JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Production Planning and Scheduling - A Data Mining Approach

[1]Nikita D. Deshpande, [2]M.S. Kalas

[1]Reaserch Scholar, [2]Professor and Head Dept
[1] Computer Science (Data Science),
[1]Kolhapur Institute of Technology's College of Engineering, Kolhapur, India

Abstract: According to experience, raw material management is a major operational difficulty in the business. The importance of raw materials to the efficient operation of a manufacturing organization cannot be emphasized; the right quality and quantity are critical. Planning and production include determining the level of activity, turn-over, and final profit in a corporation, as well as minimum and maximum stock levels. Raw material management in a manufacturing organization requires specific care and scrutiny in order to achieve uninterrupted production cycles and better operational performance. Maintaining an acceptable stock level can also enhance the amount of available operating capital that can be put to better use. Material management is defined as the coordination of efforts (planning, managing, organizing, and directing) aimed at achieving efficiency in a manufacturing organization's procurement, transportation, stocking, and utilization of inputs. The effectiveness and efficiency of material management have a direct impact on the organization's overall success. This study provides a literature review on data mining definitions as well as a categorization of existing techniques to using data mining to manage production complexity in order to assist manufacturing organizations in implementing data mining.

***IndexTerms* - Data Mining, Hierarchical Clustering Algorithm and Association Rule.**

I. INTRODUCTION

For translating data into meaningful knowledge, data mining is a natural answer. For a variety of applications, the retrieved knowledge can be utilized to model, classify, and make predictions. The essential data for analysis can be obtained during the normal operation of the manufacturing process being researched, which is a primary benefit of data mining over conventional experimental techniques.

As a result, it is rarely essential to dedicate machines or processes just to data collection. Data mining (DM) and knowledge discovery in databases (KDD) have become critical methods for achieving the goal of intelligent and automated data analysis. Data mining is a step in the KDD process that entails using specific algorithms to extract patterns (models) from large amounts of data.

Data mining has been used in a variety of industrial and logistics fields over the years, but only to a limited extent.

This research focuses on the use of data mining techniques or algorithms to the manufacturing business, namely in the areas of stock management and delivery. The Hierarchical Clustering Algorithm and Association Rule are Data Mining Algorithms/Functions that can be used for Production planning and scheduling. This study looks at how data mining techniques or algorithms can be applied to the manufacturing business, namely stock management and delivery.

II. RELATED WORK

In this section we will be discussing on the literature (of various papers) dealing with knowledge discovery and data mining applications in the broad domain of manufacturing with an special emphasis on the type of functions to be performed on data.

2.1 **As per author** [12], the goal of this research is predicting the systems performance used for producing smart materials, based on input parameter changes.

After analyzing various methods, the method they chose is neural networks for deployment of the latest data in the concluding phases

2.1.1 Limitations

Only KRIs and KPIs were utilized to ensure a Manufacturing process, but no analysis of the step before that was mentioned in the paper. The study focuses mostly on product quality, with little mention of production planning and timely delivery. Process is appropriate for systems requiring a significant financial commitment, as well as high-end machinery and manpower. Only necessary KPIs and KRIs are used to forecast the manufacturing process and system behavior.

There was no research done for companies that did not even have KRIs and KPIs or for start-up industries. A clear Future scope point highlighted in this study - All research in future can focus on studying additional to the point boundary and applying the outcomes obtained into actual systems.

2.2 **As per authors** [13], The Study focuses on floor smartification through a roll-out of sensors, which increase communication between machines and employees, could be observed - a trend commonly known as Industrie 4.0.
The paper focuses on following points: - to create a common understanding for the terms DM, ML, AI and statistics from an application point of view (i.e. production), and to support production managers to identify relevant use-cases for managing production complexity through DM.
Proposed work /Algorithms used in this paper: Application of DM methods for managing production complexity by: Evaluating and preventing new variants using Clustering and Modulization and standardization using association analysis

2.2.1 Limitations

Rather than focusing on a single issue in detail, this paper provides a broad overview of numerous algorithms and elements that affect a business. In the findings and conclusion, it was claimed that only a few DM applications were directly related to production difficulty. The study's major focus is not just on DM algorithms, but also on ML and AI in conjunction with DM.

Clustering is used here based on their production requirements rather than the client's priorities. Their main goal is to find common parts and modules that can be used to reduce product variety. Rather than working on the lead time for raw materials to arrive in the industry, decision trees are being used to anticipate lead times based on individual or staff performance.

2.3 **As per authors** [14], in this study, a data mining approach is applied to extract knowledge from a data set. The extracted knowledge is helpful for the prediction and prevention of manufacturing faults in engines.
Proposed work/ Algorithms used: Decision Tables and Decision Trees

2.3.1 Limitations

Mostly concerned with determining the source of production machine errors that result in only poor product quality. Data mining applications in raw material management and product delivery are not explored.

2.4. **As per authors** [15], this paper focuses on demonstrating the relevancy of data mining to manufacturing industry. Proposed Suggestion: The CRISP-DM and SEMMA methodologies are most widely used by the data mining community and CRISP-DM is easier to use than SEMMA as it provides detailed neutral guidelines that can be used by any novice in the data mining field.
This paper has mentioned range of data mining applications in areas:- Engineering Design, Manufacturing Systems, Decision Support Systems, Shop Floor Control and Layout, Customer Relationship Management, Data Mining in Maintenance.

2.4.1 Limitations

Rather than focusing on a single issue in detail, this paper provides a broad overview of numerous algorithms and elements that affect a business. Scheduling and Shop floor control is clearly mentioned in their Future Scope.

2.5 **As per authors** [16]**,** this paper focuses on decomposition methodology that is capable of dealing with the data characteristics associated with quality improvement. Proposed Work: A new algorithm called BOW (Breadth-Oblivious-Wrapper) has been developed. This algorithm performs a breadth first search while using a new F-measure splitting criterion for multiple oblivious trees. The new algorithm was tested on various real-world manufacturing datasets, specifically the food processing industry and integrated circuit fabrication.
The obtained results have been compared to other methods, indicating the superiority of the proposed methodology.

2.5.1 Limitations

The newly developed algorithm is mostly useful for determining product quality. The new algorithm is based on a case study of the dairy products industry, which is distinct from other manufacturing businesses in terms of expiration dates for edible items and other factors. No mention about Planning and Scheduling even if it’s a dairy Industry.

2.6 **As per authors** [17]**,** this review reveals the progressive applications and existing gaps identified in the context of data mining in manufacturing.
Proposed Suggestion: The major data mining functions to be performed include characterization and description, association, classification, prediction, clustering and evolution analysis.

2.6.1 Limitations

Rather than focusing on a single issue in detail, this paper provides a broad overview of numerous algorithms and elements that affect a business. No process mentioned about Raw Material management and it’s delivery to customers.

III. PROPOSED WORK

We came up with the concept to build a bridge by creating software that will meet the needs of a startup or newly built company after evaluating the papers and examining their constraints.

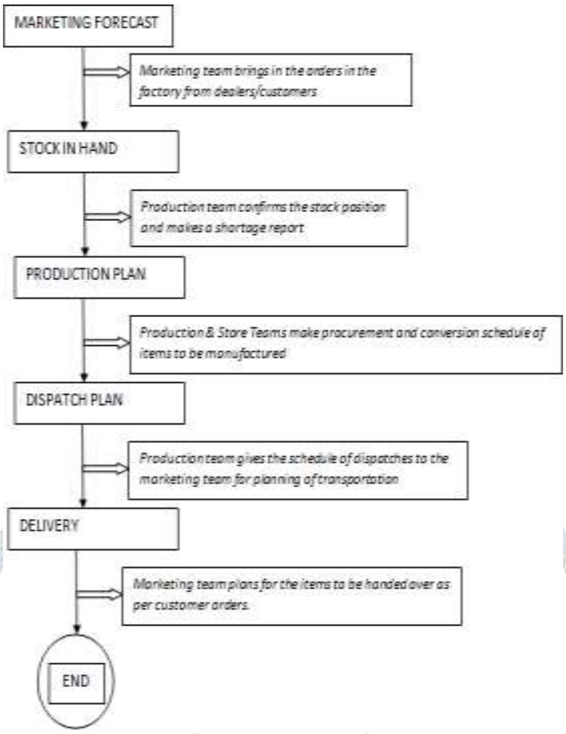

Figure1. Workflow in Manufacturing Systems Environment

Objective - This study focuses on using **data mining techniques or algorithms to the manufacturing industry**, particularly **stock management and delivery.**

Algorithms Applicable For This Project Work - Two Algorithms of Data Mining are applicable for this project.

3.1 Divisive Clustering Algorithm

A top-down method is also known as divisive clustering. This approach also eliminates the need to define the number of clusters ahead of time. Top-down clustering necessitates a method for breaking a cluster that contains all of the data and then recursively splitting clusters until all of the data has been split into singletons.

3.1.1 Steps of Divisive Clustering Algorithm:

1. Initially, all points in the dataset belong to one single cluster.
2. Partition the cluster into two least similar clusters.
3. Precede recursively to form new clusters until the desired number of clusters is obtained.

3.1.2 Divisive Clustering Algorithm:

Start
given a dataset (d1, d2, d3, ...dN) of size N
at the top we have all data in one cluster
the cluster is split using a flat clustering method e.g. K-Means etc
repeat
choose the best cluster among all the clusters to split
split that cluster by the flat clustering algorithm
until each data is in its own singleton cluster
End

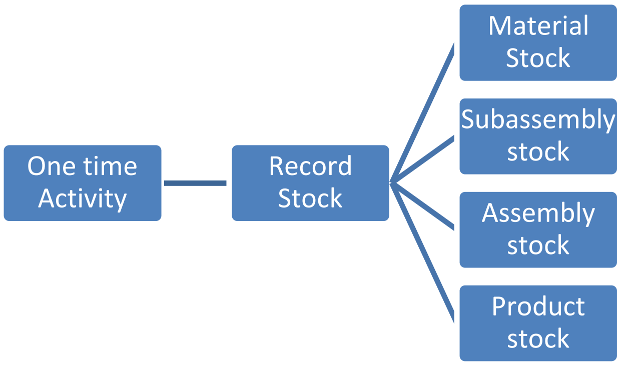

Figure2. Application of Hierarchical Divisive Clustering in manufacturing system

3.1.3 *Why we chose to split the cluster?* Data on this page was too large to maintain and hence clustering of this page is needed for knowledge discovery.

3.1.4 *How to split the above-chosen cluster?* Once we have decided to split which cluster, then the question arises on how to split the chosen cluster into 4 clusters. The criterion for choosing the clusters to split at each step is based on the optimal value of an objective function. This objective function could be "any function that reflects the investigator's purpose." In this case, it depends upon stock type parameter.

3.1.5 *How to handle the noise or outlier?* Due to the presence of outlier or noise, can result to form a new cluster of its own. To handle the noise in the dataset using a threshold to determine the termination criterion that means does not generate clusters that do not belong to any criteria.

3.1.6 *Divisive Clustering vs. Hierarchical Agglomerative:*

Divisive clustering is more complex as compared to agglomerative clustering, as in the case of divisive clustering we need a flat clustering method as “subroutine” to split each cluster until we have each data having its own singleton cluster.

Divisive clustering is more efficient if we do not generate a complete hierarchy all the way down to individual data leaves. The time complexity of a naive agglomerative clustering is O(n3) because we exhaustively scan the N x N matrix dist_mat for the lowest distance in each of N-1 iterations. Using priority queue data structure we can reduce this complexity to O(n2logn). By using some more optimizations it can be brought down to O(n2). Whereas for divisive clustering given a fixed number of top levels, using an efficient flat algorithm likes K-Means, divisive algorithms are linear in the number of patterns and clusters.

A divisive algorithm is also more accurate. Agglomerative clustering makes decisions by considering the local patterns or neighbor points without initially taking into account the global distribution of data. These early decisions cannot be undone. Whereas divisive clustering takes into consideration the global distribution of data when making top-level partitioning decisions.

3.2 Support, Confidence and Association Rule Mining

If we think of the total set of items available in our set (sold at factory, at an online website, or something else altogether, such as transactions for fraud detection analysis), then each item can be represented by a Boolean variable, representing whether or not the item is present within a given "unit." Each unit is then simply a Boolean vector, possibly quite lengthy dependent on the number of available items. A dataset would then be the resulting matrix of all possible unit vectors.

This collection of Boolean unit vectors are then analyzed for associations, patterns, correlations, or whatever it is you would like to call these relationships. One of the most common ways to represent these patterns is via association rules, a single example of which is given below:

An association rule has 2 parts: an antecedent (if) and a consequent (then).

“If a customer places an order within the range of the available stock for a day, she’s 80% likely of getting the delivery of the requested products.” ***[support = 25%, confidence = 60%]***

Support(s) – The number of transactions that include items in the {X} and {Y} parts of the rule as a percentage of the total number of transaction. It is a measure of how frequently the collection of items occurs together as a percentage of all transactions. It is a measure of absolute frequency. In the above example, the support of 25% indicates that, in our finite dataset, customer order within available stock and its timely delivery matches together in 25% of all transactions.

Confidence(c)–

It is the ratio of the no of transactions that includes all items in {B} as well as the no of transactions that includes all items in {A} to the no of transactions that includes all items in {A}. It is a measure of *correlative frequency*. In the above example, the confidence of 60% indicates that 60% of those who placed an order within available stock have got the delivery immediately.

In a given application, association rules are generally generated within the bounds of some predefined minimum threshold for both confidence and support, and rules are only considered interesting and insightful if they meet these minimum thresholds.

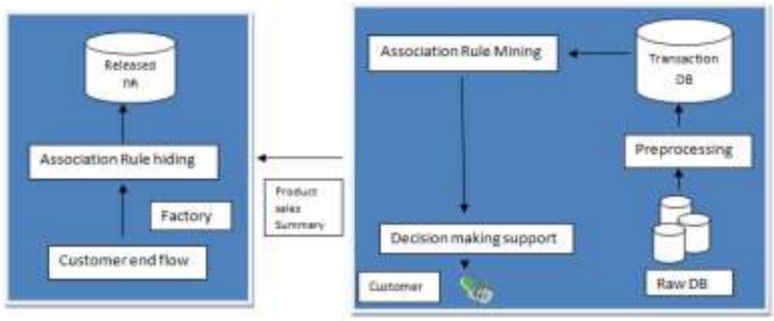

Figure3. Application of Association Mining Rule in manufacturing system

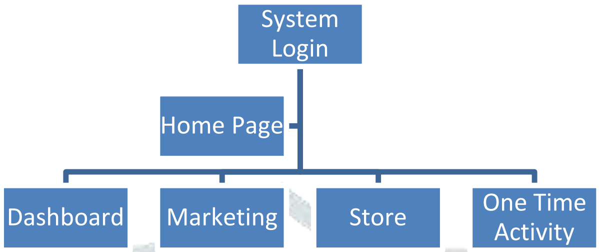

Figure4. Manufacturing Systems Environment using DM techniques

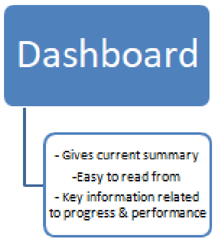

Figure4.1. Use case diagram of Dashboard module
Figure4.2. Use case diagram of marketing module

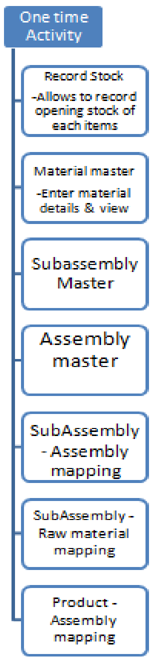

Figure4.3. Use case diagram of store module
Figure4.4. Use case diagram of one time Activity module

IV. CONCLUSION AND FUTURE SCOPE

DM can be a useful tool for managing complexity at the intersection of increased production complexity due to fluctuating market needs and enormous amounts of production data. So far, the majority of DM applications in production management have been tied to quality management. There are just handful DM applications that are directly tied to the complexity of production. Other applications of DM in other sectors of production management, on the other hand, are quite effective at managing production complexity. We've shown some of these applications, and we hope to expand the categories in future work to present a comprehensive framework of DM, as well as other machine learning and artificial intelligence (ML and AI) applications that can cover all essential areas of managing production complexity.

REFERENCES

[1] Fayyad, U. M., Piatetsky-Shapiro. G., Smyth, P., and Uthuruswamy, R., “Eds, Advances in Knowledge Discovery and Data Mining”, Menlo Park, CA: AAAI/MIT Press, 1996.
[2] Macgarry, K., “A survey of interestingness measures for knowledge discovery”,The knowledge engineering review, vol. 20, No. 1, pp 39-61, 2005.
[3]. Mitra S., Pal S.K., and Mitra P., “Data Mining in Soft Computing Framework: A Survey ”, IEEE Transaction on Neural Networks, Vol 13, No.1, pp3-14, 2002.
[4]Zhang, D. and Zhou, L., “Discovering Golden Nuggets: Data Mining in Financial Applications”, IEEE Transactions on System Man and Cyberntics- Part C Applications and Revies, Vol. 34, No. 4, pp513-521, 2004.
[5] Harding, J A, Shahbaz, M, Srinivas and Kusiak, A, "Data Mining in Manufacturing: A Review", American Society of Mechanical Engineers (ASME): Journal of Manufacturing Science and Engineering, 128 (4), pp. 969-976, 2006.
[6]. Special Issue: Data Mining and Applications in Engineering Design, Manufacturing and Logistics, International Journal of Production Research, Guest editors, Feng J. C., X., and Kusiak A, vol44, No. 14, 2006.
[7] Han, J., and Kamber, M., “Data Mining Concepts and Techniques”, MORGAN KAUFMANN PUBLISHER, 2001.
[8] Pham , D. T., and Afify, A. A., “Machine Learning Techniques and their Applications in Manufacturing”, Proc. IMechE, Journal of Engineering Manufacture: Part B, Vol. 219, pp395-412, 2005 [9] Fayyad, U. M., Piatetsky-Shapiro. G., Smyth, P., and Uthuruswamy, R., “Eds, Advances in Knowledge Discovery and Data Mining”, Menlo Park, CA: AAAI/MIT Press, 1996.
[10] Macgarry, K., “A survey of interestingness measures for knowledge discovery”,The knowledge engineering review, vol. 20, No. 1, pp 39-61, 2005.
[11] Fayyad, U., Shapiro G. P., and Smyth. P., “The KDD process for extracting useful knowledge from volume of data”, Commin. ACM, Vol. 39, pp 27-34, 1996.
[12] Nymphea Saraf Sandhu, A.K.Upadhyay, Sanjiv Sharma - Data Mining Application in Process Control of Smart Material Manufacturing 2020.
[13] Günther Schuha, Gunther Reinhartb, Jan-Philipp Protea, Frederick Sauermanna, Julia Horsthoferb, Florian Oppolzera, Dino Knollb – Data Mining Definitions and Applications for the Management of Production Complexity 2019.
[14] V.K. Jha and R.K. Singh - Application of Data Mining in Manufacturing Industry 2011.
[15] J. A. Harding, M. Shahbaz, Srinivas, A. Kusiak - Data Mining in Manufacturing - A Review - 2006.
[16] Lior Rokach , Oded Maimon - Data mining for improving the quality of manufacturing: a feature set decomposition approach 2006.
[17] A. K. Choudhary , M. K. Tiwari and J. A. Harding - Data Mining in Manufacturing: A Review Based on the Kind of Knowledge (2006).

AUTHORS PROFILE

Nikita D. Deshpande, have been graduate from D. Y. Patil College of Engineering and Technology,Kolhapur, in 2015. Currently, she is a Research Scholar and pursuing MTech in Computer Science, with specialization in Data Science at KIT’S College of Engg, Kolhapur. Her areas of interest are Data mining, system programming.

Mamata S. Kalas, have been graduate from University Of Mysore, in 1993, From B.I.E.T, Davangere, started her professional career there itself. In 1995, came to kolhapur, after her marriage and since then she has worked as lecturer at D.Y.Patil’s college of Engg., Bharati Vidyapeeth’s College Of Engg., Kolhapur. She did M.Tech(CST) Graduate and her dissertation work is based on image segmentation using parametric distributional clustering. She has been awarded with M.TECH (CST) from Shivaji University, Kolhapur of Maharashtra in June 2009. In November 2019, she received a Ph.D. in computer science and engineering from Walchand College of Engineering, Shivaji University's Research Center, under the supervision of Dr.B.F.Momin. She is currently working as Professor and Head,Dept of CSE at KIT’S College of Engg, Kolhapur. She is in her credit, 27 years of teaching experience. Ten Papers presented for international conferences, three papers presented for national conferences, seven papers published in international journals. Her areas of interest are pattern recognition and artificial intelligence, computer architecture, system programming.