International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

Predicting and Analysis Electrical Energy Consumption by Using Data Mining Algorithms

Mohamed Abdulhameed Mohamed,
Asst.Prof.Dr. Sefer Kurnaz
213720626@ogr.altinbas.edu.tr , sefer.kurnaz@altinbas.edu.tr
1,2 Electrical and Computer Engineering, Altinbaş University

Abstract

In this paper, This study accumulates and presents pertinent data in a variety of formats for the purpose of classifying power consumption by activity. It has been fascinating to determine how to graphically depict each stage and classify electrical data in a way that satisfies all requirements. This facilitates the detection of issues and anomalies and simplifies the process of comparing electricity consumption. In the coming decades, electricity will continue to gain prominence as a primary energy source. Smart circuits and smart meters offer numerous benefits to both the utility and the customer. This study combined classification (specifically five algorithms) and clustering theory, with energy consumption per hour (%) functioning as the common framework, in order to classify electricity use based on the similarities of electrical load profiles. After classifying everyone, we will be able to offer each subset advice on how to save money and energy. Consequently, individuals will be more aware of their electricity consumption and motivated to take steps to reduce it. A post-clustering and classification study that uses Weka for analysis and result generation employs an iterative technique based on computational classification calculation to identify anomalies and reallocate them to more acceptable classes. When compared, Decision Tree, Support Vector Machine, Naive Bayes, Random Forest, and Hybrid are the five classification techniques that yield identical results. Categorizing power consumption permits a deeper understanding of the relationship between human behavior and electricity consumption. It improves the quality of the energy conservation consulting service and the customer experience by providing timely, relevant advice based on the unique characteristics of each individual consumer.

Keywords: prediction, electricity consumption, smart meter, classification, clustering, intelligent system, data mining, electric usage.

Introduction

The expanding significance of energy means that its importance to the advancement of human civilization will increase. Adjusting to altering customer requirements and regulatory constraints presents significant challenges to the contemporary energy sector. Recently, the policy community has paid considerable attention to environmental issues [1]. Three key international environmental conferences, Stockholm 1972, Rio 1992, and Johannesburg 2002, offer proof [2]. From these

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

discussions arises sustainable development, an all-encompassing strategy that simultaneously addresses economic, social, and environmental challenges [3]. Figure 1 illustrates the importance of energy to all sustainable development concepts [4-5].

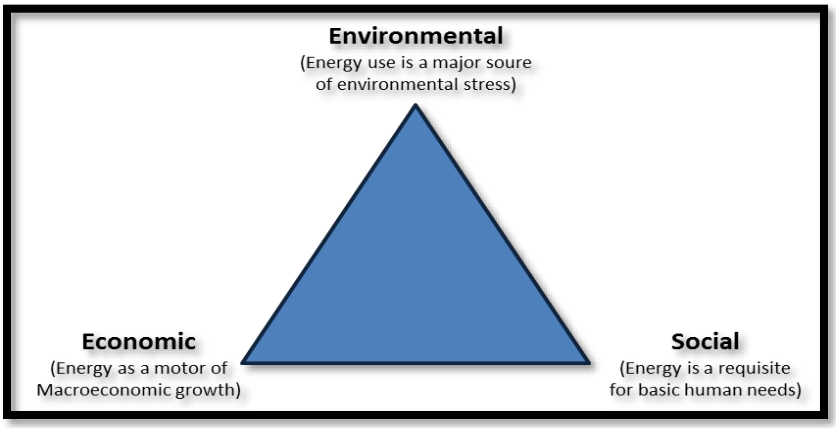


Figure 1: Energy and Sustainable Development linkages [5].

Historically, conventional energy sources, primarily fossil fuels such as oil, coal, and natural gas, have provided abundant and affordable energy through their extraction, refining, and consumption. Nonetheless, pervasive, regional, and local exploitation of finite, nonrenewable resources has resulted in extensive environmental damage. This causes climate change, acid deposition, and urban pollution [6].

Solar (photovoltaic and thermal), wind (wind and biomass), and hydropower (hydroelectric) are examples of renewable energy sources. This energy's production, conversion, and use have negligible environmental impacts, reducing the rate of global warming.

Figure 2 depicts the relationship between energy and GDP (a macroeconomic metric). Figure 2 depicts the relationship between energy and GDP (a macroeconomic metric). Industrial operations that convert basic materials into finished products or services frequently rely on the use of energy [7].

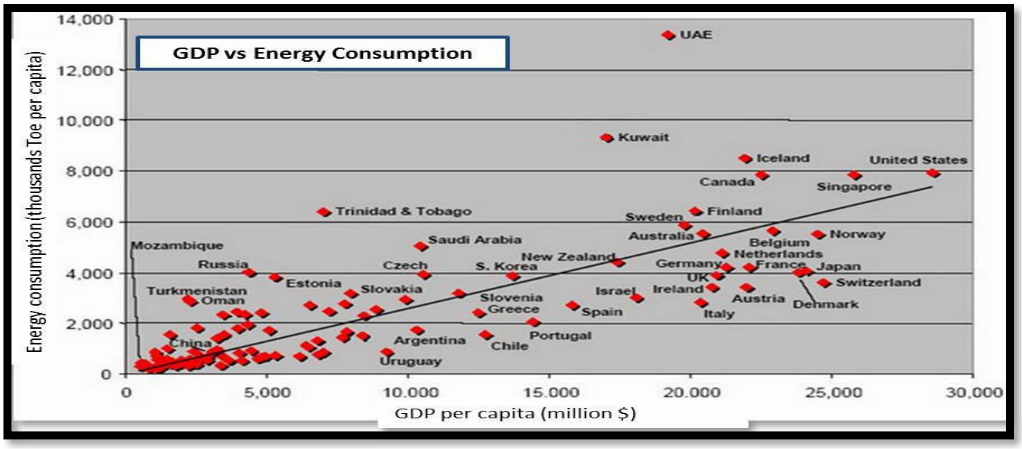


Figure 2: Correlation GDP vs Energy consumption. Energy Consumption per capita in thousands of Tons of equivalent oil (toe) and Gross Domestic Product per capita in million $[5].

The preceding link may appear obvious, but the following two pieces of evidence are even more compelling: Both industrialized and emerging nations are decreasing their energy intensity over

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

time, as shown in Figure 3 Energy intensity is calculated by dividing total energy consumption by GDP.

Energy efficiency is the polar opposite of energy intensity for a given economy. Among the top 40 economies, there are two distinct categories: developed economies, with a high GDP per capita but low efficiency, and developing economies, with a low GDP per capita and high inefficiency. [7].

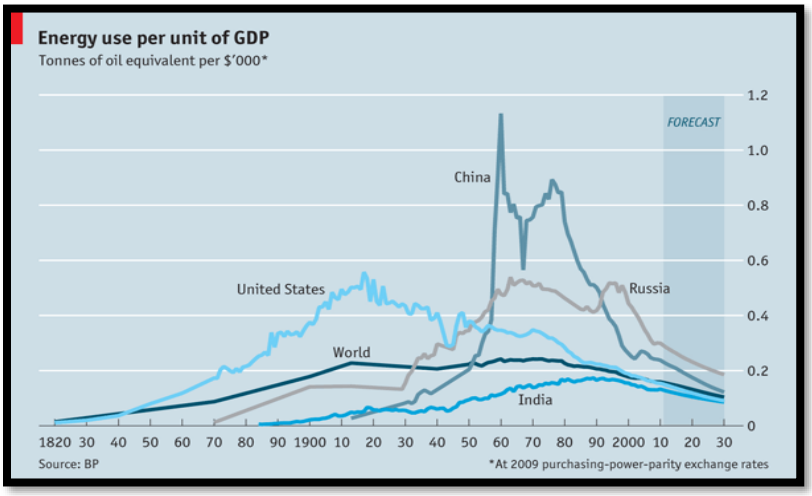


Figure 3: Energy use per unit of GDP trend. [7].

Reducing energy consumption while maintaining or increasing economic output (also known as "decoupling" energy consumption from economic growth [5]) is a recently popular concept. The Energy Transition in Germany could stand to be more energy-efficient.

The more intricate energy link addresses the social aspect of sustainable development by providing people with their fundamental needs. All of our fundamental requirements for sustenance, protection, comfort, health, education, and employment require energy. The Human Development Index (HDI) of a country is determined using variables such as per capita income, life expectancy, and educational attainment. It was created so that countries' human development progress could be compared and contrasted. In its Human Development Reports, the UNDP publishes each year the data used to calculate the HDI and equivalent indices.

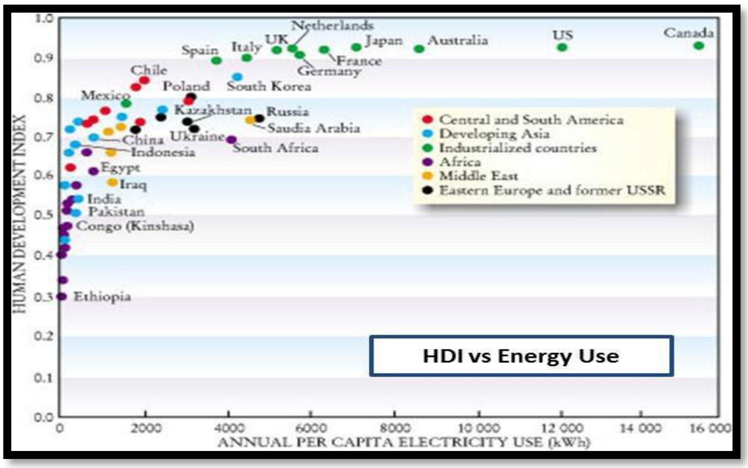


Figure 4: Energy use per country in kWh vs Human Development Index [7].

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

Numerous European studies and publications, including the 2014 Energy Supply and Demand study by the European Commission, have demonstrated the strong correlation between socioeconomic indicators and energy consumption. Comparing household energy consumption, developing nations consume substantially less energy than developed nations [8].

1. BACKGROUND

Despite its extensive application, the term "smart grid" remains undefined. Consider it a modification and upgrade of the existing electrical system to satisfy present and future needs.
The objective of "smart grids" is to improve the power grid in terms of efficiency, reliability, and safety while reducing costs and environmental impact using cutting-edge technologies such as distributed electricity producers, electricity storage, and electric vehicles, among others. Providing power generators and consumers with the means to encourage flexible output and proactive participation. Figure 2.2 illustrates that the effectiveness of the new electric grid will depend significantly on its customers.
Clearly distinguishable from the smart grid's centralized control centers that supervise electricity transmission and distribution are power plants and consumers. [19].

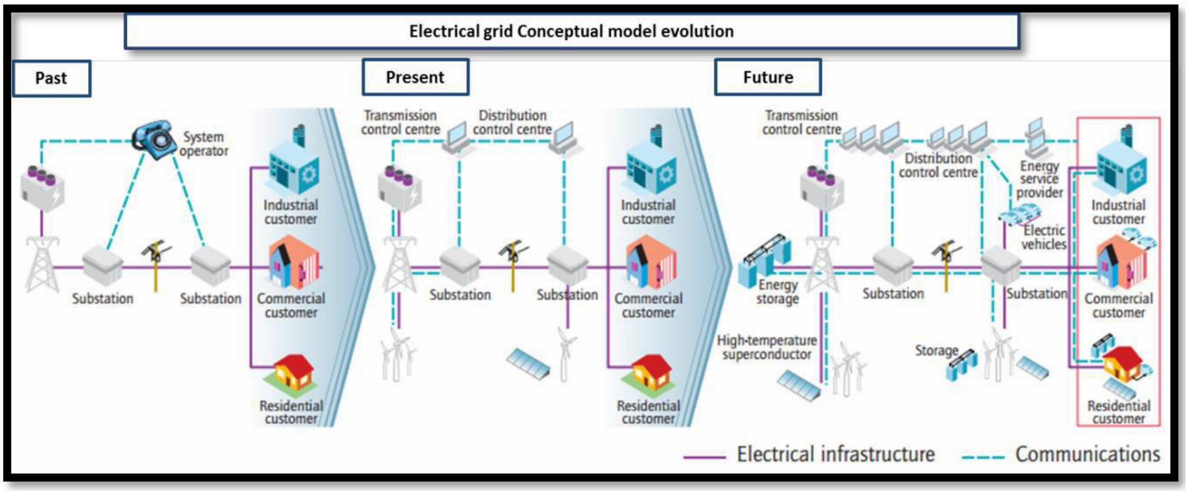


Figure 5: Smarter electricity systems, from past to future [20].

2.1. Smart Grid Characteristics, Functions, Benefits

i. Reliability, quality, and security

Today's grid is equipped with cutting-edge technologies for operating transmission systems, real-time monitoring of grid health, and self-assessments to locate, evaluate, and rectify faulty grid components.
By autonomously detecting grid interruptions and responding to unanticipated failures or attacks, the system enables the remainder of the affected zone to resume normal operations by isolating and preventing problem areas. Self-healing actions taken by the grid in response to fewer disruptions will enhance service quality and permit more effective management of the delivery infrastructure.

ii. Ability to accommodate all types of generation and storage options:

Smart grids facilitate the coordination of dispersed power plants of differing technological sophistication and size. The new infrastructure permits the connection of any "plug and play"

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

generating facility, such as rooftop solar photovoltaic arrays, combined heat and power plants, and medium-scale wind farms, to the grid.

The addition of energy storage (batteries, hydrogen, compressed air, pumped hydro, etc.) to the grid enhances its performance and reduces its operating expenses. To accommodate the growing number of electric vehicles, the grid could serve as a storage system.

iii. Consumer engagement:

Since consumers will be able to track their expenditures and consumption in real time, they will be in a strong position to influence future electric grid developments. In response to new power pricing models, such as time-of-use tariffs, which increase the price of energy during peak hours, consumers can assist the environment and save money on electricity by changing their behavior. In the future, the grid will be able to limit client consumption autonomously when electricity is expensive or scarce. Until then, turning off appliances may incur fees from utility providers.

iv. Optimize assets utilization and operating efficiency.

Distributors can pinpoint the source of power outages and allocate resources accordingly, saving money on preventative maintenance and emergency restorations, with increased network visibility.

"Smart grids" maximize the effectiveness of their power distribution networks by integrating cutting-edge technological advancements. The objective of the control devices is the efficient operation of the system; therefore, they will make adjustments to the operation to achieve this.

v. Enable new products, services, and markets.

Markets can influence energy, capacity, location, timing, and electrical quality, making them an integral component of grid management. As a result of deregulation's open-access (liberalization) market, consumers will have greater access to a variety of services.

2.2.**Smart Grid Conceptual Model**

The National Institute of Standards and Technology devised a conceptual model of a "smart grid" [22] to help readers comprehend what a "smart grid" is and how it may be subdivided into multiple sets. Figure 2.3 illustrates seven distinct categories, including "bulk production," "transmission," "distribution," "customer," "markets," "operators," and "service providers.".

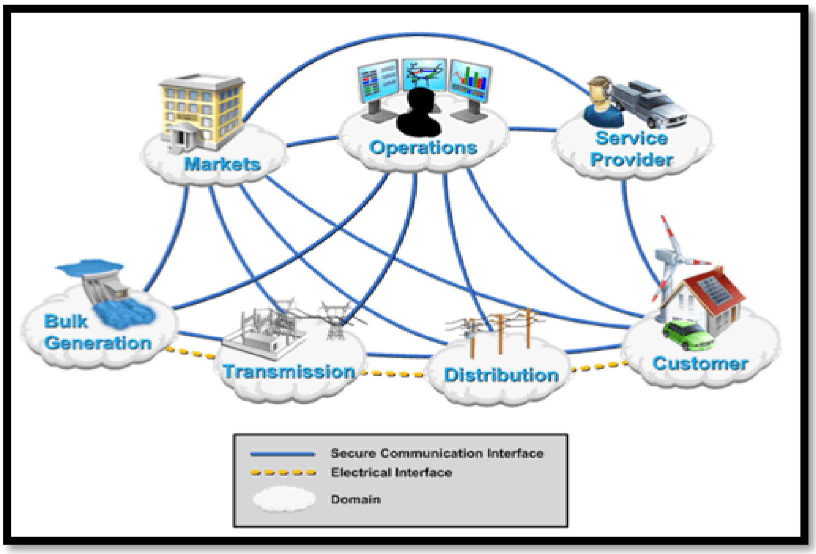


Figure 6: Conceptual illustration with all domains participating in a Smart Grid[21].

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

i. Bulk generation:

As depicted in Figure 2.4, a diverse array of energy sources and technological scales can be utilized to power large-scale, or "bulk," electrical infrastructure.

Coal, natural gas, and nuclear fusion power facilities are all conventional, nonrenewable, and unsustainable energy sources. Hydroelectric, biomass, geothermal, and pump storage are examples of dependable and durable energy sources. Solar and wind energy are examples of adaptable and renewable energy sources.

The bulk generating domain can begin exchanging data with the operations, markets, and transmission domains once power has been generated and transmitted. The ICT infrastructure enables the grid to autonomously reroute electricity flow when specific generators fail. Utilizing energy storage devices, starting and stopping power facilities, and distributing or storing energy are all viable options.

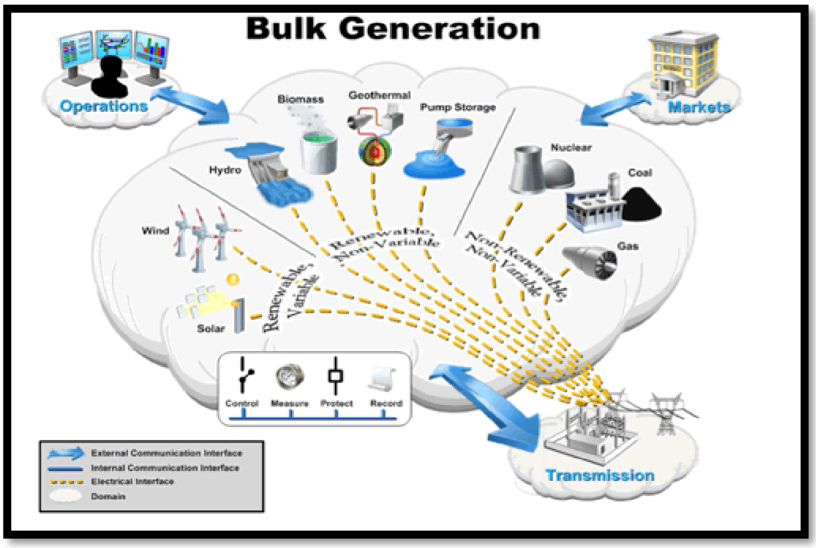


Figure 7: Deep look at the Bulk Generation domain[23].

ii. Transmission:

The network of substations that transports electricity from its source to its destination is known as transmission. It is wired to the regions of generation and distribution and communicates with the areas of operations and markets. To maintain grid reliability, transmission network operators must establish a balance between supply and demand. These sources, which may be distributed, may include storage and peaking generating units, which enable more accurate energy control and distribution.

i. Furthermore, required are physical technologies, such as those utilized to improve the transmission system.
ii. Using sensors to monitor a network segment's carrying capacity, the Dynamic Line Rating (DLR) maximizes transmission assets while introducing an overflow risk.
iii. Using flexible AC transmission systems increases the power transmission capacity and enhances transmission network control.
iv. With high voltage DC, system losses can be decreased by linking offshore wind or solar farms to load centers.
v. By decreasing transmission losses, high-temperature superconductors improve transmission efficiency.

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

vi. Distribution:

The Transmission domain communicates electronically with the Consumer domain, whereas the Distribution domain communicates with the Markets and Operations domains. Distribution grids have changed dramatically over time, transitioning from hierarchical, unidirectional interfaces with limited communication to bidirectional grids with numerous monitoring and control devices that can disperse distributed generation and storage, regulate demand response and the load, and improve reliability.

Distribution grid management enhances the intelligence of the grid at distribution and substations by automating distribution procedures and utilizing real-time data for problem detection and rapid restoration.

vii. Customer:

To be fully functional, the future smart grid will require the participation of all of its consumers. Due to the grid, everyone who wishes to contribute to the writing of this essay has access to the necessary materials. display Figure 2.5 to illustrate the relationship between the Distribution and Consumer domains, as well as the Operations, Markets, and Service Providers domains.

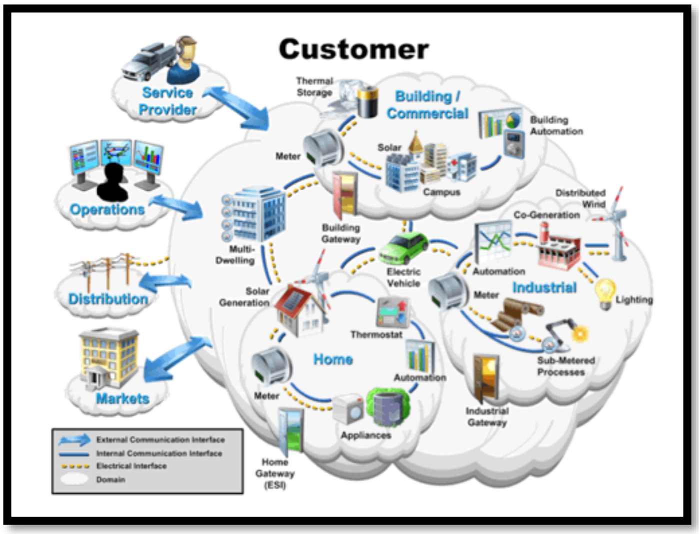


Figure 8: Deep look at the Customer domain [21]**.**

2. METHODOLOGY

The steps required to complete this study and derive the final customer segmentation for estimating power consumption based on utility categorization and clustering are described below. Obtaining data on the user's hourly power consumption, features of the residence, and occupants, adjusting the data as required for the study, and understanding how the data is displayed are the initial steps prior to analysis. The term "data cleansing" refers to the process of removing inaccurate or superfluous data from a dataset. To complete this step, a visual aid is required.

You may be able to obtain a deeper understanding of the study's findings and verify their veracity through additional reading and data visualization.

Three distinct phases comprise the process: pre-clustering, clustering itself, and post-clustering analysis. Using cluster analysis, the second phase classifies customers into more manageable subsets.

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

Before making conclusions about the electrical load profiles of users, it is essential to evaluate pertinent studies, select an appropriate input data format, and analyze the input data to determine the proportionate energy consumption per hour for each user.

During the classification phase, the prevalent data clustering techniques for energy use data are evaluated. Weka is used to perform computational clustering using the appropriate hierarchical, naive bayes, and support vector machine algorithms to repeatedly determine the optimal number of clusters and the optimal number of cluster members. Methods for clustering are evaluated based on their aesthetic and statistical outcomes, respectively.

After classifying users, they are redistributed to construct the corresponding consumer subsets. Using visualization and statistical methods, the analyst identifies outliers and manually assigns them to the most suitable group.

The third step is investigating the family and its members. Histograms are used to identify similarities between distinct consumer categories.

3.1.OBJECTIVE OF CLEANING DATA

The first phase of any data analysis is to gather the required data. The data used must also be reliable. It is essential to establish the purpose of the analysis beforehand, as this will determine the data analysis strategy. Using a data preparation service also results in more accurate conclusions. The required steps are depicted in Figure 3.1:

Figure 9: States of the data when performing a data analysis [25].

When the intent of the data is clear or when we have specific objectives in mind, we can begin to examine the unedited presentation. Although it may be difficult to view and, therefore, comprehend raw data, it is highly unlikely that we will be able to obtain the data in the exact format required to launch the investigation.

This is referred to as "processed" or "tidy" data, and it must be prepared in this manner prior to analysis. With the aid of computational data processing, processing programs are created and executed for this purpose. These scripts contain the fundamental code for data configuration.

After the data have been processed, data cleansing is carried out prior to any analysis. The term "data cleansing" encompasses both preventative and corrective measures taken to rid a dataset of errors such as duplicates, absent data, NAs, and others.

Following are descriptions of the numerous types of data that will be incorporated into this study:

- ✓ Electricity consumption data
- ✓ Households and householder’s information

3.2.ELECRICITY CONSUMOTION DATA

The ensuing investigation heavily relied on data regarding the families' electricity consumption. Figure 3.2 is a visual depiction of the steps necessary to acquire the pertinent data on electrical usage for the investigation:

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

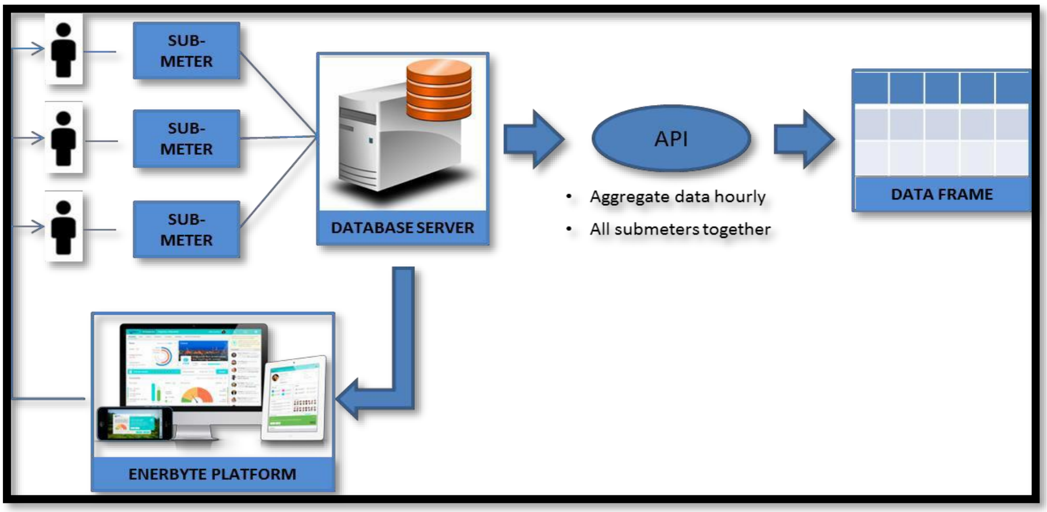


Figure 10: Diagram to explain how the data used for the analysis is obtained [27].

Submetering devices are used to monitor the quantity of energy consumption in residences and smaller businesses. These devices transfer data via a link, often Wi-Fi in the home, to the manufacturer's database server, where it is kept.

3.3 DATA EXPLORATION AND VISULIZATION

Before beginning consumer segmentation, it is essential to know the data that will be analyzed in order to have a feel for the information and identify any abnormalities or difficult areas that must be addressed. mastering both the raw data and the "Weka software" instructions necessary for its analysis.

Remember these three things when conducting data analysis:

Let's begin with the dataset format, which in the majority of instances must be modified to provide the appropriate visual representation.

The way dates are handled provides a barrier when doing analysis, as time series data necessitate creative methodologies. Especially when distinguishing between weekdays and weekends or between different months.

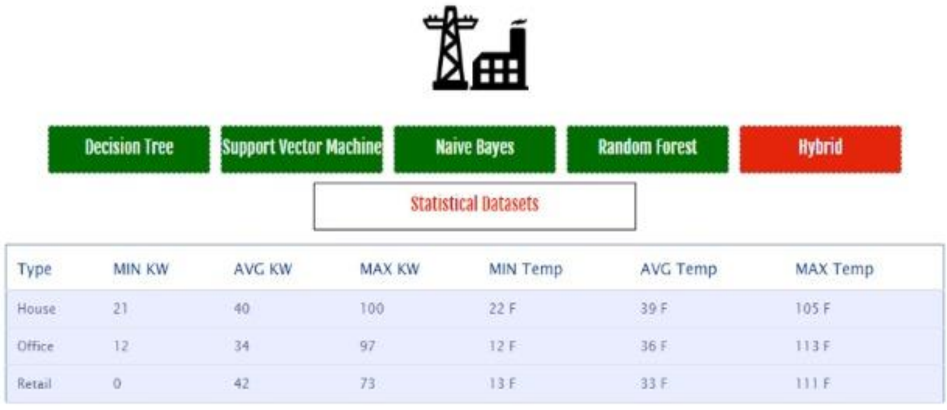
Finally, pay great attention to the specifics of the statistics on electricity use and investigate the most effective techniques of obtaining the necessary information or conclusions.

Presently, data is collected hourly. For a more in-depth analysis to establish the variables that created the profile shape and peak positions, as well as the characteristics of the home, data with a resolution of 5 or 15 minutes would be necessary.

3. RESULTS

In numerous classification schemes, the distance between two classes is defined as the smallest distance between their geometric and statistical centers.

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

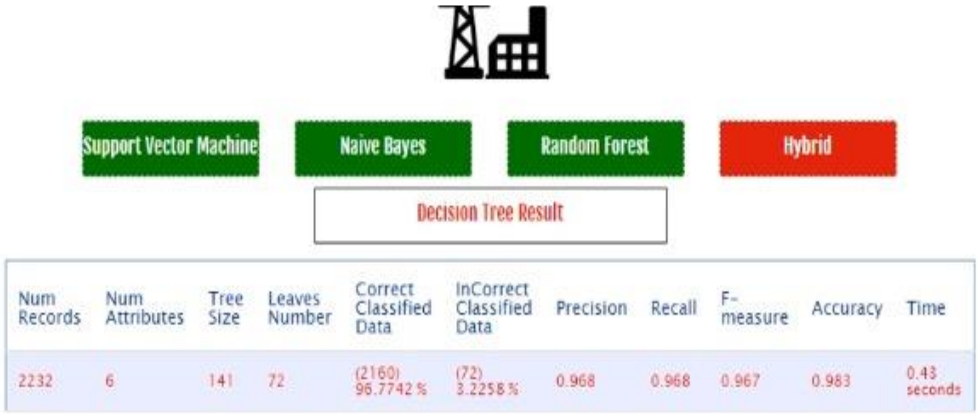


Type	MIN KW	AVG KW	MAX KW	MIN Temp	AVG Temp	MAX Temp
House	21	40	100	22 F	39 F	105 F
Office	12	34	97	12 F	36 F	113 F
Retail	0	42	73	13 F	33 F	111 F

Figure 11: The overall classification results based on approaches used on statistical dataset.

4.3.1. Decision Tree Prediction

When working with dimensionless data with time facto, such as the number of records, precision, recall, F-measure, and accuracy, the analyst must select the most appropriate input data units for the output results. Depending on the desired conclusion, classification analyses employing decision tree classification and load profile energy consumption segmentation generate diverse sets of output data.



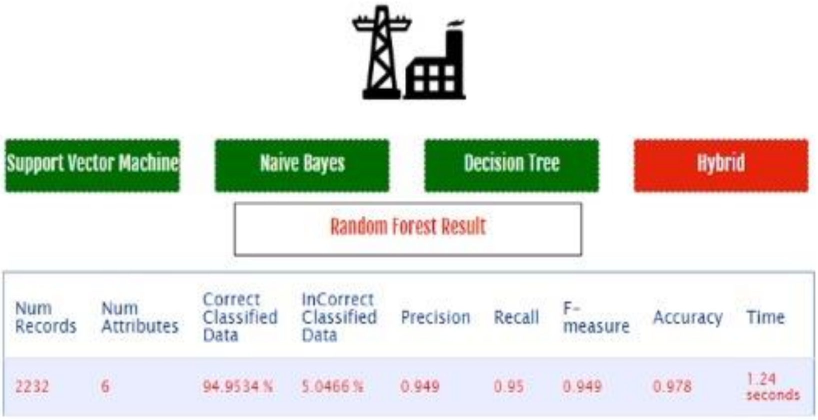
Num Records	Num Attributes	Tree Size	Leaves Number	Correct Classified Data	InCorrect Classified Data	Precision	Recall	F-measure	Accuracy	Time
2232	6	141	72	(2160) 96.7742 %	(72) 3.2258 %	0.968	0.968	0.967	0.983	0.43 seconds

Figure 12: The decision tree classification results based on approaches used.

4.3.2. Random Forest Prediction

Different load profiles have different energy consumption segmentation and classification goals when using random forest classification; using absolute values in result, such as the number of records, precision, recall, F-measure, accuracy, and time for the use of dimensionless data with time trends.

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

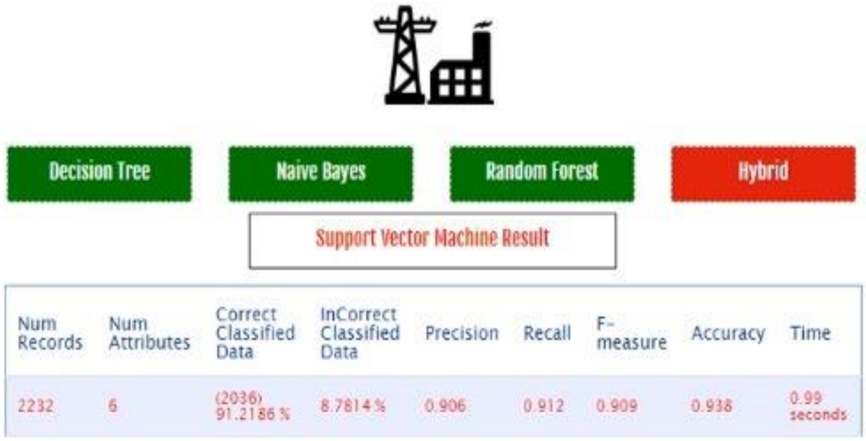


Num Records	Num Attributes	Correct Classified Data	InCorrect Classified Data	Precision	Recall	F-measure	Accuracy	Time
2232	6	94.9534 %	5.0466 %	0.949	0.95	0.949	0.978	1.24 seconds

Figure 13: The random forest classification results based on approaches used.

4.3.3. Support Vector Machine Prediction

Depending on the specific analysis goal, different output findings are used to perform the support vector machine classification analysis of the load profile's energy consumption.



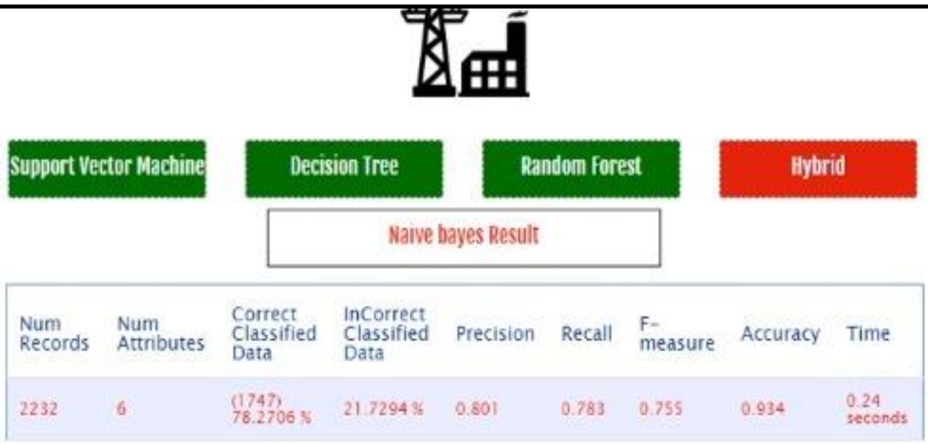
Num Records	Num Attributes	Correct Classified Data	InCorrect Classified Data	Precision	Recall	F-measure	Accuracy	Time
2232	6	(2036) 91.2186 %	8.7814 %	0.906	0.912	0.909	0.938	0.99 seconds

Figure 14: The support vector machine classification results based on approaches used.

4.3.4. Naïve Bayes Prediction

The load profile's energy consumption segmentation and classification using nave bayes classification can serve a variety of reasons; thus, the analyst must select which output data are most valuable for performing the classification study. Absolute numbers are supplied for counts, measures of precision and recall, the F-measure, the accuracy of results, and the time spent on a job for the use of dimensionless data with a time factor for generation and execution being absolute 0.24 seconds and recorded precision of 0.801..

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

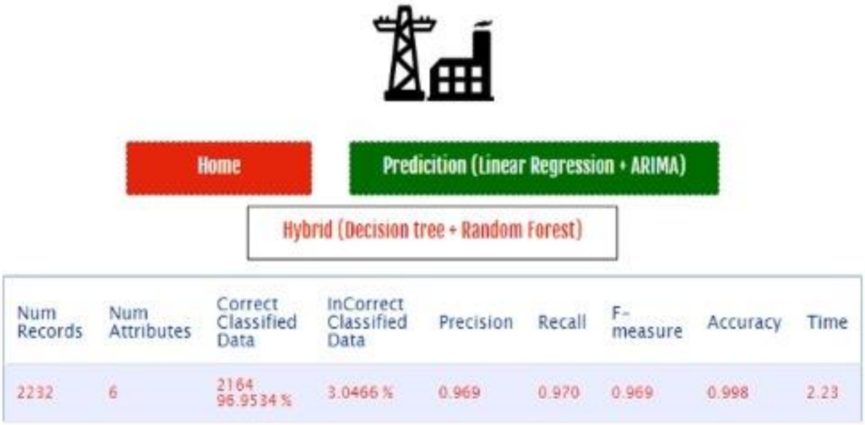


Num Records	Num Attributes	Correct Classified Data	InCorrect Classified Data	Precision	Recall	F-measure	Accuracy	Time
2232	6	(1747) 78.2706 %	21.7294 %	0.801	0.783	0.755	0.934	0.24 seconds

Figure 15: The naïve bayes classification results based on approaches used.

4.3.5. Hybrid Model Prediction

The output results used to carry out the classification analysis vary depending on the specific aim of the load profile's energy consumption segmentation plus classification using hybrid classification which knowingly involves both decision tree plus random forest, and it is up to the analyst to decide which are the most convenient input data units for output results; the use of absolute values in result including number of records, precision, recall, F-measure, accuracy and time for the use dimensionless data with time factor for generating and execution is absolute 2.23 seconds with recorded precision of 0.969.



Num Records	Num Attributes	Correct Classified Data	InCorrect Classified Data	Precision	Recall	F-measure	Accuracy	Time
2232	6	2164 96.9534 %	3.0466 %	0.969	0.970	0.969	0.998	2.23

Figure 16: The hybrid model classification results based on approaches used.

4.3.6. Prediction Result

By calculating the AVERAGE KW for the mean square error of each load profile with respect to its class mean load profile, the classification can be validated, and the load profiles that deviate the most from the mean, and thus may need to be reallocated to another prediction class between a range of months based on type, can be identified.

International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 7, July, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

Predicition Result

Type	Prediction Class	Range of Month	Range of Hours	AVG Temp	AVG KW	MSE (Mean Square Error)
House Retail Office	1 (Low)	3-6 and 9-11	4-9 and 18-21	60 F or less than 10F	less than 10per day	0.02
House Retail Office	2 (med)	7 and 12	9-12 and 12-1	40 F	less than 30per day	0.04
House Retail Office	3 (high)	8,1 and 2	4-9 and 18-21	less than 35 F or more than 100f	more than 40per day	0.03

Figure 17: The prediction of result for classifying the consumption of electricity in terms of mean square error.

4. CONCLUSION

Since home energy consumption is relatively low, energy management systems are not as prevalent in the residential sector as they are in the commercial and industrial sectors. We were able to construct a framework and an intelligent system to predict future electrical energy usage based on these findings. Homeowners may still have access to these services provided the necessary infrastructure is in place, such as when government agencies and electrical companies collaborate to build a community.

This initiative demonstrates the government's early efforts to involve customers in the effort to improve energy efficiency. It aims to educate and direct its users so that they can cut their average energy use by 10%. However, it was discovered that supplying technical consumption data in kWh to households has low impact and influence [38], thus further work is required to translate the technical knowledge into actionable steps that would assist the user in making good energy decisions (like data mining approaches). Scalability is further limited by the necessity for expensive sub-metering devices to collect consumption data. This type of endeavor cannot be expanded without access to data from smart meters installed by utilities.

Thus, the energy business stands to gain significantly from the use of smart meters and big data analytics. Methods of data mining are utilized to increase utility and customer satisfaction by generating large amounts of raw data that must be managed prior to being translated into valuable information. producing data-driven commercial opportunities to suit market needs.

For trustworthy research and policymaking, accurate data on electrical energy usage is essential. Quantitatively and qualitatively, the data used in this investigation were the best available. A substantially larger dataset with thousands of users would have better fulfilled the objectives of the study than the modest sample size that was employed.

Due to difficulties with the electrical consumption data collected by the sub-metering equipment, data cleaning and purification additionally required a pre-clustering stage. Since the sub-metering device and server communicate using the machine learning algorithms we found, consumption data is omitted and consumption is counted as if it had not occurred.

When the possibility of access to hourly electrical energy consumption data becomes a reality, it will provide a suitable framework for the current study's objectives because the number of users could be much higher and the data quality should be nearly perfect, just like the actions taken by the electricity utility to bill their customers using machine learning-based algorithms.

While the primary objective of dividing electrical energy users into groups with comparable load profiles was achieved, the secondary objective of studying the features of these groups added nothing to the project. The common framework utilized to compare and compare fairly the various load profiles, as demonstrated by computing the percentage of energy consumption per hour, served its intended function of classifying users by their load profiles.

We employed an iterative technique based on computer classification computations to arrive at the final clusters (using the Weka program). An outlier was identified, and the analyst manually reassigned it to the appropriate group using visualization and statistical techniques. No matter whatever categorization method was utilized, the outcomes were uniformly comparable (Decision tree, Random forest, Naive Bayes, Support vector machine, or the hybrid model, which combines Random forest and Decision tree). In addition, classes are utilized to aid the depiction of categorization and division possibilities.

REFERENCES

[1] Ardakanian, O. et al., 2014. Workshop Proceedings of the EDBT/ICDT 2014 Joint Conference on CEUR-WS.org.

[2] Armaroli, N. & Balzani, V., 2011. Towards an electricity-powered world. Energy Environmental Science, pp. 4, 3193-3222.

[3] Beckel, C., Sadamori, L. & Santini, S., 2012. Towards automatic classification of private households using electricity consumption data. Embedded Sensing Systems for Energy-Efficiency in Buildings: Proceedings of the Fourth ACM Workshop, (BuildSys '12), pp. pp.169-176.

[4] BloomEnergy, 2015. Fuel Cell: Distributed Generation. [Online] Available at: http://www.bloomenergy.com/fuel-cell/distributed-generation/

[5] A. Ben-Hur and J. Weston, "A User's Guide to Support Vector Machines," in Methods in molecular biology, 2010, pp. 223-239

[6] Chicco , G. & Ilie, I., 2009. Support vector clustering of electrical load pattern data. IEEE Trans. Power Syst, 24(3), pp. 1619-28.

[7] E. Goel and E. Abhilasha, "Random Forest: A Review," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 7, no. 1, pp. 251-257, 2007.

[8] Chicco , G. et al., 2005. Emergent electricity customer classification. IEE Proc Gener Transm Distrib, 152(2), pp. 164-72.

[9] Chicco , G. et al., 2004. Load pattern-based classification of electricity customers. IEEE Trans. Power Syst, 19(2), pp. 1232-9.

[10] J. Kelly and W. Knottenbelt, "Neural NILM: Deep Neural Networks," in Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, 2015.

[11] K. C. ARMEL, A. GUPTA, G. SHRIMALI and A. ALBERT, "Is disaggregation the holy grail of energy efficiency? The case of electricity," Energy Policy, vol. 52, p. 213–234, 2013.