Sapienza
Università di Roma

Short-term water demand forecasting using traditional and deep learning models

Facoltà di Ingegneria dell'informazione,informatica e statistica
Corso di Laurea Magistrale in Data Science

Candidate
Poompavai Chandraprakash
ID number 1853976

Thesis Advisor
Prof. Simone Scardapene

Co-Advisor
Dr. Pietro Dicosta

Academic Year 2021/2022

Thesis defended on January 2022
in front of a Board of Examiners composed by:

Data Science Committee Members (chairman)

Short-term water demand forecasting using traditional and deep learning models

Master's thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: cmpavaichandru@gmail.com

Abstract

With big data being used and produced in today's technological world,time series prediction plays a vital role in various aspects.Analyzing time series data with effective visualization can help to produce various insightful inferences.

The project involves short term water demand forecasting using traditional and deep learning models.This prediction helps to identify and plan for seasonality, annual patterns, production capacity, and expansion over a longer period. which in turn drives a long-term business strategy (e.g., plans to launch a facility or store internationally and expand into new markets).

External factors are also considered like weather conditions, to check the water demand forecasting in a shorter term. Both univariate and multivariate time series analysis are carried out using Long Long short-term memory (LSTM) , Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors(Sarimax), Facebook prophet(Fbprophet), Baseline persistence model.

Acknowledgements

I would like to take this time to offer my gratitude to my thesis supervisor Prof.Simone Scardapane for having allocated time on his busy schedule to supervise this thesis. The door to Prof. Simone's office was always open whenever I ran into trouble or had a question about my research. He consistently welcomed me and steered me in the right directions during my work and writing of this thesis.I further offer many thanks to my Manager Mr.Pietro Dicosta and co-advisor Miss.Miriam Di Santa from NTT Data Italia without which the thesis was not going to be possible. Without their passionate participation and input, the validation and advice, the implementation would not have been a success.

Contents

Chapter 1

Introduction

This Chapter describes the main goals of this dissertation. Adding to that we provide the objectives and motivations for carrying out our study. A time series is a sequence of data points that occurs over some period of time in successive order.In general, time series allows us to find out the influence relationship of factors with the variables over time to time

1.1 Motivation

With the use of decision making strategy and managerial skills with respect to time and money over the business sector, we could reach our goals in terms of the fact, "Time is money". Forecast is needed in making strategic decisions under uncertainty. We may not think that we are forecasting, but our choices will be directed by our anticipation of the results of our actions or in actions. Failures are caused by delays and improper/no decisions.

Today, Artificial Intelligence (AI) and Big Data have redefined business forecasting methods. The most common analysis encountered is the time series analysis. Most organizations use time series forecasting for their business strategies. These methods have been used to monitor, clarify, and predict certain 'cause and effect' behaviors. Basically time series analysis helps to understand how the past influences the future.

1.2 Objectives and Contributions

The main objective of the thesis is to perform time series forecast, for planning the business in terms of seasonality, annual patterns and so on.

Here we performed hourly, weekly daily prediction using models such as LSTM, Sarimax, fb prophet and compared with baseline resistance model. Autokeras model was also used to check if it performs better.

1.3 Thesis Outline

The dissertation is divided into different chapters. Each chapter presents the contents with focus on the title of the chapter. Chapter 2 deals with the literature review of past research and the concepts to solve the problem. Chapter 3 deals with implementation, This section deals with the methodologies to approach the problem, data collection and analysis of data set,experimental models performed to solve the problem. Chapter 4 deals with the discussion of the results. In chapter 5, we provide conclusions of our research as regards to the experimental results and we also provide details about our future research directions indicating the methods we intend to employ for future research.

Chapter 2

Literature Review

This chapter describes the models used in univariate and multivariate time series forecasting. However, we put much focus on the comparison study of different models (both traditional and deep learning) in various aspect of time intervals say hourly, weekly and monthly predictions.

2.1 Baseline model

A baseline model is essential in time series forecasting. It helps to get the knowledge on the performance of the models on the given problem. It provides the point of comparison, models which performs less than baseline models are omitted[1] However there are few techniques to be considered for creating a proper baseline model as below:

- **Simple:** A method that requires little or no training or intelligence.
- **Fast:** A method that is fast to implement and computationally trivial to make a prediction.
- **Repeatable:** A method that is deterministic, meaning that it produces an expected output given the same input.

2.2 Persistence model

It is the simplest and baseline method to predict time series forecast.It is used to calculate the future on the assumption, with respect to the unchanged condition between the present time t and future time $t + T_h$. For stationary time series, when mean and variance doesn't change over time, which is called as dull persistence and the implementation would be[1]:

$$\hat{y}(t + T_h) = y(t)$$

Persistence model uses the previous day (or the corresponding day in the previous week) as a prediction. Such a method is sensitive to a method is sensitive to rare events (e.g., the day in the last month is the Black Friday day).

2.2.1 Persistence Naive forecast

The persistence algorithm uses the previous time step(t-1), to forecast/predict the outcome at the next time step(t+1).

Steps to achieve naive forecast:

1. Convert the univariate/multivariate data set into supervised learning problem
2. Split the data set into train and test sets.
3. Create a persistence model
4. Do forecast/prediction and form a baseline performance to create a comparison with other models.

2.3 Time Series concepts

This section deals with general time series concepts such as stationarity, stochastic process, linearity in time series models, deterministic time series

2.3.1 Stationarity in time series

Analysing time series data involves checking its stationarity. Stationarity is an important concept in time series analysis. It means that the statistical properties of a time series do not change over time.Additionally, stationarity is important because many useful analytical tools and statistical tests and models rely on it. The stationarity in the data set is checked using Augmented Dickey Fuller test and Kwiatkowski-Phillips-Schmidt-Shin test. .

2.3.2 Stochastic in time series

A Stochastic process is a statistical phenomenon, it has the collection of random variables ordered in time. To analyse time series, a stochastic process is considered as a model.This process is considered to generate an infinite collection of all possible observed time series. Every member of the collection is the realization of the stochastic process.

2.3.3 Deterministic in time series

A time series is one which can be expressed explicitly by an analytic expression. It has no random or probabilistic aspects.In deterministic time series, past and future values are specific to values of a given time.

2.3.4 Linearity in time series

A model is linear if the current value of the series is a linear function of past observations. In the literature, Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) are two linear time series models as described in section 2.4. A model is non-linear, reflecting a nonlinear function of the past observations.To predict volatility changes in time series, non linear models are used.

2.4 LINEAR Time series models

To design the covariance structure in time series, linear time series models are used. This leads to autoregressive moving average models, formed by two sub-models such as: (a) the autoregressive and (a) the moving average models.

2.4.1 Autoregressive Model

This model uses the past behavior of variable y(t), to predict its future behaviour.The prediction of y(t+1) is considered by adding the weighted values, that y(t) took place in the previous time period along with error term $\epsilon(t)$.

Autoregression model

In general , the observation at various time slots, such as t-1,t-2....t-n, impacts the time period t. In such case, the coefficient

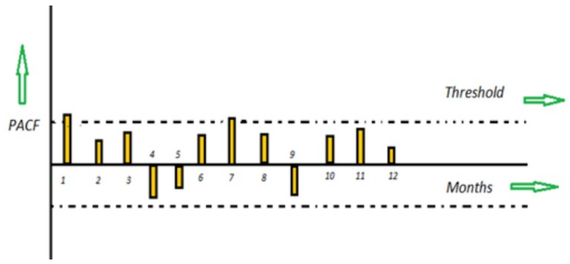

Figure 2.1. Model components

factor is used at the particular period of time to find the impact of previous time slots.
For example, the new iPhone model 'I' created by apple depends on all the previous models in the time series.So in general, Auto Regression (AR) model is defined in such a way that the model calculates the regression of past time series and calculates the present or future values in the series.

For example, we can predict the value for the next time step (t+1) given the observations at the last two time steps (t-1 and t-2). As a regression model, this would look as follows:

$$X(t+1) = b0 + b1 * X(t-1) + b2 * X(t-2)$$

Additionally, the model is noted with its order, p as the parameter. Here example: AR(p),when p=1 $\rightarrow$ AF(1), is called the first order auto regression model.This AR method is suitable for univariate time series without trend and seasonal components.

Moving average model

Here external factors play an important role in impacting time period t under different various slots t-1,t-2,t-3... In short, these external factors are termed as errors or residuals.The previous time spots impacts are calculated by the coefficient factors at the particular time slot.

For example, the stock market of the company 'I' depends on any merger with other companies that happened overnight or the company's demolition due to money factors such as bankruptcy and so on.

So, henceforth in a time series, this moving average model calculates the external factors namely errors or residuals of past and in turn calculates the present or future values in the series as shown below.

$$X_t = \alpha_1 * \epsilon_{t-1} + \alpha_2 * \epsilon_{t-2} + ...\alpha_k * \epsilon_{t-k}$$

Additionally, the model is noted with its order, q as the parameter. Here example: MA(q),when q=1 $\rightarrow$ MA(1), is called the first order moving average model.This method is suitable for univariate time series without trend and seasonal components

Auto Regressive Moving Average

It is also called the ARMA model, which is the combination of both auto regression(AR) and moving average(MA) models.So here, to forecast the future values, the impact of previous lags along with errors or residuals are considered here.

So henceforth let α and β are considered the coefficients of the MA and AR model . The representation would be as follows:

$$X_t = \beta_1 * \times_{t-1} + \alpha_1 * \epsilon_{t-1} + \beta_2 * \times_{t-2} + \\ \alpha_2 * \epsilon_{t-2} + ...\beta_k * \times_{t-k} + \alpha_k * \epsilon_{t-k}$$

Additionally, the model is noted with its order, p as the parameter for auto regression model AR(p) and q as the parameter for moving average model MA(q) respectively.For example, ARMA(p,q), when p=1 , q=1 $\rightarrow$ ARMA(1,1), is called the first order auto regressive moving average model. This model is suitable for univariate time series without trend and seasonal components.

2.4.2 Auto-Regressive Integrated Moving Average

First and foremost thing to be considered in time series is the stationary. So , we need to convert the series to stationary time series. Here comes the picture of an integrated method usage, which can be achieved by subtracting t-n value from time series t , where n = 1,2,3...This is also called a differencing method and based on the n value, it is called as first order differencing with n=1. When n=2, it is called as second order differencing and so on. In the end , the main goal is to achieve stationary time series.

So, ARIMA model is similar to ARMA model, except the fact it includes the integrated or differencing term 'I'. ARIMA model combines both AR and MA models as well as a integrated/differencing preliminary processing step of the sequence to make the time sequence stationary, This process or parameter is called as Integration(I).Here it also considers the previous lags along with residual errors for future forecasting.

Additionally, the model is noted with its order, p as the parameter for auto regression model AR(p) and q as the parameter for moving average model MA(q) and I as the parameter for integration or differencing (I) respectively.For example, ARMA(p,q,I),when p=1 , q=1 , I=1 $\rightarrow$ ARMA(1,1,1), is called the first order auto regressive moving average differencing model.This model is suitable for univariate time series with trend and without seasonal components.

Example : ARIMA(0, 1, 0) model (or I(1) model) is given by $X_t = X_{t-1} + \varepsilon_t X_t = X_{t-1} + \varepsilon_t{}^{-} whichissimplyarandomwalk.$

2.4.3 Seasonal Autoregressive Integrated Moving-Average

Sarimax stands for Seasonal Autoregressive Integrated Moving average.It is an updated version of ARIMA model.SARIMAX includes seasonal effects and eXogenous factors along with ARIMA components in the model. So basically, with prior time steps,SARIMAX model produced the output as the linear function of the difference in observations and seasonal observations, errors and seasonal errors.

This model holds a seasonal pattern,with its relation with external factors. This speciality in the model differs from other models. For instance, with respect to climatic time series, the temperature holds seasonal relationships such as low in winter or high in summers. Additionally this also includes external factors like humidity, temperature in winter is increased due to rain, with the chance of low temperature. So prediction of exact value with these factors would not be possible, if there is no cyclical or seasonal behavior. So comparatively other models can deal with this type of data.

The sarima configuration is simple with three trend elements as such :

1. Trend autoregression order(p)
2. Trend difference order (d)
3. Trend moving average order (q)

Similarly, with respect to seasonal components, there are four elements as such:

1. Seasonal autoregressive order(P)

2. Seasonal difference order(D)
3. Seasonal moving average order(Q)
4. Time steps , considered for single seasonal period(m)

So in a nutshell, Sarima holds the below syntax:

$$SARIMA(p, d, q)(P, D, Q, m)$$

Example: SARIMA(1,0,1)(1,2,2,6) here, m= 6 depicts, in a year there are 6 months seasonality, and this m value has a huge impact on PDQ too. In other words, seasonal offset observation in a model is seen with respect to P values.Additionally D=2,leads to the calculation of second order seasonal difference and Q=2, would lead to calculation of second order errors in the model.

2.4.4 Box-Jenkins Methodology

The Box-Jenkins method named and developed by two statisticians George Box and Gwilym Jenkins, is widely used in time series analysis.It's basically used in ARMA or ARIMA model to find the best fit value of time series model using the past values, in the time series data.The Box-Jenkins model is a forecasting methodology, which uses regression analysis on time series. In short, the Box-Jenkins is a mathematical model, created to do forecasting of particular range from a time series data.This model can also use different types of time series data for forecasting purposes. One such example of Box-Jenkins is the ARIMA, to predict outcome/ forecast it uses difference between data points. With respect to the trends it uses AR, MV and seasonal differencing for forecasting. Basically, it is created based on the knowledge of the occurrences which have happened in the past to influence the future values. There are some drawbacks too with time limit as this box-jenkins model is only used for forecasting events with time frames equal to one and half year or less than that.
The model approach follows three main steps :

1. **The model identification or selection** : First and primary goal is to make time series data stationary. by identifying the seasonality in the data-set. This can be achieved by checking the ACF and PACF plots respectively. These plots will give a clear picture of which techniques Auto regression (AR) or Moving Average(MA) should be used.

2. **Estimation of parameters** : This deals with computation algorithms such as maximum likelihood estimation or non-linear least square estimation, to reach or attain the coefficients to get the best fit of the model.

3. **Model checking and forecasting**: Our final goal is to get the stationary univariate process.So this step deals with checking the model to reach our final goal. So here comes the residuals of the model to the picture,these residuals hold a relationship with time with constant mean and variance. Additionally, these residuals should be independent of each other too. If the parameter estimation is not great, the steps have to be carried out to create a better model.**Null hypothesis**: White noise residuals are seen.

Auto correlation and partial auto correlation

ACF and PCF plots are used in time series data for forecasting purposes. These plots help us to find the relationship between the observations of time series at present with the previous lags.

In short, ACF plot tells the correlation relationship of time series data and its lagged values. This plot is a bar chart which holds the correlation coefficients of the respectives.This bar chart is plotted with confidence bands. So basically it tells us how the present value in the time series is correlated with past values 'n' with n = 1,2,3...k.

Adding to that, time series has general components like residual, trend, seasonality, cyclic and this ACF will take into consideration these components, while finding the correlations. So all this is put together to form an auto-correlation plot.Dealing with ACF plot, the x-axis deals with the coefficients of correlation and y-axis deals with the past lag values. If the two time series values of time t are y(t-1) and y(t), so to calculate lag 3 , it must be the correlation coefficient between y(t) and y(t 3). In statistics, correlation is defined as the relationship of two variables. Each variable takes up the Gaussian distribution. but usually the relationship between the two variables is calculated using Pearson correlation coefficients, which holds the values from -1(negative correlation) and 1(positive correlation) or 0(no correlation). So with the observations of the present time series, we can calculate the correlation relationship with the observations of previous time series , which are called time steps or lags basically. Auto correlation function is defined as the plot of the time series auto correlation by lag.

Partial auto correlation function often termed as PACF, explains the correlation relationship between time series and lags in a partial way.Here residuals which are nothing but the leftover , caused by previous lags. In comparison with ACF, these residuals form partial correlation with the future lag values, as the old found variations are already removed before finding the next correlation. This forms a linear regression analysis, in such a way x(t) is calculated from x(t-1),x(t-2) and x(t-3). Additionally, if there is any missed out residual information, that can be modeled by immediate next lag, So in order to get a good correlation , we can use the feature of immediate next lag, during modeling. One more important issue which arises here is multicollinearity, and it arises when many features are correlated. It is good practice to avoid such multicollinearity. So only the need of features which are relevant , will be considered.

2.5 FB prophet

Facebook prophet developed internally by Facebook corporation, for time series forecasting works. It is an open source platform made available to the public from 2017.It was developed by Sean J.Taylor and Ben Letham in order to overcome few issues such as inflexibility, extra assumptions and less robustness with respect to existing forecasting methodologies. With much robust tendendancy ,Facebook prophet is able to produce high quality forecasts related to business and overcome many forecasting problems as well.So basically Facebook prophet takes a novel approach and predicts forecasting using probabilistic techniques and additive models, and the final outcome is curve fitting exercise. Facebook prophet is mainly for business needs with seasonality and general knowledge is required on the events which holds a huge impact on the data set. Such events include black Friday, Easter holidays, some paid promotional events or the launch of new products in that particular week and so on. In a nutshell , Facebook prophet uses additive components, to model the time series data.

$$y(t) = g(t) + h(t) + s(t) + \epsilon(t)$$

The above equation is called prophet model components.

Explanation of the model components:

1. y(t) : time series prediction

2. g(t) : This is called a trend function. Its aim is to model non-periodic changes using a piece wise linear regression model or linear saturation growth model. This can be tuned using hyper parameter tuning.

3. h(t) : This denotes the holidays or special events which impact the data set. This can be tuned using hyper parameter tuning.

4. s(t) : This denotes the seasonality in monthly or yearly or weekly or daily or hourly. This factor models the periodic changes in the time series value.Fourier transform plays an important role for modeling and additionally we can use our own seasonalities using hyper parameter tuning.

5. $\epsilon(t)$: This is the noise component, which affects the model.

This is basically an intuitive approach, which leads to better and conceptual understanding to create predictions. Prophet is mainly designed for below business perspectives:

1. It is mainly used for time series data based on additive models.

2. Full year for historical data, with data being recorded on an hourly, weekly or daily basis.

3. Seasonality plays an important role, and it should be seen on an hourly or weekly or daily basis.

4. Outliers or unavailable data.

5. Irregularities are seen in holiday events which do not necessarily have seasonality.

6. Severe trend changes such as trend within the data set.

Facebook prophet tuning

Despite the fact, Facebook prophet provides high business forecasts with much accuracy, but customisation can be carried out in terms of adjusting or tuning Facebook prophet. Prophet is quite easy with hyper parameter tuning, so additional regressors, seasonalities or special events can be added in hyper parameter tuning.

From references[4], prophet follows additive regressive model, which holds several components such as a linear or logistics trend curve, a seasonality curve with daily;annual and weekly basis,holiday or special event curve, Re sampled user specified curves such as hourly, quarterly and so. For instance, modeling an online sales retail store for four consecutive stores, from 2002, January 1 through the end of 2003. In that case , the overall constant increase of trend with respect to time is visible from the sales of 1000 count per day to 1800 count in the time period. With respect to sales, in spring it's 50 units above average and 50 units below average in autumn season. Sales were lesser Tuesday to all days, every week and sudden rapid increase on Saturday. Throughout the hours of day, sales increase in the afternoon and fall to the lowest at the midnight.Below is the model component as shown below.

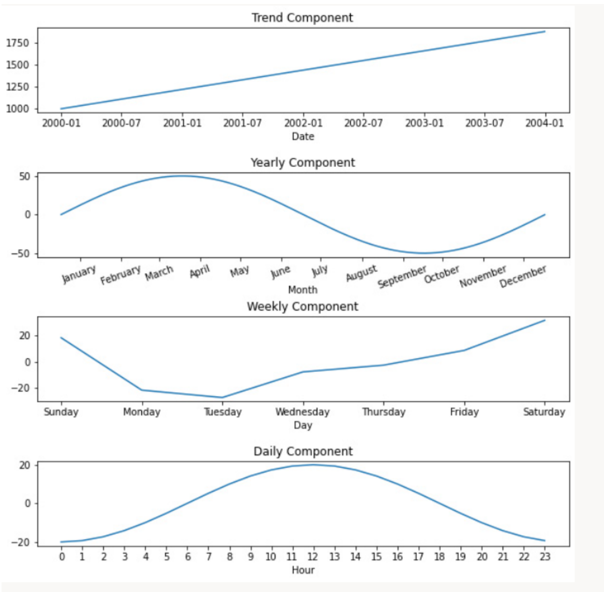

Figure 2.2. Model components

So basically the additive model results in the addition of all the curves to provide the final model at the end of year. When the components or sub components are added up , the final graph would be more complex as below. The graph displays the weekly and daily variations of the first year and the curve finally ends for four years.

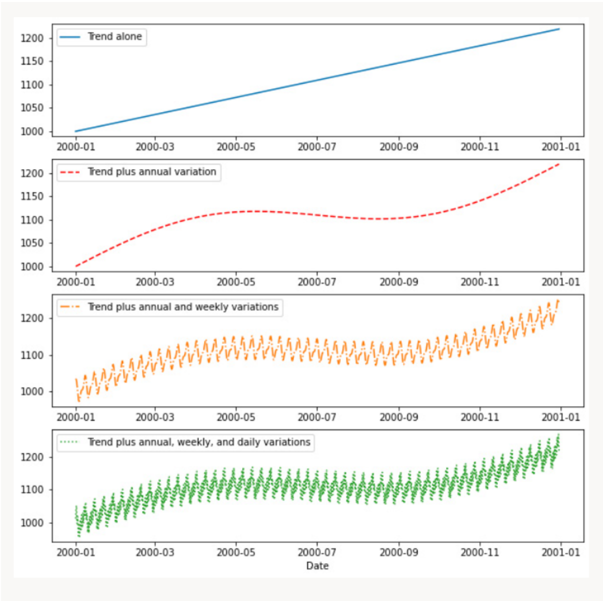

Figure 2.3. Model components

Stan and its usage in Facebook prophet

Stan is used in Facebook prophet library. In general Stan is a platform for statistical modeling and computation with high performance. Usage of Stan includes fields such as modeling with statistical data, prediction analysis in various aspects such as business,engineering, biological, social, physical and most widely used in data analysis. Stan's mathematical library provides differentiable probability function and linear algebra.It also interfaces with popular programming language in the field of data analysis such as (R, Python, shell, MAT LAB, Julia, Stata) and runs on all major platforms (Linux, Mac, Windows). Stan allows Prophet to create uncertainty levels for future predictions to create the data-driven estimate of forecasting risk and it is possible with Bayesian statistics.

Advantages of Facebook prophet:

1. Facebook Prophet Prophet is used to produce reliable forecast prediction and to meet deadlines and achieve goals and much better performance compared to other models. High-speed forecasting is done with the help of Stan library.

2. Forecast is done without manual intervention and it is done completely in an automation model. Facebook prophet is more accurate and robust in fields of outliers, missing data or dramatic changes in the time series data set.

3. With Facebook prophet, the prediction can be tun able in such a way to get the best fit model. Additionally it requires little domain knowledge in order to achieve the forecast.

4. The library of Facebook prophet is used in other R and python , at the end the model is fitted using Stan library.

5. Irregularities are seen in holiday events which do not necessarily have seasonality.

6. Severe trend changes such as trends within the data set.

7. Facebook prophet achieves typical and good results within a short span of time and it's no way lesser than complicated forecasting techniques with respect to performance. But in order to achieve better performance, hyper parameter tuning is considered.

2.6 Neural networks for time series forecasting

In today's world neural networks play an important role,as we are dealing with enormous amounts of data. This huge data set is called big data. There are various models including traditional and deep learning models to analyse and handle these data. In the modern world, alongside computer vision, and the internet of things, neural network models play a vital role in the various fields such as spam-detection, anomaly detection, time series forecasting , future predictions, fraud analysis, medical diagnosis and so on. These deep learning models tend to over-perform traditional models in an effective way in terms of result,speed and performance. These models can be used in both univariate and multivariate time series analysis. Below are the few reasons why deep learning models should be considered in time series forecasting:

1. Deep Learning can deal with huge amounts of data and they are fed into algorithms on further steps.

2. Deep learning uses an end-to-end approach to solve the problems.

3. Deep learning performs well in fields such as computer vision, natural language processing, image classification, speech recognition, time series forecasting and so on.

4. In terms of time series forecasting, recurrent neural networks (deep learning framework) is used, which in turn makes reliable predictions on time series data and problems.

5. Deep learning supports multiple inputs and outputs.

6. As we all know, they are structured and unstructured data and Deep learning has the capacity to learn and extract features automatically from structured and unstructured data set.

7. Deep learning provides more accuracy in terms of prediction and forecasting.

2.6.1 Recurrent Neural Network

Recurrent Neural Network belongs to the class of Artificial Neural Network is often termed as auto associative or feedback network. Here the units are connected to form a directed cycle, leading to showcase the dynamic temporal behavior. Generally feed forward networks deal with data points, which are independent of each other.RNN supports long term dependencies.In such case, to work on data which has a sequence or holds a dependency comes into the picture of RNN. There involve with data sequences where each input is not mapped to output , instead it creates a function to map input over time to single output. In a nutshell, RNN has the information of input or previous state in the form of memory in order to create the next output in the sequence. Advantages of RNN includes the functionality of handling sequence data, input can be of different lengths, it has the tendency to store the information of previous state. Disadvantages of RNN includes for decision making it does not consider future inputs,vanishing gradient problem and slow computation time. Another important point is the usage of activation functions. It is generally defined as how the weighted sum of the input is transferred into output from a single or multiple nodes in a network layer. There are basically three activation functions used in RNN architecture:

1. TANH : $1/(1+e^{-x})$
2. SIGMOID : $e^x - e^{-x}/e^x + e^{-x}$
3. RELU : max(0,x)

The selection of activation functions has a huge impact on the performance and outcome capability of neural networks and different activation functions can be used in different parts in neural network models.

There are different types of RNN:

1. One to One : Single input x_t produces single output y_t
2. One to Many : Single input x_t produces multi output (y_{t0}, y_{t1}, y_{t2})
3. Many to One : Many input (x_t, x_{t+1}, x_{t+2}) produces single output y_t
4. Many to One : Many input (x_t, x_{t+1}, x_{t+2}) produces many output (y_{t0}, y_{t1}, y_{t2})

One such main problem with RNN is the vanishing gradient descent, which clogs the long data sequences. Usually the information from the past input is carried to the future output and hence the gradients carry this information and there is an update visible in the RNN parameters. So usually, when the gradients become very low, the updation of parameters becomes unusual, at the end leading to the path of non-learning of the network. To overcome this error, LSTM has been implemented.

Types of RNN Architecture:

1. Bidirectional recurrent neural networks (BRNN): Inputs from future time steps are used to improve the network's accuracy.

2. Gated Recurrent Units (GRU): They mainly work on vanishing gradient problems and have reset and update gate.These gates mainly help to consider which data has to be kept for the future.

3. They mainly work on vanishing gradient problems and have three gates namely input, output and forget gates. And these gates decide which information has to be kept for the future.

2.6.2 Long-Short term Memory

Long short term Memory is short formed as LSTM, is a kind of recurrent neural network(RNN), which can learn long-term dependencies. They work mainly on vanishing gradient problems and they have three gates namely input, output and forget gates. LSTM can save the back propagation error, through each and every layers in the network across the time, without losing the information.The cell plays an important role in saving and releasing information as they decide what to store, when to store, how much to store and what to release, when to release and how much to release. As we know the information is the key in the data, and these gates in LSTM learns by itself the allowance of these information in three main process such as entering, leaving and or deletion. All these process happen in an iterative way of back-propagation error, making decisions, with the help of gradient descent adjusting weights.

From reference[] The above figure is the propagation process

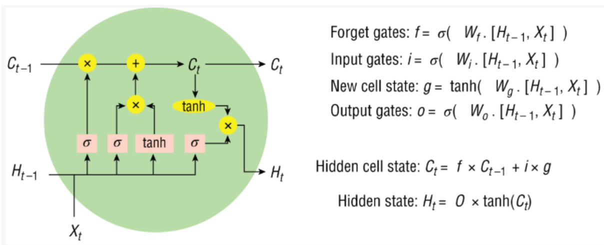

Figure 2.4. LSTM Backpropogation

in recurrent neural networks to compute gradient values. So the memory cells in LSTM involve both addition and multiplication process, with respect to the transformation of both input and information. So here comes the flow chain of LSTM working, as the addition process protects the constant error during the backpropagation event. So in order to create the following state with respect to a cell , usually LSTM considers the multiplication process of the current state with new input, but instead addition of the current state with new input takes place, while the last cell 'forget' cell still completely depends on the multiplication process. So from the figure, it is evident that the input state helps in cell state updating. Here comes the picture of long term dependencies, where the information from the current state previous state which is in hidden mode will be passed to the next state through an activation function. Here the 'sigmoid' activation function is used. This sigmoid function has its value in the range 0-1. Here values equal to 1 or closer to 1 are considered to be remembered, whereas values equal to 0 or close to 0 are meant to be forgotten. LSTM also has a forget gate, and its task is to help in the decision making strategy of which information has to be forgotten and which information has to be remembered and transferred to another state. As stated clearly, forget gates also help to transfer information to final output from long distance events. So from the graph we have our third gate,the output gate.And its ultimate goal is to find which hidden state should be. Initially the previous hidden state and present input state has been passed to sigmoid function , which is followed by the passing of updated new cell state to the tanh function. So in order to make the decision of which information the hidden state should keep, the outputs of sigmoid and tanh function are multiplied respectively.

Chapter 3

Implementation

3.1 Methodolgy

So basically the ultimate goal of the thesis is to develop a forecasting model. This model helps the client to plan and identify seasonality, annual patterns, production capacity, and expansion over the period based on requirements. This in turn leads to the development of business strategy (e.g., plans to launch a facility or store internationally and expand into new markets)

External factors are also considered like weather conditions, football events and so on to check the water demand forecasting in longer or shorter term. So, at the end, there are two kinds of analysis such as univariate time series analysis and multivariate time series analysis.

Our final goal is do prediction on the data set in three ways as below

1. **Hourly:**This includes hourly prediction on the data set for both the DMA's namely DMA 225 and DMA 468. Data is re-sampled to daily basis and prediction is carried out for each hour in a day.It is achieved using univariate and multivariate time series analysis.

2. **Weekly:**This includes hourly prediction on the data set for both the DMA's namely DMA 225 and DMA 468.Data is re-sampled to daily basis and prediction is carried out for each week. It is achieved using univariate time series analysis.

3. **Monthly:**This includes hourly prediction on the data set for both the DMA's namely DMA 225 and DMA 468.Data

is re-sampled to daily basis and prediction is carried out for each month. It is achieved using univariate time series analysis.

So based on the requirement,two methodologies have carried out as below:

1. **Univariate Time Series Analysis:** The term ‘univariate’ implies that forecasting is based on a sample of time series observations of the data.They deal with only one variable, along with the timestamp. And we need to do prediction or forecast on this only one variable. The task of this ideology involves playing with data set and aim to produce visualization and data plots,comparison of traditional and deep learning models to perform, the forecast prediction on the data set, The analysis is carried out by checking the accuracy metric mean square error.Prediction is done in weekly, hourly and monthly basis on both DMA’s 225 and 468 respectively.

2. **Multivariate Time Series Analysis:**The term ’multivariate’ implies that forecasting is based on a sample of time series observations of the multiple variables in the data set. In general they contain multiple variables, along with timestamps. And we need to do prediction or forecast on these multiple variables. External factors such as weather conditions or football events or so on to perform the water demand prediction.With respect to football events, we need to check if the football match has happened on the day and date in that location corresponding to the water demand data set. So, this would follow classification time series analy sis.The task of this ideology involves playing with data set and aim to produce visualization and data plots,comparison of traditional and deep learning models to perform, the forecast prediction on the data set, The analysis is carried out by checking the accuracy metric mean square error. With respect to external factors weather conditions are considered here. Prediction is done on Hourly basis.

3.2 Data collection

The below steps describes the data collection part of univariate and multivariate time series analysis.

1. **Univariate Time Series Analysis:** To perform univariate time series analysis Yorkshire data set is considered. This Yorkshire data set is a water data set which records flows of water every 15 minutes to each of the Distribution Management Areas across the region (DMA's).There are many DMA's found, but we consider two DMA's namely DMA 225 and DMA-468, in HU1 location of the country.This data set contains the flows in l/s for each of the 2000 DMAs that contain domestic properties A breakdown of properties by Postal Sector is given for each of the DMAs in the ex tract The dataset follows univariate time series .analysis as there each DMA's records the water demand value for the time.The Dataset has one year of data, from April to march mid with regular time intervals of 15 minutes. Here outliers are visualized with respect to the water dataset,but forecast is done without outlier removal to check the efficiency of the models.

2. **Multivariate Time Series Analysis:** To perform multivariate time series analysis, we combined Yorkshire dataset and Weather dataset.The Weather dataset is taken as the external weather factors, to perform water demand prediction of the DMA at the time interval. WorldWeatherOnline is a company providing accurate and relevant weather data. Their primary weather API types are local weather API, historical weather API, marine, sailing and surfing weather API.Specifically,their historical weather API provides hourly past weather for worldwide locations since July 2008.This dataset follows multivariate time series analysis.The Dataset has three months of data, from April mid to July mid with regular time intervals of 15 minutes.Here outliers are visualised with respect to water dataset,but forecast is done without outlier removal to check the efficiency of the models.

3.3 Data Analysis

This section involves the data analysis and observatory part.

3.3.1 Univariate Time Series Analysis Hourly Prediction

This subsection involves the data analysis part of water flows with respect to two DMA's, with corresponding timestamp.

DMA 225 univariate analysis hourly prediction

Dataset plot

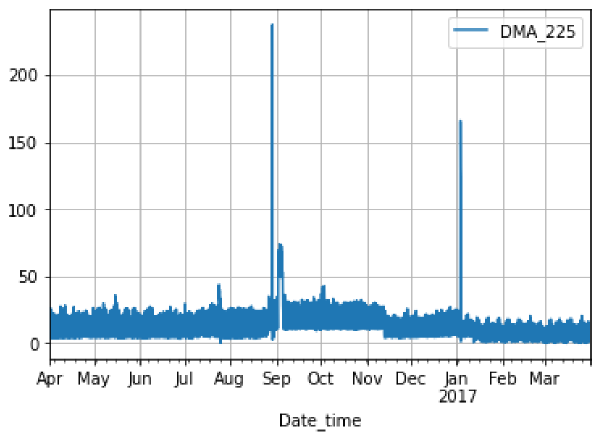

Figure 3.1. DMA-225-univariate analysis

Stationarity check

Stationarity check plays an important role in time series data.There are two hypotheses namely Null hypothesis (H0) and Alternate hypothesis (H1).These two are checked in the tests mentioned below.

(a) **Augmented Dickey-Fuller test** Null Hypothesis (H0): If failed to be rejected, meaning it is non-stationary. It has some time dependent structure. Alternate Hypothesis (H1): The null hypothesis is rejected; meaning it is stationary. It does not have a time-dependent structure.

Analysis is carried on two factors. First factor is the ADF statistic value. It is -7. The more negative this statistic, the more likely we are to reject the null hypothesis (we have a stationary data-set). Second factor is p value - 0.000058 < 0.05 .Hence these two values prove data is stationary.

(b) **Kwiatkowski-Phillips-Schmidt-Shin test:**Null Hypothesis (H0): stationary. Alternate Hypothesis (H1): non-stationary Analysis is carried on KPSS test value.It's 4.3. The positive value is a statistic, and hence we have a stationary dataset.

DMA 468 univariate analysis hourly prediction

Dataset plot

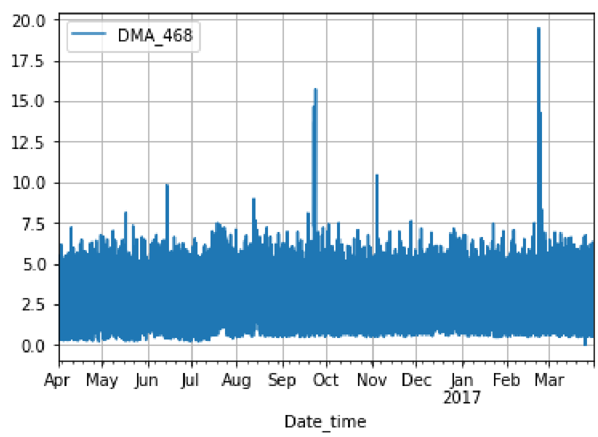

Figure 3.2. DMA-468-univariate analysis

Stationarity check

Stationarity check plays an important role in time series data.There are two hypotheses namely Null hypothesis (H0) and Alternate hypothesis (H1).These two are checked in the tests mentioned below.

(a) **Augmented Dickey-Fuller test** Null Hypothesis (H0): If failed to be rejected, meaning it is non-stationary. It has some time dependent structure. Alternate Hypothesis (H1): The null hypothesis is rejected; meaning it is stationary. It does not have a time-dependent structure.
Analysis is carried on two factors. First factor is the ADF statistic value. It is -11. The more negative

this statistic, the more likely we are to reject the null hypothesis (we have a stationary data-set). Second factor is p value - $0.000058 < 0.05$ .Hence these two values prove data is stationary.

(b) **Kwiatkowski-Phillips-Schmidt-Shin test:**Null Hypothesis (H0): stationary. Alternate Hypothesis (H1): non-stationary analysis is carried out on KPSS test value.It's 0.2. The positive value is a statistic, and hence we have a stationary dataset.

3.3.2 Univariate Time Series Analysis Weekly Prediction

DMA 225 univariate analysis weekly prediction

Time Series decomposition

The below plot depicts the time series decomposition of the DMA 225, on daily re sampling. This time series decomposition will provide the picture of components such as trend, seasonality and residuals.Data is stationary with ADF static : -2 .81 and KPSS value: 0.63.

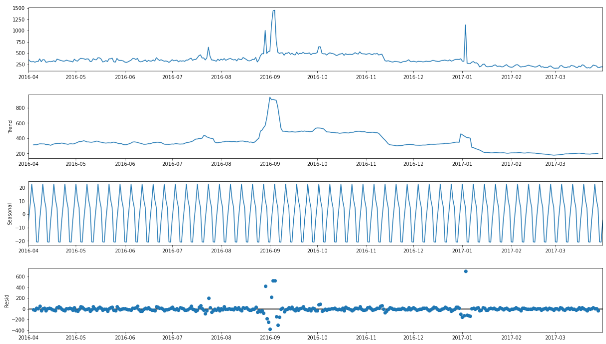

Figure 3.3. DMA-225-Time series decomposition

DMA 468 univariate analysis weekly prediction

Time Series decomposition

The below plot depicts the time series decomposition of the DMA 468, on daily re sampling. This time series decomposition will provide the picture of components such as trend, seasonality and residuals.Data is stationary with ADF static : -2 .81 and KPSS value: 0.63.Data is stationary with ADF static : -5.8 and KPSS value: 0.14

Figure 3.4. DMA-468-Time series decomposition

3.3.3 Univariate Time Series Analysis Monthly Prediction

DMA 225 univariate analysis monthly prediction

Time Series decomposition Data is resampled to daily basis, so the final shape of the dataset would be (365,1).Data is stationary with ADF static : -2.8 and KPSS value: 0.6.

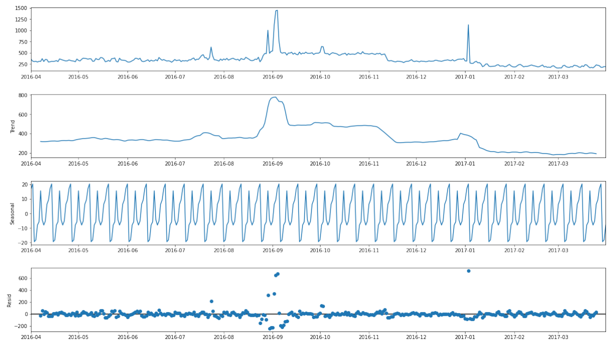

Figure 3.5. DMA-225-Time series decomposition

DMA 468 univariate analysis monthly prediction

Time Series decomposition Time Series decomposition Data is re sampled to daily basis, so the final shape of dataset would be (365,1).Data is stationary with ADF stats : -5.8 and KPSS stats : 0.14.

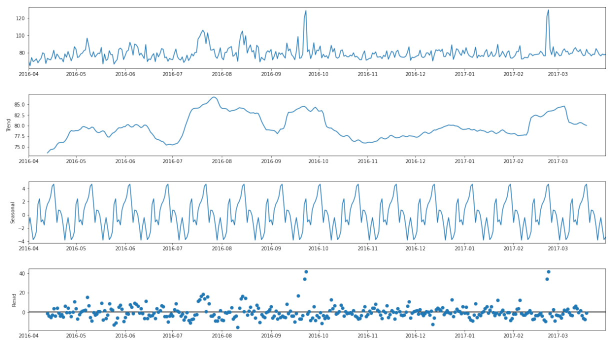

Figure 3.6. DMA-468-Time series decomposition

3.3.4 Multivariate Time Series Analysis hourly prediction

This subsection involves the data analysis part of water flows with respect to two DMA's and weather variables with corresponding timestamp.

Dataset plot of DMA 225 multivariate analysis and DMA 468 multivariate analysis

The below graph contains the multi variables, both weather and DMA values with respect to timestamp. The weather variables include tempC, HeatIndexC, FeelsLikeC,WindChillC, wind speed mph,sunHour,precipMM, humidity.

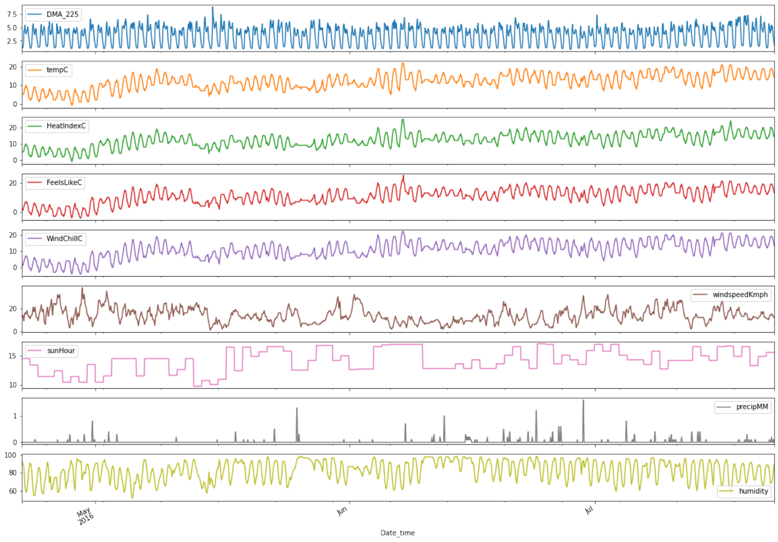

Figure 3.7. DMA-225-multivariate

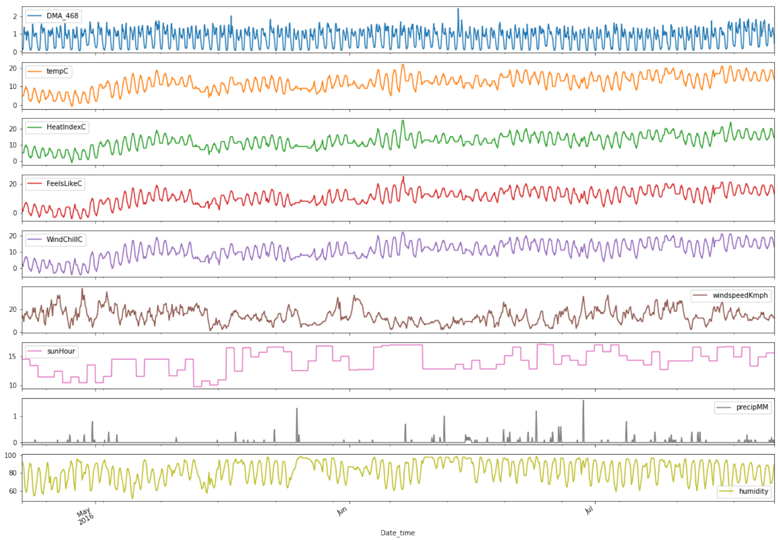

Figure 3.8. DMA-468-multivariate

3.4 Experimental Models

The below section describes the models performed for the carry out univariate and multivariate time series forecast.

(a) Univariate Time Series Analysis - hourly prediction
(b) Univariate Time Series Analysis - weekly prediction
(c) Univariate Time Series Analysis - monthly prediction
(d) Multivariate Time Series Analysis - hourly prediction

3.4.1 Univariate Time Series Analysis - hourly prediction

This includes prediction models implemented on two DMA's namely DMA 225 and DMA 468 respectively on hourly basis. Data is re sampled to hourly basis and prediction was carried out.

DMA 225 Univariate Analysis:HOURLY

To perform univariate analysis on DMA 225 , we considered 87% train set and remaining portion of data to test set.we proposed LSTM and facebook prophet on this data proportion and calculated the prediction using root mean square error metrics.

LSTM Model

We used the tensor flow library to perform LSTM model on the DMA 225 dataset. The ultimate goal of this model is to calculate the hourly predictions.In LSTM model, the look-back period is considered as 24 hours and output is predicted for 24 hours.So it follows supervised learning problem approach. We obtained the rmse score of 1.4367 as the result.

Facebook Prophet Model

We used Facebook prophet library on DMA-225 dataset.The ultimate goal of this model is to calculate the hourly predictions.We obtained the rmse score of 4.227 as the result.

Plots of LSTM and Facebook prophet - DMA 225 hourly

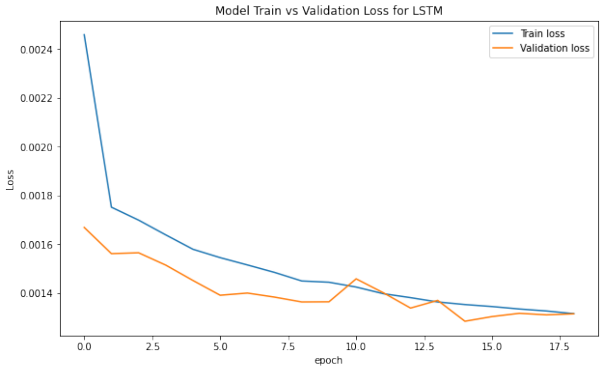

Figure 3.9. LSTM : Loss graph- DMA 225 univariate hourly prediction

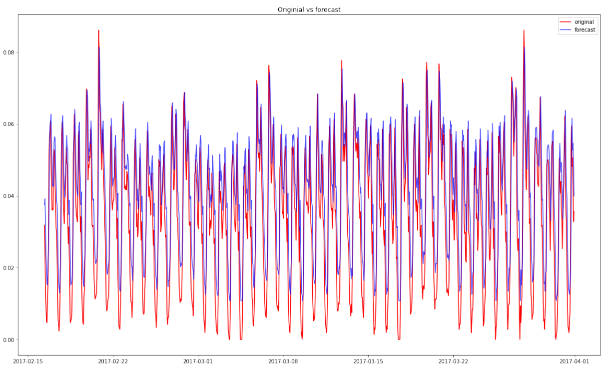

Figure 3.10. LSTM :original vs predicted-DMA 225 univariate hourly prediction

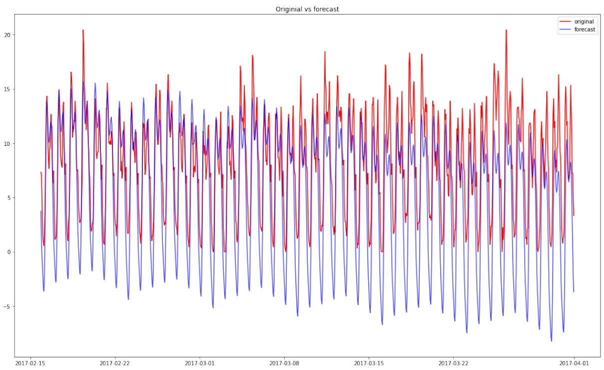

Figure 3.11. Facebook Prophet :original vs predicted-DMA 225 univariate hourly prediction

DMA 468 Univariate Analysis : HOURLY

To perform univariate analysis on DMA 468 , we considered 87% train set and remaining portion of data to test set.we proposed LSTM and fb prophet on this data proportion and calculated the prediction using root mean square error metrics.

LSTM Model

We used the tensorflow library to perform LSTM model on the DMA 468 dataset. The ultimate goal of this model is to calculate the hourly predictions.In LSTM model, the look-back period is considered as 12 hours and output is predicted for 24 hours.So it follows supervised learning problem approach. We obtained the rmse score of 0.0565 as the result.

Facebook Prophet Model

We used Facebook prophet library on DMA-468 dataset.The ultimate goal of this model is to calculate the hourly predictions.We obtained the rmse score of 1.2 as the result.

Plots of LSTM and Facebook prophet - DMA 468 hourly

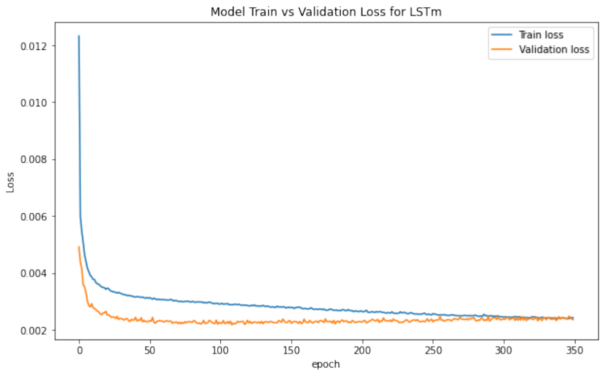

Figure 3.12. LSTM : Loss graph - DMA 468 univariate hourly prediction

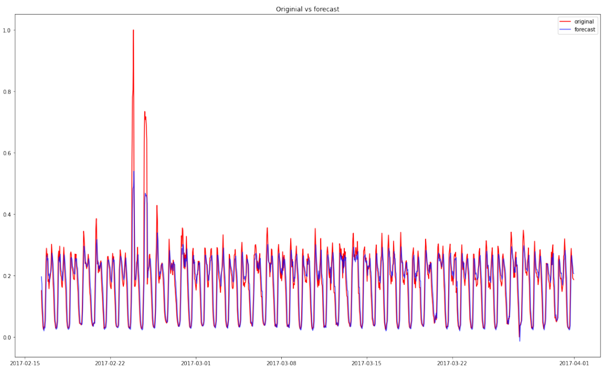

Figure 3.13. LSTM :original vs predicted - DMA 468 univariate hourly prediction

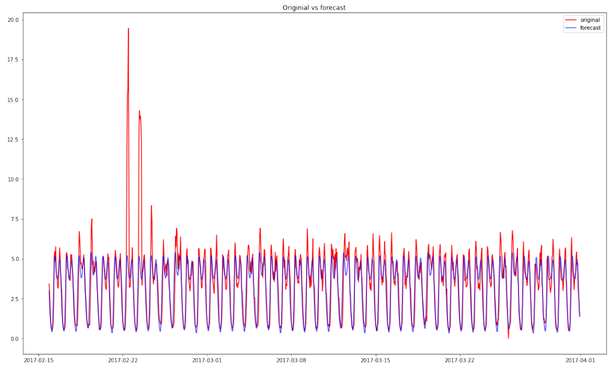

Figure 3.14. Facebook Prophet :original vs predicted- DMA 468 univariate hourly prediction

3.4.2 Univariate Time Series Analysis - weekly prediction

This includes prediction models implemented on two DMA's namely DMA 225 and DMA 468 respectively on weekly basis. Data is resampled to daily basis and prediction was carried out.Models such as LSTM , sarimax and baseline are carried out.

DMA 225 Univariate Analysis : Weekly

To perform univariate analysis on DMA 225 is resampled to daily basis and the length of dataset is 365. 70% of dataset is considered as train set and 30% is considered as test set.we proposed LSTM and Sarimax and baseline persistance on this data proportion and calculated the prediction using root mean square error metrics.

Baseline Model

Here simple baseline persistance model is carried out. The rmse value of baseline model is 113.697.

Sarimax Model We used sarimax on DMA-225 dataset. Performed grid search parameters using train set to get best parameters and using those we calculated weekly prediction, with root mean square value 132.542 as the result.

LSTM Model We used LSTM on DMA-225 dataset. Here lookback period of 3 days is considered to perform the weekly prediction. Hence finally achieved root mean square value 91.41 as the result.

Plots of Baseline, Sarimax and LSTM - DMA 225 weekly

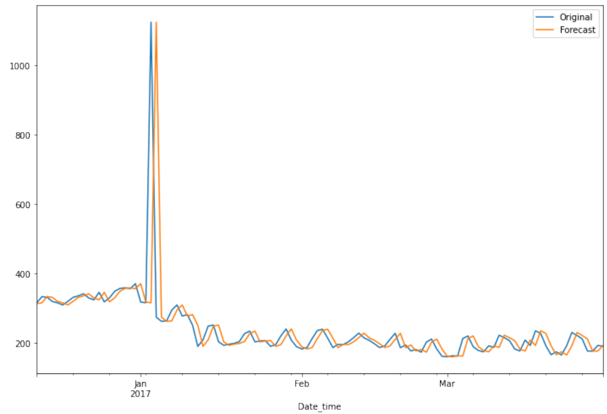

Figure 3.15. Baseline Persistence : Original vs predicted - DMA 225 univariate weekly prediction

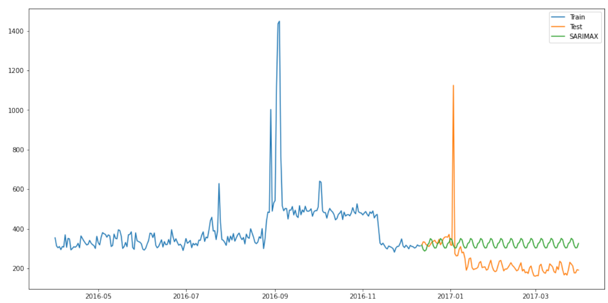

Figure 3.16. Sarimax :DMA 225 univariate weekly prediction

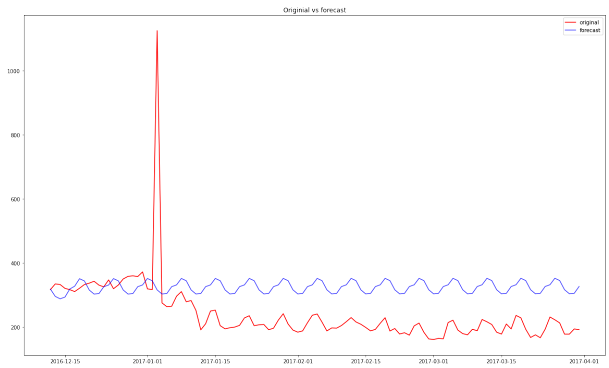

Figure 3.17. Sarimax :original vs predicted-DMA 225 univariate weekly prediction

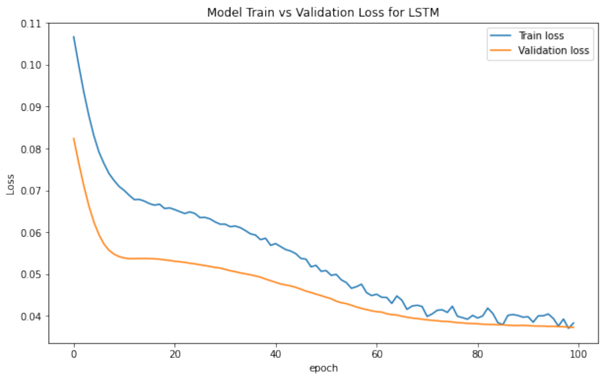

Figure 3.18. LSTM :Loss graph-DMA 225 univariate weekly prediction

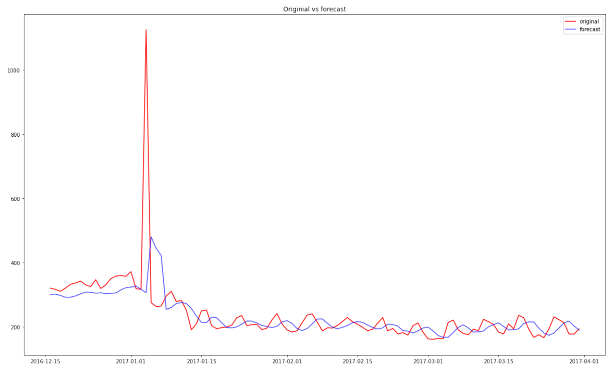

Figure 3.19. LSTM :original vs predicted graph-DMA 225 univariate weekly prediction

DMA 468 Univariate Analysis : Weekly prediction

To perform univariate analysis on DMA 468 is resampled to daily basis and the length of dataset is 365. 70% of dataset is considered as train set and 30% is considered as test set.we proposed LSTM and Sarimax and baseline persistance on this data proportion and calculated the prediction using root mean square error metrics.

<u>Baseline Model</u>

Here simple baseline persistance model is carried out. The rmse value of baseline model is 8.057.

<u>Sarimax Model</u> We used sarimax on DMA-468 dataset. Performed grid search parameters using train set to get best parameters and using those we calculated weekly prediction with root mean square value 7.263 as the result.

<u>LSTM Model</u> We used LSTM on DMA-468 dataset. Here lookback period of 3 days is considered to perform the weekly prediction. Hence finally achieved root mean square value 6.431 as the result.

Plots of Baseline, Sarimax and LSTM - DMA 468 weekly

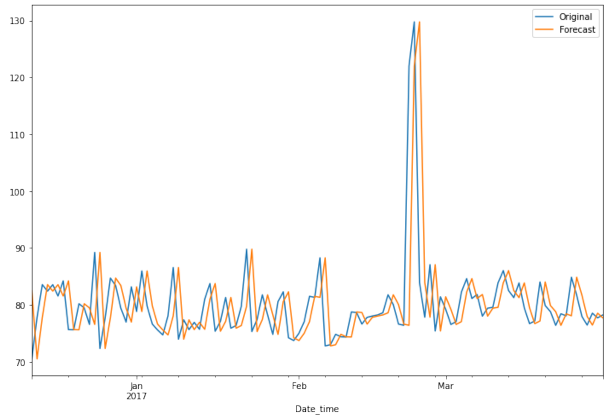

Figure 3.20. Baseline Persistence :Original vs predicted-DMA 468 univariate weekly prediction

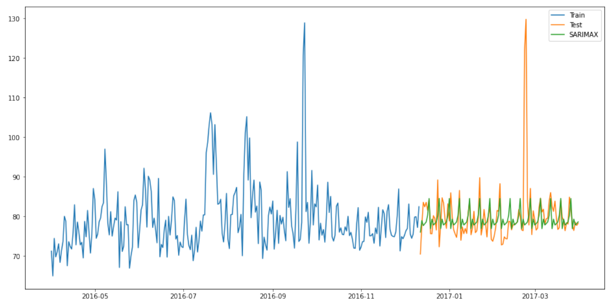

Figure 3.21. Sarimax :DMA 468 univariate weekly prediction

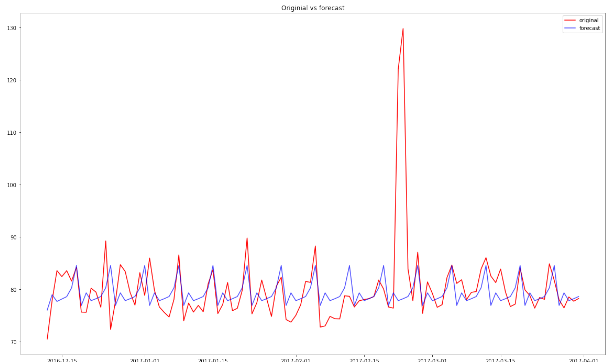

Figure 3.22. Sarimax :Original vs predicted-DMA 468 univariate weekly prediction

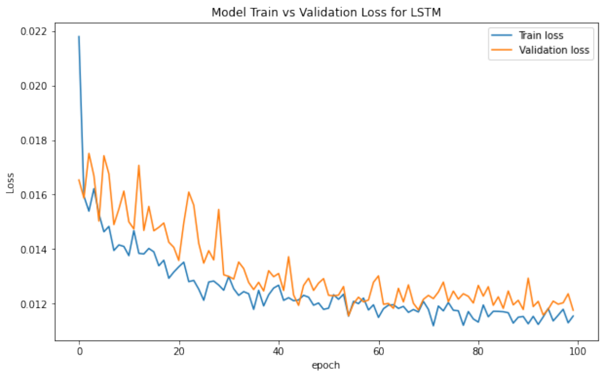

Figure 3.23. LSTM :Loss graph-DMA 468 univariate weekly prediction

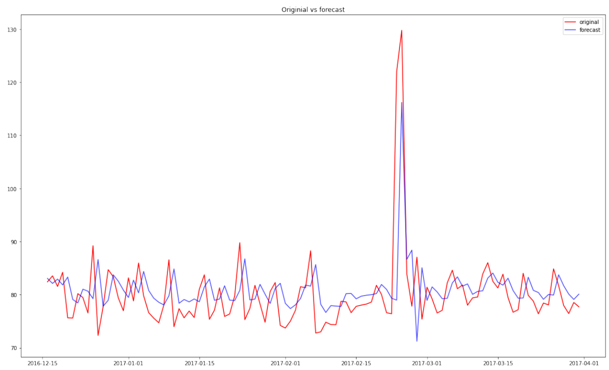

Figure 3.24. LSTM :original vs predicted graph - DMA 468 univariate weekly prediction

3.4.3 Univariate Time Series Analysis - monthly prediction

This includes prediction models implemented on two DMA's namely DMA 225 and DMA 468 respectively on monthly basis. Data is resampled to daily basis and prediction was carried out.Models such as LSTM , sarimax and baseline are carried out.

DMA 225 Univariate Analysis : Monthly prediction

To perform univariate analysis on DMA 225 is resampled to daily basis and the length of dataset is 365. Last 3 months of dataset is considered as test set and remaining is considered as train set.we proposed LSTM and Sarimax and baseline persistance on this data proportion and calculated the prediction using root mean square error metrics.

<u>Baseline Model</u>

Here simple baseline persistance model is carried out. The rmse value of baseline model is 125.565.

<u>Sarimax Model</u> We used sarimax on DMA-225 dataset. We performed grid search on train set to get the parameters and using those parameters we calculated monthly prediction with root mean square value 164.717 as the result.

<u>LSTM Model</u> We used LSTM on DMA-225 dataset. Here lookback period of 27 days is considered to perform the weekly prediction. Hence finally achieved root mean square value 21.772 as the result.

Plots of Sarimax, Baseline and LSTM - DMA 225 monthly

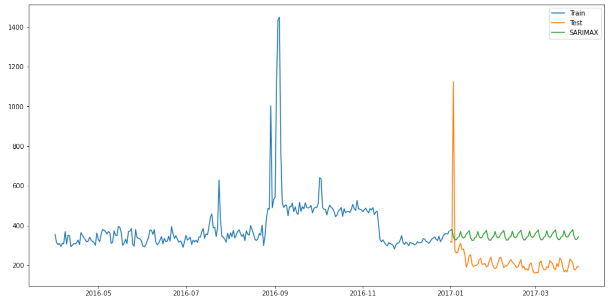

Figure 3.25. Sarimax :DMA 225 univariate monthly prediction

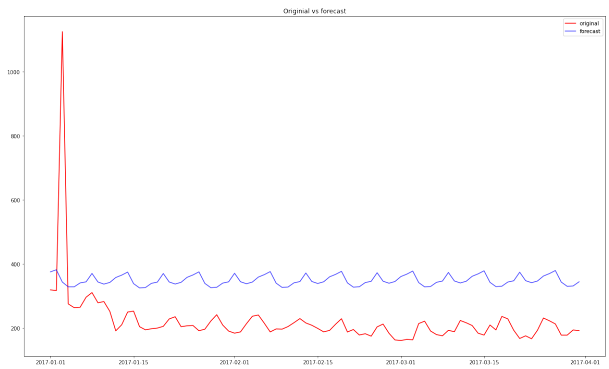

Figure 3.26. Sarimax :original vs predicted-DMA 225 univariate monthly prediction

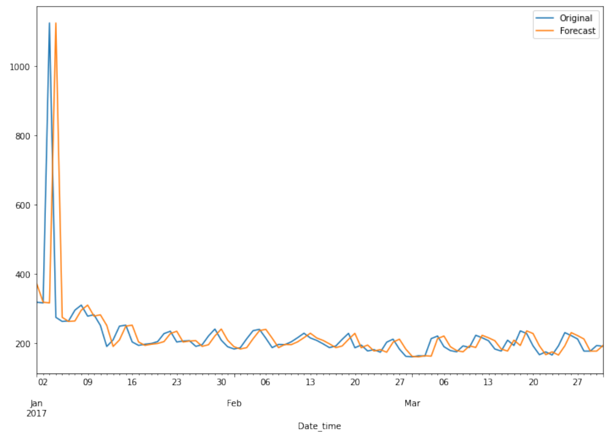

Figure 3.27. Baseline Persistence : Original vs predicted - DMA 225 univariate monthly prediction

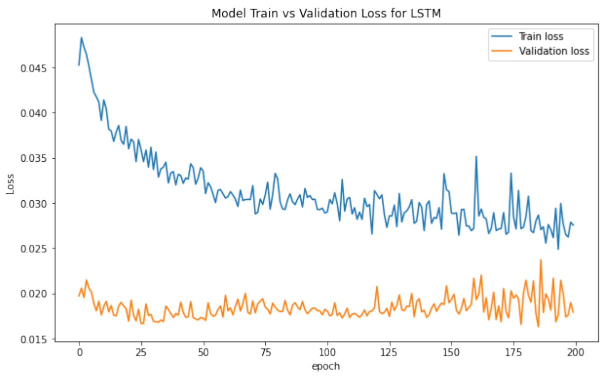

Figure 3.28. LSTM :Loss graph-DMA 225 univariate monthly prediction

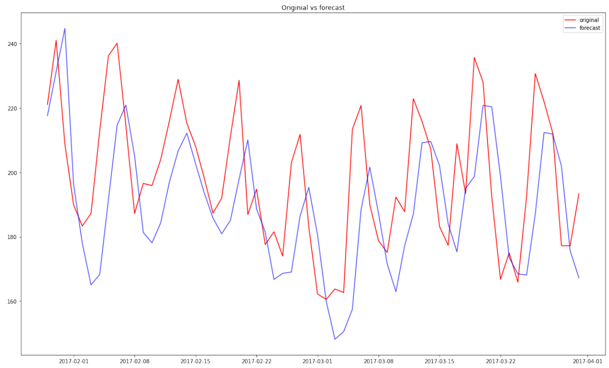

Figure 3.29. LSTM :original vs predicted graph-DMA 225 univariate monthly prediction

DMA 468 Univariate Analysis : Monthly prediction

To perform univariate analysis on DMA 468 is resampled to daily basis and the length of dataset is 365. Last 4 months of dataset is considered as test set and remaining amount of data is considered as train set.we proposed LSTM and Sarimax and baseline persistance on this data proportion and calculated the prediction using root mean square error metrics.

Baseline Model

Here simple baseline persistance model is carried out. The rmse value of baseline model is 7.796.

Sarimax Model We used sarimax on DMA-468 dataset. we performed grid search on train set to get best parameters. Using those we calculated monthly prediction with root mean square value 7.39 as the result.

LSTM Model We used LSTM on DMA-468 dataset. Here lookback period of 27 days is considered to perform the monthly prediction. Hence finally achieved root mean square value 6.836 as the result.

Plots of Baseline, Sarimax and LSTM - DMA 468 weekly

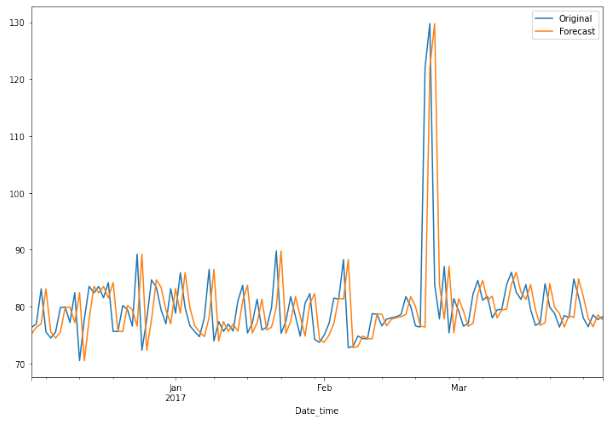

Figure 3.30. Baseline Persistence : Original vs predicted - DMA 468 univariate monthly prediction

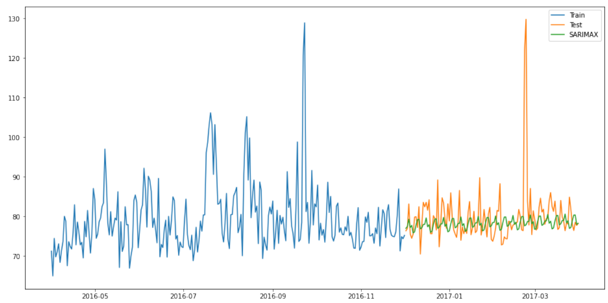

Figure 3.31. Sarimax :DMA 468 univariate monthly prediction

Figure 3.32. Sarimax :original vs predicted-DMA 468 univariate monthly prediction

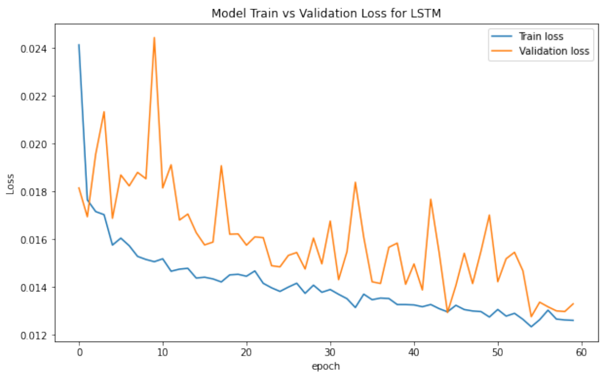

Figure 3.33. LSTM :Loss graph-DMA 468 univariate monthly prediction

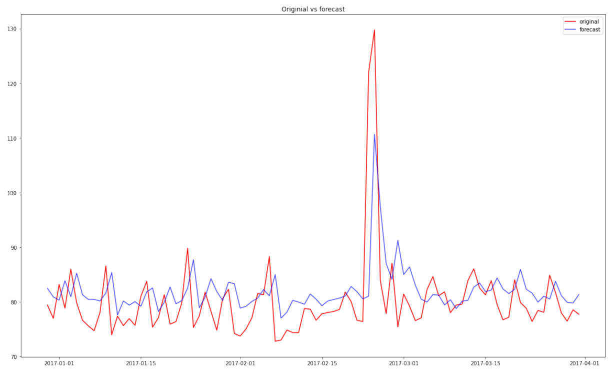

Figure 3.34. LSTM :original vs predicted graph-DMA 468 univariate monthly prediction

3.4.4 Multivariate Time Series Analysis - hourly prediction

This includes prediction models implemented on two DMA's namely DMA 225 and DMA 468 respectively on hourly basis. Data is re-sampled to hourly basis and prediction was carried out. Models such as LSTM , facebook prophet are carried out.

subsubsectionDMA 225 Multivariate Analysis:HOURLY

To perform univariate analysis on DMA 225 , we considered 87% train set and remaining portion of data to test set.we proposed LSTM and fb prophet on this data proportion and calculated the prediction using root mean square error metrics.

LSTM Model

We used the tensorflow library to perform LSTM model on the DMA 225 dataset. The ultimate goal of this model is to calculate the hourly predictions.In LSTM model, the look-back period is considered as 12 hours and output is predicted for 24 hours.So it follows supervised learning problem approach. We obtained the rmse score of 0.0951 as the result.

Facebook Prophet Model

We used Facebook prophet library on DMA-225 dataset.The ultimate goal of this model is to calculate the hourly predictions.We obtained the rmse score of 0.792 as the result.

Plots of LSTM and Facebook prophet - DMA 225 hourly mv

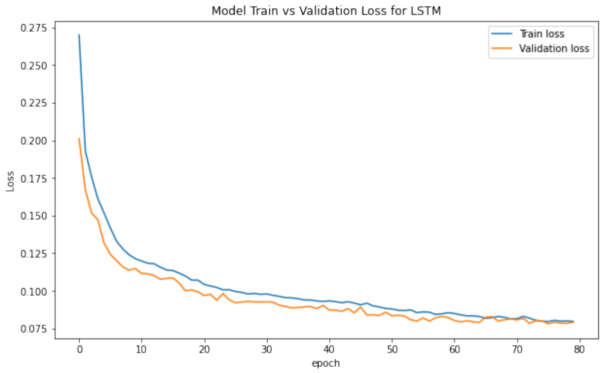

Figure 3.35. LSTM : Loss graph- DMA 225 multivariate hourly prediction

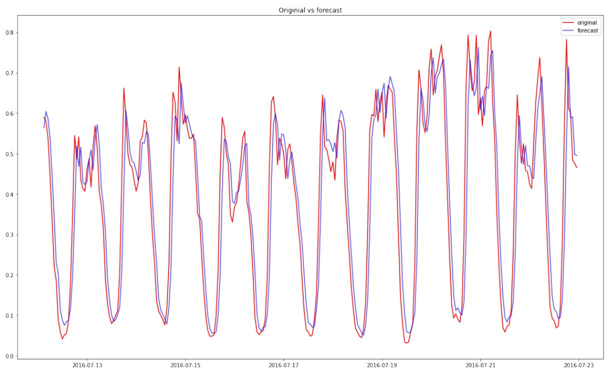

Figure 3.36. LSTM : original vs predicted-DMA 225 multivariate hourly prediction

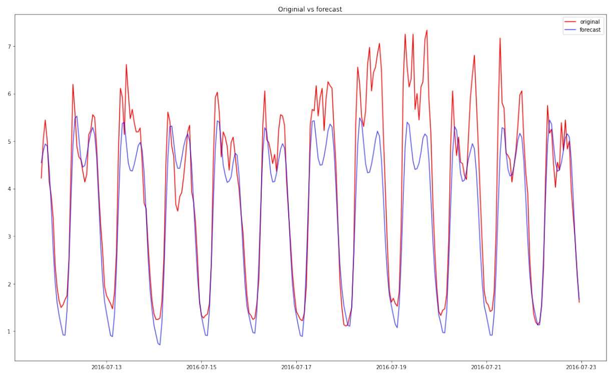

Figure 3.37. Facebook Prophet :original vs predicted-DMA 225 multivariate hourly prediction

DMA 468 Multivariate Analysis:HOURLY

To perform univariate analysis on DMA 468 , we considered 87% train set and remaining portion of data to test set.we proposed LSTM and fb prophet on this data proportion and calculated the prediction using root mean square error metrics.

LSTM Model

We used the tensorflow library to perform LSTM model on the DMA 468 dataset. The ultimate goal of this model is to calculate the hourly predictions.In LSTM model, the look-back period is considered as 12 hours and output is predicted for 24 hours.So it follows supervised learning problem approach. We obtained the rmse score of 0.101 as the result.

Facebook Prophet Model

We used Facebook prophet library on DMA-468 dataset.The ultimate goal of this model is to calculate the hourly predictions.We obtained the rmse score of 0.291 as the result.

Plots of LSTM and Facebook prophet - DMA 468 hourly mv

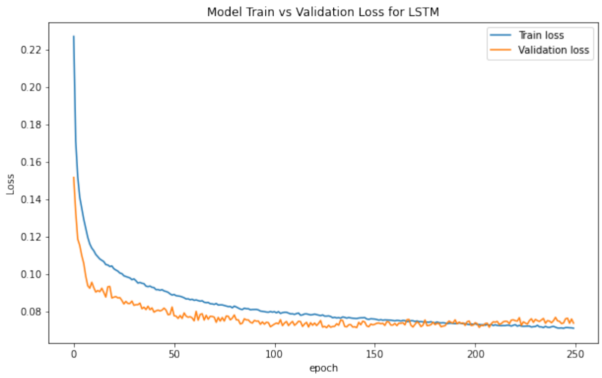

Figure 3.38. LSTM : Loss graph- DMA 468 multivariate hourly prediction

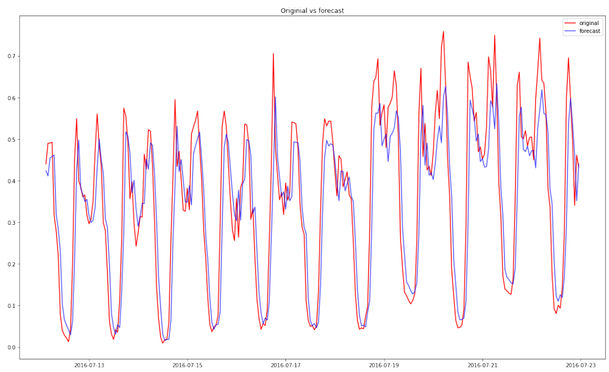

Figure 3.39. LSTM : original vs predicted-DMA 468 multivariate hourly prediction

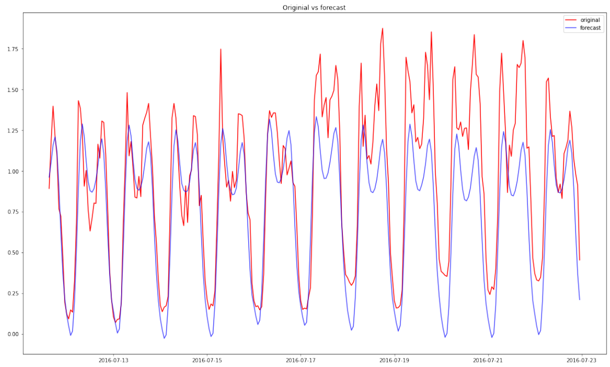

Figure 3.40. Facebook Prophet : original vs predicted-DMA 468 multivariate hourly prediction

Chapter 4

Experimental Results

This chapter describes the results which we achieved using our models. At the end, we performed the below kinds of predictions.

(a) Hourly prediction - univariate and multivariate analysis
(b) Weekly prediction - univariate analysis
(c) Monthly prediction - univariate analysis

4.1 Hourly prediction

This subsection deals with the results of hourly prediction.We performed both univariate and multivariate analysis to achieve hourly prediction.

4.1.1 Univariate analysis : Hourly prediction

DMA	**Fb prophet**	**LSTM**
DMA 225	4.227	0.0109
DMA 468	1.2	0.0565

Discussion of result : The above tabular column states that we considered two DMA's for hourly prediction and here it followed univariate analysis, since the dataset has only the column of target variable. We performed facebook prophet and LSTM on same train- test proportion. Here Root mean square error was considered as accuracy metrics. Compared to fb prophet, LSTM achieved lesser rmse score as 0.1019 and 0.0565 for DMA 225 and DMA 468 respectively.

4.1.2 Multivariate analysis : Hourly prediction

DMA	**Fb prophet**	**LSTM**
DMA 225	0.792	0.0951
DMA 468	0.291	0.1017

Discussion of result : The above tabular column states that we considered two DMA's for hourly prediction and here it followed multivariate analysis, since the dataset has both water demand and weather values for the two DMA's . We performed facebook prophet and LSTM on same train- test proportion. Here Root mean square error was considered as accuracy metrics. Compared to fb prophet, LSTM achieved lesser rmse score as 0.0951 and 0.1017 for DMA 225 and DMA 468 respectively.

4.2 Weekly prediction

This subsection deals with the results of weekly prediction.We performed univariate analysis to achieve weekly prediction.

4.2.1 Univariate analysis : Weekly prediction

DMA	**Baseline**	**Sarimax**	**LSTM**
DMA 225	113.697	132.542	91.412
DMA 468	8.057	7.263	6.431

Discussion of result : The above tabular column states that we considered two DMA's for weekly prediction and here it followed univariate analysis, since the dataset has only the column of target variable. We performed baseline persistence, Sarimax and LSTM models on same train- test proportion. Here Root mean square error was considered as accuracy metrics. LSTM outperformed sarimax and baseline models, with lesser rmse score as 91.412 and 6.431 for DMA 225 and DMA 468 respectively.

4.3 Monthly prediction

This subsection deals with the results of monthly prediction.We performed univariate analysis to achieve monthly prediction.

4.3.1 Univariate analysis : Monthly prediction

DMA	**Baseline**	**Sarimax**	**LSTM**
DMA 225	125.565	164.717	21.772
DMA 468	7.796	7.39	6.836

Discussion of result : The above tabular column states that we considered two DMA's for monthly prediction and here it followed univariate analysis, since the dataset has only the column of target variable. We performed baseline persistence, Sarimax and LSTM models on same train- test proportion. Here Root mean square error was considered as accuracy metrics. LSTM outperformed sarimax and baseline models, with lesser rmse score as 21.772 and 6.836 for DMA 225 and DMA 468 respectively.

Chapter 5

Conclusions

This research project objective was to develop short term water demand forecasting using traditional and deep learning models. This kind of forecasting will help the organisation or client to plan and identify seasonality, annual patterns, production capacity, and expansion over the period based on requirements. This in turn contributes the advent effect of business strategy(e.g., plans to launch a facility or store internationally and expand into new markets). Here we considered weather data set as external factors and we figured out if it has an adverse impact on water demand forecasting on the particular date-time. In a nutshell we performed hourly , weekly and monthly predition using univariate and mutlivariate analysis strategies. Models such as LSTM(Long short-term memory), Fb prophet(facebook prophet), Sarimax (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors), Baseline(Persistence model) were carried out. LSTM outperformed rest of the models, with respect to hourly, weekly and monthly predictions.

From the experiments and results, it 's evident that deep learning models performs better than traditional models. We used tensor-flow and fb prophet library to carry out the short term water demand forecasting with deep learning techniques, whereas we used statsmodel sarimax library, to work with traditional models. The best model is decided based on rmse square value.In Sarimax, a popular statistical method it used only past informationt to predict future. In Facebook prophet, it performed better with strong seasonal effects.It is also an additive model where non linear trends are used. Compared to those two models,LSTM performed better with lesser rmse values eventually. The reason for lower rmse values is because of its capability to use previous sequential data, by this way the model has the ability to learn from long term observations of sequence data. This will pave the way for time series forecasting.Additionally,lstm can work with long term seasonality such as yearly pattern and short term seasonality such as weekly patterns. And one more good thing with lstm, is it accepts inputs of different length and thats how it helps clients/ organisation to plan business strategy.

In future research, we hope to implement and consider different models, to perform short term / long term water demand forecasting with some other external factors. This will leads the client/organisation to plan business strategy in an effective way.

Bibliography

To add bibliography