

Prediction of Electricity Price and Units for Home Appliances using Random Forest Techniques

Vasudevan N¹, Sailesh Reddy B², Suhail Abbas A³

Assistant Professor - CSE, Anand Institute of Higher Technology, Chennai, India¹

Student, B.E Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India^{2,3}

Abstract: Conversation of Electricity is a must and it can be done by start predicting the electricity consumption. The electricity consumption is higher than the required amount. Existing system uses mysql data for datasets and mathematical calculation for the prediction process which takes longer time for processing and prediction through mathematical calculations takes more time and accuracy is lesser. The above problem can be resolved by using big data processing for large collection of data and with that data it is been used for training the machine using the random forest algorithm for predicting the electricity consumption. Through that we enter the application's usage time accordingly in the GUI which takes the input and checks with the machine and gives the predicted range of the electricity consumption in units with the price.

Keywords: Big Data, Hadoop tool kit, Random forest algorithm, GUI.

I. INTRODUCTION

Energy conservation is the effort made to reduce the consumption of energy by using less of an energy service. This can be achieved either by using energy more efficiently (using less energy for a constant service) or by reducing the amount of service used (for example, by driving less). Energy conservation is a part of the concept of Eco-sufficiency. The electricity is also a type of energy which should be conserved by using low electric consumption application or application that consumes electricity efficiently by storing the excess power in a battery and use them for future usage. Due to increase in global warming every where they have started to adopt renewable energy like solar, wind where the half of the energy is been lost while converting the solar or wind energy into electricity. Electricity can be conserved by predicting the electricity usage and check with the application independently so that the application can be used according to the need and by reducing its usage efficiently. For the prediction the data are been analysed and processed in a data set format for the machine to use the data as input for training and testing. The Hadoop tool is been used for pre-processing and analyse the large set of data. The random forest algorithm is been used for training the machine, where the random forest algorithm is a supervised learning method. The supervised learning is a method where the input and the output is been given to the system for the training and then it is compared to the test data for checking its accuracy.

II. EXISTING SYSTEM

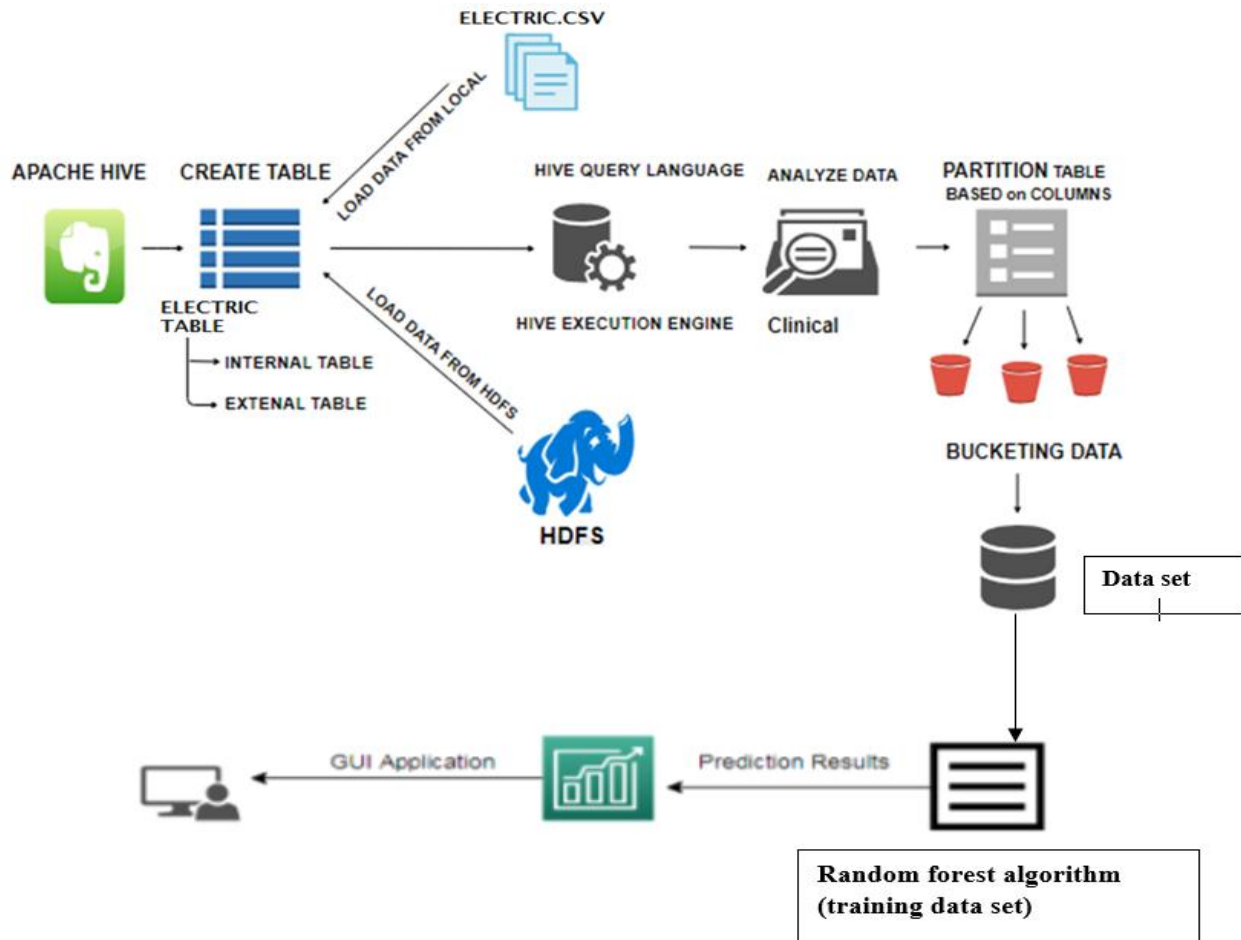
Existing System supervises the given backend by utilizing MySQL which contains heap of disadvantages i.e. information prerequisite is that preparing time is high when the information is tremendous and once information is lost, it can't recoup. In the prediction part it uses normal mathematical calculations it causes higher time consumption when large data is been processed it gives lesser accuracy.

III. PROPOSED SYSTEM

Proposed concept deals with providing database by using Hadoop tool it can analyse with no limitation of data and simply add number of machines to the cluster and get the results with less time, high throughput and maintenance cost is very less and we are using partitions and bucketing techniques in Hadoop. For the prediction and analyses part uses random forest algorithm it calculates faster with higher accuracy. The random forest algorithm is a supervised learning where we could train the machine with desired training data for the desired output.

IV. SYSTEM ARCHITECTURE

This system is designed to predict the electricity consumption of home applications. The system creates a table and loads the data from the csv and the hdfs filesystems through the hive execution engine the data is been ordered and analysed the true data only then they are been partitioned accordingly to their columns and they are bucketed for their values and stored as a csv file data. The data set is been analysed and the required dataset is been taken as a input accordingly to train the system for the required output. The trained dataset is been predicted accordingly to their trained data then the results are been checked and then it gets as the output in the GUI application.



V. SYSTEM IMPLEMENTATION

- **Pre-processing of data in the database: -**

In this module we have to create Data set for Electricity Consumption it contain set of table such that customer details, billing details and payment details for last four years .and this data first provide in MySQL database with help of this dataset we analysis this project.

- **Migration and Storage module:**

Now the dataset are been transferred into Hadoop (HDFS) for storage, Sqoop is a command-line interface application for transferring data between relational databases and Hadoop. we fetch the dataset into hadoop (HDFS) using sqoop Tool. Using sqoop we have to perform lot of the function, such that if we want to fetch the particular column or if we want to fetch the dataset with specific condition that will be support by Sqoop Tool and data will be stored in hadoop (HDFS).

- **Analyse data using Hive:**

Hive is a data ware house system for Hadoop. It runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. Hive was developed by Facebook. Hive supports Data definition Language (DDL), Data Manipulation Language (DML) and user defined functions. In here we analysis the dataset using HIVE tool

which will be stored in hadoop (HDFS). For analysis dataset HIVE using HQL Language. Using hive, we perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language

- **Validation and pre-processing**

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.

- **Prediction:**

The pre-processed data is been taken in as input for training the machine using the Random forest algorithm then the trained set is been checked using test set of data where the accuracy is been checked after testing the input for the user is been used as input and the machine runs according to its trained data set and it gives the output in the GUI application.

VI. CONCLUSION

Here the Power Distribution to Datacenters data is help to give awareness to how much power consumption in different datacenters. To run a analysis on the Power Distribution to Datacenters data in hadoop ecosystem. Then the local data set of electricity consumption is been added into the Hadoop ecosystem for analysis. Hadoop ecosystem uses hive, pig, map reduce to analysis faster than those data are been organized and been used in for the random forest for the bill prediction.

VII. FUTURE ENHANCEMENT

In future, the proposed system can add some renewable source of electrical energy into the system and predict how much the current can be saved, also able to suggest some less electricity consuming application and also predict when these applications can be switched off or kept idle to stop the wastage of the electricity.

REFERENCES

- [1]. H. P. Khomami and M. H. Javidi, "An efficient home energy management system for automated residential demand response," in Proceedings of the 13th International Conference on Environment and Electrical Engineering, Wroclaw, Poland, 2013, pp. 307–312.
- [2]. S. Shah, R. Khalid, A. Zafar, S. M. Hussain, H. Rahim, and N. Javaid, "An optimized priority enabled energy management system for smart homes," in Proceedings of the 31st International Conference on Advanced Information Networking and Applications (AINA), Taipei, China, 2017, pp. 1035–1041
- [3]. Z. Zhao, W. C. Lee, Y. Shin, and K. B. Song, "An optimal power scheduling method for demand response in home energy management system," IEEE Transactions on Smart Grid, vol. 4, no. 3, pp. 1391–1400, Sep. 2013.
- [4]. Z. R. Dogahneh and M. R. Dogahneh, "Optimizing the performance of smart grids in relation with residential energy centers equipping with Solar Power Units (PV)," Sindhological Studies, no. 1, pp. 14–22, 2017.