UNC CHARLOTTE

The WILLIAM STATES LEE COLLEGE of ENGINEERING

EMGT 6910

Technological Forecasting and Decision Making

Project Phase-3 Report

Submitted by: Mahesh Alapati

1. Introduction

In this report, I reproduced and improved the paper on "**Short Term Load Forecasting Using Multi-Parameter Regression** " [2], with different dataset. I used a dataset from npower Forecasting Challenege[3] to forecast a half hourly load of seven days ahead with MAPE less than 2%. Authors used multivariable regression method to forecast load at coming hour, but I predicted load at coming half hour.

During the last decade, electric power generation industry has undergone a significant transformation due to deregulation of the energy market. In this aspect, energy planning plays a vital role for in a competitive electric power generation industries. As everybody knows it, large amounts of energy cannot be stored, electricity must be produced as it is used. To supply electricity to the customers, their load must be predicted, which means supply should meet demand. The process of making predictions of the load is called '**Load Forecasting**'. Forecasting of load should be should be done around 3% of error, if it is underestimated, factors like reliability and security may be affected. On the other side, if the system load is overestimated, which leads to costly operation of the plant without supplying electricity to customers.

Though forecasting is not accurate, but by using better techniques and models we can improve the accuracy of load forecasting and profit share of a utility. Forecasting is the basic aspect of decision making in energy providing utilities. It has a lot of application in energy purchasing, operations, generation, and infrastructure development. Electricity has become popular and essential commodity in the business of electric utilities as other commodities.

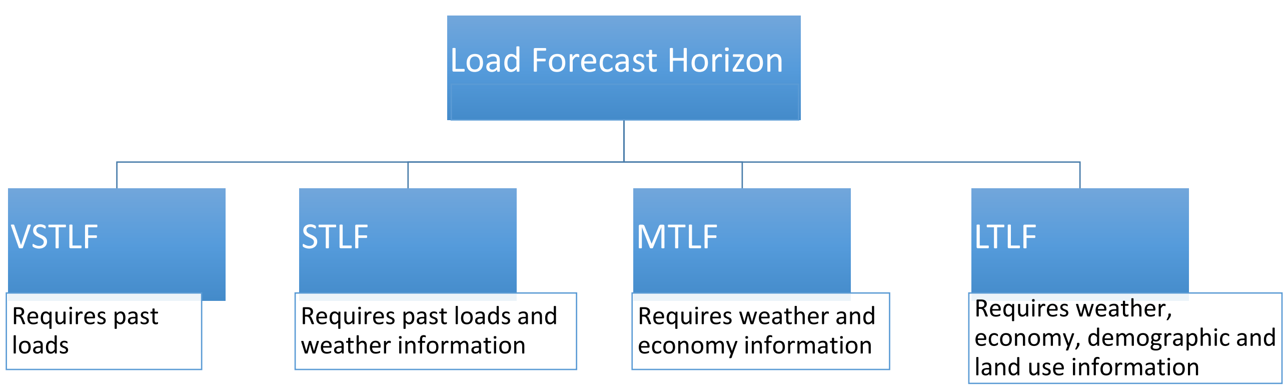

Figure 1 Load Forecast Information for different time horizons

Load forecasts can be divided into three major categories based **on input information and time horizons**: short-term load forecasting(STLF), which is a period of 1 hour to one week, the middle term load forecasting(MTLF), which ranges from a week to 3 years, long term load forecasting(LTLF) is from three to twenty years, these help companies to plan their investment, plan generation, transmission and distribution systems.

Short term load forecasts have become increasingly important since the rise of competitive energy markets and deregulation; it helps to provide a great profit and secured transmission of electricity to customers. Many techniques were used to forecast; these methods can be classified as **1. Statistical approaches and 2. Artificial Intelligence** approaches.

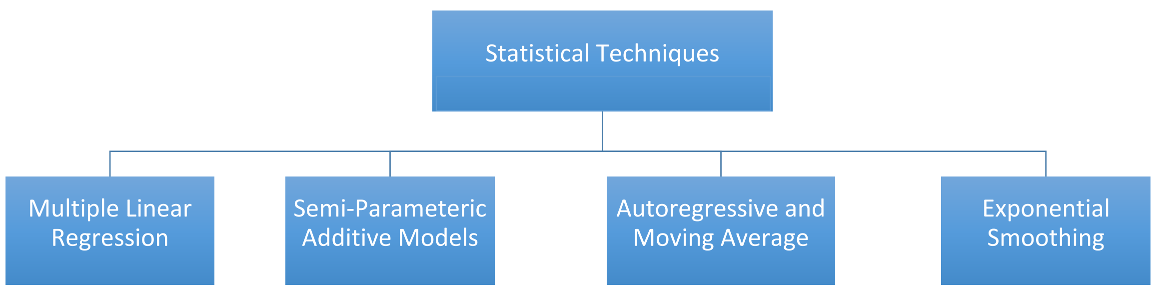

Figure 2 Type of Statistical Techniques

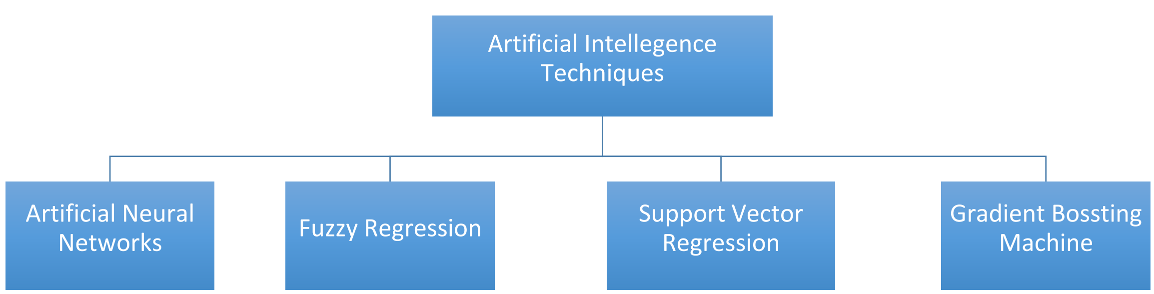

Figure 3 Types of Artificial Intelligence Techniques

2. Literature Review

Many papers are published in short-term load forecasting in various journals. In this report, literature review is focused on a summary of phase-1 of the project. Many authors and papers suggested traditional methods such as linear regression, ARIMA or any other statistical techniques have the advantage of the simple algorithm and easy implementation, but these methods are on linear analysis, in which techniques are not able to forecast accurately in nonlinear data. But, if we look at intelligent techniques such as ANN, SVMs have potential to give an optimal solution at higher speed and deals with non-linearity efficiently than statistical techniques. However, it is ambiguous to decide what kind of technique to be used to forecast, but with proper **exploratory analysis on data** produce can better forecast and right results. As we know all forecasts are wrong, but they are useful if they are modeled and analyzed as per historical data and requirements of a utility.

3. Data Description

The proposed model in this paper, authors used two sets of data: hourly load and weather variables data such as hourly temperature, wind speed, and cloud cover. However, to reproduce results half hourly data has been used from npower forecasting challenge competition , data from Jan 2012 to Jan 2013 used as history to predict parameters. Forecast of seven days have been done at half hourly load from 2 nd Jan 2013 to 9 Jan 2013.

4. Block diagram of Short Term Load Forecasting(STLF) Model

Following weather variables have been used to forecast load at coming half hour using Multiple Linear Regression technique.

Parameters used to forecast are:

- *T (Temperature at current hour)*
- *T2(Temperature at previous hour*
- *T3(Temperature at previous 2 hours)*
- *W(Wind speed at current hour)*
- *W2(Wind speed at previous hour)*
- *W3(Wind speed at previous 2 hour)*
- *C(Cloud cover at current hour)*
- *C2(Cloud cover at previous hour)*
- *C3(Cloud cover at previous 2 hour)*
- *D(Load at current hour)*
- *D1(Load at previous hour)*
- *D2(Load at previous 2 hour)*

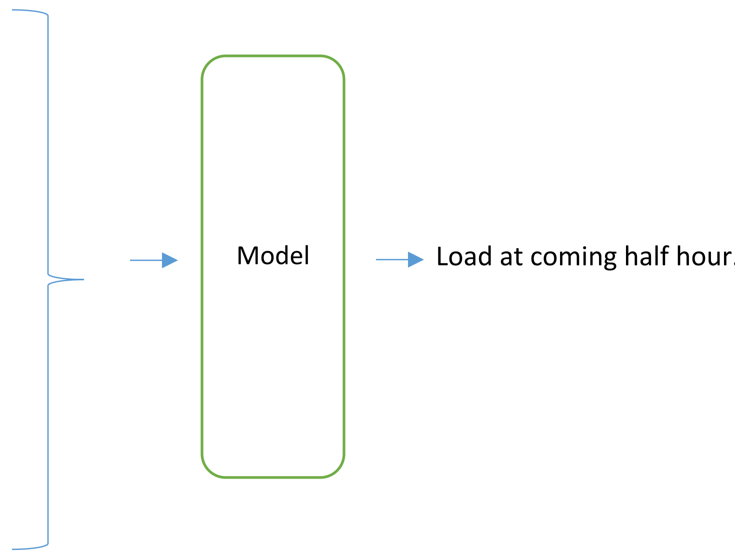

Figure 4. Schematic of STLF Model

DEPENDENT VARIABLE: Load

CLASSIFICATION VARIABLES:
Hour, Month, Day

QUANTITATIVE VARIABLES:
T, T-Square, T2, T3, W, W2, W3, C, C2, C3, D, D1, D2

MAIN EFFECTS:
Hour, Month, Day, T, T-square, T2, T3, W, W2, W3, C, C2, C3, D, D1, D2

CROSS EFFECTS:
Hour*Day, T*Month, T2*Month, T3*Month, Temperature*Hour, T2*Hour, T3*Hour, T-Square*Month, T-Square*Hour.

Methodology

In this study, three techniques have been implemented such as Multiple Linear Regression (MLR), Artificial Neural Networks(ANN) and classification. Among the above, MLR analysis has been widely used in load forecasting. Below sections discuss its implementation.

The length of data used to for parameter estimation and error statistic calculation:

Testing	Validation	Forecast
1/1/2012 – 1/31/2013	2/1/2013- 2/7/2013	2/8/2013 – 2/14/2013

Table 1 Length of data used for calculating MAPE

4.1 Multiple Linear Regression

The general linear regression model can define as**:**

$$E(Load) = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + ... + \beta_{p-1} \times X_{p-1} + e_i$$

Where β0… βp-1 are the coefficients, X1, X2... Xp-1 are the predictor variables. In this paper, they are referred as Y= Load, X1, X2 as weather variables such as temperature, wind, cloud.

Implementation of above MLR model is done in SAS. In this analysis, several MLR model is implemented in SAS Enterprise Guide. Model with less MAPE in validation set has been chosen as the best fit model of the data.

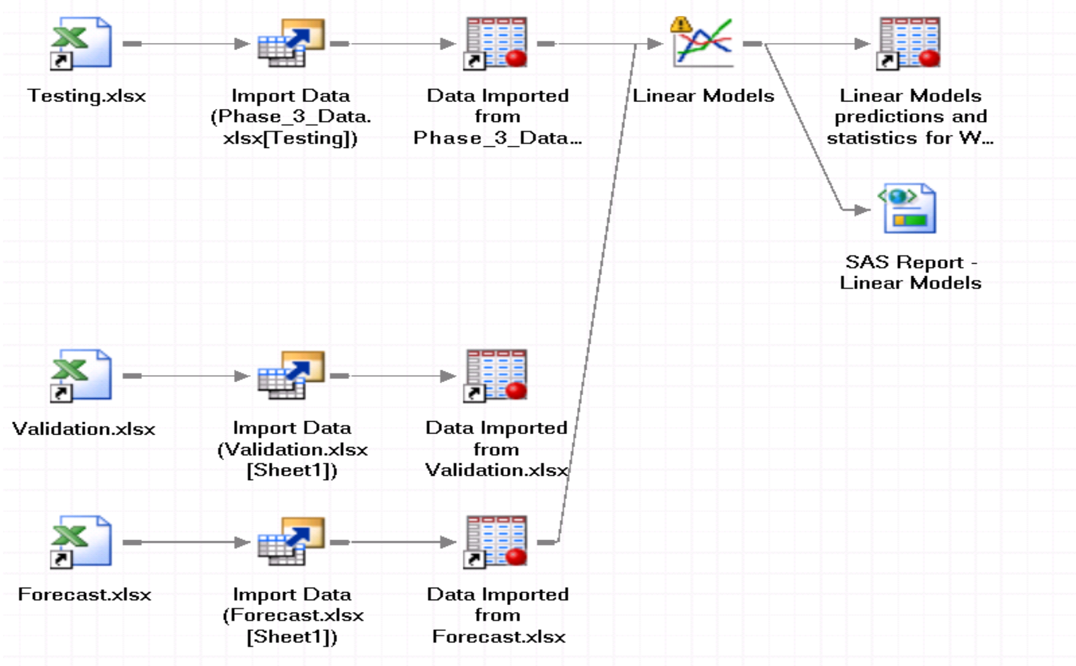

Figure 5 MLR Implementation in SAS

4.1.1 Multiple Linear Regression Implementation

Multiple Linear Regression has been performed to history data using R programming language in the Rstudio.

Steps followed:

- **Step1:** Read Input data; actual load, half hourly temperature, wind speed and cloud cover at a current half hour and previous two hours.
- **Step 2:** Multiple regression performed in Rstudio to forecast load at coming half hour based on 12 parameters.
- **Step 3:** Code is written to perform all operations from input of data to forecast of coming half hour[2].
- **Step 4:** Multiple regression is used to forecast based on any new set of parameters.

4.2 Artificial Neural networks

ANN approach is used to perform nonlinear modeling for the relationship between the load and weather variables. The algorithm is tested using npower forecasting challenge data. The algorithm has been implemented in MATLAB tool.

Artificial Neural Network technique has been used to forecast Jan 8, 2013 to Jan 14, 2013, half hourly one load ahead. Different models have been designed, but below model gave the best output. The error statistics used to measure forecast accuracy are MAPE and R-Square.

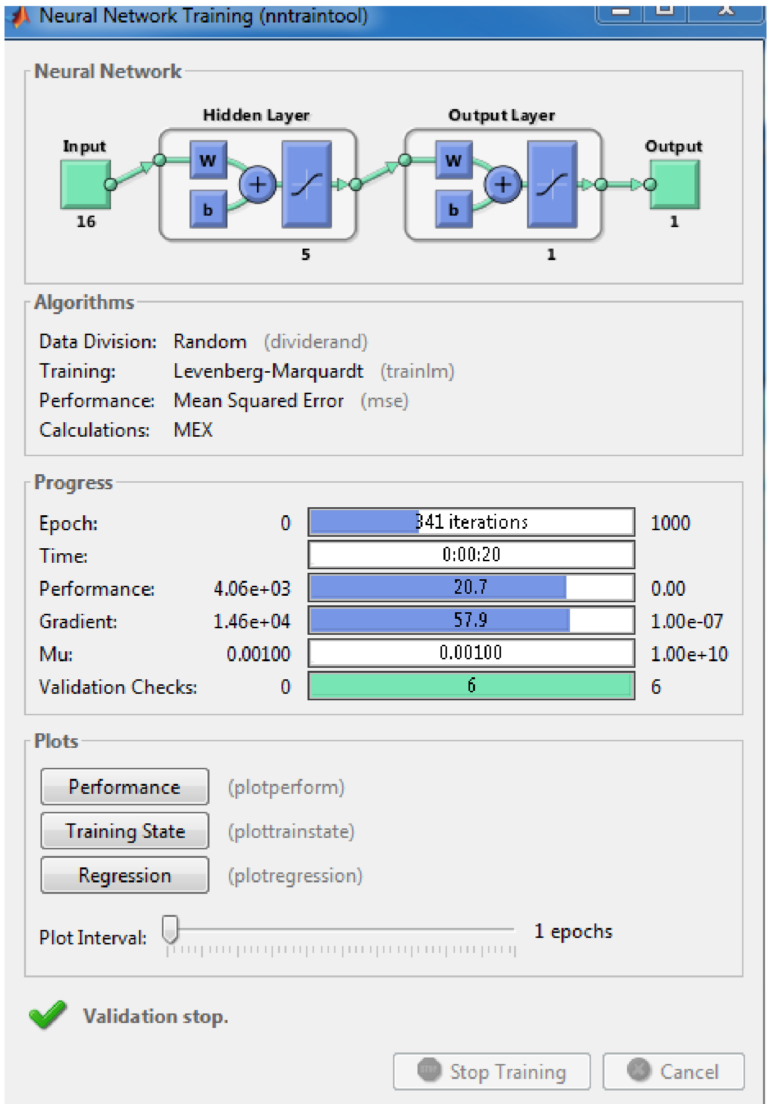

Figure 6 ANN methodology

4.3 Classification

***K-nearest* neighbor classification** method has been used to forecast seven days ahead half hourly load. The forecast has been done using 2012 data as test data. Various **NumNeighbors** have been used to find the best fit model for the given data. **NumNeighbors:3** is used to forecast 2013.

Steps followed:

- **Step1:** Read Input data; actual load and various predictor variables.
- **Step 2:** ClassificationKNN performed in MATLAB to forecast 2013 load.
- **Step 3:** Code is written to perform all operations from input of data to forecast.

Code:

```
1. model= fitcknn(X,Y,'NumNeighbors',3,'Standardize',1); % X predictor variables, Y Load
2. Forecast_2013= predict(model,F); % F Forecast predictor variables.
```

5. Error Statistics

The MAPE index is used to estimate error percentage from the actual load.

$$\frac{100}{N}\sum_{i=1}^{N}\left|(Ai-Fi)/Ai\right|$$

Where Ai= Actual Load, Fi= Forecast Load, and N is the number of observations.

6. Graphical Representation of Predicted Vs Actual load

Following figures represents a graphical representation of Actual Vs Predicted load; where X- axis denotes; Jan 8, 2013 – Jan 14, 2013, half hourly period and Y- axis represents a load.

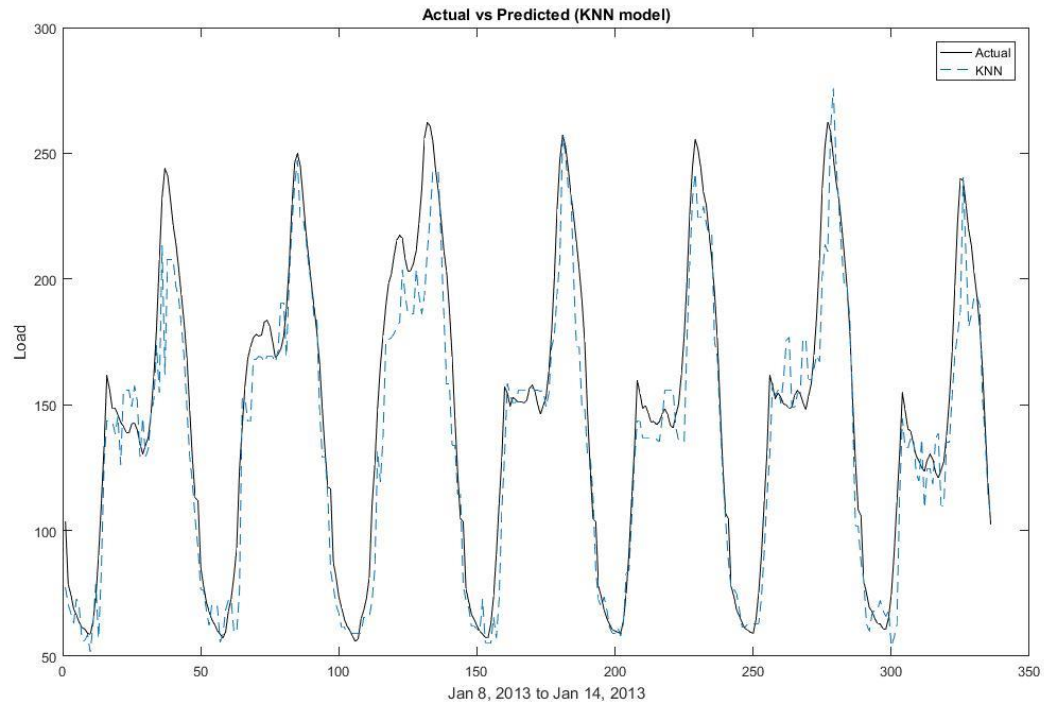

Figure 7 Actual Vs Predicted Load using KNN model

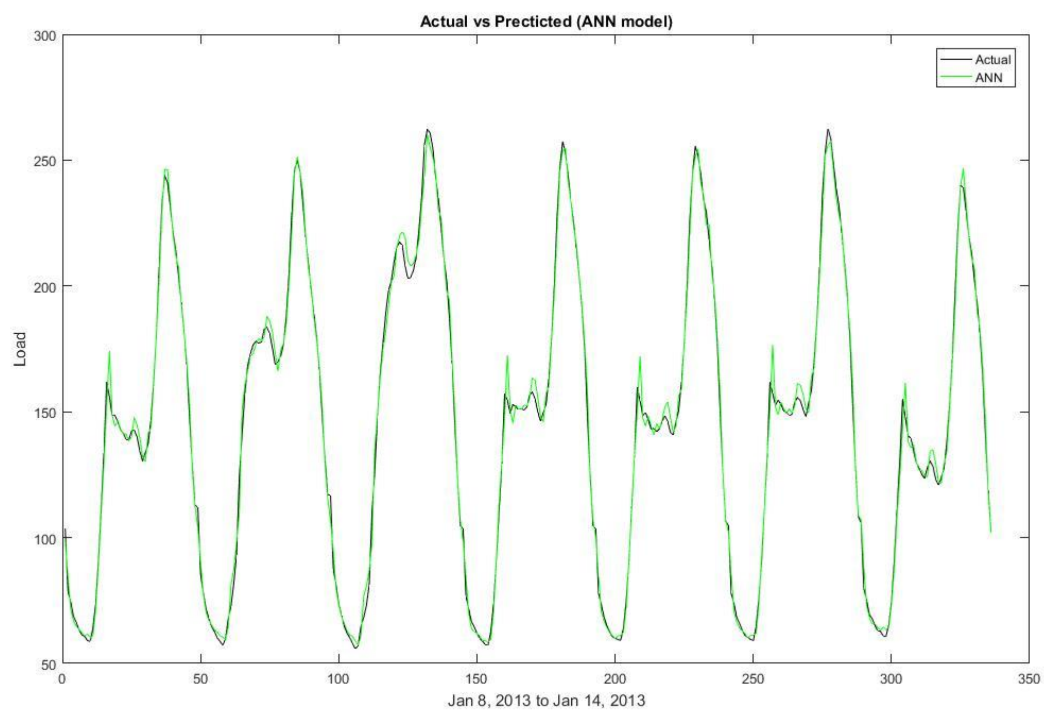

Figure 8 Actual Vs Predicted Load using ANN model

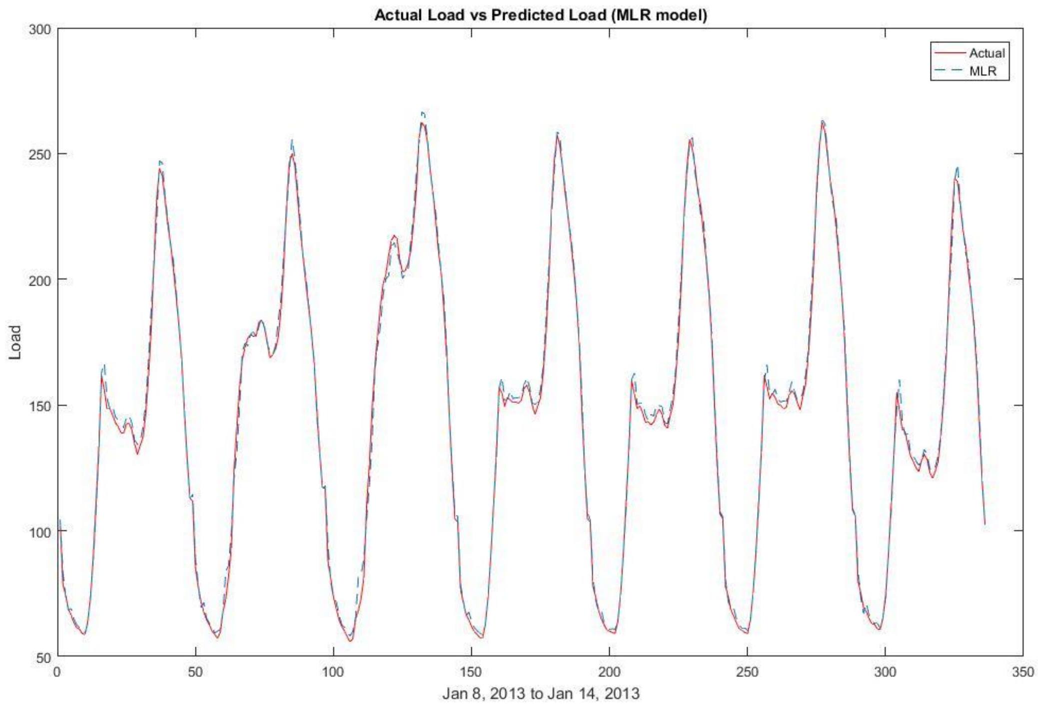

Figure 9 Actual Vs Predicted Load using MLR model

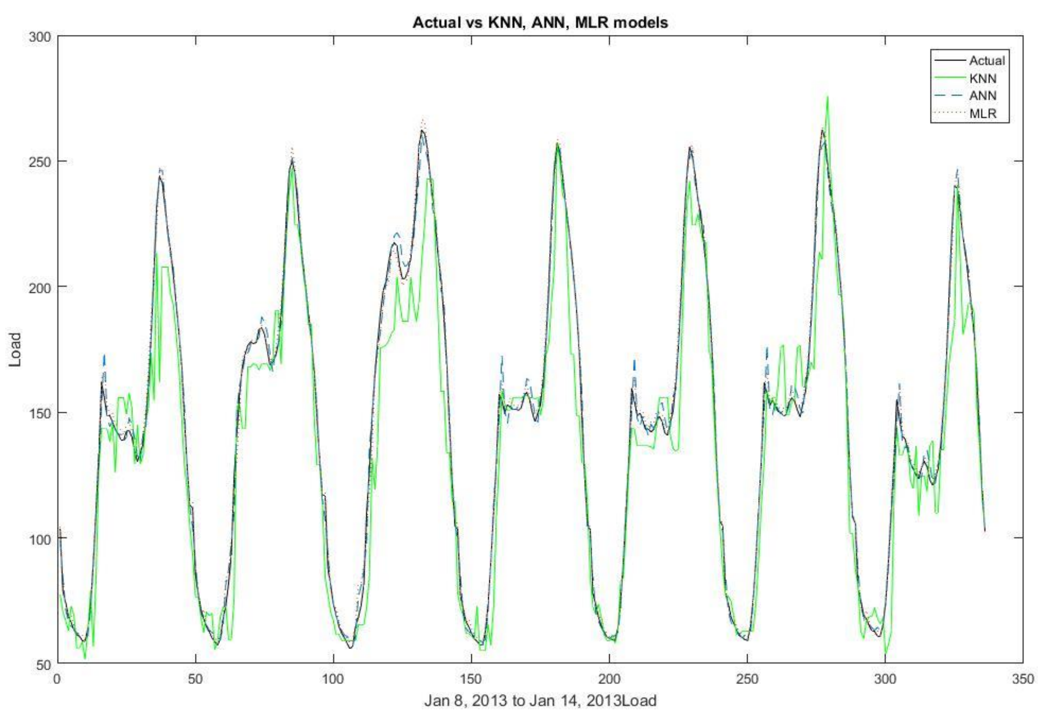

Figure 10 Actual Vs KNN, ANN, & MLR models

7. Results and Conclusion

In this project, I implemented three different techniques and chosen MLR as the best fit model for this npower forecast challenge data based on a percentage of MAPE among other models. Various analysis of MAPE has been tabulated in below table. **Accuracy has been improved from 4% to 1.95 % using MLR model.**

Category	MAPE (Validation) %	MAPE (Forecast)%
MLR	2.24	1.95
ANN	2.54	2.45
Classification	11.26	9.02

Table 2 Result Analysis of three models

8. References

1. Tao Hong, Pu Wang and H. Lee Willis," A Naïve Multiple Linear Regression Benchmark for Short Term Load Forecasting," Power and Energy Society General Meeting, 2011 IEEE.

2. Rothe, Wadhwani and Wadhwani," Short Term Load Forecasting Using Multi Parameter Regression" International Journal of Computer Science and Information Security, Vol. 6, No. 2, 2009.
3. https://www.npowerjobs.com/graduates/forecasting-challenge
4. https://www.mathworks.com/help/stats/classificationknn-class.html
5. Hesham K Alfares, Mohammad Nazeeruddin," Electric load forecasting: literature survey and classification of methods", International Journal of Systems Science.
6. Tao Hong, Hong, T. (2010). Short-term electric load forecasting. North Carolina State University.
7. Lee and Cha "SHORT-TERM LOAD FORECASTING USING AN ARTIFICIAL NEURALNETWORK", Transactions. Stems, Vol. 7. No. 1, February 1992