Industrial Science
Vol.1,Issue.3/Feb. 2014 **ISSN** : 2347-5420

Research Papers

DATA MINING INDUSTRIAL APPLICATIONS

DIVYAM SINGH

Dronacharya College Of engineering
Gurgaon, Haryana ,India

Abstract

Novel, advanced sensors, dynamic development of information technologies as well as modern high-performance computers applied in different fields of human activity result in large amount of data.

Consequently, these data, grouped in the data sets are both large and complex. The complexity come from the several mutually excluding factors like acquisition with different sensors at various times, frequencies or resolutions. The increasing size and complexity of data in different practical, often industrial branches stands the challenging problem for nowadays scientific disciplines.

KEYWORDS:
Industrial Applications , visualization and statistics.

1. INTRODUCTION

Regarding to above facts, it is notable that these techniques progressively displace many traditional methods (eg. visualization and statistics) that are no longer suitable for the analysis. Like schematic drawings and mathematical equations were formerly necessary to obtain competitive advantage, currently - data mining techniques place similar role. They enable to make scientific discoveries, gain fundamental insights into considered physical process and advance in their better understanding.

Although “data mining” term reflects somehow its idea of mining the data in general, it had a varied origins and history, evolving during time and borrowing and enhancing ideas of different fields. These domains have included statistics, image processing, machine learning, mathematical optimization, information retrieval etc. That's why data mining has multidisciplinary nature (Kamath, C., 2009).

It is worth to point out that in some disciplines like statistics, terms “data mining” or “data dredging” have negative connotation, regarding to the fact they were used to describe extensive searches through data. Statisticians tend to ignore the developments in data mining (Duebel, C., 2003), (Tang, Z., 2005).

The term “data mining” originally referred to a single stage of Knowledge Database Discovery (KDD) process. While KDD is a nontrivial process of identifying valid, useful and understandable patterns in data, data mining means which patterns are extracted and enumerated from data. This idea combines regularities finding (hidden for human) with computer's calculation speed in large amount of data (Jacobson, R. & Misner, S., 2005).

Some practitioners (Simoudis, E., 1996) refer to data mining as the process of extracting valid, previously unknown, comprehensible and actionable information form database and using it to make crucial business decisions. This approach joins data mining with data warehouse and divide the process into four actions carry out on data: selection, transformation, mining and results interpretation.

However circular definition considers it as process of extracting useful information from data. Some data miners preserve distance form terminology debate, focusing on how the ideas from different fields can be combined and enhanced to solve the problems of interest in data analysis.

It is hard to clearly distinguish data from a single component discipline on one hand. But on the other hand these technologies confluent with a new forms of data like natural language processing and comprise data mining.

Considering the golden triangle of research, knowledge and innovation, data mining approach should find also practical application in supervisory and control systems.

The chapter makes the attempt to present data mining techniques usage in various industrial applications.

2. KDD AND COMMON DATA MINING THEMES

Data mining stands one of the stages of Knowledge Database Discovery (KDD) process. Although, there are many data mining techniques, they all have their origins based on science disciplines like statistics (statistical multidimensional analysis) or machine learning. The idea of KDD combines regularities finding (hidden for human, because of time limitations) with computer's calculation speed in large amount of data (Jacobson, R. & Misner, S., 2005).

(Fayyad, U. et al., 1996) defines KDD as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of prior knowledge, and proper interpretation of the results of mining ensure that useful knowledge is derived from the data.

The necessary stages in overall KDD process applied in manufacturing are delineated in Table. 1. The KDD process is interactive and iterative involving more or less the presented stages (Fayyad, U. et al. 1996; Mitra, S. et al. 2002).

(Choudhary A., et. al., 2009) underlines that data mining is an interdisciplinary field with the general goal of predicting outcomes and uncovering relationships in data.

Bounded with data mining, sophisticated algorithms allow to discover hidden patterns associations, anomalies from large amounts of data stored in a data warehouse or other information repositories.

The authors emphasize, that in the context of industry or manufacturing apart classification, two high level primary goals of data mining are prediction and description.

First of them - descriptive data mining focuses on discovering interesting patterns to describe the data and the second – concentrate on predicting the behaviour of a model and determining future values of key variables based on existing information from available databases.

The above goals can be achieved by using a variety of data mining tools and techniques, although predictive model can be descriptive, so that boundaries between them may not be sharp.

It is crucial to determine the type of knowledge to be mined, because it determines the data mining functions.

In most manufacturing problems, it is necessary to view the summarized data in concise, descriptive terms to provide an overall picture of the manufacturing domain's data or distinguish it from a set of comparative classes. This type of data mining is called concept description and includes characterization and discrimination that are basically used to understand the process. Regarding to concept description several functions can be distinguished: quality control, job scheduling, fault diagnosis, maintenance or defect analysis.

No.	Stage name	Description
1.	Understanding the manufacturing domain	Stands the relevant prior knowledge related to manufacturing application and targeted goal.
2.	Collecting the targeted data	This stage includes the collecting raw data, selecting the data sets and focussing on the set of variables affecting the problem partly defined in the first stage.
3.	Data cleaning, pre-processing and transformation	This incorporates the pre-processing of data such as noise removal, and both replacement of missing values and data cleaning. Data are consolidated into forms appropriate for mining.
4.	Data integration	This step includes multiple manufacturing heterogeneous data sources integration.
5	Choosing the functions of data mining	Depending on the problem defined (stage 1) various data mining functions (clustering, classification, prediction, association, regression, summarization etc.) need to be performed to derive the model.
6.	Choosing the appropriate data mining algorithm	The selection of technique is important to perform the desired function for finding patterns in the data.
7.	Data mining	Includes searching for patterns of interest in a particular representational form or a set of such representations.
8.	Interpretation and visualization	Tasks that include the interpretation and visualization of patterns to derive novel knowledge.
9.	Implementation of discovered knowledge	Incorporation of discovered knowledge into the manufacturing domain performance system. The feedback is received and the knowledge can be modified further based on the achieved feedback.
10.	Knowledge storage, reuse and integration into the manufacturing system	This includes the storage of discovered knowledge for future reuse and possible integration in to the manufacturing system.

Table 1. Stages of KDD process in manufacturing (Choudhary A., et. al., 2009)

There is another important domain in industry and manufacturing which is a learning function that maps a data item into one of several predefined categorical classes. Named classification, builds to describe a predetermined set of data classes or concepts. This is known as supervised learning regarding performing analyses on the database tuples described by attributes form the training dataset. Next the learned model is represented in the form of classification rules, decision trees, or mathematical formulas.

What is important, obtained model, based on classifier accuracy, is used for classification of the future data or test data. General techniques used for classification are decision tree induction, Bayesian classification, Bayesian belief network and neural networks (NN). Othertechniques such as K Nearest Neighbour (KNN), Case Based Reasoning (CBR), GA, RST, Fuzzy Logic (FL) and various hybrid methods are also used for classification purposes (Han, J. & Kamber, M., 2001).

Classification domain is a very useful solution in many areas of manufacturing. It covers semiconductor industry, defects classification to find patterns and derive the rules for yield improvement. One of the example of the classification is online Control Chart Pattern Recognition (CCPR) for SPC. This approach is based on association of unnatural patterns displayed by a control chart with specific causes that adversely impact the manufacturing process.

Nowadays prediction in manufacturing processes, maintenance, quality improvement or defects detection become significant element. It is defined as a learning function that maps a data item to a real valued prediction variable. Prediction is usually viewed as the construction and use of a model to assess the class of an unlabelled sample. Rarely as prepared model to assess the value or value range of an attribute that a given sample is likely to have (Choudhary A., et. al., 2009).

3. INDUSTRIAL DATA MINING APPLICATIONS OVERVIEW

Since industrial systems become very complex, classical control methods become more sophisticated to lead the process more adequate according to appropriate conditions form economic (cost-effectivenes) to safety.

Both technology development as well as requirements factors are crucial to modern industry. Their main aim is advising process operators or even replace them regarding to human fault elimination and increasing both the level of quality and security.

Such approach is not new. It seems to be the continuation of computational intelligence ideas implementations, stared in early 70'. Development of scientific principles of artificial neural networks, predictive and adaptive control, become a new challenge for scientists and industry practitioners.

It is notable, that pure optimizing of known process lines stands only a part of interests. Using innovative technology allows to gain a competitive advantage on one hand, but it also opens the new possibilities to very complex, nonlinear processes, where it is very hard or impossible to gather precise, direct information from measurement equipment directly, due to high and unforeseen dynamics or extremely hard environment conditions obstacles.

Modern manufacturing facilities are in general highly automated with advanced process monitoring and data archiving systems. Large amount of process parameters and outcome variables over a number of production runs are stored in the data warehouses. Such vast amount of data stands a vital resource to comprehend the complex characteristics of several processes as well as enhance production quality, robustness or any other particular parameters.

Several studies were presented in literature according to implementation of data mining techniques. The use of data mining techniques in industry and manufacturing began in early 1990s regarding to (Lee, M. H., 1993). Generalization the algorithm, predicting the future experiments outcome under various conditions by (Irani, K. et al., 1993) opened a new chapter in of semiconductor manufacturing applications due to diagnosis and process modelling aspects.

Semiconductor industry had enormous influence on data mining appliances, because (Bertino, E. et al., 1999) in 1999 also reported successful applying data mining techniques to wafer manufacturing. Since 2000 more complex and detailed studies were applied regarding to increasing level of technology process. Directly attached computer component manufacturing uses data mining techniques to improve the process as well as to provide management staff exact/appropriate information. This computer component manufacturing was described by (Gibbons, W. et al., 2000).

The quality control automated data mining system was presented in 2001 by (Maki, H., 2001). A novel, perception-based method for automated construction of compact and interpretable models from highly noisy data sets was presented in (Last, M., 2004), where useful and understandable patterns in manufacturing data was extracted from two semiconductor products. Authors also described possible directions for the future use of automated perceptions in data mining and knowledge discovery. (Huang, H. et al., 2005) made an analysis of products quality improvement in ultra-precision manufacturing industry using data mining for developing quality improvement strategies in optical products. Their findings showed that the important factors for percentage of devices were type of processing chain, precision requirement, product classes, and raw material. It is notable, that optimum range of target group in production quality indicators was identified from gains chart.

Review of the several applications of data mining in manufacturing engineering, in determined production processes, operations, fault detection, maintenance, decision support, and product quality improvement contains (Harding, J. et al., 2006). They presented several domains of data mining grouped chronologically from early 1990s to 2005. In (Wang, L. et al., 2006) notices that data mining had created new intelligent tools for automated extracting useful information and knowledge with its profound impact on practices in manufacturing. He discusses the trends of these changes in 2006. Likewise (Rokach, L. et al., 2006) proposed the BOW (Breadh-Oblivious-Wrapper) algorithm for discovering the appropriate decomposition structure, based on idea of decision tree for each projection of subsets.

Studies continuation of data mining in the semiconductor fabrication process in 2007 can be found in (Chien, C. et al., 2007). (Gebus, S. et al., 2009) present knowledge acquisition software implemented for decision support system on electronic assembly line, that supports portability and flexibility.

Regarding to one of the most powerful management innovations in 2007 – cellular manufacturing, the new data mining algorithm for designing the conventional cellular manufacturing systems was developed in (Liu, C., 2007).

Another data mining application concerns the process control in CNC manufacturing presented in (Kumar, S. et al., 2007). The knowledge discovery was used there in designing a STEP-compliant system. Applied self-learning algorithms enabled increasing consistently producing quality of products due to knowledge acquired from previous data and results in eliminating the errors of the manufacturing system.

Two stage cluster approach based framework to generate useful patterns and rules for standard size charts was discussed in (Hsu, C., 2009). This solutions from 2008 had conducted an empirical study in an apparel industry to support manufacturing decision for production management as well as marketing with various customers' needs.

Knowledge-based artificial neural network model for monitoring of the manufacturing process was implemented by (Yu, J. et al., 2008). It was used for fault recognition of products regarding to quality categories. Due to product quality, a genetic algorithm based rule extraction approach was developed by (Yu, J. et al., 2008) to discover the independent relationship between manufacturing parameters.

Return to industry data mining origins in 2009 was described by (Kang, P. et al., 2009) according to a virtual metrology system for semiconductor manufacturing. To reduce the number of variables in developed models two variable selection methods as well as two variable extraction methods (PCA, KPCA) were employed. Another attainment was employing five regression algorithms employed to predict the metrology measurements.

Also automotive industry data mining accents can be found in (Buddhakulsomsiri, J. et al., 2009) where authors present a sequential pattern mining algorithm that allows product and quality engineers to extract hidden knowledge from large automotive warranty database.

The latest publications from 2010 concern a new approach based on particle swarm optimization algorithm for clustering problems description (Durán, O. et al., 2010) or knowledge induction from data to detect and isolate machine breakdowns in carpet manufacturing (Çiflikli, C. & Kahya-Özyirmidokuz, E., 2010) or modern manufacturing facilities for bioproducts to improve robustness of large scale bioprocesses (Charaniya, S., et al., 2010), where authors demonstrate in different stages of the process the power of mining process data in revealing hidden correlations between parameters and outcomes. Separate solution stands (Bartok, J. et al., 2010) where data mining tasks and integration is used for detection and prediction of significant meteorological phenomena due to DMM project (Data Mining Meteo).

Widespread availability of new computational methods and tools both for data analysis and predictive modelling has its successful applications traditionally in business decision making (Seng, J., & Chen, T., 2010), but also in medicine (Bellazzi, R., & Zupan, B., 2008).

4. CRUCIAL MEANING OF DATABASES

Nowadays knowledge discovery, knowledge management and knowledge engineering stands the important topics to manufacturing researchers and managers intent on exploiting current assets. Database technology is central to all these knowledge-based research and engineering topics. Combining with statistical techniques, databases have been processed to derive the underlying relationships within the data processing, previously associated with statisticians (Harding, J. et al., 2006). Regarding to presented literature overview there are plenty of methods, implemented functions or even frameworks on one hand. On the other hand they are usually dedicated to the appropriate problem solution. The fact is that, the mining techniques require stable and infallible framework. There are several data mining software vendors with leadership of Oracle, Microsoft and IBM.

The complete business intelligence solution is the advantage of Microsoft's SQL Server. Apart from relational database management system (RDBMS)(OLTP database engine), it also offers Integration Services, Analysis Services as well as Reporting Services. The Whole framework hold forth to transfer data between different systems, summary reports creation as well as advanced warehouses implementation.

The common practice shows that several stages of production are handled by different databases In the classic, relational database (OLTP), records are assigned to individual transactions. It contains measured output and control signal values as well as repository states.

Despite of continuous transaction processing from different measuring devices and actuators, the typical relational model is unable to specify the total efficiency or emergency during the particular period of production. Regular finding answer for such a question requires time consuming summaries in many tables, often using complex join operations. The time absorbing computation need to be repeated, when asked about efficiency in several departments is the main disadvantage.

Such kind of reports are improved by the MS Analysis Services, that store pre-processed data in the warehouse and they are more suitable for preparing reports. This kind of database consists of the fact table with aggregated values, named measures which are divided into time, departments etc. However, the fact table in the typical OLAP (On-Line Analytical Processing) database contains keys (not values) for measures tables, apart from aggregated values.

Data contained in the fact table constitute multidimensional cube. However, there may be more than three dimensions in an OLAP system, so the term hypercube seems to be more suitable. The arrangement of data into cubes avoids a limitation of relational databases which are not well suited for near instantaneous analysis of large amounts of data.

The typical linking the fact table with measures tables uses star schema. Such a structure facilitate queries and reports creation, using Multidimensional Expressions and Data Mining Extensions languages for data analysis.

Analysis Services use previously acquired and pre-processed data by the Integration services, but often they become source for Reporting Services and other applications (Zawadzki, M., 2005), (Żarski, A., 2006).

5. PREDICTION EXAMPLE

One of the original data mining usage for prediction analysis were marketing research. The technology lets a retailer to predict, for instance, the most preferable goods bought by particular age male at his local supermarket.

This technology can be used for similar purpose, but in completely different area. Many industrial automation systems acquire on-line process data in SCADA (Supervisory Control And Acquisition) systems. They allows for process supervising by the operators. The typical SCADA system workstation block scheme is presented in the Fig.1a, where on-line database and historical data repository was distinguished.

Although human stands the "perfect controller" on one hand, he is failure susceptible, because of his physiology on the other hand. Not optimal (even suboptimal in some cases) controller sets with it's negative process influence are referred to final product quality, and finally increase the total costs.

The solution of that problem could be an advisory system for process operators, that use "faultless operator" behaviour from knowledge base already stored in the system (see Fig. 1b). It's primal aim ought to be optimized control set prediction.

The solution of that problem could be an advisory system for process operators, that use "faultless operator" behavior from knowledge base already stored in the system (see Fig. 2b). It's primal aim ought to be optimized control set prediction.

Crucial functional purposes of considered advisory system:
storage, processing and integration data from different sources;
data analysis, ability to implement Business Intelligence (BI) implementation,
hierarchical views and several data mining techniques deploy; presentation and distribution of achieved results .

Notice, that all above can be solved by the internal services of SQL Server 2005 standard version. Scheme of the proposal advisory system is presented in the Fig. 1b, where apart standard elements analysis SQL Server database are localised.

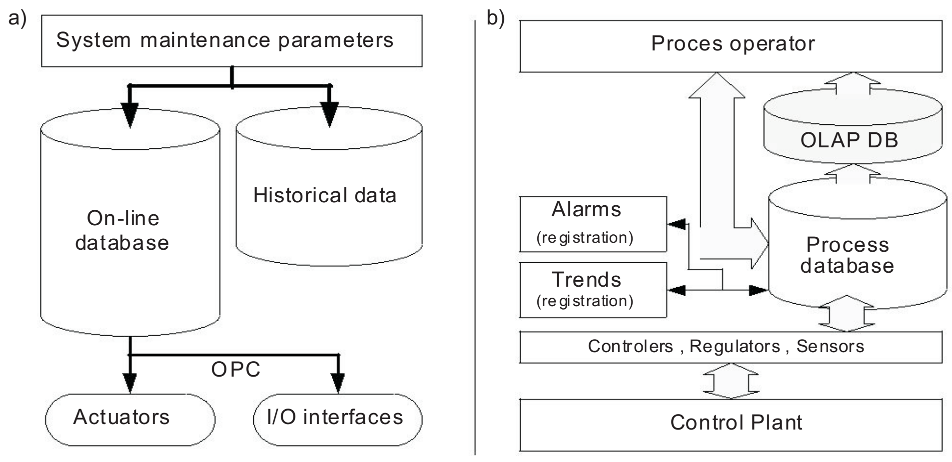

Fig. 2. Process workstation block chart, b)OLAP DB extended SCADA system scheme concept

The information gathered (and later mined) are contained in the warehouse embedded trained data model. Gathering data and making predictions on trained data models is improved by Microsoft's special query language for data mining, called DMX.

5.1 Creating and training a mining model

The prior action in this stage is as set up a connection to Analysis Services 2005 (SSAS) and create a new mining model. The purpose of the model is to predict the velocity set for beet slicer control based on some sugar production process. The DMX statement for model creation is following:

```
String CreateModel = "Create mining model BSVelocity_Prediction" + "( SampleID long key," +

"BS_DateTime text discrete, BS_No text discrete," + //which slicer "BS_Charge real, BS_Efficiency real," +
"BS_VelocityCV real," +         //measured velocity

"BS_VelocitySP real predict)" +    //setpoint
"Using Microsoft_Decision_Trees";

OleDbCommand CMD = new OleDbCommand(CreateModel, conn); CMD.ExecuteNonQuery();
```

In the above statement, after the declaration of the data type for each column, there is content type also added. This action inform the algorithm applying to the mining model (in this example, Microsoft Decision Trees) how the data in the columns is to be distributed. The last line of the Create statement uses the predict keyword, telling the algorithm that all other columns will predict the outcome of BS_VelocitySP column for the model.

Training the model consists of two stages: the data mining algorithm input cases testing and looking for correlations in the data. After correlations identification, the model is repopulated with these new patterns. Model processing starts over as new data is introduced into the model. This results in more precise predictions, because the patterns are revised over time. To populate the model with data, the DMX Insert statement is used:

```
String PipeData2Model = "INSERT INTO VelocitySP_Prediction"
+"(SampleID, BS_DateTime, BS_No, BS_Charge,"
+" BS_Efficiency, BS_VelocityCV,"
+" BS_VelocitySP) OpenQuery(btsldbsource, 'Select sampleid, datetime, "
```

+" bs_no, charge, efficiency, vel_cv, vel_sp FROM btsl')"; OleDbCommand CMD = new OleDbCommand(PipeData2Model, conn); CMD.ExecuteNonQuery();

Above query seems to be roadmap between a table called Samples in a SQL Server database, defined by the datasource btsldbsource, and the mining model. OpenQuery is a DMX function for performing DMX queries against relational databases from inside an OLE-DB session connection. Nethertheless, for this kind of tasks Integration Services in Business Intelligence Studio are frequently used.

After delivering data to the model, the algorithm may be used to test the cases as well as to identify patterns..

5.2 Control set prediction

Prepared model may be used to predict the best control set of velocity (setpoint), with appropriate efficiency level and current load of the beet slicer. For example, velocity control set is going to be determined with efficiency greater than 62%, efficiency range 55-85%, with a 80% probability or better. DMX Select query statement is used to make predictions on the model:

String PredictModel = "Select T.SampleID, VelocitySP_Prediction" +".BS_VelocitySP From VelocitySP_Prediction NaturalPredictionJoin"
+" OpenQuery(BTSL, 'select * from NewSamples) As T"
+" Where T.BS_Efficiency > 62 And T.BS_Charge Between 55 And 85"
+" And PredictProbability(BS_VelocitySP, '60') >0.8"; OleDbCommand CMD = new OleDbCommand(PredictModel, conn); OleDbDataReader myReader; myReader = CMD.ExecuteReader(); while (myReader.Read()) {//return data }myReader.Close();

Considered query introduces new cases to the mining model from BTSL datasource, con-taining the table NewSamples. The DMX function NaturalPredictionJoin allows to join the data from the NewSamples table and model without any additional specifications, because both tables have the same columns. Function, PredictProbability is used in conjunction with the Where clause to produce desired results (Smith, J., 2003), (Tang, Z., 2005).

5.3 Industrial usage assessment

To be competitive, the companies ought to adapt to the market changes very quickly, maximizing their profit with simultaneous lowering production costs. The reasonable (cheap) solution is optimization of the most important stages of the production process. The key element allowing to bring the optimization through is processing appropriate information in proper time is the. It gives a stable framework for wise and precise decision making.

Although data mining techniques are mainly addressed to IT branch, banking and stock markets, there are many arguments for industrial usage.

Described solution is intended particularly both to the enterprises that trying to keep up the market and also to these with modern processing lines. The local sugar industry may be a very good example, while it is strongly influenced by the French and German consortiums. To remain competitiveness Polish sugar industry may improve thanks to pro-cessing lines modernization, but it is very expensive and time consuming solution. Alternative idea – efficiency improvement by the production important stages optimization, using data mining seems to be necessary.

6. CONCLUSION

Data mining is blend of concepts and algorithms from machine learning, statistics, artificial intelligence, and data management. With the emergence of data mining, researchers and practitioners began applying this technology on data from different areas such as banking, finance, retail, marketing, insurance, fraud detection, science, engineering, etc., to discover any hidden relationships or patterns.

Data mining is therefore a rapidly expanding field with growing interests and importance and manufacturing is an application area where it can provide significant competitive advantage (Harding, J. et al., 2006).

The use of data mining techniques in manufacturing began in the 1990s and it has gradually progressed by receiving attention from the production community. These techniques are now used in many different areas in manufacturing engineering to extract knowledge for use in predictive maintenance, fault detection, design, production, quality assurance, scheduling, and decision support systems. Data can be analyzed to identify hidden patterns in the parameters that control manufacturing processes or to determine and improve the quality of products. A major advantage of data mining is that the required data for analysis can be collected during the normal operations of the manufacturing process being studied and it is therefore generally not necessary to introduce dedicated processes for data collection. Since the importance of data mining in manufacturing has clearly increased over the last 20 years, it is now appropriate to critically review its history and application.

Data mining techniques becomes the basic element of modern business. Although the idea is not new, new technologies and implemented standards make a contribution to their growing popularity. Regarding to mining model usage SQL Server 2005 stands breakthrough in this area. Thanks to the DMX language either programmers or database administrators are able to create Data Mining Systems in simple way.

Although economical and business publications are very fruitful of data mining approaches, the described problem is presented rather weak in the international publications. Nethertheless some industrial appliances of data mining technology were considered in (Duebel, C., 2003).

Industrial usage of data mining techniques opens new possibilities in decision making not only for top level management, but also for advisory or control systems. Several prediction, classification or even anomaly detection algorithms implementation may become lucrative tool for industrial process appropriate stages optimization, that combines diagnosis and control functions.

The reviewed literature shows that there is a rapid growth in the application of data mining in industry and manufacturing. However, there is still slow adoption of this technology in some industries for several reasons including both difficulties in determining the type of data mining function to be performed in any particular knowledge area and question of choice the most appropriate data mining technique regarding to many possibilities.

7. REFERENCES

Bartok J., Habala O., Bednar P., Gazak M. & Hluchy L. (2010). Data mining and integration for predicting significant meteorogical phenomena. International Conference on Computational Science, (ICCS 2010), Procedia Computer Science 1, Elsevier, 37-46

Bellazzi R., & Zupan B., (2008). Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics 77, 81-97

Bertino, E., Catania, B. Caglio, E. (1999). (Applying data mining techniques to wafer manufacturing), in: Zytkow, J.M., Rauch, J. (Eds.), PKDD'99, LNAI, vol. 1704, Springer-Verlag, Berlin, 41–50

Buddhakulsomsiri, J., Siradeghyan, Y., Zakarian, A., & Li, X. (2006). Association rule generation algorithm for mining automotive warranty data. International Journal of Production Research, 44(14), 2749–2770

Charaniya, S., Le, H., Rangwala, H., Mills, K., Johnson, K., Karypis, G. & Hu W. (2010).
Mining manufacturing data for discovery of high productivity process characteristics. Journal of Biotechnology, 147(3-4), 186-97

Chien, C., Wang, W. & Cheng, J. (2007). Data mining for yield enhancement in semiconductor manufacturing and an empirical study. Expert Systems with Applications 33, 192–198

Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: a review based on the kind of knowledge. Intell Manuf, 20, 501-521

Çiflikli, C. & Kahya-Özyirmidokuz, E. (2010). Implementing a data mining solution for enhancing carpet manufacturing productivity. Knowledge-Based Systems (article in press)

Duebel, C. (2003). Application of Data Mining Techniques to Industrial Processes for Improved Business Performance, APACT Conference

Durán, O., Rodriguez, N. & Consalter, L.A. (2010). Collaborative particle swarm optimization with a data mining technique for manufacturing cell design. Expert Systems with Applications 37, 1563–1567
Fayyad, U. M., Shapiro, G. P., & Smyth P., (1999). The KDD process for extracting useful knowledge from volume of data. Communications of the ACM, 39, 27-34
Gebus, S. & Leiviska, K. (2009). Knowledge acquisition for decision support systems on an electronic assembly line. Expert Systems with Applications 36 (1) 93–101
Gibbons, W., Ranta, M., Scott, T. et al. (2000). Information management and process improvement using data mining techniques. in: Logananthara j, R., et al. (Eds.), IEA/AIE 2000, LNAI, 1821, Springer-Verlag, Berlin, 93–98
Han, J. & Kamber, M. (2001). Data mining: concepts and techniques. Morgan Kaufmann, USA
Harding, J.A., Shahbaz, M., Srinivas, S. & Kusiak, A. (2006). Data mining in manufacturing: a review. Journal of Manufacturing Science and Engineering 128, 969–976
Hsu, C. (2009). Data mining to improve industrial standards and enhance production and marketing: an empirical study in apparel industry. Expert Systems with Applications 36 (3), 4185–4191
Huang, H. & Wu, D. (2005). Product quality improvement analysis using data mining: a case study in ultra-precision manufacturing industry. in: Wang, L. Jin Y. (Eds.), FSKD 2005, LNAI, 3614, Springer-Verlag, Berlin, 577–580
Irani, K.B., Cheng, J., Fayyad, U.M. et al., (1993). Applying machine learning to semiconductor manufacturing. IEEE Expert 8 (1), 41–47
Jacobson, R. & Misner, S. (2005). Microsoft SQL Sewer 2005 Analysis Services. Step by step. Promise
Kamath, C. (2009). Scientific Data Mining. A practical perspective. Society for Industrial and Applied Mathematics, Philadelphia
Kang, P., Lee H-J., Cho, S., Kim, D., Park, J., Park, C-K, & Doh, S. (2009). A virtual metrology system for semiconductor manufacturing. Expert Systems with Applications 36, 12554– 12561
Kumar, S., Nassehi, A. & Newman, S.T. et al. (2007). Process Control in CNC manufacturing for discrete components: a STEP-NC compliant framework. Robotics and Computer Integrated Manufacturing 23, 667–676
Last, M. & Kandel, A. (2004). Discovering useful and understandable patterns in manufacturing data. Robotics and Autonomous Systems 49, 37–152
Lee, M. H. (1993). Knowledge based factory. Artificial Intelligence Engineering 8, 109–125
Liu, C. (2007). A data mining algorithm for designing the conventional cellular manufacturing systems. in: Orgun, M.A. , Thornton, J. (Eds.), AI 2007, LNAI, 4830, Springer-Verlag, Berlin, 715–720
Maki, H. & Teranishi, Y. (2001). Development of automated data mining system for quality control in manufacturing. in: Kambayashi, Y., Winiwarter, W., Arikawa, M. (Eds.), DaWak, LNCS 2114, Springer-Verlag, Berlin, 93-100
Mitra, S., Pal., S. K., & Mitra P., (2002). Data mining in soft computing framework: A survey. IEEE Transactions on Neural Networks, 13(1), 3-14
Rokach, L. & Maimon, O. (2006). Data mining for improving the quality of manufacturing: a feature set decomposition approach. Journal of Intelligent Manufacturing 17 (3), 285– 299
Seng J., & Chen T., (2010). An analytic approach to select data mining for business decision. Experts Systems with Applications 37, 8042-8057
Simoudis, E. (1996). Reality check for data mining. IEEE Expert, 26–33 Smith, J., (2003). Data mining with C# and ADO.NET, www.devsource.com Tang, Z., (2005). Data Mining with SQL Server 2005, John Wiley & Sons
Wang, K. (2006). Data mining in manufacturing: the nature and implications. in: Wang, Kovacs, G., Wozny, M. et al. (Eds.), International Federation for Information Processing (IFIP) Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management, vol. 207, Springer-Verlag, Boston, 1–10
Yu, J., Xi, L. & Zhou, X. (2008). Intelligent monitoring and diagnosis of manufacturing processes using an integrated approach of KBANN and GA. Computers in Industry 59, 489–501
Zawadzki, M., (2005). SQL Server 2005, MIKOM
Żarski, A., (2006). Data mining using SQL Server 2005, www.codeguru.pl