

Demographic Population Segmentation for Predicting Employability Characteristics of Women

¹R.Priyatharshini,²Sai Prashanthi. S, ³Shanmathi. A, ⁴Sankkavai Lumumba. N.K,⁵Dr.S.Chitrakala
¹⁻⁴Information Technology. Easwari Engineering College, Chennai, India
⁵Anna University, Chennai, India

Abstract—In this paper an approach for employability characteristics prediction has been proposed. Demographic segmentation is the process of segmenting the population according to age, race, religion, gender, family size, ethnicity, income, and education. With this demographic segmentation method we segment the population based on their demographic characteristics. Data Mining is an analytic process designed to explore data in large databases and in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining. Big data is a term for massive data sets having large, more varied and complex structure with the difficulties of storing, analyzing and visualizing for further processes or results. In our paper we propose to predict the employability characteristics of women using decision tree algorithm. Here we use the data from the United Nations Statistics Division's datasets. We analyze the statistics of employed and unemployed women and propose a solution to improve the employability.

Keywords— Data Mining; Big data; Statistics; Employability.

I. INTRODUCTION

Demographic segmentation divides the market into groups based on demographic variables including age, gender, family size and life cycle. With this demographic segmentation method we segment the population based on their demographic characteristics. Self-employed women are more constrained by the lack of capital than are men in a similar category who are attempting to raise their productivity. Policies such as structural adjustment programs to reduce regulatory barriers to employment creation in the protected wage sector, and strengthen fiscal measures to accelerate growth are perhaps necessary but are not sufficient for the alleviation of poverty in rural areas. This is because these measures cannot improve the conditions of the especially rural women, who are trapped in the unprotected sector due to low levels of education and other forms of human capital. A more targeted approach is perhaps required. It is therefore necessary to complement structural adjustment by human capital development programs which can improve productivity capacity among the poor as well as strengthening measures which can

remove the legal and other regulatory constraints to employment expansion or job mobility of women. These measures are expected to benefit the rural women to allow them to gain access to protected wage jobs and also provide an efficient allocation of employment. An approach for predicting the employability characteristics of woman has been proposed according to their educational characteristics in both rural and urban areas. Data mining techniques can be used for demographic segmentation. The most commonly used data mining techniques used for demographic segmentation are clustering, classification and association rule mining techniques.

Gender-linked occupational differentiation has been seen as influenced by both the industrial structure of the economy and the gender composition of the labour force. The observed effect of industrial shifts on gender-occupational differentiation, however, is argued to be a spurious consequence of the gender-composition of the work force. Specifically, the development of tertiary industries generates greater demand for female labour. Intensive recruitment of women to the labour force in turn increases occupational differentiation because females, in gender-typed labour markets, are likely to be channeled in disproportionate numbers away from upper-status occupations. The findings demonstrate that traditional modernization theory is unable to account for this. Most women in India work and contribute to the economy in one form or another, much of their work is not documented or accounted for in official statistics. Women in rural areas plough fields and harvest crops while working on farms, weave and make handicrafts while working in household industries, sell food and gather wood while working in the informal sector. Women in urban areas are engaged in business enterprises in international platforms have greater career opportunities as a result of international network. Freer movement of goods and capital is helpful to this section. But most women continue to remain marginalized as they are generally employed in a chain of work and seldom allowed independent charge of her job.

Lot of works has been done on demographic population segmentation which is based on youth employment and unemployment, self-employment, earnings and occupational choice in labor market and child employment for finding number of people employed and unemployed and their earnings .Our approach takes in to account theeducational and employability characteristics to predict total percentage of women who are all literate and thosewho are employed or not. With this we will be able to predict the employability characteristics of woman.

II. RELATED WORKS

Lot of works has been done on demographic population segmentation.Companies and Managers have started and analyze on the social media in order to catch their target customers and form successful marketing strategies.

In [1] an approach has been proposed that provides valuable information for the marketing managers who may set their online marketing strategies especially in terms of segmentation and targeting successfully in the light of the findings of this study.

In [2] a customer segmentation approach has been applied on consumer data in order to increase the understanding of the elusive mobile services consumer markets, in a situation where few consumers are actual users of mobile services outside the early adopter category. This was done by using socio-demographic segmentation.

In [3]a technique for the heterogeneity existent within a seemingly homogenous sample of online consumer behavior in terms of their demographic profile was explored and analyzed. This was done by clustering algorithm.

In [4] an analysis of 1990 census data on the educational enrollment of 15- to 17-year-old immigrants to the United States was done which provides partial support for predictions from both the segmented-assimilation hypothesis and the immigrant optimism hypothesis.

In [5]a technique for labor market survey data is done to identify factors that determine men's and women's earnings, occupational choices, and mobility in segmented labor markets of India. The paper develops a model that considers 3 categories of labor--protected wage, unprotected wage, and self employment--representing 3 different forms of labor market segmentation according to the type of labor contract and job vulnerability.

In [6] an approach is proposed about emerging work on boundary theory by examining the extent to which individuals desire to integrate or segment their work and non-work lives. This desire is conceptualized and measured on a continuum ranging from segmentation to integration

of work and non-work roles. The fit between individuals' desires for integration/segmentation and their access to policies that enable boundary management was examined. Survey methodology is used for this segmentation.A tree based R programming application to segment a demographic population for predicting women employment characteristics with required data sets has been proposed to enroll maximum women in the employment sector.

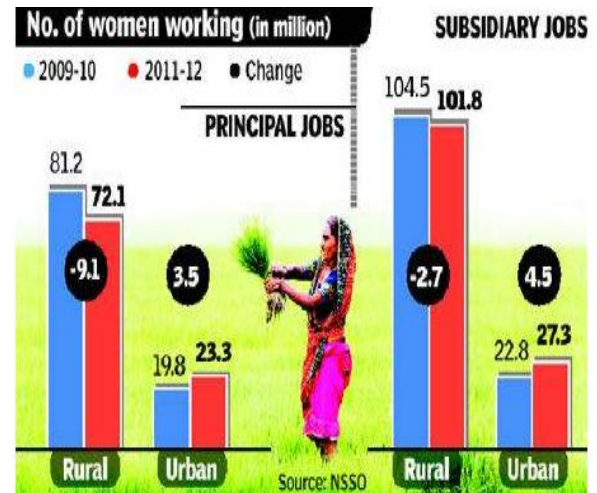


Fig.1. Statistics on women employment characteristics in [11]

III. PROPOSED SYSTEM

Women employees are motivated by different factors that are often dictated by their stage of life or personal interests. By learning what motivates each group and offering a customized program that delivers what matters most, companies find that women employees are willing to work more effectively and devote more discretionary thought and attention to their jobs. The illiterate and literate women in both urban and rural areas are calculated using data mining concepts implementing in R programming language with decision tree algorithm. With the obtained results the employment opportunities for both literate and illiterate women are identified and proposed to the corresponding sectors of women. This will further enhance the revenue of the nation and our economy. An approach for predicting the employability characteristics of woman has been proposed based on their demographic characteristics using a decision tree classification technique as in Fig.2, and Fig.3.

A. Steps Involved In Decision Tree Classification

- Get the input from any of the flat files.
- Extract the attributes in specific variables.
- Import the necessary package (Rpart) to create a decision tree.
- Declare the ctree
 - Check for age, if (age < 18) → child

- Else → adult
- Check for sex, if (sex = 'M') → male
 - Else → female
- Check for area, if (area = 1) → rural
- Else → urban
- Check for attendance, if (attendance exist = yes) → literate
 - Else → illiterate
- Check for work status, if (status = 1) → employed
- Else → unemployed

Under each category, node class gets generated.

- To make study on female working status with respect to attendance i.e., literacy.
- Plot the female unemployment vs age.
- Compare the above graph with the national economy of the country (here it's INDIA) theoretically.
- Deduce how economy gets improved if the % of active population increases.
- Display the graph.

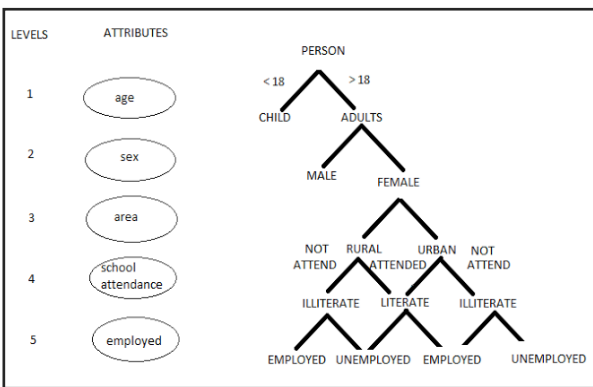


Fig.2. Proposed decision tree model

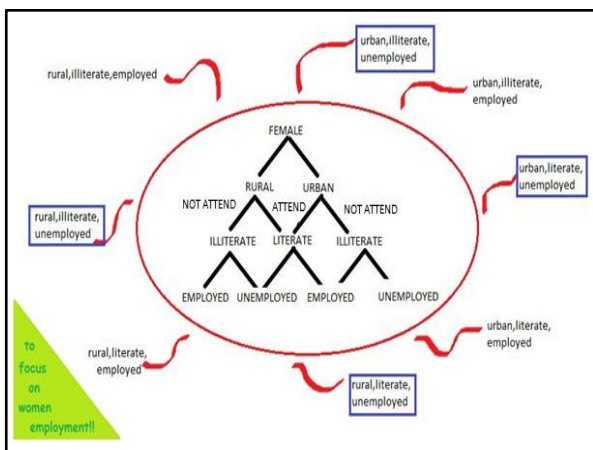


Fig.3. Pictorial Representation of Decision Tree Classification

IV. EXPERIMENTAL RESULTS

Here we use different data sets from UN Statistics Division as inputs. The inputs are derived from different datasets like Population by age, sex and urban/rural residence, Population 15 years of age and over by educational attainment, age and sex and Employed population by status in employment, age and sex. We combine the necessary fields from these datasets to form our input to the data mining process as in Fig. 4. We use R programming language to depict a statistical relationship between various attributes in the data sets. In Fig. 5, the graph explains about total amount of unemployed females in both rural and urban areas with respect to their age and also with respect to their attendance in school. From the graph we can understand that more than 75000000 females of age 50 are unemployed in rural areas. In Fig. 6, the graph explains about the unemployed women in rural areas with respect to their age. From the graph we can understand that the number of females attending school but are unemployed in the age group of 30 to 40 is high. In Fig. 7, the graph explains about the unemployed women in urban areas with respect to their age. From the graph we can understand that the number of females both attending and not attending school but are unemployed in the age group of 30 to 40 is higher than rural area.

AGE	SEX	AREA	STATUS	ACTIVITY STATUS
18	F	Urban	Attending school	Employed
19	F	Urban	Not attending school	Employed
20	F	Rural	Not attending school	Unemployed
21	F	Rural	Attending school	Unemployed
22	F	Urban	Attending school	Unemployed
23	F	Rural	Not attending school	Employed
24	F	Urban	Attending school	Employed
25	F	Rural	Attending school	Unemployed
26	F	Rural	Not attending school	Unemployed
27	F	Urban	Not attending school	Employed
28	F	Rural	Attending school	Employed
29	F	Urban	Not attending school	Employed
30	F	Rural	Attending school	Unemployed
31	F	Rural	Attending school	Unemployed
32	F	Urban	Attending school	Unemployed
33	F	Rural	Not attending school	Employed
34	F	Urban	Not attending school	Employed
35	F	Urban	Attending school	Employed
36	F	Rural	Attending school	Unemployed
37	F	Urban	Attending school	Unemployed
38	F	Urban	Not attending school	Unemployed
39	F	Rural	Attending school	Employed
40	F	Rural	Attending school	Employed
41	F	Rural	Not attending school	Employed
42	F	Urban	Not attending school	Unemployed
43	F	Urban	Attending school	Unemployed
44	F	Rural	Attending school	Employed
45	F	Rural	Not attending school	Employed
46	F	Urban	Not attending school	Employed
47	F	Urban	Attending school	Employed
48	F	Rural	Attending school	Unemployed
49	F	Rural	Not attending school	Unemployed
50	F	Urban	Not attending school	Unemployed

Fig.4 Input Dataset

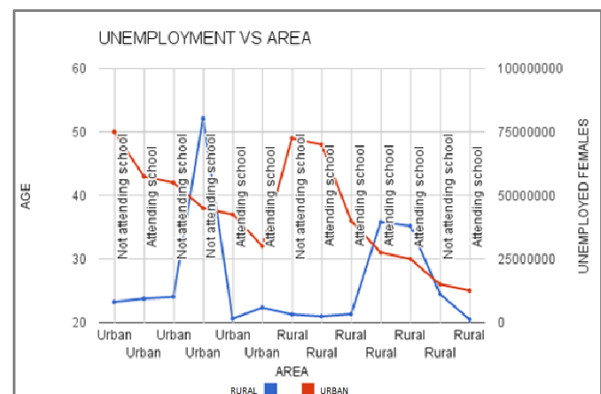


Fig.5. Unemployed Women Statistics in Urban and Rural Areas

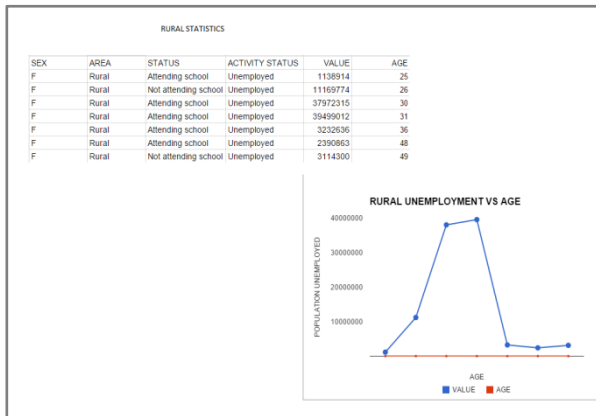


Fig.6. Rural Employment Statistics

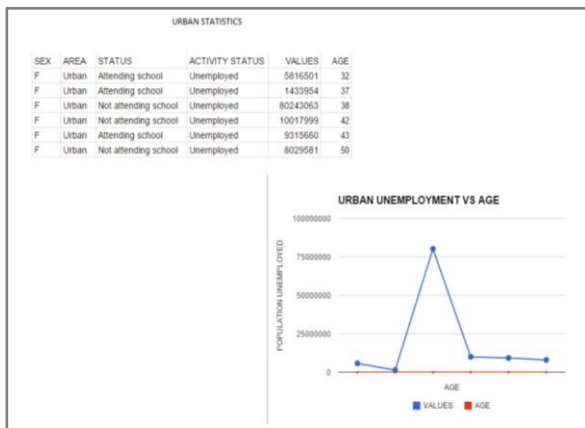


Fig.7. Urban Unemployment Statistics

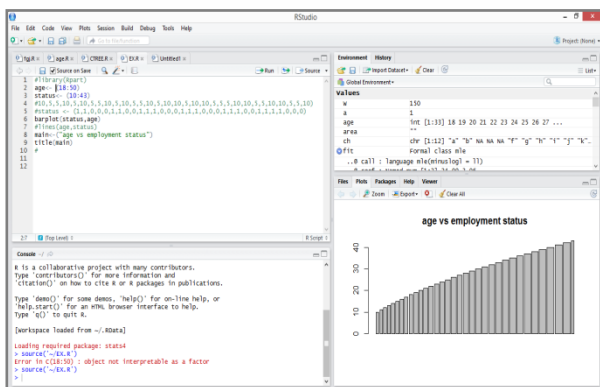


Fig.8. Screenshot of Experimental Results

V. CONCLUSION

A framework for employability characteristics prediction of woman has been developed and compared female unemployment against the literacy. We have taken a sample dataset provided by the UN dataset for our country India. Then there is a comparison made with respect to the national economy contextually. With this approach it is possible to predict the active population of a country. This work can be later extended to find many other characteristics like economic standard, living conditions in female and male, child and adult.

REFERENCES

- [1] Sell, A.Walden, P.: Segmentation Bases in the Mobile Service Market with attributes in and demographic out. System Science(HICSS),2012 45th Hawaii International Conference,January 2012
- [2] Arefin , A.S. Moscato , P. De Vries: Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference,December 2014
- [3] Cicek,M. Ozcan,S.: Control, Decision and Information Technologies (CoDIT), 2013 International Conference, May 2013
- [4] Tzyy-wen Lin, Hao-Erl Yang: Service Operations and Logistics, and Informatics, 2006. SOLI '06. IEEE International Conference,June 2006
- [5] Khandker.SR: Washington, D.C., World Bank, 1992.
- [6] Nancy P. Rothbard, Katherine W. Phillips, Tracy L. Dumas: Published Online June , 2005
- [7] SulekhaGoyat,The basis of market segmentation: a critical review of literature, European Journal of Business and Management www.iiste.org ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online) Vol 3, No.9, 2011