Data Mining Term Project

Self-Evaluation Report

- WORK DIVISION

1. AKSHAT GUPTA (2017A7PS1699H)

   Code Development (Aggregation of data, preprocessing)

   Documentation

   Plotting of graphs using Tableau.

2. ABHISHEK BHARDWAJ (2017A7PS1497H)

   Code Development (clustering, python plots)

   Documentation

   Dataset Research

3. ARUN INANI (2017A7PS0085H)

   Code Development (preprocessing)

   Documentation

   Dataset Research

- Achievements

  - We got to learn a lot about various clustering algorithm viz. K-Means, DBSCAN etc. and how DBSCAN is a better choice for our dataset.

  - We got to learn about various Preprocessing Techniques viz. Noise removal (using regular expression and other techniques), Data Reduction, Data Aggregation etc.

  - We learnt how to plot various types of graphs using Tableau.

- Short Comings

  - The Data of all the four years was exactly same. Hence the Time series analysis which planned to do is not possible. We also tried to do the same by clubbing our current dataset with any other dataset but were not able to find a suitable one.

  - There were few villages with name as alphanumeric strings (ex. 7BSD,3DW), we removed such rows using a regular expression. But for other noise in village names like false/incorrect names, it is not possible to remove any of these entries based on human judgment in this case.

  - There were few entries with negative population values or total population equal to zero, such rows were removed as a part of data preprocessing. But the cases in which value of population for any village is greater than actual population, those can't be determined.

- IMPORTANT NOTE

  We tried various types of Mining Algorithms on our data viz. Association Rule Mining, K-Means, DBSCAN etc. We weren't able to draw any useful or conclusive information using association rule mining due to quantitative nature of our Dataset. We preferred DBSCAN over K-means because

clustering is initialization dependent in K-means and it might have resulted in some biased observations.