**Problem statement:**

Predict whether the employee will stay or leave the Job(0, 1).

If target = 0, indicates the employee is not looking for a change in job and target = 1, indicates the employee is looking for a change in job

**Solution:**

Using one of the binary classification algorithms we are attempting to come up with a solution to accurately classify the given employees.

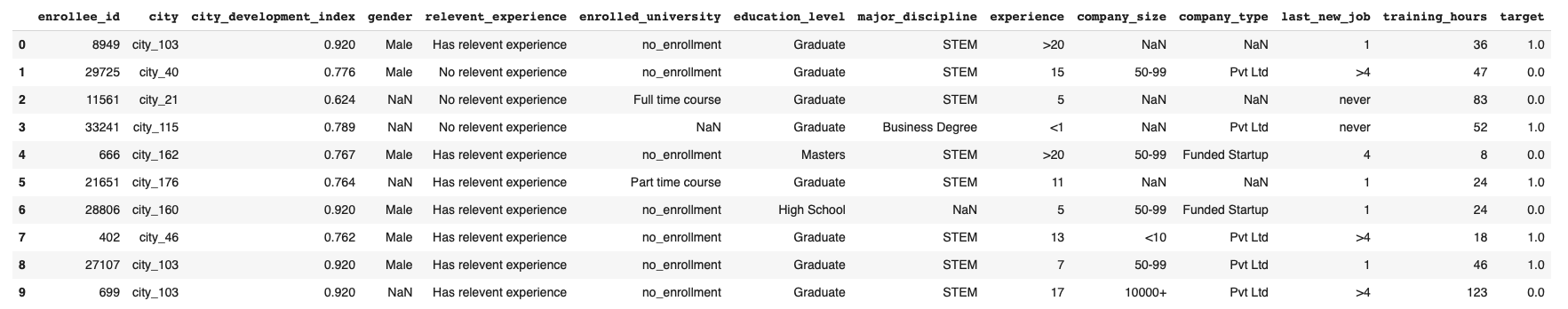
We are using binary classification as there are two target classes {0,1}

0 -> not changing

1 -> changing

**Description of data set:**

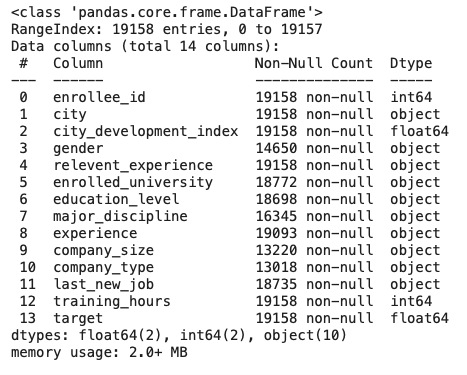
df.head():



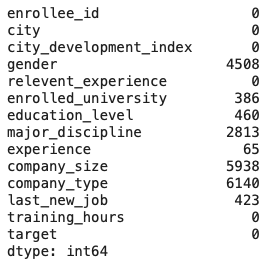
df.describe(): Gives the summary of dataset



df.info(): Gives data type and no. of non-null values



df.isnull().sum(): Gives the count of null Values



**Libraries Used:**

Numpy, pandas, seaborn , matplotlib.pyplot, sklearn

**Description of the used function:**

* **User-defined :**

def replace(column) -> discretized values (label encoding)

Function to replace and map corresponding value

* **From Libraries :**

From sklearn library :

KMeans,KNeighborsClassifier,GaussianNB,RandomForestClassifier,

DecisionTreeClassifier.

Confusion\_matrix,accuracy\_score, roc\_curve,auc

GridSearchCV

From pandas library:

* read.csv(): reads a given csv file
* head(): returns first 5 entries in the dataset
* shape: returns dimension of the dataset
* info(): returns datatype of all columns, non-null counts
* describe(): returns count, mean, standard deviation, minimum value, maximum value, quartile values, from each column
* nunique(): return number of unique values for each column
* isnull().sum(): returns the count of null values in each column
* value\_counts() : returns the total count unique values
* fillna(): fills null value , dropna(): drops null values
* loc: locis label-based,specify the name of the rows and columns that we need to filter out.
* iloc:iloc is integer index-based,specifying rows and columns by their integer index.
* apply():
* astype(): change the datatype of columns to desired datatype
* get\_dummies(): label encoding for categorical columns

From numpy library:

* dtypes(): gives the datatype of all columns
* concatenate(): merging
* reshape(): reshaping data

From seaborn/matplotlib library

* histplot(): plotting histogram
* plt.show(): to display graph
* plt.xlabel(),plt.ylabel(),plt.title(): labeling description
* plt.subplots(): plotting in grids

**Other Python in-built functions used:**

* map(): returns a map object of the results
* Lambda (): List Comprehension
* copy(): Makes a copy
* range(): Takes multiple values in the specified range
* append(): It adds values at the end.
* input(): Takes input from user
* zip(): mapping output similar to enumerate just different in syntax

Made use lists, dictionaries, tuples.

Applied the knowledge of

List slicing and list comprehension

Iteration

Inbuilt methods to work with list objects

Accessing dictionary key and values

**Description of the Used algorithm:**

**Functions that the algorithm uses:**

fit(): It fits the data in the learning algorithm

predict(): It predicts the results

Confusion\_matrix, Accuracy\_score, Best\_params\_: Evaluation parameter

**Team members:**

Archana Kalburgi

Rohit Pradhan

**Task allocation:**

**Archana**

Exploratory data analysis

Implementation of K- nearest neighbors algorithm

Implementation of Naive-Bayes algorithm

**Rohit Pradhan**

Implementation of Decision Trees algorithm

Implementation of Random Forest algorithms

Arrived at the conclusion using ROC and AUC curves