An Analysis on Finding the Influencing Factors of Preventing the Employment among the Educated Women using Data Mining Techniques

S.Dhivya
M.Phil(cs) Research scholar,
Department of Computer Science and Applications,
SCSVMV University, Enathur.
Kanchipuram.

S. Prakasam, M.C.A, Ph.D.,
Associate Professor,
Department of Computer Science and Applications,
SCSVMV University, Enathur,
Kanchipuram

ABSTRACT
The main objective of this work is to find the factors preventing the employment of educated women's. The purpose of this work is to investigate the obstacles that stand in the way of increasing women's employment. [8] The Questionnaire was designed to predict the factors preventing the employment among the educated women's. Family based reasons preventing women's employment and working environment based reasons preventing employment from these situations data is collected. The WEKA software is used for the study implementation since it is widely used for research purposes in the data mining field. In this dissertation author used Prediction techniques that predict categorical class labels. Ranking which gives binary or ordinal score.

1. INTRODUCTION
"Education is the transmission of civilization" it is an endless process which can be continued throughout our life. Education not moulds us but also built us. It may also pass on to generation to generation. Women's education is most impact and valuable thing. Educated women were bold enough to face the problem in day to day life and they find solutions to the problem. Educated women get success in each and every field in the world. Financial status of the family will increase. Women's are educated more than men they are neglected for working and the talents are suppressed by so many reasons for this major problem first and foremost thing is to take care of child and second family situation. A survey has been conducted during January-April 2014 with different category working women of 350. The author collecting data in before the factors prevent the employment of educated women's survey was distributed from face to face contact in and around Kanchipuram. The WEKA software is used for the study implementation since it is widely used for research purposes in the data mining field. In this dissertation author used Prediction techniques that predict categorical class labels. This analysis shows to identify potential drop outs of educated women's some of major affecting factors are: To take care of child, Health problem. Heavy work pressure, Financial status is satisfactory Husbands job is transferrable, Not felling comfortable with colleagues, Heavy work pressure. The prediction analysis shows the result obtain is young age women's preventing factor is to take care of child and for middle age women's preventing factor is Health problem. Ranking which gives binary or ordinal score to the attributes our ranking score is 136.699 To take care of child, 90.0384 Husbands' job is transferrable, 73.684 Not felling

comfortable with behavior of colleagues, 51.279 Heavy work pressure.

2. LITERATURE REVIEW
Discrimination against women is defined by (Mojbol Olfnk Okome) Article of the United Nations Convention on the Elimination of all forms of Discrimination Against Women of 1979 as "any distinction, exclusion or restriction made on the basis of sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, on a basis of equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field.[3] Department of Education Studies and Institute of Education of the University of Hull was asked to examine the social, economic, religious and other factors influencing the degree of female participation in formal education institutions in six carefully selected developing countries.[4] This session will focus on how and why education is the key to gender equality and to larger employment opportunities for women. It will explore links between lack of education and its impact on women's employment avenues. Taking a look at some of the trends, and more relevantly at some success stories, it will explore options that could provide women better avenues for education and employment. [9] Students' Dropout Risk Assessment case study describes the results of the educational data mining aimed at predicting the undergraduate courses of computer science (BCA and B.Tech.) students' instant dropout or after first semester. For this purpose, the preliminary data of 220 students collected randomly in prescheduled format on personal interview to find out dropout rates and dropout reasons. The simple and intuitive classifiers (decision trees) ID3 and J48 were used in this paper. The main reason recorded for dropout of students at this residential university were personal factor via; illness & homesickness, Educational factors via; learning problems & difficult courses, change of Institution with present goal and low placement rate and institutional factors such as campus environment, too many rules in hostel life and poor entertainment facilities. The information generated will be useful for better planning and implementation of educational program and infrastructure under measurable condition to increase the enrolment rate of students in ICT courses at this university.

3. METHODOLOGY

In this work, data mining technique based on Women's drop out prediction system was used. In this model we have used one of the classification algorithms called decision tree. These data analysis help us to provide a better understanding of large data. Classification predicts categorical and prediction models predict continuous valued functions. For example, we can build a classification model to categorize reasons preventing women's employment. Women's age between 20-32 are considered as young. So they must take care of child. This statement satisfy's the condition Yes. Women's age between 32-40 are considered as middle. So the chance to take care of child is less. This statement satisfy's the condition No.

3.1 Classification

Following are the examples of cases where the data analysis task is Classification:

- In the survey data analysis conclusion made that the much influencing factors preventing women's employment is to take care of child
- Young age mothers must spend their most of the time with their young one so the condition satisfies.

In this example we are bother to predict by Yes or No value. Therefore the data analysis task is example of prediction.

3.2 Building the Classifier or Model

This step is the learning step or the learning phase.

- In this step the classification algorithms is used to build the classifier.
- The classifier is built from the training set made up of database tuples and their associated class labels.

Each tuple that constitutes the training set is referred to as a category or class. These tuples can also be referred to as sample, object or data points.
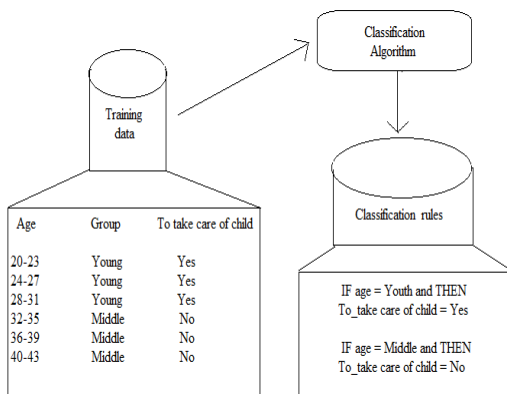

Fig 1: Classification algorithm for young and middle age people to take care of child

3.3 Using Classifier for Classification:

In this step the classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable.
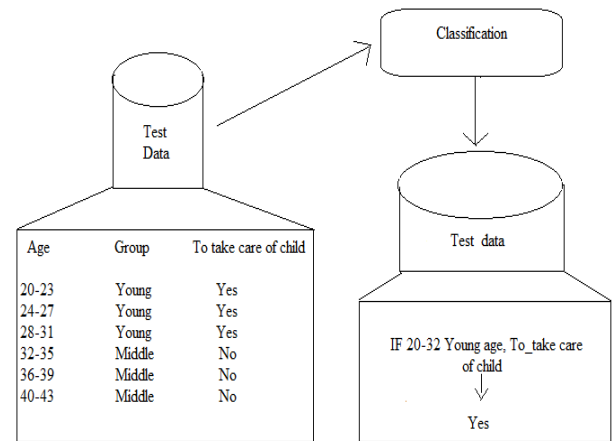

Fig 2: To predict the attribute take care of child by age in two scale opinion Yes or No

4. ANALYSIS

4.1 Experiments using WEKA for prediction (J48 Algorithm)

The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree & applying the tree to the dataset. The decision tree used in WEKA is termed as J 4.8 which is a modification of the C4.5 algorithm.J48 algorithm uses pruning method to build a tree. Pruning is a technique that reduces size of tree by removing over fitting data. The J48 algorithm recursively classifies data until it has been categorized as perfectly as possible. This technique gives maximum accuracy on training data. The overall concept is to build a tree that provides balance of flexibility & accuracy. In this we have used J48 algorithm to know the efficiency of prediction system. [1].

The confusion matrix showed the factors preventing the employment of educated women's. Here the prediction is done by these attribute present role, Financial background is satisfactory, health problem, aiming for government job, heavy work pressure, to take care of child and the predicted age result is described below. Classification on the test data is done based on the decision tree that is created.

The confusion matrix is more commonly named contingency table. In our case I have six classes, and therefore a 6x6 confusion matrix. The number of correctly classified instances is the sum of diagonals in the matrix; all others are incorrectly classified. For example the confusion matrix of J48 is shown below.

```
=== Confusion Matrix ===
 a  b  c  d  e  f  ← classified as
19  0  1  0  1  1 | a = 36to39
 6 13  1  3  3  1 | b = 32to35
 2  1 24  1  0  0 | c = 28to31
 0  0  1 27  1  0 | d = 20to23
 1  0  1  3 24  1 | e = 24to27
 1  1  1  2  0  9 | f = 40to43
```

Table 1: Prediction based on women's age

		PREDICTED						% of Correctly Predicted
		a	b	c	d	e	f	
O B S E R V E D	a=36-39	19(22)	0	1	0	1	1	86.36
	b=32-35	6	13(27)	1	3	3	1	48.14
	c=28-31	2	1	24(28)	1	0	0	85.71
	d=20-23	0	0	1	27(29)	1	0	93.10
	e=24-27	1	0	1	3	24(30)	1	80
	f=40-43	1	1	1	2	0	9(14)	64.28

Table 2: Age Vs To take care of child prediction table

Age Vs To take care of child		To take care of child					Total
		Agree	Disagree	Neutral	Strongly agree	Strongly disagree	
Age	20-23	4	8	12	9	4	37
	24-27	9	12	8	10	0	39
	28-31	14	0	6	21	0	41
	32-35	14	3	2	20	0	39
	36-39	6	0	0	22	0	28
	40-43	3	2	0	11	0	16
Total		50	25	28	93	4	200

From the above cross tabulation, women's drop out or the main reason for preventing employment of educated women is to take care of their child. Among 37 members of age from 20 to 23 have high influencing percent is 12 of them are neutral. Among 39 members of age from 24 to 27 have high influencing percent is 12 of them are disagree. Among 41 members of age from 28 to 31 have high influencing percent is 21of them are strongly agree. Among 39 members of age from 32 to 35 have high influencing percent is 20 of them are strongly agree. Among 28 members of age from 36 to 39 high influencing percent is 22 of them are strongly agree. Among 16 members of age from 40 to 43 high influencing percent is 11 of them are strongly agree. The above tabulation concludes that most of the women agree or strongly agree with the family based reasons preventing women's employment is To take care of child.[10]

Table 3: Result of study Age Vs To take care of child [16]

	Explanation	Result of study
Age	To find the person is young or middle age group (20-40)	Positive
To take care of child	young age women(20-32)	positive
	Middle age women (32-40)	Negative

Table 4: Age Vs Health problem prediction table

Age Vs Health problem	Health problem				Total
	Agree	Disagree	Neutral	Strongly agree	
Age 20-23	8	15	3	11	37
24-27	13	9	4	13	39
28-31	15	10	8	8	41
32-35	18	13	0	8	39
36-39	12	6	6	4	28
40-43	11	1	4	0	16
Total	77	54	25	44	200

From the above cross tabulation, women's drop out or the main reason for preventing employment of educated women is to take care of their child. Among 37 members of age from 20 to 23 have high influencing percent is 15 of them are Disagree. Among 39 members of age from 24 to 27 have high influencing percent is 13 of them are Agree and strongly agree. Among 41 members of age from 28 to 31 have high influencing percent is 15 of them are Agree. Among 39 members of age from 32 to 35 have high influencing percent is 18 of them are Agree. Among 28 members of age from 36 to 39 high influencing percent is 12 of them are Agree. . Among 16 members of age from 40 to 43 high influencing percent is 11 of them are Agree. The above tabulation concludes that most of the women agree or strongly agree with the family based reasons preventing women's employment is Health problem. Prediction analysis result from the above table shows. Women's age from 20 to 35 factors preventing the employment of educated women highly influenced factor is to take care of child. And for women's age from 36 to 43 highly influencing factor is Health problem [10]

Table 5: Result of study Age Vs Health problem [16]

	Explanation	Result of study
Age	To find the person is young or middle age group (20-40)	Positive
Health Problem	young age women(20-32)	Negative
	Middle age women(32-40)	Positive

4.2 Ranking

Data consists of lists of items with some partial order specified between items in each list. This order is typically induced by giving a numerical or ordinal score or a binary judgment.

Regression analysis is a statistical process for estimating the relationships among variables.

Search Method:Attribute ranking,Attribute Evaluator (supervised, Class (nominal): 1 Age):
Chi-squared Ranking Filter
Ranked attributes:
136.699   2 To take care of child
90.0384   8 Husbands' job is transferrable
67.5984 9 Ego problem between husband and wife
57.5988 6 Health problem
57.4772 7 Concentrating in higher education
57.3701 3 financial background is satisfactory
54.6801 4 Unable to manage both home and workplace
46.0899 5 Husbands salary is good enough
Selected attributes: 2,8,9,6,7,3,4,5 : 8
Result for Ranking algorithm. In this attribute evaluator method chosen is chi-squared attribute evaluator then in search method chosen is Ranker. Attribute selection mode chosen method is cross use full training set. Here output percentage for the given is clearly see the much influencing factors are To take care of child 136.699 and Husband job is transferrable 90.0384. These two are the

highly influencing factor which affects the employment of the educated women's family based reasons preventing women's employment. We Ranked these factor by the attribute age.
Ranked attributes:
73.684 3 Not feeling comfortable with behavior of colleagues
51.279 5 Heavy work pressure
38.719 2 Aiming for government job
33.355 7 Do not have proper transport facility
28.488 6 Workplace is located far from the residence
22.476 4 communication problem
Selected attributes: 3,5,2,7,6,4 : 6
Result for ranking algorithm. In this attribute evaluator method chosen is chi-squared attribute evaluator then in search method chosen is Ranker. Attribute selection mode chosen method is cross validation. Here output percentage for the given is clearly see the much influencing factors are not feeling comfortable with behavior of colleagues 73.684 and Heavy work pressure 51.279. These two are the highly influencing factor which affects the employment of the educated women's working environment based reasons preventing employment. We Ranked these factor by the attribute previous work experience. [12]

4.3 Ranked attributes in Regression

Table 6: Significant table for Age Vs to take care of child

Age * To take care of child	Estimate	Std. Error	df	Sig.
[Age = 20 to 23]	16.492	.215	1	.000
[Age = 24 to 27]	17.801	.169	1	.000
[Age = 28 to 31]	18.772	.151	1	.000
[Age = 32 to 35]	19.838	.158	1	.000
[Age = 36 to 39]	21.127	.217	1	.000
[To take care of child = Agree]	18.728	.237	1	.000
[To take care of child = Disagree]	16.796	.322	1	.000
[To take care of child = Neutral]	16.735	.320	1	.000
[To take care of child = Strongly agree]	19.085	.000	1	.
[To take care of child = Strongly disagree]	0[a]	.	.	0

Regression table shows that the significant value .000 in the case of to take care of child based on women's age.
Family based reasons preventing women's employment. Significant value is P=.000(P Value<0.05) [13].

5. CONCLUSION

The result achieved by applying selected data mining algorithms for predicting factors preventing the employment among the educated women's. We categorized our prediction analysis in two different focuses based on their age. Age from 20-32 young women's most influencing factors preventing employment is to take care of child and women's age from 32-40 most influencing factor is health problem. Result of ranking is 136.699 to take care of child, 90.0384 Husbands' job is transferrable, 73.684 Not felling comfortable with behavior of colleagues, 51.279 Heavy work pressure. Ranking is done by two situations family based situation and working environment based reasons preventing employment of educated women. The conclusion made from the conducted research will be used for lead a way in increasing women's employment rate.

6. REFERENCES

[1] A.Priyanga M.Phil(cs) Research Scholar SCSVMV University Enathur, Kanchipuram. S. Prakasam, PhD Assistant Professor Department of CSA, SCSVMV University Enathur, Kanchipuram. Effectiveness of Data Mining – based Cancer Prediction system (DMBCPS). International journal of computer applications(0975-8887) volume 83-No 10,December 2013.

[2] Department for Business and innovation and skills FEB 2013.

[3] Domestic, regional, and international protection of Nigerian women against discrimination: constraints and possibilities by Mojbol Olfnk Okome

[4] Factors affecting female participation in education in seven developing countries

[5] Factors affecting women's participation in university Management in Kenya by Jane onsongo.

[6] Factors Enabling and Constraining Women's Political Advancement in the Pacific

[7] Factors preventing young people from continuing in education

[8] Factors That Affect Women's Labour Force Participation and Suggestions for Provincial Employment and Vocational Education Boards: Ankara, Gaziantep, And Konya Hakan Ercan,Ayşe Gündüz Hoşgör,Özlem Yılmaz., Middle East Technical University, Ankara January 2010.

[9] Gender inequality in education and employment

[10] Gerben W. Dekker1, Mykola Pechenizkiy2 and Jan M. Vleeshouwers1 g.w.dekker@student.tue.nl, {m.pechenizkiy, j.m.vleeshouwers}@tue.nl. Predicting Students Drop Out: A Case Study (i) Department of Electrical Engineering, Eindhoven University of Technology, the Netherlands. (ii) Department of Computer Science, Eindhoven University of Technology, the Netherlands.

[11] J.F Superby and J.P. Vandamme N. Meskens Determination of factors influencing the achievement of the first-year university students using data mining methods Production and operations management Department, Catholic University of Mons, Chaussee de binch 151,7000 Mons, Belgium.

[12] Kawsar Ahmed , Ahsan Habib , Tasnuba Jesmin , Zamilur Rahman , Badrul Alam Miah Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Bangladesh Md. Prediction of Breast Cancer Risk Level with Risk Factors in Perspective to Bangladeshi Women using Data Mining. International Journal of Computer Applications (0975 –8887) Volume 82 –No4, November 2013 36

[13] M Prema and S Prakasam Effectiveness of Data Mining - based E-learning system (DMBELS). International Journal of Computer Applications (0975-8887) Volume 66-No.19,March 2013.

[14] Statistics Canada by government of Canada

[15] Students' Dropout Risk Assessment in Undergraduate Courses of ICT at Residential University – A Case Study.

[16] Syed Tahir Hij nd S.M.M. Raza Naqvi, azil FACTORS AFFECTING STUDENTS' PERFORMANCE. A Case of Private Colleges Bangladesh e-Journal of Sociology. Volume 3. Number 1. January 2006.

[17] The fiscal Times by Liz Peek

[18] Women's Employment in Saudi Arabia. A major challenge.