RESEARCH ARTICLE

International Journal of Computer Communication and Informatics

DOI: 10.34256/ijcci2413

Time Series Analysis for Tractor Sales using SARIMAX and Deep Learning Models

Pooja Polina [a], Swathi Ganesan [a,*], Lakmali Karunarathne[a], Nalinda Somasiri [a]

[a] Department of Computer & Data Science, York St John University, London, United Kingdom

* Corresponding Author: s.ganesan@yorksj.ac.uk

Received: 26-02-2024, Revised: 08-05-2024, Accepted: 16-05-2024, Published: 25-05-2024

Abstract: Time series forecasting is known for playing vital role in many industries to make important decisions and strategies. This study concentrates on providing accurate insights that can help manufactures and stakeholders of agriculture machinery industry on future sales of tractors by applying both traditional and deep learning models like SARIMAX which is extension of SARIMA and deep learning models. Research starts by observing history data which include years of tractor sales then preprocess the data to find its quality and stationarity further applying SARIMAX model to find trends and seasons and cycles in the data and this model is evaluated by famous metrics like Root Mean Squared Error (RMSE).Deep learning models like Gated Recurrent Unit (GRU), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional LSTM, CNN LSTM Encoder Decoder, Convolutional Neural networks (CNN). They can help in enhancing the forecasting accuracy by handling all the non-linear relationships and their dependencies in the timeseries and this study will provide comparative analysis of deep learning models and SARIMAX model. Where SARIMAX outperformed the deep learning models with RMSE score 0.01 and provide forecast of next year's tractor sales using SARIMAX model from the study and use q-q plot, residual plots and ACF and PACF graphs to make sure forecast was done accurately.

Keywords: Time series forecasting, ARIMA model, SARIMAX model, Deep learning models

1. Introduction

Machinery like tractors plays very prominent roles in the agricultural which can show very much impact on the efficiency in the agriculture, tractors are backbone for modern farming they can help farmers in doing tasks like ploughing, planting, and harvesting the crops this can directly affect the global production of the food. Tractors are very important or the primary equipment in the agriculture industry. For the manufacturers to meet the demand to satisfy the farmers with optimizing their production. So, the manufacturers and distributors of the agricultural machinery industry coming up with different strategies for marketing, inventory, and production.

Sales forecasting for the tractors can be hard and challenging in the real world because it can include all the complexities like Seasonal variations, Non-Linear relationships, data quality, data imbalance in the agricultural sector. There can many more external factors that can affect the sales like latest farming trends and advancements, changing governments and their policies all these make sales forecasting more complex and challenging. In-order to help manufactures to make important decisions on supply chain, production schedules and campaigns on marketing, both over and under stocking can be avoided, allocation of the resources can be done in better way, risks in both the financial and management sectors can be reduced; to achieve this, accurate prediction is very essential.

Very motivation of this project to use both traditional methods and deep learning methods to forecast the sales and understand the challenges while doing this. Traditional models like ARIMA and SARIMA are used to provide a strong understanding of time-dependent and seasonal patterns, trends, and cycles in the data. To capture complex elements like non-linear relationships in the data, deep-learning models were approached. Both types of models are further compared.

2. Experimental procedure

The time dependent data can be affected by different environments causing different trends in the data and spikes in the data at times. This kind of fuzzy environments [1] must be taken into consideration from this paper that can learn that sales forecasting is important strategy of any business or industry which can help to track the intense competition, constant fluctuations also help to manage the sourcing, production, and distributions these decisions are taken based on the comprehensive sales forecast. Better sales forecast can help different industries like stock market, E-commerce and many other but in the case of agricultural industries, the forecasting can help agricultural industry to maintain inventories, financial planning, marketing strategies and other risks can be managed with accurate forecasting of sales. Both traditional models like regressive models, exponential models, and Box-Jenkins models can be used for forecasting. These models can be applied to all the mentioned industries, taking into consideration production, consumption, and traffic flow. The study showed that traditional models do not provide the best results when the industry has unique characteristics like seasonality, trends, and cycles, making forecasting challenging.

So, analysing the sales trends is important for a company to achieve all the above things. Analysing sales data is just predicting the future sales using the historical sales data but it is rarely sufficient because usually agricultural products like tractor are used by different kinds of people at different locations and for different purposes in different times making the forecasting complex [2]. The major challenges faced while forecasting sales are also presented in this paper.

Forecasting can be affected when new products are introduced in the industry. This introduction brings unique data that does not match the historical data. These discrepancies can result in different market trends, which may initially be upward and then gradually downward. This paper identifies traditional models like exponential smoothing (ES) has the draw back while directly applying to the data because data can be very large, and ES can't take the casual factors into the consideration and cannot understand the trends in the data, and it is increased trends are underestimated and the decreased trends are over estimated. Other regressive method is introduced Holts Linear Method this method is introduced to cut the drawback of the ES model where here there is a parameter which can estimate the slope of the trend, this can reduce the over estimation and under estimation but cannot remove entire problem of not including casual factors. Casual factors are usually product prices, salesmen's quotes, and seasonality factors. For better forecasting of sales, other models besides regression need to be used. These models should handle non-stationary data and provide robust and dynamic sales forecasts. Then come the models like ARIMA and SARIMA can the solve the problem that cannot be solved by the regressive models.

There is significant evidence that seasonality will affect the production and sales of the agricultural industry [3], so efficient forecasting can these industries reduce the risk. This paper talks about the challenges that was faced by the manufacturers like adapting to the seasonality of the market, efficient strategies that is required when market was going through the low demand period. Also shed light on how the macro environmental factors like economic, demographic, regulatory can influence the decisions made by the organization. There should be good sales and operational planning which can be used as the tool for making strategies that is aligned with the market demand. By using this planning interdepartmental communication can be forecasted. The resources, especially in the agricultural industry, can be optimized by forecasting. This includes the workforce and manufacturing capacity.

This prediction can help adjusting the staff, production schedules which intern reduces the cost that is associated with the workforce. By forecasting overall production efficiency can be improved as well at the peak demand times companies can make sure their equipment was working efficiently which can increase the production without any problem. Forecasting can allow the company to adapt the market trend. Can analyse the underlying patterns and the changing customer behaviour can give what should the company be prepared for. The correlation between productivity and market demand can be seen from this study. Forecasting can catch this correlation, which can help companies make decisions, accordingly, maintaining the balance between productivity and market demand. To determine if the dataset was better using traditional models, it must be found if the dataset has trend, seasonality, and remainder. This can be visualized by decomposing the time series into its components. This helps to understand the underlying patterns, forecasting accuracy can be improved, early anomaly detection, modelling can be simplified, data can be interpreted well, better model selection, performance evaluation and decision making will be dependent on this. Very useful for overall understanding of the data.

Several studies on sales forecasting provided a number of solutions [4] over the years. They can be categorized into two types they are computing techniques and statistical techniques. The statistical techniques include Auto regression models (AR), Integration model (I), and Moving average model (MA), which all together are called ARIMA. Another name for the ARIMA model is the Box-Jenkins model. The ARIMA model is known for its accurate forecasting, especially with time series data. The ARIMA model is essential when the data is uncertain, and there is no prior knowledge of the underlying model or any relationships between the data. ARIMA model can be reliant on the time series analysis and specifications of the model so this model forecasting will be based on the patterns that are existing also various statistical assumptions that were made. Whereas ANN can be working like the human brain this can the learn about the data without any prior model specifications. It uses the non- linear approach. ARIMA models can be very fast compared to the deep learning models because then can use previous patterns and can also be limited to the model specifications and dependencies. On the other hand, ANN doesn't require any predefine specifications or any model dependencies but deep learning models like AA require relatively very large data sets for them to learn about the data where it can learn about the nonlinear patterns which can be difficult for the ARIMA model. It is suggested the for the stock market where the fluctuations are rapid it was said that deep learning models can be outperforming the traditional models.

ARIMA model is more dependent on the previous or historical data in the series or the previous errors to forecast [5]. In this paper prediction of the power is done here author used the short term and mid-term analysis, here wind energy is diverted into the electricity market which is crucial. Here traditional models like ARIMA, Random Forest (RF), regression and classification trees all these models were compared and worked on the integrated models like hybrid of ARIMA and random forest and ARIMA and BCART was compared with their accuracies. In the initial stage from the paper, it can be seen that single stage models like ARIMA, RF, and BCART were performing reasonably well. As the data showed the nonlinear behaviour when power generation is affected by the environmental variables. ARIMA model showed some limitations to address those limitations author performed integrated ARIMA models with the machine learning techniques to capture both linear and nonlinear behaviour of the data. In this paper finally hybrid model ARIMA-RF outperformed the other models. This shows that ARIMA alone cannot handle the complexity of the data. So, it requires support from other models to handle the seasonality.

Studies showed trained ARIMA model can give best guidelines for steady business to make decisions regarding marketing, inventory maintenance and more but ARIMA model has the draw back if the data has seasonality component which means there is a certain pattern where sales are affected by the external factors like holidays for the E-commerce, or war for the stock market, weather for the agricultural machinery industry [6]. From this study comparing the forecasting sales using different models like both traditional methods and modern forecasting methods like high level machine level techniques. Also evaluating these models using the different metrics like Root Mean Squared Error (RMSE) this was taken as the measure to compare the model's accuracy. Study shows that performing the exploratory analysis on the data is important step to understand the patterns that were underlying, trends, cycles. Traditional models like Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) these models were used a lot for the time series analysis because of their simplicity. On the other hand, modern techniques like Long Term Memory networks (LSTM) and a model developed by Facebook called FB-Prophet are used. These deep learning models are able to understand the complex patterns in the data. But in this study SARIMAX outperformed because it was able to understand the fast-moving seasonal patterns. FB-Prophet also showed equal performance on the other hand LSTM showed it the emerging potential can't outperform the SARIMA.

This can be simply fixed by the different version of ARIMA called SARIMA. SARIMA model has the function called SARIMAX (). This function can understand the seasonality component and can give the best sales forecast for non-stationary data. Sometimes classical models like ARIMA or SARIMA can outperform the deep learning models [7] like LSTM when the data was univariant and linear because these traditional models are very simple. Deep learning models can be sensitive to non-stationary data that is data which has the trend and seasonality. In the preprocessing, differencing or log transformation can easily remove this. Data loss may sometimes result from these methods. Afterward, deep learning models or hybrid models can be applied. Auto corelation functions (ACF), Partial Auto corelation functions (PACF) are used to show the relationship between the lags of the time series. [8] here the author tried to use the SARIMA to understand the characters of the traffic flow. For the monthly based time series data, they used SARIMA to find the cases of malaria in America. This can help the governments and pharma to understand the length of disease and make policies and measures that are needed to be taken.

3. Methods

3.1 Aim

The important aim of this project is to develop models which can help forecast the tractor sales with utmost confidence and help the agricultural industry. To be more precise regarding the aim there are steps which the projects aim for like data collection where historical data od tractor sales will be collected. The data is cleaned and balanced, followed by exploratory data analysis to identify the seasons and trends. Visualization of these trends is also performed.

The next step involves model selection, where both traditional methods, such as ARIMA and SARIMA, and deep learning methods, like LSTM, CNN, and RNN, are applied. Seasonal and complex dependencies are explored, and a comparison of both model [9] types is conducted using evaluation metrics. The best model is then identified through this comparison. Finally, the selected model is used to forecast tractor sales, aiding the agricultural machinery industry in making informed import decisions.

3.2 Hypothesis

3.2.1 Hypothesis for stationary data

Null Hypothesis: data of the tractors is stationary data, so it does not require any differential methods.

Alternative Hypothesis: Differential methods are required to make data stationary so ARIMA model can be applied.

3.2 2. Hypothesis for seasonality

- Null Hypothesis: there is no seasonality in the tractor sales data.
- Alternate Hypothesis: there is seasonality in the tractor sales data.

3.2.3. Hypothesis for effectiveness of model

Null Hypothesis: both AIRIMA and deep learning models have same effectiveness.

Alternative Hypothesis: The data that has been selected can result in ARIMA or Deep Learning models outperforming one another. When specific datasets are chosen, the performance of ARIMA or Deep Learning models can vary, with one model potentially surpassing the other based on the characteristics and intricacies of the dataset.

3.3 Approach

3.3.1 Data Collection and Exploration

The data is collected, and its characteristics are explored at this stage. The Time series data set consists of 144 columns and two rows. Typically, there are 12 columns for each year with the corresponding month's sales data. After the data is attained, it will be checked for missing values, outliers, trend, seasonality, and cycles.

Missing values

Checking if the data set has any missing values then asses if they are impacting the data and by implementing appropriate strategies like dropping the columns or rows that has missing values, but this might lead to data leak. The different functions such as "isnull()" and "fillna()" can be used in Python. The missing values can be filled with the mean or mode by their nature.

Outliers

In data analysis detecting the outliers is crucial part. The methods such as visualization techniques like boxplots or scatter plots are used to handle outliers. Statistical approaches such as Z-score and IQR (Interquartile Range) are also employed. Advanced machine learning techniques, like Isolation Forest, are utilized as well. Outliers can be managed by removing them, transforming them, or by imputation. Any of these methods should be applied with careful consideration of their characteristics and their effects on the data.

Checking For Trend, Seasonality and Cycle

Time series analysis is majorly used to forecast company sales, demand for the product, to know the trends in the stock market and agricultural industries and many more. The basic approach for the time series analysis is to decompose the original time series into different components which are independent. Time series can contain four usual components- trends, seasonality, cycle and irregular remainder.

Trend: the direction in which time series is heading either up or down as shown in Figure 1.

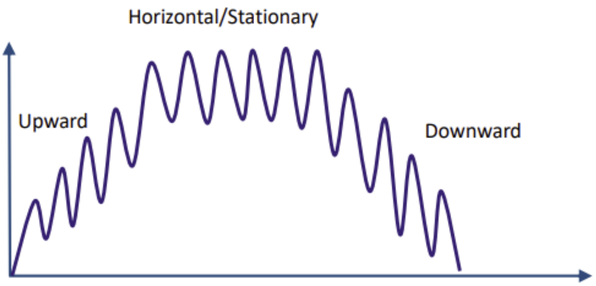


Figure 1. Trend

Seasonality: patterns that happen every month or quarter or a year or more as shown in Figure 2.

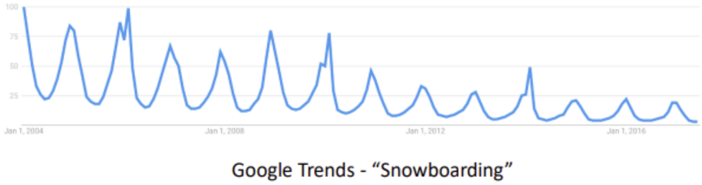


Figure 2. Seasonality

Cycle: patterns that happen long term which usually occur every 5 or 10 years as shown in Figure 3.

Figure 3. Cycle

Irregular remainder: The noise that are left after all the extraction of the components is the irregular remainder.

Decomposition

To determine if the dataset had a trend, seasonality, cycle, or irregular remainder, visualization can be done by decomposing the timeseries into its components. This approach allows for the analysis of these elements within the dataset.

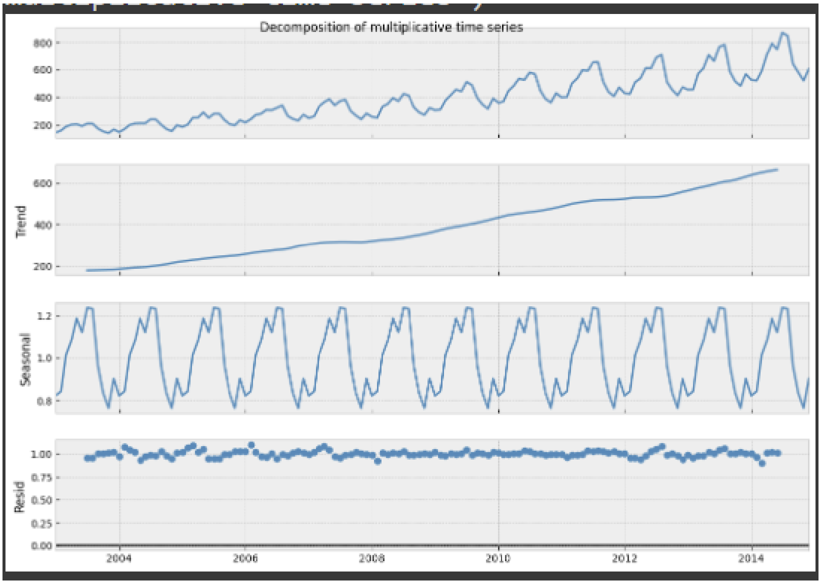


Figure 4. Example of multiplicative decomposition

This helps to understand the underlying patterns, forecasting accuracy can be improved, early anomaly detection, modelling can be simplified, data can be interpreted well, and better model selection, performance evaluation and decision-making will be dependent on this. Very useful for the overall understanding of the data. Here in this project, the used Multiplication decomposition (Figure 4) this break down the time series into trend, seasonality, and remainder

Multiplication Decomposition Equation

$Y(t) = T(t) \times S(t) \times R(t)$ [8]

Where:

$Y(t)$ is the observed value at time t.

$T(t)$ is the trend component at time t.

$S(t)$ is the seasonality component at time t.

$R(t)$ is the remainder at time t.

3.3.2 Stationarity

To use Traditional model effectively, the data must be stationary. dataset is said to be stationary only if its mean and variance is constant at any given time. In stationary data mean must be constant as it was shown in figure.5 and 6.

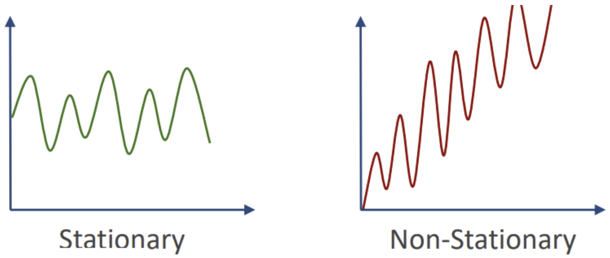


Figure 5. Constant Mean

In stationary data variance must be constant as it was shown in fig.5.

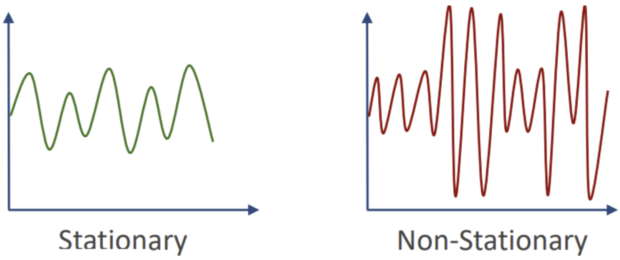


Figure 6. Constant Variance

4. Result

4.1 Data collection and Exploration

PH tractor sales data set was acquired from the Kaggle which contains twelve years from 2003 to 2006 which includes tractors of every month of every year so there are 144 columns and two rows.

4.1.1 Missing values

The missing values in the data will be checked. If any missing values are found, they can be visualized. However, it has been confirmed that the PH tractor sales data set does not have any missing values. Sales data for all months over the four years is complete.

```
dataset.isnull().sum()

Month-Year                 0
Number of Tractor Sold     0
dtype: int64
```

Figure 7. Missing values

```
import missingno as msno
msno.matrix(dataset)

<Axes: >
```

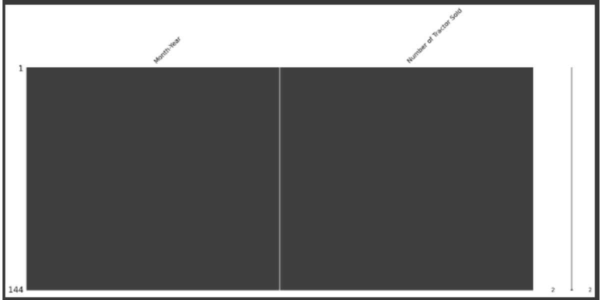


Figure 8. Visualization of no missing values

4.1.2 Outliers

Outliers that are treated as noise and fall far away from the majority of the data are checked. The visualization of the box plot (Figure 9) clearly shows that no outliers are present.

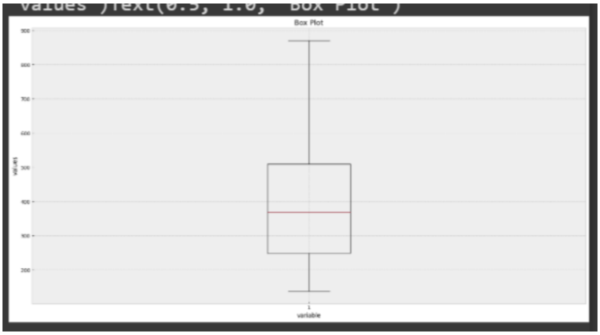


Figure 9. Outliers

4.1.3 Data visualization

The heat map was visualized from Figure 10 to better understand the confidence between all time and the number of tractors sold. The visualization displays the number of tractors sold each year, and an upward trend was shown in figure 11.

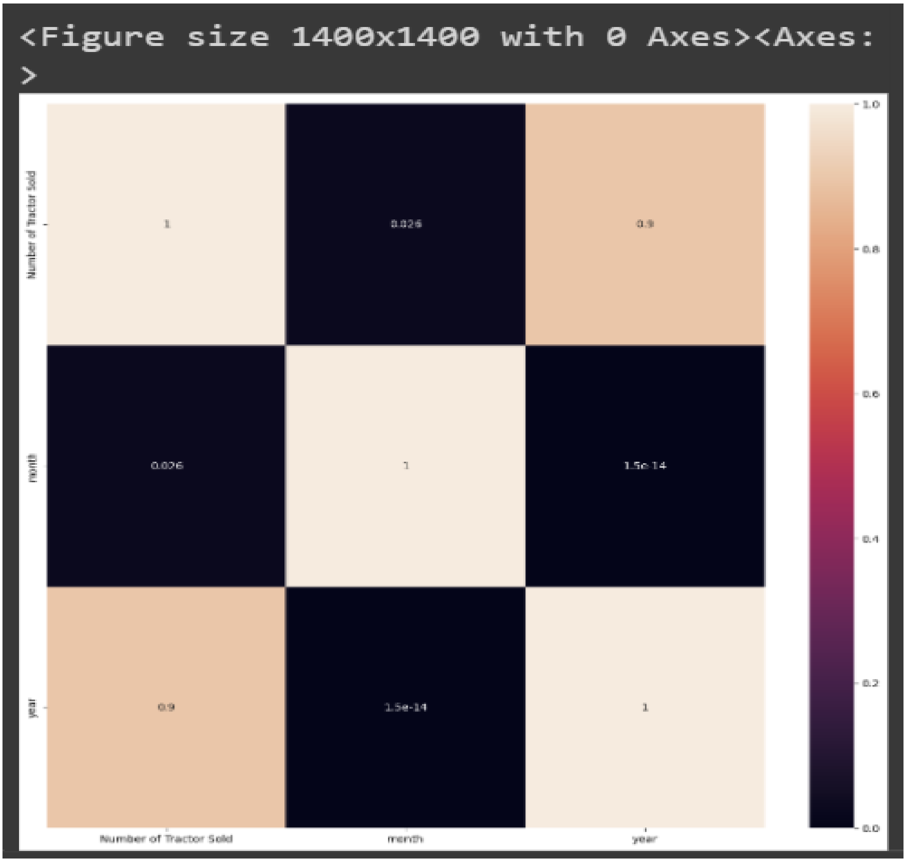


Figure 10. Heat map

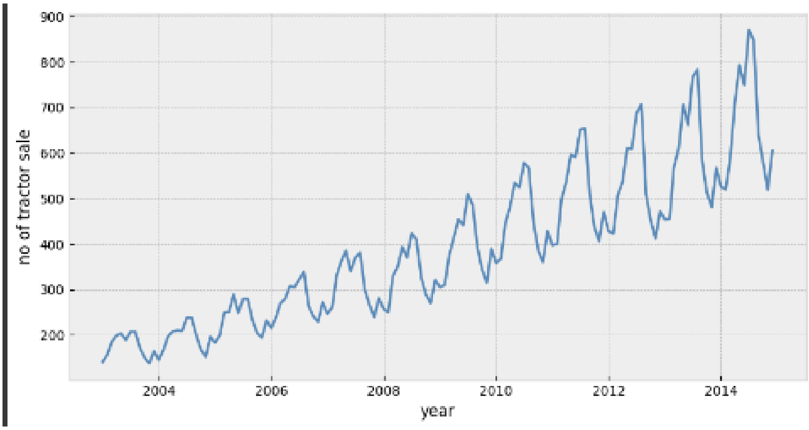


Figure 11. No. of tractors sold every year

The distribution of tractor sales is shown in Figure 12. It is evident from the density graph or bell graph that the distribution was done well.

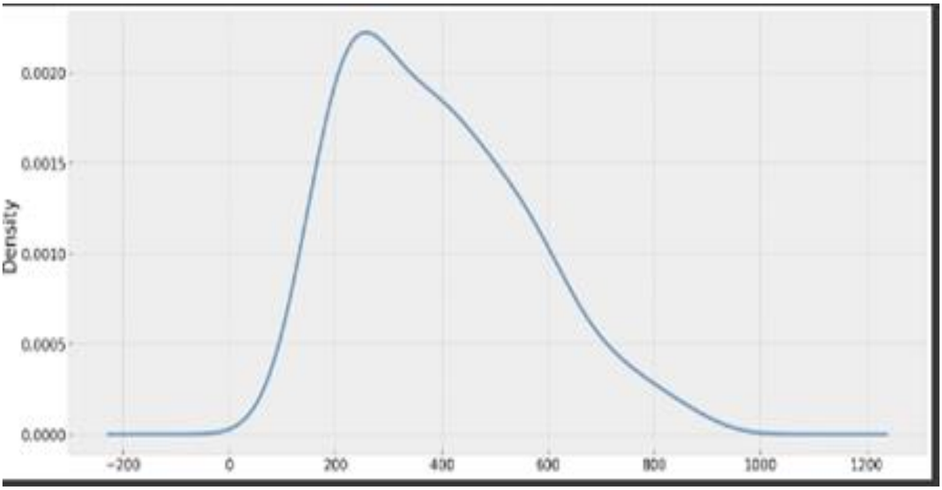


Figure 12. Distribution of tractor sales

4.1.4 Decomposition

A more wholesome understanding of the data can be provided, as it was visually determined that the data was not stationary. Additionally, underlying patterns in the PH tractor sales can be identified by building the "multiplicative time series decomposition" (Figure 13).

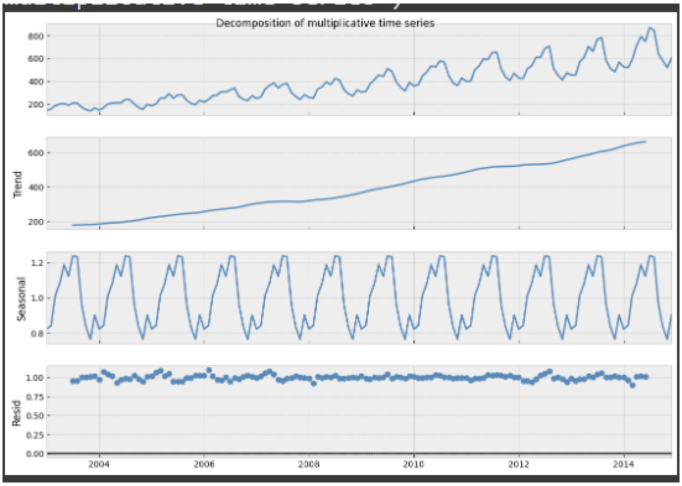
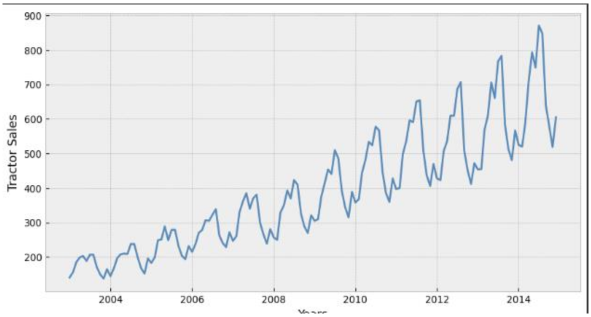


Figure 13. Decomposition of multiplicative time series

4.2 Stationarity

a) Difference data to make data stationary on mean (remove trend):



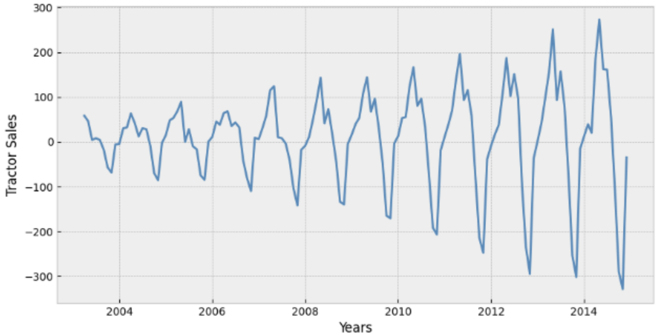


Figure 15. Time series with no trend

b) Log transformation to make data stationary on the variance:

Log transformation can be used to make a time series stationary with respect to variance. This was applied to the original data. The time series is not stationary with respect to the mean, so no difference is observed. Comparing Fig. 14 (as above mentioned), which shows the original time series, to Fig. 16, it can be seen that the time series became stationary.

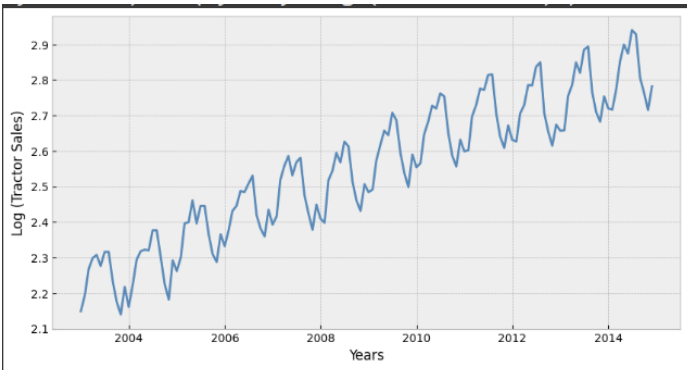


Figure 16. Time series stationary on variance

c) Difference log transform data to make data stationary on both mean and variance:

The time series was made stationary in terms of both mean and variance. This can be observed from Figure 17, which indicates that the time series lacks any trend or seasonality. The application of differencing and logarithmic transformation has effectively removed any persistent upward or downward trends, ensuring the mean remains constant over time. Additionally, the consistent amplitude of fluctuations throughout the series demonstrates a stable variance. Although initial observations might suggest some regular patterns, these patterns are not pronounced or consistent enough to signify significant seasonality. Consequently, the transformed time series is well-suited for accurate time series modelling and forecasting, as it meets the necessary stationarity conditions.

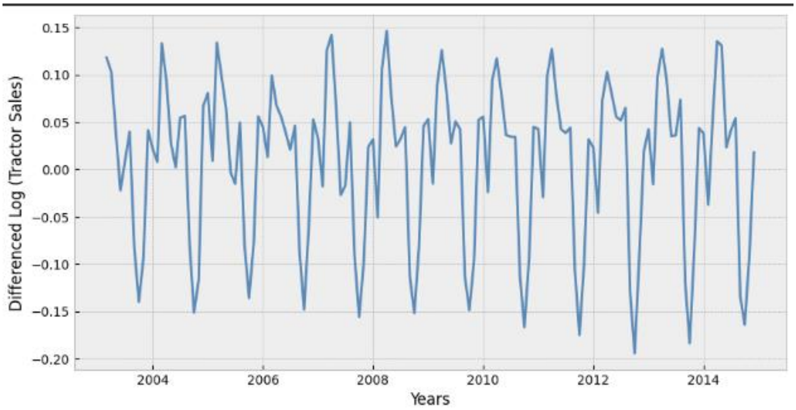


Figure 17. Time series stationary on variance

d) PH Tractor Seasonality:

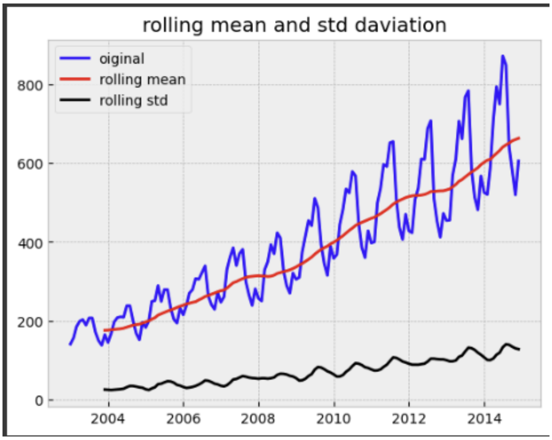


Figure 18. PH tractor seasonality

From Fig. 18, it can be seen that tractor sales have been increasing every single year without fail. The seasonal cycle occurring every 12 months can be figured out, as the mean value shows an increasing trend from the beginning of the year and drops down at the end of the year. The effect of seasonality is observed every 12 months.

e) Rolling Statistics:

- • **Remove wrinkles from the time series using moving average:**
    - The process of using a moving average helps to smooth out the time series data by eliminating short-term fluctuations and highlighting longer-term trends. This technique is useful in identifying and analyzing the underlying trend in the data over a specified period.
- **Moving average of different time periods i.e., 4, 6, 8, and 12 months from fig.19 that clearly defines the data has trend:**
    - The provided figure (Figure 19) displays the moving averages of tractor sales data over different time periods: 4, 6, 8, and 12 months.
    - In the 4-month moving average chart, the rolling mean (red line) smooths out some of the short-term fluctuations while retaining more detail of the original data.
    - The 6-month moving average provides a smoother trend line compared to the 4-month, further reducing the noise and clarifying the upward trend.
    - The 8-month moving average continues to smooth the data, making the trend even more apparent by minimizing seasonal variations.
    - The 12-month moving average offers the smoothest trend line, effectively showcasing the overall upward trend in tractor sales over the years, with minimal short-term fluctuations.

Overall, these moving averages help in removing "wrinkles" or noise from the time series data, clearly revealing the presence of an increasing trend in the number of tractors sold over the specified period.

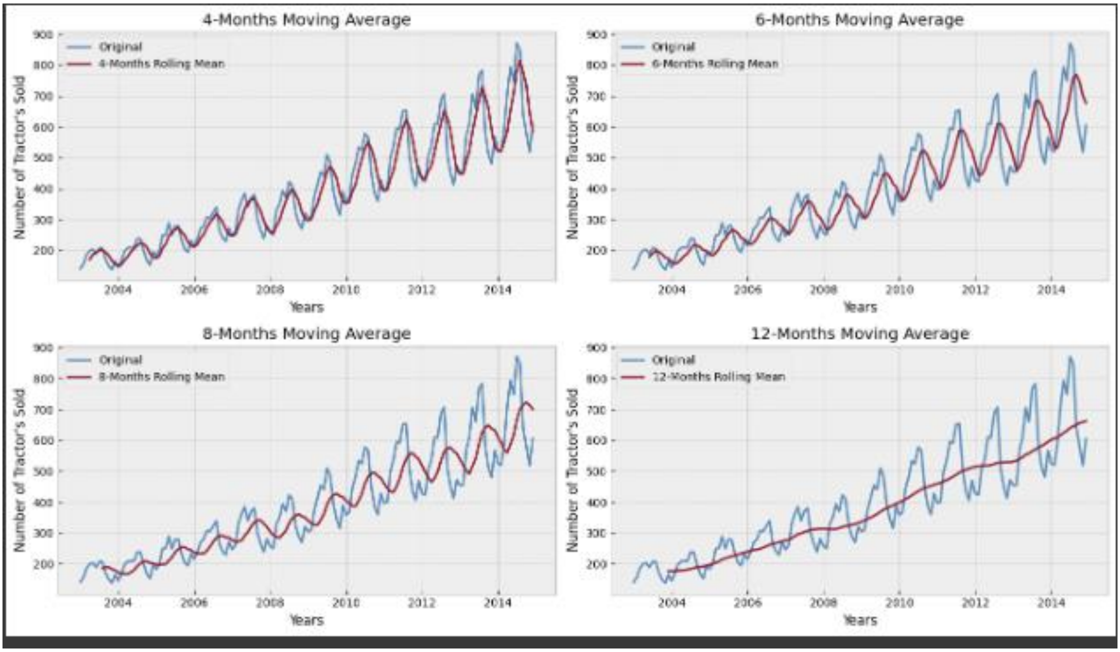


Figure 19. PH tractor Trend using moving average

f) Augmented Dickey Fuller test

ADF test was done using the regression equation given below calculated using the estimations of the coefficients which gives the critical value "p." The ADF test shows that the critical value (p-value) for the first difference is 0.99 (Figure 20), which is not less than 0.05 (the threshold). Therefore, the second difference must be used.

```
statical value              1.108825
p-value                     0.995291
#lags                      14.000000
number of obervation      129.000000
1%                         -3.482088
5%                         -2.884219
10%                        -2.578864
dtype: float64
```

Figure 20. ADF first difference

The figure, Fig. 21, indicates that the p-value is less than the threshold of 0.04. This suggests that the data is stationary. The second difference can be utilized in various ways.

```
statical value               -2.936724
p-value                       0.041241
#lags                        10.000000
number of obervation        130.000000
1%                           -3.482088
5%                           -2.884219
10%                          -2.578864
dtype: float64
```

Figure 21. ADF second difference

4.3 ACF and PACF Analysis

From the pattern observed in the graph (Figure 22), it can be determined that a seasonal component is present in the residual around lag 12. This implies that, given the monthly data, seasonality occurs approximately every 12 months. Specifically:

1. **Autocorrelation Plots (ACF):**
   - The ACF plots show significant spikes at lag 12 in the first, third, and fifth rows. This indicates a strong seasonal pattern repeating annually.
   - Other significant lags around multiples of 12 (e.g., lag 24) further support the presence of this seasonal component.
2. **Partial Autocorrelation Plots (PACF):**
   - The PACF plots also exhibit significant spikes at lag 12 in the second and fourth rows. This reinforces the idea of seasonality occurring on a yearly basis.
3. **Overall Analysis:**
   - The consistent appearance of significant spikes around lag 12 in both ACF and PACF plots across multiple subplots suggests that the time series data being analysed has a pronounced yearly seasonal cycle.
   - This seasonality is consistent with monthly data, where a full cycle occurs every 12 months.

These observations provide strong evidence of a seasonal component in the time series data, indicating that any modeling approach should account for this annual seasonality to improve accuracy and predictive power.

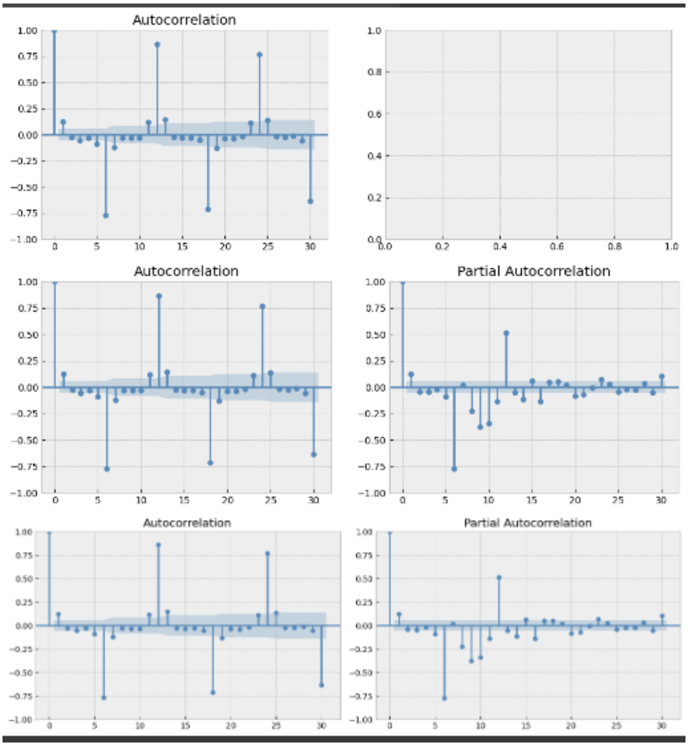


Figure 22. ACF and PCAF plots

4.4. Model Identification

From Fig. 23, it can be observed that the combination where the integrated (I) component equals one and the moving average (q) equals one was chosen. Additionally, a seasonal moving average with a lag of 12 was obtained. Both of these were as expected. The best combination will be the one with the minimum scores of AIC and BIC.

```
[(0, 0, 0),
 (0, 0, 1),
 (0, 1, 0),
 (0, 1, 1),
 (1, 0, 0),
 (1, 0, 1),
 (1, 1, 0),
 (1, 1, 1)][(0, 0, 0, 12),
 (0, 0, 1, 12),
 (0, 1, 0, 12),
 (0, 1, 1, 12),
 (1, 0, 0, 12),
 (1, 0, 1, 12),
 (1, 1, 0, 12),
 (1, 1, 1, 12)]Best SARIMAX(0, 1, 1)x(1, 0, 1, 12)12 model - AIC:-733.7719162311134
```

Figure 23. Best fit Seasonal ARIMA

4.5 Sarimax Model Estimation

Predict sales on in-sample date using the best fit SARIMAX model. The next step is to predict tractor sales for in-sample data and find out how close is the model prediction on the in-sample data to the actual truth.

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                         Tractor-Sales   No. Observations:                144
Model:             SARIMAX(0, 1, 1)x(1, 0, 1, 12)   Log Likelihood                 370.886
Date:                            Thu, 21 Sep 2023   AIC                           -733.772
Time:                                    13:16:16   BIC                           -721.921
Sample:                                01-01-2003   HQIC                          -728.956
                                     - 12-01-2014
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.3537      0.069     -5.135      0.000      -0.489      -0.219
ar.S.L12       0.9934      0.006    177.633      0.000       0.982       1.004
ma.S.L12      -0.5561      0.096     -5.764      0.000      -0.745      -0.367
sigma2         0.0003   2.74e-05      9.211      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                 1.46
Prob(Q):                              0.89   Prob(JB):                         0.48
Heteroskedasticity (H):               0.64   Skew:                             0.04
Prob(H) (two-sided):                  0.13   Kurtosis:                         3.49
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 24. Prediction of in sample sales

The sample sales data and the forecast are shown to overlap very well in Figure 25. The overlap is evident from the figure.

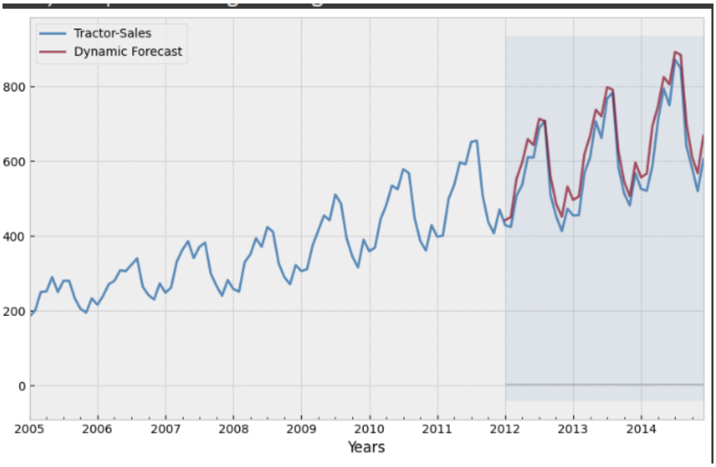


Figure 25. Plot of predictions of sample sales

4.6 Deep Learning Models Estimation

4.6.1 Preprocessing of data for deep learning models:

The data is split for training and testing. This process is explained in Figure 26.

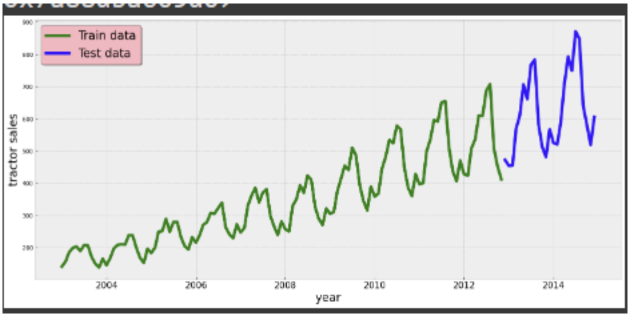


Figure 26. Splitting data into train and testing data set

4.6.2 Windowing

The sales of the tractor are predicted based on previous data. The previous data has been considered in more detail. For example, if a window of 10 is taken, the data is split into X and Y. In the first iteration the first ten (1 to 10) observations get stored in the X and the 11th observation get stored in the Y and in the second iteration (2 to 11) elements get stored in the X and the 12th observation will be stored in the Y and so on. This will help to train the data better.

```
Orignial length of train data :  119
length of train data after windowing :  89
```

Figure 27. Training dataset after windowing

4.6.3 Model 1 Gated Recurrent Unit (GRU) Fitting the GRU model.

The model was built using a function named "sequential()". GRU layers with 256 units were included, along with a dropout rate of 0.5. A dense layer, which is fully connected and has 10 layers, was also incorporated. The 'Adam' optimizer with a learning rate of 0.001 was utilized, and the model was run for 100 epochs (Figure 28).

```
Epoch 12/100
1/1 [==============================] - 0s 83ms/step - loss: 0.3505
Epoch 13/100
1/1 [==============================] - 0s 94ms/step - loss: 0.3308
Epoch 14/100
1/1 [==============================] - 0s 83ms/step - loss: 0.3489
Epoch 15/100
1/1 [==============================] - 0s 73ms/step - loss: 0.3592
Epoch 16/100
1/1 [==============================] - 0s 77ms/step - loss: 0.3598
Epoch 17/100
1/1 [==============================] - 0s 68ms/step - loss: 0.3547
Epoch 18/100
1/1 [==============================] - 0s 89ms/step - loss: 0.3501
```

Figure 28. GRU epochs

The GRU prediction visualization is shown in Figure. 29. The true values and predicted values exhibit significant errors. These errors are particularly noticeable around April 2013 and April 2014 due to the steepness of the model.

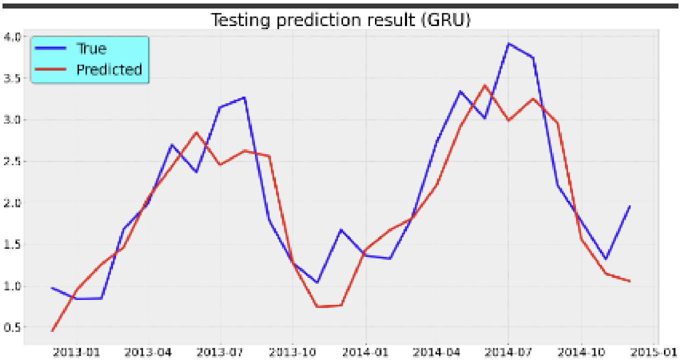


Figure 29. Testing prediction results (GRU)

4.6.4 Model 2 Simple RNN Fitting Simple RNN model.

The model was built using the function named "sequential()". RNN layers with 128 units were included, along with a dropout rate of 0.2. A fully connected dense layer with 10 layers was also incorporated. The 'Adam' optimizer with a learning rate of 0.001 was used, and the loss function MSE (mean squared error) was applied. The model was run for 100 epochs (Figure 30).

```
1/1 [==============================] - 0s 22ms/step - loss: 0.0883
Epoch 21/100
1/1 [==============================] - 0s 21ms/step - loss: 0.1103
Epoch 22/100
1/1 [==============================] - 0s 24ms/step - loss: 0.0852
Epoch 23/100
1/1 [==============================] - 0s 24ms/step - loss: 0.0626
Epoch 24/100
1/1 [==============================] - 0s 33ms/step - loss: 0.0518
Epoch 25/100
1/1 [==============================] - 0s 23ms/step - loss: 0.0676
Epoch 26/100
1/1 [==============================] - 0s 23ms/step - loss: 0.0811
Epoch 27/100
1/1 [==============================] - 0s 23ms/step - loss: 0.0797
Epoch 28/100
1/1 [==============================] - 0s 22ms/step - loss: 0.0760
Epoch 29/100
1/1 [==============================] - 0s 24ms/step - loss: 0.0653
```

Figure 30. RNN epochs

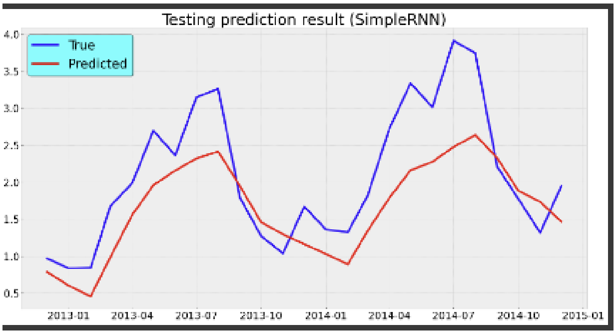


Figure 31. Testing prediction results (Simple RNN)

The prediction visualization of the Simple RNN shows that a significant error is present. This is because a drawback of the RNN is losing gradient over long prediction periods, which can result in information loss. Unlike the GRU, this leads to poor forecasting.

4.6.5 Model 3 Long Short-Term Memory (LSTM)

Fitting Long Short-Term Memory (LSTM)

The model was built using a function named "sequential()" which was initialized. LSTM layers with 128 units and a dropout rate of 0.2 were included. A fully connected dense layer with 10 units was added. The 'Adam' optimizer with a learning rate of 0.001 was used. The loss function employed was MSE (mean squared error). The model was executed for 100 epochs.

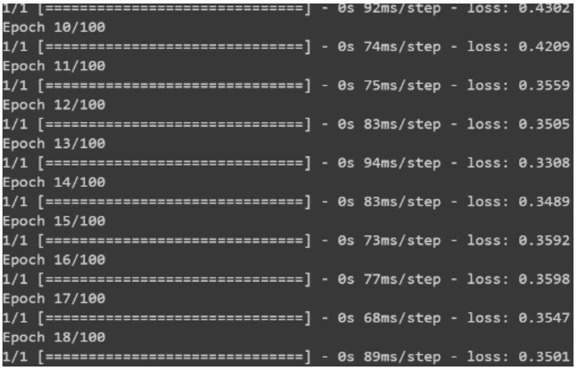


Figure 32. LSTM epochs

LSTM Prediction Visualization

From Fig. 33, it can be seen that the LSTM was not trained properly, contrary to what was expected from the model. This is because the data is not stationary. Both upward trends and a seasonality component are present in the data. LSTM is highly sensitive to non-stationary data. Differentiation might result in the loss of data.

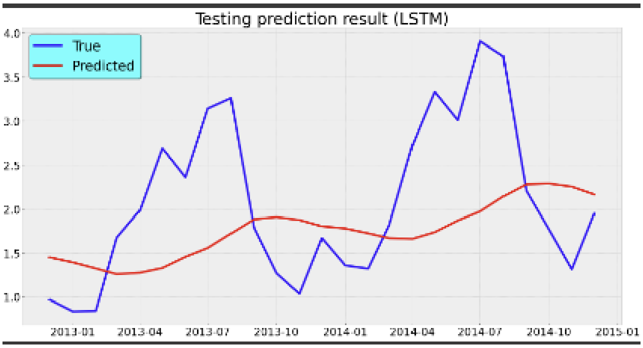


Figure 33 Testing prediction results (LSTM)

4.6.6 Model 4 Bidirectional LSTM

Fitting Bidirectional Long Short-Term Memory

The model was built using a function named `sequential()` which was initialized. Bidirectional LSTM layers with 128 units were included, along with a dropout rate of 0.2. A dense layer, fully connected with 10 layers, was also added. The 'Adam' optimizer was incorporated with a learning rate of 0.001. The loss function used was MSE (mean squared error), and the model ran for 100 epochs. The epochs that were executed in this model are shown in Figure 34.

```
1/1 [==============================] - 0s 224ms/step - loss: 0.3104
Epoch 16/100
1/1 [==============================] - 0s 216ms/step - loss: 0.3097
Epoch 17/100
1/1 [==============================] - 0s 214ms/step - loss: 0.3267
Epoch 18/100
1/1 [==============================] - 0s 239ms/step - loss: 0.2985
Epoch 19/100
1/1 [==============================] - 0s 226ms/step - loss: 0.2750
Epoch 20/100
1/1 [==============================] - 0s 183ms/step - loss: 0.2914
Epoch 21/100
1/1 [==============================] - 0s 231ms/step - loss: 0.3086
Epoch 22/100
1/1 [==============================] - 0s 210ms/step - loss: 0.3009
Epoch 23/100
1/1 [==============================] - 0s 187ms/step - loss: 0.2855
Epoch 24/100
1/1 [==============================] - 0s 226ms/step - loss: 0.2833
<keras.src.callbacks.History at 0x7a88d976d240>
```

Figure. 34 Bidirectional Epochs

Bidirectional LSTM Prediction visualization

From Figure 35, it can be seen that the bidirectional LSTM was not trained properly, which was not as expected from the model. This issue arises because the data is not stationary. The data exhibits both upward trends and a seasonal component, and the bidirectional LSTM is highly sensitive to non-stationary data. Differencing the data might result in data loss.

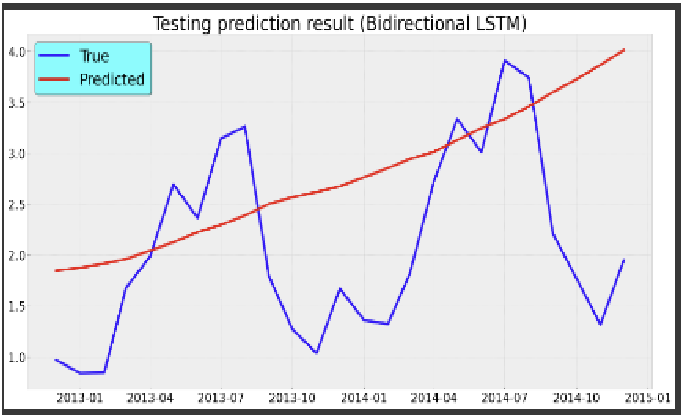


Figure. 35 Testing prediction results (Bidirectional

4.6.7 Model 5 CNN LSTM Encoder Decoder

Fitting CNN LSTM Encoder Decoder

The model is initialized with the function called "sequential()". Next, the convolutional encoder, which consists of one directional convolutional layer with 256 filters and a filter size of three, is added. The activation layer called RELU is used, and the data is trained. The second layer also does the same training operations and produces feature mappings and next layer max pooling reduce feature dimensions and flatten layer make these features into 1D vector. Next this will be sent to the LSTM layer with 128 units and RELU activation function and it was sent to the two dense layers with one hundred layers each with same activation function (RELU) this will be come out as 1d vector and serves as output vector and further sent to the training.

```
1/1 [==============================] - 0s 145ms/step - loss: 9.7889e-04
Epoch 92/100
1/1 [==============================] - 0s 134ms/step - loss: 9.4246e-04
Epoch 93/100
1/1 [==============================] - 0s 141ms/step - loss: 9.1410e-04
Epoch 94/100
1/1 [==============================] - 0s 147ms/step - loss: 8.8183e-04
Epoch 95/100
1/1 [==============================] - 0s 144ms/step - loss: 8.5430e-04
Epoch 96/100
1/1 [==============================] - 0s 144ms/step - loss: 8.3541e-04
Epoch 97/100
1/1 [==============================] - 0s 135ms/step - loss: 8.4008e-04
Epoch 98/100
1/1 [==============================] - 0s 128ms/step - loss: 8.8225e-04
Epoch 99/100
1/1 [==============================] - 0s 136ms/step - loss: 0.0010
Epoch 100/100
1/1 [==============================] - 0s 134ms/step - loss: 0.0014
<keras.src.callbacks.History at 0x7a88e1888c70>
```

Figure. 36 CNN LSTM Encoder decoder

CNN LSTM Encoder decoder Visualization

From Figure. 37, it can be seen that the model was well fitted compared to LSTM models. The complex patterns were untangled, and the data trend along with the seasonality factor was learned effectively. This model represents the best improvement.

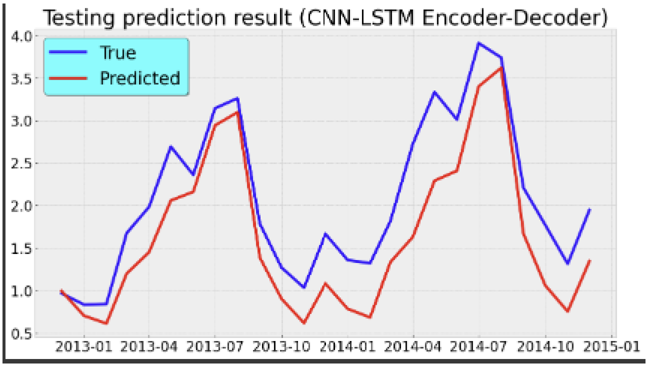


Figure. 37 Testing prediction results (CNN LSTM Encoder decoder)

4.6.8 Model 6 CNN Fitting CNN

```
1/1 [==============================] - 0s 46ms/step - loss: 0.0024
Epoch 92/100
1/1 [==============================] - 0s 54ms/step - loss: 0.0024
Epoch 93/100
1/1 [==============================] - 0s 48ms/step - loss: 0.0023
Epoch 94/100
1/1 [==============================] - 0s 58ms/step - loss: 0.0022
Epoch 95/100
1/1 [==============================] - 0s 51ms/step - loss: 0.0022
Epoch 96/100
1/1 [==============================] - 0s 59ms/step - loss: 0.0021
Epoch 97/100
1/1 [==============================] - 0s 49ms/step - loss: 0.0021
Epoch 98/100
1/1 [==============================] - 0s 47ms/step - loss: 0.0020
Epoch 99/100
1/1 [==============================] - 0s 48ms/step - loss: 0.0020
Epoch 100/100
1/1 [==============================] - 0s 52ms/step - loss: 0.0019
```

Figure. 38 CNN Epochs

The prediction of PH tractor sales data using CNN was visualized from figure.39. Errors are present in the CNN visualization, but they are not as severe as those in the LSTM. Future improvements in prediction can be achieved using different methods.

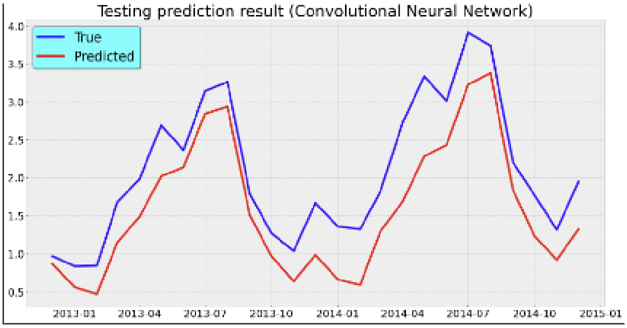


Figure. 39 Testing prediction results (CNN)

4.7 Comparing SARIMAX Model and Deep Learning Models

Tabel.1 RMSE Scores of all the models.

MODELS	RMSE SCORES
Seasonal ARIMA	0.01
GRU	0.50
Simple RNN	0.64
LSTM	0.95
Bidirectional LSTM	1.15
Encoder-Decoder LSTM CNN	0.53

CNN	0.55

From the above Tab.1, it can be said that the seasonal model is the best model for forecasting PH tractor sales for the next three years. To further demonstrate that the seasonal ARIMA model is the best, plots were drawn to provide confidence in its accuracy.

4.8 Forecasting Sales Using Best Model

After comparing all the methods traditional methods and deep learning methods the best model will be "Seasonal ARIMA" as it got the lowest RMSE score of all the models, and it is best suited model for the non-stationary data.

In this step, the tractor sales for the next three years—2015, 2016, and 2017—will be forecasted using the best model selected above. The sales will be forecasted with a 95% confidence band. This band typically results in a narrow interval prediction, which provides great precision but may ignore the true value. The upper and lower boundaries are drawn with grey lines, ensuring that the forecast is within these boundaries.

Key Observations:

If forecasting models are done every six months, they will be very useful. A good idea of what must be done will be given. The long-term forecasting model used for three years requires evaluation every six months. Changes made in the business must be added to this model. After comparing both traditional and deep learning methods, the "Seasonal ARIMA" model emerged as the most effective, achieving the lowest RMSE score among all evaluated models and proving to be the most suitable for non-stationary data.

In this step, the forecast for tractor sales for the years 2015, 2016, and 2017 will be generated using the selected model. The forecasts will include a 95% confidence band, which provides a narrow prediction interval for greater precision. However, this band may sometimes exclude the true values. The forecast boundaries are illustrated with grey lines, indicating the range within which the forecasts are expected to fall (Figure 40).

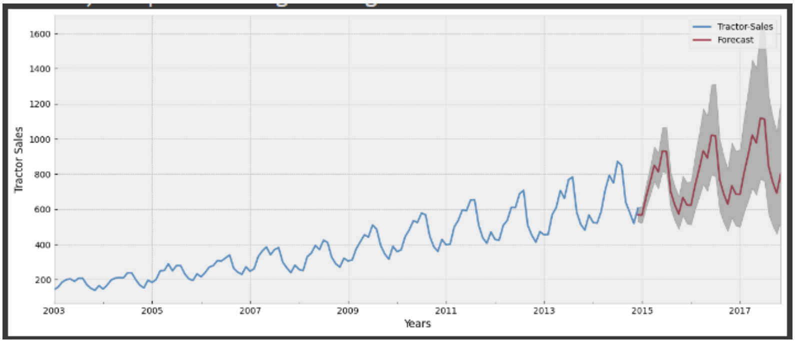


Figure. 40 Sales forecast for next three years with 95% confidence

The forecast is shown using the 99% confidence band in Figure 41. A narrow interval cannot be achieved with this band, so true values cannot be excluded. Although this band might have less precision, it is ideally used for long-term forecasts. It can be observed that the forecast is within the boundaries.

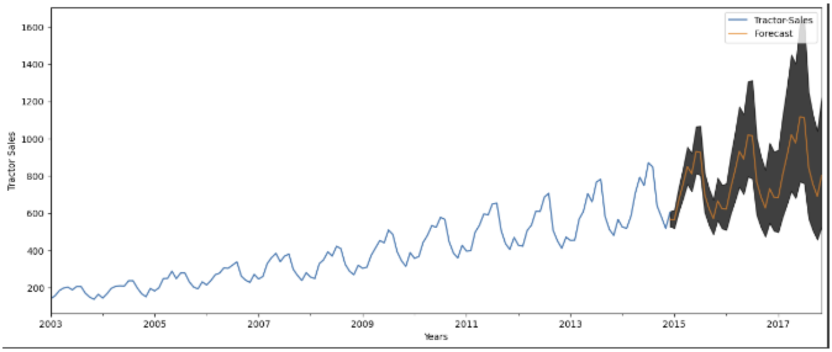


Figure. 41 Sales forecast for next three years with 99% confidence.

4.9 Forecasting Performance Metrics

ACF and PACF plots are to see the residuals for the seasonal ARIMA model to make sure the model left no information or learning or patterns to extract from the data. No patterns are shown by the residuals, indicating that they are uncorrelated and have a zero mean. If this does not occur, it suggests that the model did not perform well, and changes must be made. However, in this case, the seasonal ARIMA (0, 1, 1) (1, 0, 1) [12] model, as evidenced by the KDE plots from above, demonstrates that the residuals are normally distributed.

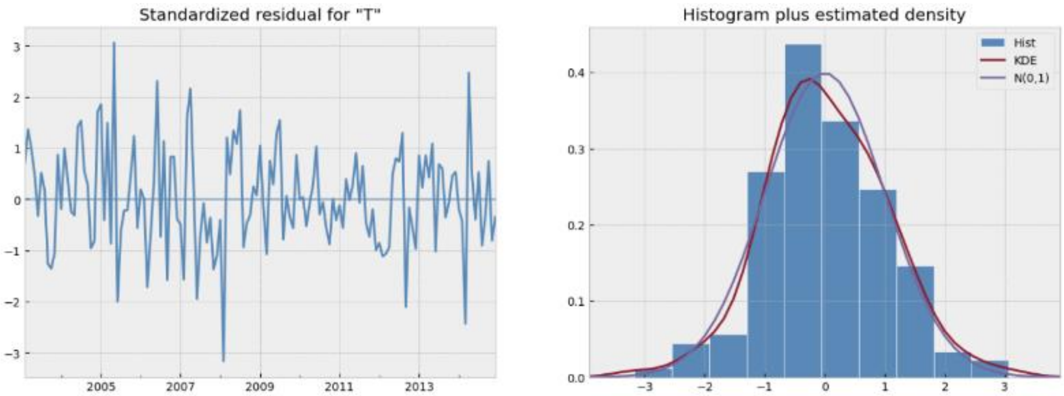


Figure. 42 ARIMA residual graphs 1

The normal Q-Q plots reveal that the residuals of the sample plotted show a linear trend. This trend indicates a strong indication of a normal distribution.

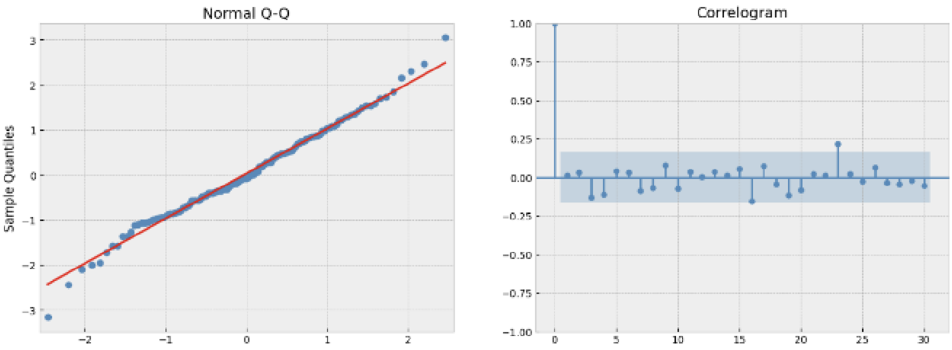


Figure. 43 ARIMA residual graphs 2

The residuals over time don't show any significant seasonality and they appear to be white noise. This can be confirmed by the autocorrelation graph (i.e., correlogram) from Figure 43, which shows that the time series residuals have low correlation with lagged versions of itself.

Those observations coupled with the fact that there are no spikes outside the insignificant zone for both ACF and PACF plots lead to conclude that residuals are random with no information or patterns in them, and the model produces a satisfactory fit that could help to understand the time series data and forecast future values. It means that the ARIMA model is working fine.

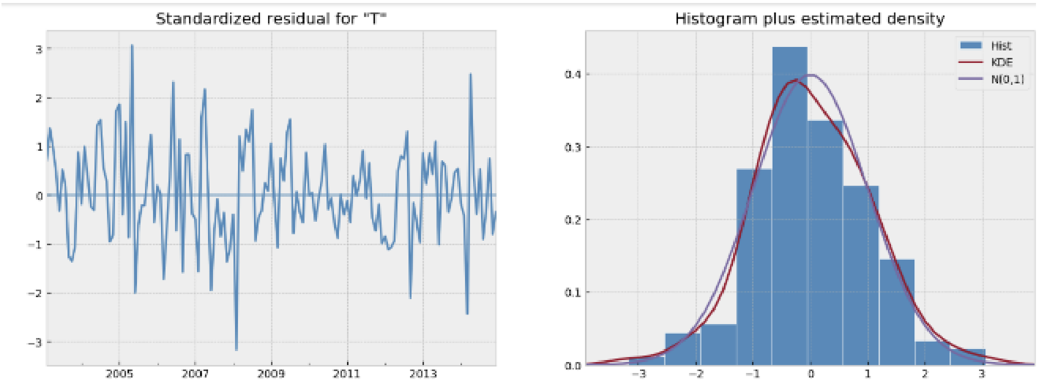


Figure. 44 ARIMA residual graphs 3

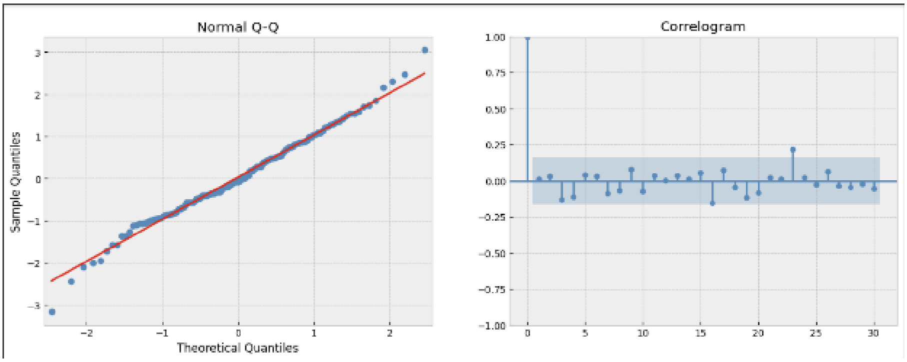


Figure. 45 ARIMA residual graphs 4

5. Discussion

The hypothesis was considered, and it was observed that the tractor sales data set was not stationary. The presence of a trend and seasonality was indicated. To address this, different methods such as differencing, log transformation, and rolling statistics were employed to make the data stationary. A strong seasonality component was evident in the data, which led to the use of the SARIMAX model to account for seasonality and other environmental factors that could impact sales. It was observed that traditional models and deep learning models did not perform equally. The SARIMAX model outperformed the deep learning models.

Residuals are shown to have no patterns, indicating that they are uncorrelated and distributed with zero mean. If this does not occur, a poor performance of the model is implied, necessitating changes. In this case, the seasonal ARIMA (0, 1, 1) (1, 0, 1) [12] model was used. The KDE plots demonstrate that the residuals are normally distributed. The Normal Q-Q plots reveal that the residuals of the plotted sample exhibit a linear trend, which strongly suggests a normal distribution. The residuals over time, as shown in the top-left plot, do not display significant seasonality and appear to be white noise. This can be confirmed by the autocorrelation graph (i.e., correlogram) from fig.47, which shows that the time series residuals have low correlation with lagged versions of itself. Those observations coupled with the fact that there are no spikes outside the insignificant zone for both ACF and PACF plots lead to conclude that residuals are random with no information or patterns in them, and the model produces a satisfactory fit that could help to understand the time series data and forecast future values. It means that the ARIMA model is working fine. From the above observations, it can be said that SARIMAX is the best model for forecasting future sales of tractors. It is evident from the data that this model provides the most accurate predictions.

References

[1] Y. Liu, L. Lieli, (2009). Sales forecasting through fuzzy neural networks. *In Proceedings of the 2009 International Conference on Intelligent Systems and Knowledge Engineering (ISKE), IEEE,* China. https://doi.org/10.1109/ICECT.2009.65

[2] L. Wu, J.Y. Yan, Y.J. Fan, (2012). Data mining algorithms and statistical analysis for sales data forecast. *In 2012 Fifth International Joint Conference on Computational Sciences and Optimization IEEE,* China. https://doi.org/10.1109/CSO.2012.132

[3] E.E. Frey, P.H. Braz, M.E. Camargo, M.R. Cruz, A. Galelli, S.L. Russo, G.F. Silva (2012). Planning and production indicators facing seasonal agricultural machinery market. *In Proceedings of the 2012 International Conference on Robotics and Automation Science and Engineering (ICBASE), IEEE,* Indonesia. https://doi.org/10.1109/ICMIT.2012.6225893

[4] L.G.N. De Leon, R.C. Gomez, M.L.G. Tacal, J.V. Taylar, V.V. Nojor, A.R. Villanueva, (2022). Bitcoin price forecasting using time-series architectures. *In 2022 International Conference on ICT for Smart Society (ICISS), IEEE,* Indonesia. https://doi.org/10.1109/ICISS55894.2022.9915199

[5] A.K. Biswas, S.I. Ahmed, T. Bankefa, P. Ranganathan, H. Salehfar, (2021). Performance analysis of short and mid-term wind power prediction using ARIMA and hybrid models. *In 2021 IEEE Power and Energy Conference at Illinois (PECI), IEEE,* USA. https://doi.org/10.1109/PECI51586.2021.9435209

[6] Y. Ali, S. Nakti, (2023). Sales Forecasting: A Comparison of Traditional and Modern Times-Series Forecasting Models on Sales Data with Seasonality. *In 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE,* India.

[7] Y. Ensafi, A.H. Saman, Z. Guoqing, S. Bharat, Time-series forecasting of seasonal items sales using machine learning–A comparative analysis. *International Journal of Information Management Data Insights,* 2(1), (2022) 100058. https://doi.org/10.1016/j.jjimei.2022.100058

[8] ankur raja, “Time Series: Decomposition (Part I) — Trend Cycle Computation,” Analytics Vidhya, Nov. 01, 2019. https://medium.com/analytics-vidhya/time-series-decomposition-part-i-trend-cycle-computation-29fac227896a

[9] V. Velde, F.A. Parvez, J. Chaitanya, A Blockchain Enabled System for Security, Non-Repudiation, and Integrity of Judiciary Proceedings. *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), IEEE,* India. https://doi.org/10.1109/ICEEICT53079.2022.9768427

Funding

No funding was received for conducting this study.

Conflict of interest

The Author's have no conflicts of interest to declare that they are relevant to the content of this article.

About The License