Campus Placement Prediction

Objective: The main goal is to predict whether the student will be recruited in campus placements or not based on the available factors in the dataset.

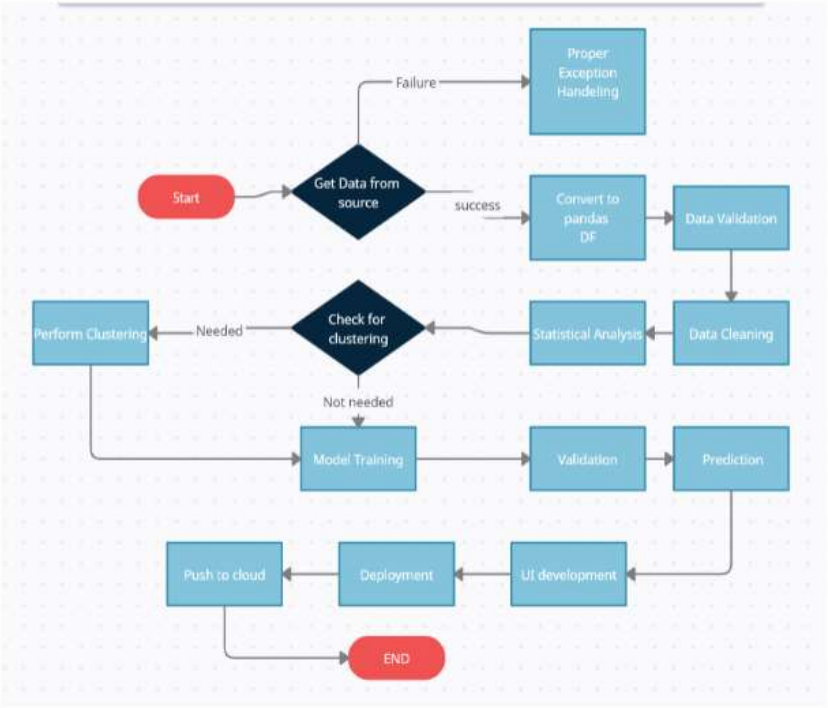
Benefits:

This Campus Placement Prediction Project will always be helpful to both the students, as well as the institution.

Data Sharing Agreement :

- Sample file name
- Length of date stamp(8 digits)
- Length of time stamp(6 digits)
- Number of Columns
- Column names
- Column data type

Architecture



Data Validation and Data Transformation :

- Checking for missing values: Identify any missing data points and decide on appropriate strategies for handling them, such as imputation or deletion.
- Handling outliers: Identify and analyze outliers in the dataset and determine whether they should be removed or transformed.
- Ensuring consistency: Check for inconsistencies or errors in the data, such as conflicting information or formatting issues.
- Feature scaling: Scale numerical features to a similar range to prevent certain features from dominating the model training process.
- Encoding categorical variables: Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding.

Model Training:

Data Export from Db :

The accumulated data from db is exported in csv format for model training

Data Preprocessing

Performing EDA to get insight of data like identifying distribution , outliers ,trend

among data etc.

Check for null values in the columns. If present impute the null values.

Encode the categorical values with numeric values.

Perform Standard Scalar to scale down the values.

Model Selection

Here we selected two algorithms LogisticRegression and RandomForestClassifier.

Prediction:

Once the logistic regression model is trained and evaluated, use it to make predictions on new or unseen data. Here we can use the trained model to predict the placement status of students based on their attributes such as 'gender', 'ssc_p', 'hsc_p', 'degree_p', etc.