

A Data Mining Techniques for Analysis of Indian Census Data for Girl Child Education

Mrs. Swaroopa Shastri¹, Usha M. Madiwal²

¹Assistant Professor, ²PG Student

Department of Computer Applications, Visvesvaraya Technological University Postgraduate Centre Kalaburagi

Abstract - The project Analysis of Indian Census Data for Girl Child Education provides reasons which are effecting the girl child education, to the Government of India. In the current system the government collect the date and gives the public and access to the data. However the data is completely meaningless and that census data doesn't give any interpretation, a data can give interpretation only when it is cross checked with other data through some mathematical expressions. The Indian census is the most credible resource of information on Demography like Economic Activity, Literacy and Education, Housing & Household Amenities, languages, and so on. Every ten years the government of India carries the census, they will go each and every home and ask questions like how many members are there, how many males are there and how many are females, and they ask about house related questions, that is how many cylinders are there and fill the Census form. In proposed system using filtering, clustering, classification and visualization techniques we build system, that takes Census data as input and implements some data mining technique on that data, and that data will be in the table form, next it will compare each data with other data to find out the causes which are effecting the women care and education, Further our system comes with sophisticated visualization technique, by means of which visualization as statistical technique, but not only the findings are given as the text message but at the same time they depicted graphically.

Keywords: Filtering, Association rule mining and Visualization.

I. INTRODUCTION

Indian census means every ten years Government of India carries out a census, so people will go from home to home then ask how many members are there ,how many old people are there, how many females are there, how many children s are there and where they work, government employee or private, what is the salary they earn, and the wife is working or housewife and which school the children s are going government or private and how many cylinders are there one or two and anyone drop out from the school, when he or she drop out the school, like that all the question that are related to a people, so this census carried out literally in every home in India and then ask cost, how much income and which religion so this whole data is aggregated and put inside a website called census.govt.in. The census process involves visiting each and every household and collecting particulars by asking

questions and filling up Census Forms The first Indian census will be conducted in 1871 and till now the Indian census will be conducted 15 times and the last census will conducted in 2011 and it was the 15th census and 7th census after the independence, and next it will be conducted in 2021. The census will carried out by the Government servants appointed as Enumerators and they have identity card with them that will provided by government of India. The census form includes two forms first form relates to the House listing and Housing Census. In this 35 questions relating to Building material, Use of House, Drinking water, Availability and type of latrines, Electricity, possession of assets etc will be canvassed. and the second form related to National Population Register [NPR], this will be a comprehensive identity database this include details such as Name, Date of Birth, Place of Birth, Sex, Marital Status, Present, Permanent Address, Name of Father, Mother and Spouse, Occupation, Nationality as Declared, Educational Qualification, Relationship to Head of family will be collected by visiting each and every household. All usual residents will be eligible to be included irrespective of their Nationality. The Census will conduct different dates for states, in Karnataka it will conduct on 15th April and also in Arunachal and Chandigarh.

II. SYSTEM MODEL

This paper mainly discussed about the women development in India, especially in Education field. This pa per done analysis based on some issues of gender and they are senior education in human being progress and other fields like financial side, Education, civilization and politics benefits. We all know that women's are the half of the world, and this paper is helpful to women to increase their Empowerment and this empowerment requires the whole childhood behaviour to build the human Development and we know other state have good women human development achievements rather than in India, because in India there is no equality in men and women and society does not give equal respect to men and women. This paper really helped me a lot to knowing the women education and how it will relate to promoting the women empowerment and human Development [1].In

this paper the author discussed about the inequalities in educational opportunity the author ask that why the most of the children's are still out of School what are the reasons behind those children's who have not coming school. After studying this paper it will give three important issues why inequality in educational opportunity. The first inequality educational opportunity is found in the sex ratio and regions and income group, how much they earn and second one is the checking of adult women empowerment in education and improving their school attendance and third one is the estimation of explanatory power. This paper is useful to this project because this project calculate the female literacy and it will compared with all parameters those are help to improve the women education and this project give the reasons why the female literacy is low in some region, however this paper also give why inequality in men and women education and what are the issues that effecting the women Empowerment. [2]. this paper discussed about the Education, the education is one of the important part of the every human being to build a good development. This paper gives all the information about Education how it will important and how it will be increased. This paper gives some important points that points are very useful to improve the Education. Those are first one is it will discussed about how many Educational Institutions are there in one District and those schools have primary, Primary with upper primary and higher secondary and it will also give information about teachers, that how many teachers are appointed in one district. And this paper also discussed about the infrastructure of school, that is how many schools are have the building facility and how many are having drinking water facility and how many have sanitation facility and how many have blackboard facility in classrooms. This paper is really helped this project a lot, through this paper it will clearly mentioned that how the Education improved, For improvement of the Education this paper mentioned some points and those points are really helped to this Project to increase the Female Literacy. [3]. This paper mainly tells about child labour in India, this paper tells how children's are in labour field and how they work, the work may be many types some are most dangerous to them and as well as country also we all are know that India is the topmost country and this facing child labour issue is very high in India. For this the author want to the Government of India take some actions that are helpful to avoid this child labour issue. And this paper is very helpful to this project because this project is mainly for finding the girl child education so this paper is child labour issue than it is tells that the reason behind the child labour issue that reasons are also helpful to this project for finding female literacy because the child labour may be girls also that's why this paper is very helpful to this project. [4]. the paper

provides information about Government funds that are sanctioned for girl's child education in Rural India for public schools, because the female literacy is low in the villages. In villages the parents will not send their girls to school they will think that the girl education not important, in villages the schools are also not good and the schools are not got all the facilities, for that the Government of India provide the funds and facilities to encourage the girls to come school. For that this paper will help to increase the girl's enrolment in the upper primary schools in rural areas of India. And this paper will assist the project, in the project it will find out the female literacy, where the girl's education is high and in which state it is low. For that this paper will really very useful to find out the female literacy and why the literacy is low in some state or villages. [5]. this paper is mainly for women empowerment. For this education is the milestone of women empowerment, because it allows them to take actions to the problems, to face their traditional character and change their life. This paper also help them to decrease the inequalities between the families, the girl child education is one of the important main issue in the India. The women's growth in the rural areas is very slow; this means that still in our country the women are illiterate. For this our project is very helpful to women's to solve their problems which are effecting their growth, because our project is mainly for girl child education and the main objectives of the project is to find out the reasons that are effecting the women growth and another one important objective is identified the issues which are effecting the women's education and it will give a solutions to those problems in the text message and also displayed in the graph format. The Government of India take the appropriate action to solve those problems and sanctioned money to women education. [6]. There are major challenges to determine whether improvements in child survival are seen in both males and females where females are unnecessarily dying. And this project using the Making Comparison challenging technique. This project marks a significant advance towards profiling and subsequently tackling the issue of gender inequities in child health mortality. This disaggregated data that incorporates age, region within a country; wealth and education of the family are important covariates to be studied in relation to gender when looking at child health care access and outcome. This paper is very useful to this project because the project contains the census data in that the female's data will be there, so we can say that it will help to know the female sex ratio. [7]. The present study was carried out to identify the beneficiary level factors of utilization of Janani Suraksha Yojana scheme in urban slums and resettlement colonies of trans-Yamuna area of Delhi. The 469 mothers interviewed, 333 (71%) had institutional delivery, 128 (27.3%) had benefited from

JSY scheme and 68 (14.5%) had received cash benefits of Janani Suraksha Yojana. There is a need to improve the awareness among urban slum population about the utilization of Janani Suraksha Yojana scheme. Targeting difficult to access areas with special measures and encouraging more antenatal visits were essential, prerequisites to improve the impact of JSY. The Project Census data contains the population data so this paper is related to this project, in urban areas there is no knowledge of government yojana. It will help to government to know the how much the janani suraksha yojana is useful to urban slum population. [8].

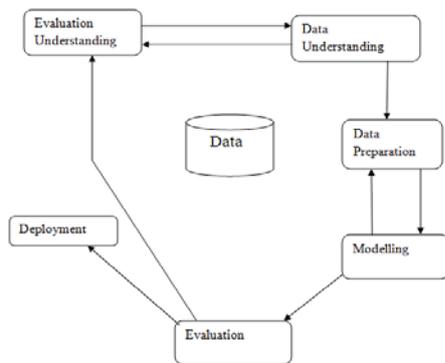


Fig.1 Architecture Diagram

This is the Architecture diagram of a project, in this first it will done data preparation that is filtering the data after filtering the data the data will be send to the modelling state after modelling the data that is extracting the rule then it will be evaluating that is evaluation of the data then it will be repeating the steps that are Evaluation understanding, preparation, modelling, evaluation, deployment. These all steps are repeat up to complete all the data.

III. PREVIOUS WORK

In existing system the government just collect data, converts it in to software keeps in to a website, the data is called raw data, and raw data in itself doesn't reveal any cause or any information. Therefore it need to mine this data through certain mathematical process, such that it can find out the hidden cause in this data, so that presence system doesn't reveal any cause, it doesn't reveal any analysis and it doesn't reveal any relationship.

IV. PROPOSED METHODOLOGY

This block diagram discuss about the modules of the project. They are

1. Data Gathering
2. Data Pre-processing
3. Dataset

4. Clustering
5. Classification
6. Visualization
7. Knowledge

1. Data Gathering:

The data gathering module first collect or gather the data from Excel sheet that are present in the Indian Census website. The government of India stored the all the information or Data in the website, applying some Data mining tecnique we gathering the data from Excel Sheet.

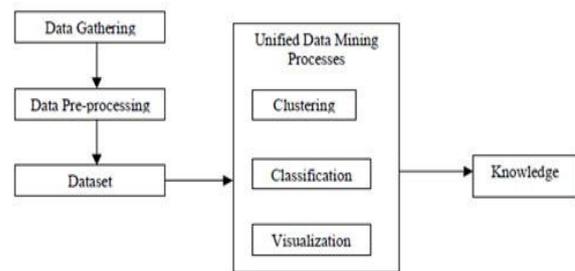


Fig 2 Block Diagram

2. Data Pre-processing:

The Data pre-processing means filtering the data, in this project the low pass filtering will be used. The filtering means removing outliers, that is occurrence of high frequency data, that is the data which is occurring long time period, that is it will not occur again and again after filtering the data we get a filtered data and that will give to the Data Modelling Process. In the modelling process the filtered data is assign to the classification algorithm this algorithm state that which country or state is high and which one is low.

3. Clustering:

Clustering is the one of the important module in the Project, the clustering means it will group or clustered the data based on some parameters, like in the project many type of data will be there. For example the data may be Population data, Education data, Health data and Income data based on these data the clustering will be done.

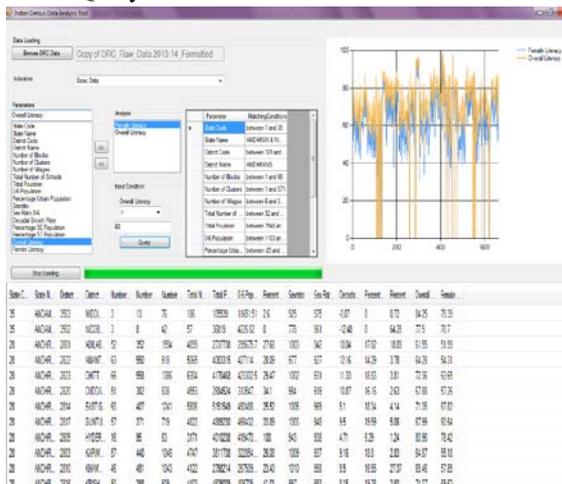
4. Classification:

Classification will be done on some rules, this classification is classify the data based on some rules that is it classify that which state is high and which one is low in Population and it will classify that sex ratio of men and female based on comparison of men versus female and later it will be compared with other state data to know which one is high and which one is low in the sex ratio of men and female.

State	State ID	State Name	State Code	Number	Number	State ID	Total P	65%	Percent	Percent	Percent	Percent	Percent	Percent
20	ANP	ANDAMAN & NICOBAR ISLANDS	AN	13	76	100	100	100	100	100	100	100	100	100
20	AP	ANDHRA PRADESH	AP	13	106	100	100	100	100	100	100	100	100	100
20	AR	ARUNACHAL PRADESH	AR	13	106	100	100	100	100	100	100	100	100	100
20	AS	ASSAM	AS	13	106	100	100	100	100	100	100	100	100	100
20	BR	BIHAR	BR	13	106	100	100	100	100	100	100	100	100	100
20	CH	CHHATTISGARH	CT	13	106	100	100	100	100	100	100	100	100	100
20	DN	DADRA AND NAGAR HAVELI	DN	13	106	100	100	100	100	100	100	100	100	100
20	DL	DELHI	DL	13	106	100	100	100	100	100	100	100	100	100
20	GA	GOA	GA	13	106	100	100	100	100	100	100	100	100	100
20	GJ	GUJARAT	GJ	13	106	100	100	100	100	100	100	100	100	100
20	HR	HARYANA	HR	13	106	100	100	100	100	100	100	100	100	100
20	HP	HIMACHAL PRADESH	HP	13	106	100	100	100	100	100	100	100	100	100
20	JK	JAMMU AND KASHMIR	JK	13	106	100	100	100	100	100	100	100	100	100
20	KA	KARNATAKA	KA	13	106	100	100	100	100	100	100	100	100	100
20	KL	KERALA	KL	13	106	100	100	100	100	100	100	100	100	100
20	MP	MADHYA PRADESH	MP	13	106	100	100	100	100	100	100	100	100	100
20	MH	MHARASHTRA	MH	13	106	100	100	100	100	100	100	100	100	100
20	MZ	MEGHALAYA	MZ	13	106	100	100	100	100	100	100	100	100	100
20	NP	NIJZORAM	NP	13	106	100	100	100	100	100	100	100	100	100
20	NL	NAGALAND	NL	13	106	100	100	100	100	100	100	100	100	100
20	OR	ODISHA	OR	13	106	100	100	100	100	100	100	100	100	100
20	PU	PUDUCHERRY	PU	13	106	100	100	100	100	100	100	100	100	100
20	PJ	PUNJAB	PJ	13	106	100	100	100	100	100	100	100	100	100
20	RJ	RAJASTHAN	RJ	13	106	100	100	100	100	100	100	100	100	100
20	SK	SIKKIM	SK	13	106	100	100	100	100	100	100	100	100	100
20	TN	TAMIL NADU	TN	13	106	100	100	100	100	100	100	100	100	100
20	TR	TRIPURA	TR	13	106	100	100	100	100	100	100	100	100	100
20	UP	UTTAR PRADESH	UP	13	106	100	100	100	100	100	100	100	100	100
20	UK	UTTARAKHAND	UK	13	106	100	100	100	100	100	100	100	100	100
20	WB	WEST BENGAL	WB	13	106	100	100	100	100	100	100	100	100	100

This screen shot show the in the villages how many schools are there and they all are greater than 60%. In this screen shot it will clearly show that how many states are there in India they all are having greater than 60% of schools in the villages and those states are Andaman & Nicobar Islands, Andhra Pradesh, Arunachal Pradesh, Assam, Bihar, Chandigarh, Chhattisgarh, Dadra & Nagar Haveli, Delhi, Goa, Gujarat, Haryana, Himachal Pradesh, Jammu And Kashmir, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Manipur, Meghalaya, Mizoram, Nagaland, Odisha, Puducherry, Punjab, Rajasthan, Sikkim, Tamil Nadu, Tripura, Uttar Pradesh, Uttarakhand, West Bengal.

4. Comparing Female Literacy and Overall Literacy with Query



This Screen shot show the where the female literacy is highest or greater than 60 % compared to Overall literacy. By giving the Query to a system or tool we got an output from the system accordingly what we query the system. In this output the user want to know the where the female literacy is greater than 60%. According to the query the user got the output is Andaman & Nicobar Islands, Andhra Pradesh, Arunachal Pradesh, Assam, Bihar, Chandigarh, Chhattisgarh, Dadra & Nagar Haveli, Daman & Diu, Delhi, Goa, Gujarat, Haryana, Himachal Pradesh, Jammu And Kashmir, Jharkhand, Karnataka, Kerala, Lakshadweep, Madhya Pradesh, Maharashtra, Manipur, Meghalaya, Mizoram, Nagaland, Odisha, Puducherry, Punjab, Rajasthan, Sikkim, Tamil Nadu, Tripura, Uttar Pradesh, Uttarakhand, West Bengal.

VI. CONCLUSION

Every year various Department of Government of India and the State Government invest significant amount of public money in to gathering and collecting data about public health and public education. Such data is available as public records, however most of the Government Schemes still does not utilized the findings from this data, even though the record is made available as public due to immense number of parameters and dependency of this data on various other departmental data and observation, it is not easy to find out the acute observations from these datasets.

In this world, we have proposed a noble data mining technique based on associate rule mining to find out the index or indices effecting the girl's child education in India. Our technique takes the advantage of public data available by both state departments as well as human resource department and we run a data filtering and associate rule mining and this associate rule mining provides a relationship between different parameters and it gives us an easy way of understanding different regions effecting girl's education.

Further our system comes with sophisticated visualization technique, by means of which visualization as statistical technique, by means of which not only the findings are depicted but not only the findings are given as the text message but at the same time they depicted graphically.

VII. FUTURE SCOPES

This system can further be improved by interlinking the data coming from various other departments like health department and finding out the relationship of girls education on other aspects like, public health, the medical facility availability and the employability of the family member and so & so. Such a system can easily be developed by incorporating the data domains of the other department and then interpolating that data with the proposed system.

REFERENCES

- [1] Geeta Nair, "Gendered Impact of Globalization of Higher Education: Promoting Human Development in India", 2015.
- [2] Saswati Das, "Inequality in Educational Opportunity in India: Evidence and Consequence of Social Exclusion", March 2016.
- [3] Rajarshi Majumder, Jhilmam Ray, "Education: Capacity Building for Human Development", 30 November 2015.
- [4] Gupta Naresh Kumar, "Child Labour in India: A Brief Study of Low and its Implementation", 2 December 2015.

[5] Marian Miller, Stephan Litschig, “Adapting the Supply of Education to the Needs of Girls: Evidence from a Policy Experiment in Rural India”, 30 November 2015.

[6] Smt.S.J.Sandhya, “Impact of Education in Women Empowerment of in Bagalkot District, Karnataka”, May 2015.

[7] John Jungpa Park, Luciana Brondi, “Why are girls still dying unnecessarily? The need to address gender inequity in child health in the post-2015 development agenda”, April 30 2015.

[8] K Vikram, A.K Shurma, “Beneficiary level factors influencing *Janani Suraksha Yojana* utilization in urban slum population of trans-Yamuna area of Delhi”, September 2013.