



Data Mining to Elicit Factors Causing Infertility in Women's in Rural and Urban Areas

Mrs. Anamika Arora
Research Scholar
Indore, India

Dr. Pradeep Chouksey
Vice Principal, TIT
Bhopal, India

Abstract- In India each family had five to six children but from last decade's low fertility is growing rapidly in both urban and rural areas of India. Economic and social modernization in India with gain in education, obesity and age led to lower fertility in every region of India. In this paper Data mining techniques like statistical analysis, classification, Naives bayes, J48 has been used to analyse and mine knowledge on significant factors causing infertility in women. Even though there are many factors were causing infertility in women, but only three factors Age, Obesity (BMI) and Education taken for analysis. Samples are taken from both rural and urban areas of Indore. Out of several independent attributes collected about outpatients, only three attributes consider being significant have been taken up for preliminary study. The aim of the study is to assess both in the rural and urban areas they said factors in the light of fertility in women. Common attributes have been considered among an equal sample size of fertile and infertile outpatients of rural and urban areas. The result of the study shows that the attributes considered are significant in determining fertility of women both individually and together. In the research found that the education effects age of the marriage of women and age influence body mass index (BMI) and obesity triggers changes in hormonal levels which is the cause of the infertility in urban area's whereas in rural area's the percentage of infertility is low due to education and the age of the marriages is earlier compare to urban areas.

Keywords- Data mining, Statistical analysis, classification, Naive bayes, J48

I. INTRODUCTION

Fertility is the ability of a person, animal or plant to reproduce by natural means. In humans infertility may describe women who are unable to carry a pregnancy to full term. There are many biological women who are unable to carry a pregnancy a full term. There are many biological & other causes of infertility in both rural and urban areas. Common causes of infertility in females include ovulation problems tubal blockage, pelvis inflammations, obesity, age related factors, education, endometriosis or advanced maternal age. Infertility is a life crises with invisible lose and its Consequences are manifold. World health organisation (WHO) defines infertility as failure to conceive after completion of two years of marriage in India in last 1970 to 1980 every family have four to five children's but according to survey of NSFG (National Survey of Family Growth) in 2000 every family have one or two children and NSFG find the growth of infertility in families. The survey found

that drop in children in the family from average 4:1 to 5:0 in mid 1970 to 1980 to 2:04 in 1990[1] this is defined as an average of 2:1 children per women. If this fertility level is maintained population growth will slow and eventually population size will stable.

In this paper Data mining techniques like statistical analysis, classification and association rule mining has been used to analyse and mine knowledge on significant factors causing infertility in women both in rural and urban areas. Even though there are many factors causing infertility in women,

but only three factors are taken Age Obesity (BMI) and Education for analysis, samples taken from both rural and urban areas of Indore. Out of several independent attributes collected about outpatients, only three attributes consider being significant have been taken up for preliminary study. Among these the most significant ones leading to infertility are identified. Characteristics of each significant factor are studied in fertile and infertile women leading to knowledge discovery of cause of infertility in each case. The results of the study show that the attributes considered are significant in determining fertility of women both individually and together. In the research found that education effects age of marriage and age influence body mass index and obesity triggers changes in hormonal levels in urban areas whereas in rural areas the percentage of infertility is low due to education and age of marriage is earlier as compare to urban areas. Statistical Analysis has been used to analyze and mine knowledge on significant factors causing infertility in women. Association rules that govern infertility are also generated using Associates Rule Mining

II. LITERATURE REVIEW

There are many causes of infertility Education, Age and BMI play a significant role. Impact of education has been recognized as a crucial factor influencing women's childbearing patterns. The availability of data from large no of families that participated in the world family survey (WFS) in 1970 to till 2000 there is a dropped in the childbearing pattern of women's in late 1970,s the age of marriage around 18 to 20 as this pattern changes as women devoted to education which is significantly associated with

low fertility [1]. As the higher the age of marriage the infertility will rapidly develop. In table 1 I have shown the changes in fertility rates from 1970 and in 2014

Table I Total fertility rates 1970 and 2014

total fertility rates 1970 and 2014	
YEAR	
1970	2014
4.53	1.16
Source-(NSFG) in India	

There are some key dimension of education and subsequently explore their implications for fertility behaviour.

2.1 Education as a source of knowledge – The transmission of knowledge is the first goal of schooling. Schooling imparts literacy skills, enables pupils to process a wide range of information and stimulates the shape of the behaviour and interaction with the surrounding world.

2.2 Education is a vehicle of socioeconomics advancement – Education opens up economic opportunities and social mobility. In today's societies education is treated as for entry into formal employment and for sorting individuals into hierarchy of occupations.

2.3 Education as a transformer of attitudes – education transforms the level of questioning traditional beliefs and the society's structures.

We can assume or say that these three dimensions of education have an impact on women's thoughts, reproductive desires and behaviour. In table 2 illustrates the three major drawbacks of infertility growth i.e., time of college completion, labor force participation and marriage among women's by all these there is impact in age of women's which is the major lead of infertility. Schooling completion significantly associated with the age of marriage among women in India.

Table II College completion, labor force participation and marriage among women's ages 25 -35 in India in 1970, 1990, 2014

College completion, labor force participation and marriage among women's ages 25-35 in India in 1970, 1990, 2014.								
Completed four years of college			Employed			Married		
1970	1990	2014	1970	1990	2014	1970	1990	2014
10	65	90	6	45	75	80	60	45
Source-survey with the help of data mining								

Age is one of the most common infertility risk factors for women [2]. As women ages her fertility starts to decline as a result of her ovaries producing fewer eggs. A woman's egg quantity and quality begins to sharply decline in her mid – 30's and deterioration accelerates rapidly in her late 30's making conception more challenging [3]. There's also an increased risk of miscarriage in later years. Statistics show that about 1/3 couples in which the women is older than 35 have fertility problems. It is likely that the number of ova is limited after reproductive age; that factor alone can reduce fertility and increase the risk of birth defects in the foetus and newborn. In addition men over 35 compared with men under 35 have less chances of fertility and the risk of chromosomal abnormalities is higher in their children.

According to global public health the obesity rate is increasing very fast is one of the major factors of infertility. Approximately 65% adults are overweight i.e. BMI is greater than 30[4]. It is also recognized that obese is also factor of infertility in women. The relative risk of anovulatory infertility is 2.9 in women with BMI >32 kg/m² at the age 18 while in ovulatory the chances of spontaneous conception decreases by 7% for each unit increase in the BMI [4]. Obesity occurs in 35% of women causes PCOS increase the magnitude of hormonal and metabolic dysfunction in these women [5]. Hormonal misbalance in women's lead to disruption of the normal secretion of luteinizing hormone(LH) and follicular stimulating hormone(FSH) to pulsatile secretion of gonadotropin releasing hormone is evidenced in a number of reproductive disorders in women. TSH are considered important components of the infertility. Thyroid dysfunction rise to aspects of reproduction and pregnancy. Many researchers found TSH play major role in fecundity and increased morbidity during pregnancy.

III. METHODOLOGY

A. Statistical Analysis

Waikato environment for knowledge analysis (Weka) [6] is a suite of machine learning software written in Java, developed at the University of Waikato, New Zealand .it is free software licensed under the GNU General Public Licence. Weka contains a collection of visualization tools and algorithms for data analysis and predictive modelling, together with graphical user interfaces for easy access to these functions. Weka support several standard data mining tasks, more specifically, data pre-processing, clustering, classification, regression, visualization and feature selection. Weka is a popular data mining tool. It is used to analyze the most significant factors causing infertility. It is also used to perform statistical analysis of each individual attribute [8].

B. Naives Byes

The Bayesian classification represents a supervised learning method as well as a statistical method for classification [9]. It comes under probabilistic model and allows capturing uncertainty about the model in a principled way by determining probabilities of the outcomes. Bayesian classification solves diagnostic and predictive problems [10].

This algorithm use the bayes formula which calculate the probability of a data record having the class label ci

$$P(\text{label}=ci/Y) = P(Y/\text{label}=ci) * P(ci)$$

$$P(Y)$$

The class label ci with the largest conditional probability value determines the category of the data record. Dominator P(Y) can be safely eliminated as it does not depend on the label.

The actual values of features a^1, a^2, \dots, a^n for data record Y be equal to a^1, a^2, \dots, a^n .

The above probability can be rewritten as

$$P(\text{label}=ci/Y) = P(ci) * P(a^i=a^i/ci)$$

Where $P(a^i=a^i)$ is the ratio of the samples that have a^i for the i^{th} feature, among all the samples with class label ci and $P(ci)$ is the ratio of the samples with class label ci to all the available samples.

C. J48

J48 is an open source java implementation of the c4.5 algorithm in the WEKA data mining tool. C4.5 is a program that creates a decision tree. This algorithm was developed by Ross Quinlan.

In WEKA [11] the implementation of a particular learning algorithm is encapsulated in a class, and it may depend on other classes for some of its functionality. J48 class does not actually contain any code for building a decision tree. It includes references to instances of other classes that do most of the work.

IV. EXPERIMENTAL RESULTS

A. Using the WEKA the statistical analysis done; the result revealed the following facts:

1. Patients with higher education are more likely to be infertile in urban areas as compare to rural areas.
2. Higher education delay the marriage age i.e. patient with age > 30 is prone to infertility.
3. At the increase in age lead to increase in BMI, may be revealed to infertility

B. Here we did 10-fold cross validation for the classifiers and the result summarizes figure 1 shows the experimental result we have carried out in our research using classification algorithms for predicting the cause of infertility in women's both urban and rural areas.

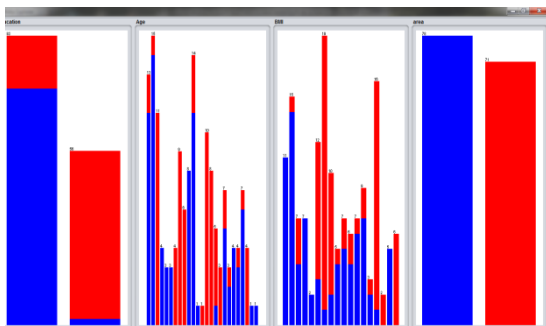


Figure 1: visual form of dataset using all attributes

Table II shows the experimental result .we have carried out some experiments in order to evaluate the performance and

usefulness of different classification algorithms for predicting infertility in women's of rural and urban areas

TABLE III PERFORMANCE OF THE CLASSIFIER

Evaluation Criteria	classifiers	
	Naive Bayes	J48
Timing to build model	0.02	0.05
Correctly classified instances	134	136
Incorrectly classified instance	15	13
Accuracy	89.93%	91.25%

From the above table we can conclude that j48 is more accurate than naive bayes algorithm j48 showing more accuracy than naives bayes.

In table IV Kappa statistic [7], mean absolute error and root mean squared error will be in numeric value only. in the table we have shown the relative absolute and root relative squared error in percentage for reference and evaluation. The result of the simulation is shown in tables IV.

TABLE IV. TRAINING AND SIMULATION ERROR

Evaluation Criteria	Classifiers	
	Naive Bayes	J48
Kappa Statistics(KS)	0.797	0.8239
Mean absolute error(MAE)	0.1281	0.1713
Root Mean Squared error(RMSE)	0.2941	0.3193
Relative absolute error(RAE)	25.67	34.33
Root Relative Square Error(RRSE)	58.86	63.92

TABLE V. COMPARISON OF ACCURACY MEASURE

Classifier	TP	FP	PRECISION	RECALL	CLASS
Naive Bayes	0.949	0.155	0.871	0.949	URBAN
	0.845	0.051	0.59	0.84	RURAL
J48	0.987	0.169	0.865	0.831	URBAN
	0.831	0.013	0.175	0.178	RURAL

Confusion matrix displays the frequency of correct and incorrect predictions. It compares the actual values in the test dataset with the predicted values in the trained model. The columns represent the predictions, and the rows represent the actual class. To evaluate the robustness of

classifier, the usual methodology is to perform cross validation on the classifier. The confusion matrix shown in table

Table VI. CONFUSION MATRIX

Classifier	Urban	Rural	CLASS
Naive Bayes	77	1	URBAN
	12	59	RURAL
J48	74	4	URBAN
	11	60	RURAL

V. CONCLUSION

In this paper statistical analysis and classification techniques is evaluated based on the selected classifiers algorithm, we used two common data mining algorithms naive bayes and J48 algorithm have been used to analyze and mine knowledge on significant factors causing infertility in women. There are many factors causing infertility but we have taken only three factors Education, Age and BMI .Sample data are collected from rural and urban areas of Indore. out of several independent attributes collected about outpatients, only three attributes considered to be significant have been taken up for preliminary study. Common attributes and equal size of sample collected from rural and urban areas of Indore. The results of the study shows the education affects the age of women marriage due to which the age is greater than 30 which influences the BMI, BMI triggers hormone levels which is the cause of infertility .the result show that in urban areas the infertility percentage is high as compare to rural areas.

VI. REFERENCE

- [1] World fertility survey, united nation, Department of international economic and social affair."Fertility behaviour in the context of development:"united nation, 1987, pp 383, 16 August 2008.
- [2] NSFG "increased infertility with age in men and women" 2004 Reproduction, vol.24 (10), pp 226-32 Jan 2009.
- [3] Vander steeg J W, Steves P, Eijkemans, M MJ, Habbema, J D, Hompes, PG, Burggraaf, JM, et.al."Obesity affects spontaneous pregnancy chances in sub fertile, ovulatory women."Human reproduction, Vol.23 (2), pp 324-328, Feb 2008.
- [4] Nohr EA, vaith M, Rasmussen S, Ramlav-hansen CH, Olsen J."Waiting time to pregnancy according to maternal birth weight and pregnancy BMI."Human Vol 8(6), pp 515-23, Nov2007.
- [5] Metwally M.L, TC, Ledger WL."The impact of obesity on female reproductive function"; obesity review, Vol 8(6), pp 515-23, Nov2007.
- [6] K Meena, N Vijyalakshmi, "Analysis of Factors causing infertility in women using statistical Analysis and association Rule Mining", Indian public of health Research and Development, Vol 6(2), pp 112-117, april-June 2015.
- [7] Sim, J Wright C. C (2005), "The Kappa Statistic in Reliability studies: Use, Interpretation and sample size requirement."85, pp 257-268.
- [8] L. Breiman, "Statistical Analysis," machine Learning 26, 1996, pp 123-140.
- [9] John, George H; Langley, Pat (1995), "Estimating in continuous Distribution in Bayesian Classifier". Proc. Eleventh Conference on uncertainty in Artificial Intelligence. Morgan Kaufman Vol 2(8), pp 338-385.
- [10] Ramana, B.V.M.S.P. Babu and N.B. Venkateswarlu, "a critical evaluation of Bayesian classifier for liver diagnosis using boosting methods." International Journal of engineering sciences and technology, 2011. Vol.3.
- [11] Rohit Arora, Suman. "Comparative Analysis of classification algorithms on different datasets using WEKA"; IJCA (0957-8887) Vol.54, pp13-15.