Annals of Emerging Technologies in Computing (AETiC)
Vol. 1, No. 1, 2017

Article

Analysis of Home Energy Consumption by K-Mean

Fahad Razaque[1], Nareena Soomro[1,*], Javed A. Samo[1], Huma Dharejo[1] and Shoaib Shaikh[1]

[1]Department of Computing, Indus University Karachi, Sindh, Pakistan
(fahad‖nareena‖j.ahmedsam‖huma)@indus.edu.pk, skshaikh@outlook.com
*Correspondence: nareena@indus.edu.pk

Received: 02/08/2017; Accepted: 15/09/2017; Published: 01/10/2017

Abstract: The smart meter offered exceptional chances to well comprehend energy consumption manners in which quantity of data being generated. One request was the separation of energy load-profiles into clusters of related conduct. The Research measured the resemblance between groups them together and load-profiles into clusters by k-means clustering algorithm. The cluster met, also called "Gender (Male/Female), House (Rented/Owned) and customers status (Satisfied/Unsatisfied)" display methods of consuming energy. It provided value information aimed at utilities to generate specific electricity charges and healthier aim energy efficiency programs. The results show that 43% extremely dissatisfied of energy customer is achieved by using energy consumption.

Keywords: Load-Profiles; K-means; Clusters; Data Science, Data Set

1. Introduction

In general accord humanoid actions was face bad influence of the environment and have to speed up both weather alteration and global warming in the world nowadays. The energy was necessary for the illumination and warming, ventilation and air-conditioning (HVAC) systems in house creations through ecological terrorizations was concentrated. Housing and profitable construction was accountable for up to 32% of the entire last energy intake, from the IEA (International Energy Agency)[1].

Data shows that elderly constructions joint with growing house action on an industrial country side were reason energy ingesting to soar in the close upcoming contain inefficient energy management and enhance the bad influences related with consumption. Furthermore, keyword energy cost needs the implementation of intellectual policy to adjust and decrease energy ingesting besides to discovery another and maintainable energy base. The improvement of present energy administration processes and substructures, also it was comprised use of inexpensive energy causes. One of the most significant problems for energy firm, the concluding comprise of transportation and optimization of energy generation on basis of user request [2].

The computer-aided methods were freshly derived into limelight. Researched that was how to automatically find out non-trivial information as of Data Science, Data, which involves extensive range of techniques and more composite datasets, specially improved data consciousness in firms was led to the growth of answers on the basis of Data Mining. Data Science was constructing energy management by address difficulties, for instance: (i) the energy of prediction request so that dissemination and get used to making. (ii) The examination of constructing maneuvers besides of tools status and maintenance costs and failures to optimize maneuvers. (iii) The energy ingesting configurations by detection of fraud and make customized commercial suggestions. It required gathering data related to user behavior and building operation user behavior. The data were interpreted to implement adopted energy management strategies [3,4,5].

This Research was Data Science (DS) methods and clarify how they were members to pact with complex dare countenance via house energy. The classification and clustering approach were frequently used, such as temporal and frequent pattern discovery by load prediction.

1.1 Objectives

- To analysis the data of energy approach that establishes reliable house energy consumption which was used by people.
- To learning house occupant behavior, such as quantitatively identifying the effect of occupant behavior on house energy consumption, and identifying the occupant behavior that may be modified to save energy
- To examine the classification among house energy data, and extracts beneficial knowledge from them to better reduce energy consumption and comprehend system operation.

2. Results

The research was interpreted and managed the information in classify to acquire respected comprehension. The method begins through group of unrefined data, showed in Figure.1. Later, it was required for data clean, and selects the compartment to pertinent info. Used for determination, the researcher relates cleans to devise query to eradicated immaterial info. It was as extra basis of info may be merged moreover integrated among innovative data toward deliver additional comprehension. After information was organized for utilize, investigative analysis like visualization tool was assist choose which approaches was most operational to acquire the wanted information. The concluding procedure was guide towards the result to facilitate executive, which once more, may depend on visualization. DS includes set of tools and methods which pursue dissimilar aims and proceed from diverse circumstances. The absolute most well-known techniques are regression, clustering, classification and association rule mining, also beneficial in provided that solutions for house energy difficulties.

2.1. Data Science (DS)

Data science (DS), technological tools was profited broad assortment of areas, and Management and power Productivity was no exemption. Growths in different zones of ICT (information and communications technology), for instance manage and computerization, real-time monitoring, smart metering, also Data Science, was a tremendous influence on this arena. DS constructs algorithms and systems to detect patterns, discover knowledge, and create beneficial predictions and perceptions from extensive info. It engage entire procedure of data analysis, which arises among DC (data cleaning) and, mining and widen to DA(data analysis), summarization and explanation. The consequence was forecast of innovative standards and imagining. DS accordingly comprises statistical analysis and mathematical, collective with information technology implements.

2.2 Data classification (DC)

It was a procedure of organizing data into categories for its most operative and effectual use. DC system creates crucial data easy to retrieve and find. It was define what kinds and standards the house would use to specify the roles and classify data and duties of members within the house. It comprises of predicting a certain outcome based on given input[6,18]. The algorithm procedures a training set comprised the respective outcome and, a set of attributes usually known as goal or prediction attribute. SVM (Support Vector Machine)[7] was classification by discovery hyper plane that maximizes the margin between two classes. The vectors describe the hyper plane was supported vectors. A model of SVM analysis ought to create hyper plane that fully separated the vectors into two non-overlapping classes. It may outcome in model with various cases that model was not categorize appropriately. The SVM find the hyper plane that maximizes the margin and minimizes the misclassifications. The Naive Bayes classification was easy probabilistic that found on Bayes Algorithm among sturdy and naïve autonomy hypothesis. It was individual of necessary text classification method among diverse application, sexually explicit content detection, and private electronic mail sorted, manuscript classification, and email spam, language and emotion recognition[8, 9, 10]. Although naïve devise and more than generalized hypothesis to method used, Naive Bayes executed fine in a lot of complicated real-world tribulations.

2.3 Regression

It was data mining function that predicted numeral such that distance, temperature, Age, weight, income, or sales might all be predicted by regression techniques. The regression task was initiated thru data set in which the targeted value was identified[11]. It was verified in calculating different statistics that measured the difference between expected and predicted values. The Temporal data for regression development was classically separated into two data sets: one house model, the other for testing model. KNN (K-nearest neighbor)[12] was utilized for both regression and classification predictive issues. It stored the whole training dataset which was used as its representation-near neighbor creates prediction just-in-time by computing the parallel between an input sample and each training case. It was a good idea to rescale data, for example with normalization, when used KNN.

2.4 Clustering:

It was a data mining approach used to place element data into an associate group without advance knowledge of group definitions. Clustering algorithm used associations that exist in the data and within in the clusters that the algorithm identified[13]. K-means was one of the simplest unsupervised learning algorithms that were exploit as it contain without a label info. The aim of algorithm was to discover cluster inside data, through quantity of group denoted by K (variable). The algorithm worked iterative to give every point of data to any of K grouped found on featured that were delivered. Data dot was characteristic resemblance as cluster. The K-means cluster method as result: The cluster K as center of mass, which was applied toward tag latest data. K-median technique to clustering data attempted to minimize the 1-norm distances between each fact and it was closest cluster midpoint. The minimization of distances was obtained by location the center of each cluster to be the median of all points in that cluster.

2.5 Association

It was inference term of type ($X \rightarrow Y$), as X and Y was item set[14]. The association regulation might was considered as term of supports and confidence. Support defines how frequently regulation was pertinent to specified data sets, while confidence define how frequently objects into Y appeared in businesses that include X.

2.6 Sequence discovery

Sequence discovery contains finding every subsequence which appeared regularly in a longer time sequence. It was transferred from genetic factor analysis in bioinformatics[15].

2.7 Anomaly Detection

Anomaly detection was technique used to identify unusual patterns that was not conform to expected behavior, called outliers. The detection of anomalies seeks to find abnormal subsequences in a series[16].

2.8 System Architecture

Data set was databases stored the utilization data in over all, comprise houses info intended for all sample instance gender, rented and owned house. Pre-processing Data to cleaned database which was removed incorrect interpolated, as probable, data omitted and contain layout toward input. Outlier Detection was execution of method to remove and identify erroneous data, discerning as of accurate data which illustrate anomalous value. Clustering represented collective appliance of scikit-learn and k-mean. Report was key chore to drive info of resulted, to be displayed graphically.

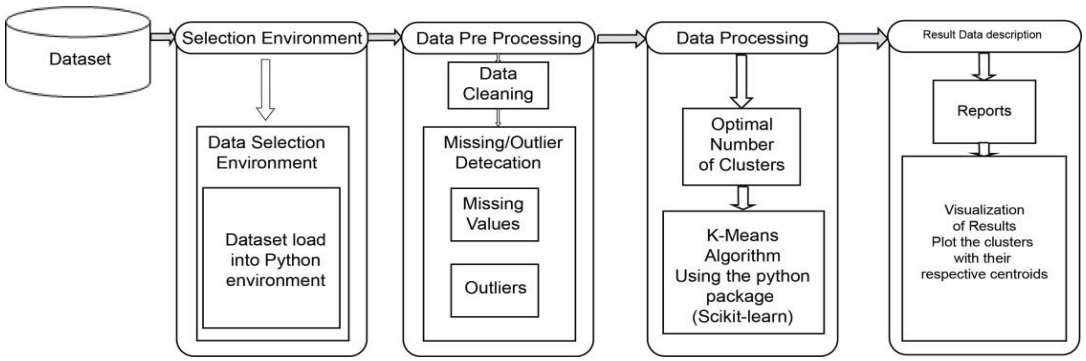

Figure 1. System Architecture.

2.9 Data Selection Environment

Anaconda was fermium open source allocation of programming languages R and Python used for extensive predictive analytics, data processing, and scientific computing, that goals toward simplified package executive and utilization. Python was technical calculated and computational modeling; it required further libraries such as package that was not standard library element. Using as, generate plot, derived matrices, and used specialized mathematical process. Its packages following; linear algebra and matrices as Numeric Python (numpy), generate plot data as Plotting Library (matplotlib), arithmetical as Scientific Python (scipy). The packages numpy, scipy and matplotlib were assemble pebbles of computational effort and very extended[17].

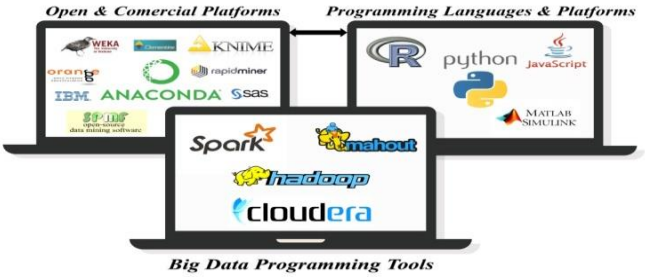

Figure 1. Selected platform & programming tools which implemented data science technique.

3. Results

3.1 House Energy Consumption as Dataset

The energy consumption was measured by dataset was downloaded from Kaggle Datasets, was platform used for analytics competitions and predictive modeling wherein researchers post data, data miners and statisticians compete to produce the best models for describing and predicting the data. The best place to discover and seamlessly analyze open data [n]. Dataset comprises of information of district, Gender (Male/Female), House (Rented/Owned) and Customers status (Satisfied/Unsatisfied). In order to work on dataset that is huge in size so the author needs to create samples from the entire population. Only limited data (a few records from) has been processed in order to achieve quick and reliable results. The selected dataset is processed using K-mean clustering technique in order classify the data.

3.2 K-Means using Clustering

The evaluations results of the clusters specified in k-mean illustrate to two clusters. Figure 3 grants every dot belong to clusters. It was simple to understand the high resemblance among dots of identical cluster, and considerable distinction while evaluate through other clusters.

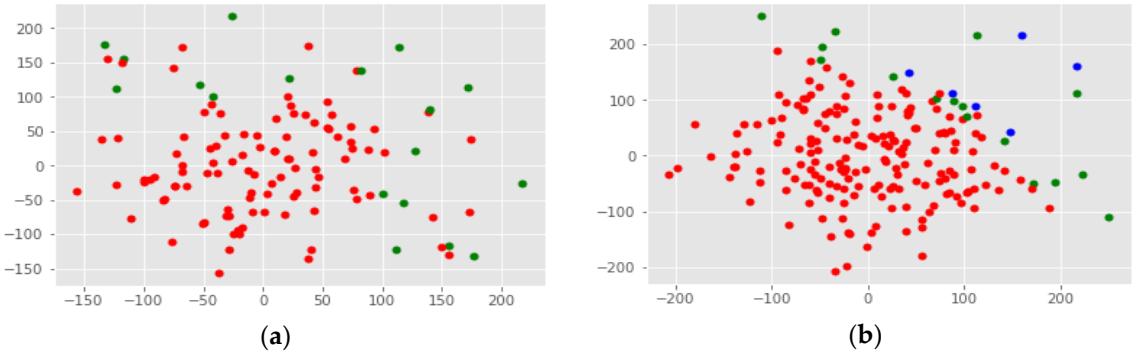

Figure 3. Load dot for the two clusters, red dots as male, green dots as female and blue dots as family members. (a) Gender-wise Cluster: K-mean clustering algorithm classifies the gender-wise consumption of the energy and calculate sum of male and female users. (b) House Owner/Rented Cluster: the power consumption of rented and owned house was categorized.

As shown in Figure 4, 93.4% owner users were satisfied from the services of Electricity Company.

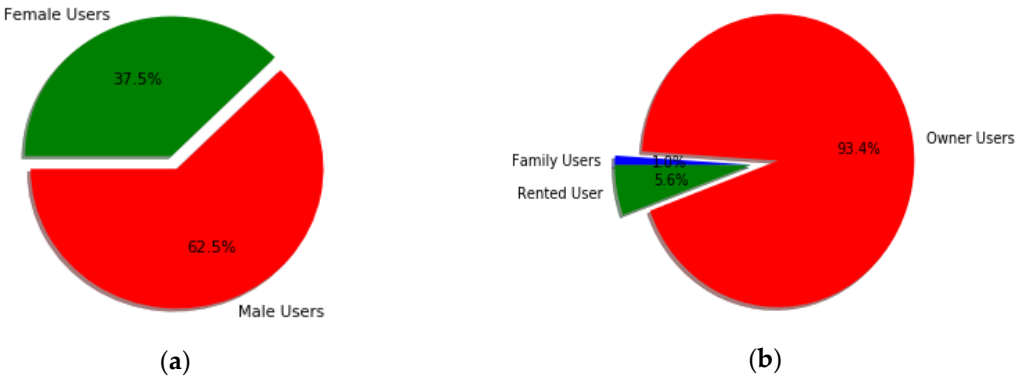

Figure 4. Distribution of electricity end-use depending on the gender, family, rented, owner user in the household (a) Male/Female Energy Service satisfaction Ratio. Shows that 62.5% male users were satisfied from the services of Electricity Company (b) House Rented/Owner Energy Service satisfaction Ratio.

Figure 5 Shows that distribution of electricity end-use depends on the Customer in the household, 43.9 extremely dissatisfied.

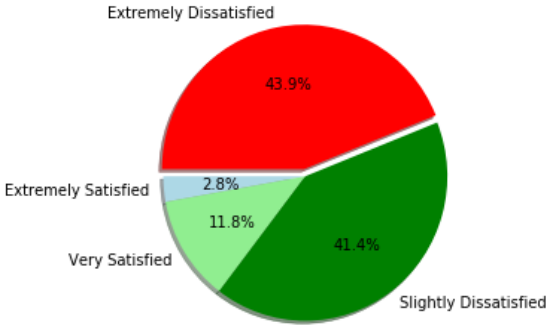

Figure 5. Overall Energy Service provider's satisfaction Ratio.

4. Conclusion

The current important problems affected depressingly on evolvement of Pakistan increase of consumption of electricity. The performance of energy sector was altered by such prevalent problem

as a small enterprise was closing their organizations because they could not afford alternative sources of electricity. Let the consumer beware of varying prices, which Average firms were shifted priced load on the assumed of the customer. There was 6000-7000MW lack of electricity supply in each season contains alteration was owed to huge the load shedding of nearly 8-12 hours in urban areas and 14-16 hours in rural areas. Ensuing keen on the customers and economic loses which turns their income to collapse. The solution as running energy zone well on earth was through decreasing power lack. Power lack was reduced using fostering investment. The research suggested method and efficiently assist to developed technique as data analysis intended for mine concealed knowledge as of house related data, classify toward report intended for communications among influence and house energy utilization issue. Apparent and thorough indulgent of such communications might give necessary direction into reduced utilization house energy. Humanized and growing bind among upcoming energy wealthy state was not mistreated.

References

[1] WBCSD. Energy efficiency in buildings: business realities and opportunities. Technical report, The World Business Council for Sustainable Development, 2007.

[2] Allcott H, Mullainathan S. Behavior and energy policy. Science 2010;327(5970):1204-5.

[3] Kusiak A, Li M, Zhang Z. A data-driven approach for steam load prediction in buildings. Appl Energy 2010;87(3):925-33

[4] Prahastono I, King D, Ozveren CS. A review of electricity load profile classification methods. In: Proceedings of the 42nd international universities power engineering conference, 2007. UPEC 2007. Sept 2007 pp. 1187-91.

[5] Yu Z, Haghighat F, Fung BCM, Yoshino H. A decision tree method for building energy demand modeling. Energy Build 2010;42(10):1637- 46.

[6] Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D. Top 10 algorithms in data mining. Knowl Inf Syst 2008;14(1):1-37.

[7] Bennett KP, Campbell C. Support vector machines: hype or hallelujah?. ACM SIGKDD Explor Newsl 2000;2(2):1-13

[8] Andrew YN, Michael IJ. On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. In: Dietterich TG, Becker S, Ghahramani Z. editors, Advances in neural information processing systems, vol. 14. MIT Press, 2002. pp. 841-8.

[9] Banzhaf W, Francone FD, Keller RE, Nordin P. Genetic programming: an introduction: on the automatic evolution of computer programs and its applications. Morgan Kaufmann Publishers Inc; 1998.

[10] Kriesel D. A Brief introduction to neural networks. Available at ⟨http://www. dkriesel.com⟩, 2007.

[11] Chatterjee S, Hadi AS. Regression analysis by example. John Wiley & Sons; 2013.

[12] Nitin Bhatia V. Survey of nearest neighbor techniques. Int J Comput Sci Inf Secur 2010;8(2):302-5.

[13] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv 1999;31(3):264-323.

[14] Agrawal R, Imielinski T, Swami A. Mining associations between sets of items in massive databases. :In: ACM-SIGMOD international conference on data. 1993.pp. 207–16.

[15] Mabroukeh NR, Ezeife CI. A taxonomy of sequential pattern mining algorithms. ACM Comput Surv 2010;43(1):1-41.

[16] Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv 2009;41(3)

[17] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst 2005;20(2):596–602.

[18] Verdú SV, García MO, Senabre C, Marín AG, Franco FJG. Classification, filtering, and identification of electrical customer load patterns through the use of selforganizing maps. IEEE Trans Power Syst 2006;21(4):1672-82.