ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

Analysis on Households Registered/Working through Data Mining Techniques on NREGS (National Rural Employment Guarantee Scheme) Data of Andhra Pradesh

Dr.M.Usha Rani

Abstract: ***Data mining merged the ideas from statistics, machine learning, databases and parallel computing. The data mining techniques such as characterization, discrimination, classification, clustering, outlier and trend analysis, etc. are applied on National Rural Employment Guarantee Scheme (NREGS) data of Andhra Pradesh. From 2006 to 2011 year caste-wise households registered and working from these registered households are collected for all 22 districts of Andhra Pradesh. Data mining tools are used to extract the knowledge from the databases created. Analysis on households registered/working through data mining techniques is presented in this paper.***

Keywords: **NREGS Database, Association, Clustering, Outlier Analysis, Data mining Techniques, Trend Analysis.**

I. DATA MINING

In database technologies, new research has allowed the improvement of databases. And as for computers, new powerful architectures have made possible the elaboration of huge data volumes [22]. Knowledge Discovery in Databases (KDD), also referred to as data mining, is the search for usable intelligence in large volumes of raw data. KDD is an umbrella term describing a variety of activities for making sense of data [6]. The field of data mining addresses the question of how best to use the historical data to discover general regularities and improve the process of making decisions [14]. The patterns and rules might provide guidelines for decision-making, or they might identify the issues upon which the decision-maker should focus during the choice process [19]. The finding of useful patterns in data is known by different names (including data mining) in different communities (e.g., knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing) [11]. Data mining outcomes include classification, clustering, prediction, estimation, and affinity grouping. Peacock [15][16] proposed that five foundation-level analysis tasks are the "reasons why" of data mining: summarization, predictive modeling, clustering /segmentation, classification, and link analysis. Procedures for using data mining involve learning the application domain, creating a target dataset, data cleaning and preprocessing, data reduction and projection, choosing the function of data mining, choosing the data mining algorithm(s), data mining, interpretation, and using discovered knowledge [11].

II. NATIONAL RURAL EMPLOYMENT GUARANTEE ACT (NREGA)

The policy of employment guarantee scheme (EGS) is well known since the 1817 Poor Employment Act and the 1834 Poor Law Amendment Act in Great Britain [7][8], and the New Deal programs of the 1930s in the United States [12][5], and it was considered an important element of relief policies during natural disasters and economic downturns. In recent past, this kind of schemes were followed both in developed and developing countries with multiple objectives such as poverty reduction, building and maintaining public infrastructure, improving bargaining power of the workers etc. Among others, some notable examples of such schemes mainly for poverty alleviation are as followed by Chile (1987), India (1978), Pakistan (1992), Bangladesh (1983), Philippines (1990), Botswana (1960), Kenya (1992). The EGS is a policy of direct transfer to the poor through the provision of public works [10][13][21] satisfying the property of self selection [3][4] and the magnitude of such direct transfer benefits was empirically studied [18]. The earlier studies on rural public works programs emphasized its role as income insurance in the presence of seasonality in agrarian labor market [1], for building longer term capital assets [2], preventing dislocation of families in search of jobs and food, affecting rural-urban migration [17]. Though the policies of employment guarantee schemes are commonly followed by governments in both developing and developed countries alike as a relief policy, there is very little by way of formal theoretical modeling to understand the issue. Recently Basu et al [9] provides a theoretical framework to analyze both the positive and normative implications of such a policy in a spectrum of labor market structures. While the introduction of such a scheme introduces a contestability in the labor market where government is another employer, however the outcome with respect to wages and overall level of employment in the labor market would depend on the "degree of distributional and/or efficiency concerns of the planner". Consider the scenario where there exists a labor market with demand and supply of labor and a wage rate. Even if the wage is competitively determined, it may not be adequate for the poor household to reach their target income; what they consider as means of a decent living for their households. Envisaging situations like these, the Indian government from time to time has implemented different policies to complement the income of the poor.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

The National Rural Employment Guarantee Act (NREGA) is one such policy of EGS which is introduced in recent past. Passed by the Lok Sabha on August 23, 2004 and signed by the President of India on September 5, 2005, NREGA has been hailed as a major initiative in the Government of India's commitment to providing an economic safety net to India's rural poor. The NREGA extends to all rural areas of India, including Fifth and Sixth Schedule areas, except the State of Jammu and Kashmir.

A. The Scheme: National Rural Employment Guarantee Act of India

The Parliament enacted an Act No. 42 of 2005 called the National Rural Employment Guarantee Act. The Act provides a guarantee for rural employment to households whose adult members volunteer to do un-skilled manual work not less than 100 days of such work in a financial year in accordance with the scheme made under the Act. **The scheme** has been launched on February 2nd 2006 in 200 districts of the Country. It is expected to enhance people's livelihood on sustained basis by developing economic and social infrastructure in rural areas. It is a direct attack on the causes of chronic poverty such as drought, deforestation and soil erosion. Rural Employment Guarantee Scheme is demand-driven instead of being supply-driven.

The focus of the scheme shall be on:-

1. Water conservation and water harvesting.
2. Drought proofing (including a forestation and tree plantation).
3. Irrigation canals including micro and minor irrigation works.
4. Provision of irrigation facility to land owned by households belonging to the SCs and STs or to land beneficiaries of land reforms or that of the beneficiaries under the Indira Awas Yojana.
5. Renovation of traditional water bodies including desalting of tanks.
6. Land development.
7. Flood control and protection works including drainage in water logged areas.
8. Rural connectivity to provide all-weather access.
9. Any other work which may be notified by the Central Government in consultation with the State Government.

III. ANALYSIS ON HOUSEHOLDS REGISTERED/WORKING

The number of job cards issued to households, caste-wise households are working information is collected for all the districts from the secondary data [23-32] for the years 2006 to 2011. These data is integrated in CasteWiseEmploy database. The CasteWiseEmploy schema is presented in Fig. 1, the structure of database is shown in table 1. The partial part of data base is shown in table 2 included with the trend in 2012 and 2013. This trend is calculated from the past history and is listed as 7th & 8th rows for each district.

A Classification/Clustered Analysis

Linear Growth Rate is calculated for each district on (1) households working and their wages, (2) SC households working and their wages and (3) ST households working and their wages are presented in table 3. Based on Characterization, Discrimination and classification the clusters observed are visualized. It is illustrated from the Fig.2 that, 4 clusters are identified from the households working from the registered households (in Lakhs) in the districts of Andhra Pradesh. The decision tree on which the formation of clusters with the centroids is also shown in Fig.2. It is observed from the Fig.2 that the linear growth rate is negative in W.Godavari, Anantapur, Kadapa, and Chittoor & Nellore. The wage amounts that are spending in these areas are least due to the people are not coming forward to work. It is observed from the figure that in west Godavari district Government is spending more amounts towards wages to households and attracting the people to make them to work. The maximum amount that the govt. is spending on wages to households is Krishna. Even though the Govt. not spending more amount under NREGS works, but the people registered in Srikakulam district is increasing. It is illustrated from the Fig.3 that, 4 clusters are identified from the SC households working from the registered SC households (in Lakhs) in the districts of Andhra Pradesh. The decision tree on which the formation of clusters with the cancroids is also shown in Fig.3. It is observed from the figure that clusters C1, C2 with negative centroids. The districts under these two clusters are Nellore, Chittoor, Kurnool, kadapa, Khammam, Anantapur, W.Godavari and Mahaboobnagar. Even the Govt. is coming forward to spend more amount on W.Godavari district but the west Godavari SC people are not coming forward to register under NREGS works. The Govt. should take the necessary measures to fill up the gaps in between Govt. officials and SC public. It is illustrated from the Fig.4 that, 4 clusters are identified from the ST households working from the registered ST households (in Lakhs) in the districts of Andhra Pradesh. The decision tree on which the formation of clusters with the centroids is also shown in Fig.4. **Outlier Analysis:** It is observed from Fig.4, C1 cluster with single district named Nellore became an outlier. Its linear growth rate is -27.898. ST households are not coming forward to work under NREGS. The total wage amount that is spending to the works in Krishna district is high and this district became a outlier from the cluster C3. In the C2 cluster West Godavari district became outlier based on amount spending on works is more.

B.Trend Analysis on House holds Registered are Working

Based on the past history i.e. from 2006 to 2011 years, the no.of households who registered in their respective districts may work in the future i.e. in years 2012 & 2013 – the trend is calculated and listed in 7th & 8th rows for each district. The partial sample data is showned in table 2. Expected no.of households who registered may work in the future(trend) years i.e. 2012 & 2013 are showned along

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

with 2011 data in the table 4. How the trend changes from 2011 to 2013 for all 22 districts i.e. Adilabad to West Godavari are showned in Fig. 5. It is observed from the figure that more no. of households registered may work in Srikakulam district and less at Nellore. The Fig. 5 provides guidelines to Govt. that how much to spend on what district.

IV. CONCLUSION

Data mining tool - Rapid miner is used to discover the interested patterns on the data of caste wise households that are registered and caste wise households that are working in NREGS works. Caste Wise Employee database is created from NREGS data. Rapid miner finds the clusters based on other data mining techniques like Characterization, Discrimination, and Classification on linear growth rate of households working and wages paid to them. The results are interpreted in figures. So, the outcome from this work is useful to the government to take the decision that how many are dropping. So, that the funds to that respective districts may be divert to where there is actual need.

REFERENCES

[1] Basu Arnab K., "Impact of rural employment guarantee schemes on seasonal labor markets: optimum compensation and workers' welfare," College of William and Mary, mimeo, 2007.

[2] Basu Kaushik, "Food for work programmes: beyond roads that get washed away," Economic and Political Weekly, Jan 1981, pp. 3–10.

[3] Besley Timothy, Coate Stephen, "Workfare vs. welfare: incentive arguments for work requirements in poverty alleviation programs," American Economic Review 82 (2), 1992, pp.249–261.

[4] Besley Timothy, Kanbur Ravi, "Principles of targeting." In: Lipton, Michael, van de Gaag, Jacques (Eds.), Including the Poor: Proceedings of a Symposium Organized by the World Bank and the International Food Policy Research Institute. The World Bank, Washington D.C., 1993, pp. 67–90.

[5] Bernstein Irving, "Turbulent Years: A History of the American Worker," Houghton Mifflin, Boston, 1970, pp.1933–1941.

[6] Brachman, R.J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G., Simoudis, E. Mining business databases," Communications of the ACM 39(11), Nov 1996, pp. 42-48.

[7] Blaug, Mark. "The myth of the old poor law and the making of the new." Journal of Economic History 23, 1963, pp.151–184.

[8] Blaug Mark, "The poor law report re-examined. Journal of Economic History 24," 1964, pp. 229–245.

[9] Basu Kaushik, Sanghamitra Das, Bhaskar Dutta, "Child labor and household wealth: Theory and empirical evidence of an inverted-U," Journal of Development Economics 91, 2010, pp. 8–14.

[10] Drèze Jean, Sen Amartya, "Strategies of entitlement protection." Hunger and Public Action, 1991, pp. 104–121.

[11] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. "The KDD process for extracting useful knowledge from volumes of data," Communications of the ACM 39(11), Nov 1996, pp. 27-34.

[12] Kesselman Jonathan R., "Work relief programs in the Great Depression." In: Palmer, J.L. (Ed.), Creating Jobs: Public Employment Programs and Wage Subsidies. Brookings Institution, Washington, D.C, 1978.

[13] Lipton Michael, "Success in anti-poverty." Issues in Development Discussion Paper, vol. 8. International Labour Office, Geneva, 1996.

[14] Mitchell, T.M. "Machine learning and data mining," Communications of the ACM 42(11), Nov 1999, pp.30-36.

[15] Peacock, P. R. "Data mining in marketing: Part 1," Marketing Management 6(4) winter 1998, pp. 8-18.

[16] Peacock, P. R. "Data Mining in Marketing: Part 2," Marketing Management 7(1), 1998, pp.14-25.

[17] Ravallion, Martin, 1991. "Reaching the rural poor through EGS employment: arguments, lessons, and evidence from South Asia." World Bank Research Observer 6 (1), 1991, pp.153–176.

[18] Ravallion, Martin, Datt, Guarav, Chaudhuri, S. "Does Maharashtra's employment guarantee scheme guarantee employment?" Effects of the 1988wage increase. Economic Development and Cultural Change 41 (2), 1993, pp.251–275.

[19] Sauter, V. L. "Intuitive decision-making," Communications of the ACM 42(6), 1999, pp.109-115.

[20] Usha Rani, M., Ramasree, R.J. "Superficial Overview of Datamining Tools", Discovery Publications, 2010.

[21] Von Braun Joachim, "Employment for Poverty Reduction and Food Security." International Food Policy Research Institute, Washington DC, 1995.

[22] Zanasi, A. "Competitive intelligence through data mining public sources," Competitive Intelligence Review 9(1), Jan-Mar 1998, pp. 44-54.

[23] http://nrega.nic.in.

[24] http://nrega.ap.gov.in/.

[25] http://www.nregs.org/.

[26] http://nregsodisha.org/.

[27] http://www.seri.ap.gov.in/nregs/mgnregs.asp.

[28] http://www.rd.ap.gov.in/NREGS_GOs.html.

[29] www.sasanet.org/curriculum_final/downlaods/CB/Case St...

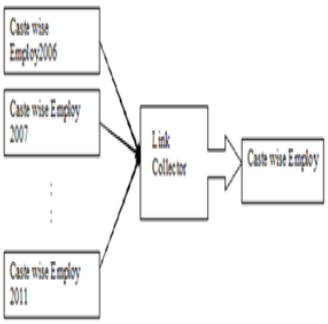
[30] http://www.sasanet.org/curriculum_final/downlaods/CB/Case St...

[31] http://www.nregaconsortium.in/downloads/NREGS_AP_WASSAN.doc.

[32] http://www.aponline.gov.in/Quick Links/Departments/Panchayat.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

Fig.1 House Holds Registered/Working Schema



Attribute Name	Data Type	Description
District Name	Nominal	District Name
Total HHs Issued Job Cards	Numerical	Total House Holds Issued Job Cards
Total HHs	Numerical	Total House
Working		Holds Working
Percent HHs Working	Numerical	Percentage of HHs Working
Total SCs Working	Numerical	Total SCs Working
Percent Of SCsEmp	Numerical	Percentage Of SCs Employed
Total STs Working	Numerical	Total STs Working
Percent Of STs Emp	Numerical	Percentage Of STs Employed
Total BCs Working	Numerical	Total BCs Working
Percent Of BCsEmp	Numerical	Percentage Of BCs Employed
Total Others Working	Numerical	Total Others Working
Percent Of OthersEmp	Numerical	Percentage Of Others Employed

Table 1: Structure of Caste Wise Employee Database

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

	District	Year	No. of Job cards issued	No of HHs working	% of HHs working	SC-No registered HHs	SC HH working	% of SC working	ST-No of regd HHs	ST HH working	% of ST working	BC-No of regd HHs	B wo
1	Adilabad	2006	469609	125208	26.66	107247	35430	33.04	132597	50740	38.27	201682	3
2	Adilabad	2007	469609	250279	53.30	107247	64062	59.73	132597	85553	64.52	201682	9
3	Adilabad	2008	469609	269103	57.30	107247	67409	62.85	132597	87777	66.20	201682	1
4	Adilabad	2009	469609	272589	58.05	107247	67460	62.90	132597	85847	64.74	201682	1
5	Adilabad	2010	469609	276155	58.81	107247	66150	61.68	132597	85012	64.11	201682	1
6	Adilabad	2011	469609	171365	36.49	107247	45420	42.35	132597	60118	45.34	201682	6
7		**2012**		**258639.7**	**55.1**		**63281.7**	**59.0**	**132597.0**	**80174.9**	**60.5**		**10**
8		**2013**		**267551.1**	**57.0**		**64889.2**	**60.5**	**132597.0**	**81413.1**	**61.4**		**11**
	Anantapur	2006	717513	226919	31.63	132753	51324	38.66	33030	11471	34.73	384236	1
	Anantapur	2007	717513	350645	48.87	132753	68290	51.44	33030	16768	50.77	384236	1
	Anantapur	2008	717513	315255	43.94	132753	63641	47.94	33030	16611	50.29	384236	1
	Anantapur	2009	717513	343079	47.82	132753	70595	53.18	33030	17647	53.43	384236	1
	Anantapur	2010	717513	294758	41.08	132753	60403	45.50	33030	14898	45.10	384236	1
	Anantapur	2011	717513	206957	28.84	132753	47279	35.61	33030	10943	33.13	384236	1
		2012		**265637.5**	**37.0**		**56562.1**	**42.6**		**14001.6**	**42.4**		**15**
		2013		**258790.4**	**36.1**		**55506.9**	**41.8**		**13795.5**	**41.8**		**15**
	Chittoor	2006	607971	231952	38.15	168201	79289	47.14	34210	12683	37.07	229304	8
	Chittoor	2007	607971	271248	44.62	168201	85966	51.11	34210	14815	43.31	229304	1
	Chittoor	2008	607971	236808	38.95	168201	77311	45.96	34210	12381	36.19	229304	8
	Chittoor	2009	607971	252029	41.45	168201	82469	49.03	34210	12874	37.63	229304	9
	Chittoor	2010	607971	202400	33.29	168201	70598	41.97	34210	9654	28.22	229304	8
	Chittoor	2011	607971	146767	24.14	168201	55021	32.71	34210	6408	18.73	229304	5
		2012		**161809.2**	**26.6**		**58880.4**	**35.0**		**6832.7**	**20.0**		**65**
		2013		**144173.5**	**23.7**		**54243.7**	**32.2**		**5508.0**	**16.1**		**59**

Table 2: Partial Part of House Holds Registered/Working Data (District-wise)

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

District	LGR-No of HHs working	LGR-HHs total wage (in Lakhs)	LGR-SC HH working	LGR-SC-Total wage (Rs in Lakhs)	LGR-ST HH working	LGR-ST-Total wage (Rs in Lakhs)
Adilabad	3.918	10.6337	2.7883	9.7458	1.6326	8.8731
Anantapur	-2.3643	8.41	-1.7512	12.1939	-1.3999	10.5772
Chittoor	-7.8895	0.1464	-6.1734	2.4077	-11.5502	-5.5361
East Godavari	10.0733	16.8018	14.1052	22.7479	13.9467	21.6115
Guntur	8.4485	24.4994	12.1730	31.4572	12.6140	37.2723
Kadapa	-2.5539	7.8791	-3.4871	8.4371	-10.0555	2.8185
Karimnagar	5.3444	10.2302	2.3241	7.1460	5.7885	11.6447
Khammam	-0.37	7.1539	-2.5453	1.5272	2.0071	12.9036
Krishna	13.4325	54.2569	12.6466	53.1770	18.2130	70.9021
Kurnool	-5.1208	-0.8841	-4.2775	3.2745	-6.4399	9.8960
Mahabubnagar	2.2144	13.8025	0.3131	12.0063	3.6902	18.6673
Medak	6.7632	12.202	3.9815	9.4555	9.5425	16.7657
Nalgonda	4.5552	13.2706	1.9890	10.4960	7.4169	20.9431
Nizamabad	5.3661	15.4296	2.7450	11.7006	7.4828	20.7696
Prakasam	16.4483	28.8927	14.9578	30.9186	4.2775	20.5109
Ranga Reddy	7.7334	16.7207	5.8963	14.2933	11.1293	24.9798
S.P.S Nellore	-11.2076	7.1996	-10.8760	9.6450	-27.8979	-6.6057
Srikakulam	21.4402	31.9127	21.2047	34.0586	2.9066	33.1980
Visakhapatnam	3.5804	35.4617	0.3936	30.4047	14.6483	53.0995
Vizianagaram	10.3195	25.7975	6.2978	25.6994	7.5123	35.5409
Warangal	5.9518	17.283	2.1172	13.5914	7.6058	20.7804
West Godavari	-2.2007	28.8358	-0.1121	30.7123	-3.7481	33.3393

Table 3: Linear Growth Rate – House Holds Registered/Working

District	House Holds Registered working in Year 2011	Expected HH Registered may work in Year 2012	Expected HH Registered may work in Year 2013
Adilabad	171365	258639.7	267551.1
Anantapur	206957	265637.5	258790.4
Chittoor	146767	161809.2	144173.5
East Godavari	257212	384102.9	412709.1
Guntur	169910	175881.5	186105
Kadapa	158906	189852	184527.4
Karimnagar	212747	335161.1	350250.8
Khammam	177191	256195	255234.7
Krishna	169456	198165.5	212665.1
Kurnool	222647	243387.5	231476.6
Mahabubnagar	236080	322877.1	329512.5
Medak	164260	244463.5	257832.4
Nalgonda	300662	402865.1	418692.9
Nizamabad	145109	221521.7	231529.2
Prakasam	234662	341967.4	375182.4
Ranga Reddy	100720	127549.8	135312.6
S.P.S Nellore	160693	143345.5	124746.2
Srikakulam	269564	402679.3	452002.3
Visakhapatnam	241298	283073.5	292376
Vizianagaram	279166	349893.7	376420.1
Warangal	266698	392141.7	411457.6
West Godavari	154989	180411	177663.4

Table 4: Trend Analysis for House Holds Register

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

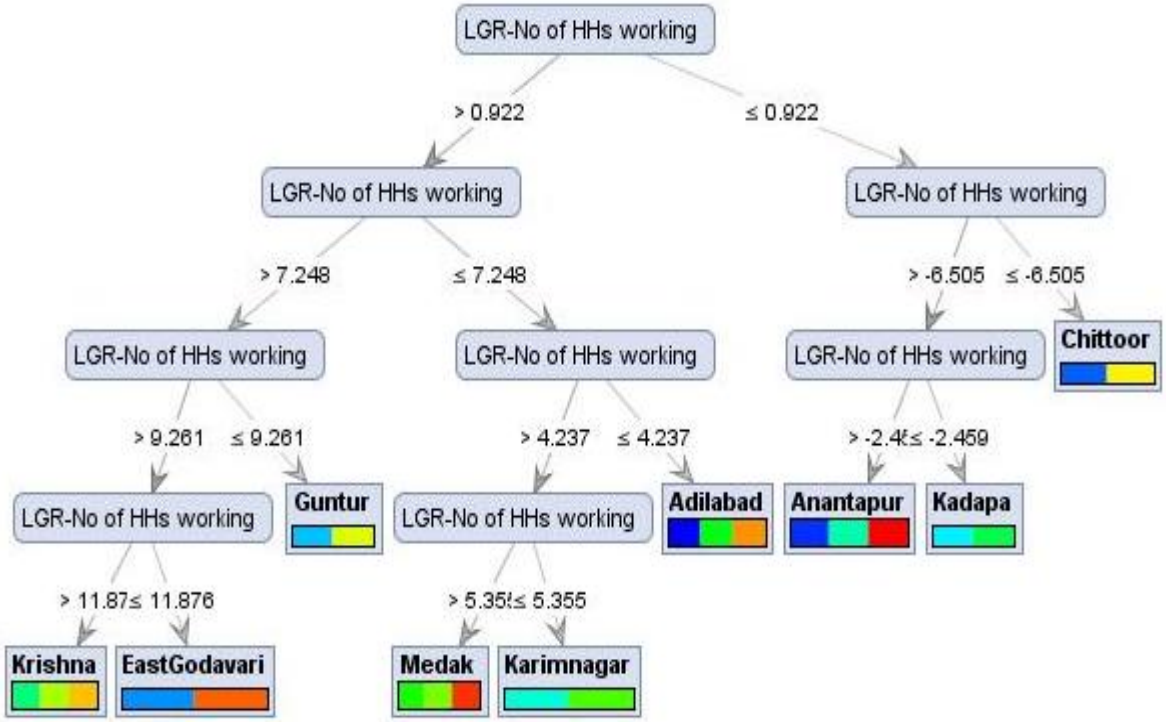
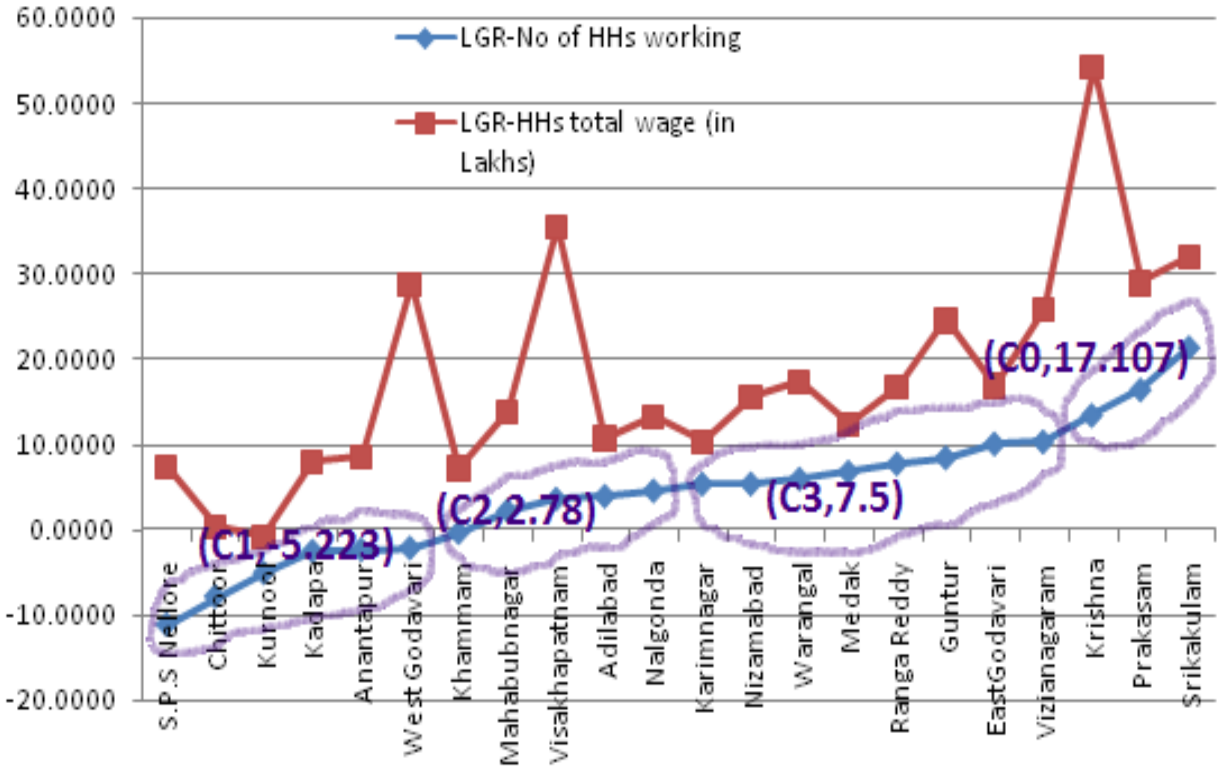


Fig.2: House Holds Working from the Registered House Holds



ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

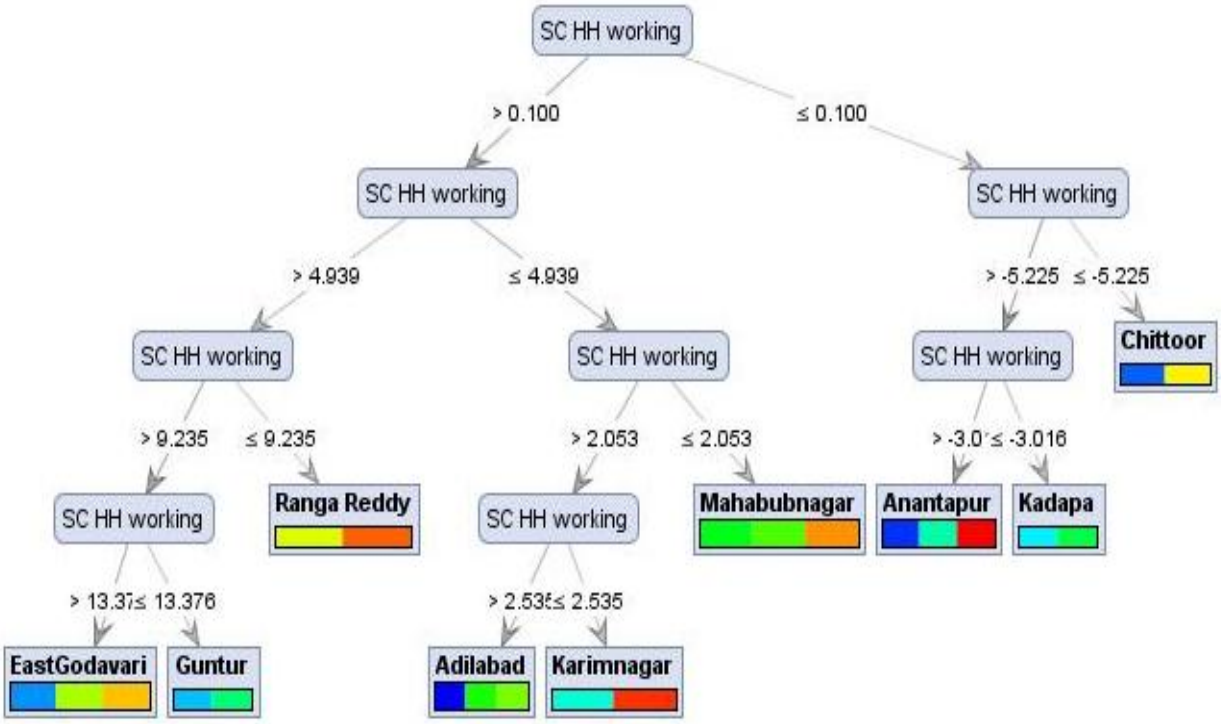
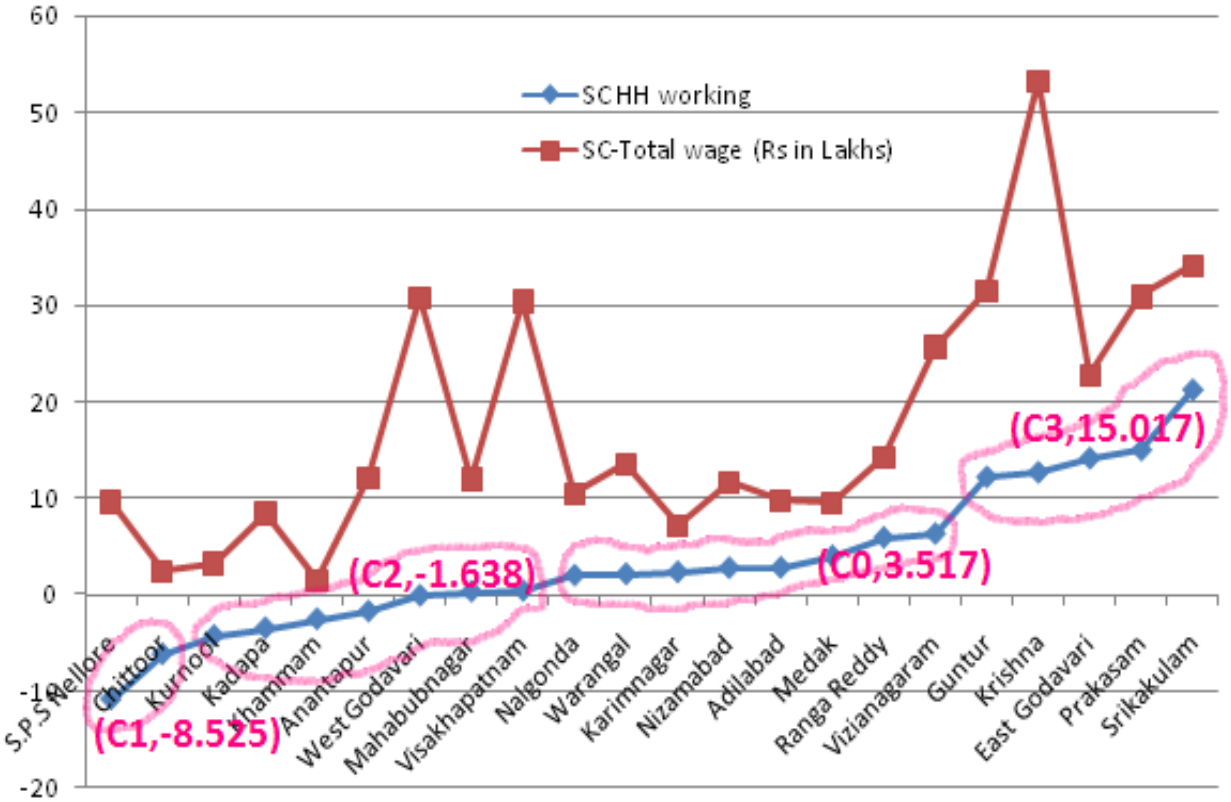


Fig.3: SC House Holds working from the Total Registered House Holds



ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

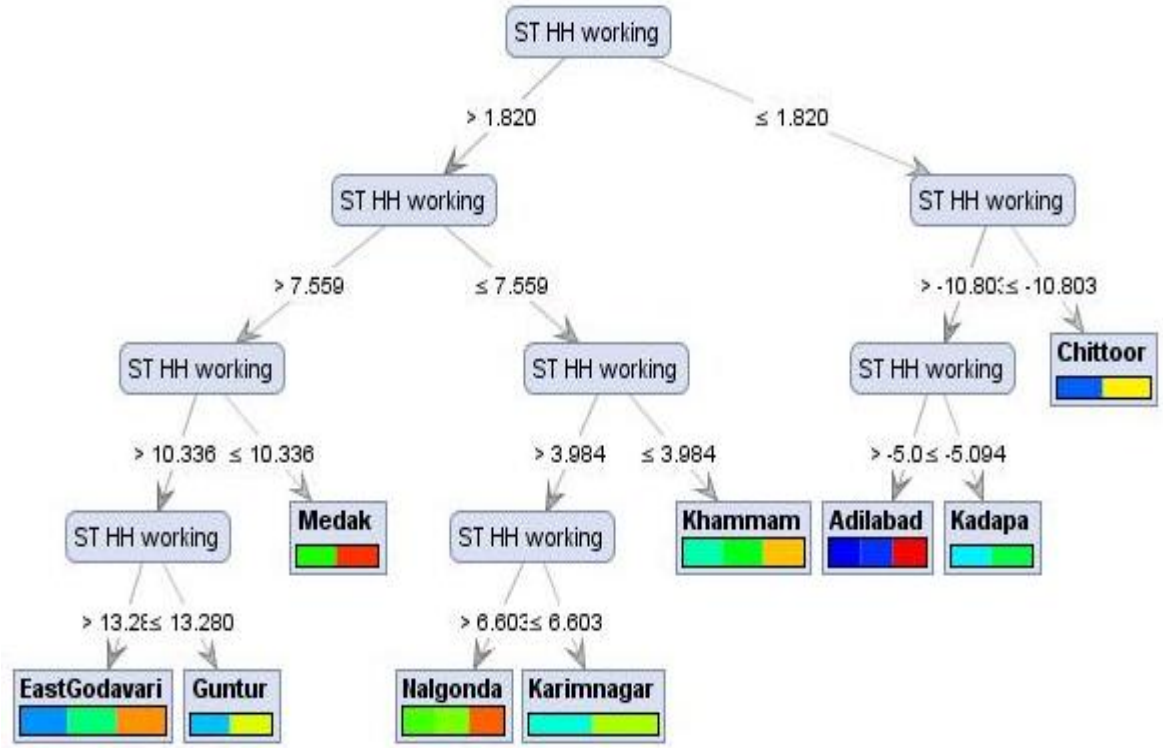
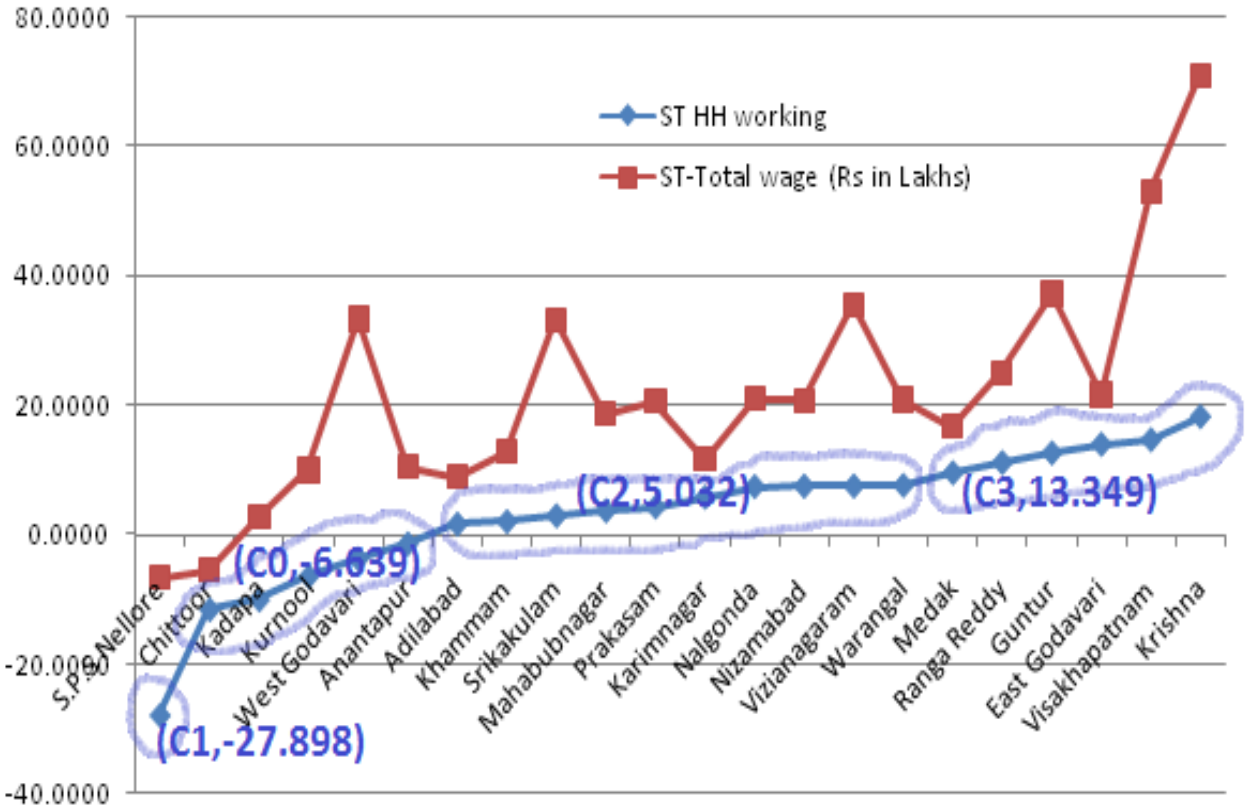


Fig.4: ST House Holds working from the Total Registered House Holds



ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 2, August 2012

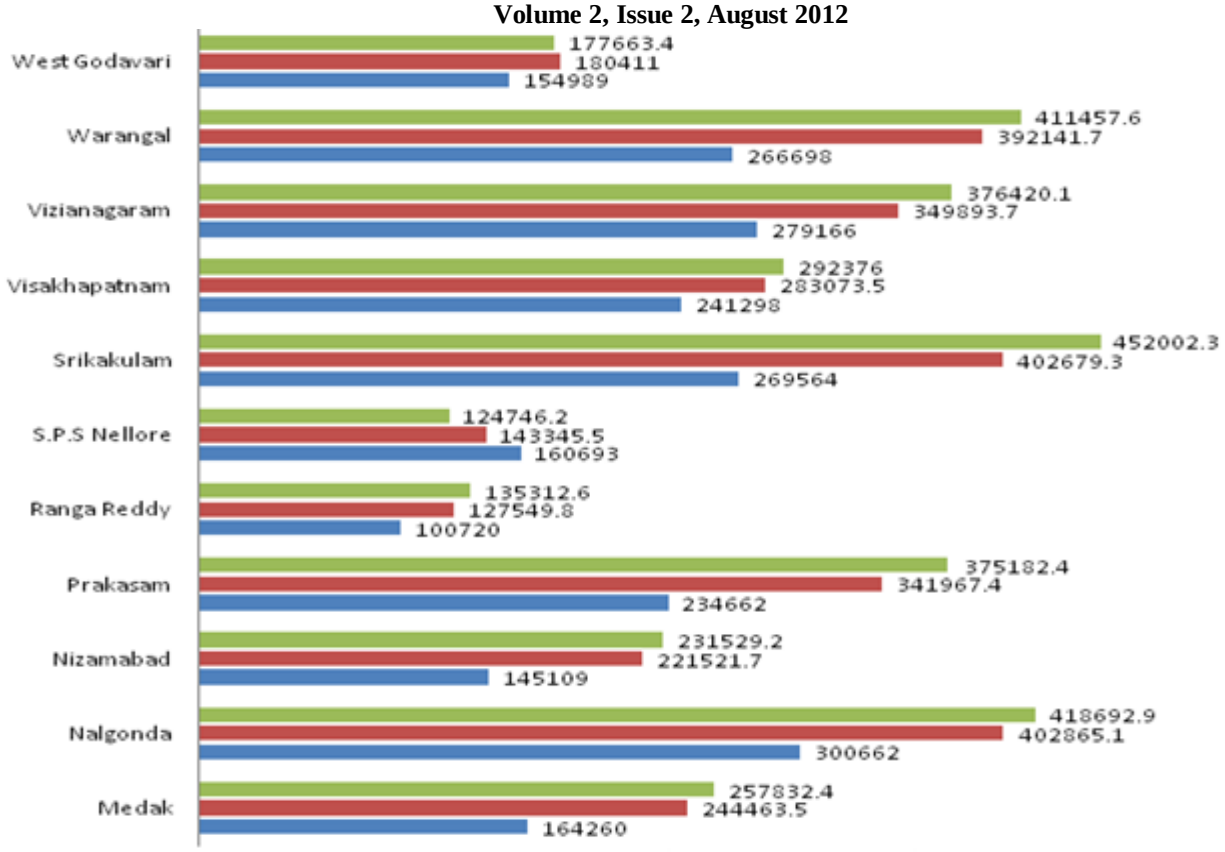


Fig.5 : Trend Analysis on House Holds Register