

ANALYSIS OF ELECTRICITY CONSUMPTION AT HOME USING K-MEANS CLUSTERING ALGORITHM

Submitted in partial fulfillment of the
requirements for the award of
Bachelor of Engineering degree in Computer Science and

Engineering by

CH. SRI REVATH(36110238)

CH. MADHAN MOHAN REDDY(36110266)



**DEPARTMENT OF COMPUTER
SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119

MARCH - 2020



SATHYABAMA

INSTITUTE OF SCIENCE AND
TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600119



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **CH. SRI REVATH(36110238), CH.MADHAN MOHAN REDDY (36110266)** who have done the Project work as a team who carried out the project entitled **"ANALYZING OF ELECTRICITY CONSUMPTION AT HOME USING K-MEANS CLUSTERING** under my supervision from **NOVEMBER 2019 to APRIL 2020.**

Internal Guide

Mrs .E. BRUMANCIA.,M.E.,(Ph.D).

Head of the Department

Dr. S. ViGNESHWARI, M.E., Ph.D.

Submitted for Viva voce Examination held on

Internal Examiner

External Examiner

DECLARATION

I **CH. SRI REVATH (36110238)** hereby, declare that the Project Report entitled **ANALYSIS OF ELECTRICITY CONSUMPTION AT HOME USING K-MEANS CLUSTERING ALGORITHM** done by me under the guidance of **E. BRUMANCIA., M.E.,(PH.D)** is submitted in partial fulfilment of the SCIENCE AND ENGINEERING

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY MANAGEMENT** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.SASIKALA M.E., Ph.D., Dean**, School of Computing , **Dr. S. VIGNESHWARI M.E., Ph.D.** and **Dr. L. LAKSHMANAN M.E., Ph.D.** , Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs.E. BRUMANCIA.,M.E.,(Ph.D)** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non - teaching staff members of the Department of **COMPUTER SCIENCE AND ENGINEERING** who were helpful in many ways for the completion of the project

ABSTRACT

The energy consumption of residential and commercial buildings has risen steadily in recent years, an increase largely due to their HVAC systems. Expected energy loads, transportation, and storage as well as user behavior influence the quantity and quality of the energy consumed daily in buildings. However, technology is now available that can accurately monitor, collect, and store the huge amount of data involved in this process. Furthermore, this technology is capable of analyzing and exploiting such data in meaningful ways. Not surprisingly, the use of data science techniques to increase energy efficiency is currently attracting a great deal of attention and interest. This paper reviews how Data Science has been applied to address the most difficult problems faced by practitioners in the field of Energy Management, especially in the building sector. The work also discusses the challenges and opportunities that will arise with the advent of fully connected devices and new computational technologies.

TABLE OF CONTENTS

CHAPTER No	TITLE NAME	PAGE No
	ABSTRACT	iv
	LIST OF FIGURES	ix
	LIST OF TABLES	x
	LIST OF ABBREVIATIONS	xi
1	INTRODUCTION	01
	1.1 Introduction	01
	1.2 Application of java	02
	1.3 Features of java	03
	1.4 Collection of framework	04
	1.5 Mysql	05
	1.6 Feasibility study	06
	1.7 Economical feasibility	06
	1.8 Technical feasibility	07
	1.9 Social feasibility	07
2	LITERATURE REVIEW	08
	2.1 Literature review	08
3	METHODS AND MATERIALS	12
	3.1 Existing system	12
	3.1.1 Advantages of existing study	12
	3.2.1 Advantages of proposed of algorithm	12
	3.3 proposed algorithm	12

3.3.1	Proposed of k-mean algorithm	12
3.3.2	Details of k-mean	13
3.4	Naïve bayes algorithm	14
3.4.1	How naïve bayes works	14
3.5	Application of naïve bayes	16
3.6	System architecture	17
3.7	System requirments	17
3.7.1	Hardware requirments	17
3.7.2	Software requirments	17
3.8	Requirment analysis	18
3.9	Functional requirments	18
3.9.1	Usability	18
3.9.2	Robustness	18
3.9.3	Security	18
3.9.4	Reliability	19
3.9.5	Compatibility	19
3.9.6	Flexibility	19
3.9.7	Safety	19
3.10	Non functional requirments	19
3.10.1	Portability	19
3.10.2	Performance	19
3.10.3	Accuracy	19
3.10.4	Maintability	19
3.11	Modules	20
3.11.1	Energy dataanalytics	20
3.11.2	Energy consumption data prediction	20

3.11.3	Energy data processing	21
3.12	UML Diagrams	21
3.12.1	Use case diagram	22
3.12.2	Sequence diagrams	23
3.12.3	Class diagrams	23
3.12.4	Activity diagram	24
3.13	System design and testing	24
3.13.1	Input design	24
3.13.2	Output design	25
3.14	System design	25
3.14.1	Testing process	25
3.15	Types of testing	25
3.15.1	Unit testing	25
3.15.2	Integration testing	26
3.15.3	Functional testing	26
3.15.4	System testing	27
3.15.5	White box testing	27
3.15.6	Black box testing	27
3.16	Test strategy and approach	28
3.16.1	Test objectives	28
3.16.2	Integration testing	28
3.16.3	Acceptance testing	29
3.16.4	Alpha testing	29
3.16.5	Beta testing	29

4	RESULT AND DISCUSSION	30
----------	------------------------------	-----------

5	SUMMARY AND CONCLUSION	33
6	REFERENCES	34
7	APPENDIX	36
	A. Sample code	36
	B. Screen shots	41
	C. Plagarism report	43

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
1	System architecture	17
2	Use case diagram	22
3	Sequence diagram	23
4	Class diagram	23
5	Activity diagram	24
6	Unit testing	26
7	White box testing	27
8	Black box testing	27
9	Comparison of electricity Consumption between the Four season	31
10	Visualization of electricity consumption in each season	32

LIST OF TABLES

TABLE NO	TITLE	PAGE NO
3.4.1	Frequency table	14
3.4.1	Likelihood table	14
4	The data simple after dimension reduction	32

LIST OF ABBREVIATIONS

MTBF - Mean Time Between Failure

RFE - Request For Evidence

GA - Genetic Algorithm

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Energy has become essential for human existence and development. The imbalance between dwindling energy supply and increasing energy demand has intensified gradually. As a result, energy conservation has become a topic that attracts global attention. Energy consumption in buildings has an enormous influence on sustainability. According to the statistics from literature, buildings account for 32% of the total amount of energy consumed worldwide. In the US and other developed nations, this figure is 41%, making the construction industry the biggest energy consumer. The ratio of the buildings energy consumption approximates to one-third in some developing nations. With accelerating urbanization and higher living standards of the people in our country, the ratio of the building energy consumption will continue to rise, and may top the list of energy consumers. The Building Energy Consumption Monitoring Platform and Conservation-Minded Campus Energy Conservation Supervision Platform have sprung up in 2008, thanks to the substantial support from the Chinese government. According to China's report on energy saving of buildings in rural and urban areas in 2013, real-time energy consumption monitoring was performed on 5,000-plus large public buildings by the end of 2013. The energy-saving campus construction projects were piloted in 210 colleges as shown in Fig. 1. The development of the system for monitoring energy consumption in buildings can provide large-scale data information helpful in improving energy management and reducing energy consumption. It is also the most intuitive source of data concerning the operational status of buildings. BECMP collects the temperature, humidity, pressure, electrical signals and control signals of building systems. These are critical aspects of building system operation and management, performance diagnosis, fault analysis and energy efficiency evaluation. The energy consumption data of buildings is affected by the building structure, climate environmental parameters, construction, operating modes, and user behavior patterns. Traditional methods of analysis are not sufficient in handling large-scale data which are generated during the operation of the energy consumption monitoring systems. In addition, management personnel may not be able to find and process abnormal data that may occur. This results in a serious

discrepancy between the energy consumption data “on paper” and the actual energy consumption. The main reason for this is that the traditional methods adopt the energy consumption simulation scheme. This requires the users to have a good command of the building and system architecture. Furthermore, it is limited, inaccurate and cannot process useful information quickly for practical applications. As an emerging technology, data mining has immense strength in big data processing, and it has been utilized in machine learning, human intelligence, pattern recognition, statistics, informatics, and data visualization. It has drawn growing attention from various industries and has become a bigger player in management decision-making. The data mining technology focuses on data rather than computing with complicated formulas, thereby it is easier to integrate in many applications. It is an emerging powerful technology with great potential to discover hidden and useful knowledge in large data sets and while it has used in the construction field of the existing "rich data and poor information" dilemma , its application in the field of building energy systems, particularly in the energy consumption data, has been minimal.

Java is one of the world’s most important and widely used computer languages, and it has held this distinction for many years. Unlike some other computer languages whose influence has weared with passage of time, while Java's has grown.

1.2 APPLICATION OF JAVA

Java is widely used in every corner of world and of human life. Java is not only used in softwares but is also widely used in designing hardware controlling software components. There are more than 930 million JRE downloads each year and 3 billion mobile phones run java.

Following are some other usage of Java:

1. Developing Desktop Applications
2. Web Applications like LinkedIn.com, Snapdeal.com etc
3. Mobile Operating System like Android
4. Embedded Systems
5. Robotics and games etc.

1.3 FEATURES OF JAVA

The prime reason behind creation of Java was to bring portability and security feature into a computer language. Beside these two major features, there were many other features that played an important role in moulding out the final form of this outstanding language. Those features are;

1) Simple

Java is easy to learn and its syntax is quite simple, clean and easy to understand. The confusing and ambiguous concepts of C++ are either left out in Java or they have been re-implemented in a cleaner way.

Eg: Pointers and Operator Overloading are not there in java but were an important C++.

2) Object Oriented

In java everything is Object which has some data and behaviour. Java can be easily extended as it is based on Object Model.

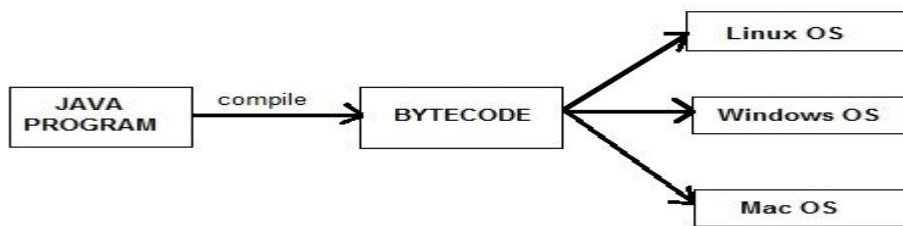
3) Robust

Java makes an effort to eliminate error prone codes by emphasizing mainly on compile time error checking and runtime checking. But the main areas which Java improved were Memory Management and mishandled Exceptions by introducing automatic Garbage Collector and Exception Handling.

4) Platform Independent

Unlike other programming languages such as C, C++ etc. which are compiled into platform specific machines. Java is guaranteed to be write-once, run-anywhere language.

On compilation Java program is compiled into byte code. This byte code is platform independent and can be run on any machine, plus this byte code format also provide security. Any machine with Java Runtime Environment can run Java Programs.



5) Secure

When it comes to security, Java is always the first choice. With java secure features it enable us to develop virus free, temper free system. Java program always runs in Java runtime environment with almost null interaction with system OS, hence it is more secure.

6) Multi-Threading

Java multithreading feature makes it possible to write program that can do many tasks simultaneously. Benefit of multithreading is that it utilizes same memory and other resources to execute multiple threads at the same time, like While typing, grammatical errors are checked along.

7) Architectural Neutral

Compiler generates byte codes, which have nothing to do with a particular computer architecture, hence a Java program is easy to interpret on any machine.

8) Portable

Java Byte code can be carried to any platform. No implementation dependent features. Everything related to storage is predefined, example: size of primitive data types.

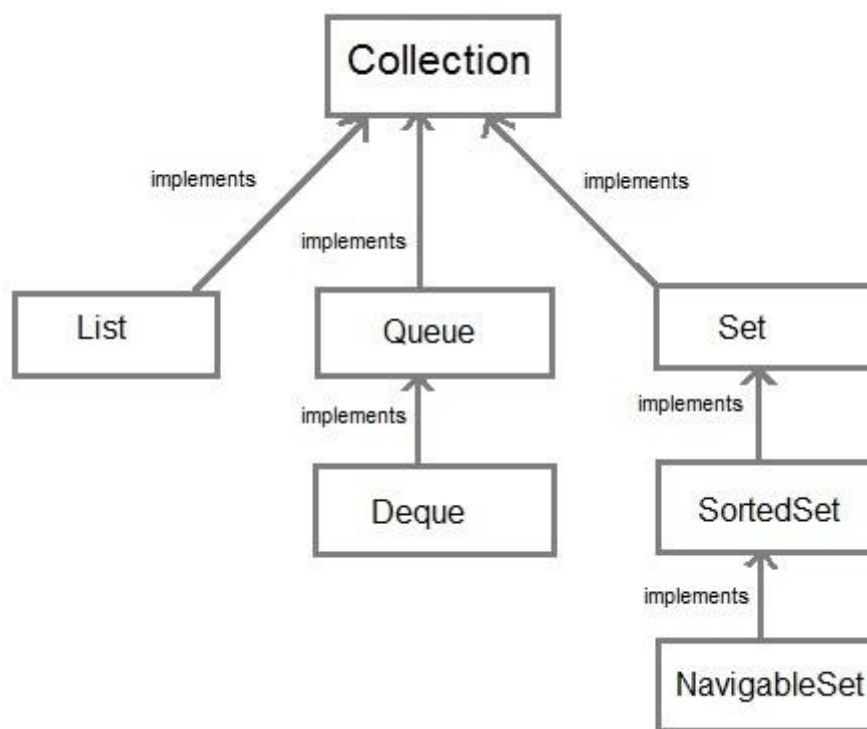
10) High Performance

Java is an interpreted language, so it will never be as fast as a compiled language like C or C++. But, Java enables high performance with the use of just-in-time compiler.

1.4 COLLECTION FRAMEWORK

Collection framework was not part of original Java release. Collections was added to J2SE 1.2. Prior to Java 2, Java provided adhoc classes such as Dictionary, Vector,

Stack and Properties to store and manipulate groups of objects. Collection framework provides many important classes and interfaces to collect and organize group of alike objects.



1.5 MYSQL

MySQL, officially, but also called "My Sequel" is the world's most widely used open-source relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases, though SQLite probably has more total embedded deployments. The SQL phrase stands for Structured Query Language.

The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation. MySQL is a popular choice of database for use in web applications, and is a central component of the widely used LAMP open source web application software stack (and other 'AMP' stacks).

LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python." Freesoftware-open source projects that require a full-featured database management system often use MySQL.

For commercial use, several paid editions are available, and offer additional functionality. Applications which use MySQL databases include: TYPO3, MODx, Joomla, WordPress, phpBB, MyBB, Drupal and other software. MySQL is also used in many high-profile, large-scale websites, including Wikipedia, Google (though not for searches), Facebook, Twitter, Flickr and YouTube.

1.6 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

1.7 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

1.8 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

1.9 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

CHAPTER 2

REVIEW OF LITERATURE

2.1 LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

Building enovation and modernisation in Europe: State-of-the-art review (R).Netherlands: OTB Research Institute for Housing

The final purpose of the project “State of the art review of building renovation and modernisation in Europe”, launched by Erabuild, the European Research Area for the sustainable construction and operation of buildings, is to identify the most promising fields for future activities within the trans-national Erabuild research programme on Sustainable Renovation. To meet this final purpose, four work packages have been defined and are treated successively in the present report. 1. Mapping of building typologies and stakeholder interests (WP1). 2. Analysis of existing incentives and their impact on the renovation rate (WP2). 3. Mapping of modernisation and renovation

research (WP3). 4. Recommendations about the most promising fields for future activities (WP4). The countries covered in this study are Austria, Finland, France, Germany, the Netherlands, Sweden, Switzerland and the United Kingdom. Residential and nonresidential buildings are covered.

Action plan for energy efficiency: Realising the potential (R)

Energy is essential for Europe to function. But the days of cheap energy for Europe seem to be over. The challenges of climate change, increasing import dependence and higher energy prices are faced by all EU members. Moreover the interdependence of EU Member States in energy, as in many other areas, is increasing – a power failure in one country has immediate effects in others. Europe needs to act now, together, to deliver sustainable, secure and competitive energy. In doing so the EU would return to its roots. In 1952 with the Coal and Steel Treaty and 1957 with the Euratom Treaty, the founding Member States saw the need for a common approach to energy. Energy markets and geopolitical considerations have changed significantly since then. But the need for EU action is stronger than ever. Without this, the EU's objectives in other areas, including the Lisbon Strategy for growth and jobs and the Millennium Development Goals, will also be more difficult to achieve. A new European Energy Policy needs to be ambitious, competitive and long-term – and to the benefit of all Europeans.

Development of an energy monitoring system for large public buildings (J)

The preference of public building owners to the energy saving incentive Instruments of building determines the influence of policy tools, which is formulated and implemented by the government. In this paper, firstly we define the types of building energy-saving incentive policies, and identify the influence factors of public building owners to incentive policy tools. Secondly, we analyze the characteristics of the factors affecting the policy preferences of the public buildings, the preference of the public building owners to those policy tools and the relationship between them. Finally, in order to improve the energy efficiency of public buildings, the paper puts forward some policy suggestions for the government.

Data Mining: Concepts and Techniques (M)

Analyzing large amounts of data is a necessity. Even popular science books, like “super crunchers,” give compelling cases where large amounts of data yield discoveries and intuitions that surprise even experts. Every enterprise benefits from collecting and analyzing its data: Hospitals can spot trends and anomalies in their patient records, search engines can do better ranking and ad placement, and environmental and public health agencies can spot patterns and abnormalities in their data. The list continues, with cybersecurity and computer network intrusion detection; monitoring of the energy consumption of household appliances; pattern analysis in bioinformatics and pharmaceutical data; financial and business intelligence data; spotting trends in blogs, Twitter, and many more. Storage is inexpensive and getting even less so, as are data sensors. Thus, collecting and storing data is easier than ever before. The problem then becomes how to analyze the data. This is exactly the focus of this Third Edition of the book. Jiawei, Micheline, and Jian give encyclopedic coverage of all the related methods, from the classic topics of clustering and classification, to database methods (e.g., association rules, data cubes) to more recent and advanced topics (e.g., SVD/PCA, wavelets, support vector machines). The exposition is extremely accessible to beginners and advanced readers alike. The book gives the fundamental material first and the more advanced material in follow-up chapters. It also has numerous rhetorical questions, which I found extremely helpful for maintaining focus. We have used the first two editions as textbooks in data mining courses at Carnegie Mellon and plan to continue to do so with this Third Edition. The new version has significant additions: Notably, it has more than 100 citations to works from 2006 onward, focusing on more recent material such as graphs and social networks, sensor networks, and outlier detection.

Data processing method for public building energy consumption monitoring systems based on data mining

Smart streetlamp as an important part of smart cities aiming at constructing cities with a higher level of intelligence, automation, information and network has been applied in urban construction widely and successfully. This paper presents a data processing method for smart streetlamp energy consumption analysis systems based on data mining, which is not only used to classify, recognize and correct problem data in

energy consumption data, but also used in smart streetlamp energy consumption analysis systems for abnormal alarm, troubleshooting advice, energy consumption prediction, etc. It is necessary to process energy consumption data before energy consumption analysis, because the problem data might interfere the analysis and cause deviations and errors. Moreover, the method proposed is used to realize the functions of processing historical data and monitoring in this paper. Finally, the experimental result proves the availability and reliability of the method.

Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques

This paper presents a data mining (DM) based approach to developing ensemble models for predicting next-day energy consumption and peak power demand, with the aim of improving the prediction accuracy. This approach mainly consists of three steps. Firstly, outlier detection, which merges feature extraction, clustering analysis, and the generalized extreme studentized deviate (GESD), is performed to remove the abnormal daily energy consumption profiles. Secondly, the recursive feature elimination (RFE), an embedded variable selection method, is applied to select the optimal inputs to the base prediction models developed separately using eight popular predictive algorithms. The parameters of each model are then obtained through leave-group-out cross validation (LGOVCV). Finally, the ensemble model is developed and the weights of the eight predictive models are optimized using genetic algorithm (GA).

The approach is adopted to analyze the large energy consumption data of the tallest building in Hong Kong. The prediction accuracies of the ensemble models measured by mean absolute percentage error (MAPE) are 2.32% and 2.85% for the next-day energy consumption and peak power demand respectively, which are evidently higher than those of individual base models. The results also show that the outlier detection method is effective in identifying the abnormal daily energy consumption profiles. The RFE process can significantly reduce the computation load while enhancing the model performance. The ensemble models are valuable for developing strategies of fault detection and diagnosis, operation optimization and interactions between buildings and smart grid.

CHAPTER 3

MATERIALS AND METHODS

3.1 EXISITING SYSTEM

In existing system, delves into the data mining technology to determine its application in the analysis of building energy consumption data including energy consumption prediction, fault diagnosis, and optimal operation. Recent literature are reviewed and summarized, the problems faced by data mining technology in the area of energy consumption data analysis are enumerated, and research points for future studies are given.

3.1.1 DISADVANTAGES OF EXISITNG SYSTEM

- Insufficient.
- Efficiency is low.

3.2 PROPOSED SYSTEM

In proposed technology is now available that can accurately monitor, collect, and store the huge amount of data involved in this process. Furthermore, this technology is capable of analyzing and exploiting such data in meaningful ways. Not surprisingly, the use of data science techniques to increase energy efficiency is currently attracting a great deal of attention and interest. This paper reviews how Data Science has been applied to address the most difficult problems faced by practitioners in the field of Energy Management, especially in the building sector. The work also discusses the challenges and opportunities that will arise with the advent of fully connected devices and new computational technologies.

3.2.1 ADVANTAGES OF PROPOSED SYSTEM

- It is more sufficient.
- Efficiency is high.

3.3 PROPOSED ALGORITHM

3.3.1 K-MEANS CLUSTERING ALGORITHM:

What is K-means?

1. Partitional clustering approach

2. Each cluster is associated with a centroid (center point)
3. Each point is assigned to the cluster with the closest centroid
- 4 Number of clusters K must be specified

<p>Input: k (the number of clusters), D (a set of lift ratios)</p> <p>Output: a set of k clusters</p> <p>Method: Arbitrarily choose k objects from D as the initial cluster centers;</p> <p>Repeat:</p> <ol style="list-style-type: none">1. (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;2. Update the cluster means, i.e., calculate the mean value of the objects for each cluster <p>Until no change;</p>
--

3.3.2 Details of K-means

1. Initial centroids are often chosen randomly. - Clusters produced vary from one run to another
2. The centroid is (typically) the mean of the points in the cluster.
3. 'Closeness' is measured by Euclidean distance, cosine similarity, correlation, etc.
4. K-means will converge for common similarity measures mentioned above.
5. Most of the convergence happens in the first few iterations. - Often the stopping condition is changed to 'Until relatively few points change clusters'

3.4 NAIVE BAYES ALGORITHM

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$. Look at the equation below:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$

Above,

- $P(c|x)$ is the posterior probability of class (*c, target*) given predictor (*x, attributes*).
- $P(c)$ is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor.

3.4.1 WORKING NAÏVE BAYES ALGORITHM

Let's understand it using an example. Below I have a training data set of weather and corresponding target variable 'Play' (suggesting possibilities of playing). Now, we need to classify whether players will play or not based on weather condition. Let's follow the below steps to perform it.

Step 1: Convert the data set into a frequency table

Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Problem: Players will play if weather is sunny. Is this statement is correct?

We can solve it using above discussed method of posterior probability.

$$P(\text{Yes} \mid \text{Sunny}) = P(\text{Sunny} \mid \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$$

Here we have $P(\text{Sunny} \mid \text{Yes}) = 3/9 = 0.33$, $P(\text{Sunny}) = 5/14 = 0.36$, $P(\text{Yes}) = 9/14 = 0.64$

Now, $P(\text{Yes} \mid \text{Sunny}) = 0.33 * 0.64 / 0.36 = 0.60$, which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes.

Pros:

- It is easy and fast to predict class of test data set. It also perform well in multi class prediction
- When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.

- It perform well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).

Cons:

- If categorical variable has a category (in test data set), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. This is often known as “Zero Frequency”. To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.
- On the other side naive Bayes is also known as a bad estimator, so the probability outputs from predict_proba are not to be taken too seriously.
- Another limitation of Naive Bayes is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

3.5 Applications of Naive Bayes Algorithms

- **Real time Prediction:** Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.
- **Multi class Prediction:** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
- **Text classification/ Spam Filtering/ Sentiment Analysis:** Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)
- **Recommendation System:** Naive Bayes Classifier and Collaborative Filtering together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not

3.6 SYSTEM ARCHITECTURE

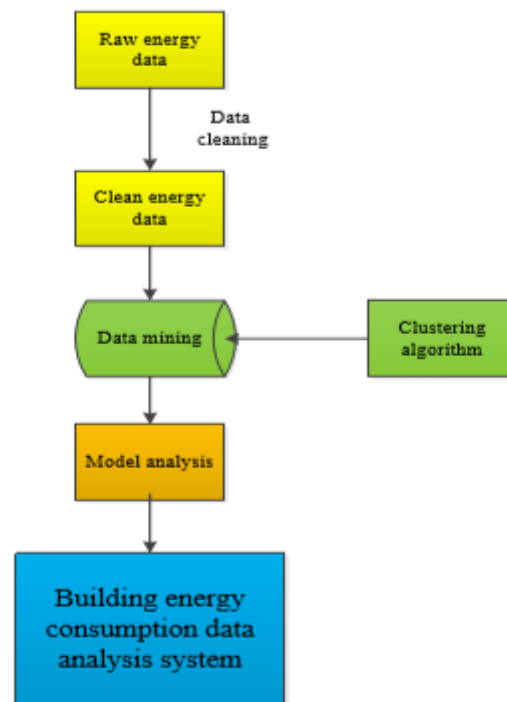


FIG.1 SYSTEM ARCHITECTURE

3.7 SYSTEM REQUIREMENTS

3.7.1 HARDWARE REQUIREMENTS:

- System - Pentium-IV
- Speed - 2.4GHZ
- Hard disk - 40GB
- Monitor - 15VGA color
- RAM - 512MB

3.7.2 SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding language - Java

- IDE - Net beans
- Database -MYSQL

3.8 REQUIREMENT ANALYSIS

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analysing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

3.9 FUNCTIONAL REQUIREMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

3.9.1 Usability

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

3.9.2 Robustness

It refers to a program that performs well not only under ordinary conditions but also under unusual conditions. It is the ability of the user to cope with errors for irrelevant queries during execution.

3.9.3 Security

The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security.

3.9.4 Reliability

It is the probability of how often the software fails. The measurement is often expressed in MTBF (Mean Time Between Failures). The requirement is needed in order to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and survive and even capable of working around any failure.

3.9.5 Compatibility

It is supported by version above all web browsers. Using any web servers like localhost makes the system real-time experience.

3.9.6 Flexibility

The flexibility of the project is provided in such a way that it has the ability to run on different environments being executed by different users.

3.9.7 Safety

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one's personal information.

3.10 NON- FUNCTIONAL REQUIREMENTS

3.10.1 Portability

It is the usability of the same software in different environments. The project can be run in any operating system.

3.10.2 Performance

These requirements determine the resources required, time interval, throughput and everything that deals with the performance of the system.

3.10.3 Accuracy

The result of the requesting query is very accurate and high speed of retrieving information. The degree of security provided by the system is high and effective.

3.10.4 Maintainability

Project is simple as further updates can be easily done without affecting its stability. Maintainability basically defines that how easy it is to maintain the system. It means that how easy it is to maintain the system, analyse, change and test the

application. Maintainability of this project is simple as further updates can be easily done without affecting its stability.

3.11 MODULES

- Energy Data Analysis
- Energy Consumption Data Prediction
- Energy Data Processing

3.11.1 Energy Data Analysis

Building energy consumption data usually has high volume and high dimension. It is very difficult to find and summarize information contained in these data. Based on these data Data collection and preprocessing, Clustering and mining, Model analysis, Model application. Based on the hour-wise energy consumption data of an office building and a large store during a month, experiments are conducted using an improved version of Weka which incorporates the Chameleon algorithm. The proposed model is utilized to partition energy consumption in the office building and large store into three periods, shown as in Figs. 3 and 4, respectively. Energy consumption requirements are imposed on each time period to achieve energy conservation. The proposed model is also helpful in diagnosing energy consumption diagnosis in similar building types.

3.11.2 Energy Consumption Data Prediction

Researchers have developed an effective approach to constructing the prediction models of the next-day energy consumption and peak power demand [18] which can be used to eliminate deficiencies evident in this current study. Abnormal building energy consumption profiles are first identified and are removed using feature extraction, clustering analysis, and the generalized extreme studentized deviate (GESD). Base models are then developed using eight popular predictive algorithms. A data-driven input selection algorithm called the recursive feature elimination (RFE) is applied to individually find inputs to the eight base models respectively. The flowchart of data mining methodology is shown in Fig. 5. The ensemble models are constructed by combining eight base models. A genetic algorithm (GA) is used to

optimize the weights of eight base models in the final ensembles. The proposed approach is applied in the analysis of the vast energy consumption data of the tallest building in Hong Kong. The performances of individual base models and the ensemble models, together with their computation times, are then compared

3.11.3 Energy Data Processing

Considering the state of the data from energy consumption monitoring platform in public buildings, a hierarchical data processing method is proposed. Steps are as follows:

- (1) Data classification: Analyze and classify causes of data anomalies in the energy consumption-monitoring platform, per the building's energy consumption curves and field investigation.
- (2) Abnormal data recognition: Formulate abnormal data recognition methods based on features of various abnormal data.
- (3) Data processing: Devise abnormal data processing schemes. Delete and label different types of abnormal data. Compensate for the deletion of abnormal data through data patterns acquired from data mining.

3.12 UML DIAGRAMS

UML is simply another graphical representation of a common semantic model. UML provides a comprehensive notation for the full lifecycle of object-oriented development.

ADVANTAGES

- To represent complete systems (instead of only the software portion) using object oriented concepts
 - To establish an explicit coupling between concepts and executable code
 - To take into account the scaling factors that are inherent to complex and critical systems
 - To creating a modeling language usable by both humans and machines
- UML defines several models for representing systems

- The class model captures the static structure
- The state model expresses the dynamic behavior of objects
- The use case model describes the requirements of the user
- The interaction model represents the scenarios and messages flows
- The implementation model shows the work units
- The deployment model provides details that pertain to process allocation

3.12.1 USECASE DIAGRAM

Use case diagrams overview the usage requirement for system. They are useful for presentations to management and/or project stakeholders, but for actual development you will find that use cases provide significantly more value because they describe “the meant” of the actual requirements. A use case describes a sequence of action that provides something of measurable value to an action and is drawn as a horizontal ellipse.



FIG 2. USE CASE DIAGRAM

3.12.2 SEQUENCE DIAGRAM

Sequence diagram model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and commonly used for both analysis and design purpose. Sequence diagram are the most popular UML artifact for dynamic modeling, which focuses on identifying the behavior within your system.

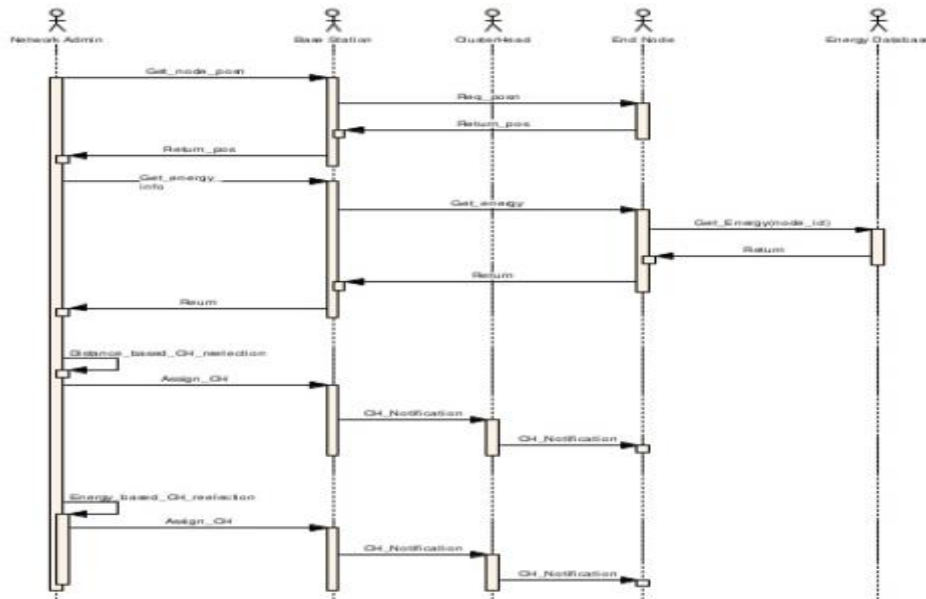


FIG 3.SEQUENCE DIAGRAM

3.12.3 CLASS DIAGRAM

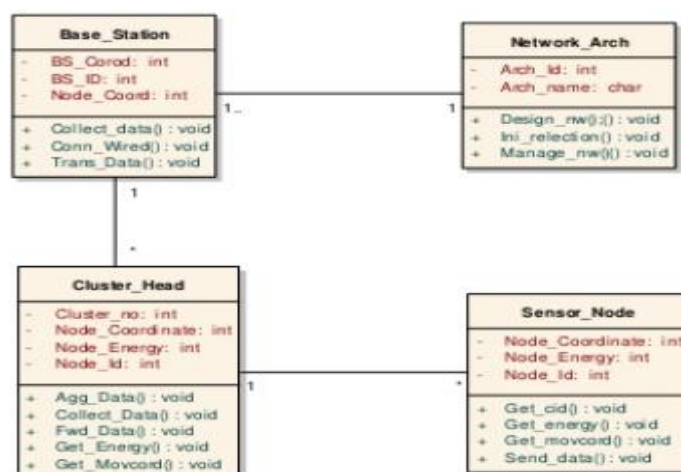


FIG 4.CLASS DIAGRAM

3.12.4 ACTIVITY DIAGRAM

Activity diagram are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. The activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. Activity diagram consist of Initial node, activity final node and activities in between.

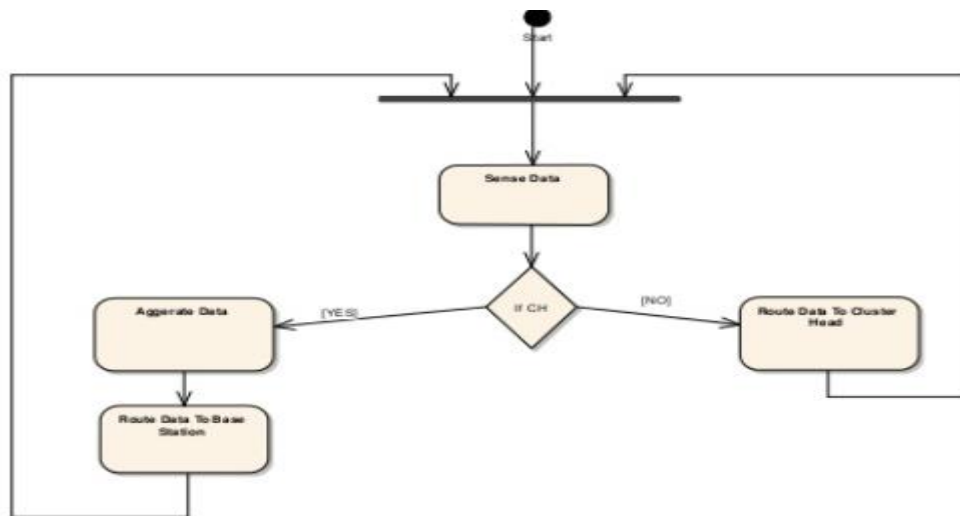


FIG 5 .ACTIVITY DIAGRAM

3.13 SYSTEM DESIGN AND TESTING

3.13.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?

- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

3.13.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.
-

3.14 SYSTEM TESTING

3.14.1 TESTING PROCESS

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

3.15 TYPES OF TESTS

3.15.1 Unit Testing

Unit testing involves the design of test cases that validate that the internal

program logic is functioning properly, and that program input produces valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

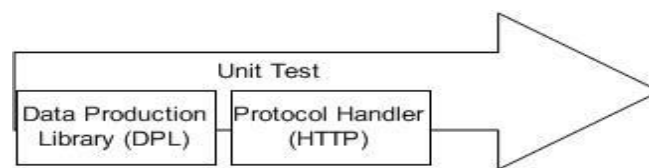


Fig 6. Unit Testing

3.15.2 Integration Testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

3.15.3 Functional Testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation and user manuals.

- Functional testing is centered on the following items:
- Valid Input is used to identified classes of valid input must be accepted.
- Invalid Input is used to identified classes of invalid input must be rejected.
- Functions is used to identified functions must be exercised.

- Output is used to identify classes of application outputs.

Systems/Procedures is used to interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive Processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

3.15.4 System Testing

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

3.15.5 White Box Testing

White Box Testing is a testing in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a black box level.

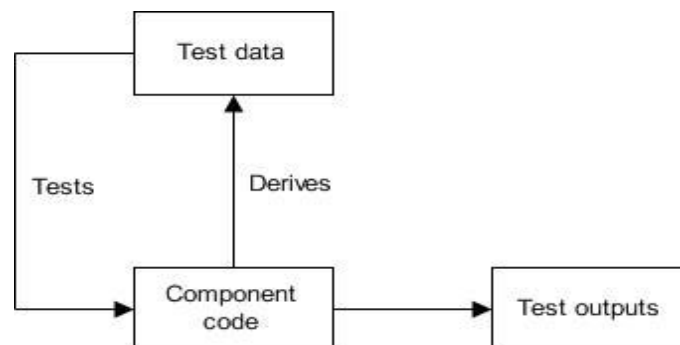


Fig 7. White box Testing

3.15.6 Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as

specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without

Considering how the software works.

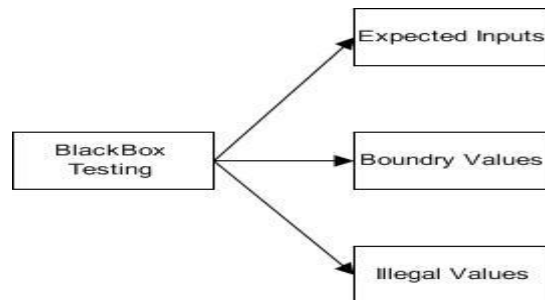


Fig 8. Black box Testing

3.16 TEST STRATEGY AND APPROACH

Field testing will be performed manually and functional tests will be written in detail.

3.16.1 test Objectives

- All field entries must work properly.
- ☐ Pages must be activated from the identified link.
- ☐ The entry screen, messages and responses must not be delayed.
- ☐ Features to be tested
- ☐ Verify that the entries are of the correct format
- ☐ No duplicate entries should be allowed
- ☐ All links should take the user to the correct page.

3.16.2 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by

interface defects.

The task of the integration test is to check that components or software applications.

3.16.3 Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

3.16.4 ALPHA TESTING

In software development, alpha test will be a test among the teams to confirm that your product works. Originally, the term alpha test meant the first phase of testing in a software development process. The first phase includes unit testing, component testing, and system testing. It also enables us to test the product on the lowest common denominator machines to make sure download times are acceptable and pre loaders work.

3.16.5 BETA TESTING

In software development, a beta test is the second phase of software testing in which a sampling of the intended audience tries the product out. Beta testing can be considered "pre-release testing." Beta test versions of software are now distributed to curriculum specialists and teachers to give the program a "real world" test.

CHAPTER 4

RESULTS AND DISCUSSION

In this research, we introduced a method to analyze the electricity use pattern of university buildings using the K-means clustering method. K-means clustering is popular because it is able to cluster large amounts of data both quickly and efficiently. It remains a basic framework for developing numerical or conceptual clustering through various possibilities of distance and prototype choice. However, K-means clustering has one disadvantage: the number of clusters must be chosen before the analysis is conducted. The result of the clustering algorithm depends on the value of the initial centroids chosen for clustering. So, in this research, we proposed a method to select better clustering centroids using the percentile method based on empirical cumulative distribution.

Building electricity use in universities show similar patterns of increasing during the day time and decreasing at the night; so, in this research, the initial centroids were chosen based on the hourly distribution of one year of electricity use. The proposed method was tested for accuracy in terms of number of accurately classified patterns in case of four different real world datasets. In all the 0 10 20 30 40 50 60 Mon Tue Wed Thu Fri Sat Sun No. of days in each cluster Day of the week 1 2 3 4 5 6 Cluster Energies datasets, the proposed method was found to be more accurate..

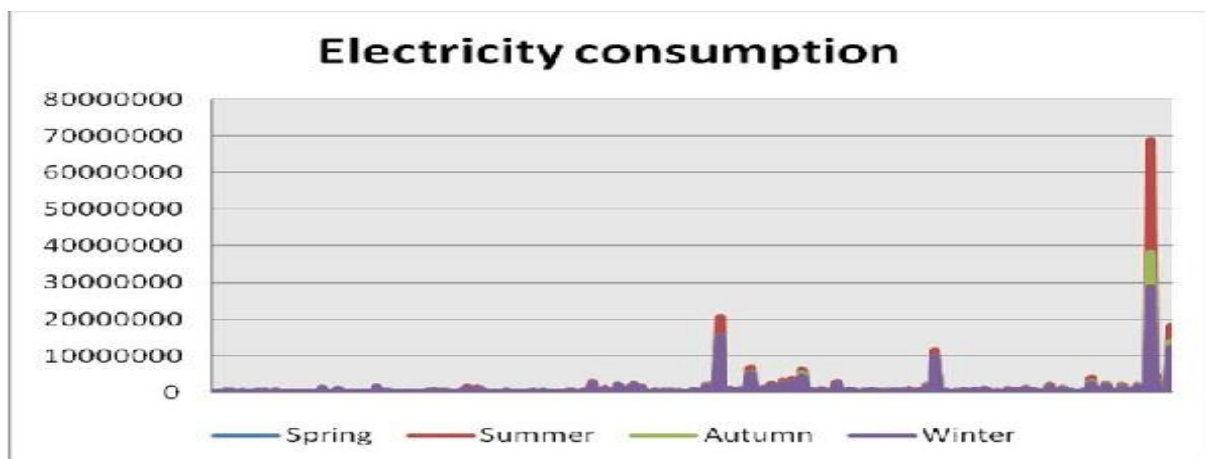


Figure 9. Comparison of electricity consumption between the four seasons.

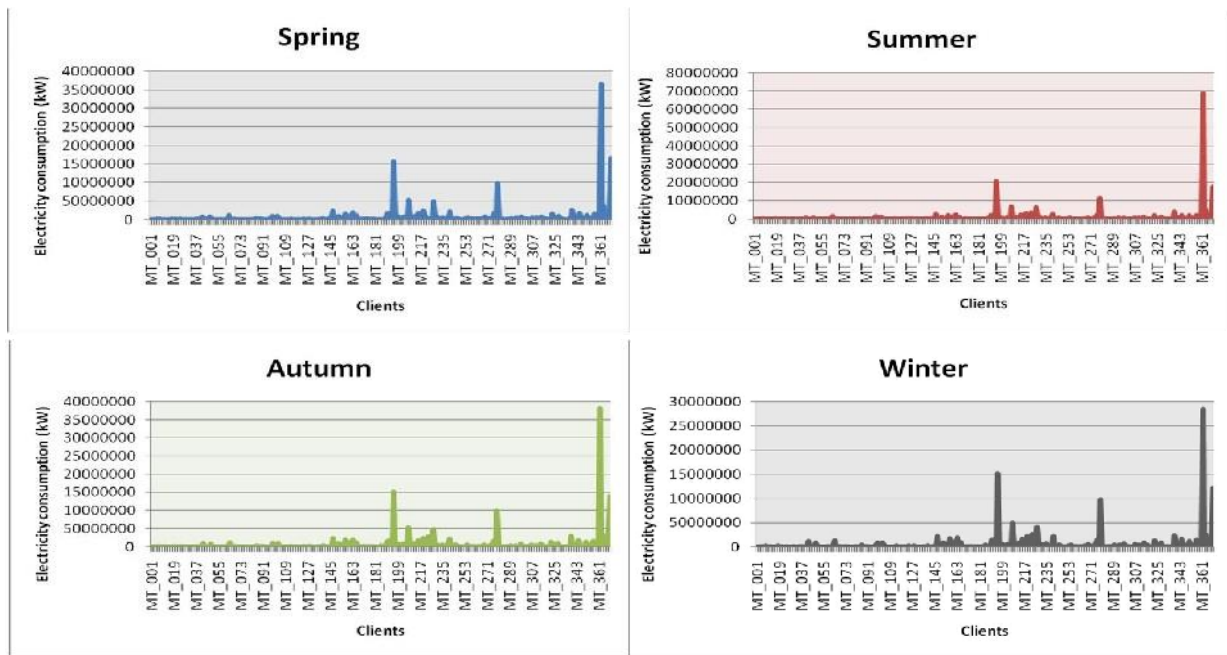


Figure 10. Visualization of electricity consumption in each season

For conservation of energy in buildings, it is necessary to know when and how much electricity is consumed. Analyzing the electricity use pattern manually is a time consuming process. Since clustering is a machine learning process, building electricity use analysis using clustering techniques can improve the efficiency of the work. Without analyzing the electricity consumption of each day, peak electricity use, base electricity use, days with unusual electricity use, etc., can be extracted easily using clustering, saving the analyst's time. Deciding the proper number of clusters is important for obtaining accurate results. In this research, when analyzing the clustering result using three to six clusters, six clusters was found to be appropriate. Thus, K-means clustering using six clusters could analyze the electricity use pattern of Chubu University. The cluster center and calendar plot information regarding when and how much electricity is consumed can be obtained without concerning the university schedule with noticeably good accuracy. Once daily electricity use has been categorized, base electricity consumption, electricity consumption by human activities, and energy consumption by air-conditioning can be determined. As energy consumption by usage is clarified, measures for energy consumption in university buildings can be proposed.

Results

The example of result dataset after the process of dimension reduction

	row.names	V1	V2	V3	V4
1	MT_001	36017.77	64649.75	98902.28	19529.19
2	MT_002	251569	308256	240322.9	225683.5
3	MT_003	17935.71	16034.75	14913.12	14728.06
4	MT_004	839878	944004.1	981579.3	1212486
5	MT_005	352457.3	427929.3	460042.7	577119.5
6	MT_006	1503557	1546244	1738848	1951997
7	MT_007	31565.86	157960.4	33873.94	45366.87
8	MT_008	2151296	2434896	2236761	2277626
9	MT_009	381304.2	447903.8	516484.3	496896.9
10	MT_010	548790.3	438402.2	535288.2	508466.7
11	MT_011	293096.9	335868.9	350378.5	421040.2
12	MT_012	0	333385.1	1259196	1513834
13	MT_013	602344.9	714411.3	679792.7	652699.8
14	MT_014	356758.8	336899.2	407737.7	473116.8
15	MT_015	58722.62	176699.4	163892.6	448886.9
16	MT_016	254970.4	313702.2	281698.2	336478.9
17	MT_017	393219.2	454653.2	427071.9	449489.7

Table4: *The data sample after dimension reduction.*

CHAPTER 5

CONCLUSION AND SUMMARY

This paper has reviewed recent developments in information technologies and their influence on Building Energy Management. We examined the usefulness of various data science techniques that have been applied or could be applied to solve energy problems. Given the current challenges that must be addressed in energy management, it is evident that data science techniques will be widely applied in the near future. In all areas, the discovery and exploitation of the information hidden within collected data is extremely useful. However, in the case of energy consumption, this is even more so because of the economic and environmental implications. In Building Energy Management, the identification of equipment and user consumption patterns will doubtlessly save money, improve comfort, and reduce contaminant emissions. The economic impact of this sector is reflected in the number of new companies that are currently applying Data Science to the analysis of energy consumption data, user habits, and building infrastructure.

SUMMARY:

This is leading to synergies between energy companies and information technology enterprises who are beginning to work together towards more efficient energy management. This new context actively challenges researchers to develop solutions for the management of huge amounts of heterogeneous data in real time, as well as to find ways to deal with its associated uncertainty. Data science techniques have shown themselves to be valuable tools capable of extracting and exploiting the knowledge and information inherent in user data. In the near future, Big Data techniques will expand these possibilities and democratize them. This will enhance energy awareness, since users will have access to more data and be able to understand their own energy consumption habits. In this regard, companies have begun to realize that energy savings are not only a question of optimizing components, but also of understanding and acting on user behaviors.

REFERENCES

- [1] Itard L, Meijer F, Vriens E, et al. Building renovation and modernisation in Europe: State-of-the-art review (R). Netherlands: OTB Research Institute for Housing, Urban and Mobility, 2008.
- [2] Commission E. Action plan for energy efficiency: Realising the potential (R). Brussels: Commission of the European Communities, 2006.
- [3] Zhao L, Zhang J, Liang R. Development of an energy monitoring system for large public buildings (J). *Energy and Buildings*, 2013, 66: 41-48.
- [4] Han J W, Kamber M. *Data Mining: Concepts and Techniques* (M). 3rd ed. The Morgan Kaufmann Series in Data Management Systems, 2011.
- [5] Fan C, Xiao F, Wang S W. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques (J). *Apply Energy*. 2014, 127: 1-10.
- [6] Seem J E. Using intelligent data analysis to detect abnormal energy consumption in buildings (J). *Energy and Buildings*, 2007, 39(1): 52-58.
- [7] Li X L, Bowers C P, Schnier T. Classification of energy consumption in buildings with outlier detection (J). *IEEE Transactions on Industrial Electronics*, 2010, 57(11): 3639-3644.
- [8] Qing X X, Xiao D, Wang B. A real-time monitoring method of energy consumption based on data mining (J). *Journal of Chongqing University*, 2012, 35(7): 133-137.
- [9] O. Maimon, L. Rokach, *Data Mining and Knowledge Discovery Handbook*, 2nd ed., Springer, New York, 2010.
- [10] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region (J). *Energy and Buildings* 37 (2005) 545–553.
- [11] M.R. Amin-Naseri, A.R. Soroush, Combined use of unsupervised and supervised learning for daily peak load forecasting (J). *Energy Conversion and Management* 49 (2008) 1302–1308.
- [12] A. Kusiak, M.Y. Li, F. Tang, Modeling and optimization of HVAC energy consumption (J). *Applied Energy* 87 (2010) 3092–3102.

- [13] A. Ahmed, N.E. Korres, J. Ploennigs, H. Elhadi, K. Menzel, Mining building performance data for energy-efficient operation (J). *Advanced Engineering Informatics* 25 (2011) 341–354.
- [14] Z. Yu, F. Haghighat, C.M. Fung, H. Yoshino, A decision tree method for building energy demand modeling (J). *Energy and Buildings* 42 (2010) 1637–1646.
- [15] Z. Yu, F. Haghighat, C.M. Fung, L. Zhou, A novel methodology for knowledge discovery through mining associations between building operational data (J). *Energy and Buildings* 47 (2012) 430–440.

APPENDIX

SAMPLE SOURCE CODE:

```
import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.Statement;
import javax.swing.JFileChooser;
import org.jfree.ui.RefineryUtilities;

/*
 * To change this license header, choose License Headers in Project Properties.
 * To change this template file, choose Tools | Templates
 * and open the template in the editor.
 */

/**
 *
 * @author Benin sowmi
 */
public class Home extends javax.swing.JFrame {
    Statement st2,st3,st4;
    ResultSet rs1,rs2,rs3;
    public static double pcount=0.0,ncount=0.0,acount=0.0;
    public static int i1=0;
    /**
     * Creates new form Home
     */
    public Home() {
        initComponents();
    }
}
```

```

/**
 * This method is called from within the constructor to initialize the form.
 * WARNING: Do NOT modify this code. The content of this method is always
 * regenerated by the Form Editor.
 */
@SuppressWarnings("unchecked")
// <editor-fold defaultstate="collapsed" desc="Generated Code">
private void initComponents() {

    jPanel1 = new javax.swing.JPanel();
    jLabel1 = new javax.swing.JLabel();
    jButton1 = new javax.swing.JButton();
    jLabel2 = new javax.swing.JLabel();
    jButton2 = new javax.swing.JButton();
    jButton3 = new javax.swing.JButton();
    jComboBox1 = new javax.swing.JComboBox<>();
    jLabel7 = new javax.swing.JLabel();
    jPanel2 = new javax.swing.JPanel();
    jLabel3 = new javax.swing.JLabel();

    setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);

    jPanel1.setBackground(new java.awt.Color(204, 255, 255));

    jLabel1.setText("Load Dataset");

    jButton1.setText("Select CSV File");
    jButton1.addActionListener(new java.awt.event.ActionListener() {
        public void actionPerformed(java.awt.event.ActionEvent evt) {
            jButton1ActionPerformed(evt);
        }
    });

    jLabel2.setText("Path");

```

```
jButton2.setText("Process");

jButton2.addActionListener(new java.awt.event.ActionListener() {

    public void actionPerformed(java.awt.event.ActionEvent evt) {

        jButton2ActionPerformed(evt);

    }

});


jComboBox1.setModel(new javax.swing.DefaultComboBoxModel<>(new String[] { "Item 1", "Item 2", "Item 3", "Item 4" }));


jLabel7.setText("Area");


javax.swing.GroupLayout jPanel1Layout = new javax.swing.GroupLayout(jPanel1);
jPanel1.setLayout(jPanel1Layout);
jPanel1Layout.setHorizontalGroup(
    jPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
        .addGroup(jPanel1Layout.createSequentialGroup()
            .addGap(25, 25, 25)
            .addGroup(jPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
                .addComponent(jLabel1, javax.swing.GroupLayout.PREFERRED_SIZE, 80, javax.swing.GroupLayout.PREFERRED_SIZE)
                .addGroup(layout.createSequentialGroup()
                    .addGap(18, 18, 18)
                    .addComponent(jPanel2, javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE)
                    .addGap(18, 18, 18))
            )
        )
);

pack();

} // </editor-fold>

st.executeUpdate("insert into dt
values("+dt1[0]+","+Integer.parseInt(dt1[1])+","+Integer.parseInt(dt1[2])+","+Double.parseDouble(dt1[3])+","+Integer.parseInt(dt1[4])+","+Double.parseDouble(dt1[5])+"");
```

```

    }
}
i1=0;
rs1= st3.executeQuery("select DISTINCT Area from dt");
jComboBox1.removeAllItems();
jComboBox1.addItem("All");
while(rs1.next())
{
    jComboBox1.addItem(rs1.getString(1));
}
} catch (Exception e) {
    e.printStackTrace();
}
}

/**
 * @param args the command line arguments
 */
public static void main(String args[]) {
    /* Set the Nimbus look and feel */
    //<editor-fold defaultstate="collapsed" desc=" Look and feel setting code (optional) ">
    /* If Nimbus (introduced in Java SE 6) is not available, stay with the default look and feel.
     * For details see http://download.oracle.com/javase/tutorial/uiswing/lookandfeel/plaf.html
     */
    try {
        for (javax.swing.UIManager.LookAndFeelInfo info : javax.swing.UIManager.getInstalledLookAndFeels()) {
            if ("Nimbus".equals(info.getName())) {
                javax.swing.UIManager.setLookAndFeel(info.getClassName());
                break;
            }
        }
    } catch (ClassNotFoundException ex) {
        java.util.logging.Logger.getLogger(Home.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (InstantiationException ex) {
        java.util.logging.Logger.getLogger(Home.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (IllegalAccessException ex) {

```

```

        java.util.logging.Logger.getLogger(Home.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (javax.swing.UnsupportedLookAndFeelException ex) {
        java.util.logging.Logger.getLogger(Home.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    }
}

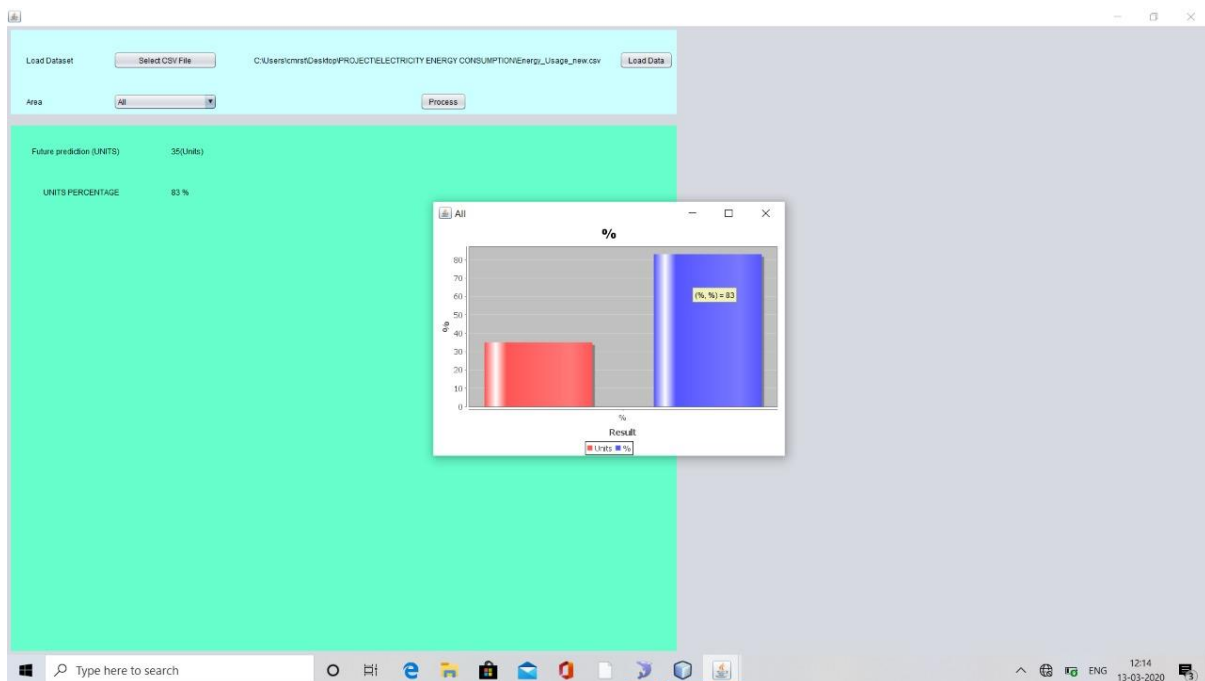
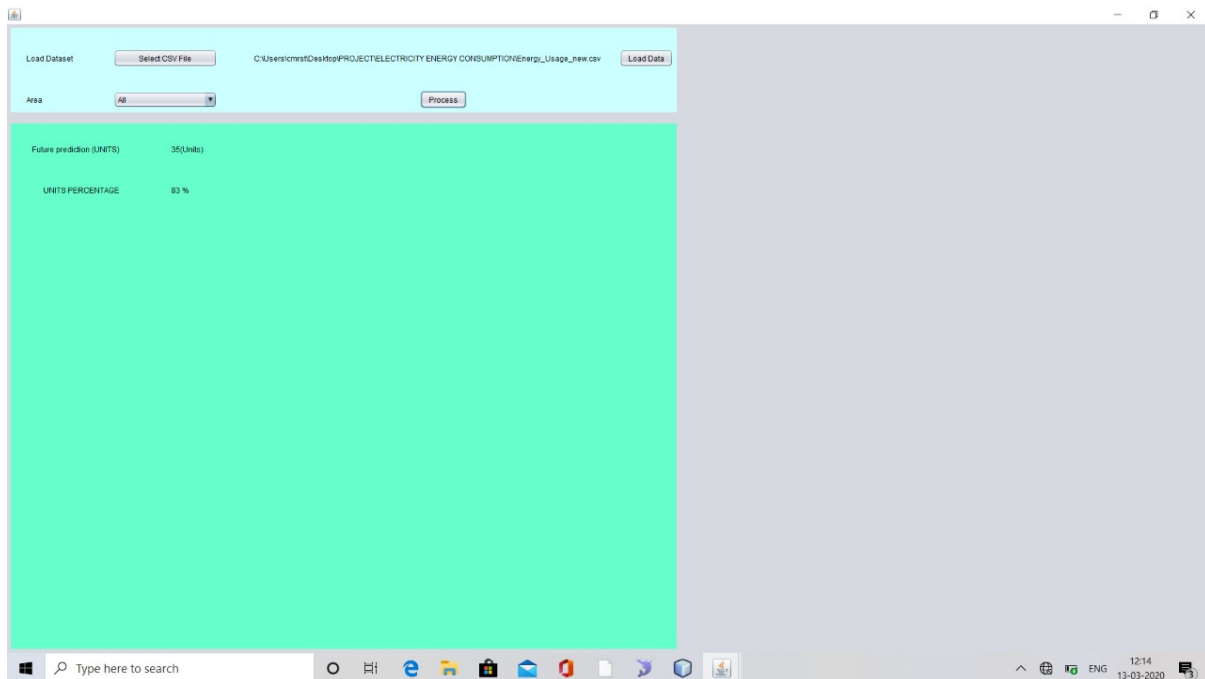
//</editor-fold>

/* Create and display the form */
java.awt.EventQueue.invokeLater(new Runnable() {
    public void run() {
        new Home().setVisible(true);
    }
});
}

// Variables declaration - do not modify
private javax.swing.JButton jButton1;
private javax.swing.JButton jButton2;
private javax.swing.JButton jButton3;
private javax.swing.JComboBox<String> jComboBox1;
private javax.swing.JLabel jLabel1;
private javax.swing.JLabel jLabel2;
private javax.swing.JLabel jLabel3;
private javax.swing.JLabel jLabel4;
private javax.swing.JLabel jLabel5;
private javax.swing.JLabel jLabel6;
private javax.swing.JLabel jLabel7;
private javax.swing.JPanel jPanel1;
private javax.swing.JPanel jPanel2;
// End of variables declaration
}

```

SCREEN SHOTS:



PAPER REPORT:



Plagiarism Checker X Originality Report

Similarity Found: 13%

Date: Thursday, February 27, 2020

Statistics: 612 words Plagiarized / 4808 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

ANALYSIS OF ELECTRICITY CONSUMPTION AT HOME USING K-MEANS CLUSTERING ALGORITHM ABSTRACT The vitality utilization of private and business structures has risen consistently as of late, an expansion generally because of their HVAC frameworks. Expected vitality burdens, transportation, and capacity just as client conduct impact the amount and nature of the vitality expended day by day in structures. Be that as it may, innovation is presently accessible that can precisely screen, gather, and store the colossal measure of information associated with this procedure.

Besides, this innovation is equipped for breaking down and misusing such information in significant manners. As anyone might expect, the utilization of information science systems to expand vitality proficiency is right now drawing in a lot of consideration and intrigue. This paper surveys how Data Science has been applied to address the most troublesome issues looked by experts in the field of Energy Management, particularly in the structure area.

The work likewise talks about the difficulties and openings that will emerge with the appearance of completely associated gadgets and new computational advancements INTRODUCTION Vitality has gotten basic for human presence and advancement. The unevenness between lessening vitality supply and expanding vitality request has escalated bit by bit. Therefore, vitality preservation has become a point that draws in worldwide consideration. Vitality utilization in structures has a gigantic impact on supportability. As per the measurements from writing, structures represent 32% of the aggregate sum of vitality expended around the world.

In the US and other created countries, this figure is 41%, making the development business the greatest vitality customer. The proportion of the structures vitality

As indicated by China's report on vitality sparing of structures in rustic and urban regions in 2013, continuous vitality utilization observing was performed on 5,000 or more enormous open structures before the finish of 2013. The vitality sparing grounds development ventures were steered in 210 universities as appeared in Fig. 1. The improvement of the framework for observing vitality utilization in structures can give huge scope information data accommodating in improving vitality the board and decreasing vitality utilization.

It is additionally the most natural wellspring of information concerning the operational status of structures. BECMP gathers the temperature, moistness, pressure, electrical signals and control signs of building frameworks. These are basic parts of building framework activity and the executives, execution conclusion, shortcoming examination and vitality effectiveness assessment. The vitality utilization information of structures **is influenced by the** structure, atmosphere natural parameters, development, working modes, and client standards of conduct.

Customary techniques for investigation are not adequate in taking care of huge scope information which are produced during the activity of the vitality utilization observing frameworks. Furthermore, the board work force will most likely be unable to discover and process strange information that may happen. This outcomes in a genuine inconsistency between the vitality utilization information "on paper" and the real vitality utilization. **The primary explanation behind this is** the customary strategies embrace the vitality utilization recreation conspire. This requires the clients to have a decent order of the structure and framework engineering.

Moreover, it is constrained, off base and can't process helpful data rapidly for reasonable applications. As a developing innovation, information mining has massive quality in large information handling, and it has been used in AI, human insight, design acknowledgment, measurements, informatics, and information representation. It has drawn developing consideration from different ventures and has become a greater player in the board dynamic.

The information mining innovation centers around information as opposed to registering with confounded recipes, in this way it is simpler to incorporate in numerous applications. It is a rising amazing innovation with incredible potential to find covered up and valuable information in enormous informational collections and keeping in mind that it has utilized in the development field of the current "rich information and poor data" predicament, its application in the field of building vitality frameworks, especially in the vitality utilization information, has been negligible. LITERATURE SURVEY Writing overview is the most significant advance in programming improvement process.

Before building up the device it is important to decide the time factor, economy and friends quality. When these things are fulfilled, at that point the following stage is to figure out which working framework and language can be utilized for building up the apparatus. When the software engineers begin assembling the apparatus the developers need parcel of outside help. This help can be gotten from senior developers, from book or from sites.

Before building the framework the above thought are considered for building up the proposed framework. The significant piece of the venture improvement area considers and completely overview all the necessary requirements for building up the task. For each venture Literature study is the most significant part in programming advancement process. Before building up the apparatuses and the related structuring it is important to decide and study the time factor, asset prerequisite, labor, economy, and friends quality.

When these things are fulfilled and completely studied, at that point the following stage is to decide about the product particulars in the individual framework, for example, what kind of working framework the venture would require, and what are largely the vital programming are expected to continue with the subsequent stage, for example, building up the instruments, and the related activities. Building remodel and modernisation in Europe: State-of-the-workmanship survey (R).Netherlands: OTB Research Institute for Housing The last reason for the undertaking "Best in class audit of building remodel and modernisation in Europe", propelled by Erabuild, the European Research Area for the practical development and activity of structures, is to distinguish the most encouraging fields for future exercises inside the trans-national Erabuild look into program on Sustainable Renovation.

To meet this last reason, four work bundles have been characterized and are dealt with progressively in the present report. 1. Mapping of building typologies and partner interests (WP1). 2. Examination of existing impetuses and their effect on the redesign rate (WP2). 3. Mapping of modernisation and remodel look into (WP3). 4. Proposals

about the most encouraging fields for future exercises (WP4). The nations shrouded right now Austria, Finland, France, Germany, the Netherlands, Sweden, Switzerland and the United Kingdom. Private and nonresidential structures are secured.

Activity plan for vitality productivity: Realizing the potential (R) Vitality is basic for Europe to work. Be that as it may, the times of modest vitality for Europe appear to be finished. The difficulties of environmental change, expanding import reliance and higher vitality costs are looked by all EU individuals. Besides the relationship of EU Member States in vitality, as in numerous different zones, is expanding – a force disappointment in one nation has quick impacts in others. Europe needs to act now, together, to convey economical, secure and serious vitality.

In doing so the EU would come back to its underlying foundations. In 1952 with the Coal and Steel Treaty and 1957 with the Euratom Treaty, the establishing Member States saw the requirement for a typical way to deal with vitality. Vitality markets and geopolitical contemplations have changed altogether from that point forward. In any case, the requirement for EU activity is more grounded than at any other time. Without this, the EU's destinations in different zones, including the Lisbon Strategy for development and occupations and the Millennium Development Goals, will likewise be increasingly hard to accomplish.

Another European Energy Policy should be driven, serious and long haul – and to the advantage everything being equal. Improvement of a vitality checking framework for huge open structures (J) The inclination of open structure proprietors to the vitality sparing motivating force Instruments of building decides the impact of approach tools, which is planned and actualized by the government. In this paper, firstly we characterizes the sorts of building vitality sparing motivator policies, and recognize the impact variables of open structure proprietors to impetus arrangement tools. Secondly, we break down the qualities of the components influencing the strategy inclinations of the open structures, the inclination of the open structure proprietors to those arrangement apparatuses and the connection between them. Finally, so as to improve the vitality effectiveness of open buildings, the paper advances some approach proposals for the legislature. Information Mining: Concepts and Techniques (M) Breaking down a lot of information is a need.

Indeed, even famous science books, similar to "super crunchers," give convincing situations where a lot of information yield disclosures and instincts that shock even specialists. Each endeavor profits by gathering and breaking down its information: Hospitals can spot patterns and irregularities in their patient records, web crawlers can improve positioning and promotion arrangement, and ecological and general wellbeing

offices can spot examples and anomalies in their information.

The rundown proceeds, with cybersecurity and PC organize interruption recognition; checking of the vitality utilization of family unit apparatuses; design examination in bioinformatics and pharmaceutical information; money related what's more, business insight information; spotting patterns in web journals, Twitter, and some more. Capacity is reasonable and getting even less along these lines, as are information sensors. In this way, gathering and putting away information is simpler than any time in recent memory. The issue at that point turns out to be the manner by which to break down the information. This is actually the focal point of this Third Edition of the book.

Jiawei, Micheline, and Jian give broad inclusion of all the related techniques, from the exemplary subjects of bunching and order, to database strategies (e.g., affiliation rules, information 3D shapes) to later and propelled themes (e.g., SVD/PCA, wavelets, bolster vector machines). The article is very open to tenderfoots and propelled perusers the same. The book gives the basic material first and the further developed material in follow-up sections. It additionally has various facetious inquiries, which I discovered very supportive for looking after core interest.

We have utilized the first two versions as reading material in quite a while **mining courses at Carnegie Mellon and plan to** keep on doing as such **with this Third Edition.** **The new** form has noteworthy augmentations: Notably, it has in excess of 100 references **to works from 2006** forward, concentrating on later material, for example, charts and informal communities, sensor systems, and anomaly identification. Information handling technique for open structure vitality utilization checking frameworks dependent on information mining Shrewd streetlamp as a significant piece of savvy urban areas targeting building urban communities with a more elevated level of insight, robotization, data and system **has been applied in** urban development generally and effectively.

This paper shows an information preparing technique for shrewd streetlamp vitality utilization examination frameworks dependent on information mining, which isn't just used to order, perceive and address issue information in vitality utilization information, yet in addition utilized in savvy streetlamp vitality utilization investigation frameworks for irregular alert, investigating counsel, vitality utilization forecast, and so on. **It is important to** process vitality utilization information before vitality utilization examination, in light of the fact that the issue information may meddle the investigation and cause deviations and blunders.

Additionally, the technique proposed is utilized to understand the elements of preparing

authentic information and observing right now. At last, the exploratory outcome demonstrates the accessibility and unwavering quality of the technique. Improvement of expectation models for following day building vitality utilization and pinnacle power request utilizing information mining systems This paper introduces an information mining (DM) based way to deal with creating troupe models for foreseeing following day vitality utilization and pinnacle power request, with the point of improving the expectation precision. This methodology for the most part comprises of three stages.

Right off the bat, anomaly recognition, which unions include extraction, bunching examination, and the summed up extraordinary studentized go astray (GESD), is performed to expel the strange every day vitality utilization profiles. Also, the recursive component disposal (RFE), an installed variable choice technique, is applied to choose the ideal contributions to the base forecast models grew independently utilizing eight mainstream prescient calculations. The parameters of each model are then gotten through forget about gathering cross approval (LGOCV). At long last, the gathering model is created and the loads of the eight prescient models are improved utilizing hereditary calculation (GA).