**Atul Johnson**

**Introduction –**

In order to shuttle people at a rapid pace, one of the public rail transportation systems is called Jet rail. With greater use of jet rail, the transportation sector has seen a significant increase in the requirement for forecasts concerning transport. For data analysts and transportation investors, precisely forecasting future data is a crucial undertaking and one of the most exciting fields. Time series data has always been on the rise across a variety of fields. Data from a time series is a chronologically organised stream of real-time value information. The time series data plays a huge role in the future of jet rails.

So as a formal definition to time series - it is a series of data, where the observations are recorded at regular interval, and these values has a order or sequence. There are several characteristics of timeseries data –

* It has a data-time component
* The subsequent observations depend upon the previous values
* The observations in the data are collected at equal interval of time

**Background Work –**

LI LI.et.al [1], In the study, many time series models are discussed. Additionally, he discusses the capabilities of the Intelligent Transportation System (ITS) and offers solutions for abnormal data detection, missing data and imputation in the dataset, data compression, and traffic prediction. The ITS has handled missing values and imputed the correct time-based value, which is entirely dependent on the missing period of time.

Priyamvada.et.al [2], discusses how to anticipate using the Holt-Winter model in several domains and forecast data based on this model to determine if a forecast is short-term or long-term.

(Sak et al., 2014) [3], discuss about the LSTM model (Long Short-Term Memory Model. It is a RNN model (Recurrent Neural Network) used in Deep learning. This model has good performance in the data with seasonality and particularly in sequence prediction tasks, and are capable of learning long-term dependencies.

**Need for the Solution -**

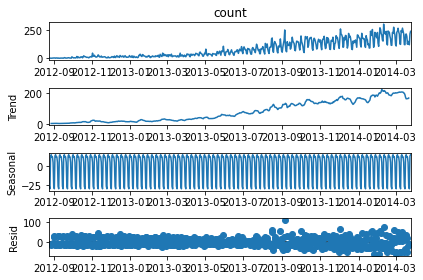
Jet rail is a new form of transportation, which uses jet propulsion technology to run the rails. That is to transport people from one place to another like trains, local trains and metros. So, over the years, the number of passengers using the jet rail has increased and this has resulted in various problem. Forecasting in the current system is based on set constant values applied to every day. Therefore, just a fixed constant will be used to compare the newly collected data. As a result, the current system lacks the intelligence needed to forecast future data.

They had limited resources and limited manpower, and they did not expect their traffic or count of people to increase at such an alarming rate. So, maintaining and managing this increasing traffic has been a very challenging task for them. So, for the same reason jet rails wants to know the expected traffic in the coming months.

This will help them to optimize the process, also can schedule more jet rails as per the forecast. or they can increase the capacity, or mange the man power accordingly.

**Proposed Solution –**

Decomposed Chart



This graph shows the trend, seasonality and the residual of the data separately. This actually gives us an idea which model to use and how to reach an optimum solution. Here we can understand that, there is trend in the data and there is high seasonality seen in the data. So, from this we can infer that the models like Holt Winters, SARIMA, Prophet and LSTM will be best models to try.

Holt-Winter-

The predictions may be made using holt winters by considering trend, seasonality and average value. It is also known as triple exponential smoothing. The model can have the seasonal component in both additive and multiplicative form.

Additive form – used when the peaks and troughs are the same roughly size

Multiplicative form - used when the peaks and troughs have significant amplitudes.

SARIMA –

ARIMA model takes the past values to make the forecast, but it doesn’t capture the seasonality well. So, in that situations SARIMA is used (Seasonal ARIMA).

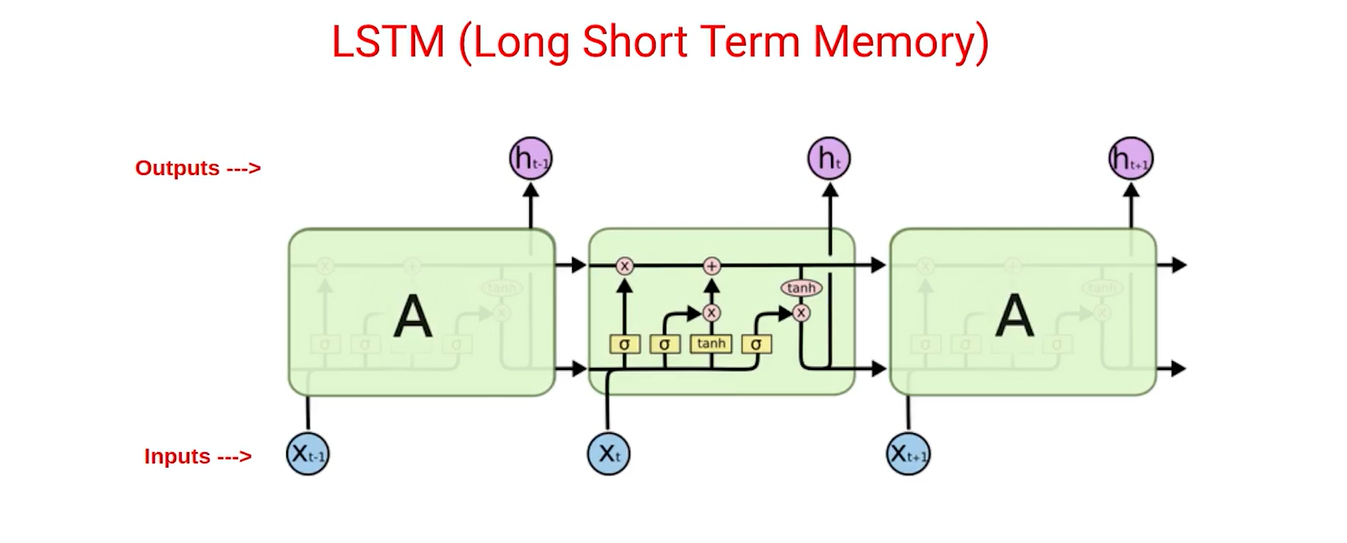
Parameters of SARIMA model –

* (p, d, q)
* (P, D, Q, m)

Prophet –

This time series forecasting algorithm was introduced by Facebook research team. The prophet algorithm mainly consists of three component – trend, seasonality and holiday effect. The holiday effect is a particular effect which is unusual period in the data, which is unexpected and doesn’t have any effect on the other data points.

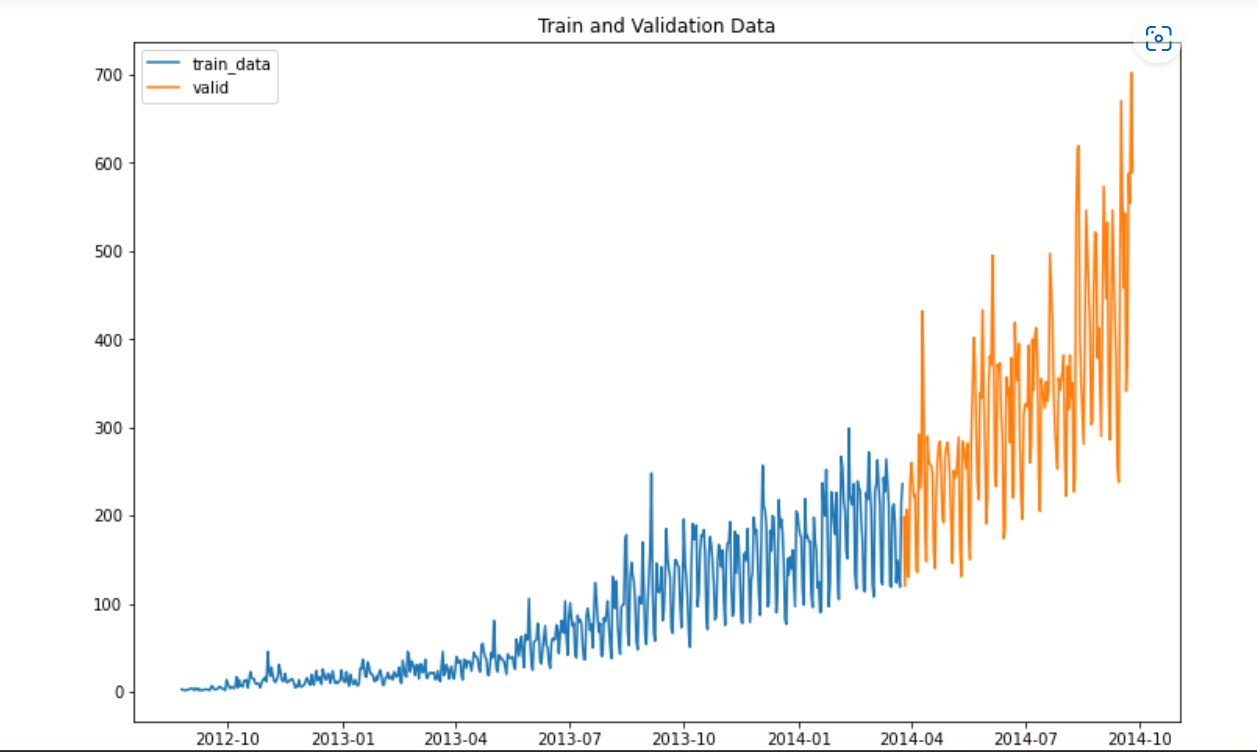
LSTM –

In multilayer perceptron, there is no transfer of sequential information happens. But in the LSTM, it connects the hidden layers of different time steps which results in more accuracy of the predictions. The long short term sequential memory helps to grab the trend and seasonality of the data. 

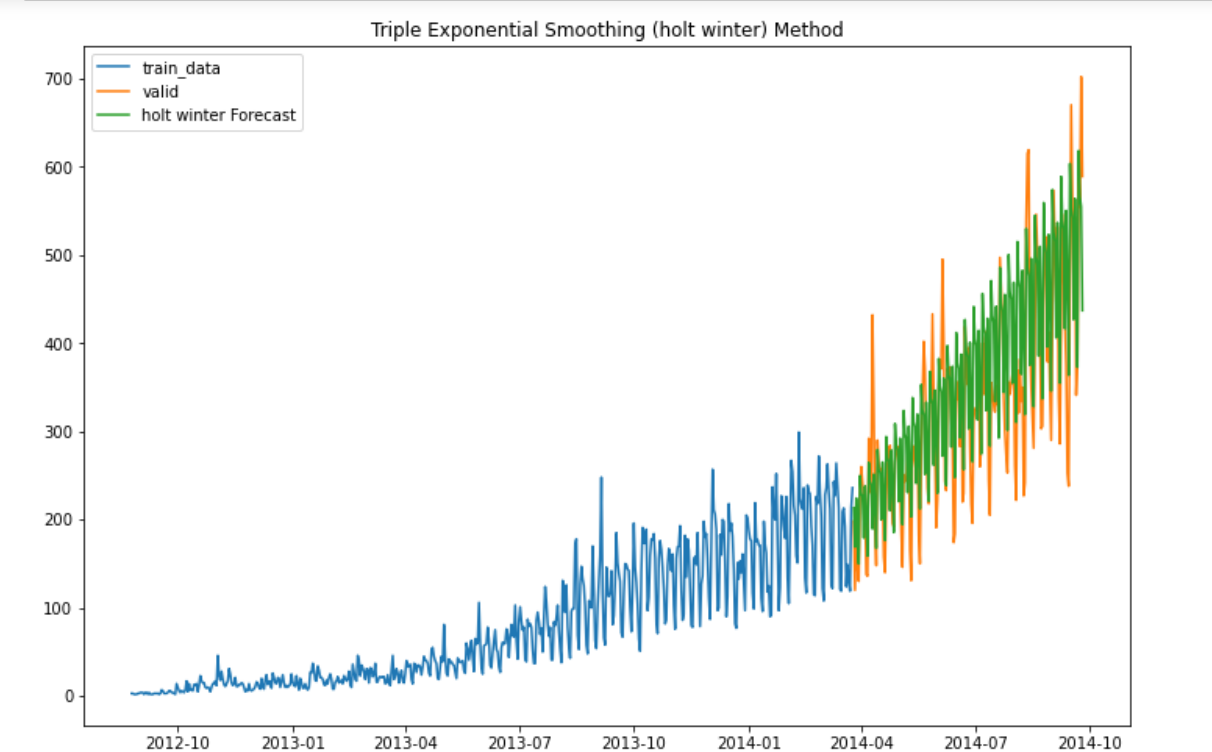
Here we can understand that, there are 2 inputs for each time step in the LSTM. One is the output of the last time step and another one is the hidden layers of the last time step.

**Analysis of solution –**

The diagram below shows the plot of the data of jet rail.

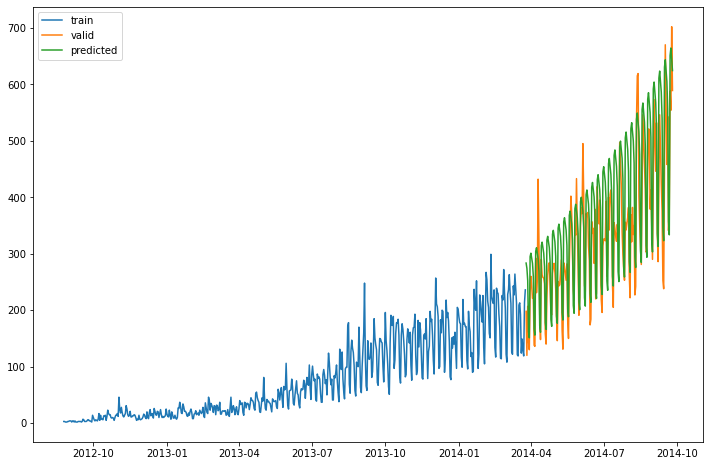


**Holt-Winters Model**



The RMSE value for Holt Winter Method is 78.59684424893284. Here we can see that, the triple exponential smoothing technique has successfully identified the trend and seasonality of the data. This has given a nice RMSE value of the 78. From the resultant graph we can understand that, the seasonality component of this data is captured and predicted very well.

**SARIMA**



The above diagram shows the fitting of SARIMA model in the data. The parameters of the model are mentioned below.

P -1

D - 1

Q - 1

m – 7

p - 2

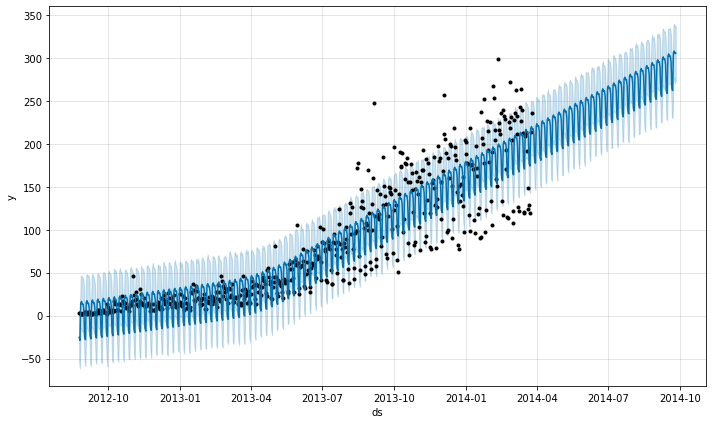
d - 1

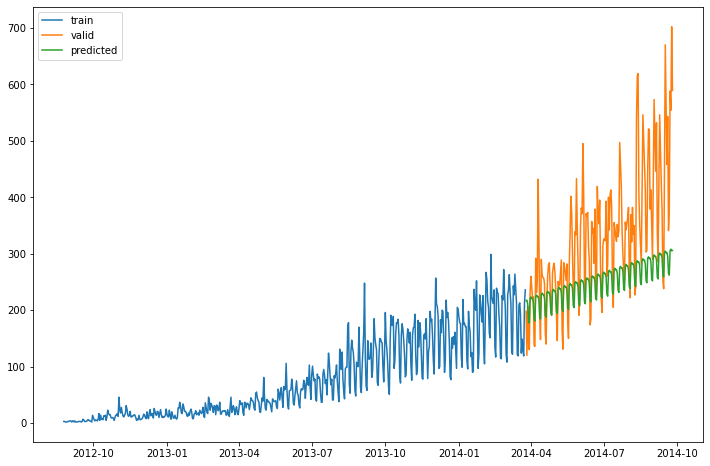
q - 2

The RMSE value for SARIMA is 78.9664046303094.

Here we can see that, the SARIMA model has successfully identified the trend and seasonality of the data. This has given a nice RMSE value of the 78. From the resultant graph we can understand that, the seasonality component of this data is captured and predicted very well. The RMSE value is considered as the evaluation metric here, so though there was more fitting for the data visually for SARIMA, the RMSE value doesn’t change with holts-winter.

**Prophet**

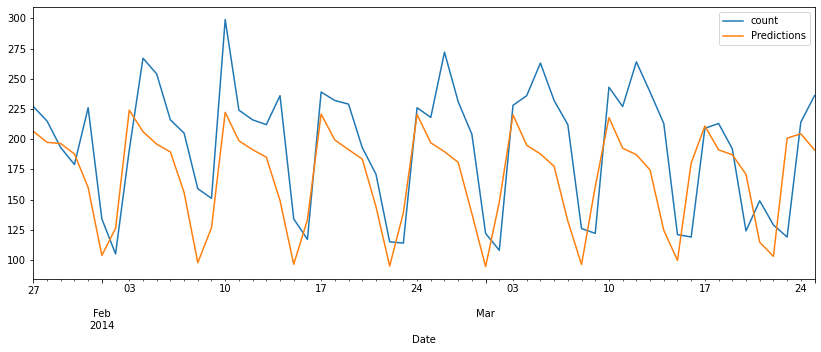




The RMSE value for Prophet is 118.18631808116592

The above diagram shows the fitting of prophet model with the data. Here we can understand that, the model hasn’t successfully identified the trend or seasonality of the data. This has made a significant impact on the RMSE value also. The model has a high RMSE value of 118. The highest values in the data are in the range of 600-700, the RMSE of 118 cannot be a good forecast. If we go with this forecast, then the error rate will be high and the accuracy will be very low. Which will ultimately lead to problems in business decision making.

**LSTM** (Long Short-Term Memory)



Here the dataset is given more training data. The data is split into 90:10, This is because the hidden processing layers in these neural networks are numerous, and as the number of layers rises, so do the instances that must be learned from. Here the data is scaled in order to reduce the variability in the data. After that, the number of features is given as 1 and number of inputs in the sequence is given as 12. This is mainly because of the seasonality show in the plots. The seasonal trend is shown here like this.

* Number of neurons – 100
* Activation Function – Relu

The RMSE value for LSTM is 44.867792063131326.

Here we can see that, the LSTM model has successfully identified the trend and seasonality of the data. This has given the best RMSE value of the 44. From the resultant graph we can understand that, the seasonality component of this data is captured and forecasted very well.

**Performance Evaluation –**

|  |  |
| --- | --- |
| **Model** | **RMSE Value** |
| Holt Winters Model | 78.596 |
| SARIMA | 78.966 |
| Prophet | 118.186 |
| LSTM | 44.867 |

So, from the above table, we can understand that, the LSTM model is the best fit for the data. The RMSE value for the LSTM model is the least - 44.86. This indicates that the model can be deployed using LSTM model.

**Recommendations –**

* From the analysis, we can understand that, LSTM model is the best model to deploy the jet rail customer count data
* By using this deep learning technique company can optimize the arrangement for arriving customers.
* Try to optimize the number of neurons, number of hidden layers, activation function etc.
* Business can certainly use this model because this model give good accuracy and less error rate.

**Conclusion and future scope of work**

Here we've implemented the Holt winters, SARIMA, PROPHET model, LSTM models, we've discovered that the LSTM model's predicted values are better and that it makes significantly fewer forecasting errors than other time series models and techniques. This paper's primary objective was to forecast the number of passengers using jet rail from 2014 to 2015.

As a future scope of work, the model can be deployed with a more accurate model. More tuning parameters can be taken into consideration and optimizing them will build a great model. Finding more information about the domain and going further more from time series analysis also can be considered as future scope.

**References –**

1. L. Li, X. Su, Y. Zhang, Y. Lin and Z. Li, "Trend Modeling for Traffic Time Series Analysis: An Integrated Study," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, pp. 3430-3439, Dec. 2015, doi: 10.1109/TITS.2015.2457240.
2. Priyamvada and R. Wadhvani, "Review on various models for time series forecasting," 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017, pp. 405-410, doi: 10.1109/ICICI.2017.8365383.
3. Sak, H., Senior, A., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. *Interspeech 2014*. <https://doi.org/10.21437/interspeech.2014-80>
4. Fattah, Jamal & Ezzine, Latifa & Aman, Zineb & Moussami, Haj & Lachhab, Abdeslam. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management. 10. 184797901880867. 10.1177/1847979018808673.