

Team members: Team 05

1. Achyuta Krishna
2. Ajith P J
3. Alvin Ronnie
4. Anirudh G

Project Title: Analysis of Suicide Statistics of India

Team achievements:

1. Initially with Exploratory data analysis, we were able to identify redundancies in dataset, relation between various attributes and draw rough conclusions from the dataset.
2. The chosen dataset was pre-processed using techniques like :
  - a. Data Cleaning
  - b. Aggregation
  - c. Attribute Subset Selection
  - d. Encoding
  - e. Normalization
3. We used 3 data analysis techniques to derive meaningful insights from the dataset.
  - a. Cluster Analysis : We identified 6 clusters and unique properties for each cluster. From this, we were able to identify how the “reason for suicide” affected different sections of the population in a varying degree.
  - b. Outlier Detection: We were able to identify the points which varied significantly from the other points on the basis on number of suicides. Insights were drawn on analysing the properties of these outliers.
  - c. Classification: We were able to create a model to predict the range of possible values for the number of male suicides given the total number of suicides, location, cause for suicide and year.

Team reflection:

1. As a result of working with real world datasets, we were exposed to the difficulties of dealing with noise and redundancy in dataset and were able to learn data cleaning and pre-processing techniques.
2. After trying some techniques like normalization and association analysis, we were unable to get meaningful values due to some constraints which come with the chosen dataset. Hence, we were able to identify appropriate techniques which may work for one dataset but won't for another.
3. Our python language proficiency and knowledge of many python libraries became better as a result of learning and using python for the project.
4. We were able to understand the importance of data mining in this world which has vast quantity of data from which it is not easy to derive conclusions.

Teamwork division:

- Achyuta Krishna: Aggregation, Encoding, Attribute subset selection and Outlier Analysis.
- Ajith P J: Aggregation, Normalisation and Encoding.
- Alvin Ronnie: Exploratory data analysis, Plots and Classification Analysis.
- Anirudh G: Exploratory data analysis, Data cleaning, Aggregation and Clustering Analysis.