International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

To Development Manufacturing and Education using Data Mining: A Review

Aye Pwint Phyu, Khaing Khaing Wai

Department of Information Technology Support and Maintenance,
University of Computer Studies, Mandalay, Myanmar

How to cite this paper: Aye Pwint Phyu | Khaing Khaing Wai "To Development Manufacturing and Education using Data Mining: A Review" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.2168-2173, **https://doi.org/10.31142/ijtsrd27910**

IJTSRD27910

ABSTRACT

In modern manufacturing environments, vast amounts of data are collected in database management systems and data warehouses from all involved areas. Data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. It is the extraction of information from huge volume of data or set through the use of various data mining techniques. The data mining techniques like clustering, classification help in finding the hidden and previously unknown information from the database. In addition, data mining also important role and educational sector. Educational Data Mining (EDM) is a field of analysis and research where various data mining tools and techniques are used to optimize the applications in education sector. The paper aims to analyze the enormous data from the education sector and provide solutions and reports for specific aspects of education sector such as student's performance and placements. Moreover, this paper reviews the literature dealing with knowledge discovery and data mining applications in the broad domain of manufacturing with a special emphasis on the type of functions to be performed on the data. The major data mining functions to be performed include characterization and description, association, classification, prediction, clustering and evolution analysis.

KEYWORDS: *Data Mining, EDM, Manufacturing, Literature review*

1. INTRODUCTION

In most sectors, manufacturing is extremely competitive and the financial margins that differentiate between success and failure are very tight, with most established industries needing to compete, produce and sell at a global level. To master these trans-continental challenges, a company must achieve low cost production yet still maintain highly skilled, flexible and efficient workforces who are able to consistently design and produce high quality and low cost products. In higher-wage economies, this can generally only be done through very efficient exploitation of knowledge (Harding and Popplewell 2006; Choudhary et al. 2006). In modern manufacturing, the volume of data grows at an unprecedented rate in digital manufacturing environments, using barcodes, sensors, vision systems etc.

The huge amounts of data in manufacturing databases, which contain large numbers of records, with many attributes that need to be simultaneously explored to discover useful information and knowledge, make manual analysis impractical. All these factors indicate the need for intelligent and automated data analysis methodologies, which might discover useful knowledge from data. Knowledge discovery in databases (KDD) and data mining (DM) have therefore become extremely important tools in realizing the objective of intelligent and automated data analysis. Data mining is a particular step in the process of KDD, involving the application of specific algorithms for extracting patterns (models) from data.

1.1 Data Mining for Manufacturing

Knowledge discovery in databases (KDD) and data mining (DM) have therefore become extremely important tools in realizing the objective of intelligent and automated data analysis. The additional steps in the KDD process, such as data preparation, data cleaning, data selection, incorporation of appropriate prior knowledge and proper interpretation of the results of mining, ensure that useful knowledge is derived from the data (Mitra et al. 2002). these fields provide specific data mining tools that can be used in various steps of a KDD process. Recently, with the growth of data mining technology, researchers and practitioners in various aspects of manufacturing and logistics have started applying this technology to search for hidden relationships or patterns which might be used to equip their systems with new knowledge. Early applications of data mining were mostly applied to financial applications, for example Zhang and Zhou (2004) described data mining in the context of financial applications from both technical and application perspectives. In this area, the competitive advantage gained through data mining included increased revenue, reduced cost, much improved market place responsiveness and awareness. A recent sur-vey carried out by Harding et al. (2006) and a special issue published on "data mining and applications in engineering design, manufacturing and logistics" (Feng and Kusiak 2006) clearly indicated the potential scope of data mining in these areas to achieve competitive advantages. A major advantage of data mining over other experimental techniques is that the required data for analysis can be collected during the nor-mal operation of the manufacturing process being studied. Therefore, it is generally not necessary to specially dedicate machines or processes for data collection.

1.2 Data Mining for Manufacturing Literature Review

Han and Kamber (2001) classified data mining systems based on various criteria such as *kind of database mined, kind of knowledge mined, kind of technique utilized and application areas adopted*. Pham and Afify (2005) reviewed machine learning techniques in the manufacturing domain.. Harding et al. (2006) surveyed data mining systems in different application areas of manufacturing, including some less considered areas such as manufacturing planning and shop floor control. However, in the last few years, data mining research in manufacturing has increased at an exponential rate. Han and Kamber (2001) mentioned that the kind of knowledge to be mined determines the data mining functions to be performed. Possible kinds of knowledge include *concept description (characterization and discrimination), association classification, clustering, and prediction*. The aim of this paper is therefore to consolidate the existing state-of-the art research efforts concerning the current practices in data mining applications in manufacturing based on the *kind of knowledge mined* and *the kind of technique* utilized, thereby identifying promising areas for study. The remainder of the paper is organized as follows. briefly discusses about KDD, data mining, and the kinds of knowledge that particularly occur in manufacturing contexts. Section "Concept description (characterization and discrimination) in manufacturing" will discuss concept descriptions which include characterization and discrimination in manufacturing. Classification in manufacturing is discussed in section "Classification in manufacturing," followed by clustering in manufacturing in section "Clustering in manufacturing". Section "Prediction in manufacturing" discusses prediction in manufacturing, and association in manufacturing is discussed in section "Association in manufacturing". Details of our novel text mining approach are given in section "Detailed analysis and discussion: a text mining perspective on reviewed literature" and this is followed by conclusions in section "Conclusion".

1.2.1 KDD, data mining and knowledge types

KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad et al. 1996a). The KDD process is interactive and iterative involving more or less the following steps (Fayyad et al. 1996b; Mitra et al. 2002).

- -Understanding the manufacturing domain
- -Collecting the targeted data
- -Data cleaning, pre-processing and transformation
- Data integration
- -Choosing the functions of data mining
- -Choosing the appropriate data mining algorithm
- -Data mining
- -Interpretation and visualization
- -Implementation of discovered knowledge
- -Knowledge storage, reuse and integration into the manufacturing system

Data mining is an interdisciplinary field with the general goal of predicting outcomes and uncovering relationships in data. It makes use of automated tools and techniques, employing sophisticated algorithms to discover hidden patterns, associations, anomalies and/or structure from large amounts of data stored in a data warehouse or other information repositories. In the context of manufacturing, two high level primary goals of data mining are prediction and description. Descriptive data mining focuses on discovering interesting patterns to describe the data. Predictive data mining focuses on predicting the behaviour of a model and determining future values of key variables based on existing information from available databases. The boundaries between, descriptive and predictive data mining are not sharp, e.g. some aspects of the predictive model can be descriptive, to the degree that they are understandable and vice versa. The goals of prediction and description can be achieved by using a variety of data mining tools and techniques. The next section there-fore describes a range of functions and reviews their applicability in manufacturing domains.

1.2.2 Concept Description (Characterization and Discrimination) in Manufacturing

Characterization can be used to identify the features that significantly impact the quality. Characterization provides a concise and succinct summarization of the given collection of data, while concept or class discrimination or comparison provides descriptions that compare two or more collections of data. In manufacturing contexts, these functions are basically used to understand the process. Huyet (2006) proposed an evolutionary optimization and data mining based approach to produce the knowledge of systems behaviour in a simulated job shop based production process. Assigning proper dispatching rules is an important issue in enhancing the performance measures for a flexible manufacturing system (FMS). Lee and Ng (2006) presented a hybrid case based reasoning (HyCase) system for online technical support of PC fault diagnosis. Romanowski and Nagi (1999) applied a decision tree based data mining approach on a scheduled maintenance dataset and a vibration signal dataset. Subsystems which are most responsible for low equipment availability are recognized in the scheduled maintenance data and a recommendation for preventive maintenance interval is made.

1.2.3 Classification in Manufacturing

Classification is a useful functionality in many areas of manufacturing, for example, in the semiconductor industry, defects are classified to find patterns and derive the rules for yield improvement. Online control chart pattern recognition (CCPR) is another example of classification for SPC, because unnatural patterns displayed by a control chart can be associated with specific causes that adversely impact the manufacturing process. Classification is a learning function that maps (classifies) a data item into one of several predefined categorical classes. Generally, classification is performed in two steps. In the first step, a model is built to describe a predetermined set of data classes or concepts, and this is done by analyzing the database tuples described by attributes, which collectively form the training dataset. Rokach and Maimon (2006) applied a feature set decomposition methodology for quality improvement. They developed the Breadth Oblivious Wrapper (BOW) algorithm and showed its superiority over existing tools on datasets from IC fabrication and food processing. The idea is to find the classifier that is capable of predicting the quality measure of product or batch based on its manufacturing parameters. Braha and Shmilovici (2002) presented three classification based data mining methods (decision tree induction, neural network and composite classifier) for a new laser based wafer cleaning process called advanced wafer cleaning. The purpose of the data mining based classifier is to enhance understanding of the cleaning process by categorizing the given data into a given predefined number of categorical classes and determine to which the new data belongs. A fractal dimension based classifier was proposed by Purintrapiban and Kachitvichyanukul (2003) for detection of unnatural patterns in process data. Kusiak (2002a), Kusiak (2002b) applied data mining to support decision making processes by using different data-mining algorithms to generate rules for a manufacturing system. A subset of these rules was then selected to produce a control signature for the manufacturing process where the control signature is a set of feature values or ranges that lead towards an expected output.

From this review, the major application areas where data mining tools and techniques are used for classification include fault diagnosis, quality control and condition monitoring. In order to perform the classification task, decision tree, rough set theory, hybrid neural network and other hybrid approaches have been successfully used. In hybrid approaches, Fuzzy logic is used often in combination with other techniques to deal with noise and uncertainty in the data. The next section will deal with clustering and its performance on manufacturing databases.

1.2.4 Clustering in Manufacturing

Clustering is an important data mining function that can be performed on specified manufacturing data such as order picking in logistics and supply chain. For example order picking is routine in distribution centers and before picking a large set of orders, orders are clustered into batches to accelerate the product movement within the storage zone. Clustering is also useful in the formation of cells in cellular manufacturing where it is used for the simultaneous design of the part families and machine cells.

Clustering is also known as unsupervised learning. Unlike classification (supervised learning), in clustering the class object of each data object is not known. Clustering maps a data item into one of several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models (Mitra et al. 2002; XuandWunsch2005). Sebzalli andWang (2001) applied principal component analysis and fuzzy c means clustering to a refinery catalytic and fuzzy c means clustering to a refinery catalytic and fuzzy c means clustering to a refinery catalytic and fuzzy c means clustering to a refinery catalytic process to identify operational spaces and develop operational strategies for the manufacture of desired products and to minimize the loss of product during system changeover. Kim and Ding (2005) proposed a data mining aided optimal design method for fixture layout in a four station SUV side panel assembly process. Clustering and classifications are carried out to generate a design library and design selection rules, respectively. Torkul et al. (2006) showed the outperformance of fuzzy c means clustering over crisp methods on a selected data set. Romanowski and Nagi (2001) proposed a design system which supports the feedback of data mined knowledge from life cycle data to the initial stages of the design process. Romanowski and Nagi (2005) and Romanowski and Nagi (2004) also applied a data-mining approach for forming generic bills of materials(GBOMS) entities that represent the different variants in a product family and facilitate the search for similar designs and the configurations of new variants. Lee et al. (2001) proposed an intelligent inline measurement sampling method for process excursion monitoring and control in semiconductor manufacturing. The average diagnostic accuracy of 80% showed that this hybrid model is promising for an EMI diagnostic support system. Hui and Jha (2000) investigated the application of data mining techniques to extract knowledge from the customer service database for decision support and fault diagnosis. Predictability of manufacturing processes, quality, maintenance, defects, or even within manufacturing systems is of vital importance. For example in the context of maintenance, predictions can be made about what condition maintenance will be required or how equipment will deteriorate based on the analysis of past data. Feng and Kusiak(2006), Feng et al. (2006) showed that there is no significant statistical advantage of using fivefold CV over threefold CV and or of using a two hidden layer neural network over a one hidden layer neural network for turning surface roughness data. Pasek (2006) used the rough set theory based classifier for the prediction of cutting tool wear. For tool condition monitoring Sun et al. (2005) applied a neural network for recognition of tool condition in a monitoring system. Sylvain et al. (1999) used different data mining techniques including decision trees, rough sets, regression and neural networks to predict component failure based on the data collected from the sensors of an aircraft. Their results also led to the design of preventive maintenance policies before the failure of any component. Lin and Tseng (2005) introduced a cerebellar model articulation controller (CMAC) neural network based machine performance estimation model. Tsai et al. (2006) presented a case based reasoning (CBR) system using intelligent indexing and reasoning approaches for PCB defect prediction. Knowledge elicitation is a technique that is generally used for producing rules based on human expertise. A method was developed by Browneet al. (2006) to fuse knowledge elicitation and data mining using an expert system.

1.2.5 Association in Manufacturing

Association rules mining was first introduced in 1993, and is used to identify relationships between a set of items in a database (Agrawal et al. 1993). These relationships are not based on inherent properties of the data themselves (as with functional dependencies), but rather are based on co-occurrence of the data items. In design contexts, the associations between requirements may provide additional information useful for the design. For example, technical specifications might state that a car that has two doors and a diesel engine requires a specific speed transmission. In such cases, knowing the number of cars with two doors and the number of cars with a diesel engine is not relevant whilst the number of cars with two doors and a diesel engine is useful, for example to determine the capacity of a manufacturing process. The nature of this association can be extracted by applying data mining algorithms on the database.

Agard and Kusiak (2004b) applied data mining to customer response data for its utilization in the design of product families. Jiao and Zhang (2005) developed explicit decision support to improve the product portfolio identification issue by using association rule mining from past sales and product records. This review shows that the major areas where association as a data mining function has been applied include product design, process control, mass customization, cellular design etc. Association rule mining has been applied as a dominating tool to identify the associations among variable.

2. Data Mining For EDM

Similarly; Education sector is one of the sectors where data mining is relatively new as compared to other sectors and hence it is under-utilized. The International Educational Data Mining Society defines EDM as follows: "EDM is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in" (Baker, 2015). "The EDM process converts raw data coming from educational systems into useful information that could potentially have a greater impact on educational research and practice" (Romero and Ventura, 2010). EDM by and large comprises (Baker, 2010; and Romero and Ventura, 2010) four phases:

1. Data Collection:

The first phase of EDM is to explore the interrelations between the data of educational sector using data mining techniques, i.e., classification, clustering, regression etc. This phase focuses on grouping the data and also preprocesses them for mining. Data size is enormous and hence needs a lot of preprocessing in order to obtain a desired outcome.

2. Validating Relations:

The second phase of EDM is validation of found inter-relations between data with the goal that uncertainty can be evaded. The relations are then validated based on the training dataset.

3. Predicting the Future Progress:

The third phase is to make predictions for future on the basis of validated relationships in learning environment.

4. Decision Making:

The fourth phase is utilizing the gathered information and making calculated decisions using techniques like prediction and classification.

Educational institutes use data mining techniques for purposes like analyzing and visualization of data, predicting student's performances. Data mining techniques like clustering can be used to group students based on the parameters decided by the analyst. Data mining helps in identifying unwanted behaviors provides feedback to instructors on student's performance with information to support the evaluation.

Data mining utilizes many techniques and algorithms, and they can be classified into the following categories:

➢ **Prediction:**

It aims at generating a single target attribute of the data by analyzing all the other attributes and generating patterns from them (Romero and Ventura, 2013). Types of prediction techniques are classification, clustering, etc.

➢ **Classification**:

Groups information/data into a few predefined attributes. The techniques utilized for classification are:

- Decision tree
- Naive-biased classification
- Generalized Linear Models (GLM)
- Support vector machine etc.

➢ **Clustering**:

In clustering technique, the dataset is divided into various groups, known as clusters. As per clustering phenomenon, the data point of one cluster should be more similar to other data points of same cluster and more dissimilar to data points of another cluster. There are two ways of initiation of clustering algorithm: Firstly, clustering algorithm has to be started with no prior assumption; and secondly clustering algorithm has to be started with a prior postulate.

➢ **Relationship Mining:**

It helps in finding relations between values in a data corpus and organizing them as rules. There are various relationship mining procedures such as association rule mining, sequential pattern mining, correlation and causal data mining. In EDM, relationship mining is utilized to recognize connections between the understudy's web exercises and the last outcomes and to display student's critical thinking movement successions.

➢ **Discovery with Models:**

It uses an approved model of a method utilizing expectation, grouping, or information building as a segment ahead of time examination, for example, forecast or relationship mining. It is utilized as a part of circumstances to get a kick out of the chance to recognize the connections between the understudy's history and qualities.

➢ **Outlier Detection:**

The point of outlier detection is to distinguish characteristics that are unfathomably interesting than whatever is left of information. An exception is an alternate occasion that is normally more prominent or lesser than alternate esteems in information corpus. In EDM, exception identification can be used to recognize varieties in the students or instructor's activities or practices, unpredictable learning forms, and for Distinguishing understudies with learning troubles (Dominguez *et al.,* 2010; and Baker, 2015).

2.1 Data Mining for EDM Literature Review

Pruthi and Bhatia (2015) utilized the data mining technique to predict the student's performance in the placement activity of the computer science students and also predict the company they are going to be placed in (name and type of company). They used the classification process based on the parameters like their overall result and specific student's marks. The main issue with their process was that they used a limited amount of data only available with the University for the training and testing purposes. They identified the parameters as marks in many cores IT subjects.

Dominguez *et al.* (2010) developed a process and feedback generation engine that generated feedback based on the current performance or the performance of similar class of students. They used student information, current performance and the performance history of other users as parameters to predict the performance of the student and thus provide real-time feedback for them. This is a real-time generation of the evaluation of student and hence prediction of student performance. Educational data mining methods are based on statistics, machine learning and database theory. The main activities of this area are: data mining usage for Intelligence Tutoring Systems support, analysis of education processes, visual data mining and visual education process pattern. The analysis of the scientific literature in the field of using the methods of data mining showed that this problem is interesting to many modern researchers. For example, in (Ceylan 2015) the authors propose a searching model system related to student success in the form of classifiers, each of them is learned with different dataset with hundreds of thousands of lines in relation to sections. Received classifiers would serve as an advisory system for students who want to choose courses prior the registration in the semester. In (Herlina 2017), the role of the K-Means algorithm for classifying students learning activities using e-learning was showed. This algorithm helped to form student activity and improving student abilities cluster. An approach based on minimal spanning tree for clustering e-learning resources is proposed in (Wu 2016). The developed clustering method can classify students into groups so that a homogeneous classification can increase the learning effectiveness.(Rawat 2019) justified the use of cluster analysis for classifying a new student into the corresponding class and recommending relevant courses using various evaluation metrics. In addition, global trends, dynamic environment, difficulty of the problems requiring greater efficiency, adaptability, integration and coordination of all of relevant design process and implementation of the e-learning systems.

2.2 Factors Affecting EDM

There are many factors affecting the aspects of EDM. The main issues that EDM focuses on are placement, admissions, and branch or career selection and student performances. There are many factors that affect the areas of education. Although there are many factors, almost all of them can be classified into the following factors:

➢ **Interest of Student:**

Career of an individual depends on the choices he makes. These choices are above averagely influenced by the interest of the student towards any area. The area in which a student has interest in can help him perform better in terms of academics as well as in his corporate life. If a student chooses to pursue an occupation or academics in a topic which he is not interested in, it can lead to a difficult life as he would not be performing well. Interest also includes his habits and hobbies. For example, if a person has a hobby of traveling around, then he can choose his future in that field. Hence interest, hobbies and habits can affect above averagely all the factors of education.

➢ **College Facilities:**

College facilities are the things that student pays fee for— for better infrastructure, better faculties, library, residential facilities, food availability and other things. All this combine to make the basic need of a student for education. College is responsible for providing all these facilities along with academic knowledge which is their primary work.

➢ **Schooling**:

Children with good schooling present good academic results in higher education. They have experienced an educational environment that takes more interest in the practical view of studies. They tend to be more mature and regular in their assignments. Most of the students adapt to the learning material and methodology quickly with ease. Similarly, schooling, medium and tutoring are imperative as students with English medium foundation generally make more inquiries amid showing learning process. These students who are actively participating in the classrooms activity tend to have a strong base of technical and nontechnical skillset which helps them in placement activities. Such parameters as medium of schooling and student's skillset help in predicting the performance of students in the academics.

➢ **College Reputation:**

Google's CEO is from IIT and even Microsoft's CEO is from Manipal University. Why not from a normal institute? Yes, that is due to the reputation which these institutions have made. Hence, even if a student is an above average student from a local, non-reputed college, no MNC is going to offer a job directly via college campus. That is just because the college is not reputed.

3. Discussion

The reviewed literature shows that there is a rapid growth in the application of data mining in manufacturing, particularly in the semiconductor industry. In this research, we have briefly discussed data mining concept and its techniques for development of knowledge management in organizations. The next section discusses the text mining experiments undertaken using the abstract and keywords of the 150 published works reviewed in this paper.

➢ Knowledge discovery in text and text mining applications on the literature review. Following the definition of KDD by Fayyad et al. (1996a), Karanikas and Theodoulidis (2002) defined KDT as "the non trial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in unstructured data". Text Mining (TM) is also a step in the KDT process consisting of particular data mining and natural language processing algorithms that under certain computational efficiency and limitations produce a particular enumeration of patterns over a set of unstructured textual data. KDT in reviewed literature mainly consists of three steps as follows:

A. Abstract and keyword collection:

In our experiments, the abstracts and key words of the literature reviewed in this paper have been collected. Where necessary, additional key words have also been identified from the papers and added to the abstract for text mining. This is important as the published abstracts often did not include full details of the type of data mining function (s) and areas of application discussed in the paper.

B. Retrieving and pre-processing documents:

Abstracts have only been taken from papers which deploy data mining methodology to solve problems of manufacturing. The additional key words have been identified based on knowledge area, function performed and technique used. The major knowledge areas examined include manufacturing system, quality control, fault diagnosis, maintenance, job shop, yield improvement, manufacturing process, fault diagnosis, product design, production control, and supply chain management. Similarly, the functions considered include concept description, classification, clustering, prediction and association. Major techniques used include rough set theory, decision tree, statistics, neural network, association rule, fuzzy c means clustering, and regression analysis and hybrid algorithms. In this context, the term "hybrid algorithm" indicates that either a group of algorithms have been used in combination to solve a particular problem, or a group of algorithms have been used at different stages of data mining.

C. Text mining:

For the current purpose, text analysis and link analysis were used to extract patterns, trends, useful knowledge and meet the listed benefits. The text mining was performed as an automatic process with manual interventions during the pre-processing stage. Poly analyst, which is one of the leading data/text mining software package in the market was used for this purpose. All the results shown and interpretations made were automatically generated using this software. The following subsections describe how the abovementioned objectives were achieved. Equally, job description mining can reveal actionable insight for students, employers and the institution. The institution can provide students with a better understanding of co-op opportunities in various disciplines and therefore help them select the right academic program and career. Additionally, the institution may use frequently appearing words and the clustering of jobs in various disciplines to produce more effective promotional material for its co-op programs and to help attract strong students. Furthermore, students can find out what types of jobs are available to them and what soft and technical skills are required. In particular, clustering can be used to segment the job descriptions to make it easier for students to find jobs they are interested in and institutions can align their curricula with job market needs.

4. CONCLUSION

Knowledge discovery and data mining have created new intelligent tools for extracting useful information and knowledge automatically from manufacturing databases. The present article provides a survey of the available literature on data mining applications in manufacturing with a special emphasis on the kind of knowledge mined. The types of knowledge identified indicate the major data mining functions to be performed include characterization and description, association, classification, prediction, clustering in data. This paper reviewed A novel text mining approach has been applied on the reviewed literature to identify the popular and successful research tools and existing research gaps, examine the under looked and overlooked areas, identify good practices in data mining in manufacturing and some key features unknown to data mining practitioners. EDM and manufacturing for using data mining as an area of research. The paper discussed various techniques, factors and applications of EDM and manufacturing. There are many factors that affect the aspects of EDM. The paper highlighted some of them and also compared many of them based on their impact on placement outcomes, academic performance, and college and branch selection.

REFERENCES

[1] Agard, B., & Kusiak, A. (2004a). Data mining for subassembly selection. *Journal of Manufacturing Science and Engineering, 126*, 627– 631. doi:10.1115/1.1763182.

[2] A. K. Choudhary. J. A. Harding. M. K. Tiwari. Data mining in manufacturing: a review based on the kind of knowledge, J Intell Manu(2009)20:501-521 DOI 10.1007/s10845-008-0145-x.

[3] Agard, B., & Kusiak, A. (2004b). Data-mining based methodology for the design of product families. *International Journal of Production Research, 42*, 15, 2955–2969.Agrawal, R., Imielinski, T.,&Swami, A. N. (1993).Mining association

[4] Rules between sets of items in large databases. In *Proceedings of the1993 ACM SIGMOD International Conference on Management of Data* (pp. 207–216). Washington, D.C., May 1993.Backus, P., Janakiram, M., Mowzoon, S., Runger, G. C., & Bhargava, A. (2006). Factory cycle time prediction with a data-mining approach. *IEEE Transactions on Semiconductor Manufacturing, 19*(2). doi:10.1109/TSM.2006.873400.

[5] Batanov, D., Nagarur, N., & Nitikhumkasem, P. (1993). Expert—MM: A knowledge based system for maintenance management. *Artificial Intelligence in Engineering, 8*, 283–291. doi:10.1016/0954-1810(93)90012-5.

[6] Belz, R., & Mertens, P. (1996). Combining knowledge based systems and simulation to solve rescheduling problems. *Decision Support Systems, 17*, 141–157. doi:10.1016/0167-9236(95)00029-1.

[7] Baker R (2010), "Data Mining for Education", in B McGaw, P Peterson and

[8] E Baker (Eds.), *International Encyclopedia of Education*, 3rd Edition, Vol. 7, pp. 112-118, Elsevier, Oxford, UK. Baker R (2015), *Educational Data Mining and Learning Analytics,* Springer International Publishing.

[9] Blagojević M and Micić Ž (2013), "A Web-Based Intelligent Report e-Learning System Using Data Mining Techniques", *Computers & Electrical Engineering,* Vol. 39, No. 2, pp. 465-474, Elsevier BV.

[10] Gafarova, L. M., Zavyalova, I. G., Mustafin, N. N. (2015) 'On the features of the application of the Pearson consensus criterion *X*2', *ESGI*, no. 4(8), pp. 63-67.

[11] Gorlushkina, N. N., Kotsyuba, I. Yu., Khlopotov, M. V. (2015) 'Tasks and methods of intellectual analysis of educational data for decision support], *GTR*, no. 1, pp. 472-482.

[12] Hanna, M. (2004) 'Data Mining in the e-learning domain', *Campus-Wide Information Systems*, no. 21(1), pp. 29-34.

[13] He, W. (2013) 'A survey of security risks of mobile social media through blog mining and an extensive literature search', *Information Management and Computer Security*, no. 21(5), pp. 381-400.

[14] Herlina, Latipa Sari, Dewi, Suranti Mrs.and Leni, Natalia Zulita (2017) "Implementation of k-means clustering method for electronic learning model" // *International Conference on Information and Communication Technology (IconICT) IOP Publishing IOP Conf. Series: Journal of Physics: Conference Series*, Volume 930.

[15] Hussain, M. et al. (2018) "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores" // *Computational Intelligence and Neuroscience,* vol 2018.

[16] Ilyina, T. S., Zakharov, N. Yu. (2016) 'Management of educational risks', *Vestnik VGUIT*, no. 4(70), pp. 290-295.

[17] Kamisli, Ozturk, Z., Erzurum, Cicek, Z. I. and Ergul, Z (2017), "Sentiment Analysis an Application to Anadolu University," *Acta Physica Polonica A*, vol. 132, no. 3, pp. 753–755.

[18] *Smeet M Thakrar*, Navjyotsinh Jadeja** and Nikunj Vadher***,Educational Data Mining:A Review*, The IUP Journal of Information Technology, Vol. XIV, No. 1, 2018