

# Data Mining for Engineering Sector in Pakistan: Issues and Implications

Muhammad Shahbaz<sup>\*</sup>, Nadeem ur Rahman<sup>±</sup>

**Abstract** - In recent years data mining has become a popular technique for extracting information from the databases of many types in different areas of commerce and industry. The strength of data mining comes from its flexibility of working on any kind of database and its ability to discover previously unknown and sometimes surprising results. Competitiveness increasingly depends on improving the quality of decision making from past information. Improved knowledge of engineering capabilities and products enables engineers to better target future production strategies. To move from quality control to quality assurance and reduced error occurrence, companies need to exploit their existing knowledge and previous experiences more effectively. Data mining analysis offers many potential benefits in this context. This paper briefly highlights the benefits that can be gained through the application of data mining technology in different sectors of Pakistan in general and in engineering sector in particular

**Index Terms:** Data Mining, Artificial Analysis, Databases, Manufacturing Engineering, Production

## I. INTRODUCTION

The progress in data acquisition and successful development of storage technology at cheaper rates, along with limited human capabilities in analyzing and understanding big databases have tempted scientists and researchers to move forward towards the specific field of knowledge discovery in databases (KDD). This recently emerged discipline, is the synergy of data management, artificial intelligence, machine learning and statistics. Data Mining is the search for valuable knowledge in large volumes of data [1]. It is a through potential efforts of human being using computational techniques. Humans design systems and databases, identify problems and set targets & goals; computers explore the data, looking for relationships and patterns that match these goals.

<sup>\*</sup>Dr. Muhammad Shahbaz: Department of Computer Science & Engineering, University of Engineering & Technology, Lahore, Pakistan  
[M.Shahbaz@uet.edu.pk](mailto:M.Shahbaz@uet.edu.pk)

<sup>±</sup>Nadeem ur Rahman: Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University  
UK, LE11 3TU  
[N.ur-Rehman@lboro.ac.uk](mailto:N.ur-Rehman@lboro.ac.uk)

The central step within the overall KDD process is data mining, the application of artificial intelligence algorithms in the task of finding patterns and models in data [2]. The major areas enjoying the benefits of KDD include banking, finance, business and medical sciences. Many companies, including engineering and manufacturing enterprises all over the world, are now giving attention to the utilization of KDD technology for the improvement in their current status.

To date, KDD techniques have not commonly been used in Pakistani industries in general and in engineering sector in particular. The reasons for this are not certain, but it may be because the long time scales and expenses involved in introducing new techniques in this area and specially also because of lack of awareness of the benefits offered by this new technology. An alternative possible reason might be the complexity and diversity of different engineering processes as these make it very difficult to devise a generic data mining model that can be used for all kinds of engineering processes and can handle all types of engineering problems. Developing countries like Pakistan have another significant problem of lacking with automated data recording and storages devices within their equipments and system. Few engineering sectors have some data recording practices but in a very rough form and format which hardly can help to predict trends and behaviours of manufacturing product design and optimizing process. This data is recorded manually having exceptionally high number of a missing and/or unrealistic values and show quite unexpected or abnormal behaviour during the analysis of data.

In this paper we will however discuss that how engineering product, process and services can be improved and how data warehousing can help in data analysis and knowledge extraction especially in the perspective of any developing countries like Pakistan. In the end potential application of data mining technology in resolving different industrial problem has been discussed. These areas are selected randomly keeping into the needs and requirements of industrial setup in Pakistan which has been through an era of great need to revolutionise to meet the international standards of both manufacturing and services sectors.

This paper has been concluded with remarks that there is a great need to built a relationship between industry and

academic activities to vitalize the activities that has been carried out in the industrial sector of Pakistan.

## II. PROCESS AND SERVICES IMPROVEMENT STRATEGIES

Competitive improvement can be achieved in many ways, for example by improving the quality of products or by reducing the material waste, production and overhead costs, or by reducing time to launch new or improved product. Data Mining can support these improvements, through the extraction of knowledge from either existing data warehouses, or from current production data. Applying this knowledge can help to improve the quality of products by better controlling the engineering processes and methodologies, and by keeping product and production parameters in range [3].

Competitiveness increasingly depends on the quality of decisions and improving it through information and experience [4]. Improved knowledge of production capabilities and products enables engineers to better understand their targets and design their future activities according to that. Searching through data enables to identify unknown or hidden information or patterns that are resultant product of exploiting relationships among different attributes in existing databases. This knowledge thus extracted is then utilised to improve the engineering processes and quality of decision system which ultimately affect the controlling strategies of the whole system.

The engineering sector in Pakistan such as textile industry, manufacturing engineering, process industry etc. have professionals and engineers with capabilities to exercise their potentials to bring changes in manufacturing industries and provide industry with solutions but most of their talent wasted in fixing problems at production line without proper implementation of information technological efforts. They bent on finding problems and fixing those problems with manual efforts. The cause of problems and their effects are not recorded in the forms of databases. Though events are recorded but they are in the form of log books or on machine record history cards/charts. Although these records help in finding the brief history of events happened in the past but they do not give any clue of the exact reason of the machine failure or any other damage. It is virtually impossible to know the influence of the external factors in the machine failure or production loss etc. It is therefore necessary to make arrangements to start building data warehouses so that any details of events occurred in past and their relationship would be recorded. Although it is not compulsory to extract knowledge or information from the builtin data warehouses but in order to best utilize data mining technology in Pakistan, industry need to start building their data warehouses where any abnormalities, events, and dependency relationships could be saved, if attributes are chosen intelligently.

## III. DATA WAREHOUSING FOR INDUSTRY

The existence of any data warehouse is very important in an enterprise where data mining is to be used as a knowledge retrieval tool. A data warehouse saves substantial time that would be wasted in collecting data from multiple sources during the initial stages of data mining. This is particularly true in multi-site operations, which are common in large industries in Pakistan, where manufacturing processes do not all reside at one site and the product has to move from one site to another during its manufacture (as shown in the Figure-1). In such a complicated production environment the data should be collected from each individual site and then after cleaning the integrated information should be stored in a data warehouse. If data has to be collected from the individual sites before it is used for mining, considerable effort will have to be put into pre-processing. But if the process requires the mining to be done online then a network needs to be established which shares the information extracted or any governing rules discovered on one stage or site, with the other stages or sites using data mining tools such as decision trees, association rules, neural networks etc [5]. For static data mining where historical data is analysed to determine any relationships between the different processes or machines to improve quality of the product, or to decrease the production time or to introduce a more flexible manufacturing system then a common data warehouse will be the best choice. Data from all the engineering sites and processes should be archived at a common storage location after cleaning and integrating to avoid any duplications and errors. If the requirements of data-mining implementations are considered when information is stored in the data warehouse, pre-processing effort will be reduced and in consequence, the time of the whole knowledge discovery process will also be reduced.

## IV. AREAS TO BE EXPLOITED THROUGH DATA MINING TO EXTRACT KNOWLEDGE

Data mining analysis tends to work from the data up and the best techniques are those developed with an orientation towards large volumes of data, making use of as much of the collected data sets as possible to arrive at reliable conclusions and decisions. The analysis process starts with a set of data and uses a methodology to develop an optimal representation of the structure of the data, during which time knowledge is acquired. Once knowledge has been acquired the process can be extended to larger sets of data working on the assumption that the larger data set has a structure similar to the sample data. Again this is analogous to a mining operation where large amounts of low-grade materials are sifted through in order to find something of value.

An enterprise needs to redesign its aims that have not been achieved or to define strategies through which these aims can be acquired. Good strategy and a well structured enterprise result in profit whereas bad strategy or a business that does not meet the competitors' challenges, damages the

company in the marketplace. Therefore management decisions should be based on accurate and reliable information that is structured within a data warehouse and a

factory data model [6]. A factory model focuses on operation and infrastructure in contrast to a data warehouse provides information about the behaviour of the existing

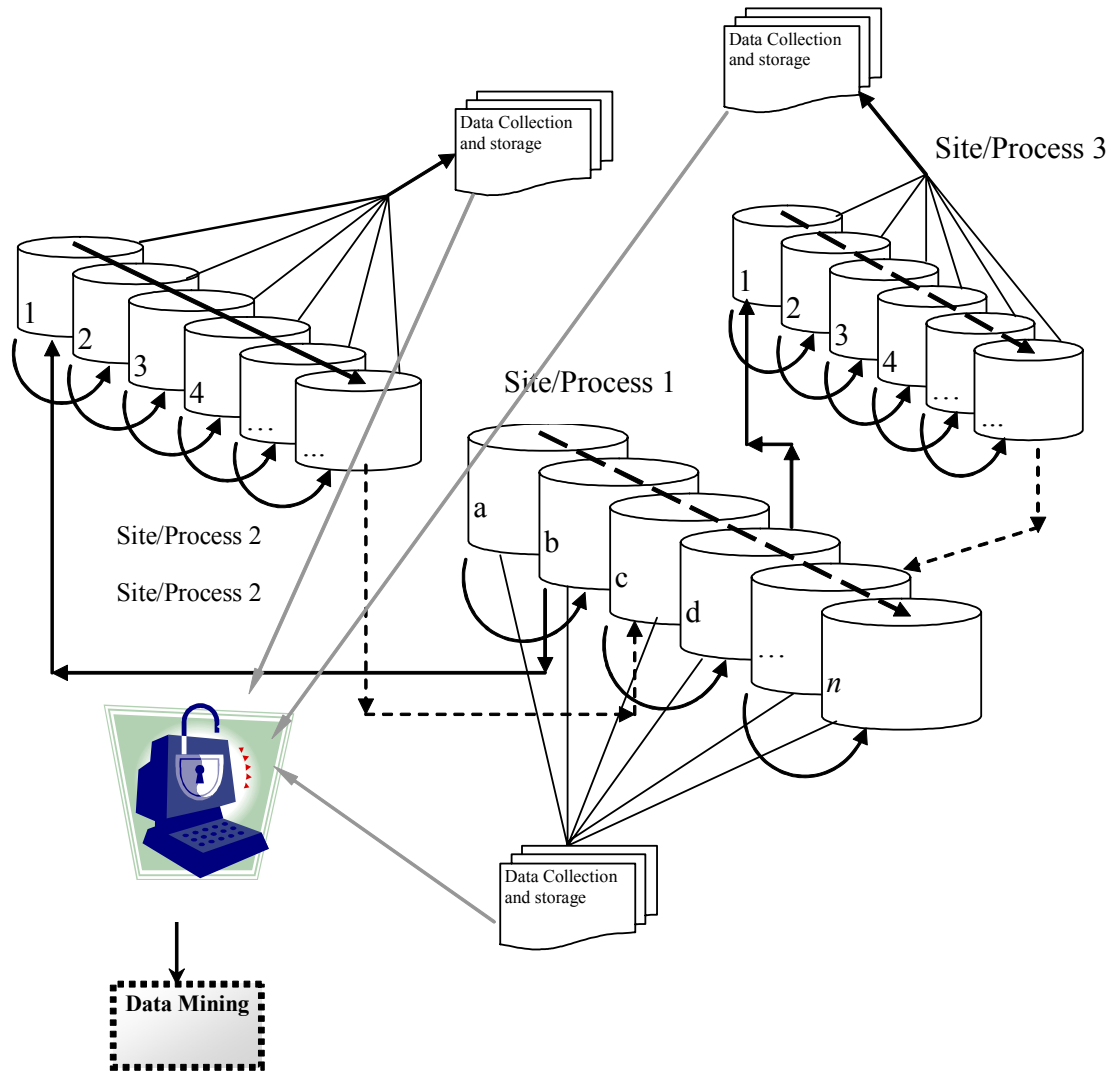


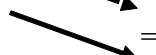

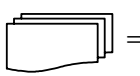




Fig1: Data warehouse supported data mining in a manufacturing process

**Legend:**

-  = Data Collecting stage of a Manufacturing Process
-  = A manufacturing process flow where product can be moved to another site and then return
-  = A manufacturing process flow where product continuously flows till the end of the process
-  = Movement of the in process product from one stage to the other
-  = Data Collection and recording at the local storage
-  = Product is being transferred from one site to the other
-  = Finalized product from one site brought back to the original production line for further operations

enterprise. Both these source of information are vital for the design and redesign of an enterprise and for performance evaluation.

However, the existence of this useful information is only the part of solution. How it can be utilized effectively to produce the required results is equally important. Simple statistics normally work well to give a very good picture of the current overall engineering process but there may be much more hidden knowledge needed to be discovered. Machine learning and artificial intelligence tools can be used to gain insight into the data and to discover hidden patterns and trends.

The above analysis regarding the ways of extracting information may not appear to provide adequate solutions as the results of the mining process cannot be predicted. Hence, we cannot be sure that the value of the knowledge that may be discovered in the data will be greater than the time and resources that need to be spent in the mining process. However implementations of the same technology in other areas of human sciences like banking, finance, marketing, insurance, telecommunication, health care etc. have given very good results [7, 8] and people are now benefiting from the knowledge they have gained in their respective fields. This naturally begs a question that how engineering enterprises in Pakistan could be benefited from the application of these techniques?

Now we give an overview of the areas where data mining techniques can be applied effectively and analysis of various industrial issues can be made successfully to save industry from going to be collapsed in coming years of time

#### *A. Textile*

The textile sector is well established industry in Pakistan and plays a vital role in the development of the manufacturing sector in Pakistan. The industry is boosted up and it gave an impetus to the production industry in Pakistan. It comprises of major part of the manufacturing sector, its output, employing large numbers of skilled people equipped with engineering knowledge and earns a major part of foreign exchange for the country. Textile Industries have to face several obstacles in the market. Such obstacles can be over come through improved information management capabilities and better decision making strategies. Textile industry in the world in general and in Pakistan in particular is not utilizing the advance tools of data analysis and artificial intelligence. Research shows that textile sector is not taking advantage of information technology effectively and efficiently. Efforts should therefore be made in order to develop awareness in the textile sector to use advance tools to increase the production rate and improve the quality of the products [9]. The full benefits of IT will only be realized by a system that can manipulate captured data to ensure that only appropriate

information need to be directed in a timely fashion to decision makers.

Data mining in textile manufacturing has the potential to improve the manufacturing process and to enhance the usefulness of existing data. There are several intermediate or sub stages that can get benefit of data analysis to check the reliability of the previous process and also any problems or faults anchor in the previous and/or current stage. In this way, for example problem of printing can be traced back in weaving or even at spinning stages or with the chemical composition and the environment of the printing area etc. The quality assurance problems at the stitching stage can have some causes from the fabric, thread or from the stitching machines. The quality of weaving, colour printing and acceptance of new patterns at public level are issues of major concern. These problems are due to mishandling of machines or some careless attitude or inefficiency of machines which puts a lot of barriers in bringing good quality clothes and the colour of fabric either goes to be faded or unappealing to the customer that cause big loss in terms of wastage of efforts and money. These issues can be identified by proper collection of data from the production and product's life cycle and integration with the production process in order to take advantage of this useful technology in Pakistan.

#### *B. Manufacturing*

Manufacturing Industry in Pakistan which has gone through a lot of sufferings due to inception of information and computer technology (ICT) and it has not been in practice in Pakistani Industry at large basis which might bring heavy loss in next couple of years. Both small or medium level enterprises and high level industries are striving hard to give maximum output with their best quality products. In this competitive environment neighbouring country like China, India and far east countries are in the race to produce product at cheaper rates and in short time to meet the requirements of their customers. Thus Pakistani manufacturing sector is struggling to cope with the high standards of quality, keeping costs and/or prices at minimum level.

In this rapidly changing scenario data mining techniques can be exploited to predict their manufacturers needs and can reduce product time to market by maintaining the international standards as well. Data mining can help engineers and professional in the manufacturing sector to discover and analyse the weaknesses present in the manufacturing sector. With the applications of artificial intelligence techniques like data mining costs can be reduce to minimum with the increase in the quality of the products.

Production rate can also be increased without compromising with the quality of the products. There is a need to train the engineering professionals to take advantage of modern data analysis techniques so that they

can use their own experience and knowledge along with the machine learning algorithms to find out the complex relationships present between different controlling parameters.

### C. Telecommunication

Telecom business, like other parts of the world, is also flourishing in Pakistan. Companies in this sector are trying their best to attract customers by providing them with best services at cheaper rates and at their doorstep. The big advantage of telecom business is the availability of the huge piles of data. This data may or may not be ready to use but data warehouse can purposely built from such industries with comparatively less effort.

Increasing competitive pressures require that telecom industry of the country should develop new and innovative ways to satisfy the increasing demands of customers. To develop these new ideas information about the needs of their customers and turning these information into valuable knowledge must be carried through data and customer's behaviour could be predicted. The data should be seen as one of the greatest asset of the business enterprises. However, the challenging issue faced by most of the telecoms organizations is that the volumes of data collected is huge in size and the range of customer's need is so diverse that it seems impossible to rationalize the situation of what is happening, and how this could be encountered thorough some technological efforts. The road from data to information, and finally to the decision making process itself, is not an easy task.

A numbers of inferences can be made through use of data mining technology and possible questions that can be answered can be enlisted as exemplified that

- What types of e characteristics are common in the customers?
- How churn rate of the customer can be predicted?
- How customers can be valued in terms of their behaviour?
- How fraudulent issues can be addressed while dealing with their customers?
- Can I identify the characteristics of customers who need special action as suspension or deactivation?
- Can I identify new prospects using demographic data?
- Do I have groups of customers for whom I can develop new products or services?
- Which customers are potential candidates for cross-selling, up-selling etc..

Another imperative outlook is how to identify the fraudulent behaviour of your customers. Data mining can be employed to identify these behaviours using the clustering data mining technique. The results of this type of data mining can be used, for example, to determine which

types of behaviours should be considered as fraud, or to produce an automatic alarm system for the fraud.

In a similar way data mining can also be used at the technical side of the telecom business. The problems that regularly occurring at their BSCs, MSCs etc, at exchanges and other areas including the system & power failure and environmental reasons, can be corrected by utilizing data mining tool for knowledge and relationship extraction. This can help in minimizing the costs and increasing the profits which can then be delivered to the customer and increase the business.

### D. Process Industry

Pakistan has many continuous process industries that are partially collecting data mostly for their parts warehouses and for marketing their products. Although process details are also being recorded but these details are mostly incomplete. Data need to be recorded for each and every component that is being used in the process. Some time it can increase the cost exponentially but efforts can be made using the expertise of the experienced people to check the importance of each and every component and then data recording can be on done on selected equipments. These details can help in identifying the causes of failure or abnormalities in the process.

The data from the parts warehouse can help in developing relationships between the machine failures and help in preventive maintenance plans.

It is therefore important to spend some funds on research and development to extract knowledge from their current data and to further improve it.

### E. Agriculture

Agriculture sector of Pakistan contributes to more than 24% of our GDP [10], and has the most profound impact on our national economy. Use of data mining in agriculture sector can change the future and can strengthen the national economy. Recently a data mining based system "Agro-Informatics" is in use which is the development and application of data mining based solutions for gathering, managing and analyzing data produced by agricultural systems and to develop models and forecasting systems. Data Mining can be used effectively in analysing different agricultural related databases. Pakistan is basically an agricultural country where more than 70 percent of the people are directly or indirectly related with the agriculture. Some work has been done on agricultural databases in the past [11] but there is still an extensive need to use modern data analysis techniques in this area to maximize the production under the local environment and weather conditions. Data Mining can also be used to soil quality analysis or in genetic engineering to obtain good quality

seed or to help in evolving new crop breeds. Data Mining can also help in discovering the relationships for best utilization of the cold storages and use of canal water. Model farms, that already exists in different part of the country can be used to collect huge amounts of data about the production and effective use of land to increase the yield.

## V. CONCLUSION

We have discussed in this paper some of instances from industries where data mining techniques can take us a long way to meet the requirements of rapidly changing production environment over the globe. But the problem which is of grave concern is the availability of data which is collected is either so rough that can not be exploited or available in a poor representable format. But still there is a potential to make some substantial changes in the manufacturing industrial environment so that industries can be benefited from the available tools and techniques provided by the data mining technology. The break even can be achieved very quickly as there is more room of improvement as compared with the modern industries in the developed countries. Traditional engineering techniques can help to some extent but there is a need to work on our system using advanced techniques to reduce the gap both in the production line and knowledge experts using data mining technology.

There is also a need to establish close relationship between industrial and academic research. Government should take some measures in order to establish effective link so that new methods could be developed keeping in mind the local environment, behaviours, system and limitations. This will help in not only developing new methods but also for the reuse of available methodologies that have been used effectively in the developed nations.

## REFERENCES

- [1] Han, J. and Kamber, M., *Data Mining: Concepts and Techniques*. 2001: Morgan Kaufmann Publishers. 1558604898
- [2] Usama, M. F., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., *Advances in Knowledge Discovery and Data Mining*. 1996: AAAI/MIT Press.
- [3] Shahbaz, M., Srinivas, Harding, J. A., and Turner M W, "Product Design and Manufacturing Process Improvement Using Association Rules", Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 220(2), 2006, pp 243-254.
- [4] Amir Netz et al, "Integration of Data Mining and Relational Databases" Proceedings of the 26<sup>th</sup> International Conference on Very Large Databases, Cairo, Egypt.
- [5] Shahbaz, M. and Harding, J.A., "An Integrated Data Mining Model for a Manufacturing Enterprise" , *Advances in Manufacturing Technology XVII* , Y. Qin and N. Juster (eds), Professional Engineering Publishing Ltd, UK, Proceedings of the International Conference on Manufacturing Research , Strathclyde, UK, September 2003, pp 539-545, ISBN 1-86058-412-8
- [6] J.A. Harding, B. Yu, "Information-centred enterprise design supported by a factory data model and data warehousing". *Computers in Industry*, Elsevier, 40 (1999) 23-36
- [7] Hashimot, K, Matsumoto, K, Terabe, M. "Applying Data Mining Techniques to Telecommunication Services" Japanese Society for Artificial Intelligence, vol 17, part 3, p. 320-325 ISSN: 09128085, 2002
- [8] Omer F. A., Ertan K., Piero M., "Data Mining for Database Marketing at Garanti Bank", *Data Mining II*, Edeken, NFF et al (Editors), WIT Press, 2000.
- [9] George H., William O., Yatin K., Neil C., "Information Engineering: Textile Industry's Value-adding key to Effective Decision making" (I99-S10), National Textile Center Annual Report: November 1999, <http://www2.ncsu.edu/unity/lockers/project/nteprojects/projects/I99-S10/>
- [10] Dr. S.M. Alam and Dr. M.H. Naqvi, <http://www.pakistaneconomist.com/database2/cover/c2003-28.asp>, Access data March 28, 2007
- [11] Ahsan Abdullah et al, "Learning Dynamics of Pesticide Abuse through Data Mining", Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation - Volume 32, Dunedin, New Zealand, Pages: 151 - 156