Expenditure Analysis Through Data Mining Techniques on NREGS(National Rural Employment Guarantee Scheme) Data of Andhra Pradesh

Dr.M.Usha Rani, Assoc. Professor, Dept. of Computer Science, SPMVV, Tirupati.
musha_rohan@yahoo.com

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. Expenditure on different districts for unskilled labor, material and contingency are collected separately from January 2010 to December 2011(month-wise data). The hidden patterns, and information find from all these techniques about the expenditure on different districts for unskilled labor, material and contingency are presented in this paper.**

Keywords: Expenditure Analysis, NREGS Data, Association, Clustering, Outlier Analysis

1. 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].

2. NATIONAL RURAL EMPLOYMENT GUARANTEE ACT(NREGA)

2.1 Introduction

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 policy of employment guarantee scheme 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 analyse 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.

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.

2.2 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. Is expected to enhance people's livelihood on sustained basis by developing economic and social infrastructure in rural areas? 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.

3. EXPENDITURE ANALYSIS

As a part of NREGS scheme, expenditure spend by government on different districts for unskilled labor, material and contingency are collected separately from January 2010 to December 2011. The schema for expenditure analysis is shown in fig.1 and the high dimensional "ExpendAnalysis" database is created with the attributes listed in table 1 for the month-wise secondary data collected from different sources[23-32]. From the "ExpendAnalysis" the cumulative expenditure of Unskilled, cumulative expenditure of Material and cumulative expenditure of Contingency are identified for each district. The Linear Growth Rate for these three are calculated and listed in table 2.

It is illustrated from the fig.2 that, 4 clusters [20] are identified from the cumulative expenditure on unskilled wage (in Lakhs). The decision tree on which the formation of clusters with the centroids are also shown in fig.2. It is observed from the fig.2 that the expenditure on unskilled-labor is raising in Visakhapatnam, Krishna and Guntur districts of Andhra Pradesh. Whereas least expenditure is identified at Nizamabad district. So, the Govt. should take the correct decision by observing the difference.

Fig.1: Expenditure Analysis Schema
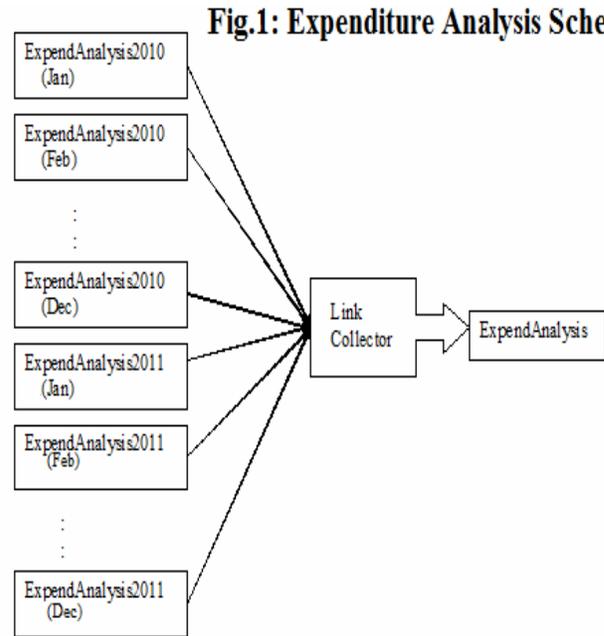
Attribute Name	Data Type	Description
DistrictName	Nominal	District Name
CumulExpendOnUnskilled	Numerical	Cumulative expenditure on unskilled wage
CumulExpendOnMatrl	Numerical	Cumulative Expenditure On Material
CumulExpendOnCntngncy	Numerical	Cumulative Expenditure On Contingency
ContingencyPercent	Numerical	Contingency Percentage

Table 1: Structure of ExpendAnalysis Database

District	LGR-Cumulative Expenditure on Unskilled Wage (in Lakhs)	LGR-Cumulative Expenditure on Material (in Lakhs)	LGR-Cumulative Expenditure on Contingency (in Lakhs)
Adilabad	-5.746	-1.812	-1.667
Anantapur	-2.963	-3.843	-0.366
Chittoor	-2.426	-1.874	0.528
East Godavari	-0.315	-2.107	0.487
Guntur	2.529	-2.465	0.019
Kadapa	-2.900	-3.663	0.128
Karimnagar	-6.188	-4.248	-3.058
Khammam	-4.161	-0.428	-1.070
Krishna	0.942	-6.966	-0.808
Kurnool	-2.062	0.808	-2.520
Mahabubnagar	-4.106	-4.614	-0.867
Medak	-5.455	-5.564	-0.419
Nalgonda	-4.598	-5.216	0.059
Nizamabad	-7.407	-6.776	-2.765
Prakasam	-2.721	-5.440	-0.199
Ranga Reddy	-1.013	-3.105	-4.480
S.P.S Nellore	-0.499	-6.395	-0.130
Srikakulam	-3.206	-3.712	0.414
Visakhapatnam	0.657	-1.631	-0.897
Vizianagaram	-1.213	-2.763	-0.391
Warangal	-4.801	-5.767	-2.007
West Godavari	-2.480	-7.165	-2.660

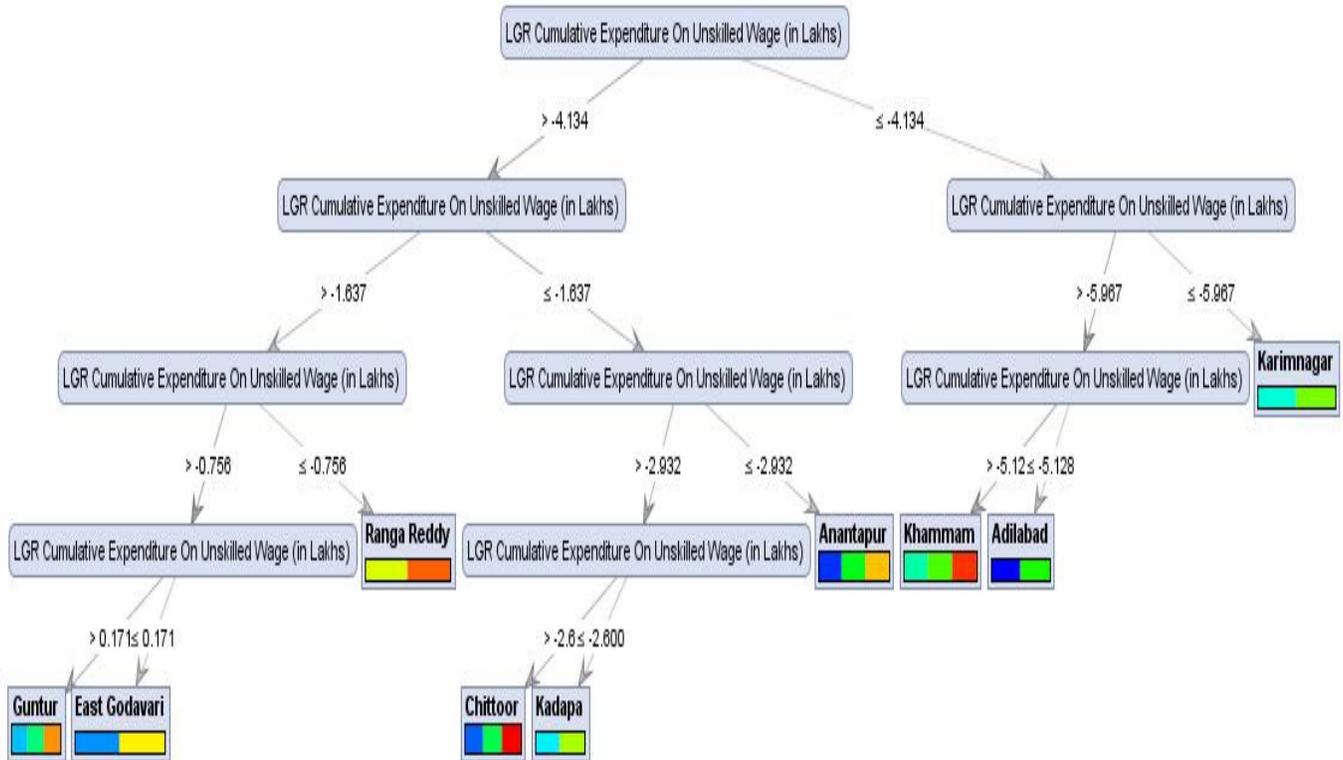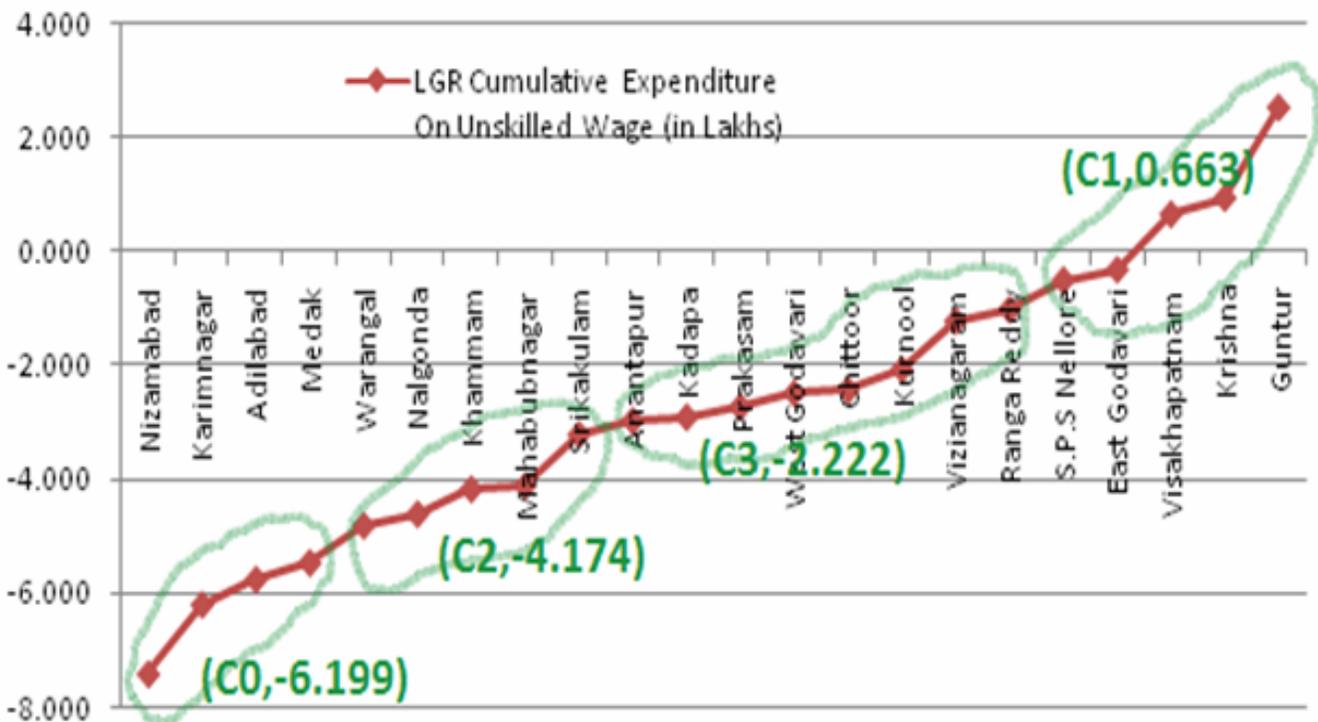Table 2: Expenditure Analysis – Linear Growth Rate

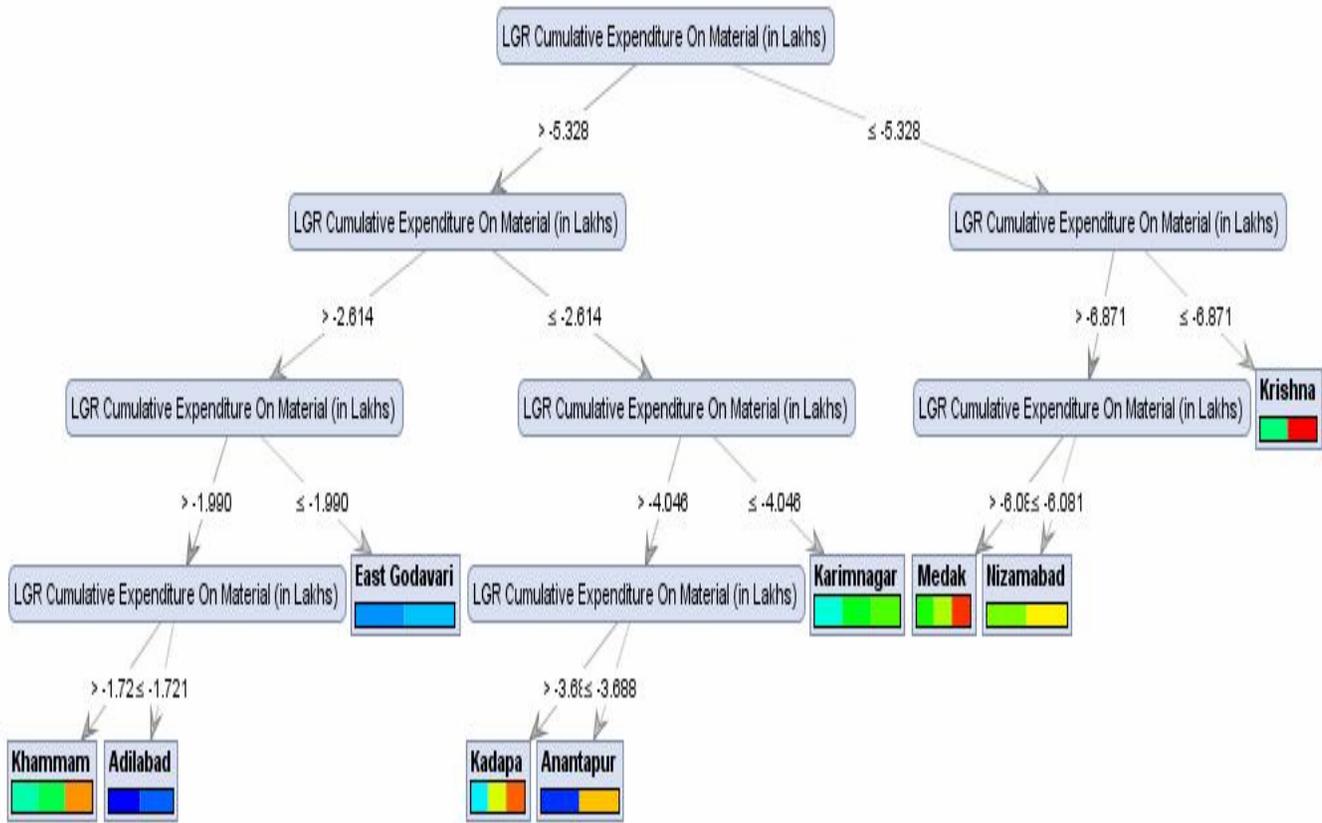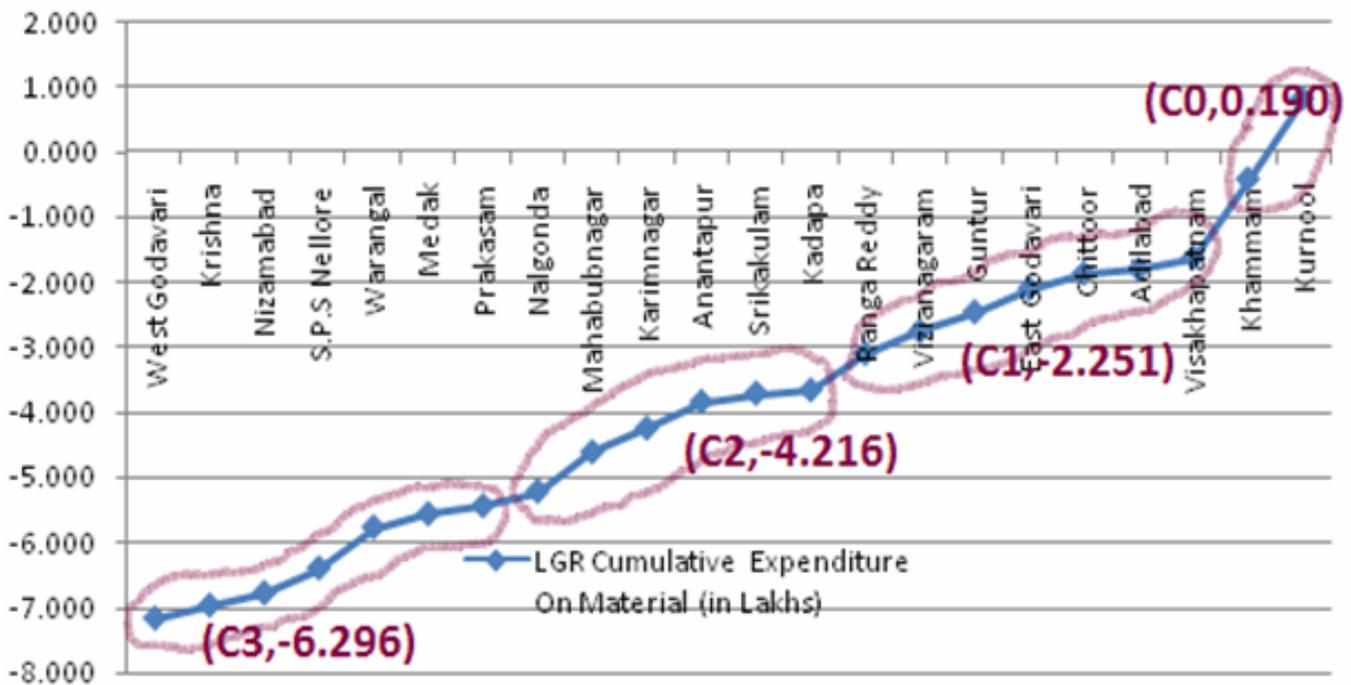Fig.2: LGR Cumulative Expenditure On Unskilled Wage (in Lakhs)

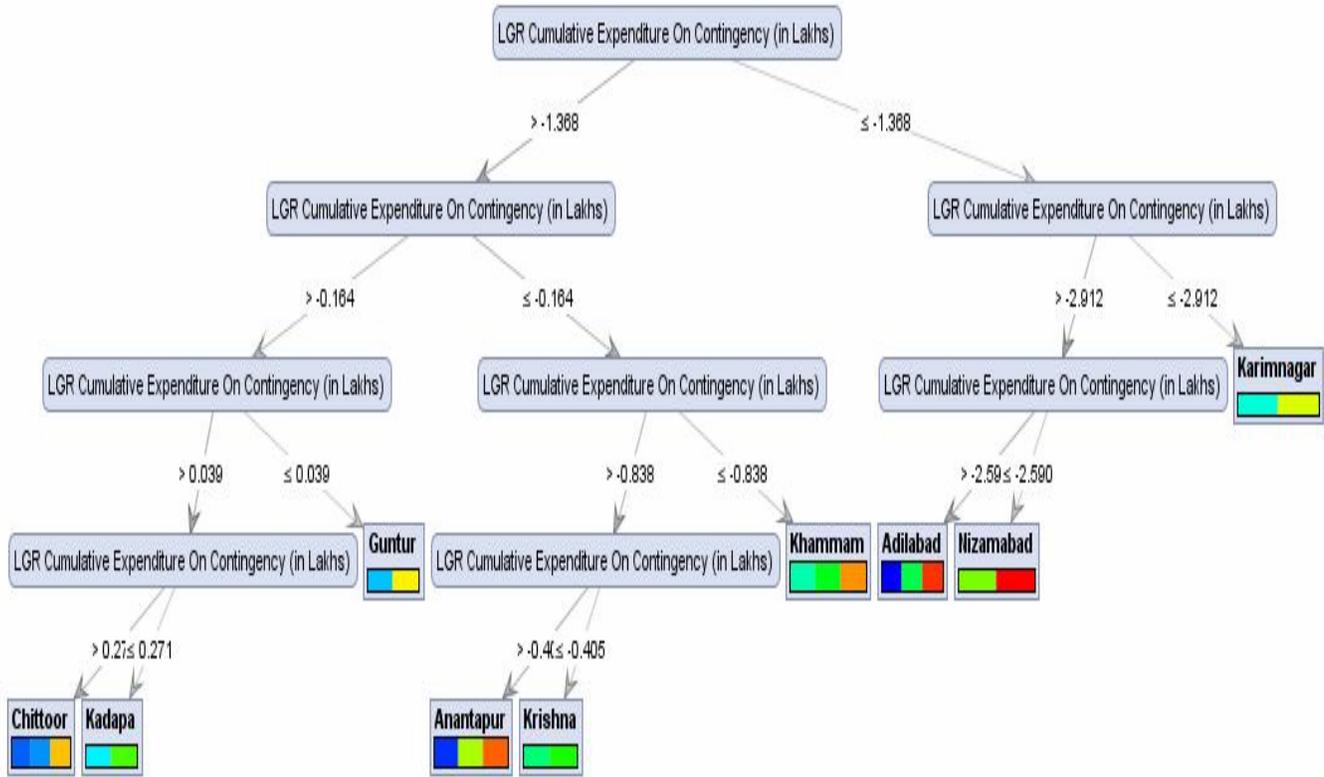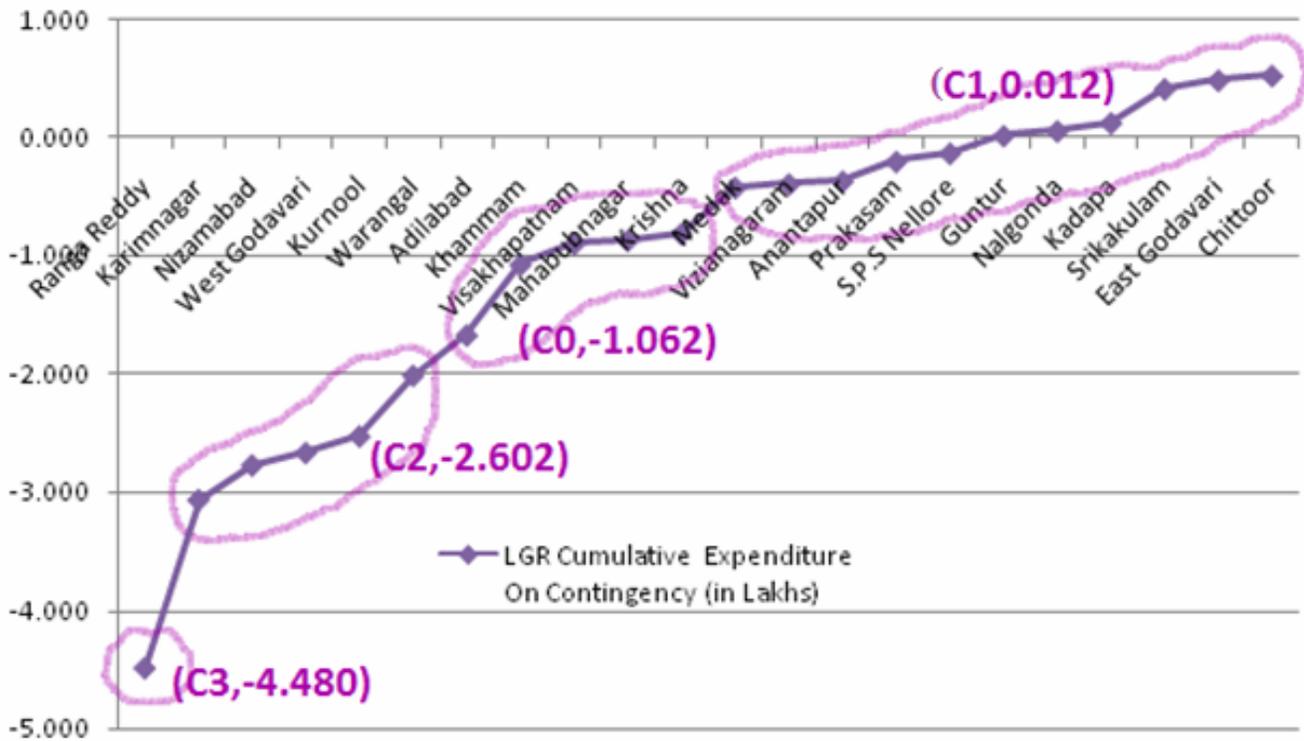# Fig.3: LGR Cumulative Expenditure On Material (in Lakhs)

Fig.4:LGR Cumulative Expenditure On Contingency (in Lakhs)

It is illustrated from the fig.3 that, 4 clusters are identified from the cumulative expenditure on material (in Lakhs) for districts of Andhra Pradesh. The decision tree on which the formation of clusters with the centroids are also shown in fig.3. It is observed from the fig.3 that the expenditure on material is positive and slightly raising only in Kurnool. It is observed from the cluster 3 with centroid -6.296 least amount of expenditure in west Godavari, Krishna, Nizamabad, Nellore Warangal, medak & Prakasam districts.

It is illustrated from the fig.4 that, 4 clusters are identified from the cumulative expenditure on contingency (in Lakhs) for districts of Andhra Pradesh. The decision tree on which the formation of clusters with the centroids are also shown in fig.4. It is observed from the fig.4 that the expenditure on contingency is raising in Nalgonda, Kadapa, Srikakulaum, East Godavari and Chittor districts of Andhra Pradesh. It is identified cluster 3 with centroid -4.480 is outlier and the expenditure on contingency is least for Rangareddy district.

4. CONCLUSIONS

Data mining tool Rapid miner is used to discover the interested pattern on Expenditure Analysis database which is created from NREGS data. Rapid miner, find the clusters on "Expenditure Analysis" database based on other data mining techniques like Characterization, Discrimination, Classification. The results are interpreted in figures. So, the outcome from this work is useful to the government to take the decision on which district the amount has to spend, where to minimize and for what type of works.

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] http://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