Mining Industrial Engineered Data of Apparel Industry: A Proposed Methodology

Md Shamsur Rahim
Department of Computer Science
American International University-Bangladesh
Dhaka, Bangladesh

Mashiour Rahman
Department of Computer Science
American International University-Bangladesh
Dhaka, Bangladesh

Azm Ehtesham Chowdhury
Department of Computer Science
American International University-Bangladesh
Dhaka, Bangladesh

ABSTRACT

Data mining and knowledge discovery play a significant role in the field of industrial engineering as the vast amount of generated data help to reveal previously unknown interesting patterns and knowledge. Many industries have already adopted data mining techniques for better productivity by following clear and concise methodologies. But apparel industries are yet waiting to adopt data mining techniques due to the absence of a data mining methodology which meets the particular requirements and business objectives. The objective of this research is to develop such a mining methodology that will be able to fulfill the requirements of apparel industries. This research paper has proposed a methodology for mining industrial engineered manufacturing data of apparel industries. This methodology covers from analysis of apparel industrys manufacturing unit to implement and evaluate mining model. It also includes the analysis of different departments in manufacturing to identify correlation and dependencies among the departments which is absent in the existing methodologies. Furthermore, the proposed methodology provides a clear and unambiguous transitions among different steps to perform data mining.

General Terms

Data Mining, Industrial Engineering

Keywords

Apparel industry, industrial engineering, data mining, data mining methodology, manufacturing data

1. INTRODUCTION

To differentiate a manufacturing enterprise from its competitor and to achieve the best of its ability in efficiency and effectively, knowledge is the most valuable asset for the business. Although knowledge exists in every aspect of business functions but knowledge can be difficult to identify, extract and manage due to the absence of following a proper methodology. Recognizing the importance of data mining methodology, researchers are vigorously developing new data mining methodologies [1] and thousands of businesses are applying data mining tools and techniques every day in order to manipulate, identify, and extract useful knowledge from the stored records. Like other businesses, apparel industries are also using data mining techniques on predicting size chart [2], sales forecasting and marketing but no techniques have been applied to industrial engineered (IE) manufacturing data till now. As a result, apparel industries around the world are still facing several difficulties like low productivity, longer production lead time, poor line balancing, high rework and rejection, changes in style [3]. All these issues reduce the efficiency of productivity and cause the loss of million dollars. To solve the issues behind lower productivity in the apparel industry, analysis of data from different departments can play a vital role.

Applying data mining tools and techniques on IE data of apparel industry to get the insights of productivity chain is a difficult task. Because the practice of implementing these data mining techniques is inconsistent since the software vendors focus on specific business application. But the industrial engineers require an application-neural methodology to develop good data mining strategies [4]. As apparel industry is a labor incentive industry and the industrial engineering practices differ from others, so a specific, application-neutral data mining methodology is required in order to reveal hidden patterns or parameters that control manufacturing processes or to determine and improve the productivity and efficiency that may help to cut down the million dollars lose. According to the data provided in 2012, the worth of global garment industry is 1.7 trillion US Dollar (1.3 trillion Euro) [5]. This means a tiny improvement by applying data mining techniques on IE data of apparel industries may have an enormous impact. According to Malcolm Newbery [6], if any business can able to achieve a 2.5% improvement in the use of fabric and a 10% improvement in the cost of sewing labor then profit turns into 8% from 5% which results in a massive 60% improvement in profit for the owners. Data analytics can be a game changer if it can be used properly in the apparel industries.

But the application of data mining in manufacturing practice is rare in comparison with the various successful data mining applications

in the service industry, for instance in banking, telecommunications or retailing [7]. As a result, a methodology for mining IE data of apparel industry has not been developed yet.

Developing a data mining methodology for apparel industry is a challenging task because of the various engineering practice among industries. These various engineering practices produce heterogeneous data that causes the different objectives and goals from industry to industry. In this study, the authors have developed a new data mining methodology for apparel industry after analyzing different data mining methodologies. The authors have executed the proposed methodology on real world scenario and discussed the result in Section V.

The following sections discussed in this paper are structured as follows: the manufacturing process and process flow of apparel industry in section II, analysis of previous works related to data mining methodology in section III, description of the collected IE data from manufacturing unit of apparel industry in section IV, description of the proposed methodology in section V, result and analysis in section VII and finally conclusion and further research direction in section VII.

2. BACKGROUND STUDY

This section discusses regarding the manufacturing process in the apparel industry. It's necessary to understand the problem in order to identify the solution.

To transform yarn into clothes, an apparel manufacturing industry contains several departments. A typical apparel industry may contain around 12 departments which can be: Merchandising, Sampling, Fabric Store, Trims and Accessories Store, Spreading and Cutting, Sewing, Washing, Quality Assurance, Finishing and Packaging, Maintenance, Finance and Account, Industrial Engineering and HR/Admin Department.

Each of these departments produces a large volume of data. These data are different from each other in terms of structure and content. In Bangladesh, the second-largest global apparel exporter country [8], most the industries store their valuable data in an unstructured format. The authors believe that the situation is almost same around the globe. As a result, it is difficult to combine the data originating from different departments to produce fruitful knowledge.

Unlike the structure, the process flow in apparel manufacturing is also divided into several steps. The process starts from Design and ends with Shipment [9]. In between 14 different activities are performed which are shown in Fig. 1.

The overall production process of a garments factory is controlled by the Industrial Engineering department at present. The major objective of industrial engineering department is to increase the productivity by reducing waste and non-value adding operations and improving effective utilization of resources [10]. This department is a great source of data like daily individual effciency report, line and factory effciency report, line and factory lost time report, incentive working report [11]. Understanding the value of IE department in an apparel industry and due to the different practices between apparel and other industries, it is required to develop a methodology for mining industrial engineered data of apparel industry.

3. RELATED WORKS

The application of data mining in manufacturing practice is rare in comparison with the service industry [7]. Existing data mining approaches in manufacturing address the following fields of application:

—Quality analysis of products to correlate output quality and system parameters [12], [13].

—Failure analysis of production resources [14], [15].

—Maintenance analysis to enhance the availability of production resources [16], [17].

—Production planning and scheduling analysis to improve planning quality [18], [19].

From the above discussion, the authors conclude that due to the rare practice of application of data mining in the apparel industry, the necessity for developing and using methodology especially designed for mining IE data of apparel industry has been overlooked till now.

However, a generic methodology has been proposed for mining industrial engineered data by Solarte Jose [4]. But the proposed methodology was not tested or executed. It also suffers from the lack of some required steps of mining industrial engineered data of apparel industry like analyzing role and relation of departments, data preprocessing, implementing suitable data storage and improper flows of activities. Although he suggests that, the methodology can be used on any industrial engineered data but due to the absence of these steps its not appropriate for the apparel industry. Besides, the methodology contains some inconsistent transitions among different steps like, move to choose a new tool based on only one evaluation criteria among many, no clarification where to start from when changes are performed.

There are other different methodologies for data mining, analytics or data science related projects. Among these, according to the survey conducted by kdnuggets.com in 2014 [20], the two most popular methodologies are SEMMA [21] and CRISP-DM [22]. The survey shows that CRISP-DM is the most widely used data mining methodology with 43% votes. Methodologies prepared by the practitioners holds the second position with 27.5% popularity. On the other hand, SEMMA holds the third position with 8% votes. Based on the findings from the survey, the authors have studied the appropriateness of CRISP-DM, SEMMA for mining IE data of apparel industry.

SEMMA [21] is a data mining methodology proposed by the SAS institute which is a leader in data mining solutions. In SEMMA, five steps for mining data has been proposed by SAS which are: sample, explore, modify, model, and assess. According to J. Solarte [4], some fundamental parts of any information systems project, including analysis, design, and implementation phases are missing in this methodology. Furthermore, the organization and the stakeholders are not considered in this methodology which may result in an unexpected output from the project for apparel industry as the structure of organization and interest of the stakeholders determine the success. In addition, SEMMA is specifically designed to help the user of the SAS enterprise miner software, the data mining software of the SAS institute. As a consequence, it may lead to ambiguity if its applied outside of the software [23]. In terms of flow among different stages, SEMMA methodology suffers from great inefficiency. For instance, after the modification step, it can happen that the sample data may not lead to required answers. In this case, it would be efficient to traverse to the first step, Sampling rather than go through the model and assess step.

CRISP-DM [22] methodology is a non-proprietary, application or industry neutral where the mining process is broken down into six
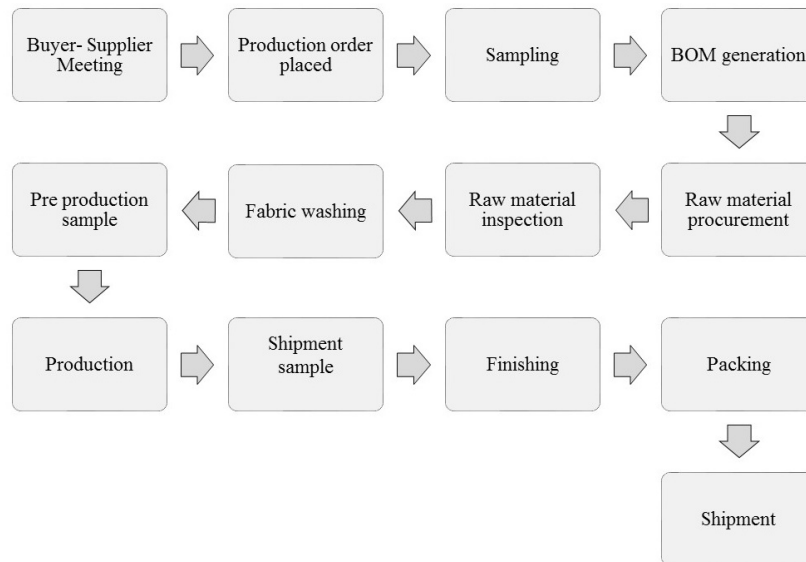
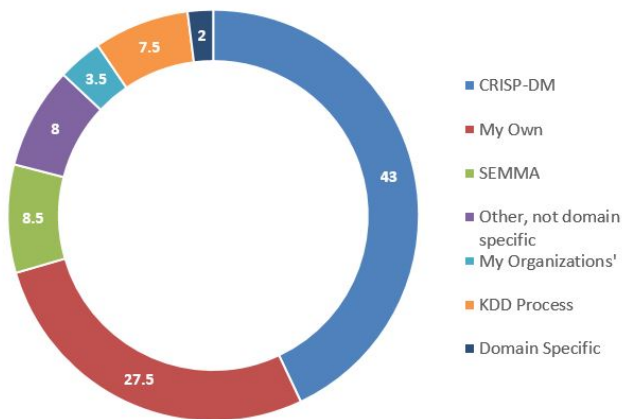Fig. 1. Process flow of Apparel Manufacturing.

Fig. 2. The result of the survey to find out mostly used Data Mining Methodologies.

phases. The phases are business understanding, data understanding, data preparation, modeling, evaluation, deployment. Its major drawback is that it combines tools (software packets) and techniques in the same category [4]. This may cause the organizations goals and requirements to be under-analyzed and biases the study. Furthermore, in the case of switching among steps, it is unclear what to do after deployment and in which step to go if any changes occur.

In this age of Big data and Data Science, another methodology has been proposed by John B. Rollins [24] to construct a model for prediction or discover patterns. This a generic methodology consists of 10 stages. Although this methodology has emphasized on new practices like use of the large volume of data, integration of text analytics but it is missing the essential part of determining the appropriate data mining tool and its feasibility.

From the above discussion, the authors sum up that the apparel industry has got less attention compared to other service sector in terms of performing data mining research. Due to the nature of the apparel industry, some important steps like analysis of apparel manufacturing unit, understanding the roles and data of each department, analysis of IE data are missing in the previously proposed methodologies. the authors have identified the research gap for a domain specific data mining methodology for apparel industry which is currently absent. But a specific methodology for apparel industry can support this sector to apply data mining techniques in order to increase productivity, reduce waste of time and money for maximizing profit.

4. PROPOSED METHODOLOGY

A structured methodology is required by engineers to solve problems which enable them to replicate results or errors at any step in the process. As a result, they are confident on their recommended solutions. For this reason, a data mining methodology has been offered in this section for solving problems related to industrial engineered data of apparel industry.

This proposed methodology holds some similarities to other well-known methodologies [21], [22], [24], but it has proposed some new steps like analyzing apparel manufacturing unit, understand the roles and data of each department, analysis of IE data and determine the tools and its feasibility. Besides, the proposed methodology has tried to solve the inconsistent transitions from one step to another found in the existing methodologies. The inclusion of all these new steps and clear transitions in a methodology makes it a domain specific data mining methodology for mining IE data of apparel industry.

Different steps of this methodology are described as follows:

(1) Analysis of Apparel Manufacturing Unit
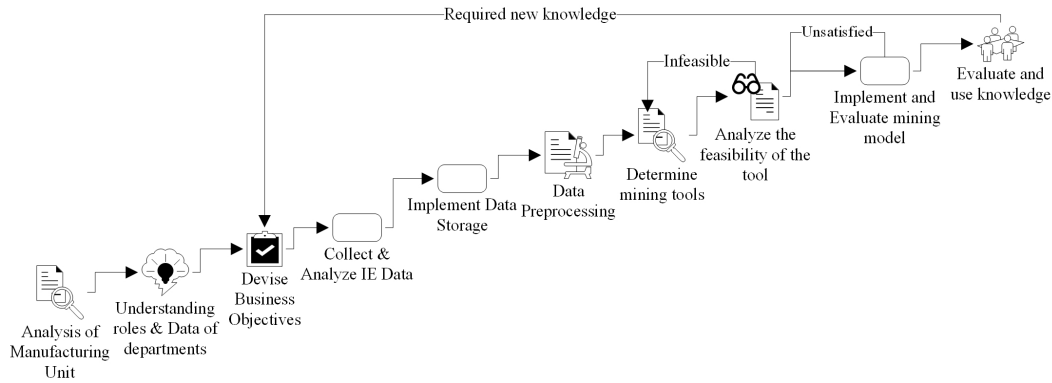(2) Understand the Roles and Data of Each Departments
(3) Devise Business Objectives

Fig. 3. Methodology for Mining Industrial Engineered Data of Apparel Industry.

(4) Collect and Analyze IE Data

(5) Implement Data Storage

(6) Data Preprocessing

(7) Determine Data Mining Tool

(8) Analyze the feasibility of Data Mining Tool.

(9) Implement and Evaluate Mining Model

(10) Evaluate and Use Discovered Knowledge

These steps are described briefly in the following subsections and shown in Fig. 3

4.1 Analysis of Apparel Manufacturing Unit

Before applying data mining tools and techniques on industrial engineered data, the analyst should analyze the apparel manufacturing unit first. The analyst needs to understand the process flow of manufacturing as this differs from company to company. During analyzing the process flow, its required to identify the departments under the manufacturing unit.

When the departments are identified then its time to identify the stakeholders to successfully implement the data mining project. Stakeholders including major owners, users, analysts, designers and developers are considered an essential element for analyzing and designing any information system. Once the stakeholders are identified then its time to understand the point of interest, requirements and expectations of the stakeholders. Its a well-known fact that the success of an information project depends on the satisfactions of stakeholders expectations.

4.2 Understand the Roles and Data of Each Departments

Manufacturing unit of apparel industry comprises of many departments e.g. merchandising, sampling, fabric store, spreading, cutting and sewing. Each department has its own role and produces different types of data. Understanding the roles and data of each department helps the analyst to gain in-depth knowledge about the relevance and importance of a department to meet the requirements. At this step, the analyst eliminates irrelevant departments in accordance with the point of interest and identify the co-related departments. For example, in order to predict the productivity or efficiency of each worker under the sewing department, along with the sewing department, data from HR and IE department are required to get the effective outcomes.

4.3 Devise Business Objectives

The goals and objectives must be clear and specific to form up an action plan; these should be understood completely by all the participants involved in the project. These goals and objectives are dynamic in nature as they correspond to the needs and requirements of business or factory. The analyst should collect and understand the industrial engineered data in order to devise the expected output from the project. The business objectives should be revised whenever new knowledge is required.

4.4 Collect and Analyze IE Data

In modern days, the production process of an apparel industry is controlled by the industrial engineering department. This department is responsible for increasing productivity by eliminating the unwanted operations and ensuring efficient usage of resources. Significant data concerning daily efficiency reports, line and factory efficiency reports, line and factory lost time reports, incentive working reports are generated by IE department. In order to get an effective insight of productivity, collection and analysis of IE data is an obvious task for the analyst. The authors have observed that the data are not structured in a consistent manner to analyze. The data entry operator needs to structure it manually or create an automation process to construct the properly structured data.

4.5 Implement Data Storage

For storing and managing data, organizations can use Database Management System (DBMS) or Data Warehouse. A DBMS which is also known as online transaction processing (OLTP) system, provides its users a clear, systematic approach to create, store, update and delete data. DBMS is used to cover the day to day operations like banking transactions, registrations. On the other hand, data warehouse also known as online analytical processing (OLAP) helps the users in data analysis and decision making. According to Han, a well-known author on data mining defines data warehouse as a data repository that is maintained separately from an organizations operational databases [?]. A data warehouse system consists of essential preprocessing steps like data cleaning, data integration, data transformation for data mining.

In the case of the apparel industry, choosing data warehouse for performing analytics would be beneficial for practitioners due its higher performance, support in decision making, based on the multidimensional data model and as targeted for knowledge workers

like managers, executives and analysts. When it comes to data modeling, there exist several schemas. In the case of DBMS, the entity-relationship (ER) data model is a popular choice. On the other hand, star schema, snowflake schema or fact constellation schema are most popular for the multidimensional data model.

Star schema is the most common data modeling schema for multidimensional data modeling where the data warehouse contains one large central or fact table and a set of smaller attended table (dimension table).

The Snowflake schema is another variant of star schema where dimension tables are normalized, means splitting the data into additional smaller tables. Although Snowflake schema offers easier maintenance than Star schema but it can reduce the efficiency of browsing as more joins are required to execute queries.

On the other hand, Fact constellation schema contains several fact tables which make the schema a collection of several stars. In fact, the constellation schema, dimension tables are shared among fact tables.

As the performance of snowflake schema reduces, in case of multiple joins to execute query and fact constellation schema shares dimension tables among fact tables which are not required for storing manufacturing data of apparel industry, so based on the analysis for each data modeling schema above, we suggest that star schema would be the smart choice for the data warehouse for IE data analytics and mining.

4.6 Data Preprocessing

The real world data is often incomplete, missing certain behavior, inconsistent or noisy due to their massive volume and origin from multiple, heterogeneous sources [1]. These facts reduce the quality of the data but low-quality data results in low-quality mining which may fail to fulfill the objectives. So it is vital to improve the quality of the data before applying data mining and this can be achieved by performing preprocessing techniques on data.

Data cleaning is one kind of preprocessing technique that is used to remove or replace missing, noisy data and removes inconsistencies from data. On the other hand, data integration is used to combine data from multiple sources to a compact data source like a data warehouse. Data reduction can be used to reduce the size by dimension, numerosity, and compression. Data transformation can be applied where data can be scaled to fall in a smaller range in order to increase the accuracy and efficiency of the mining algorithm.

4.7 Determine Data Mining Tool Based on Task and Techniques

Once the action plan is formed and a data warehouse is implemented, it is now possible to select the appropriate data mining tasks, techniques, and tools. The selection of data mining tasks depends primarily on the goals of the project. So it is unwise to select the tasks after selecting the techniques and tools as the tools and techniques may limit or influence the data mining task. In this research, data mining tools are considered as the specific software packages or solutions presently offered by vendors. For the selection of data mining package, several authors have suggested a number features that should be analyzed but they have also concluded that "there is no one best data mining tool for all purposes" [24].

4.8 Analyze the Feasibility of the Tool

The determined data mining tool in the previous step may not be feasible always. There are many factors on which the feasibility of a data mining tool depends. In [4], the author has proposed 6 factors for analyzing the feasibility of a data mining tool. The factors are designated as tasks, price, performance, functionality, usability and support. If the selected tool in the previous step is not feasible, then the analyst should select an alternative tool.

4.9 Implement and evaluate mining model

A mining model is constructed by applying data mining algorithm on data. But the mining model is more than a container of an algorithm or metadata. It is defined as a set of data, statistics, and patterns which can be used further to generate prediction and identify relationships among variables for new data [25].

The architecture of a mining model consists of three parts: metadata, patterns, and bindings. Metadata contains the name of the model, storing address and definition of the model. The pattern can be generated by applying algorithms on training data set. Bindings point to the data cached in the mining model. Initially, a mining model is always empty unless data is provided. After processing the data, the mining model contains metadata, results, and bindings to the mining structure.

After generating mining model outcomes, it is evaluated by a human supervisor because the outcomes may not always fruitful. Different methods and criteria can be used to evaluate a model depending on model type. For example, coincidence matrix can be used to evaluate classification models, means error rate to evaluate regression models.

Based on the evaluation result or new business objectives, it may be required to perform updates or changes mining model and this process continues until the project fulfill all the requirements of the project.

4.10 Evaluate and Use Discovered Knowledge

By interpreting the patterns found in the previous step, valuable knowledge can be discovered. This discovered knowledge may help the stakeholders in decision making. When new knowledge is required but existing mining model cannot provide it, then the mining process should start again from the devised business objectives step. This process continues until the stakeholders require no more knowledge or unable to carry the project further.

5. RESULT AND DISCUSSION

Our proposed methodology has mainly focused on the manufacturing unit and understanding the role and data of each department of this unit. The methodology has been used experimentally to discover pattern from hourly production data of swing department in order to verify the effectiveness of the methodology.

Based on the historical productivity data of 6 months, generated regression model shows that the efficiency in productivity for apparel industry is lowest at the first hour in the morning. The efficiency increases every hour and this uptrend continues up to lunch time. After the lunch (01-02), the efficiency drops a bit but continues to rise till the last working hour. The derived equation for the simple Regression Model is:
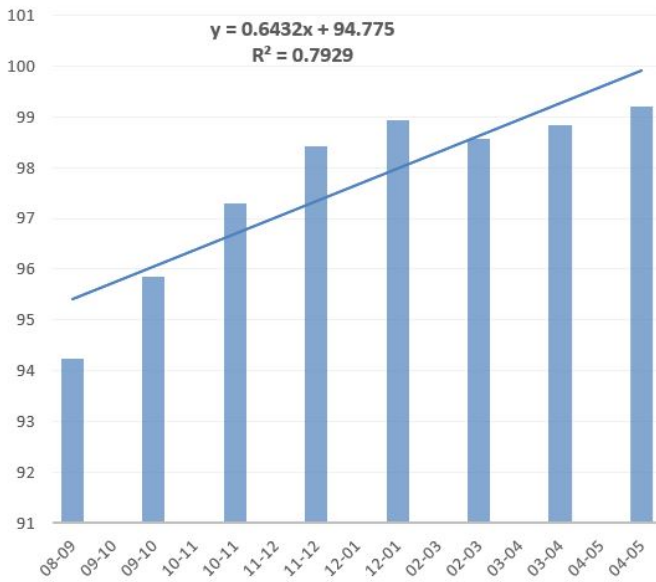
$$y = 0.6433x + 94.775$$

Fig. 4. Efficiency in Productivity in time series.

where x= working time (in 12hour time format) and y = productivity in percent.

After further investigation, two factors are identified as the reason behind the lowest productivity in the first hour of the morning. One reason is the workers need to prepare their workstation every day and they do it normally in the morning. For this case, its recommended to instruct the workers to keep ready their workstation in the previous day so that they can save their time in the morning. The main reason behind the lower productivity in the morning is the synchronization between man and machine. As apparel industry is a labor-incentive industry that requires high human-machine interaction, so its very important to establish synchronization between these two. Sewing machines should be designed in such a way so that workers can easily synchronize with the machines in less time.

6. FUTURE DIRECTION AND CONCLUSION

The proposed methodology in this research paper has extended the new scope of doing further research in the application of data mining techniques in the apparel industry. As for future work, it is recommended to apply and generalize the proposed methodology for other sectors of the apparel industry and other industrial engineering sectors.

This methodology emphasizes the manufacturing unit of apparel industry along with understanding the roles and data of the constituent departments. The methodology has attempted to provide clarity on transitions among the substantial steps compared to other existing methodologies. The authors believe that the proposed methodology can be considered as a domain specific data mining methodology for mining industrial engineered data of apparel industry due to the introduced new steps and clear transitions between various steps. This methodology can be greatly helpful to the leaders in the field of industrial engineering management and decision makers involved in the manufacturing unit.

7. REFERENCES

[1] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[2] Chih-Hung Hsu. 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, 2009.

[3] Naresh Paneru et al. Implementation of lean manufacturing tools in garment manufacturing process focusing sewing section of mens shirt. 2011.

[4] Jose Solarte. A proposed data mining methodology and its application to industrial engineering. 2002.

[5] Melissa Breyer. 25 shocking fashion industry statistics, Sep 2012.

[6] Malcolm Newbery. *Apparel manufacturing technology*. Aroq Limited, 2005.

[7] Christoph Gröger, Florian Niedermann, and Bernhard Mitschang. Data mining-driven manufacturing process optimization. In *Proceedings of the world congress on engineering*, volume 3, pages 4–6, 2012.

[8] Bangladesh is second-largest global apparel exporter - retailers' hub: India's 'dollar city' tirupur on a rise, but bangladesh reigns supreme, Jul 2013.

[9] Flow chart of garments manufacturing process, Jul 2015.

[10] Martand Telsang. *Industrial engineering and production management*. S. Chand, 2006.

[11] Prasanta Sarkar. Duties and responsibilities of industrial engineering department.

[12] Sewon Oh, Jooyung Han, and Hyunbo Cho. Intelligent process control system for quality improvement by data mining in the process industry. In *Data mining for design and manufacturing*, pages 289–309. Springer, 2001.

[13] Victor A Skormin, Vladimir I Gorodetski, and Leonard J Popyack. Data mining technology for failure prognostic of avionics. *IEEE Transactions on Aerospace and Electronic Systems*, 38(2):388–403, 2002.

[14] Wei-Chou Chen, Shian-Shyong Tseng, and Ching-Yao Wang. A novel manufacturing defect detection method using association rule mining techniques. *Expert systems with applications*, 29(4):807–815, 2005.

[15] Lixiang Shen, Francis EH Tay, Liangsheng Qu, and Yudi Shen. Fault diagnosis using rough sets theory. *Computers in industry*, 43(1):61–72, 2000.

[16] Dentcho Batanov, Nagen Nagarur, and Prapan Nitikhunkasem. Expert-mm: A knowledge-based system for maintenance management. *Artificial intelligence in engineering*, 8(4):283–291, 1993.

[17] Carol J Romanowski and Rakesh Nagi. Analyzing maintenance data using data mining methods. In *Data mining for design and manufacturing*, pages 235–254. Springer, 2001.

[18] Reimund Belz and Peter Mertens. Combining knowledge-based systems and simulation to solve rescheduling problems. *Decision Support Systems*, 17(2):141–157, 1996.

[19] JA Srinivas and Muhammad Shahbaz. Agent oriented planning using data mined knowledge. In *10th International Conference on Concurrent Engineering, Adaptive Engineering for Sustainable Value Creation*, pages 301–307, 2004.

[20] Gregory Piatetsky. Kdnuggets, Oct 2014.

[21] Ana Isabel Rojão Lourenço Azevedo and Manuel Filipe Santos. Kdd, semma crisp-dm: a parallel overviee. *IADS-DM*, 2008.

[22] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rudiger Wirth. Crisp-dm 1.0 step-by-step data mining guide. 2000.

[23] Seyyed Soroush Rohanizadeh and Mohammad Bameni Moghadam. A proposed data mining methodology and its application to industrial procedures. *Journal of Industrial Engineering*, 4(1):37–50, 2009.

[24] John B. Rollins. Foundational methodology for data science. 2015.

[25] Mining models (analysis services - data mining).