# Load Curve Analysis and Short-Term Load Forecasting – A Big Data Analytics Approach

**Dinesh Ghanta**
MBA in Business Analytics,
REVA University,
Bengaluru, India
Dineshg.BA01@reva.edu.in

**Pavanraj Talawar**
MBA in Business Analytics,
REVA University,
Bengaluru, India
Pavanrajt.BA01@reva.edu.in

***Abstract* – *Many of the electrical power distribution companies face challenges in being able to forecast the amount of electricity to be bought for a particular period of time to meet the demand of the consumers. The emergence of smart grid technologies in recent year has seen a paradigm shift in redefining the electrical system of the future [1]. The power distribution companies install smart plugs in the households which play a role of a proxy between the wall power outlet and the device connected to it. Smart meters are used to record consumption of electric energy and to communicate the information to the electricity supplier for monitoring and billing. The smart meter users can be households, office buildings, restaurants etc. The dataset used for our analysis is based on the recordings originating from smart plugs, which are deployed in private households. The data is collected for many households every minute for a period of 30 days.***

***In this paper, we provide a big data analytics approach to assist the customers as well as the power distribution companies by building a short-term load forecasting model which creates real time alerts when the power consumption reaches a threshold. This helps in analyzing consumer's usage on a real-time basis. It not only ensures power systems operate safely and economically, but also lays the foundation of formulating electricity dispatching schedule and transactions [4].***

***Keywords*: *Load Curve Analysis, Load forecasting, Big Data Analytics, Energy Analytics, Smart Grids, Smart Meters, Smart Plugs, Data Pipeline***

## I. INTRODUCTION

The power distribution companies have always faced challenges in predicting the customer's demand. The companies were not able to predict the amount of power required for the next month or quarter and therefore could not predict the amount of electricity to be bought for a particular period of time. If the company buys less and the demand is high, they will have to bid again and will have to buy at a very high price. If it buys more and the demand is less, the company will end of wasting electricity and will incur losses. In many countries, the company might have to pay fine for wastage of electricity as well. Many countries have realized the importance of power grids and are making huge investments in the power sector to build smarter power grids which will help their countries in the long run. The Institution for Electric Efficiency, USA has released a statistic that by the end of 2012, approximately 36 million smart meters were installed in households across the country and the number is estimated to be 80 million for China.

The smart plugs play a role of a proxy between the wall power outlet and the device connected to it. Every household contains multiple smart plugs and a smart plug has a load sensor measuring current load with Watt as unit. These smart plugs have the ability to record consumption data on a second by second basis. However, for our analysis, the load values have been aggregated at minute level. Also, the smart plug data for a household has been aggregated at the household level. In this paper, we have provided a big data analytics approach to analyze this data in real time and also alert consumers and power companies whenever the consumption reaches a certain threshold. This will help the power companies in managing the demand better. The power companies can also create customized plans for certain sector of consumers and introduce smart billing plans. The overall objective is to build a solution which would help companies and consumers alike to reduce wastage of electricity and to ensure optimal utilization of available electricity.

## II. LITERATURE REVIEW

Researchers around the world have different approaches for short term load forecasting. A few of the researchers Zhang, P., Wu, X., Wang, X., & Bi, S [4] have proposed their framework based on big data technologies. The paper suggests an approach of performing cluster analysis to classify daily load patterns. The next step would be to perform association analysis to identify the influential factors. Next decision tree is applied to create classification rules. Depending on the type of load patter, the relevant forecasting model is then chosen. By aggregating the individual forecasting results, the overall system load is forecasted.

The paper published by Li, Y., Guo, P., & Li, X [5] also suggests a similar approach for load forecasting. The paper mines the electricity usage patterns of different consumers to improve the forecasting accuracy. First, for different date types the typical day load is calculated for various households. Consumers with similar pattern of usage are clustered together. The final load forecasting model based on Online Sequential Extreme Learning Machine (OS-ELM) is applied to different clusters to conduct load forecasting and the load forecast is summed to obtain the system load. Simulation experiments are then performed using MATLAB to prove the validity of the proposed model.

In another research, Khalid, Y. I., Hassan, N. U., Yuen, C., & Huang, S. [6] have developed an algorithm for reduction of peak load to reduce the impact of increased usage of air conditioner in residential smart grid community. A demand response management (DRM) plan is created to highlight the maximum severity and the maximum severity of inconvenience. The air conditioner is modelled as a power throttling device and impact increasing the number of power states on the resulting peak load reduction is studied for each DRM.

The paper submitted by Shen, J., Jiang, C., & Li, B. [7] discusses on the approaches to control load management approaches in smart grids. The traditional controllable load approaches such as the end users' controllable appliances, storage battery, Vehicle-to-Grid (V2G), and heat storage are reviewed. The "broad controllable loads" management, such as the microgrid, Virtual Power Plant (VPP), and the load aggregator are also presented. Furthermore, the load characteristics, control strategies, and control effectiveness are analyzed.

## III. DATA DESCRIPTION AND PREPARATION

The data used for analysis is based on the recordings originating from smart plugs installed in the various private households. These plugs were installed in 40 different houses. A hierarchical structure exists in the dataset with a house, identified by a unique house id, being the topmost entity. Every house contains one or more households, identified by a unique household id (within a house). Every household contains multiple smart plugs and a smart plug has a load sensor measuring current load with Watt as unit. The smart plug records data every second. However, for the purpose of our analysis, the load data is aggregated at minute level.

The dataset contains the power load data for 40 houses over a period of 30 days. The data is aggregated at a minute level from the seconds data captured by the smart plug. Each house has upto 18 households identified by a unique ID. The dataset has 4 features:

**house_id** – a unique identifier of a house where the household with the plug is located

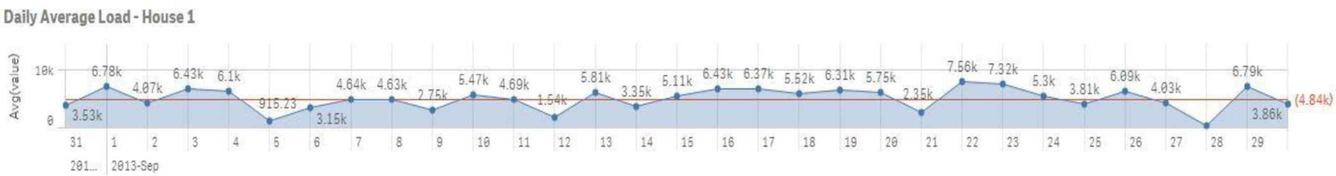
**household_id** – a unique identifier of a household (within a house) where the plug is located

**timestamp** – timestamp of measurement (number of seconds since January 1, 1970, 00:00:00 GMT)

**value** – the measurement (in Watts)

The total number of observations available in the dataset for 40 houses is more than 11 million (11,904,222). As the dataset is huge, for our analysis we have considered the data for 4 houses only for the same duration of 30 days. The final number of observations is approximately 1.1 million (1,117,379). Also, as the timestamp is represented in seconds, this has been changed to date format and has been included as a new feature for our analysis.

The daily average consumption of the each of these 4 households have been plotted in the graphs below. The overall average has been highlighted in red for comparison.




We can observe that the overall average varies for different households based on their pattern of utilization. Also, the data for $28^{th}$ Sept was not captured due to issues with the smart plugs and hence this date has been excluded.

## IV. BIG DATA APPROACH

In order to process real time data from the smart power grid, we need powerful computation systems which have the flexibility to evaluate the collection of all the power system and consumption data in near real time.

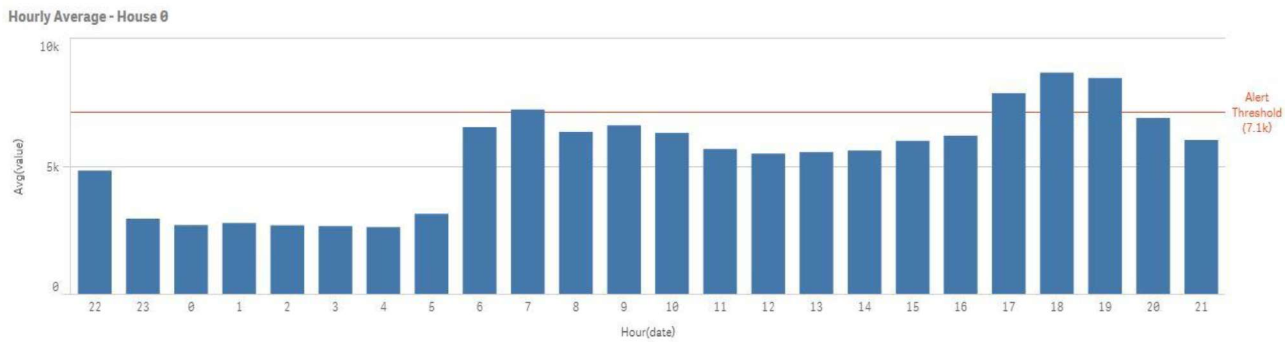
The Big Data Analytics platforms are designed for this very purpose to meet the requirement of very high power and also flexibility. There are different types of processing techniques used in Big Data analysis. The main types are batch processing, stream processing and iterative processing. Therefore, we require such a Big Data Analytics platform to store this vast distributed data and to perform all these types of analysis.

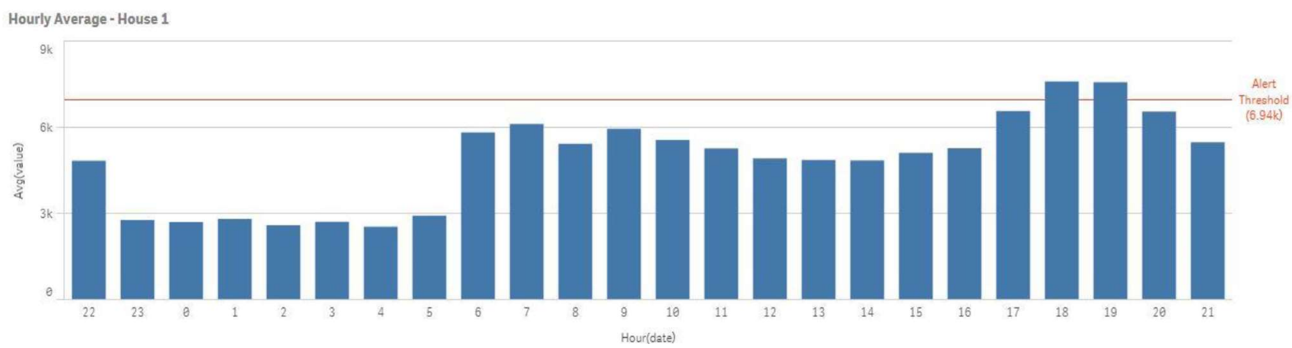
In this study, the processing type that we have used for our research is stream processing. The power consumption data is streamed in real-time using a csv file, which is processed by Apache Kafka for real time data processing and Apache Spark to process the data and also to generate alerts. The data is also visualized using dashboard for real time monitoring of load and generating alert messages.

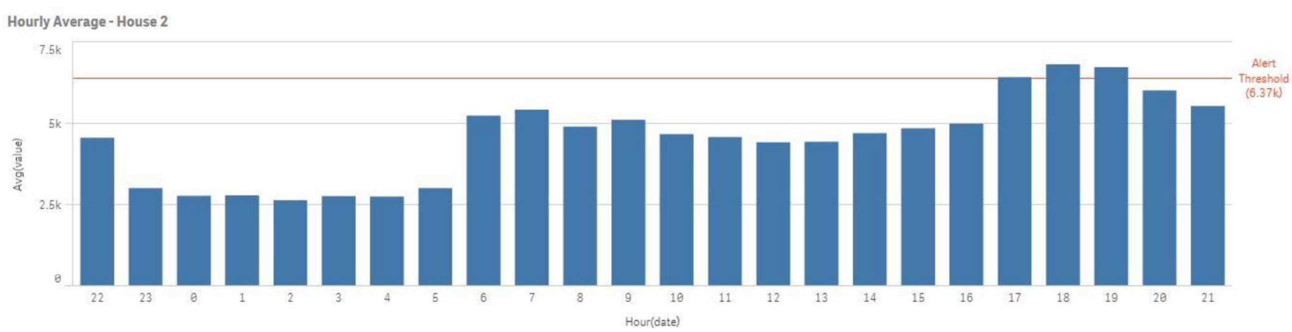
The data is ingested from various sources on real-time basis. This data from the comma separated file is consumed in a streaming fashion through Kafka. The Kafka consumer is started with code to aggregate the sum and count for each batch of RDD, based on house-household-id, date and hour combination. All the entries from the file are consumed and the average value is obtained based on cumulative sum and count per house-household-id, date and hour combination. This result is then stored in csv file. For the first alert, the output file is taken from earlier step as stream (this can be a Kafka stream too), and for each input value group in format "house-household-id"_"hour",date and value. This value is sent via flatMapGroupsWithState to compute mean and

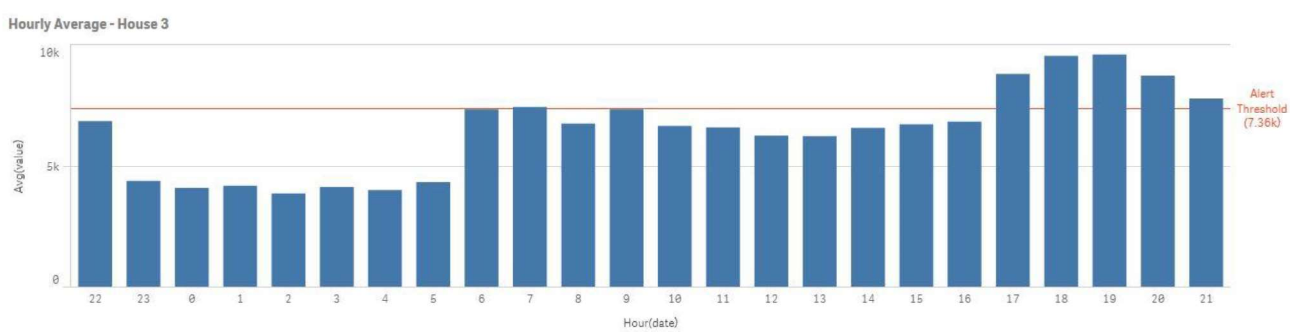
Standard Deviation (SD). The state and values are temporarily stored with "house-household-id"_"hour" combination. This will be used to append future value of this group to calculate running average and SD.

The following bar charts show us the hourly load for each house and the alert threshold calculated using 1 Standard Deviation based on the household's historic mean consumption.



## CONCLUSION

The big data approach proposed in this paper provides the required tools for consumers and power distribution companies to forecast the power required over a duration of time. By implementing this solution, electricity can be optimally used and wastage can be minimized. Future scope of work would be to refine the solution using newer advanced big data technologies and also to implement the solution using the entire dataset as opposed to the sample of dataset used in our analysis.

## REFERENCES

[1] Broeer, T. (2015). Analysis of smart grid and demand response technologies for renewable energy integration: operational and environmental challenges (Doctoral dissertation).

[2] Price, P. (2010). Methods for analyzing electric load shape and its variability.

[3] El-Hawary, M. E. (2014). The smart grid—state-of-the-art and future trends. Electric Power Components and Systems, 42(3-4), 239-250.

[4] Zhang, P., Wu, X., Wang, X., & Bi, S. (2015). Short-term load forecasting based on big data technologies. *CSEE Journal of Power and Energy Systems*, *1*(3), 59-67.

[5] Li, Y., Guo, P., & Li, X. (2016). Short-term load forecasting based on the analysis of user electricity behavior. *Algorithms*, *9*(4), 80.

[6] Khalid, Y. I., Hassan, N. U., Yuen, C., & Huang, S. (2014, November). Demand response management for power throttling air conditioning loads in residential smart grids. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on* (pp. 650-655). IEEE..

[7] Shen, J., Jiang, C., & Li, B. (2015). Controllable load management approaches in smart grids. *Energies*, *8*(10), 11187-11202.

Github Link for dataset:
https://github.com/pavanrajt/short_term_load_forecasting/blob/master/household_data.zip