International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume 2, Issue 2 (February '2016)

# LIFE OF GARMENT FEMALE WORKERS: A DATA MINING CASE STUDY

**K. Sujeenthra,**
PG Student,
Department of Computer Science,
Jayaraj Annapackiam College for Women (Autonomous),
Periyakulam,Theni,Tamilnadu,India.

**V. Annalakshmi,**
Assistant Professor,
Department of Computer Science,
Jayaraj Annapackiam College for Women (Autonomous),
Periyakulam,Theni,Tamilnadu,India.

**M. Krishnaveni,**
PG Student,
Department of Computer Science,
Jayaraj Annapackiam College for Women (Autonomous),
Periyakulam,Theni,Tamilnadu,India.

**Abstract:** The aim of this study is to describe the experiences of some Young Women workers who have encountered exploitation in their workplaces in garment industries. It equally looks at the effects of worker exploitation on Young Women workers, and suggested possible ways that workers could make themselves less vulnerable to workplace exploitation. The findings were divided into three parts. The first part of the finding uncovered the various kinds of exploitative practices the migrants face in their workplaces such as denial of various workers' rights.The second part showed the personal feelings of the exploited persons in response to their experiences. The third part of the finding showed the various negative effects thatYoung Women workers exploitation produces in the life of the exploited persons such as psychological, physiological and emotional as well as health effects. The study was able to discover the reasons for migrant workers' vulnerability to Young Women workers exploitation and the actions that could be taken to reduce the vulnerability of migrant workers in their workplaces.

***Keywords:*** Data Mining Analysis, Garments, Exploitation, Migration, Young Women Workers.

## I. INTRODUCTION

India is the world's second largest producer of textiles and garments after China. It is the world's third largest producer of cotton—after China and the USA—and the second largest cotton consumer after China. The textile and garment industry in India is one of the oldest manufacturing sectors in the country and is currently it's largest. The textile and garment industry fulfils a pivotal role in the Indian economy. It is a major foreign exchange earner and, after agriculture, it is the largest employer with a total workforce of 35 mm. In 2009 textiles and garments accounted for about16 per cent of industrial production and 18 per cent of export earnings. The industry covers a wide range of activities. These include the production of natural raw materials such as cotton, jute, silk and wool, as well as synthetic filament and spun yarn. In addition an extensive range of finished products are made. The Indian textile industry accounts for about 24 percent of the world's spindle capacity, making it the second highest after China and around eight percent of global rotor capacity. Also, it has the highest loom capacity—including hand looms—with a 63per cent share. India accounts for about 14 per cent of the world's production of textile fibers and yarns.This includes jute, of which it is the largest producer. The country is the second largest producer of silk and cellulose fiber and yarn and the fifth largest producer of synthetic fiber and yarn. The Indian textile garment industry is an enormous complex entity. There is organized sector, decentralized sector and down the line weavers, the artisans as well as the farmers. The spectrum of technology is widespread right from handmade to semi-mechanical, mechanical and highly sophisticated information based technology and microprocessor based technology.

This case study presents data mining in lifestyle environment that identifies the problem of the workers. The identified patterns are analyzed to offer helpful and constructive recommendations to the poor women workers in garment industry. It enhances their own problems in daily life. This will also used to improve the workers environment and their comfortless of the industry. This study is undertaken to analyses the structure, conduct and performance of women and their impact on the women in Theni district, Tamilnadu.

## II. PROBLEM DEFINITION AND DESCRIPTION

The textile industry today 450,000 jobs exist in about 6,000 firms, whereby only 14 per cent can be regarded as medium or larger companies, most firms being small or very small in size. At least 10 per cent of the textile industry is controlled by foreign capital. The significant trend towards concentration in this branch is shown clearly by the fact that large firms only employ 10 per cent of the workforce but account for 84 per cent of turnover in this branch of industry. The key region is the south-east where 73 per cent of the firms are concentrated, half of which are located in the industrial Centre of Sao Paulo. The garment industry (including shoe production and textile accessories) comprises 35,000 companies in which 650,000 workers are employed. As a result, the garment industry is, after the building industry, the second largest sector in Brazil.

 

Several statistical tools have been used to analyze workers performance from different points of view. This case study presents data mining in lifestyle environment that identifies the problem of the workers. The identified patterns are analyzed to offer a helpful and constructive recommendations to the poor women workers in garment industry. It enhance their own problem in daily life. This will also use to improve the workers environment and their comfortness of the industry.

## III. RESEARCH METHODOLOGY

The case study will consist of different stages, roughly following the cross industry standard procedure CRISP-DM. Firstly, the business understanding phases as to be carried out. In this phase, the project objectives and requirements are stated and reined and the resulting data mining problem is formulated. The results of this phase are summarized in the previous sections. Although the collection of additional data results in a richer data set and is therefore likely to give better results, model acting on a data set that is already automatically kept up-to-date is potentially a much more useful tool.

**a) Design for Methodology**

A multiple case study design with historical and cross-sectional perspective was adopted. According to Patton (1987) a project is a case study design when it seeks to capture rich and unique experiences as they happen in reality. Merriam (1998) indicates that a historical case study is essentially descriptive and deals with developments in the past. Yin (1994) points out that a case study design is cross sectional to the extent that it examines "how" and "why" questions in a contemporary set of events. To the extent that this study attempts to reflect upon historical and cross-sectional events as reconstructed by subjects or respondents it is a survey research, and since it aims to include investigations from more than a single case it is a multiple case study design. The implementation of the study has been framed around the following procedures:

- Developing a research proposal and instruments for data collection,
- Creating contact with respective educational bureaus,
- Selecting research settings and sampling the research population or respondents,
- Pre-testing instruments,
- Administering instruments and completing data collection,
- Analyzing and interpreting data, and writing a report.

**b) Algorithm used**

**Cluster Analysis**

Cluster analysis is a multi variety analysis that attempts to form groups or "clusters" of objects (sample plots in our case) that are "similar" to each other but which differ among clusters. The exact definition of "similar" is variable among algorithms, but has a generic basis. The methods of forming clusters also vary, but follow a few general blueprints.

**Similarity, Dissimilarity and Distance**

Similarity is a characterization of the ratio of the numbers of attributes two objects share in common compared to the total list of attributes between them. Objects which have everything in common are identical, and have a similarity of 1.0. Objects which have nothing in common have a similarity of 0.0. As we have discussed previously, there is a large number of similarity indices proposed and employed, but the concepts are common to all.

Dissimilarity is the complement of similarity, and is a characterization of the number of attributes two objects have uniquely compared to the total list of attributes between them. In general, dissimilarity can be calculated as similarity K-means clustering. The most common partitioning method is the K-means cluster analysis.

Conceptually, the K-means algorithm:

1. Selects K centroids (K rows chosen at random)
2. Assigns each data point to its closest centroid
3. Recalculates the centroids as the average of all data points in a cluster (i.e., the centroids are p-length mean vectors, where p is the number of variables)
4. Assigns data points to their closest centroids
5. Continues steps 3 and 4 until the observations are not reassigned or the maximum number of iterations (R uses 10 as a default) is reached.

Implementation details for this approach can vary. R uses an efficient algorithm by Hartigan and Wong (1979) that partitions the observations into k groups such that the sum of squares of the observations to their assigned cluster centers is a minimum. This means that in steps 2 and 4, each observation is assigned to the cluster with the smallest value of:

$$SS(k) = \sum_{i=1}^{n} \sum_{i=n}^{\mu} (x_{ij} - \bar{x}_{kj})^2$$

Where k is the cluster, $x_{ij}$ is the value of the $j^{th}$ variable for the $i^{th}$ observation, and $\overline{Xkj}$ bar is the mean of the $j^{th}$ variable for the $k^{th}$ cluster.

K-means clustering can handle larger datasets than hierarchical cluster approaches. Additionally, observations are not permanently committed to a cluster. They are moved when doing so improves the overall solution. However, the use of means implies that all variables must be continuous and the approach can be severely affected by outliers. They also perform poorly in the presence of non-convex (e.g., U-shaped) clusters.

The format of the K-means function in R is kmeans(*x*, *centers*) where *x* is a numeric dataset (matrix or data frame) and *centers* is the number of clusters to extract. The function returns the cluster memberships, centroids, sums of squares (within, between, total), and cluster sizes.

Since K-means cluster analysis starts with k randomly chosen centroids, a different solution can be obtained each time the function is invoked. The seed() function to guarantee that the results are reproducible. Additionally, this clustering approach can be sensitive to the initial selection of centroids. The kmeans() function has an nstart option that attempts multiple initial configurations and reports on the best one. For example, adding nstart=25 will generate 25 initial configurations. This approach is often recommended.

Unlike hierarchical clustering, K-means clustering requires that the number of clusters to extract be specified in advance. Again, the NbClust package can be used as a guide. Additionally, a plot of the total within-groups sums of squares against the number of clusters in a K-means solution can be helpful. A bend in the graph can suggest the appropriate number of clusters.

**c) Tools for the study**

**The R environment:**

R is a free software environment for statistical computing and graphics. It provides a wide variety of statistical and graphical techniques. R can be extended easily via packages. R is an integrated suite of software facilities for data manipulation, calculation and graphical display.

**Validating cluster solutions:**

The function cluster stats () in the FPC package provides a mechanism for comparing the similarity of two cluster solutions using a variety of validation criteria.

**Data sources and Methodology:**

**Target population:**

This survey covers all the villagers of Theni district.

**Instrument design:**

This questionnaire collects data on the attitude of the people regarding dropping out of women's. The items and reasons on the questionnaire have remained unchanged for several years. However, should modifications become necessary, proposed changes would go through a review committee and a field test with respondents and data users to ensure it's relevant.

**Sampling:**

This survey is a census with a cross-sectional design. Data are collected for all units of the target population, therefore no sampling is done.

**Data sources:**

Responding to this, survey is mandatory. Data are collected directly from survey respondents. Data are compiled from the responses the researcher collected by the questionnaire.

The researcher performs the data capture activities, and follow-up of non-respondents. Contact with respondents is maintained for subsequent follow-up.

**Error detection:**

There are edits built into the data capture application to check the entered data for unusual values, as well as to check for logical inconsistencies. Whenever an edit fails, the interviewer is prompted to correct the information (with the help of the respondent when necessary). For most edit failures the interviewer has the ability to override the edit failure if necessary.

**Imputation:**

A 100% response rate is attained; therefore imputation is not necessary.

**Quality evaluation:**

Prior to the data release, combined survey results are analyzed for comparability; in general, this includes a detailed review of individual responses, general economic conditions, and historical trends. The data is examined at a macro level to ensure that the long-term trends make sense when compared to publicly available information in media reports, and etc. Large Subject matter officers follow up with the academicians to confirm the data and to document reasons for large functions in sales or inventories.

**Disclosure control:**

Releasing any data would divulge information obtained under the Statistics Act that relates to any identifiable person, business or organization without the prior knowledge or the consent in writing of that person, business or organization. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosure of identifiable data.

## IV. ANALYSIS OF DATA

**a) Data Mining Process**

Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. With data mining, a retailer could use point-of-sale records of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop products and promotions to appeal to specific customer segments.

**Data mining consists of five major elements:**

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

**b) Analysis of Data**

The researcher collects more than 400 responses from samples all over Theni district. She spent nearly 2 months to collect the fresh data from end users. After collecting the information, all the details are fed into the software and checked for outlier.

The cleaned data was analyzed using single attribute and multiple attributes.

## V. DATA MINING FINDINGS

The initial studies unveiled a number of relationships between variables as well as threshold values that justify further analysis. The several values of several attributes are useful predictors of retention and or attrition. These explanations increase our confidence that the values of these attributes will continue to be predictors in the future. We have classified few reasons for life of garment female workers by using R Tool.

**a) Sample Input Data**

| Name | Age | Shift | Salary | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A.selvarani | 23 | 1 | 5000 | 2 | 1 | 2 | 2 | 2 | 1 | 3 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 |
| A.Rajathi | 21 | 2 | 4000 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 |
| G.viji | 32 | 1 | 7000 | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 |
| B.MathiAlagu | 18 | 1 | 4000 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 |
| C.sakthi | 15 | 1 | 4000 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 |
| D.suganya | 23 | 2 | 8000 | 2 | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 |
| S.kayathiri | 25 | 1 | 8000 | 2 | 1 | 1 | 1 | 2 | 2 | 3 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 |
| s.MuthuSelvi | 19 | 1 | 4000 | 2 | 2 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| P.PandiSelvi | 26 | 2 | 7000 | 2 | 1 | 1 | 2 | 2 | 2 | 3 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 |
| P.Devi | 31 | 2 | 8000 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |

**Table1. Sample Data**

**Step1:**

After apply all the details into the R Tool we get the following cluster dendrogram with Au/BP values the result will be shows in the follwing fig.1

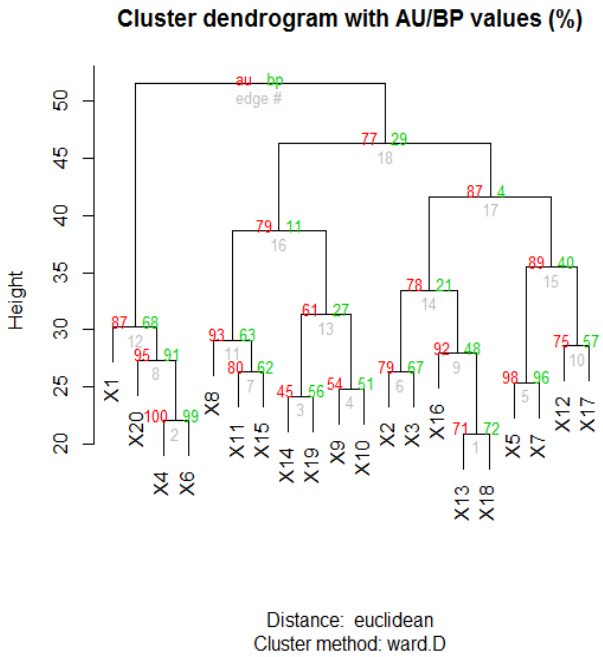


**Figure 1. Cluster dendrogram with AU/BP Values**

**Step2:** Here we find out the p-value and standard Error. Fig2. Shows the p-value vs standard error plot value details

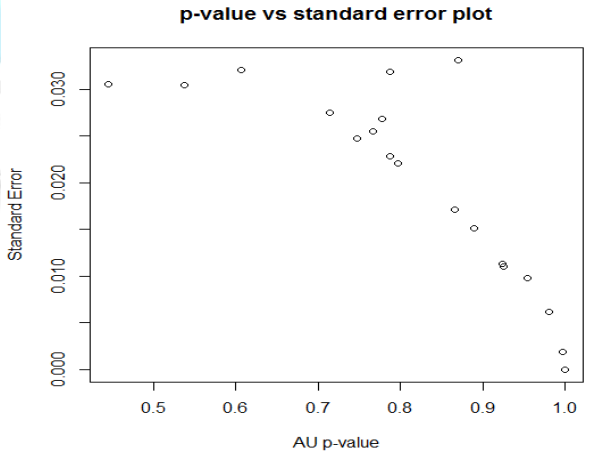


**Figure 2. P-Value vs. Standard Error Plot**

**Step3:** The clustered output details will be shown in the follwing figure 3.

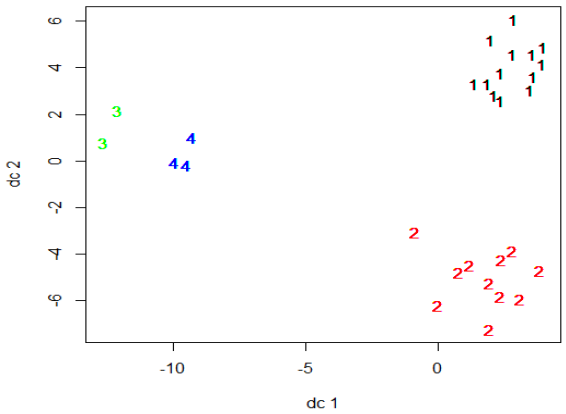


**Figure 3. Output data**

## VI.CONCLUSION

The goal of clustering is to build a set of models that can correctly predict the class of the different objects. It is clear that for many of the women at least in the earlier decades of the industry, the decision to take up garments work was difficult because it involved conflict or tension with (generally men) family members, or with women's and society's own values and ideologies about women's mobility and work. Average and percentage analysis was carried out to draw meaningful interpretation of the results. Garret ranking technique was used to find the reasons for joining the garment Industry. Even if they can refuse, the workers are not able to refuse as they need the money that they make with the paid overtime. The workers are routinely denied leave, and overstaying of leave granted is treated as an offence meriting termination of services. Before the case study was done, it is very difficult to cluser the reason for the above hypothesis. Our case study reveals that the reasons for migrant workers' vulnerability to Young Women workers exploitation and the actions that could be taken to reduce the vulnerability of migrant workers in their workplaces.

## VII. REFERENCES

[1] Kantor, Paula, 1999, 'Estimating Numbers in Household Industry: The Case of Home based Garment Makers in Ahmedabad', Ahmedabad, (mimeo).

[2] Lalbhai, Arvind, SamvegLalbhai and Samar Verma, undated, 'Indian Textile Industry, Strategy for Achieving Strong Global Presence', (mimeo).

[3] Mishra, C (2005). SHGs in the Unorganized Garments Sector – A Case Study of Madurai, Kurukshetra, 53(9), 43-47.

[4] Manmohan, G; Tushir, M and Chadha, M (2008). Rural Banking and Microfinance. Southern Economist, 47(2) 9-12.

[5] Gunasekaran, A. (2010). Micro Credit and Women Employment through Self Help Groups: A Case study of Chennai City Rural Credit in the Era of Globalization, Mabhav Books, Haryana.

[6] Arya, P.P (1984), "Work Satisfaction and its Correlates", Indian journal of Industrial Relations, Vol. 20, 89-100.

[7] Eaton, A.E., Gordon, M.E., and Keefe, J.H., (1992), "The impact of quality of work life programs and grievances system effectiveness on union commitment", International and Labor Relations Review, Vol. 45, No. 3

[8] Kothari.C.R.,Research Methodology, Vikas Publication House, New Delhi, 1967.