I J C T A, 9(32), 2016, pp. 243-250 ISSN: 0974-5572

A Comparative Analysis of Applications of Census Data Using Data Mining Techniques in E-Governance

B. Tulasi Kumari*, and D. Rajeswara Rao**

ABSTRACT

E-governance is the major unified software framework in the major portion of present government to accommodate services thorough Information and Communications Technologies (ICT) to transform the efficiency, responsibility of informational & transactional exchanges with in government to allow citizens through access & use of information. Governments deal with large amount of census data across states in the country. To make ensure that comparative analysis of census data is input to an effectual use in facilitating policy-making on finding required child labor statistics across cities and finding pattern of similarly data to construct meaningful needy census data starting from 1993 to 2009/10 years of population. It has shown the several types of problems regarding child labor trafficking across countries and having more illiteracy to be analyzed on these four years data to be addressed by policy-makers like MLAs and MLCs in constituency of the country. Existing data mining algorithms are showcase the solution bundle to determine clusters that fit some static models for e-governance software. In the paper deals with proportional analysis and use of Data mining in every the dimension of e-governance like literacy rate across states, child labor and finding meaning information about population with respective to different applications of census data. In the paper authors are deals some methodologies like E-pragathi is a e-governance software used between government and people to find out keep track of status of people who posed problems to the government and obtain results more accuracy .In this context two keys of data mining tools i.e. OLAP (Online Analytical Processing), Data mining used for finding meaning pattern. Data mining is a suitable data analytics for extracting good information from raw facts. OLAP tool is contemporary based tool for usage where all types of data to be scrutinized in solving these frequent patterns of old data and produce appropriate data to end user to deploy on target location. In the paper, the researcher throws a light on justification of improving literacy rate for next couple of year and decrease the percentage of child labor by applying rigorous Data mining technique, automatic discovery of blueprint & other interesting ways are find out close consensus of data.

Keywords: E-governance, literacy rate, Data Analytics, Census, OLAP, Data warehousing, Data Mining Techniques.

1. INTRODUCTION

E-governance involves the application of Information and Communications Technologies(ICT) by Indian government agencies for information and service delivery to people, business and government employees. In the paper discusses most wanted data mining tools and their usage [1] were used in large data to eliminate duplication on large data sets in recent days. They have started predicting required information and services online from state to another state in country and corporate organizations to enhance their literacy rates to the individual in the state of art .The concept of e-governance come into life in the beginning of 21st century with a focus on vision 2020 in Indian government where all types of applications like census 2011 details collection, economic monitoring, planning and the inclusion of Information Technology. In the paper authors put several efforts made by National Informatics Center (NIC) in the last two decades. Authors works on grinding the knowledge of all sections of people through census analytics and produce the fruitful results

* Research Scholar, Dept. of CSE, K L University, *Email: Guntur,tulasi2k@yahoo.com*

** Proffesor, Dept. of CSE, K L University, Guntur, *Email: rajeshduvvada@kluniversity.in*

from child population and literacy rate data analytics. There has been participation of non gazette office's organization and non benefitted forms works towards development of unified framework of e-governance for development to mitigate the issues of non e-governance laws and technologies in developing countries like India. For good governance, the possibility of paradigm shift gradually required from manual implementation to software framework through analysis to procure operations that are reliable to good administration and quality service, but this shift can be conceived as a meaningful capital with potential for returns. Authors discusses Clustering techniques using data mining is a process of predicting patterns from raw data that groups a set of data such that the intra-cluster similarity is maximized and the inter-cluster similarity is minimized in comparing literacy and illiteracy in rural and urban regions in India.

E-governance is to enhance the process of service delivery in population growth and maintenance in information dissemination to citizens (end user) using electronic means providing advantages over the conventional tools are:

- Increased information dissemination in various Governmental processes.
- Transparency and anticorruption in all banking transactions.
- Empowerment of citizens of India and encouragement of their participation in governance though proper analysis of system.

The main aim of E-Governance is to change the parameter of public administration through man power into e-organization by electronic data transfer. An e-organization needs to focus on the following things to develop automation product for customer orientation during population growth. Those are:

1. To Develop automation product for customer orientation during population growth.
2. To Maintain customer relationships through automation.
3. To Impetus the software consultants in transparency way.
4. To Streamline Literacy rates among consecutive years.
5. To Make better decisions.
6. To Coordinate activities better

2. IMPORTANCE FOR E-GOVERNANCE

2.1. Prerequisites of E-Governance on Census Analytics

- Apply good analytics on good administration to set new rules to stake holders to enumerate census details.
- For good public administration by writing good software to correlate details of parts of census data through proper analytics like Google analytics to adopt unified framework according source of good census details.
- De-layering or re-layering of decision-making of levels.
- Security and privacy are the two major concerns.

2.2. Factors necessary for successful e-governance

In the normal progress of population of country there is tremendous gap[2] in literacy and illiteracy rates over last couple of years. Authors are address the literature gap need to understand to improve the better literacy over illiteracy and make it India to be digital India in coming vision 2050. The following ways are used for successful e-governance rate among countries in the world.

- Enumerators commitment
- Effective administrative leadership
- Effective handling of HR issues
- Involvement of staff at design stage
- Innovative funding strategy and revenue model
- Appropriate administrative structure
- Common infrastructure and database creation
- Training & Motivation to enumerators

3. IMPLEMENTATION ISSUES OF LITERACY RATE AND CHILD LABOR IN E-GOVERNANCE

The present government of India put forth efforts on investing 100 crores in Information and Communication Technology (ICT). The aim of these investment is to improve the efficiency of government through automation, to make decision and their results reach to people those who suffers lack of education in government schools, although their parents thinking to send their children to work place to earn money. There are some lots of technical issues which need to be discussed apart from above mentioned issue. The above mentioned issue can be resolved by the government but as far as technical issues are concerned they need more focus to obtain results about this issue.

Some of technical issues related to e-governance are:

- Technical infrastructure while enumerators loading details of people related to poor economic conditions.
- Collection of large amount of census data.
- Analytics on census so that accurate information can be made.
- Online support to all departments of Government organization
- Retrieval of meaningful Data.
- Presentation of meaningful data, so fast decision can be made

E-governance results in solving to avoid duplicate patterns while collecting population details in electronic-governance has evolved [3] as an information centric model of governance that access to getting

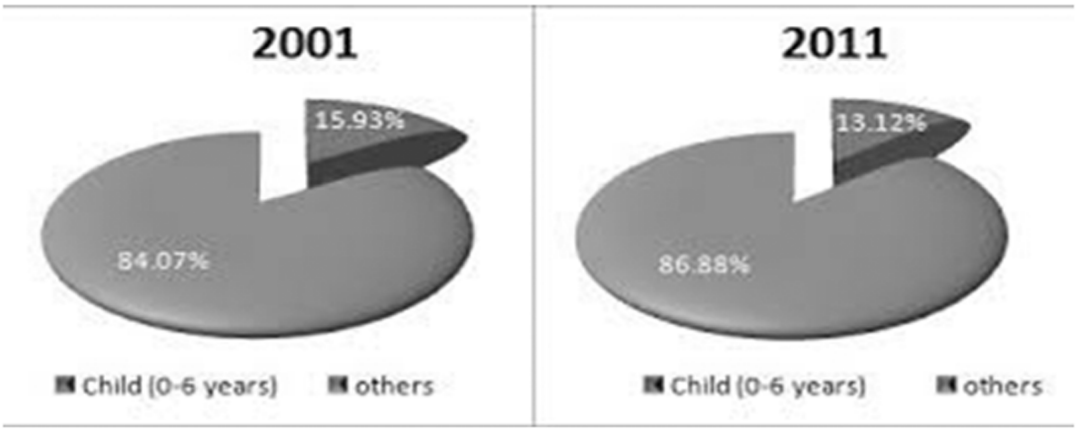

Figure 1: Child labor analysis over age in last decade(source from 2001-2011)

knowledge from process of literacy rates and child labor living across 23 states and made structuring on data set for maintaining the potentialities of ICT at various levels of public sector. In fig.1 explains the details of comparison of impact of child labors between 2001 and 2011 in the age group of 0 to 6 years to find out dropout of literacy from 15.93 to 13.12 percentages. In this context data mining help to find out what parameter to point out the literacy rate.

In fig 2 the census of data analyze the patterns of how much people living around the unit lived in sq.m. in this context how data mining help to guide how much child going to labor when economic conditions are worst based on average income of poor people in the density during 2004 to 2014. In the fig emphasize the details of how much age of child enter into work for money towards survival of their family when parents' income is below 40000 per year.

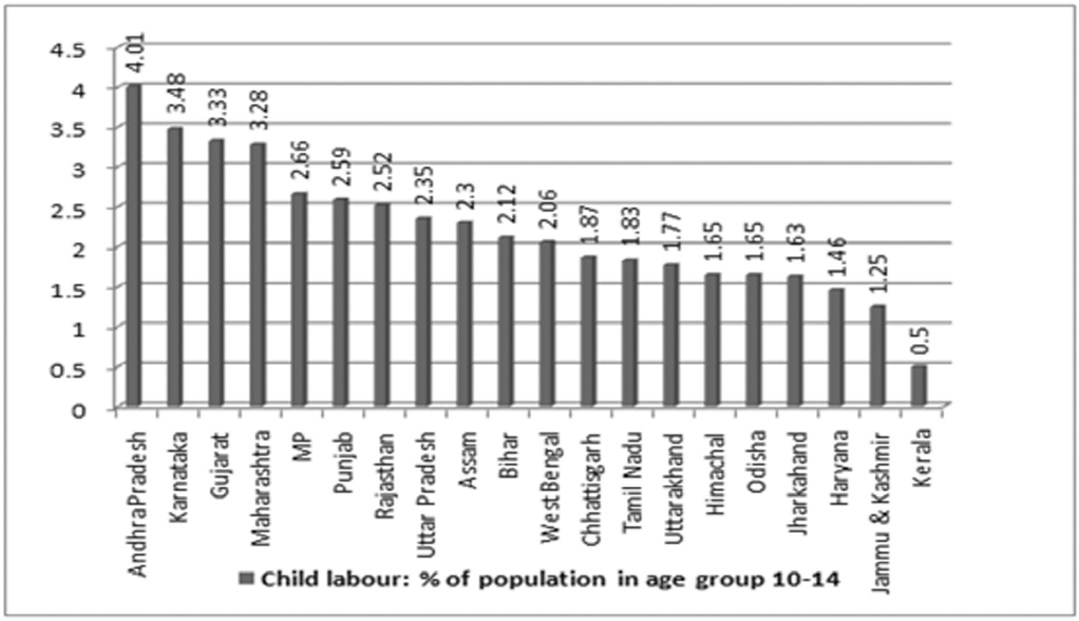

Figure 2: Census on Child labor over last decade

E-governance is the ultimate unified framework to utilize unified frameworks technologies to enhance governmental relationships in order to encourage the fair & efficient delivery of services to poor illiteracy through e-governance .The ICT model uses the new technologies to maintain the data in government organization .Some of these technologies are discussed in this paper which is very important now a days.

The central idea is on interactive analytical model for census implemented by the integration of filtered information about literacy information through pre processing of given model is fairly designed.

Two key parameters are for analytical processing of model:

1) OLAP
2) Data Mining

4. COMPARATIVE ANALYSIS OF LITERACY RATES AND CHILD LABOR IN E-GOVERNANCE

4.1. Need of data mining on child labour and Literacy census

Governments deal with enormous amount of data around 132 crores according source of 2015 census information in the country. In order to find out complete details of ration of literacy and illiteracy information is put to an effective use in facilitating decision-making on census data, a suitable data mining is constructed

over the census data. It permits several types of services through analysis on census to be addressed by policy-makers.

According to the Census 2001 the contribution of children in the 5-14 years age group in the total work force in our country is 3.15 percent totaling 12.6 million children. While there has been a decline within the numbers of children as main child workers from 9 million to 5.7 million there has been a considerable increase in the numbers of children engaged as marginal workers from 2.2 million to 6.8 million.

India is agricultural based country, so naturally more people including child labour are having chance to work in agricultural based sectors. We found max no of child labour in agriculture based sector rather than other fields. The table1 shows maximum number of child labour is in agriculture sector.

4.2. Benefit of a Data Mining with Data Warehousing for e-governance

Overcome poor literacy rate is the key objective of any Government body. For overcome the poor literacy rate in the country, it is the tradeoff between increase the literacy rate and reducing the child labor when we find out meaning patterns thoroughly about grinding section of people. Hence this unified framework can build for e-Governance can typically have data related to person and inhabitant area. The software framework can also be used for all sections of the people to streamline government schemes to poor people exercised in the eleventh five year planning commission. This analytics provide roadmap in making unified software model that are centered and the benefit of the government policies can reach the intended group. The various types and number of questions that can be handled by the data mining are limited only by the

Table 1
The details of the child labour in different sectors in the year 2009-10.

State	Agricultural sector	Mining and Quarry	Mfg.	Const.	Trade and Hotel	Transport	Others
A.P.	68.96	2.1	10.5	4.5	8.1	1.1	4.7
Assam	65.4	1.9	11.1	2.3	8.1	0.05	10.9
Bihar	67.32	0	10.45	2.1	14.9	0.04	5.6
Chhattisgarh	88.3	1.1	3.2	0.5	5.7	0	1.2
Delhi	0	0	9.3	0	54.3	0	36.4
Goa	0	2.1	0	6.4	0	64.5	26.9
Gujrat	76.69	1.04	2.58	0.28	17.77	0.16	1.48
H.P.	83.42	0	0	0	8.8	1.4	6.3
Haryana	67.21	0	4.3	8.3	7.4	0	12.6
Jharkhand	62.21	0	13.45	6.4	11.32	0.8	5.7
Karnataka	79.8	1.1	11.1	3.4	4.5	0.6	0.3
Kerala	16.5	0	30.4	0	34.9	0	18.11
M.P.	79.21	0	10.1	2.5	6.3	0	1.9
Maharastra	76.4	0	4.5	3.2	6.8	0.6	8.5
Orissa	66.6	2.1	18.2	4.8	3.9	1.1	3.2
Punjab	64.6	0	13.4	2.1	8.8	1.9	9.1
Rajasthan	73.4	0	10.5	3.6	8.1	1.1	2.9
T.N.	33.2	0	50.1	6.5	7.5	0.8	1.9
U.P.	57.8	0	22.3	3.2	11.4	1.1	4.2
Uttaranchal	78.2	0	4.3	5.8	10.3	0	1.4
W.B.	30.23	0	41.23	4.5	11.2	2.1	10.5

Source: Estimated from Unit Level Records of NSSO, 2009-10

intelligence of the person using the data warehouse and the data fed to it. Attentive targeted center query quarry

4.3. Implementation of Classification technique

Classification is a data mining function it assigns items in a collective set to target categories or classes. The goal of classification is to perfectly predict the target class for each case in the records. A decision tree is a classifier uttered as a recursive partition of the instance space. The decision tree consists of nodes form a rooted tree, meaning it is a directed tree with a node called "root" that has no incoming edges. All other nodes have exactly one arriving edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves. In a decision tree, each internal joint splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. In the simplest and most frequent condition, each test considers a single attribute, such that the instance space is partitioned according to the attribute's value. In the case of numeric attributes, a condition refers to a range. Each leaf is assigned to one category representing the most appropriate target value. Alternatively, the leaf may hold a possibility vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them of the root of the tree down to a leaf, according to the outcome of the test along the path.

For doing classification, first data should be normalized, so that attributes with large values do not dominate those with lower values. Then can be apply classification using decision tree induction by identifying those in particular intervals. From this we can identify which states have number of working children in particular interval. Similarly we can study about the number of illiterate children. First we can perform decision tree induction on all the states then we can take each interval and perform decision tree induction at final level.

This can be done in two stages one with respective child workers in the years in the above table2. In second case, we can compare net enrolled students in primary as well as upper primary groups for the above years for all the states of India.

Table 2
Estimate of Trends in India's Child Labour by Rural-Urban, 1993-2009/10 (in millions)

	Age Group		
Year	*5 to 9*	*10 to 14*	*5 to 14*
Rural			
1993-94	1.13	11.03	12.16
1999-00	0.6	8.05	8.65
2004-05	0.26	7.18	7.44
2009-10	0.2	4	4.2
Urban			
1993-94	0.12	1.52	1.64
1999-00	0.07	1.32	1.39
2004-05	0.08	1.44	1.52
2009-10	0.02	0.68	0.7
Combined			
1993-94	1.27	12.59	13.86
1999-00	0.63	9.5	10.13
2004-05	0.35	8.72	9.07
2009-10	0.22	4.68	4.9

Source: Derived from Respective Unit Level Records of NSS

5. PRESENT SYSTEM IN GOOD GOVERNANCE USING DATA MINING TECHNIQUES

In present, the Chief Minister of Andhra Pradesh announced about the software e-Pragathi in the year 2016 for people who need enormous services of different regions using Information Technology (IT) enabled services to accommodate good governance [4]. It is one of the reliable assets for public administration in proper channel. The role of IT in E-governance is well defined procedure to accommodate services to different levels of people through online. A formal definition of Predicting patterns from raw data in databases is given in data mining is the two fold task of extraction of implicit previously unknown details and potentially useful information about analytics.

5.1. Extracting Required Census details From Predicting patterns from raw data in E-governance

Data mining is a concept of predicting patterns from raw facts. In recent years it has attracted great deal of interest in Information industry. From E-Pragathi usually finding required details of census from Predicting patterns from raw facts encompasses details of iterative sequence of information dissemination, data redistribution, data selection, data mining pattern recognition and knowledge presentation. Data mining on analytics is discovery driven over lots of raw facts. Without data mining it is difficult to realize the full potential of data collected within organization as data under analysis is massive, highly dimensional, distributed and uncertain.

For better governance to succeed in fig 3 describes the data mining cycle is discussed. They must have the ability to capture, store and analyze raw facts. One of tool OLAP is adopted for finding necessary facts from proper grinding of raw facts to analyze processed information. With the help of e-governance application to leverage the low level knowledge to high level when proper public administration taken place. E-governance software is most accessible to all cadres of secretariat to enhance services to poor people using good Data analytics. Data analytics can be supported[6] in supporting filtered information for categorizing actual child labors from nearby information. The replacement of contemporary data analysis through manual approach is vague process then to introduce computer assisted analysis.

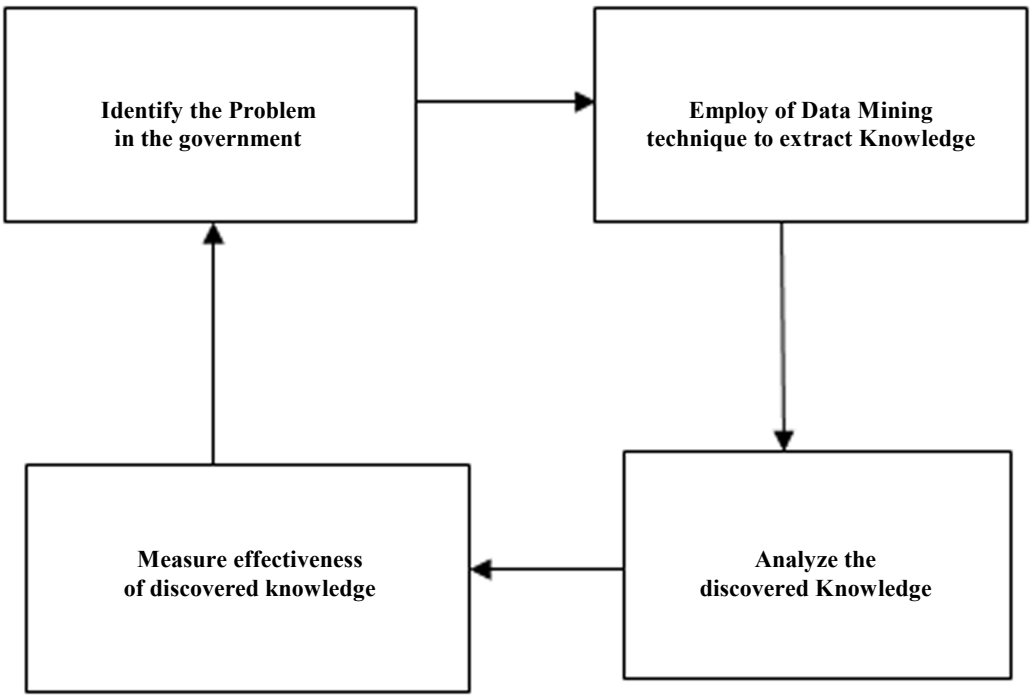

Figure 3: Real Data Mining Cycle in Census Data Calculation

5.2. Data Mining technique For E-governance

There are various data mining techniques[7] available with their suitability dependent on the domain application. Statistics provide a strong fundamental background for quantification and evaluation of results. However, algorithms based on statistics need to be modified and measured before they are applied to data mining.

6. CONCLUSION

Until now there is no proper e-Governance software is used so far to grinding actual literacy rate in population through census enumeration in the states and at the centre. In the paper data mining infrastructure has been used for find out suitable software frame work to get literacy details more appropriately. This is the time for idea of studying data mining tools and make it comparative study in the e-System like e-pragathi like software designed in Andhra Pradesh in 2016. Finally the paper issues merits and demerits of comparative study of such data mining tools have been used from general census details of 2014, then the same can be replicated in other sectors of the government. In the paper the detailed knowledge about statistical information of child labor in census. Researchers are addressing the solutions of how to overcome increasing the child labor and literacy rate.

REFERENCES

[1] Ester M., Wittmann R.: "Incremental Generalization for Mining in a Data Warehousing Environment", Proc. 6th Int. Conf. on Extending Database Technology, Valencia, Spain, 1998, in: Lecture Notes in Computer Science, Vol. 1377, Springer, 1998, pp. 135-152.

[2] Fayyad U., Piatetsky-Shapiro G., and Smyth P.: "Predicting patterns from raw data and Data Mining: Towards a Unifying Framework", Proc. 2nd Int. Conf. on Predicting patterns from raw data and Data Mining, Portland, OR, 1996, pp. 82-88.

[3] Gueting R. H.: "An Introduction to Spatial Database Systems", The VLDB Journal, Vol. 3, No. 4, October 1994, pp. 357-399.

[4] Han J., Cai Y., Cercone N.: "Data-driven Discovery of Quantitative Rules in Relational Databases", IEEE Transactions on Knowledge and Data Engineering, Vol.5, No. 1, 1993, pp. 29-40.

[5] Kaufman L., Rousseeuw P. J.: "Finding Groups in Data: An Introduction to Cluster Analysis", John Wiley & Sons, 1990.

[6] Mumick I. S., Quass D., Mumick B. S.: "Maintenance of Data Cubes and Summary Tables in a Warehouse", Proc. ACM SIGMOD Int. Conf. on Management of Data, 1997, pp. 100-111.

[7] Ng R. T., Han J.: "Efficient and Effective Clustering Methods for Data Mining", Proc. 20th Int. Conf. on Very Large Data Bases, Santiago, Chile, 1994, pp. 144-155.

[8] Review of child labour, Education and Poverty Agenda – India country Report -2006

[9] Child Labour, always right in front you, But Why? Constantly hidden from your view. Copyright © Prognosys e -Services Pvt. Ltd. 2012. All right reserved.

[10] Saharia, Priyam, "LABOR VERSUS LEARNING: EXPLAINING THE STATE-WISE VARIATION OF CHILD LABOR IN INDIA" (2014).Theses and Dissertations—Political Science. Paper 10.