Support Vector Model showed a sign of over fitting or poor forecasting accuracy.The Random Forest Model didn't classify the dataset as a feature and was predicting the same values for any inputs given by the user.

To minimize the over fitting and improve the accuracy, we used Time series forecasting models.The Date attribute has been converted into dd-mm-yyyy format.Dicky Fuller test is been performed to test the dataset is stationery or not stationery, On Analyzing tit was found that it was seasonal.Now Auto correlation and Partial Auto correlation graph has been plotted and p,d,q value is found .

Identification of an AR model is often best done with the PACF.For an AR model, the theoretical PACF "shuts off" past the order of the model. The phrase "shuts off" means that in theory the partial auto correlations are equal to 0 beyond that point. Put another way, the number of non-zero partial auto correlations gives the order of the AR model. By the "order of the model" we mean the most extreme lag of x that is used as a predictor.

Identification of an MA model is often best done with the ACF rather than the PACF.For an MA model, the theoretical PACF does not shut off, but instead tapers toward 0 in some manner. A clearer pattern for an MA model is in the ACF. The ACF will have non-zero auto correlations only at lags involved in the model. p AR model lags,d difference and q MA lags.

Next the closest and accurate p,d,q value has been found and it was p=1,q=1 and d=3.Now a graph was plotted between the original value (blue line) and forecasted value (orange line).ARIMA model was to poor in forecasting the values.

The efficient model here was SARIMA and SARIMAX comparatively showed better results and had a good forecasting accuracy.The Mean Squared Error and Root Mean Squared Error and in the table below.
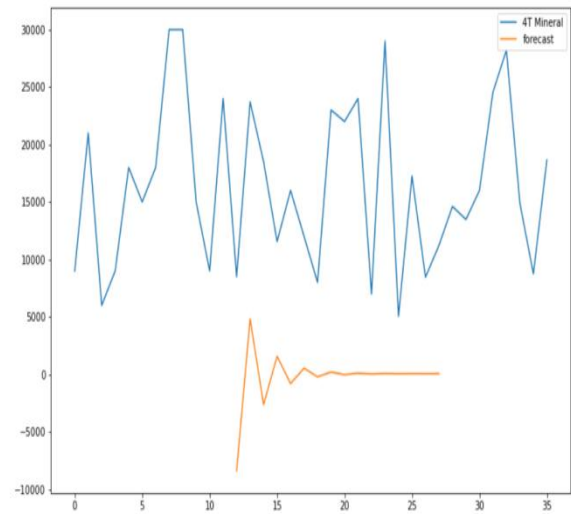

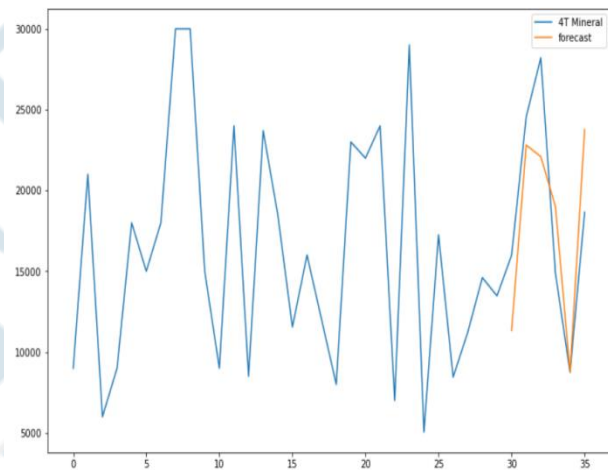**Figure 1:**ARIMA model predictions for engine oil.

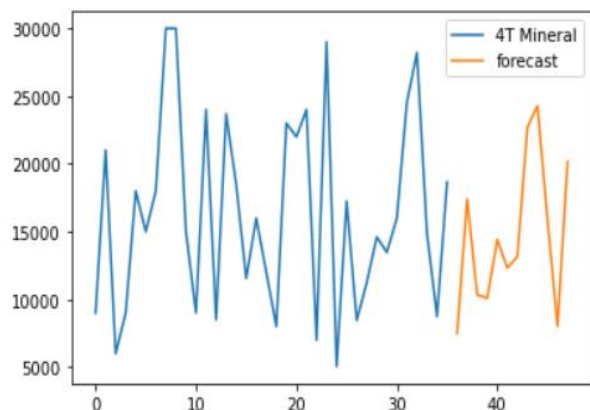**Figure 2:**SARIMAX model prediction for engine oil

Figure 3:SARIMAX predictions for future dates

Which shows that time series models have been performing better than classical machine learning models like Polynomial Regression,Random Forest,Support Vector Machine models.

MODEL	MEAN SQUARED ERROR	ROOT MEAN SQUARED ERROR
POLYNOMIAL REGRESSION	11618420.421	3408.58
RANDOM FOREST	2600061.844	1638.28
SUPPORT VECTOR MACHINE	20732994.023	4553.34
TIME SERIES(1,1,3)	14647.28	121.02

V. CONCLUSION

It has been observed that uni-variate and multivariate time series forecasting has been better in forecasting the values for a period of time than the classical machine learning models like Support Vector Machine,Random forest,Decision Tree,and Polynomial regression Among the uni variate and multivariate time series forecasting ARIMA was effective in forecasting the stationery datas,which does not any seasonality and SARIMA or SARIMAX model has been observed to be performing good with data having trends and seasonality.

REFERENCES

[1]. Prapanna Mondal,Labani Shit and Saptarsi Goswami-"Study Of Effectiveness Of Time Series Modeling (ARIMA) In Forecasting Stock Prices", International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.4, No.2, April 2014:14-29 pp 13-24.

[2]. Xin James He-"Crude Oil Price Prediction Time Series vs SVR Models-"Journal of International Technology and Information Management,Volume 27 Issue 2 January 2018.pp-25-42.

[3]. ShengweiWang,Juan Feng,Gang Liu, Application Of Seasonal time series model in the precipitation forecast" on Mathematical and Computer modeling,Volume 58,Issue 3-4 on August 2013 pp 68-73.

[4]. Abubakar Alkali, Ibrahim Atan Bin Sipan, Muhammad Najib Razali-"The Impact Of Macroeconomic Variables On Real Estate Price Forecasting Modelling In Abuja Nigeria "International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-5C, May 2019,pp-67-72.

[5]. Bohdan M. Pavlyuchenko-"Machine-Learning Models for Sales Time Series Forecasting"from 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2018,Data 2019, 4, 15,pp 1-11.

[6]. Aarati Gangshetty,Gurpreet Kaur,Uttam Sitaram Malunje-"Time Series Prediction of Temperature in Pune using Seasonal ARIMA Model" fromInternational Journal of Engineering Research & Technology (IJERT) Vol. 10 Issue 11, November-2021,pp 235-240.

[7]. N Viswam and G Satyanarayana Reddy-"Stock market prediction using time series analysis",International Journal of Statistics and Applied Mathematics on 2018 ISSN: 2456-1452 in December 2017,pp 465-469.

[8]. Soham Talukdar-"Application energy prediction using time series forecasting:A Comparative study of different Machine learning algorithms",International trends of Advanced Trends in Computer Applications Volume 08 Number 01,February 2021,pp 7-18.

[9]. Shambulingappa-"Crude Oil Price Forecasting Using Machine Learning" International Journal of Advanced Scientific Innovation Volume 01 Issue 01 2020 ISSN 2582-8436 pp 6-11.

[10]. S. Vasantha Kumar,Lelitha Vanajakshi-"Short-term traffic flow prediction using seasonal ARIMA model with limited input data",in European Transport Research Review,Volume 7,Article number 21,June 2015,pp 1-9 .

[11]. Navya Sri Kalli , Harsha Teja Pullagura -"Predicting Total Business Sales using Time Series Analysis" International Journal of Scientific Research in Computer Science,Engineering and Information Technology ISSN:2456-3307 ,Volume 6, Issue 4 pp-475-482.

[12]. Peng Chen,HongYong Yuan,Xueming Shu-"Forecasting crime using ARIMA model",published on November 2008 .ISBN:978-0-7695-3305-6 in Fifth International Conference on Fuzzy Systems and Knowledge Discovery in November 2008,pp:627-630.

[13]. Nari Sivanandam Arunraj,Diane Ahrens,Michael Fernandes"Applications of SARIMAX models to forecast daily sales in Food Industry",International Journal of Operations Research and Information Systems ,April 2016,Volume 7,Issue 2,pp 1-21.

[14]. Catherine McHugh,Sonya Coleman,Dermott Kerr,Daniel McGlynn ,"Forecasting Day-ahead Electricity Prices with SARIMAX,Vol 1,International Conference on Time Series and Forecasting,Volume l, 27 Sep 2019,pp-310-320.

[15]. Adhistya Erna Permanasari,Indriana Hidayah, Isna Alfi Buston,"SARIMA (Seasonal ARIMA) Implementation on Time Series to Forecast The Number of Malaria Incidence", on International Conference on Information Technology and Electrical Engineering,ICITEE 2013,pp 203-207.