INNOMATICS RESEARCH LABS

Data App Development

Title :**Data Science Automation Tool**

Domain : Application Development

Prerequisites : **Python, Flask, Web Technologies, EDA and ML**

Tasks:

Given any dataset, your application should be able to perform below mentioned tasks

1. Read Dataset
2. Display Dataset information

   Data Cleaning

   1. Type Conversion
   2. Missing Value Treatment
   3. Outlier Treatment
   4. Transformations
   5. Data imbalancing

3. Model building
4. Hyper parameter tuning
5. Evaluation metrics
6. Testing

Project Description :

In this project we are going to develop a tool, which can automate all the operations on the given dataset to build the machine learning model.

Tasks to be performed:

Sprint - 1

- Create a layout page with **upload** button to upload any kind of dataset
- Once the dataset has been uploaded, you need to display the buttons like **info, shape** and **Columns**, and those buttons should perform their respective operations.
- You should display the output on the right side once I click on any button.
- And there is one more button in the bottom with **Strat Data Analysis** for Data Analysis on the uploaded dataset

Sprint - 2

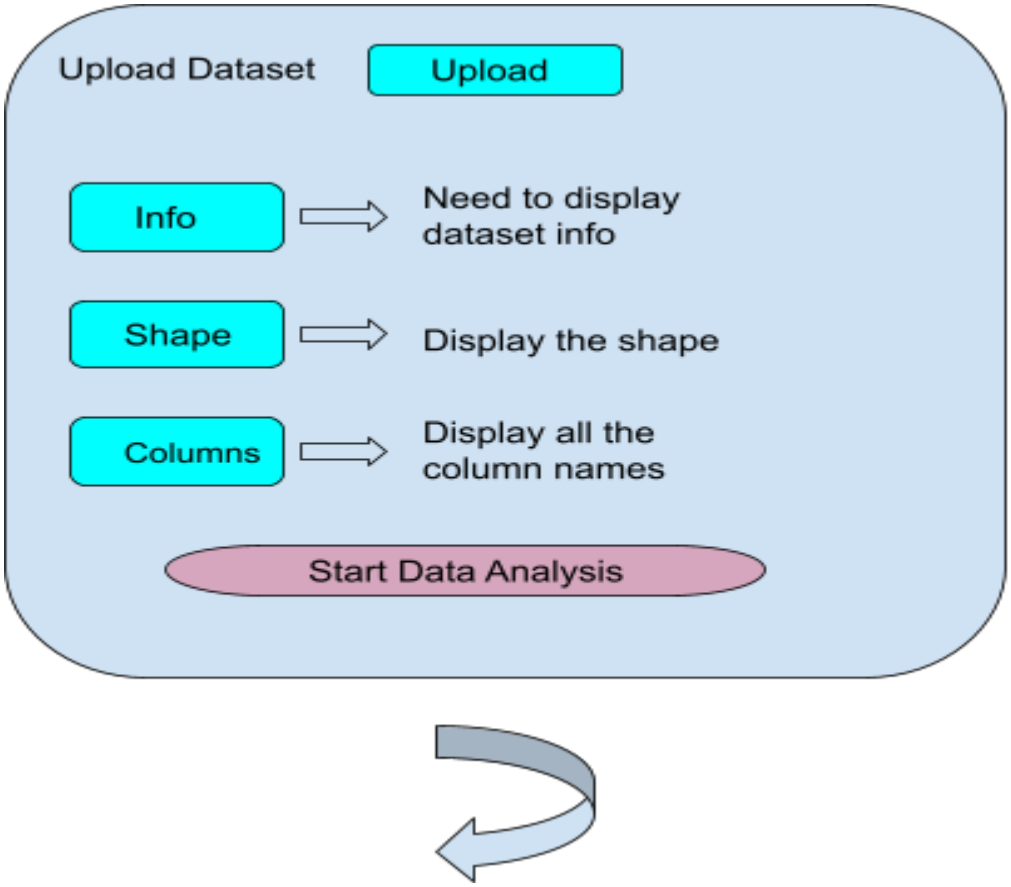
- ★ Create functions for all the operations of EDA like **type_conversion(), missing_value_treatment(), outlier_treatment(), transformations(), data_imbalancing()** and so on…
- ★ Function should be applicable for any kind of dataset like for example if you want to handle missing values for one column, first it should check for the datatype for the column and if it is an integer type, then it should ask whether you want to drop the data or replace. If replace shows us different methods to replace the null values.
- ★ Each function should contain all the possibilities to handle the data.
- ★ After that finally should show the button with **Train-Test split** for splitting the data with the given percentage

Sprint - 3

- ➔ Need to build the model on cleaned data
- ➔ Should show three options like **Regression, CLassifications** and **Clustering**
- ➔ Based on the selected option should show all the algorithms
- ➔ If we select any algorithm it should show all the metrics results like **accuracy, recall, precision** and so on.

Sample Architecture :

INNOMATICS
RESEARCH LABS



INNOMATICS
RESEARCH LABS