International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

A Review of Data Mining and its Methods Used in Manufacturing and How Warehousing Impacts Manufacturing

C S Nagashree[1], Dr. Shanta Rangaswamy[2]

[1]Dept. of Computer Science and Engineering, RV College of Engineering, Bangalore-560059
[2]Associate Professor, Dept. of Computer Science and Engineering, RV College of Engineering, Bangalore-560059

Abstract: *Huge data being generated every instant need management and processing. Knowledge being one important aspect of any process, it can be obtained from data and information processing where data can be from any processes including purchasing, manufacturing, designing, marketing, maintenance and even distribution. Knowledge Discovery is a major outcome of datamining which discovers patterns present in data. Manufacturing is process of goods making which are in turn is sold to end customers. Warehousing being a process that allows organizations to control and administer the operations of warehouse beginning from goods entering until they are being shipped out. Warehouse management is an integral part of manufacturing to improve its effectiveness. This paper focus on review of datamining in manufacturing and how warehousing has been easing manufacturing process to be performed efficiently.*

Keywords: **Data Mining, Manufacturing, Machine Learning, Warehousing, Knowledge Discovery**

1. INTRODUCTION:

Data mining being dealing with data where patterns in data are being mined. It is a non-trivial extraction of novel, implicit, and actionable knowledge from huge data sets which increasingly uses computer databases for storing and retrieving information effectively and this is referred to as Knowledge Discovery in Databases (KDD). [1] basically, enables exploring of data, analysis of data, and visualization of data present in huge databases. Data mining techniques are results of research processes and development of product and these includes artificial neural networks, decision trees and genetic algorithms and many more. [2] can find usage in many fields out of which manufacturing being one. (datamining process paper to be included)

Data mining fits into an iterative process which consists of the following steps illustrated in Figure 1:

1. Data selection: is the process of resolving the correct data. Here, the relevant data for the analysis is selected and retrieved from the collection of data. It also identifies the instrument to collect the accurate and legitimate data.
2. Data cleansing: known to be data scrubbing where noisy, unwanted, incomplete and irrelevant data are removed from the collection or even replaced.
3. Data transformation: known to be data consolidation. In this phase the selected data is transformed, summarized, normalized following data-preprocessing into fitting representation suitable for the mining procedure [3].
4. Data integration: is merging of heterogeneous data from varied data sources into one common source [4].
5. Data mining: A crucial step in which data is analyzed from varied perspectives and useful information is brought out. These smart methods are applied to extract patterns that are potentially useful.
6. Pattern evaluation: method that enables to identify interesting patterns that represent knowledge that are identified based on predetermined score measures, where these scores have numerical significance.
7. Knowledge representation: the final phase is about representing the knowledge obtained after mining. It is making sense of the patterns obtained. This knowledge discovered is represented visually to the consumer. It is an essential step and uses extensive number of visualization techniques to help consumers easily understand and interpret the results of mining [5].

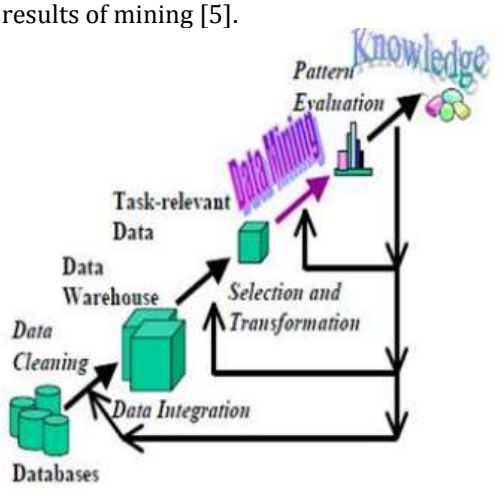

Fig -1: Data mining process [5].

Manufacturing is a process of transforming raw materials to finished products or semi-finished products and it makes use of machineries for the process. This process allows the

products to be made available to the customers to satisfy their requirements and needs. These products carry a worth in the market. These processes can be classified as assembly and processing operation where the latter is remodeling from one state to other progressive state where operation value is added to the work material by transforming its shape properties, geometry, appearance and more. In former, few components are affixed to create a new entity and this new entity yielded is called the assembly or sub-assembly based on the state of the product. A job can be defined as the total work or duties performed by the worker and a station being the location or area where tasks and jobs are performed.

[6] presents an approach for consistent manufacturing process for producing products meeting the quality necessary and the goals of reliability. This approach evaluate the performance of the work by the process and ensures that proper measurements are used for process performance allowing movement from appraisal-type philosophy to preventive-type philosophy and ensuring provisioning of zero-defect.

Warehouse is a storage place that can hold goods and products in stock for further usage. This needs inventory management for maintaining its product information and its location.

2. DATAMINING IN MANUFACTURING INDUSTRY:

Data in large databases are mined to extract knowledge. This can also give the information about manufacturing process that is appropriate. The methodology for selecting the manufacturing process be seen as 3 parts: learning, exploitation and specialization and it is taken part by part. The learning phase is discovering the knowledge from the process and exploitation is improving the above process efficiency by using clustering and specialization is adapting the relevant process to the part. This allows higher utilization of data on mining hence gaining better knowledge [7]. There are several data mining methods for knowledge gain [8].

The automation of assembly in manufacturing industries improves quality, rate of production and decreases re-work. The data mining is performed on these industrial data present as part of the database containing details such as bill-of-materials, product-design, techniques-of-manufacturing that are majorly used part of manufacturing enterprises. This process reveals patterns that can detect assemblies in sub-assemblies. This assemblies influence the important design factors [9]. The data mining method can also help detect faulty parts using apriori algorithm which identifies causal relationship that exists between manufacturing parameters and product quality. It gives abnormal warnings, identifies assignable causes and helps to optimize process parameters [10].

The manufacturing processes perhaps controlled by data driven models where these models have been of benefitting advantage in decision making for optimization of complex manufacturing procedures. These representations allow predictive modelling and on-line process monitoring for optimizing the production process and these predictive modelling provides early signals for operators on the production-lines using various parameters of machines as well as using different material properties. It finds outliers in heterogeneous data-sets using measures of density that examines the global neighborhood while identifying the outliers in the data subspaces [11]. When the data that is characterizing the manufacturing process are collected using electrical machines, must be analyzed using data mining for quality assessment. There are various techniques for assessing the quality which needs the analyst to know which tool performs better according to data set. One of the techniques being Machine (SVM) although other methods could have been an option as well for the process of mining [12].

[13] reviews datamining models and applications in manufacturing. The CRISP-DM (Cross Industry Standard Process for Data Mining) and SEMMA (Sample, Explore, Modify, Model, and Assess), SolEuNet, Kensington Enterprise Data Mining, Data Mining Group have established methods, languages and tools for standardizing the datamining industrial applications whose prime-focus being mining algorithms and application development. Their methodologies are most often used manufacturing methods in data mining community which is because of the step-by-step guidance for datamining implementation. CRISP-DM being easier than the SEMMA due to its neutral guidelines that keeps it clear for novice to implement datamining. The latter is developed as a set of functional tools for SAS's software. It also talks about various methods in various domains including engineering, manufacturing, decision support systems, shop floor control and layout, fault detection and quality improvement, maintenance, customer relationship management. These are specific to a problem rather than manufacturing wide datamining. [14] briefs about machine learning algorithms of datamining that are in use in manufacturing and talks about the efficiency that these machine learning frameworks can give and how these can be applied in datamining for decision making and how the frameworks can be used widely in other domains such as pharmaceutical, medical and industrial domains. This presents a data driven paradigm for manufacturing and service applications where the approach is seen as learning phase and decision-making phase restricting its usage to only certain industries. [15] reviews machine learning methods that play a major role in data mining are used in manufacturing processes which include supervised, un-supervised and semi-supervised methods. 90–95% of the applications in the industry make use of supervised and un-supervised learning methods but the semi-supervised methods have been in use in recent times. It considers 8

supervised methods and 10 un-supervised methods and compares among them to deduce that Principal Component Analysis (PCA) an un-supervised learning method and Principle Component Regression (PCR) and Partial Least Squares (PLS) as supervised learning method are most widely used methods in usage. Figure 2 shows usage summary of the methods in the application.

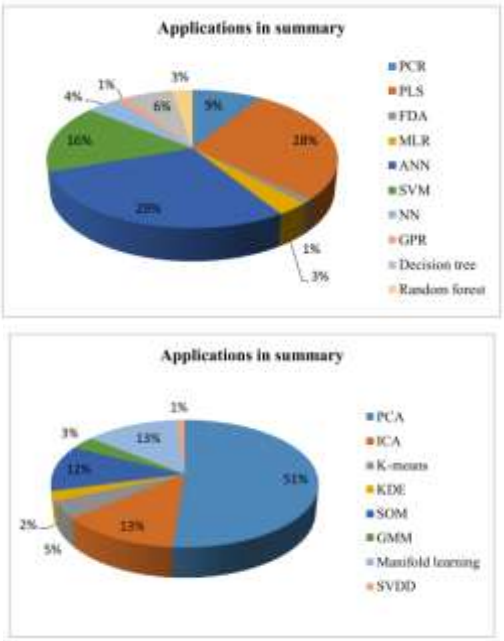

Fig -2: Usage summary of un-supervised and supervised learning methods respectively in the applications. [15]

Patterns in data can be used to improve manufacturing quality, but the data in the manufacturing plants can have unique features such as unbalanced distribution of target attribute and curse of dimensionality and training set is relative to the input features. For handling these, a feature set decomposition methodology can be used that makes use of BOW (Breadth-Oblivious-Wrapper) algorithm for examining the data. The algorithm performs breadth first search while using F-measure splitting criteria and this methodology is focused on quality improvement to solve classification problems but lacks the backtracking capability and feature decomposition is giving smaller feature sets and aims at using other methods such as support vector machines [16]. The manufacturing industry to survive in market is to be agile when making quality improvement decisions and should use appropriate data mining tools. The knowledge learned from this mining allows to identify quality issues. A decision tree approach as suggested by [17] allows optimization of parameters on which data mining techniques applied allows data analysis.

Datamining also allows data summarization and predictions to be made by discovering information from the data hidden within. This also allows manufacturers to find suitable answers to their manufacturing questions about demands in market easing decisions and strategy planning. [18] predicts future car market demand allowing launch analysis and slow turning analysis. It also compares different mining methods according to type of databases available and type of knowledge that needs to be obtained and as to which of the available techniques that could be adopted. This although covers macro-market analysis does foresee to focus on micro-market segmentation, sales prediction, loyalty rates, market share forecasting tracking investments and optimizing on-line interaction. The prediction process of data mining in [19] also allows manufacturing process behavior prediction according to production data. It uses different data mining methods taking into consideration classification and regression methods such as Boosting Tree (Boost), Standard Classification and Regression Trees (CART), Multivariate Adaptive Regression Splines (MARS) , Random Forest (RF), Multiple Regression (MR) , Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Neural Networks (NN), out of them NN performing the best for predicting the behavior.

	NN	Rand	CART	KNN	SVM	Boost	MARS
Overall accuracy	1	2	3	4	5	6	7
Correctly classified TRUE	1	3	2	4	5	6	7
Correctly classified FALSE	1	3	2	4	5	6	7
Error rate	1	2	3	4	5	6	7
Lift chart	1	3	2	4	5	6	7
Chi-square	1	3	2	4	5	6	7
G-square	1	3	2	4	5	6	7
Percentage disagreem.	1	3	2	4	5	6	7
Summary	8	22	18	32	40	48	56

Fig -3: Summary of classification methods evaluation. [19]

The figure 3 summarizes the classification methods where the points are assigned to different methods between 1-7 from highly successful to least successful. The error rates in figure 4 also indicate that NN method is most suitable for prediction.

	Error rate
NN	0.001600
KNN	0.018314
CART	0.021583
RF	0.026241
SVM	0.052129
Boost	0.144016
MARS	0.166651
MR	0.166850

Fig -4: Error rates of the regression methods. [19]

3. WAREHOUSING IN MANUFACTURING INDUSTRY:

Warehousing plays a major role in the organizations. Its operations include: Receiving, Put-away, Storage, Picking, Packing, Shipping [20] as shown in figure 5. Selecting appropriate warehouse strategies becomes a major concern as each has its own impact on efficiency of the organization.

Considerations for selecting strategies can include cost, time, space availability in warehouse and labor resource [21]. Each individual organization focus on Return on Investment (ROI) and hence requiring reducing the warehousing cost which can be done by automating the warehouse management. Warehousing also is about improving productivity and decreasing operational cost which has many influential factors like storage availability, demand trends, storage requirement, layout of storage system, distance from the inventory system, time to fetch, waiting time between processes, complexity of warehouse management and many more. RFID (Radio Frequency Identification) is used as part of warehousing and it allows full-time tracking of the product and its movement through-out supply chain [22]. RFID has a tag that is embedded in a physical device and sends radio-frequencies to its reader's antenna which then initiates business processes. This accurate data is sent to mobile data repositories in time. The usage of RFID eases warehousing tasks therefore contributing to performance. Warehouse performance can be measured by metrics such as: inventory management, order fulfillment, warehouse productivity. Automated warehouse management benefits than manual management of warehouse which includes, scheduled operations, algorithm-based storage, sequenced picking and packing, reduced total round-trip time, traceability of products, higher production rate, reduced manpower dependency. [23]

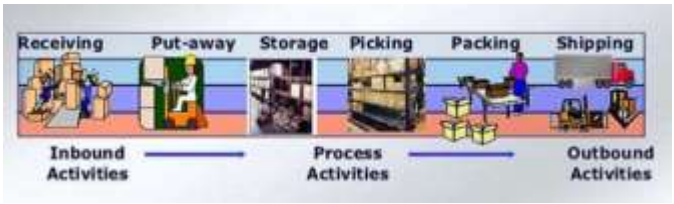

Fig -5: Warehousing operations [20]

Warehouse management can face certain issues that causes performance imbalance in the industry. Few issues are: Redundant processes, Poor layout facility, Demand variations, High cost for labor, Inaccurate inventory. The redundant processes issue can be overcome by using scanners like RFID scanners that will improve resource utilization. Layout related issues can be managed by re-structuring of layouts into vertical and horizontal layouts within the warehouse and using fork-lifts to reach horizontal heights. Variable demands can be met by product re-organizing within warehouse and changing picking patterns and proper usage of transportation services and transportation networks. Labor-cost can be managed by replacing labors by automated systems. Inventory inaccuracy can be prevented by use of proper forecasting methods and using other appropriate methods to manage a better margin for production. To summarize, these issues can be overcome by transitioning towards smart warehousing.

Warehousing in developing countries use traditional method as compared to smart method of automation of process due to the cost incurred to transform from former to latter and the lack of skilled workers present in smart warehouse. Rather the semiskilled or un-skilled workers are present in traditional warehouses. These workers contribute a major part of economy as well. The smart warehousing has an upper hand than that of traditional warehousing due to its higher throughput ability and yielding higher efficiency. On the contradictory not always are smart warehousing methods any better as compared to traditional warehouse as it requires re-training of potential employees for the new process and a small technical glitch can cause the entire process to become standby. The use case discussed in [24] illustrates that the industry defines the method of warehouse to be used. There are several options in smart warehousing like, blockchains, IOT, Automatic Identification Technology (AID), Cloud technology, Enterprise Resource Planning (ERP) to name a few and each having their own advantage and disadvantages. IOT is using data from all devices and allows integration and analysis of this data for business decisions, but it has privacy issues as network can be hacked into. ERP offers tools for various enterprise functions. [25]

Warehouse optimization is important in warehouse management and there are different optimization methods that can be used in practice according to [26]. Optimal can be in terms of customer satisfaction, in-time operations, minimal cost, optimal utilization of resources, layout of warehouse, proper utilization of labor, system assignment, routing methods, type of conditions to be applied instead of application specific conditions. Industry specific challenges can impact warehousing and can influence its performance. The performance measurement benchmark comprises of quality, speed, dependability, reliability, flexibility and time [20].

4. ISSUES AND CHALLENGES IN DATAMINING:

There have been several issues in the process of mining. To enlist a few:

1. Datamining can cause abusage of data by exploitation of vulnerable people by unethical mob.
2. Improper sources of data that can be because of data diversity and challenge in integration of conflicting/redundant data.
3. Inaccurate techniques for mining that directly shadows on decision making process.
4. Cost of data collection and maintenance. [27] [5].

Challenges that influences datamining could be:

1. Quality of Data
2. Data Scalability
3. Dimensionality of Data
4. Complex and Heterogeneous Data
5. Data Distribution and Data Ownership

6. Preserving Data Privacy
7. Streaming Data [27]

5. CONCLUSION:

The discussion in this paper is about how data mining and warehousing is taking a major role in manufacturing and how improvements in industry has seen contributions from data mining and warehousing. The data mining has allowed several crucial decisions to be made generating significant profit and allowing predictions to made for example regarding market trends. Warehousing being integral part of manufacturing has huge impact on the returns. It can be made efficient by using smart warehousing, but not all companies can transition depending on margins.

REFERENCES:

[1] L. B.N. and R. G.H., "A Conceptual overview of Data Mining," in *National Conference on Innovations in Emerging Technology*, Erode, Tamilnadu, India, 2011.

[2] E. Trybus and G. Trybus, "A brief introduction to field of Data Mining," in *PICMET '01. Portland International Conference on Management of Engineering and Technology. Proceedings Vol.1: Book of Summaries (IEEE Cat. No.01CH37199)*, Portland OR USA, 2001.

[3] K. Swati and D. S. Kumar, "A Comparative Study of Various Data Transformation Techniques in Data Mining," *International Journal of Scientific Engineering and Technology,* vol. 4, no. 3, pp. 146-148, 2015.

[4] P. Ziegler and K. R. Dittrich, "Data Integration — Problems, Approaches and Perspectives," in *Conceptual Modelling in Information Systems Engineering*, Zurich, Springer, 2007, pp. 39-58.

[5] S. Hemlata, S. Shalini and G. Seema, "A Brief Overview on Data Mining Survey," *International Journal of Computer Technology and Electronics Engineering (IJCTEE),* vol. 1, no. 3, pp. ISSN 2249-6343, 2012.

[6] M. Elbert and R. Howe, "Manufacturing process study and certification," in *IEEE 34th Annual Spring Reliability Symposium, 'Reliability - Investing in the Future'.*, Boxborough, MA, USA, 2002.

[7] B. Agard and A. Kusiak, "Data Mining for Selection of Manufacturing Processes," *Springer,* pp. 1159-1166, 2005.

[8] K. Wang, "Applying data mining to manufacturing: the nature and implications," *Springer Science+Business Media, LLC,* pp. 487-495, 2007.

[9] D. A. T. Mathew, B. M. Barkachary, N. Selvaraj and c. s. p. Rao, "Application of Data Mining in Assembly Planning," in *3rd International & 24th All India Manufacturing Technology, Design and Research (AIMTDR) Conference 2010*, 2010.

[10] W. Kanarkard, D. Chetchotsak, D. Hormdee, R. Adams and N. Davey, "Data Mining to Recognize Fail Parts in Manufacturing Process," *ECTI TRANSACTIONS ON ELECTRICAL ENG., ELECTRONICS, AND COMMUNICATIONS ,* vol. 7, no. 2, pp. 122-127, 2009.

[11] B. V. Stein, M. V. Leeuwen, H. Wang, S. Purr, S. Kreissl, J. Meinhardt and T. Bäck, Towards Data Driven Process Control in Manufacturing Car Body Parts, Las Vegas, NV, USA, 2016.

[12] R. Hamideh, D. Jean-Yves and H. Lazhar, "Review of data mining applications for quality assessment in manufacturing industry: Support Vector Machines," *International Journal of Metrology and Quality Engineering,* vol. 401, no. 6, pp. 1-18, 2015.

[13] J. Harding, M. Shahbaz and A. Kusiak, "Data Mining in Manufacturing:A Review," *Journal of Manufacturing Science and Engineering,* vol. 128, no. 4, pp. 969-976, 2006.

[14] A. Kusiak, "Data mining: manufacturing and service applications," *International Journal of Production Research,* vol. 44, no. 18-19, pp. 4175-4191, 2006.

[15] Z. Ge, B. huang, Z. Song and S. X. Ding, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *Data-Driven Monitoring, Fault Diagnosis,* vol. 5, pp. 20590 - 20616, 2017.

[16] L. Rokach and O. Maimon, "Data mining for improving the quality of manufacturing: a feature set decomposition approach," *Journal of Intelligent Manufacturing,* vol. 17, no. 3, pp. 285-299, 2006.

[17] S.-g. He, Z. He, G. A. Wang and L. Li, "Quality Improvement using Data Mining in Manufacturing Processes," in *Data Mining and Knowledge Discovery in Real Life Applications*, I-Tech, Vienna, Austria, 2009, pp. 357-372.

[18] M. Alnoukari and W. Alhussan, "Using Data Mining Techniques for Predicting Future Car market Demand; DCX Case Study," *IEEE International Conference on Information & Communication Technologies: From Theory to Applications,* pp. 929-930, 2008.

[19] D. Janikova, P. Vazan, T. Pavol, Z. Cervenanska and M. Kebisek, "Using data mining methods for manufacturing process control," *ScienceDirect,* pp. 6178-6183, 2017.

[20] F. Saleheen, M. H. Miraz, M. M. Habib and Z. Hanafi, "Challenges of Warehouse Operations: A Case Study in Retail Supermarket," *Int. J Sup. Chain. Mgt* , vol. 3, no. 4, pp. 63-67, 2014.

[21] M. Kłodawski, M. Jacyna, K. Lewczuk and M. Wasiak, "The Issues of Selection Warehouse Process Strategies," *10th International Scientific Conference Transbaltica 2017,* vol. 187, pp. 451-457, 2017.

[22] L. Lefebvre, É. Lefebvre, Y. Bendavid, S. Wamba and H. Boeck, "The potential of RFID in warehousing activities in a retail industry supply chain," *Journal on Chain and Network Science,* vol. 5, no. 2, pp. 101-110, 2007.

[23] Ramaa.A, K.N.Subramanya and T.M.Rangaswamy, "Impact of Warehouse Management System in a Supply Chain," *International Journal of Computer Applications,* vol. 54, no. 1, pp. 0975-8887, 2012.

[24] D. A. M. A. Kamali, "Smart Warehouse vs. Traditional Warehouse - Review," in *CiiT International Journal of Automation and Autonomous System,*, 2019.

[25] H. Walaa, M. Noha and E. Hesham, "Towards a Smart Warehouse Management System," in *Proceedings of the International Conference on Industrial Engineering and Operations Management* , Washington DC, USA, 2018.

[26] J. Karasek, "An Overview of Warehouse Optimization," *International Journal of Advances in Telecommunications Electrotechnics, Signals and Systems,* vol. 2, no. 3, pp. 1-7, 2013.

[27] S. S. Ghuman, "A Review of Data Mining Techniques," *International Journal of Computer Science and Mobile Computing,* vol. 3, no. 4, pp. 1401-1406, 2014.

[28] R. R. Deshmukh and V. Wangikar, "Data Cleaning: Current Approaches and Issues," in *IEEE International Conference on Knowledge Engineering*, Aurangabad, 2011.

BIOGRAPHIES

Ms. C S Nagashree is pursuing Post-Graduation (MTech) in Department of Computer Science and Engineering at RV College of Engineering, Bangalore. Research interests include Data Mining and Machine Learning.