Assessment of Forecasting Strategies on Univariate Time Series Data

IIT2016039 Sai Charan Teja
IIT2016040 Anmol Singh Sethi
IIT2016042 Manavdeep Singh

Introduction

Time Series: A time series is a series of data points indexed in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time.

Load Forecasting: Load forecasting is a technique used by power or energy-providing companies to predict the power/energy needed to meet the demand and supply equilibrium

Need: Accurate models for electric power load forecasting are essential to the operation and planning of a utility.

Types: Very Short term load forecasting, Short term load forecasting, Medium term load forecasting, Long term load forecasting

Benchmark Models: Auto Regression(AR), Moving Averages(MA), Auto Regression Integrated Moving Averages(ARIMA), Auto Regression Moving Averages(ARMA), etc.

Problem Statement

Using a Non-Conventional Technique - Long Short Term Memory (LSTM) on Univariate Time Series data from different sources and compare the findings with benchmark models like Auto Regression(AR), Auto Regression Integrated Moving Averages(ARIMA) to predict Electricity Load demands. Where Auto Regression and Auto Regression Integrated Moving Averages are statistical models for time series data, LSTM is a machine learning approach to the same.

To validate the output of LSTM model in a generic way, we have taken 5 different data-sets and made expansive experiments to determine an efficient method of solving the problem of Electricity Demand Load Prediction.

Data Sets Used

FRED ECONOMIC DATA

Data set for industrial production provided by Board of Governors of the Federal Reserve System US. It has monthly frequency.

OPEN POWER SYSTEM DATA

This is a time series data-set giving load, wind and solar, prices in hourly manner. The data is available for 37 European countries. It is provided by Open Knowledge Foundation. Column used is *GB_EAW_load_actual_tso*

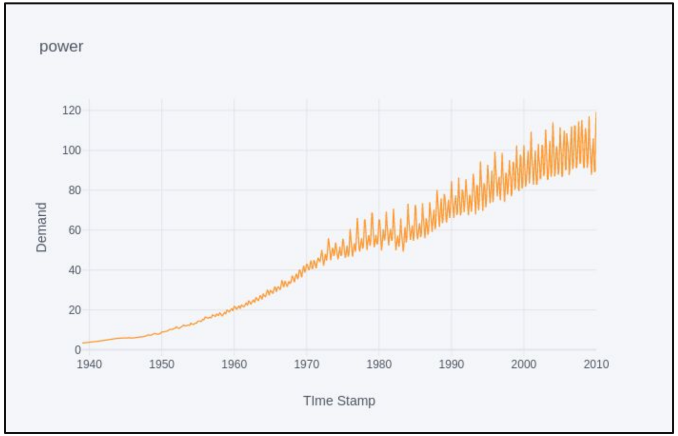


Fig: FRED Data

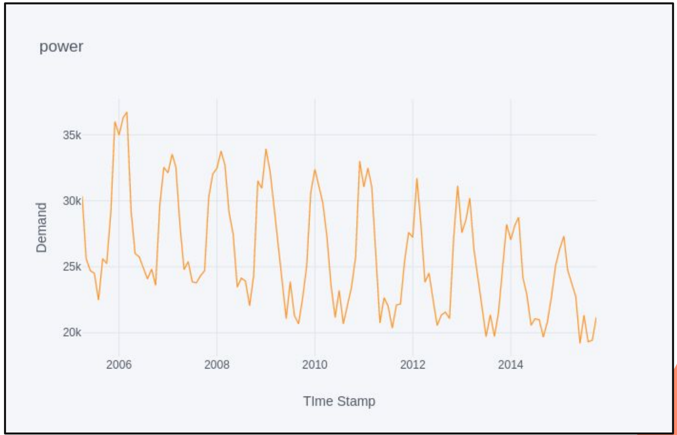


Fig: OPSD Data

SMARD

A data set provided by Federal Network Agency, Germany, contains the electricity market data. The frequency of the data is daily. The data set contains the value of load for every 15 minutes per day.

UC, IRVINE DATA

hedata-set contains 2075259 instances for one household with one minute sampling located in France for a period of 47 months. Nearly 1.25 percent of the values are missing in the data-set.

UTTAR PRADESH STATE LOAD DISPATCH DATA

This is a real time data provided by Uttar State Load Dispatch Centre. The data has been web-scrapped for a period of 1 day.

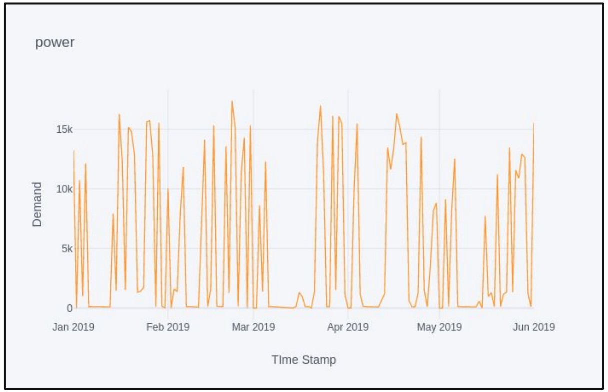


Fig: SMARD Data

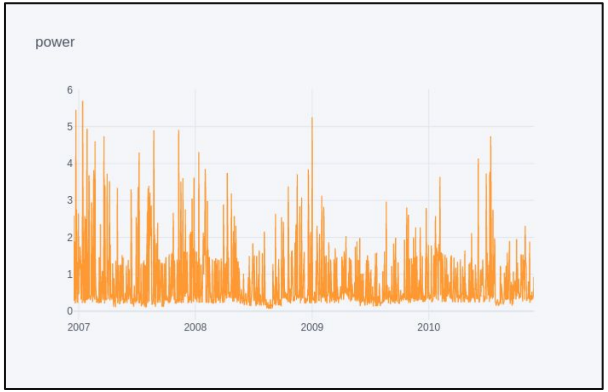


Fig: UCI Data

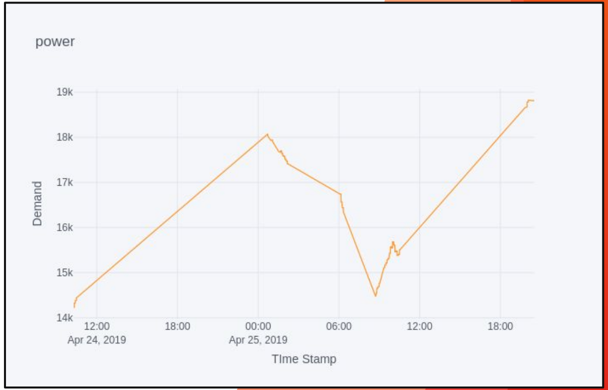


Fig: UPSLDC Data

Literature Review

Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market, Gao Gao, Kwoklun Lo, Fulin Fan, 2017

introduces the models of autoregressive integrated moving average (**ARIMA**) and artificial neural network (**ANN**) which are applied to the price forecasts for up to 3 steps **8 weeks** ahead in the **UK electricity market**. compared in terms of the **root mean square errors** (RMSEs) of price forecasts.[2]

Short Term Load Forecasting Using Time Series Analysis: A Case Study for Karnataka, India, Nataraja.C, M.B.Gorawar, Shilpa.G.N., Shri Harsha.J, 2012

development of Short Term Load Forecasting Models Using Time series Analysis for **Karnataka Demand** and hence comparison of different models, such as **AR**, **ARMA**, **ARIMA**, with errors of 13.03%, 11.03%, 6.15% respectively.[3]

Study & Development of Short Term Load Forecasting Models Using Stochastic Time Series Analysis, V.Venkatesh, Shilpa G N ,Nataraja.C, 2014

study & development of various time series models for Short Term Electrical Load Forecasting Using Time series approach. one year load data, first six months data is used for model development. The methodology involves Initial **Model Development Phase, Parameter Tuning Phase and Forecasting Phase**

Methodology

STATISTICAL APPROACH:

AUTO REGRESSION

Auto regression is a special type of regression in which prediction of a variable is done with help of the past or previous values of the same variable.

For example we can predict the price of gold annually or we can predict the load of electricity over a period of time based on their previous values.

$$Price(t+1) = a_0 + a_1 * Price(t-1) + a_2 * Price(t-2)$$

ARIMA

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

There are seasonal and Non-seasonal ARIMA models that can be used for forecasting.

SEASONAL ARIMA

As the name suggests, this model is used when the time series exhibits seasonality. This model is similar to ARIMA models, we just have to add in a few parameters to account for the seasons. We write SARIMA as ARIMA(p,d,q)(P, D, Q)m,

p — the number of autoregressive terms

d — degree of differencing

q — the number of moving average terms

m — refers to the number of periods in each season

(P, D, Q)— represents the (p,d,q) for the seasonal part of the time series

Data-Set	p	d	q	P	D	Q	m
FRED	2	1	2	1	0	0	12
OPSD	1	0	0	2	0	0	12
SMARD	0	0	0	0	0	0	1
UCI	0	1	1	0	0	1	24
UPSLDC	0	1	0	0	0	0	24

Parameter values for our ARIMA models

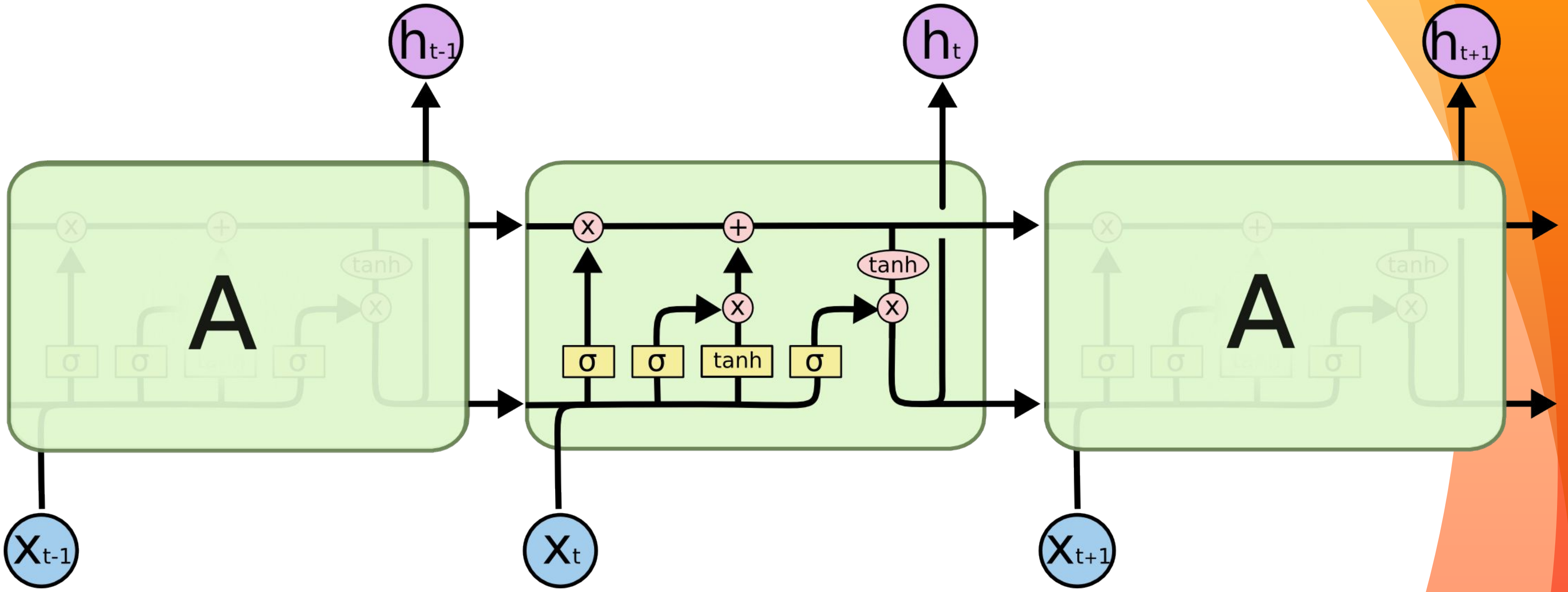
MACHINE LEARNING(NEURAL NETWORKS) APPROACH:

LSTM

LSTM Neural Networks, which stand for Long Short-Term Memory, are a particular type of recurrent neural networks.

LSTM networks have some internal contextual state cells that act as long-term or short-term memory cells.

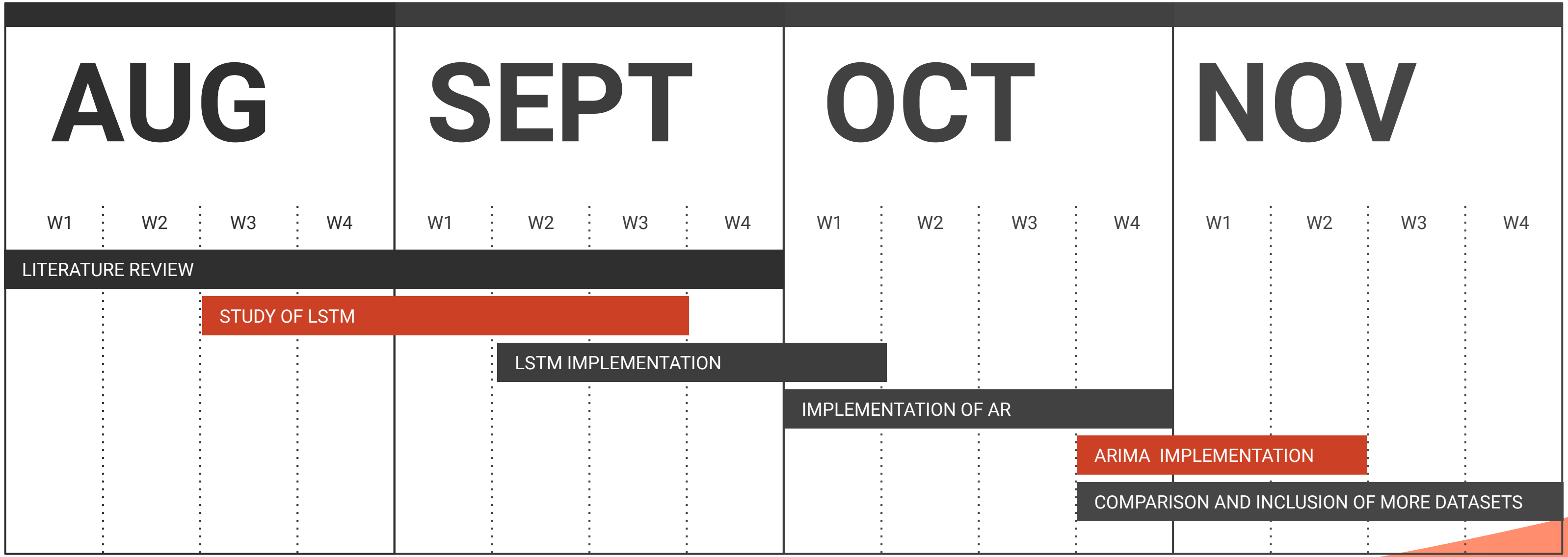
The output of the LSTM network is modulated by the state of these cells.



LSTM Repeating Module

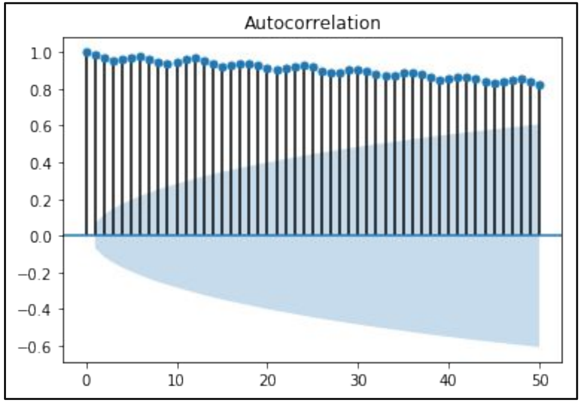
Our LSTM Pipeline

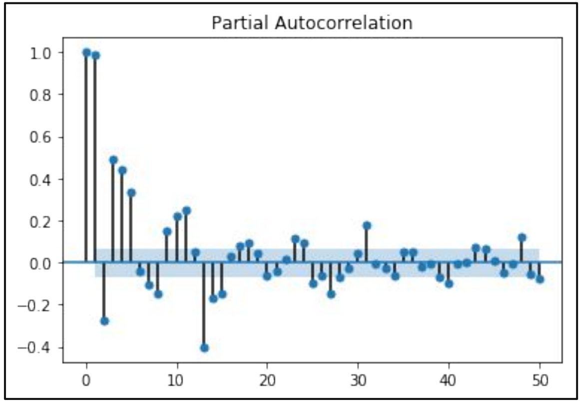
Implementation Timeline



Results

FRED Data





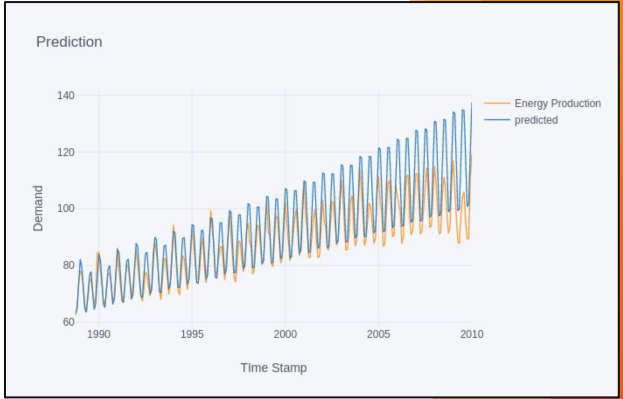


Fig: AR

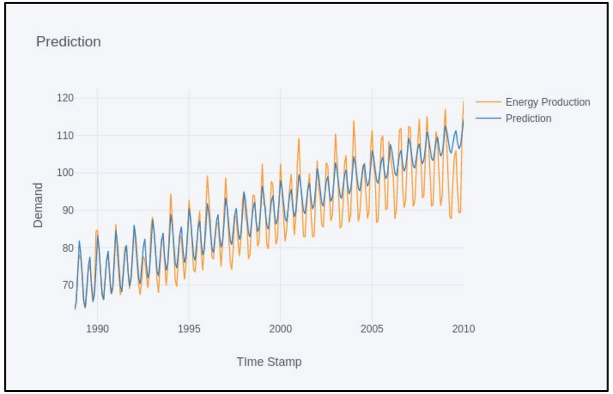


Fig: ARIMA

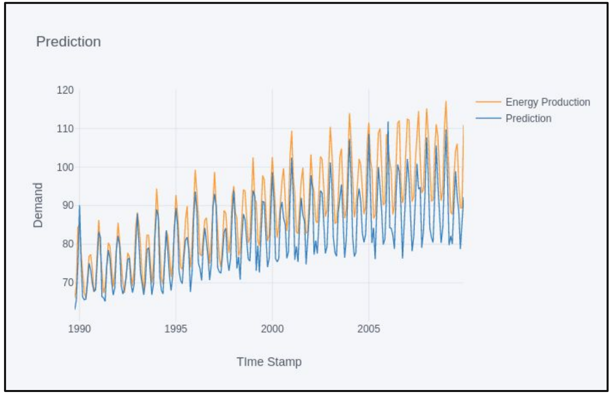
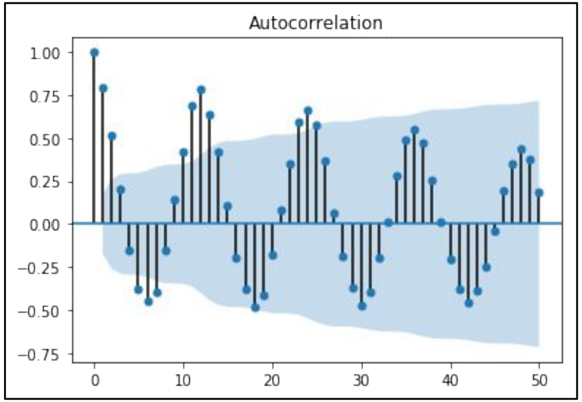
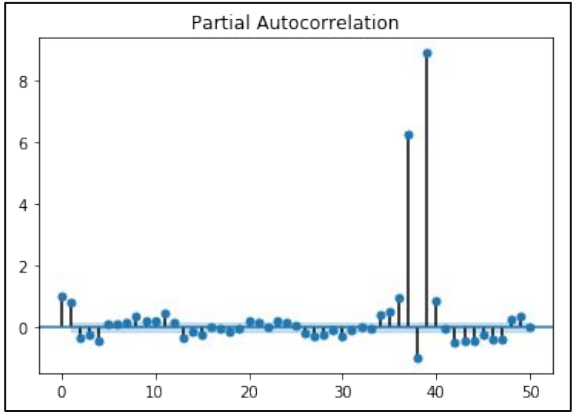


Fig: LSTM

	AR	ARIMA	LSTM
RMSE	8.24	5.19	7.99
R2	0.55	0.82	0.36

OPSD Data





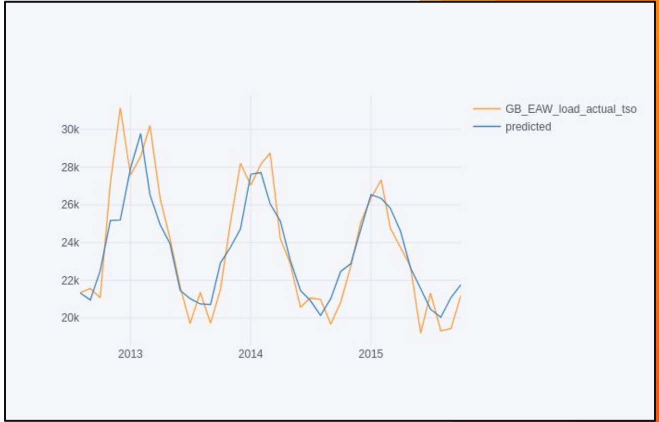


Fig: AR

Fig: ARIMA

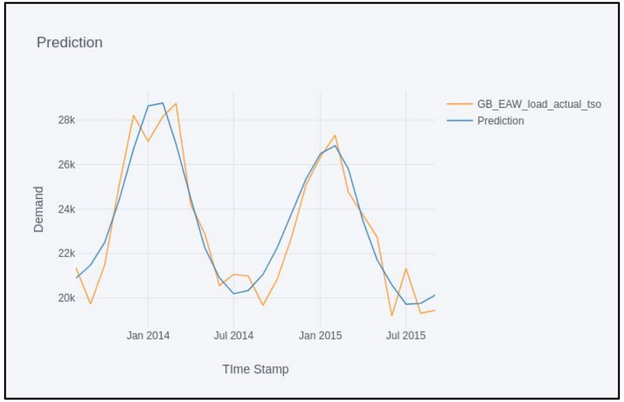
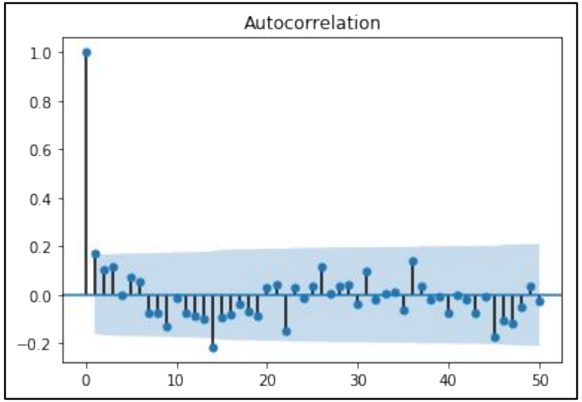
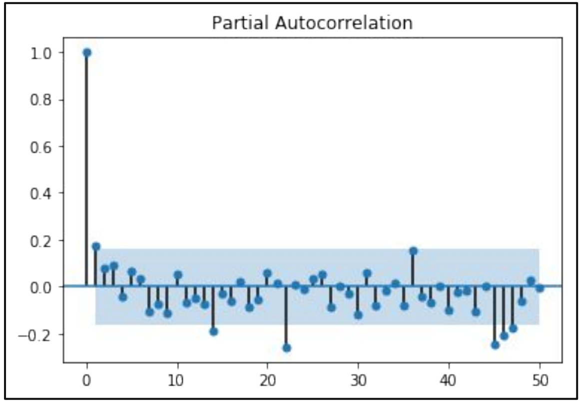


Fig: LSTM

	AR	ARIMA	LSTM
RMSE	1646	3710	1035
R2	0.76	-0.21	0.86

SMARD Data





	AR	ARIMA	LSTM
RMSE	-	5864	5184
R2	-	-1.3	-56.8

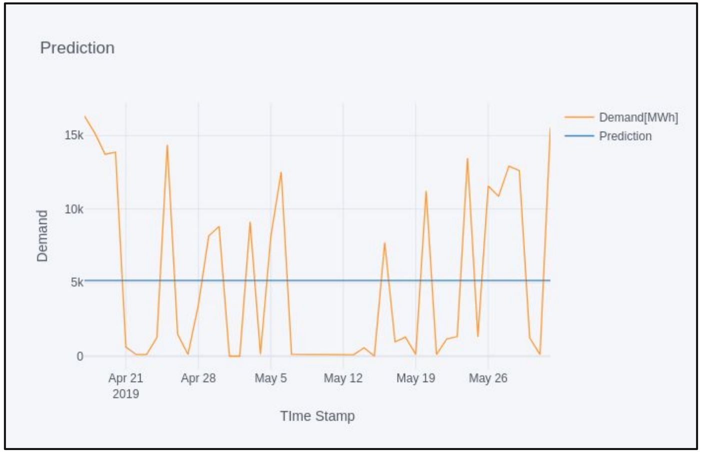
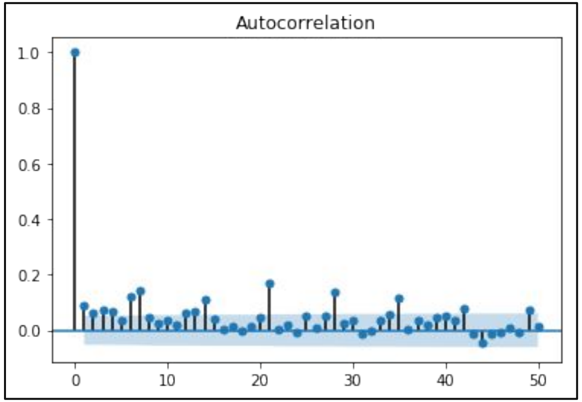
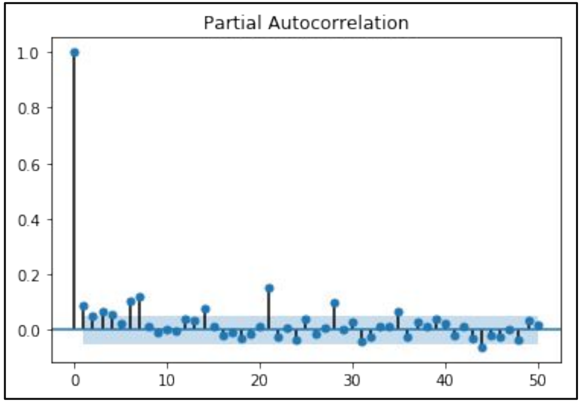


Fig: ARIMA

Fig: LSTM

UCI Data





	AR	ARIMA	LSTM
RMSE	-	0.78	0.65
R2	-	-0.4	-13.0

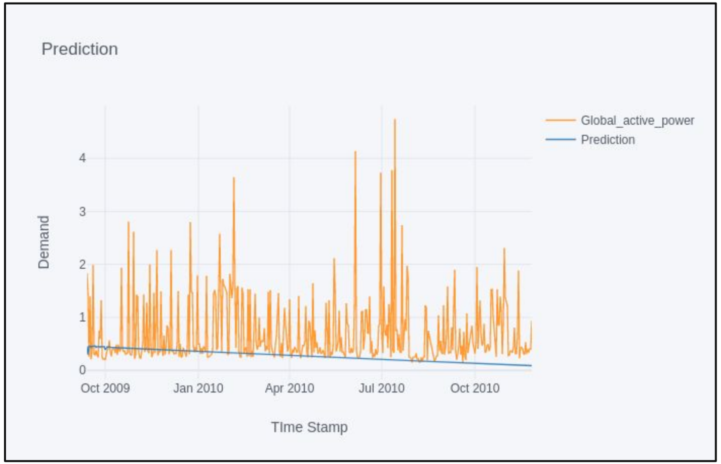


Fig: ARIMA

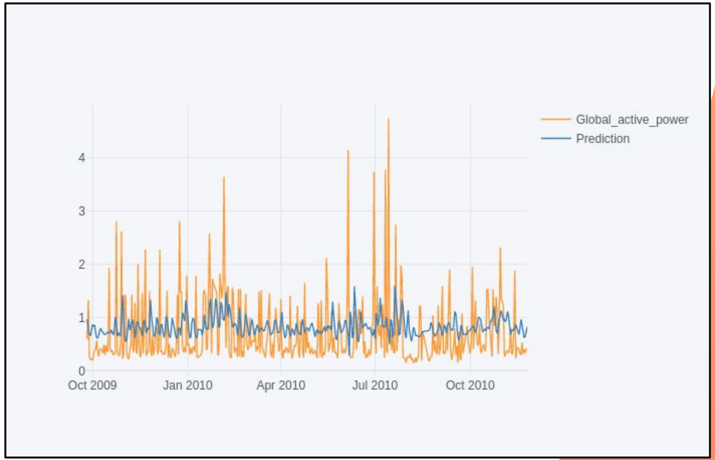
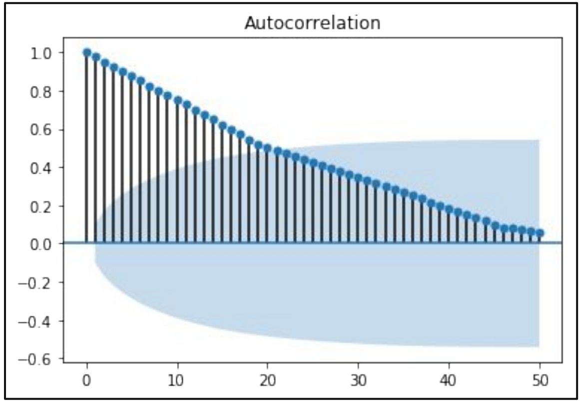
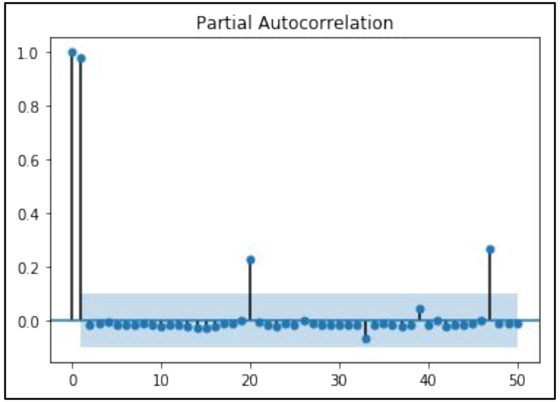


Fig: LSTM

UPSLDC Data





	AR	ARIMA	LSTM
RMSE	-	1780	334
R2	-	-0.2	0.95

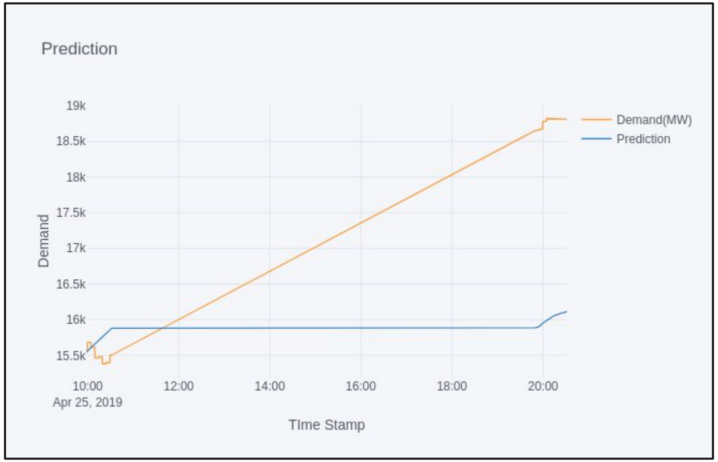


Fig: ARIMA

Fig: LSTM

Comparison

Comparison (R2 Score)			
Data-set	AR	ARIMA	LSTM
FRED	0.558	0.825	0.360
OPSD	0.761	-0.212	0.8647
SMARD	-	-1.347	-56.4874
UCI	-	-0.481	-13.052
UPSLDC	-	-0.2292	0.954

Comparison (Root Mean Square Error)			
Data-set	AR	ARIMA	LSTM
FRED	8.247	5.193	7.992
OPSD	1646.63	3710.969	1035.05
SMARD	-	5864.92	5184.04
UCI	-	0.780	0.6526
UPSLDC	-	1780.71	334.540

Disposal of comments

COMMENT 1: Three different data-sets should be taken

We have used five different datasets for this project.

COMMENT 2: Comparison of LSTM with ARIMA is required

Comparison using two parameters namely Root Mean Square Error and R2 Score have been used. For all data-sets except FRED, LSTM gives a better result as compared to ARIMA. Fred data-set has a very good pattern because of which better results are shown in ARIMA.

References

[1] Nataraja, C. Gorwar, Mahesh Shilpa, G.N. Harsha J, Shri. (2012). Short term load forecasting using time series analysis: A case study for Karnataka, India. International Journal of Engineering Science and Innovative Technology. 1. 45-53.

[2] Gao, Gao Lo, Kwoklun Fan, Fulin. (2017). Comparison of ARIMA and ANN Models Used in Electricity Price Forecasting for Power Market. Energy and Power Engineering. 09. 120-126. 10.4236/epe.2017.94B015.

[3] Chikobvu, Delson Sigauke, Caston. (2012). Regression-SARIMA modelling of daily peak electricity demand in South Africa. Journal of Energy in Southern Africa. 23. 23-30. 10.17159/2413-3051/2012/v23i3a3169.

[4] Miswan, Nor Mohd Said, Rahaini Anuar, S.H.H.. (2016). ARIMA with regression model in modelling electricity load demand. 8. 113-116.

[5] Iwok, Iberedem Okpe, A. (2016). A Comparative Study between Univariate and Multivariate Linear Stationary Time Series Models. American Journal of Mathematics and Statistics. 2016. 203-212. 10.5923/j.ajms.20160605.02.

[6] Jalil, N.A. Ahmad, Maizah Mohamed, N.. (2013). Electricity load demand forecasting using exponential smoothing methods. World Applied Sciences Journal. 22. 1540-1543. 10.5829/idosi.wasj.2013.22.11.2891.

[7] \Load Forecasting Bibliography", Phase I, IEEE Transactions on Power Apparatus and Systems, Vol.PAS-99, No.1 January/February 1980.

[8] Short term load forecasting using time series modeling with peak load estimation capability", IEEE Transactions on Power Systems, Vol.16, No.3 August 2001

[9] Hippert, H.s Pedreira, Carlos Souza, Reinaldo. (2001). Neural Networks for Short-Term Load Forecasting: A Review and Evaluation.

[10] Load forecasting using support vector machines: A study on EUNITE competition 2001", IEEE Transactions on Power Systems, Vol.19, No.4, November 2004.

[11] \Short term load forecasting using fuzzy neural networks", IEEE Transactions on Power Systems, Vol.10, No.3 August 1995.

[12] \Short term load forecasting for fast developing utility using knowledgebased expert systems", IEEE Transactions on Power Systems, Vol.17, No.4, May 2002.

[13] \Short term power system load forecasting using the iteratively reweighed least squares algorithm", Electric power system research, 19(1990) pp.11-12

[14] M. A. Abu-El-Magd and N. K. Sinha, "Short-Term Load Demand Modeling and Forecasting: A Review," in IEEE Transactions on Systems.

Thanks!

Correlation Formula

$$\rho_{xy} = \frac{\mathrm{Con}\,(r_x\,,\, r_y)}{\sigma_x\,\sigma_y}$$

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

AutoCorrelation Function(ACF)

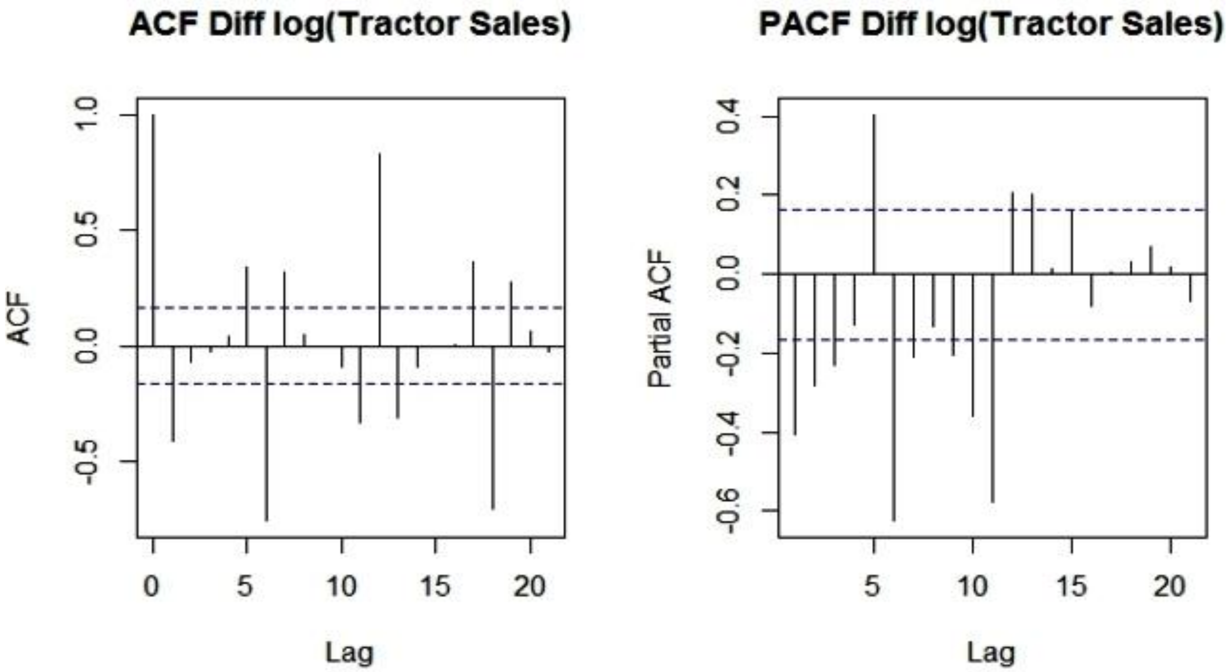
We can calculate the correlation for time series observations with observations with previous time steps, called lags. Because the correlation of the time series observations is calculated with values of the same series at previous times, this is called a serial correlation, or an autocorrelation. A plot of the autocorrelation of a time series by lag is called the AutoCorrelation Function, or the acronym ACF. This plot is sometimes called a correlogram or an autocorrelation plot.

Partial AutoCorrelation Function (PACF)

A partial autocorrelation is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed.

The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.

Example -


ACF Diff log(Tractor Sales)
ACF
1.0
0.5
0.0
-0.5
0
5
10
15
20
Lag
PACF Diff log(Tractor Sales)
Partial ACF
0.4
0.2
0.0
-0.2
-0.4
-0.6
5
10
15
20
Lag