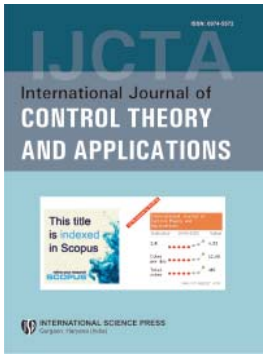

International Journal of Control Theory and Applications

ISSN : 0974–5572

Volume 9 • Number 40 • 2016

Analyzing Electricity Energy Consumption Behaviour based on SAX Algorithm using Hadoop with R

Jothi B[a] Sridev K[a] Rayudu Chaitanya Santhoshi B[a] Krishnaveni S[a] and Pushpalatha M[b]

[a]*Department of Software Engineering, SRM University, Chennai, Tamil Nadu, India*

[b]*Department of Computer Science & Engineering, SRM University, Chennai, Tamil Nadu, India*

Abstract: The administration offices and furthermore the enormous worldwide firms over the globe have practical experience in vitality protection and temperate use of vitality. Need the need the prerequisite of misuse vitality in an extremely prudent approach is that the need of creating nations like Asian country and China. The development of sensible framework meters gave North American country access to vast amount of vitality utilization learning. This information gave by sensible meters might be utilized speedily to supply experiences into vitality preservation measures and activities. Various vitality dissemination firms bridle this learning and acquire unusual outcomes concerning client's utilization design. They then when acting examination anticipate the request and utilization of clients. This investigation helps them to settle on a choice the levy at totally unique motivation behind your time. The organizations attempt to beat the bottleneck in capital speculation cost of learning [3]. Further, handle immense information for outline era and examination might be a moderate strategy and isn't sufficiently fast to bolster timeframe choosing. Our paper exhibits a Business Intelligence apparatus that utilizations Apache Hadoop to quickly deal with the common issues. Taking the benefit of this device, vitality dispersion firms will curtail the venture by misuse group equipment that runs Hadoop [5]. The utilization of appropriated processing devices conjointly diminishes the time interim extensively to change timeframe perception and choosing. This device additionally will curtail carbon impression and distinctive associated issues in vitality circulation together with loses and crime. In future, this same examination might be done on various utility assets like gas and conduits used in soothing administration time limitations

Keywords: *Vitality, Dissemination, Capital speculation cost, Carbon impression*

1. INTRODUCTION

Investigation of vitality utilization information to accomplish experiences into customer use examples is the thing that vitality wholesalers attempt to achieve for some objective applications like time-of utilization duty, request reaction administration and asking exactness. This sensible meter gathers learning every moment which winds up in producing incredible arrangement of learning of learning of data though the current mechanical meter gathers information by hourly or month to month. These tremendous learning stockpiling abilities and consequently the nature of data process knowledge fluctuate significantly with totally unique applications. Antiquated RDBMS[6] of utility firms might be a bottleneck in beating this approach. Subsequently, for the

business to really have the advantage of the sensible lattice speculation, it's vital that the vast amount of data made possible by sensible meters be taken care of speedily in partner sorted out way that helps network administrators make auspicious determinations to work framework securely, financially and dependably. Apache Hadoop is that the determination possible to handle on top of issues that keeps running on antique machines exclusively. It's circulated figuring device that have goliath stockpiling likewise as process ability. We tend to territory unit exploitation Apache Hadoop system that empowers for the disseminated procedure of gigantic learning sets crosswise over groups of PCs. Hadoop MapReduce might be a framework for information handling of tremendous learning sets. Antiquated RDBMS or option applications zone unit plenteous slower and wasteful in taking care of gigantic learning created by sensible meters when contrasted with Hadoop system. in this manner for the business and clients to accomplish edges like asking precision, vitality taking location, breaking down customer use examples and request reaction administration and so on it's never-endingly beneficial to utilize Hadoop that keeps running on minimal effort curio equipment. With the development of sensible meters for sensible circulation and conservative utilization of vitality, power, the created control should be used appropriately with legit economy increases to merchants and in this way the clients. Along these lines with this concentration of vitality appropriation inside the area of vitality utilization, which can prompt to diminishment of carbon prints, the investigation for the data got from the sensible meters should be finished. This immense size of examination can might want mammoth calculation which might be done the help of circulated process system, Hadoop. The system's utilization can give helpful valuable yields that include: soliciting precision, time-from utilize levy arranges and so forth along these lines this thought, sensible meter learning examination, is upheld with a perused of future utilize.

2. LITERATURESURVEY

There is creating excitement for watching practices of force customers in both the private and business divisions. With the approach of high-assurance time-course of action control ask for data through advance metering mining this data could be excessive from the computational point of view. This is proposed by Ramon Granell, Colin J. Axon on 2015 in 'Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles'[1]. One of the notable frameworks is clustering, however depending upon the computation there course of action of the data can have a basic influence on the consequent gatherings. This paper exhibits how common assurance of drive demand profile influences the way of the gathering method, the consistency of pack support (profiles indicating near lead), and the efficiency of the gathering methodology. This work uses both rough data from family use data and built profiles. The motivation for this work is to upgrade the gathering of force load profiles to help perceive customer sorts for collect diagram and trading, fault and blackmail recognizable proof, ask for side organization, and essentialness efficiency measures. The key control for mining extensive enlightening lists is the methods by which little information ought to be used to get a strong result, while keeping up assurance and security.

With the development of savvy network, loads of recharge capable vitality assets, for example, wind and sunlight based are conveyed in power framework, which is why Pei Zhang, Xiaojun Wang, and Sheng Bi proposed a power framework stack on 2015 with differed complex than before which will acquire difficulties here and now stack guaging territory on 'Short-Term Load Predicting Based on Big Data Technologies'[2]. To beat this issue, this paper proposes another transient load guaging system in view of enormous information advancements. To start with, group an assumed name for each shaped to characterize every day stack designs for individual burdens utilizing keen meter information. Next, an affiliation examination is utilized to decide basic influential variables. This is trailed by the utilization of a choice tree to set up classification rules. At that point, fitting estimating models are decided for various load designs. At long last, the determined aggregate framework load is acquired through a conglomeration of an individual load's anticipating comes about. Contextual analyses utilizing genuine load information demonstrate that the proposed new structure can ensure the precision of here and now stack anticipating inside required points of confinement.

There are a few examples based bunching strategies which are utilized for various applications, for example, design acknowledgment, information mining, and so on. As of late M. K. Sheikh-El-Eslami, and S. M. Bidaki on 2009 said some of these strategies are executed in power framework contemplates, particularly to cluster stack bends for planning appropriate duties, request reaction programs determination, and so forth in the paper 'Improving WFA K-means Technique for Demand Response Programs Applications'[3]. Decision of the best bunching technique for certain application is a standout amongst the most imperative issues which is case subordinate and ought to be considered in utilizing of grouping burden bends. Request reaction projects are broadly utilized as a part of force framework for various applications, for example, crests cutting, request diminishment, and so on since request reaction projects are highlighted with various attributes. Along these lines, choice of appropriate projects for various client classes is of extraordinary significance. In this paper, an enhanced weighted fluffy normal (WFA) K-implies with the end goal of interest reaction programs applications is produced. This technique is actualized on 316 load bends of Tehran dispersion arrange and the outcomes are examined.

This paper by Carlos León &RocíoMillán on 2011 proposes a far reaching system to identify non-specialized misfortunes (NTLs) and recoup electrical vitality (lost by variations from the norm or misrepresentation) by method for an information mining investigation, in the Spanish Power Electric Industry on 'Variability and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies'[5]. It is partitioned into four areas: information determination, information preprocessing, enlightening, and prescient information mining. The creators demand the significance of the learning of the specific qualities of the Power Company client: the primary components accessible in databases are depicted. The paper presents two inventive measurable estimators to append significance to changeability and pattern investigation of electric utilization and offers a prescient model, in view of the Generalized Rule Induction (GRI) show. This prescient investigation finds affiliation manages in the information and it is supplemented by a twofold Quest tree classification technique. The nature of this structure is delineated by a contextual investigation considering a genuine database, provided by Endues a Company.

A keen home is likely sooner rather than later, therefore 'A Time Based Markov Model for Automatic Position-Dependent Services in Smart Home'[6] was introduced by Yi Yang, Zhiliang Wang for a critical fixing in a savvy domain, for example, a house is programmed administrations, which implies home framework itself could know or foresee what the occupant need to do, thus give tenant the administrations naturally. Many inquiries about uncover that the greater part of the administrations in savvy home are area subordinate so the programmed administrations must be based on the area mindfulness. In this paper, we display occupant area design as a period based markov show (TMM). The reenactment result demonstrates that contrasted with alternate models, the TMM has an arrangement of advantages, for example, less time unpredictability, higher forecast exactness and speedier unions rate. These advantages make TMM meets the necessities of programmed administrations in brilliant home.

The expanding US organization of private progressed metering framework (AMI) has made hourly vitality utilization information broadly accessible through 'Household Energy Consumption Segmentation Using Hourly Data'[7] which was proposed by JungsukKwac, June Flora, and Ram Rajagopal on 2014 utilizing CA shrewd meter information, we explore a family unit power division procedure that uses an encoding framework with a pre-handled load shape lexicon. Organized methodologies utilizing highlights got from the encoded information drive five specimen program and strategy significant vitality way of life division systems. We additionally guarantee that the systems created scale to substantial informational indexes.

Bunching strategies are progressively being connected to private savvy meter information, which gives various vital open doors for circulation organize administrators (DNOs) to oversee and arrange low-voltage systems. The paper titled 'Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data'[8] by Stephen Haben, Colin Singleton on 2012 addressed this issue. Grouping has various potential points of interest for DNOs, including the identification of appropriate contender for request reaction

and the change of vitality profile displaying. In any case, because of the high stochastic city and inconsistency of family level request, definite investigation is required to define proper credits to bunch. In this paper, we show inside and out examination of client keen meter information to better comprehend the pinnacle request and real wellsprings of changeability in their conduct. We find four key eras, in which the information ought to be broke down, and utilize this to shape important characteristics for our grouping. We display a finite blend demonstrate based bunching, where we find ten particular conduct bunches portraying clients in view of their request and their inconstancy. At long last, utilizing a current bootstrap system, we demonstrate that the bunching is dependable. To the creators' learning, this is the first time in the power frameworks writing that the example strength of the bunching has been tried.

3. IMPLEMENTATION

Proposed concept deals with providing database by using Hadoop tool we can analyze no limitation of data and simple add number of machines to the cluster [Figure 1 & 2] and we get results with less time, high throughput and maintained cost is very less and we are using joins, partitions and bucketing techniques in Hadoop

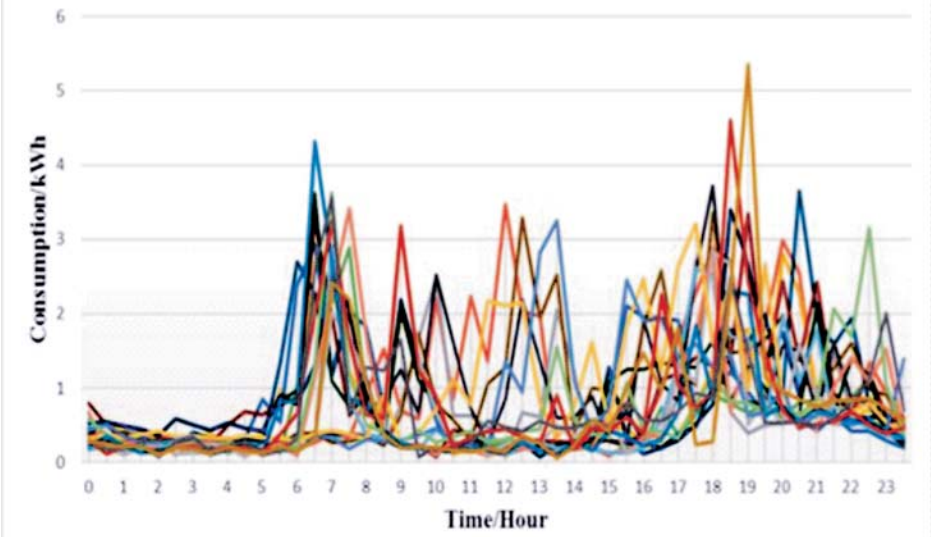

Figure 1: Time/hour consumption(consumer 1)

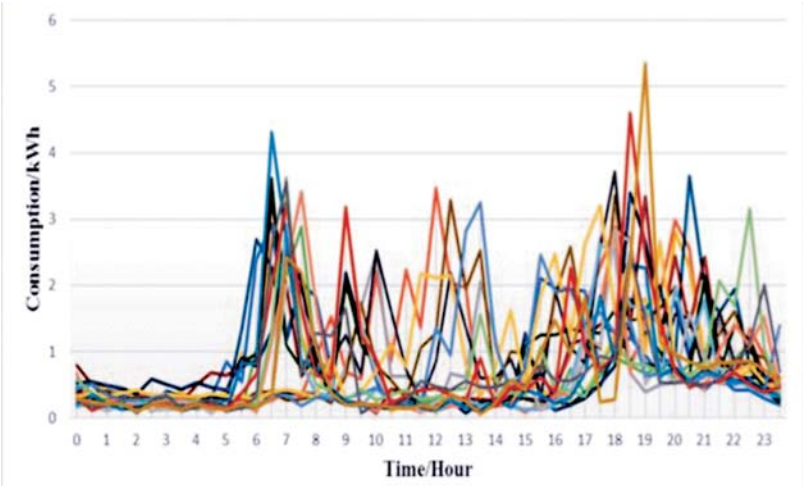

Figure 2: Time/hour consumption (Consumer 2)

4. ALGORITHMS

4.1. Symbolic mixture approximation

SAX is the first symbolic representation for time series that allows for dimensionality reduction through and indexing with a lower-bounding distance measure. In classic data mining tasks such as the clustering, classification, index, etc., SAX is as good as well-known representations such as the Discrete Wavelet Transform (DWT) and Discrete Fourier Transform (DFT), while requiring low amount of storage space. In addition, the representation allows researchers to avail of the wealth of data structures which are available and algorithms in bioinformatics or text mining, and also provides solutions to many challenges associated with current data mining tasks. One example is motif discovery, a problem which we have defined for time series data. There is great potential for extending and applying the discrete representation on a wide class of data mining tasks.

$$\vec{x}_i = \frac{1}{k_i - k_{i-1}} \sum_{j=k_{i-1}+1}^{k_i} x_j'$$

where j is the index of the normalized load data; i is the index of the transformed PAA load data; k_i is the i^{th} time domain breakpoint; and $x_1^{'}$ is the average value of the i^{th} segment

5. METHODOLOGY

The proposed policy performs well in the all-inclusive community and moreover in sub-populaces [Figure 3]. Comes about validate that the planned show altogether enhances expectations over set up gauge strategies dissecting power utilization. The independent of this analysis was to observe the sum of units expended in most recent four years year as the estimate for the next year. and how much sum they paid past four

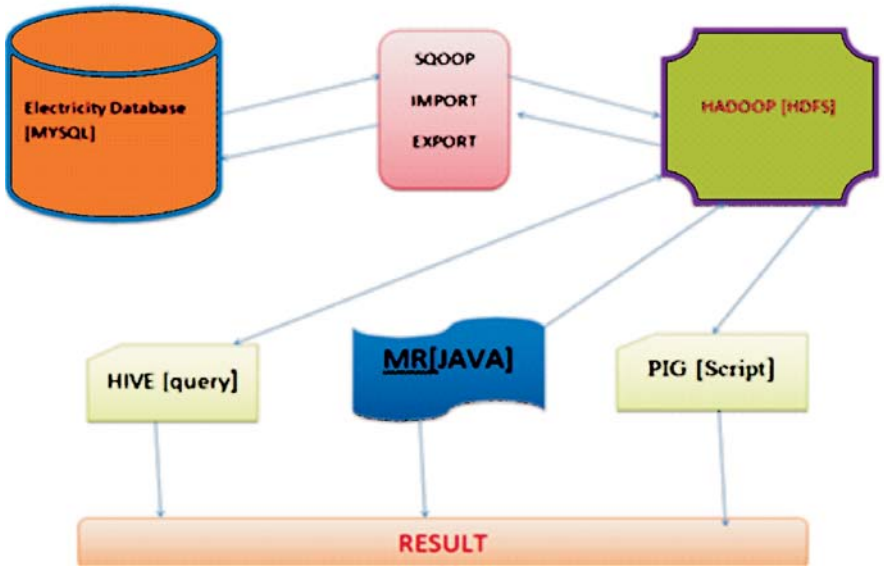

Figure 3: General methodology

5.1. Data Preprocessing Module

In this module [Figure 4] we have to create Data set for Electricity Consumption it containsset of tables 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.

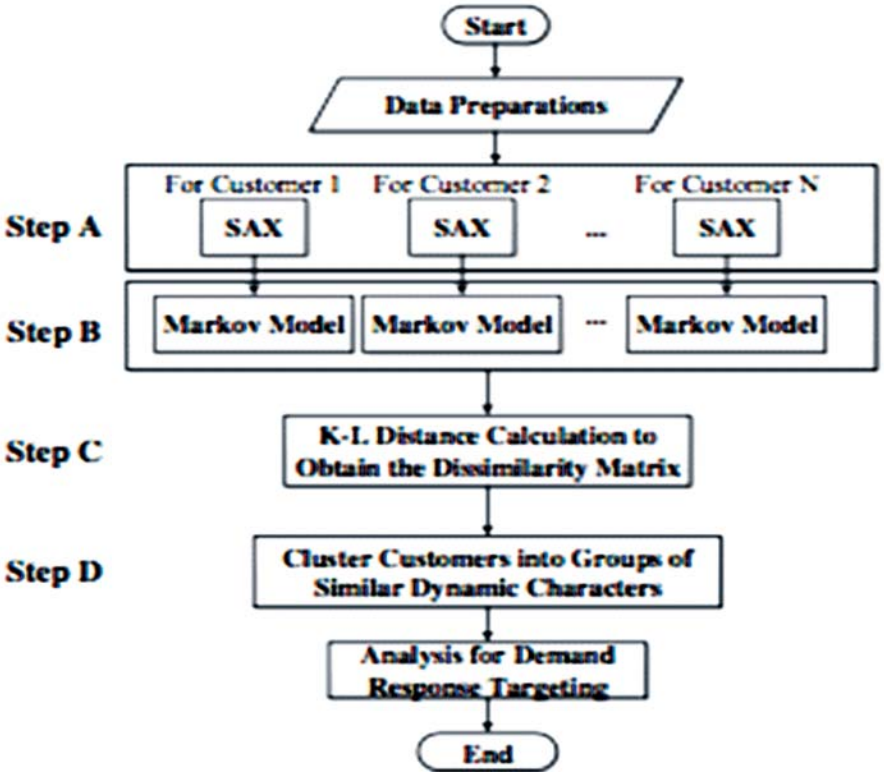

Figure 4: Complete setup

5.2. Data Migration Module with Sqoop

Now we are ready with dataset. So now our aim is transfer the dataset into Hadoop (HDFS) that will be happen in this module.Sqoop is an application for transporting data among personal stores and Hadoop.In this module, 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 and with specific condition that will be support by Sqoop Tool and data will be stored in Hadoop (HDFS).

5.3. Data Analytic Module with Hive

Hive is an informative product house framework for Hadoop. It runs SQL like questions called HQL (Hive inquiry dialect) which get inside changed over to delineate occupations. Hive was created by Facebook. Hive underpins Data DefinitionLanguage, Data Manipulation Language and client characterized functions.In this module, we need to investigation the dataset utilizing HIVE device which will be put away in Hadoop (HDFS). For examination dataset HIVE utilizing HQL Language. Utilizing Hive, we perform Tables manifestations, joins, Partition, Bucketing idea. Hive investigation the main Structure Language.

5.4. Data Analytic Module with Pig

Apache pig is an abnormal state information stream stage for execution of Map Reduce projects of Hadoop. The dialect for Pig will be Pig Latin. Pig handlesboth structure and unstructured dialect. It is additionally top of the guide decrease handle running foundation. In this module likewise utilized for breaking down the Data set through Pig utilizing Latin Script information stream language.in this additionally we are doing all administrators, capacities and joins applying on the information see the outcome.

5.5. Data Analytic Module with MapReduce

MapReduce is a preparing strategy and a program demonstrate for appropriated processing in light of java. The MapReduce calculation contains two vital undertakings, in particular Map and Reduce.In this module likewise utilized for dissecting the informational index utilizing MAP REDUCE. Outline Run by Java Program.

5.6. Analyzing data through r language

R is a dialect and environment for factual processing [table 1]and i6 illustration [Figure 5 & 6]. It is a GNU venture which is like the S dialect and environment which was produced at Bell laboratories by john chambers and colleagues.

Bill_No	*Ser_No*	*Pre_Red*	*Pri_Red*	*Units*	*Con_Cha*	*Cust_Cha*	*Ele_Char*
1	BI001 SC-000101	5000	4800	200	400	10	20
2	BI002 SC-000102	5000	4700	300	600	10	20
3	BI003 SC-000103	5000	4500	500	1000	10	20
4	BI004 SC-000104	5000	4300	700	1400	10	20
5	BI005 SC-000105	5000	4400	300	600	10	20
6	BI006 SC-000106	5000	4200	800	1600	10	20

Figure 5: Sample data used

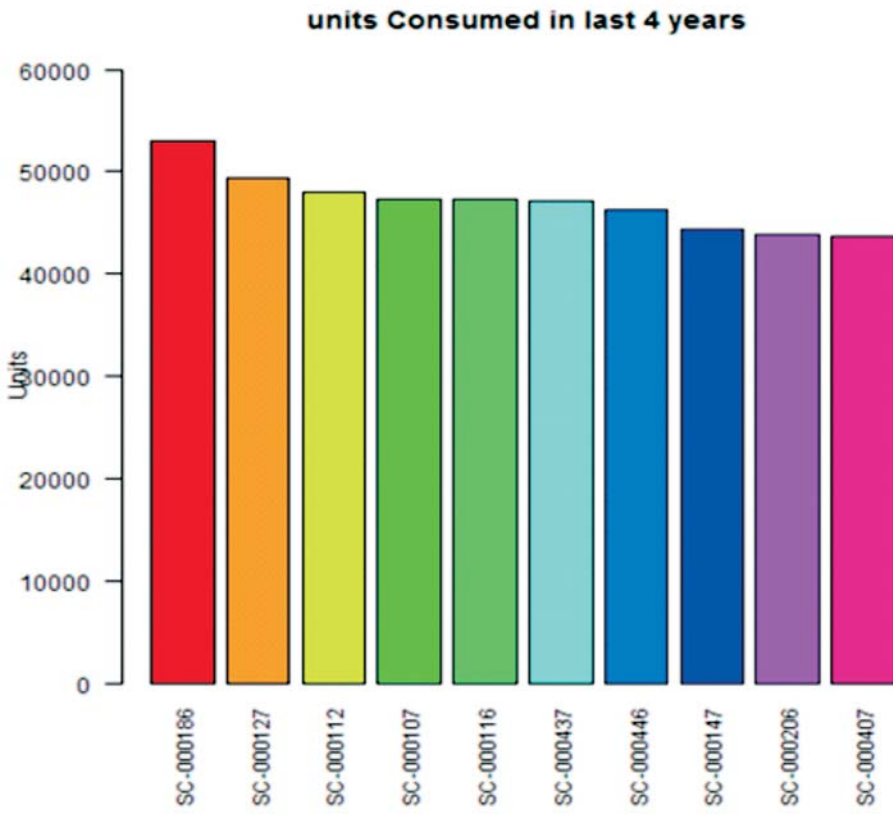

Figure 6: Pie Chart representation

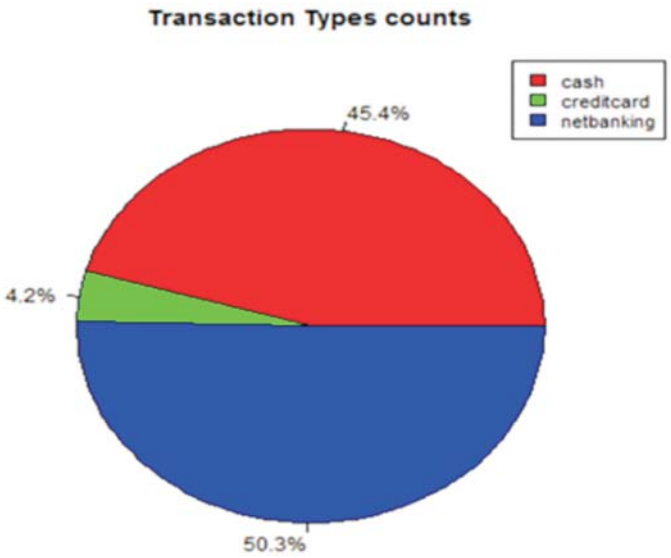

Figure 7: Graph representation

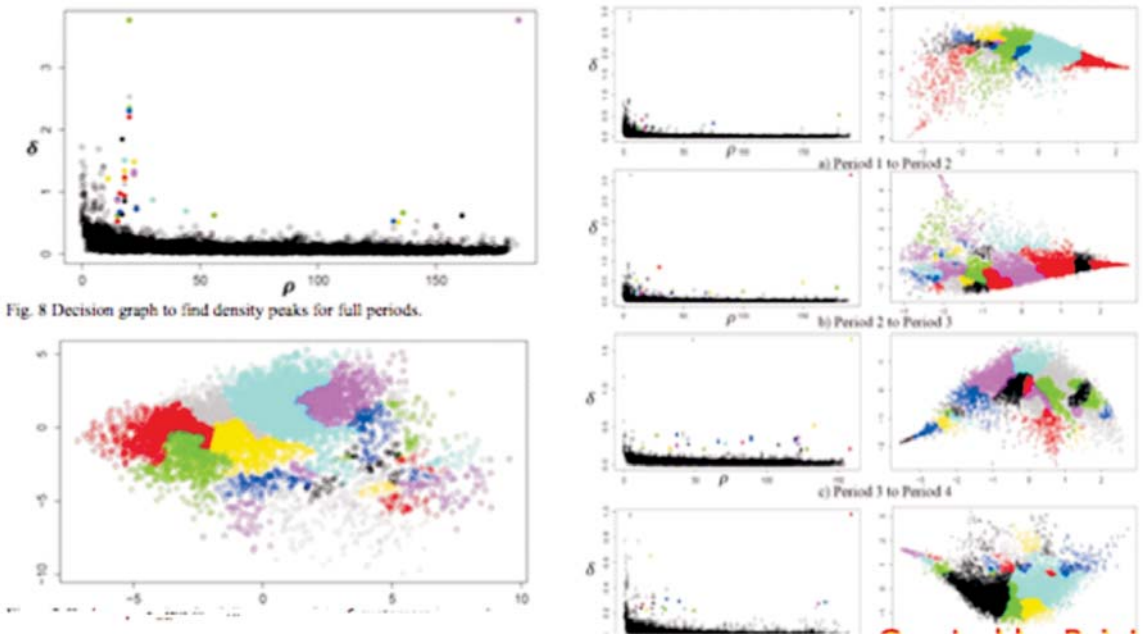

Fig. 8 Decision graph to find density peaks for full periods.

Figure 8: Density peak identification

6. CONCLUSION

A method for predicting future consumption in electricity has been developed using features extracted [figure 7] fromcustomer previous consumption for the following year. Rational energy use could be for a bigger cluster of firms, municipalities and public organizations attributable to the gainin importance of the energy prices and environmental problem, thus correct data regarding their consumption of good meter sends energy consumption knowledge at tiny intervals leading to generating huge knowledge. Time and storage are vital factors that have an effect on lots on building any application. The answer for handling such huge knowledge is Hadoop.

The goal of this study was to analyze how much of units consumed in last four years and how much amount they paid previous four year as the forecast is required by them along with and its distribution between totally different activities. a complete image of their energy use, potential for savings, in conjunction with prices is given to them by good meter knowledge analytics, enabling effective energy management.

REFERENCES

[1] USA Division of Energy, Smart Network/Branch of Energy, http://energy.gov/oe/innovation advancement/savvy matrix, 2015

[2] I. P. Panapakidis, M. C. Alexiadis, " Stack profiling in the deregulated power showcases: An audit of the applications," in European Energy Market (EEM), 2012 ninth International Conference on the, 2012,pp. 1-8.

[3] R. Granell, C. J. Axon and D. CP.J H. Wallom, " Effects of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles," IEEE Trans. Power Systems, vol. 30, pp. 3217-3224, 2015.

[4] C.M Leon, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, and R. Millan, "Inconstancy and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies," IEEE Trans. Power Systems, vol. 26, pp., 2012.

[5] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, " Stack profiling and its application to request reaction: A survey," Tsinghua Science and Technology, vol. 20, pp. 117-129, 2015.

[6] R. Li, C., F. Li, G. Shaddick, and M. Dale, " Advancement of Low Voltage Network Templates-Part I: Substation Clustering and Classification," IEEE Trans. Power Systems, vol. 30, pp. 3036-3044, 2015.

[7] K. Zhou, S. Yang and C. Shen, " A survey of electric load grouping in brilliant lattice environment," Renewable and Sustainable Energy Reviews, vol. 24, 2014.

[8] G. J. Tsekouras, P. B. Kotoulas, C. D. Tsirekis, E. N. Dialynas, and N. D. Hatziargyriou, " An example acknowledgment technique for assessment of load profiles and commonplace days of expansive power clients s," Electric Power Systems Research, vol. 78, pp. 1494-1510, 2008.

[9] M.V. Verdu, M.L. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, " Arrangement, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps," IEEE Trans. Power Systems, vol. 21, pp., 2009.