

---

# Neuro Data Engineering Online EDA Automation

---

Low Level Design (LLD)

Vamsi Alla

October 23, 2022  
INEURON

## 1. INTRODUCTION

The initial and the most crucial part of any model building is making the data set ready for next step. However it consumes 70-80% of the time which can be reduced. An initial knowledge of the data is very much required to go for advanced data mining. This work discusses the implementation of unmanned basic exploration of the data in order to reduce the time of the scientist so they can work on other important part of the work. And this work also explains the behind implementation of data mining and can be very much useful for new data analytics.

The solution proposed here is an Automated EDA which can be implemented to perform above mentioned use cases. Initially the Automated EDA will take the dataset from the client, will process the dataset into DataFrame and on this processed DataFrame further operations will be performed. The choice of operations are fairly dependent on the user and the copy of the performed operations will be made available to the client on the provided email.

## PROBLEM STATEMENT

Create a web app application to perform Data cleaning, Feature engineering, and EDA. The web application should allow the user to perform various data transformation operations on the dataset with help of prebuild component. Users must be able to drag and drop the existing component at UI to perform any operation.

## 2. DATASET INFORMATION

Select the File form UI w.r.t file type and perform the following operations

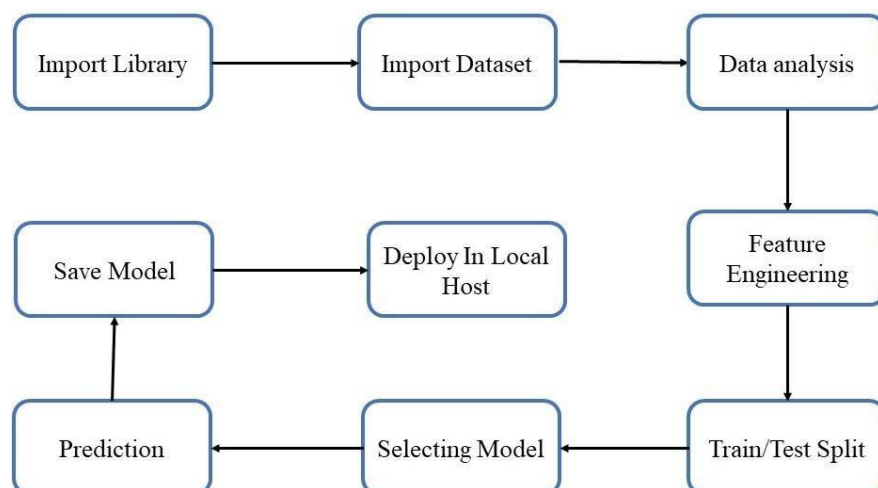
To create an automated exploration of data at the initial stage will make following processes faster:

To reduce coding so the exploration of the data can be made possible for management

To reduce time involved in basic exploration of data set and data mining

To demonstrate the working of exploring data to the data science enthusiast

To make the code readily available to the data scientists



## 1. Architecture Description.

### 1.1. Data Description:

The dataset was taken from any source and type of file is csv,tsv and mongo db.

#### **4.2 Data Pre-processing.**

This included importing of important libraries such as seaborn, matplotlib, pandas etc.

#### **4.3. Data Analysis**

Here we handled the null values, changed the column names, plotted multiple graphs in seaborn, matplotlib and other visualization library for proper understanding of the data and the distribution of information in the same. As there were no null values in the data, we proceeded with the visualization and analysis.

For each specific feature we analysed the data using visualization, and jotted down the important key points which can impact the final predictions.

#### **4.4. Feature Engineering**

Merging 2 or more columns to get in-depth knowledge and information regarding the data.

#### **4.5 EDA automation**

To create an automated exploration of data at the initial stage will make following processes faster:

To reduce coding so the exploration of the data can be made possible for management

To reduce time involved in basic exploration of data set and data mining

To demonstrate the working of exploring data to the data science enthusiast

To make the code readily available to the data scientists

#### **4.8 Deploy in Local Host**

We created a HTML template and deployed the model through Flask.

Here is the image of the same:

