

A Comparative Study on Poverty Mapping by Using Data Mining Technique: A Clustering

Sen, Umesh Kumar

Guest Faculty, Dept.of Comp.Sc. & Application, Govt.Mgm Pg College, Itarsi, M.P., India

Abstract

Since we know that poverty is burning issue to across India. To reduce the poverty ratio the Planning Commission estimates levels of poverty in the country on the basis of consumer expenditure surveys conducted by the National Sample Survey Office (NSSO) of the Ministry of Statistics and Programme Implementation. The percentage of the population living below the poverty line in India decreased to 22% in 2011-12 from 37% in 2004-05, according to data released by the Planning Commission in July 2013. This article presents data on recent poverty estimates and goes on to provide a brief history of poverty estimation in the country. The aim of the project is to use data mining technique viz clustering for poverty mapping based on Schedule 63_1.0 consumer expenditure of Socio Economic Survey Sixty third round (July 2006 – June 2007) (Report of NSS) . The main focus is to find the percentage of people below poverty line. This research paper will help the Government to take necessary steps to alleviate poverty in different states and try to uplift the community which is below poverty line in majority of states. A recent survey says that near about 100 crore people's are poor across India.

Keywords

Data mining, NSSO, Planning commission, Clustering.

I. Introduction

The current methodology for poverty estimation is based on the recommendations of an Expert Group to Review the Methodology for Estimation of Poverty (Tendulkar Committee) established in 2005. The Committee calculated poverty levels for the year 2004- 05. Poverty levels for subsequent years were calculated on the basis of the same methodology, after adjusting for the difference in prices due to inflation. The Planning Commission has periodically estimated poverty lines and poverty ratios for each of the years for which Large Sample Surveys on Household Consumer Expenditure have been conducted by the National Sample Survey Office (NSSO) of the Ministry of Statistics and Programme Implementation. These surveys are normally conducted on quinquennial basis. The last quinquennial survey in this series was conducted in 2009-10 (NSS 66th round). However, since 2009-10 was not a normal year because of a severe drought, the NSSO repeated the large scale survey in 2011-12 (NSS 68th round). The summary results of this survey were released on 20th June 2013.

II. Dataset

It based its calculations on the consumption of the following items: cereal, pulses, milk, edible oil, non-vegetarian items, vegetables, fresh fruits, dry fruits, sugar, salt & spices, other food, intoxicants, fuel, clothing, footwear, education, medical (non-institutional and institutional), entertainment, personal & toilet goods, other goods, other services and durables.

Table 1: Shows national poverty levels for the last twenty years, using methodology suggested by the Tendulkar Committee. According to these estimates, poverty declined at an average rate of 0.74 percentage points per year between 1993-94 and 2004-05, and at 2.18 percentage points per year between 2004-05 and 2011-12.

Table 1: National poverty estimates (% below poverty line) (1993 – 2012)

| Year | Rural | Urban | Total |
|-----------|-------|-------|-------|
| 1993 – 94 | 50.1 | 31.8 | 45.3 |
| 2004 – 05 | 41.8 | 25.7 | 37.2 |
| 2009 – 10 | 33.8 | 20.9 | 29.8 |
| 2011 – 12 | 25.7 | 13.7 | 21.9 |

Source: Press Note on Poverty Estimates, 2011 – 12, Planning Commission; Report of the Expert Group to Review the Methodology for Estimation of Poverty (2009) Planning Commission; PRS. State-wise data is also released by the NSSO.

Table 2: Shows state-wise poverty estimates for 2004-05 and 2011-12. It shows that while there is a decrease in poverty for almost all states, there are wide inter-state disparities in the percentage of poor below the poverty line and the rate at which poverty levels are declining.

Table 2: State-wise poverty estimates (% below poverty line) (2004-05, 2011-12)

| State | 2004-05 | 2011-12 | Decrease |
|-------------------|---------|---------|----------|
| Andhra Pradesh | 29.9 | 9.2 | 20.7 |
| Arunachal Pradesh | 31.1 | 34.7 | -3.6 |
| Assam | 34.4 | 32 | 2.4 |
| Bihar | 54.4 | 33.7 | 20.7 |
| Chhattisgarh | 49.4 | 39.9 | 9.5 |
| Delhi | 13.1 | 9.9 | 3.2 |
| Goa | 25 | 5.1 | 19.9 |
| Gujarat | 31.8 | 16.6 | 15.2 |
| Haryana | 24.1 | 11.2 | 12.9 |
| Himachal Pradesh | 22.9 | 8.1 | 14.8 |
| Jammu and Kashmir | 13.2 | 10.4 | 2.8 |
| Jharkhand | 45.3 | 37 | 8.3 |
| Karnataka | 33.4 | 20.9 | 12.5 |

| | | | |
|----------------|------|------|------|
| Kerala | 19.7 | 7.1 | 12.6 |
| Madhya Pradesh | 48.6 | 31.7 | 16.9 |
| Maharashtra | 38.1 | 17.4 | 20.7 |
| Manipur | 38 | 36.9 | 1.1 |
| Meghalaya | 16.1 | 11.9 | 4.2 |
| Mizoram | 15.3 | 20.4 | -5.1 |
| Nagaland | 9 | 18.9 | -9.9 |
| Odisha | 57.2 | 32.6 | 24.6 |
| Puducherry | 14.1 | 9.7 | 4.4 |
| Punjab | 20.9 | 8.3 | 12.6 |
| Rajasthan | 34.4 | 14.7 | 19.7 |
| Sikkim | 31.1 | 8.2 | 22.9 |
| Tamil Nadu | 28.9 | 11.3 | 17.6 |
| Tripura | 40.6 | 14.1 | 26.5 |
| Uttar Pradesh | 40.9 | 29.4 | 11.5 |
| Uttarakhand | 32.7 | 11.3 | 21.4 |
| West Bengal | 34.3 | 20 | 14.3 |
| All India | 37.2 | 21.9 | 15.3 |

Source: Review of Expert Group to Review the Methodology for Estimation of Poverty (2009) Planning Commission, Government of India; Press Note on Poverty Estimates, 2011 – 12 (2013) Planning Commission, Government of India; PRS.

Note: A negative sign before the number in column four (decrease) indicates an increase in percentage of population below the poverty line.

History of Poverty Estimation In India

Pre Independence Poverty Estimates: One of the earliest estimations of poverty was done by Dadabhai Naoroji in his book, 'Poverty and the Un-British Rule in India'. He formulated a poverty line ranging from Rs 16 to Rs 35 per capita per year, based on 1867-68 prices. The poverty line proposed by him was based on the cost of a subsistence diet consisting of 'rice or flour, dhal, mutton, vegetables, ghee, vegetable oil and salt'. Next, in 1938, the National Planning Committee (NPC) estimated a poverty line ranging from Rs 15 to Rs 20 per capita per month. Like the earlier method, the NPC also formulated its poverty line based on 'a minimum standard of living perspective in which nutritional requirements are implicit'. In 1944, the authors of the 'Bombay Plan' (Thakurdas et al 1944) suggested a poverty line of Rs 75 per capita per year.

Post Independence Poverty Estimates: In 1962, the Planning Commission constituted a working group to estimate poverty nationally, and it formulated separate poverty lines for rural and urban areas – of Rs 20 and Rs 25 per capita per year respectively. VM Dandekar and N Rath made the first systematic assessment of poverty in India in 1971, based on National Sample Survey (NSS) data from 1960-61. They argued that the poverty line must be derived from the expenditure that was adequate to provide 2250 calories per day in both rural and urban areas. This generated debate on minimum calorie consumption norms while estimating poverty and variations in these norms based on age and sex.

Alagh Committee (1979): In 1979, a task force constituted by the Planning Commission for the purpose of poverty estimation, chaired by YK Alagh, constructed a poverty line for rural and urban areas on the basis of nutritional requirements. Table 3 shows the

nutritional requirements and related consumption expenditure based on 1973-74 price levels recommended by the task force. Poverty estimates for subsequent years were to be calculated by adjusting the price level for inflation.

Table 3: Minimum calorie consumption and per capita consumption expenditure as per the 1979 Planning Commission task force on poverty estimation

| Area | Calories | Minimum consumption expenditure (Rs per capita per month) |
|-------|----------|---|
| Rural | 2400 | 49.1 |
| Urban | 2100 | 56.7 |

Source: Report of the Expert Group on Estimation of Proportion and Number of Poor, 1993, Perspective Planning Division, Planning Commission; PRS

Lakdawala Committee (1993): In 1993, an expert group constituted to review methodology for poverty estimation, chaired by DT Lakdawala, made the following suggestions: (i) consumption expenditure should be calculated based on calorie consumption as earlier; (ii) state specific poverty lines should be constructed and these should be updated using the Consumer Price Index of Industrial Workers (CPI-IW) in urban areas and Consumer Price Index of Agricultural Labour (CPI-AL) in rural areas; and (iii) discontinuation of 'scaling' of poverty estimates based on National Accounts Statistics. This assumes that the basket of goods and services used to calculate CPI-IW and CPI-AL reflect the consumption patterns of the poor.

Tendulkar Committee (2009): In 2005, another expert group to review methodology for poverty estimation, chaired by Suresh Tendulkar, was constituted by the Planning Commission to address the following three shortcomings of the previous methods: (i) consumption patterns were linked to the 1973-74 poverty line baskets (PLBs) of goods and services, whereas there were significant changes in the consumption patterns of the poor since that time, which were not reflected in the poverty estimates; (ii) there were issues with the adjustment of prices for inflation, both spatially (across regions) and temporally (across time); and (iii) earlier poverty lines assumed that health and education would be provided by the State and formulated poverty lines accordingly. It recommended four major changes: (i) a shift away from calorie consumption based poverty estimation; (ii) a uniform poverty line basket (PLB) across rural and urban India; (iii) a change in the price adjustment procedure to correct spatial and temporal issues with price adjustment; and (iv) incorporation of private expenditure on health and education while estimating poverty. The Committee recommended using Mixed Reference Period (MRP) based estimates, as opposed to Uniform Reference Period (URP) based estimates that were used in earlier methods for estimating poverty.

The Committee computed new poverty lines for rural and urban areas of each state. To do this, it used data on value and quantity consumed of the items mentioned above by the population that was classified as poor by the previous urban poverty line. It concluded that the all India poverty line was Rs 446.68 per capita per month in rural areas and Rs 578.80 per capita per month in urban areas in 2004-05. The following table outlines the manner in which the percentage of population below the poverty line changed after the application of the Tendulkar Committee's methodology.

Table 4: Percentage of population below poverty line calculated by the Lakdawala Committee and the Tendulkar Committee for the year 2004-05

| Committee | Rural | Urban | Total |
|---------------------|-------|-------|-------|
| Lakdawala Committee | 28.3 | 25.7 | 27.5 |
| Tendulkar Committee | 41.8 | 27.5 | 37.2 |

Source: Report of the Expert Group on Estimation of Proportion and Number of Poor, 1993, Perspective Planning Division, Planning Commission; Report of the Expert Group to Review the Methodology for Estimation of Poverty, 2009, Planning Commission; PRS

The Committee also recommended a new method of updating poverty lines, adjusting for changes in prices and patterns of consumption, using the consumption basket of people close to the poverty line. Thus, the estimates released in 2009-10 and 2011-12 use this method instead of using indices derived from the CPI-AL for rural areas and CPI-IW for urban areas as was done earlier. Table 5 outlines the poverty lines computed using the Tendulkar Committee methodology for the years 2004-05, 2009-10 and 2011-12.

Table 5: National poverty lines (in Rs per capita per month) for the years 2004-05, 2009-10 and 2011-12

| Year | Rural | Urban |
|---------|-------|--------|
| 2004-05 | 446.7 | 578.8 |
| 2009-10 | 672.8 | 859.6 |
| 2011-12 | 816.0 | 1000.0 |

Source: Report of the Expert Group to Review the Methodology for Estimation of Poverty(2009) Planning Commission; Poverty Estimates 2009-10 and Poverty Estimates 2011-12, Planning Commission; PRS

Rangarajan Committee: In 2012, the Planning Commission constituted a new expert panel on poverty estimation, chaired by C Rangarajan with the following key objectives: (i) to provide an alternate method to estimate poverty levels and examine whether poverty lines should be fixed solely in terms of a consumption basket or if other criteria are also relevant; (ii) to examine divergence between the consumption estimates based on the NSSO methodology and those emerging from the National Accounts aggregates; (iii) to review international poverty estimation methods and indicate whether based on these, a particular method for empirical poverty estimation can be developed in India, and (iv) to recommend how these estimates of poverty can be linked to eligibility and entitlements under the various schemes of the Government of India. The Committee is expected to submit its report by 2014.

IV. Data Mining Technique Used

For poverty mapping, clustering technique which is a popular unsupervised learning technique has been used. Details of clustering are given below:

A. Clustering

Clustering is a data mining task used to cluster observations into groups of related observations without any prior knowledge of those relationships. The k-means algorithm is one of the popular clustering techniques and it is commonly used in variety of

applications.

B. K-Means Algorithm

The k-means algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters observations into k groups, where k is provided as an input parameter. It then assigns each observation to clusters based upon the observation's proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. Here's The algorithm works as follows:

Each point in the dataset is assigned to the closed cluster, based on the Euclidean distance between each point and each cluster center. Each cluster center is recomputed as the average of the points in that cluster. The above steps are repeated until the clusters converge. It is often a good idea to experiment with different values of k to identify the value that best suits your data. You also may wish to explore the use of other data mining algorithms in your quest for machine-learned knowledge. For the present study, K=15 was found to be optimum to find the people below poverty line.

V. Conclusion

A detailed study has been undertaken for Poverty Mapping in Indian States using clustering which is a data mining technique. Poverty mapping has been done region and state wise. This will definitely help the Government of india to take necessary steps to alleviate poverty in different states.

References

- [1]. Government of India, Planning Commission, July 2013.
- [2]. National Sample Survey Organisation (NSSO) figures of the Ministry of Statistics.
- [3]. B. Chandra, Professor, Department of Mathematics, Indian Institute of Technology, Delhi.
- [4]. Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.
- [5]. Dept. of Statistics, across India.
- [6]. Fayyad, U: Data Mining and Knowledge Discovery in Databases .
- [7]. Frawley and G. Piatetsky - Shapiro, Knowledge Discovery in Databases: An Overview. Published by the AAAI Press/ The MIT Press, Menlo Park, C.A 1996.