Module 3: Introduction to Machine Learning with Python

Case Study

edureka!

edureka!

Case Study

Objective:

1. Make the learner load the data using pandas.
2. Create new columns in dataset from existing columns.
3. Use pandas to answer questions of interest.
4. Plot variables of interest.

Dataset used: Prisoners dataset sourced from data.gov.in

Questions:

1. Data Loading:
   a. Load the dataset "prisoners.csv" using pandas and display the first and last five rows in the dataset.
      [Hint: Refer to read_csv, head and tail methods in pandas]

   b. Use describe method in pandas and find out the number of columns. Can you say something about those rows who have zero inmates?
      [Hint: Use the loc attribute of dataframe combined with conditional checks]

2. Data Manipulation:
   a. Create a new column -'total_benefitted' that is a sum of inmates benefitted through all modes.
      [Hint: Use sum method with appropriate axis]

   b. Create a new row - "totals" that is the sum of all inmates benefitted through each mode across all states.

3. Plotting:

a. Make a bar plot with each state name on the x -axis and their total benefitted inmates as their bar heights. Which state has the maximum number of beneficiaries?
[Hint: Use bar method of pyplot]

b. Make a pie chart that depicts the ratio among different modes of benefits.
[Hint: Use pie method of pyplot]