ELECTRICITY LOAD FORCASTING IN SMARTGRID

C Sai Pranathi
B.Tech IV ECM Student, Sreenidhi Institute of Science and Technology, Hyderabad, TS, INDIA.

Bonagiri Varun
B.Tech IV ECM Student, Sreenidhi Institute of Science and Technology, Hyderabad, TS, INDIA.

Ravikanti Sahithi
B.Tech IV ECM Student, Sreenidhi Institute of Science and Technology, Hyderabad, TS, INDIA.

Dr. Attili Venkata Ramana
Associate Professor-ECM, Sreenidhi Institute of Science and Technology, Hyderabad, TS, INDIA.

Abstract: Smart Grid, named after the next generation of the power grid, has a flow between electricity and consumption and offers the delivery of an automated energy-efficient network. Despite being a power grid and efficient it does require some attention when it comes to load analysis as it is an automated system and system components are dependent on each other. This paper proposes a novel method for load factor analysis using two machine learning algorithms (RFE) and Logistic Regression (LR). These methods perform daily and monthly load analysis on Smart Grid using the feature selection process; Results are shown as a graphic format i.e., weekly and daily.

Index Terms— Random Forest, Logistic Regression, feature selection process, load forecasting, smart grid

I.INTRODUCTION

The grid is used to describe an electrical system that supports other functions such as generating electricity, moving electricity, distribution, electricity control. The term Smart Grid (SG) also called the power grid or smart grid is a 20th Century Traditional Grid (TG) enhancer. Traditional Grid infrastructure is outdated, which eventually slows down and reduces power. As a result of the use of outdated infrastructure, the energy loss is experienced. This high energy loss results in overload, which is one of the major problems of today's world. The traditional power grid is used to carry power from central creators to a large number of users or customers. Smart Grid uses two-way flow of electricity and information to create a distributed and automated network of networks [1].

Table 1: comparing the smart grid and the traditional grid [1]

Smart Grid	**Traditional Grid**
Digital	Electromechanical
Pervasive control	Limited control
Distributed generation	Centralized generation
Self-monitoring	Manual monitoring

SG uses information from two-way technology, cyber-secure, and computational intelligence in an integrated manner in the electricity generation, transmitting, installing, distributing and implementing a clean, safe, secure, reliable, efficient, and sustainable.

The rise in energy prices over the past decade and the growing fear of the impact of global climate change have provided the latest research into sustainable building and housing.

Forecasts play a major role in complex SG management, especially in short-term forecasts such as days and weeks to make definite short-term decisions regarding components of generation distribution, energy savings and consumers, which provide safe and economical operation of SGs. [2].

This paper focuses on the load forecasting of the smart grid. The smart grid consists of a lot of appliances like phase meters, generators etc., Feature selection is one of the eminent techniques which is to be performed on datasets that contains lot of features. As smart grid contains a lot of appliances connected to it feature selection technique can be used. The Forecasting technique is used to analyze the future prediction of the total load which can be obtained from the appliances.

II. EXISTING SYSTEM

Varshney et al. has developed an integrated model for predicting the electricity market that preceded the day in terms of temperature and load, using neural structure and single-spectrum analysis.

Mousavian et al. advance the probabilistic method for hourly energy forecasting [3], where bootstriling technology is used to study uncertainties and to propose an extreme learning method developed for wavelet neural networks. Kobayashi et al. [4] developed a modified supply chain model to solve the problem of high electricity prices in real time supporting the development work, which looks at the trade-off between consumer consumption and energy savings.

III. PROPOSED SYSTEM

The proposed system is used to predict the future load of appliances. The model consists load data of a residential smart grid. It should perform 3 basic steps, i.e., normalization of data, feature selection using Random Forest Regressor, and eventually forecasting of the load using Logistic Regression (LR). This makes the forecasting better as the features are reduced during predicting.

Advantages:

- Elimination of correlated features makes prediction easier.
- Minimum cost maintenance as the load can be predicted.
- It helps in deciding and planning for the maintenance of the power systems.

IV. ANALYSIS

Fig 1: Flowchart

V. IMPLEMENTATION

a. Environment

The experiment computer has Intel Core i5-4440
processor clocked at 3.10 GHz. Neural networks were trained on NVIDIA GeForce GT1030 GPU having a CUDA compute capability of 6.1 with 2 GB dedicated GPU memory, and an overall CPU RAM of 8 GB.

b. Dataset

The dataset contains 49 columns where every column represents the total load taken by each appliance in terms of kwatts. Load is obtained every hour for 365 days i.e., one full year. These parameters combining together help in predicting the load in smart grid. From the dataset we take the sum of the values for every 24 samples as it makes whole day; we perform the forecasting for the 365 samples. To perform the forecasting on weekly basis we sum up the day values for every 7 samples as it makes 53 samples. Figure 2 shows the samples of dataset containing 49 parameters and 8760 samples.

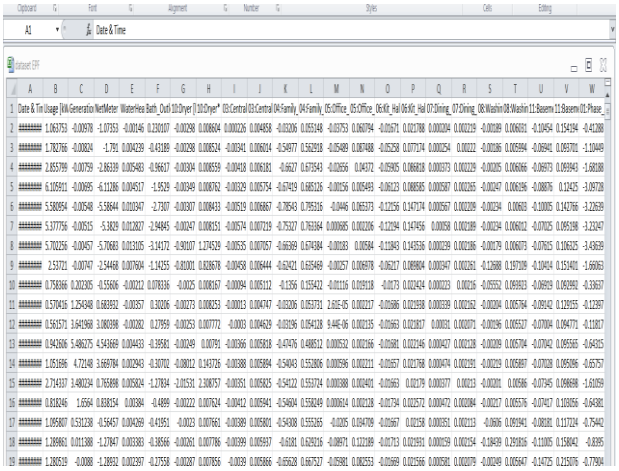

Fig 2: Dataset

c. Data Preprocessing

Data processing is a data mining method that converts data into logical format.
Real-world data is often incomplete, incompatible, and / or decreased in some behavior and may contain many errors.

The dataset mainly consists of inconsistent data and empty values those are needed to be removed. The correlation among the dataset is needed to be removed as the number of columns is decreased and resultant columns make forecasting easier and efficient. As the term correlation refers to a mutual relationship or association between quantities. A correlation is that the starting moves to understanding relationships and later it builds better statistical models.

d. Random Forest Regressor for feature Selection

A random forest as a name describes is an ensemble containing many decision trees. Each tree within a random forest disperses the phase prediction [7].

A large number of unrelated trees defined as models serve as a group to build any model for each area. The low correlation between the models is that the key [7]. Instead a low-risk investment improves the creation of a portfolio that is much larger than the sum of its parts, unbiased models can produce more common forecasts than one perspective. The purpose behind this brilliant outcome is that the trees protect each other from their errors. While a few trees might be good and bad, numerous different trees will move the correct way, so as related trees, they can move in a straight line.

e. Forecasting using Logistic Regression

LR, the type of mathematical model which is used for rewriting, is used to perform analysis on a given dataset, which makes guesses using independent variables. The end result of LR is in the form of binary terms. The main purpose of LR is to explain the association between dependent and independent variables. There are two important categories of LR: sigmoid function and loss function [7]. Properties ought to be in double organization to utilize the LR calculation. In this way, standardization of the information is required before the applying LR model to the accessible information. The sigmoid loss function is defined by the accompanying condition, which is gotten from the loss function. Loss function (t) = 1m (−ytlog (h) - (1 - y) tlog (1 - h)).

Probability estimates require conversion to binary format (0 or 1), making estimation possible. Structural decomposition can be a straightforward process, but estimates can be modified using the logistic function [7]. The result of this is that, as a mathematical assumption, we now cannot interpret expectations as direct combinations of input, as the above model implies:
p (X) = e ^ (c0 + c1 * X) / (1 + e ^ (c0 + c1 * X))

Structural decomposition is a way of finding composite combinations to separate two or more groups. It has no limiting consideration for the DFA.

VI. WORKING

The dataset contains 49 parameters of all the appliances. It is used for predicting using the different machine learning techniques. “Pandas” is an open source data analytics tool which is very flexible to implement. It is used to obtain data from Comma Separated Values (csv) files.

Data cleaning and normalizing is performed using different functions of Numpy. Numpy is an N dimensional object, then the feature selection is performed using Random Forest Regressor algorithm. RFR is an ensemble which is used for extracting the important features using scikit learn API.

Finally, the prediction is performed using the selected features using Logistic Regression algorithm. The predicted output is depicted in the form of a graph.

VII. RESULT ANALYSIS

The result is analyzed in the form of line graphs. As the actual test data and model evaluated test data is compared based on the difference.

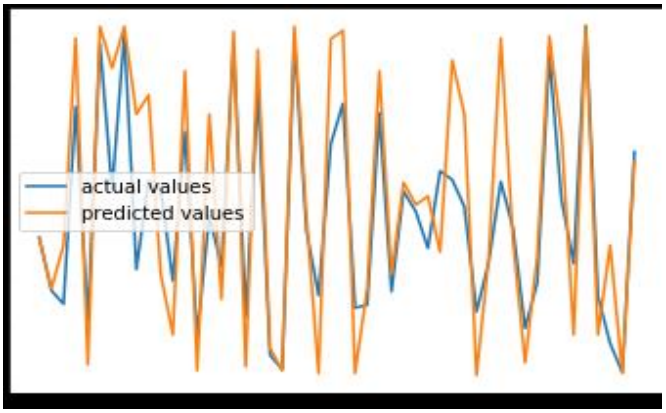

Fig3: Prediction of load on daily basis

The graph above shows the values on daily basis in which the actual values are compared with the forecasted values.

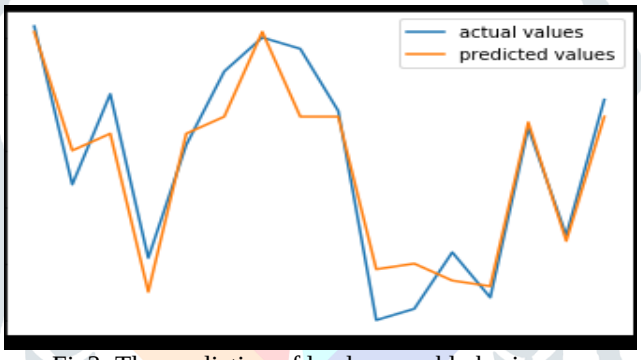

Fig3: The prediction of load on weekly basis

The graph above shows the values on a weekly basis in which the actual values are compared with the forecasted values.

VIII. CONCLUSION

We present the forecasting model based on the novel model which is based on the feature selection process in this paper. The features while forecasting on the weekly and daily basis depends upon the various parameters, the values of the load and hence the same features are opted for both. The conclusions and the comparisons were presented which depend on the ordinary least-squares based techniques. The result showed that the load predicted at the last quarter of the year shows better performance. Even the forecasting on a weekly basis shows a better performance in last quarter of the year.

REFRENCES:

[1] Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Smart Grid — The New and Improved Power Grid: A Survey. IEEE Communications Surveys & Tutorials, 14(4), 944–980. doi:10.1109/surv.2011.101911.00087

[2] Bracale, A., Caramia, P., Carpinelli, G., Di Fazio, A. R., & Varilone, P. (2013). A Bayesian-Based Approach for a Short-Term Steady-State Forecast of a Smart Grid. IEEE Transactions on Smart Grid, 4(4), 1760–1771. doi:10.1109/tsg.2012.2231441

[3] Wang, K., Xu, C., Zhang, Y., Guo, S., & Zomaya, A. (2017). Robust Big Data Analytics for Electricity Price Forecasting in the Smart Grid. IEEE Transactions on Big Data, 1–1. doi:10.1109/tbdata.2017.2723563

[4] Wang, H., & Gao, Y. (2019). Distributed Real-Time Pricing Method Incorporating Load Uncertainty Based on Nonsmooth Equations for Smart Grid. Mathematical Problems in Engineering, 2019, 1–18. doi:10.1155/2019/1498134

[5] Jaiswal, J. K., & Samikannu, R. (2017). Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. 2017 World Congress on Computing and Communication Technologies (WCCCT). doi:10.1109/wccct.2016.25

[6] Shaukat, N.; Ali, S.M.; Mehmood, C.A.; Khan, B.; Jawad, M.; Farid, U.; Ullah, Z.; Anwar, S.M.; Majid, M. A survey on consumers empowerment, communication technologies, and renewable generation penetration within Smart Grid. Renew. Sustain. Energy Rev. 2018, 81, 1453–1475. [CrossRef]

[7] Hanwu Luo, Xiubao Pan East Inner Mongolia Electric Power Co. Ltd Hohhot, China Logistic Regression and Random Forest for Effective Imbalanced Classification 2019 IEEE 43rd Annual Computer Software and ApplicationsConference (COMPSAC)