Industrial Energy Audit Web Application using Data Mining Model

Noor M. Maricar*

Electrical Engineering Department

University of Hail

Abstract-- **The excessive use of energy in industrial sectors necessitates the decision maker to always question on how the energy is being used efficiently. Energy used for air-conditioning and lighting in a medium industry counts for almost 60% of the total energy used. The small percentage of energy use reduction relates to the lower product cost and higher profit margins. Therefore, it is important to the decision makers of an industry to have a proper method to audit the building plant and to come up with the practical actions needed in optimizing the use of energy, while at the same time to improve the comfort and product quality. This paper shows the data mining web application for energy audit that can be used in a typical industrial site.**

Index Terms—**Air-conditioning, Data mining, Energy audit, Energy conservation, Energy management system, Energy saving, Lighting, Web application.**

I. INTRODUCTION

Air-conditioning and lighting are two of the major uses in electricity in every sector and building type across Malaysia and accounts for approximately 60% of national electricity use [1]. Thus there is great potential for saving electricity, reducing the emission of pollutant gases associated with electricity production, and reducing consumer energy costs through the use of more efficient air-con and lighting technologies as well as advance air-con and lighting design practices and control strategies.

Energy audit is a measure to check the energy utilization on how much energy in the building is being used and should be used, and to what extend can energy requirement be reduced by using energy efficient procedures and devices.

Data mining is a process of employing one or more computer learning techniques to automatically analyze and extract knowledge from data contained within a database [2]. The purpose of data mining application is to identify trends and patterns in data [7]. The use of web application in this process enables users to access from remote location and guarantees the data security and integrity.

This paper addresses the data mining web application in air-con and lighting energy audit of an industrial building. It consists of the introduction, data mining process model, web application-based data mining, energy audit methodology, a case study example of data mining web application in energy audit, and the conclusions.

II. DATA MINING PROCESS MODEL

The knowledge gain from a data mining session is given as a model or generalization of the data. Several data mining exist, however all data mining methods use induction-based learning [2]. Induction-based learning is the process of forming general concept definitions by observing specific examples of concepts to be learned. Here are some examples of knowledge gained through the process of induction-based learning.

- Do you know why the air-conditioning and lighting energy audit are the most two common audit taken place? The use of combined air-conditioning and lighting energy in premises normally around 60% to 70%.
- How could we make almost 20% energy usage reduction just from a walk through audit in the old premises? Over design of equipment installed in an old building and changing the old equipment with the new technology that have double to triple efficiency improvement are the two main reasons.
- Does it make sense for you to know that the energy audit practices become popular very fast in the beginning stages when it starts? Normally the early energy audit projects have rapid payback, high probability of success, and few negative consequences.

Those examples make it easy to see why data mining is fast becoming preferred technique for extracting useful knowledge from data. Knowledge discovery in databases (KDD) is a term frequently used interchangeably with data mining. Technically KDD is the application of the scientific method to data mining. In addition to performing data mining, atypical KDD process model included a methodology for extracting and preparing data as well as making decisions about actions to be taken once data mining has taken place. When a particular application involves the analysis of large volumes of data stored in several locations, data extraction and preparation become the most time-consuming parts of the discovery process.

A. Computer Learning

As the definition implies, data mining is about learning. Learning is a complex process. Four levels of learning can be differentiated [5].

- Facts: a simple statement of truth.

---

*The author can be contacted by email: nmaricar@vt.edu or nmmaricar@yahoo.com

- Concepts: a set of objects, symbols or events grouped together because they share certain characteristics
- Procedures: a step-by-step course of action to achieve a goal.
- Principles: general truths or laws that is basic to other truths. It represents the highest level of learning.

Computers are good at learning concepts. Concepts are the output of a data mining session. The data-mining tool dictates the form of learned concepts. Common concept structures include trees, rules, networks, and mathematical equations. Tree structures and production rules are easy for humans to interpret and understand. Networks and mathematical equations are black-box concept structures in that the knowledge they contain is not really understood.

B. A Four-step Process Model

In a broad sense, the data mining can be defined as a four-step process.
1. Assemble a collection of data to analyze;
2. Present these data to a data mining software program;
3. Interpret the results;
4. Apply the results to a new problem or situation.

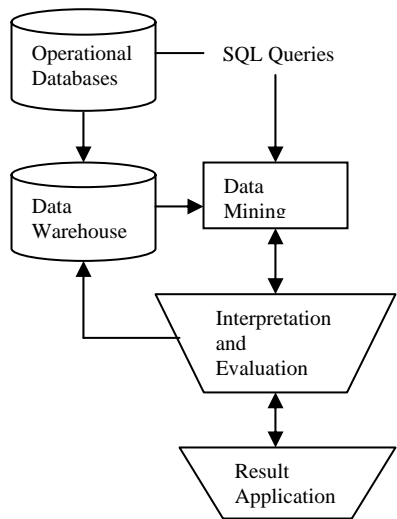A proposed data mining process model that incorporates these four steps is shown in Fig 1.

Fig. 1. Data Mining Process Model

1)   Assembling the Data

Data mining requires data. The data may be represented as volumes of records in several database files or some records in a single file. After a problem has been defined, a first step in data mining process is to extract or assemble a relevant subset of data for processing. This step requires a great amount of human time and effort. There are three common ways to access data in data mining: from a data warehouse, from a relational database, and from a flat file or spreadsheet.

Fig 1 shows data being transferred from the operational environment of operational database to a data warehouse. Operational databases are transaction-based, designed using relational database and having several normalized tables to

reduce redundancy and promote quick access to individual records. The data warehouse is a historical database designed for decision support rather than transaction-processing [6]. Therefore only data useful for decision support is extracted from the operational environment and entered into the warehouse database. Data transfer from the operational to the warehouse database is an ongoing process and usually done daily. Before each data items enter the warehouse, the item is time-stamped, transformed as necessary and checked for errors.

If a data warehouse does not exist, a database query language such as SQL can be used to create a table suitable for data mining. Finally, when a database structured has not been designed and the amount of collected data is minimal, the data can be stored in a flat file or spreadsheet.

2)   Mining the Data

Data mining is used to build generalized models to represent unstructured data. The data classification is an important step, where the *nearest neighbor classification method* can be used [2]:
- Create a classification table containing all data items whose classification is known.
- For each new instance:
  o   Compute a score representing the similarity of each table item.
  o   Give the same classification as the table item to which it is most similar.
  o   Add the new classified instance to the table.

Mining the data is done using the data mining tools which has the following options to choose:
- Which type of learning should be used, supervised or unsupervised?
- Which instances in the assembled of data will be used for building the model and which instances will test the model?
- Which attributes will be selected from the list of available attributes?
- What parameter settings should be used to build a model to best representing the data?

3)   Interpreting the Results

Result interpretation requires the output examination of the data-mining tool to determine if what has been discovered is both useful and interesting. Fig 1 shows that if the result are less than optimal, the data-mining step can be repeated using new attributes and/or instances. Alternatively, the decision is to return to the data warehouse and repeat the data extraction process:

4)   Result Application

The ultimate goal is to apply what has been discovered to new situations. Data mining has been used to build models for predicting the *intrinsic value*. An intrinsic value is the expected value based on historical value of similar objects. Once an object's intrinsic value is determined, an appropriate strategy can be applied. Three possibilities for each new object can be derived:

- A new object's intrinsic value is in line with their actual value.
- A new object's intrinsic value is greater than their actual value.
- A new object's intrinsic value is less than their actual value.

The following sections will brief out the lighting concepts, followed by lighting audit flow diagram and a case study example in using the four-step process model introduced in this section.

III. WEB APPLICATION-BASED DATA MINING

Large-scale and medium-scale Web applications designed for supporting the energy business must ensure a high level of availability, security and scalability, because they can be exposed to millions of concurrent users in the potentially hostile Internet environment [8]. To ensure the required level of service, enterprise Web applications must have a modular architecture, where each component can be easily replicated, to increase performance and avoid single points of failure.

The requirements of scalability and reliability have fostered the success of application server, which its flow of requests and responses can be seen in Fig 2. Technically speaking, an application server is a software platform, distinct from Web server, dedicated to the efficient execution of process components for supporting the construction of dynamic pages.

The client request (1), formatted in HTTP, is received by the Web server, which transform it into a request to the scripting engine (2). The scripting engine executes the program associated with the requested URL, which may include calls to process components hosted in application server (3). The components managed by the application server dispatch the query to the data source (4) using data-mining model, collect the query result (5), possibly elaborate it and hand it back to the scripting engine (6). Query results are integrated into the HTTP response by the scripting engine, to obtain the result HTML page (7), routed by the HTTP server to the client (8).
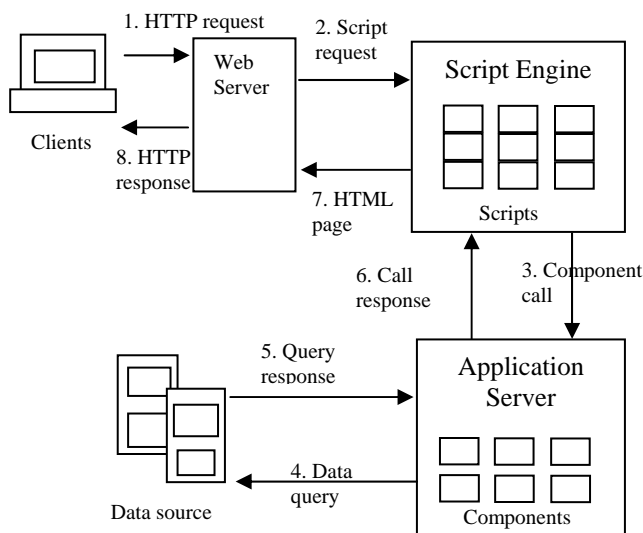

Fig. 2. Application Server Architectures Flow of Requests and Responses

The main purpose of the application server is to provide a feature-rich execution environment for the process components, which facilitates the construction of scalable and reliable applications. This is done by the deep integration at the application server [9]. This execution environment, often called a *managed runtime environment* includes the following services [8]:

- Transparent component functioning;
- Failure recovery;
- Transaction management;
- Resource pooling;
- Interoperability;
- Multi-protocol

A. Transparent Component Functioning

The transparent component functioning includes component distribution, replication and load balancing. The process objects programmed by user are installed into the manage runtime environment, which may distributed on multiple processes and physical machines. The application server automatically manages the creation of processes, replication of objects and the allocation to the available processes (load balancing) in such a way that are totally transparent to the calling client, which can behave as if interacting with a single instance of a process object.

B. Failure Recovery

The application server monitors the active hosts, process objects, detect hardware, software and network failure, and automatically avert client requests addressed to a failed component and route them a available replicas of the same object.

C. Transaction Management

The application server provides the capability of defining units of work (called *transactions*), which are either executed successfully from start to end, or rolled back completely in case of failure of any of the included operations. Transactions are typically offered by database management systems for sequences of database update operations.

D. Resource Pooling

The application server handles pools of expensive resources, like database connections, and shares these resources among multiple process objects in an optimized way.

E. Interoperability

The application server is equipped with predefined gateways or software development kits for exchanging messages and data with applications developed on obsolete platforms or with surpassed technologies (legacy applications).

F. Multi-protocol

The application server integrates multiple application distribution protocols and programming languages into a uniform development environment, and facilitates cross-

platform application development and migration.

IV. ENERGY AUDIT METHODOLOGY

Decision makers or utility managers in companies will often need to conduct or commission energy survey to identify energy efficiency opportunities in their buildings. They need a useful energy audit to assist them in the energy decision-making process. The main considerations include operational energy level needs, hours of operation, task and general energy and consumption over time (annual, seasonal and monthly). The overview of energy audit methodology is as shown in Fig 3.
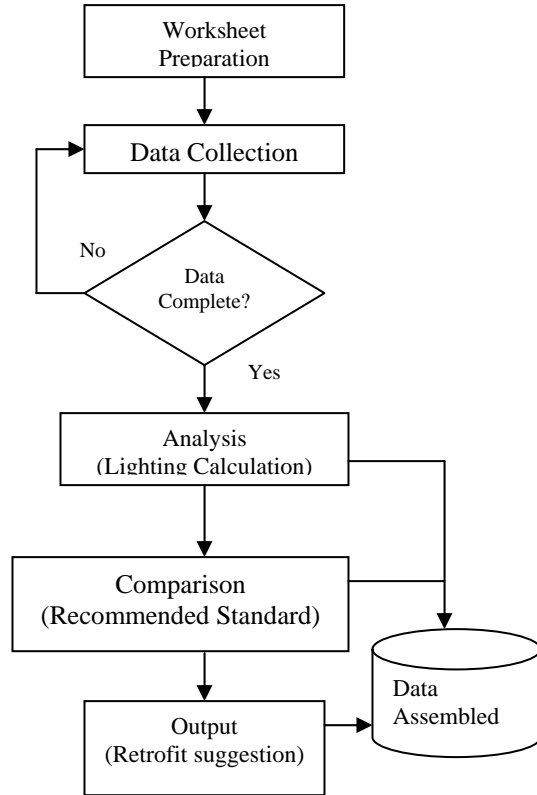

Fig. 3. Flow Chart of Energy Audit

The first step is to do the worksheet preparation for energy audit purpose. Auditor needs to learn about the existing building system by simply walks through the building and will take the necessary considerations into account. The good system study will produce a good worksheet for the data collection purposes. Only the complete data set is used for data analysis process. The analysis covers the energy calculation. The comparison with the standard values and retrofit suggestions are following the analysis process, while all the data assembled are stored in the databases.

In combining the energy audit method and the web application-based data mining process model, the proposed model will use the assembled data from Fig 3 to be the operational database with the web-server architecture of Fig 2 and the data warehouse from Fig 1. The following section lists this proposed web-based data-mining model for building energy audit in a case study example.

V. CASE STUDY EXAMPLE

The case study is done in buildings at Faculty of Electrical Engineering, Kolej Universiti Teknikal Kebangsaan Malaysia (KUTKM). The buildings includes of Administration block, Lecturer room, Lecture hall and Laboratory.

A. Worksheet Preparation

TABLE I
BUILDING DIVISION

No	Building	Levels	Rooms
1.	Administration Block	AB1	20
2.	Lecturer Room	LR1	11
		LR2	13
		LR3	12
3.	Lecture Hall and Laboratory	LH1	6
		LB2	3
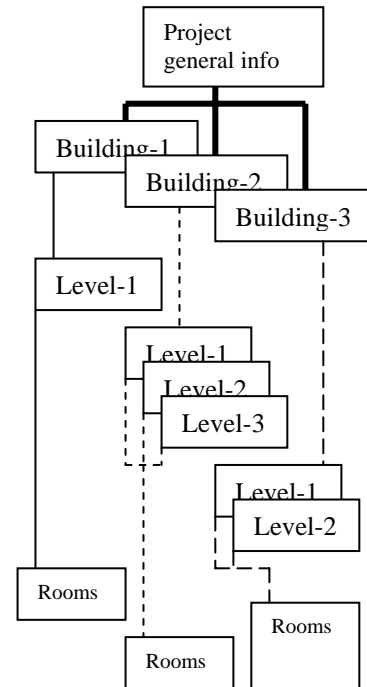
Fig. 4. Database Link Structure

TABLE II
PARAMETERS FOR DATABASE

No	Data Items	Parameters
1.	General info	Person in charge name Address No of buildings
2.	Building	Building name No of level Utility info
3.	Level	Level name No of room
4.	Room	Room name Room type Room dimension Window size Working table height Lux measured Temperature measured No of light fitting No of lamp/light fitting Lamp type No-of-air-con unit Air-con type

The worksheet is prepared by taking consideration of the project general info and building info. The prepared worksheet is used to generate the database structure. Further, the buildings are subdivided into levels and rooms. The Administration block, Lecturer room, Lecturer hall and Laboratory are categorized as separate building. Table I shows the building division. Fig 4 shows the database link structure. Table II shows the parameters collected for general info, building, level, and room.

B. Data Collection

Data collection is based on the parameters needed for the analysis part of lighting audit, as an example [7]. It includes the characteristic of activity in the room, the dimension of the room, light level readings in lux, number of luminaries, number of lamp per luminaries, lamps fixture and the lamp wattage, which listed in Table II. Color of the wall of each room was also recorded for analysis purposes.

Some other parameters must be known prior to the calculations as shown in Table III: The room index can be calculated using equation (1).

$$RI = \frac{l \times w}{h(1 + w)} \qquad (1)$$

Where: l = length of room; w = width of room; h = mounting height, i.e. the vertical distance between the working plane and the luminaries.

TABLE III
PARAMETERS FOR LUMEN METHOD

No.	Parameters	Description
1.	Room index	Use equation (5)
2.	Room reflectance	Depends on the brightness of the wall, ranging from 0 to 1
3.	Utilization factor	Given in the manufacturers data sheet, ranging from 0 to 1
4.	Light loss factor	Check CIBSE "Code for Lighting" [4]

C. Lumen Method Calculations

The Malaysian Standard [3] lists three most common used methods for calculating illumination from an electrical source. Among the three, *lumen method* suits the best into this lighting audit scenario. The lumen method is applicable to a large number of real-life situations, such as offices, laboratory and classrooms [7].

After the data collection process has been done, lumen method formula is used to determine the optimum number of luminaries in each room in the form described in equation (2).

$$N = \frac{E_m \times A}{UF \times MF \times IL} \qquad (2)$$

Where: N = number of luminaries; E_m = average illuminance over the horizontal working plane in (lux); A = area of the horizontal working plane; UF = utilization factor for the

horizontal working plane; MF = maintenance factor; and IL = initial bare lamp luminous (lumen).

D. Comparison Process

The lumen method calculations give the recommended numbers of luminaries in the room. The number is then compared with the existing system's luminaries. Further, the installed load efficacy ratio (ILER) can be determined using equation (3). Table IV shows the load efficacy ratio assessment. Based on the ILER ratio, proper retrofit action can be taken to maximize the energy usage.

$$ILER = \frac{\text{installed or actual } lux/w/m^2}{\text{target } lux/w/m^2}$$
$$= \frac{average\_lux\_measured}{standard\_illuminance} \qquad (3)$$

Where standard illuminance is taken from lighting manufacturer's handbook.

TABLE IV
LOAD EFFICACY RATIO ASSESSMENT

ILER	Assessment
0.75 above	Satisfactory to good
0.51-0.74	Review suggested
Below 0.5	Urgent action

The calculated lux for an existing room is derived from equation (2) to equation (4).

$$LCal = \frac{LN \times UF \times MF \times IL}{A} \qquad (4)$$

Where LCal is the lux calculated (lux); and LN is number of lamp in the room.

E. Suggestions Output

TABLE V
SUGGESTION LOGIC SCHEME; NOTE: N/A INDICATES NOT APPLICABLE OF CHECKING MEANING VALID FOR THE WHOLE RANGE OF ITS VALUE

P	ILER	Suggestions
When Lamp required < Existing lamp		
> 0.25	N/A	Lux too high Reduce tube
<= 0.25	> 1.25	Turn of lights Reduce tube
<= 0.25	0.75 – 1.24	Within specs
<= 0.25	0.51 – 0.74	Install reflector Clean tube
<= 0.25	< 0.5	Add tube
When Lamp required > Existing lamp		
< -0.25	N/A	Lux too low Add tube
>= -0.25	> 1.25	Turn of lights Reduce tube
>= -0.25	0.75 – 1.24	Within specs
>= -0.25	0.51 – 0.74	Install reflector Clean tube
>= -0.25	< 0.5	Add tube
When Lamp required = Existing lamp		
N/A	> 1.25	Turn of lights Reduce tube
N/A	0.75 – 1.24	Within specs
N/A	0.51 – 0.74	Install reflector Clean tube
N/A	< 0.5	Add tube

The difference between calculated lux and the standard lux as in equation (5), together with ILER are used for giving suggestions as listed in Table V.

$$P = \frac{\text{LCal} - standard\_illuminance}{standard\_illuminance} \quad (5)$$

In giving the suggestions, more aspects from the experts or decision makers can be added to extend the detail of the analysis, but the example that been listed in Table V is selected for the purpose of this case study example only.

F. Data Assembling & Mining

The subset of results as listed in Table VI shows the data assembled from the analysis process using suggestion logic scheme from Table V.

TABLE VI
SUBSET OF RESULTS; NOTE: RM IS ROOM NUMBER, LR IS LAMP REQUIRED, LE IS LAMP EXISTING, P IS FROM EQUATION (8), AND ILER IS FROM EQUATION (6)

Rm	LR/LE	P	ILER	Assessment
1	Lo	Ok	Hi	Satisfactory
2	Lo	Lo	Lo	Urgent action
3	Hi	Hi	Hi	Satisfactory
4	Hi	Hi	Lo	Urgent action
5	Ok	Hi	Ok	Review suggested

Note: LR/LE has values Lo (< 1), Hi (>1) and Ok (=1); P has values Lo (<-0.25), Hi (>0.25) and Ok (-0.25 – 0.25); and ILER has values Lo (< 0.5), Hi (> 1.25) and Ok (0.75 – 1.24).

TABLE VII
OTHER SUBSET OF ANALYSIS DATA WITH THEIR AUTOMATIC ASSESSMENT

Rm	LR/LE	P	ILER	Assessment
6	Ok	Ok	Ok	Review suggested
7	Lo	Hi	Lo	Urgent action
8	Hi	Lo	Hi	Satisfactory

The data mining algorithm uses assessment results as in Table VI and can concludes that ILER's value plays an direct impacts to the assessment results, therefore if other set of data as in Table VII come from the analysis, the assessment result can be given automatically.

The failure of the data-mining algorithm in giving the assessment for room 7 and 8 in Table VII is due to unavailable data set for training that does not have Lo-Hi and Hi-Lo combination for LR/LE-P set of data. This failure can be corrected by adding the dataset as in Table VIII for training the data-mining algorithm.

TABLE VIII
SUBSET OF ANALYSIS DATA WITH THEIR CORRECT ASSESSMENT

Rm	LR/LE	P	ILER	Assessment
6	Ok	Ok	Ok	Review suggested
7	Lo	Hi	Lo	Lux too high
8	Hi	Lo	Hi	Lux too low

G. Web Application Data Mining

The deep integration at the application server focuses on data exchange services, in an XML format. It requires to develop website (portal) from a user-centric perspective. These five common parameters (themes) are being checked:

- Focus on the process objective;
- Emphasize on easy of use;
- Deep integration of applications;
- Scalability of services;
- Well-develop security models.

VI. CONCLUSION

The work has addressed the data mining web application in energy audit of an industrial building with a case example of lighting energy audit. The method can be used for minimizing the time in analyzing a large amount of analysis data. The work has shown the new combined energy audit method with the four-step method in web-based data mining algorithm using application server architecture. After tested with the case study example, the method has shown its capability to give the correct assessment using the existing calculated data. Further work extensions in calculating the savings measures and in other type of energy audit in buildings can be extended to gain the maximum benefits with more emphasize in data security model.

VII. REFERENCES

[1] Schiler, *Simplified Design of Building Lighting*, John Wiley & Sons, New York, USA, 1992.
[2] Roiger, R. J.., and Geatz, M. W., *Data Mining – A Tutorial-based Primer*, Addison Wesley, Boston, 2003.
[3] MS 603 : 1979, *Code of Practice for interior lighting*, SIRIM, Malaysia, 1996.
[4] CIBSE, *Code for Lighting*, United Kingdom, 1996.
[5] Merril, D.M., and Tennyson, R.D., *Teaching Concepts: And Instructional Design Guide*, Englewood Cliffs, NJ Educational Technology Publications, 1977.
[6] Ackerman, W.J., and Block, W.R., "Understanding Supervisory Systems." IEEE Computer Applications in Power, October 1992, pp. 37-40.
[7] NM Maricar, CK Gan, and MN Jamal, "Data Mining Application in Industrial Energy Audit for Lighting," IASTED Europe Energy and Power Systems, Spain, June 2005.
[8] Ceri, S., et al., *Designing Data-Intensive Web Applications*, Morgan Kaufmann, USA, 2003.
[9] Sullivan, D., *Proven Portals – Best Practices for Planning, Designing, and Developing Enterprise Portals*, Addison-Wesley, USA, 2004.

VIII. BIOGRAPHIES

Noor M Maricar (IEEE S'87, M'03) is an assistant professor at Hail Community College, Applied Electrical Engineering Department, King Fahd University for Petroleum and Minerals, Saudi Arabia. He earned his PhD from Virginia Polytechnic Institute and State University, Alexandria - Virginia, USA in Electrical Engineering, 2004. He received his MS. in Electrical Engineering from Illinois Institute of Technology, Chicago, USA and MSc. In Computer Integrated Engineering from Nanyang Technological University, Singapore, in 1998 and 1997, respectively. His interest research areas are renewable technologies, power system planning, data support system and energy conservation.