BOILER EFFICIENCY EVALUATION AND MONITORING THROUGH DATA MINING TECHNIQUES IN TEXTILE INDUSTRY

[1]S.K. TYAGI, [2]Dr. B. K. SHARMA, [3]SANJEEV SAXENA

[1]Professor , Department of CSE, Lord Krishna College of Engineering Ghaziabad (India), Ph.D. Scholar – Singhania University,

[2] Sr. Scientific Officer & Head, Software Development Centre, Northern India Textile Research Association, Ghaziabad (India)

[3] Sr. Scientific Officer & Head, Engineering Division, Northern India Textile Research Association, Ghaziabad (India)

ABSTRACT

Cost of energy is the highest among all other components in the production of textile products. So energy efficiency analysis and control has assumed paramount importance. Boiler consumes measure chunk of fuel in any processing industry. Efficiency of any Boiler depends upon minimization of various indirect losses of the boiler so that amount of energy input in the boiler by burning the fuel can be maximum utilized for generation of steam and cost of steam can be minimized ultimately. The proposed data mining technique can prove to be a very effective tool for evaluating and maintain cluster wise boiler efficiency and indirect losses. Also it is very helpful to meet the objective of energy conservation and fuel saving by curbing the losses with the help of various check points. Cluster wise evaluation of boiler efficiency is a major highlight of data mining technique. Effective use of data mining will ear mark various areas where energy and there by precious fuel can be saved.

Keywords: *Data Mining,* Clustering, Knowledge Discovery, Information Technology & Knowledge Management, Boiler *Efficiency, Indirect losses, Steam, Software, Heat, and Costing*

1. INTRODUCTION

In textile processing, major emphasis is usually laid upon productivity and quality, whereas energy is considered as a second priority. However, due to the alarming increase in energy cost, every effort should be given to minimize it. As we are aware that Boiler is main source of fuel consumption in any textile and non-textile industry. It is a matter of great concern to all that boiler should run at its maximum efficiency with minimum indirect losses. For attaining maximum boiler efficiency, exact assessment of boiler efficiency and all indirect losses are very important. Boiler consumes measure chunk of fuel in any processing industry. Efficiency of any Boiler depends upon minimization of various indirect losses of the boiler so that amount of energy input in the boiler by burning the fuel can be maximum utilized for generation of steam and cost of steam can be minimized ultimately. The direct efficiency of the boiler is based on fuel consumption and steam generation for a particular time period as per standards. The following indirect losses can be minimized for efficient boiler efficiency.

- Dry Flue Gas Loss
- Fuel Moisture Loss
- Blow Down Losses
- Incomplete Combustion Loss
- Air Moisture Loss
- Radiation and Convection Loss

After knowing the various heat losses it is possible to take action to improve boiler efficiency. A model report format of the boiler efficiency is shown in Table-1 where all the input details can be fed to computer like fuel analysis, calorific value, steam pressure, enthalpy, %CO_2, TDS etc. Subsequently the boiler efficiency and the indirect losses are calculated and displayed. Graphical representation of Boiler efficiency can be displayed as shown in Fig.1.

Table 1 : Model Report Format For Boiler Efficiency & Indirect Losses System

CODE: AL1 **DATE**: 19/12/11 12:00 AM **FUEL**: OIL

TYPE OF OIL : FO	
CARBON (C)	84.00
HYDROGEN (H2)	11.00
SULPHUR (S)	3.50
WATER (H2O)	1.00
ASH	0.50
GCV	10200.00
WATER CONSUMED PER HOUR	1325.00
COAL CONSUMED	127.00
STEAM PRESSURE (1 TO 15)	7.50
STEAM ENTHALPY	660.20
CARBON DIOXIDE (CO2)	10.00
TDS (FEED WATER)	1300.00
FEED WATER TEMPERATURE	50.00
FEED WATER ENTHALPY	50.00
FLUE GAS TEMPERATURE	170.00
ATMOSPHERIC TEMPERATURE	30.00
CARBON MONOXIDE	0.00
CARBON CONTENT (Kg/Kg Fuel)	0.00
OXYGEN	0.00
BLOW DOWN WATER TDS	0.00
WETBULB TEMPERATURE	0.00
DRYBULB TEMPERATURE	0.00
MOISTURE IN AIR (Kg/Kg Fuel)	0.00
BOILER EFFICIENCY	**69.35**
INDIRECT LOSSES	**30.65**
(I) RY FLUE GAS LOSS	7.42
(II) FUEL MOISTURE LOSS	6.31
(III) BLOW DOWN LOSSES	14.77
(IV) INCOMPLETE COMBUSTION LOSS	0.00
(V) AIR MOISTURE LOSS	0.00
(VI) RADIATION & CONVECTION LOSS	2.15

Similarly, we can generate boiler efficiency of any type of fuel.

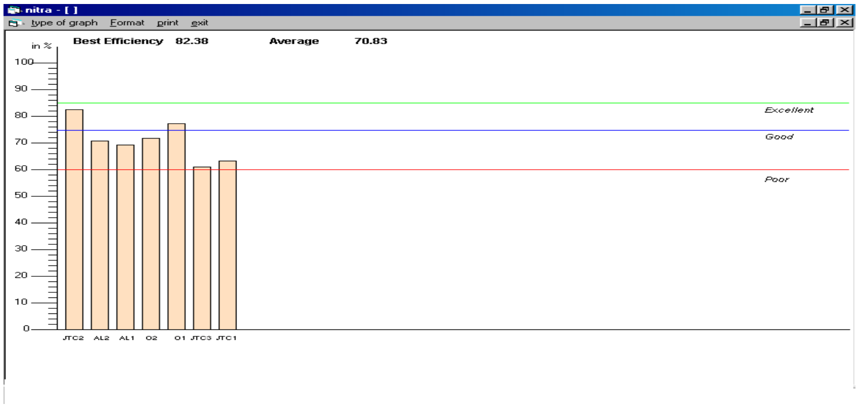

Fig 1 : Graphical Presentation of Boiler Efficiency

2. ELEMENT AND USES OF DATA MINING

Most textile process houses today have some sort of IT based application to log and manage data. The use of IT application like SQL query, Excel sheet and customize software are providing the textile energy and utility related information i.e. power consumption, down time, energy conservation.

Every progressive organization has the only major objective to increase its profitability. However, in today's competitive marketplaces, profitability is not only depends on increasing sales but also just as importantly on reducing cost and improving the quality. Introducing Data Mining Tools and Techniques into textile production processes can achieve a substantial increase in energy efficiency and productivity. It is here that Data Mining plays a vital role.

Data Mining can be defined as a technique for extracting the "meaning" contained in information to allow the understanding needed by a user to make a "right" decision. Another definition could be providing the right information, in the right form, at the right time, so as to enable the manager to efficiently and effectively perform his/her job. It is Data Mining that allows a computer to digest the constant stream of data being generated by the computerized sensors and monitors of the plant, and then extract from that information that has some meaning content. Data mining tools and techniques can be used for rationalizing the data so as to reduce the overload that tends to occur and make it simple for the personnel to make a right decision in textile industry.

Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. The more interesting way to use a data mining model is to get the user to actually understand what is going on so that they can take action directly. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining consists of five major elements:

1. Extract, transform, and load transaction data onto the data warehouse system
2. Store and manage the data in a multidimensional database system.
3. Provided data access to business analysts and IT professionals
4. Analyze the data by application software.
5. Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

Artificial neural networks: non-linear predictive models that learn through training and resemble biological neural networks in structure.

Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and selection in a design based on the concepts of natural evolution.

Decision trees: tree- shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decisions tree methods include Classification and regression trees (CART) and chi Square Automatic interaction Detection (CHAID). CART and CHAID are decisions tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2- way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.

Nearest neighbor method: A technique that classifies each record in a data set based on a combination of the classes of the k record (s) most similar to it in a historical dataset (when k = 1). Sometimes called the k-nearest neighbors technique.

Rule induction: The extraction of useful if-then rules from data based on statistical significance. Data visualization: The extraction of useful if-then rules from data based on statistical significance.

Data visualization: the visual interpretation of complex relationship in multidimensional data Graphics are used to illustrate data relationships. The point of data visualization is to let the user understand what is going on. Since data mining usually involves extracting "hidden" information from a database, this understanding process can get somewhat complicated. In most standard database operations nearly everything the user sees is something that they knew existed in the database already. A report showing the breakdown of sales by product and region is straightforward for the user to understand because they intuitively know that this kind of information already exists in the database. If the company sells different products in different regions of the county, there is no problem translating a display of this information into a relevant understanding of the business process. The primary benefit of data mining is the ability to turn feeling into facts. Data mining can be used to support or refute feelings of people have about how businesses is going. It can be used to add credibility to these feelings and warrant dedication of more resource and time to the most productive areas of a company's operations. This benefit deals with situations where a company starts the data mining process with an idea of what they are looking for. This is called targeted data mining. Data mining can discover unexpected patterns in behavior, patterns that were not under consideration when the mining exercise commenced. This is called out of the blue data mining.

In existing, some modern textile industry are maintaining and analyzing the MIS reports by SQL query and data mining software. Some vendors of data mining software are given in Table-2.

COMMERCIAL	OPEN SOURCE
1. ORACLE 2. MICROSOFT 3. IBM 4. BUSINESS OBJECTS 5. MICROSTRATEGY 6. SAS 7. TERADATA 8. SAP	1. PENTAHO 2. JASPERSOFT 3. ACTUATE 4. COGNOS

Table-2: Vendors of Data Mining Software

3. USE OF DATA MINING APPLICATION IN ENERGY EFFICIENCY

In a textile process house, the utility department checks boiler efficiency and indirect losses system on regular basis. The details of boiler efficiency and indirect losses system for ten days have been provided in Table-3.

Day	Code ID	Boiler Efficiency	Indirect Losses System						
			A	**B**	**C**	**D**	**E**	**F**	**Total**
1	A1	**69.35**	7.42	6.31	14.77	0.00	0.00	2.15	**30.65**
2	A2	**63.33**	10.36	6.67	6.25	0.42	9.89	3.08	**36.67**
3	A3	**61.00**	19.88	6.88	7.41	0.77	2.06	2.00	**39.00**
4	A4	**81.80**	6.75	6.25	2.34	0.21	0.64	2.01	**18.20**
5	A5	**77.32**	8.19	6.39	4.17	1.65	0.78	1.50	**22.68**
6	A6	**72.78**	9.20	5.12	4.10	4.34	2.23	2.23	**27.22**
7	A7	**71.68**	10.34	6.55	6.25	1.78	0.90	2.50	**28.32**
8	A8	**70.74**	5.72	6.26	14.78	0.00	0.00	2.50	**29.26**
9	A9	**64.34**	10.23	6.34	6.16	0.38	9.53	3.02	**35.66**
10	A10	**62.17**	18.50	6.67	7.32	0.34	2.01	2.99	**37.83**

Table 3: Boiler Efficiency and Indirect Losses System

Where;

A = Dry Flue Gas Loss
B = Fuel Moisture Loss
C = Blow Down Losses
D = Incomplete Combustion Loss
E = Air Moisture Loss
F = Radiation and Convection Loss

Let the three seeds be the first three boiler efficiency particulars shown in Table-4.

Code ID	Boiler Efficiency	Indirect Losses System						
		A	**B**	**C**	**D**	**E**	**F**	**Total**
A1	**69.35**	7.42	6.31	14.77	0.00	0.00	2.15	**30.65**
A2	**63.33**	10.36	6.67	6.25	0.42	9.89	3.08	**36.67**
A3	**61.00**	19.88	6.88	7.41	0.77	2.06	2.00	**39.00**

Table 4: The three seeds for boiler efficiency

Now compute the distance using seven attributes and using the sum of absolute difference for simplicity (i.e. using the K-median method). The distance values for all the objects are given in Table-5, wherein column 9,10 & 11 give the three distance from the three seeds respectively. Based on these distance, each efficiency is allocated to the nearest cluster. We obtain the first iteration result as shown in Table-5.

	Boiler Efficiency	Indirect Losses System						Distance from Cluster			Allocation to the Nearest Cluster
C1 C2 C3	**69.35** **63.33** **61.00**	7.42 10.36 19.88	6.31 6.67 6.88	14.77 6.25 7.41	0.00 0.42 0.77	0.00 9.89 2.06	2.15 3.08 2.00	From C1	From C2	From C3	
A1	**69.35**	7.42	6.31	14.77	0.00	0.00	2.15	0.00	29.08	31.72	C1
A2	**63.33**	10.36	6.67	6.25	0.42	9.89	3.08	29.08	0.00	22.48	C2
A3	**61.00**	19.88	6.88	7.41	0.77	2.06	2.00	31.72	22.48	0.00	C3
A4	**81.80**	6.75	6.25	2.34	0.21	0.64	2.01	26.60	36.94	41.62	C1
A5	**77.32**	8.19	6.39	4.17	1.65	0.78	1.50	22.50	30.44	34.40	C1
A6	**72.78**	9.20	5.12	4.10	4.34	2.23	2.23	23.72	26.74	31.50	C1
A7	**71.68**	10.34	6.55	6.25	1.78	0.90	2.50	17.04	19.42	24.38	C1
A8	**70.74**	5.72	6.26	14.78	0.00	0.00	2.50	3.50	31.88	35.22	C1
A9	**64.34**	10.23	6.34	6.16	0.38	9.53	3.02	27.24	2.02	23.66	C2
A10	**62.17**	18.50	6.67	7.32	0.34	2.01	2.99	29.26	18.42	4.32	C3

Table 5: First Iteration allocating each object to the nearest cluster

The first iterates leads to six code id of boiler efficiency in the first cluster and two each in the second and third clusters. Table-6 compare the cluster means of cluster found in Table-5 with the original seeds.

	Boiler Efficiency	Indirect Losses System						
		A	B	C	D	E	F	Total
C1	**73.95**	7.94	6.15	7.74	1.33	0.76	2.15	26.05
C2	**63.84**	10.30	6.51	6.21	0.40	9.71	3.05	36.16
C3	**61.59**	19.19	6.78	7.37	0.56	2.04	2.50	38.41
Seed1	**69.35**	7.42	6.31	14.77	0.00	0.00	2.15	**30.65**
Seed2	**63.33**	10.36	6.67	6.25	0.42	9.89	3.08	**36.67**
Seed3	**61.00**	19.88	6.88	7.41	0.77	2.06	2.00	**39.00**

Table 6: Comparing new cenroids and the seeds

Use the new cluster means to again compute the distance of each of the mean, again allocating each object to the nearest cluster. Table-7 shows the second iteration result.

	Boiler Efficiency	Indirect Losses System						Distance from Cluster			Allocation to the Nearest Cluster
C1 C2 C3	**73.95** **63.84** **61.59**	7.94 10.30 19.19	6.15 6.51 6.78	7.74 6.21 7.37	1.33 0.40 0.56	0.76 9.71 2.04	2.15 3.05 2.50	From C1	From C2	From C3	
A1	**69.35**	7.42	6.31	14.77	0.00	0.00	2.15	14.40	28.16	30.35	C1
A2	**63.33**	10.36	6.67	6.25	0.42	9.89	3.08	26.02	1.00	20.37	C2
A3	**61.00**	19.88	6.88	7.41	0.77	2.06	2.00	27.96	23.06	2.15	C3
A4	**81.80**	6.75	6.25	2.34	0.21	0.64	2.01	15.92	35.94	40.45	C1
A5	**77.32**	8.19	6.39	4.17	1.65	0.78	1.50	8.42	29.48	33.67	C1
A6	**72.78**	9.20	5.12	4.10	4.34	2.23	2.23	11.66	25.78	30.35	C1
A7	**71.68**	10.34	6.55	6.25	1.78	0.90	2.50	7.50	18.70	22.65	C1
A8	**70.74**	5.72	6.26	14.78	0.00	0.00	2.50	15.02	30.96	33.15	C1
A9	**64.34**	10.23	6.34	6.16	0.38	9.53	3.02	24.26	1.02	21.55	C2
A10	**62.17**	18.50	6.67	7.32	0.34	2.01	2.99	26.36	18.96	2.17	C3

Table 7: Second Iteration allocating each object to the nearest cluster

The number of code id of boiler efficiency in cluster 1 is again 6 and other two clusters still have two code id of boiler efficiency each. A more careful look shows that the clusters have not changed at all. Therefore the method has converged rather quickly for this very simple dataset.

The cluster membership is as follows;

Cluster 1 – A1, A4, A5, A6, A6, A7
Cluster 2 – A2, A9
Cluster 3 – A3, A10

The consistency of boiler efficiency in each cluster can be evaluated by the statistical Co-efficient of variation method. The comparison of dispersion for above mentioned distribution can be made by calculating Co-efficient of variation. The detail description is as follows:

CV% = (Standard Deviation/ Mean) * 100

The values of CV% for three clusters are as follows:

Cluster 1 – 5.8158
Cluster 2 – 0.7911
Cluster 3 – 0.9499

Since coefficient of variation is less for cluster-2 as compare to cluster-3 and cluster-1, hence boiler efficiency of cluster-2 is more consistence as compare to cluster3 and cluster1. by above method of cluster wise data mining technique, the boiler efficiency can be improved by consistency of data, Evaluation of indirect losses and incorporating various check points to curb the losses In a more scientific manner.

4. CONCLUSION

In the present scenario, the energy cost is the highest among all other cost component in the production of textile goods. So far there have not been much efforts to evolve a systematic approach to analyze and monitor energy consumption in each process and translate into financial terms. The proposed case study for analysis of boiler efficiency and indirect losses system are a step forward towards those objectives. By proper use of this data mining technique it is possible to analyze cluster wise boiler efficiency and also consistency of boiler efficiency can also be maintained within a cluster. This may serve as an important MIS tool for the management to exercise effective energy conservation and cost control measures. In order to compete with international products, there is no other alternative but to go for automation in near future. This approach may act as a precursor to that.

5. ACKNOWLEDGEMENT

Authors express sincere thanks to the management of Sheena Exports and National Woolen, Panipat, for their help and cooperation in taking practical field trials.

REFERENCES

[1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. Verkamo, *Fast Discovery of Association Rules, in Advances in Knowledge and Data Mining*, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy Eds, MIT Press, 1996.
[2] H. Toivonen. *Sampling large databases for association rules*. Proc. of the Int'l Conf. on Very Large Data Bases (VLDB), 1996.
[3] R. Agrawal and R. Srikant. *Fast algorithms for mining association rules*. In Proceedings of the 20th VLDB Conference, Santiago, Chile, 1994.
[4] R. Agrawal, T. Imilienski, and A. Swami. *Mining Association Rules between Sets of Items in Large Databases*, Proc. Of the ACM SIGMOD int'l Conf. on Management of Data, pages 207-216, May 1993.
[5] R. Agrawal, T. Imilienski, and A. Swami. *Data base Mining: A Performance Perspective*, IEEE Transactions on Knowledge and Data Engineering, 5(6):914-925, December 1993.
[6] Patricia L. Carbone, Expanding the Meaning of and Applications for Data Mining 2000 IEEE
[7] Dr. B.K. Sharma, D.K. Sharma, Application of Information Technology in Textile Wet Processing for Strategic Decision Making, International Journal of Management and System, Australia
[8] Prof. S.M. Ishtiaque, Dr. B.K. Sharma, D.K. Sharma, Management Information System in Textile Wet Processing for batter decision-making, International Conference on IT in Textile Sector organized by Ministry of Textile, Govt. of India
[9] Dr. B.K. Sharma & Dr. A. Das, Application of information Technology to manage the textile quality control text data, The Textile Industry and Trade Journal, Vol. No. 44, No. 5-6, May-June 2006
[10] Soler, S. and D. Yankelevich (2001). "Quality Mining: A data Mining Based Method for Data Quality Evaluation", Processing of the Sixth international Conference on Data Quality, MIT.
[11] Dongsong Zhang and Lina Zhou Discovering Goden Nuggets: Data Mining in Financial Application 2004 IEEE.
[12] Dr. B.K. Sharma, Prof. S.K.B & Abhay Bansal, Data Mining Tools and Techniques in Textile Industry for Effective Decision Making and Corrective Action, Asian Textile Journal, Vol No. 15, No. 8, August 2006
[13] C. Brunk, J. Kelly, and R. Kohavi, "MineSet: An Integrated System for Data Access, Visual Data Mining, and Analytical Data Mining," Proceedings of the Third Conference on Knowledge Discovery and Data Mining (KDD-97), Newport Beach, CA, August 1997.
[14] G.K. Gupta, A book on introduction to Data Mining with Case Studies.
[15] Prof. S.K. Tyagi & Dr. B.K. Sharma "Evaluation of energy efficiency, monitoring and improvement through data mining tools and techniques for oil conservation in textile industry, International Journal of computer sciences software engineering and electrical communication.
[16] D.K. Bhattacharya, Dr. B.K. Sharma & Sanjeev Saxena "Energy balance & accounting for boiler and process house in textile industry through software approach at 43rd Joint technology conference, IIT Delhi on 2-3rd march 2002.
[17] Sanjeev Saxena & Dr. B.K. Sharma " Evaluation of Energy Efficiency, Monitoring and improvement through software for oil conservation at 5th international petroleum conference (PETROTACH-2003) organized by ministry of petroleum and natural gas on 9-12 january,2003at vigan bhawan, New Delhi