Problem Statement: Predict the churn risk rate

Approach

* Library used are pandas, numpy and sklearn.
* Extracted the csv file in pandas Dataframe.
* Data Cleaning Procedure:
  + Check for null values and filled the null values with numeric value mean and non-numerical value forward propagation fillna ffill method
  + Selecting some features based on their importance and discarded the features which has no contribution towards the dataset.
  + All the categorical features which are in string are converted to labels using Label Encoder of sklearn
  + All the numerical features are cleaned by removed the outliers from the dataset.
  + Values less than 0 and negative values has been removed also string values within numerical value has been removed as well.
* Machine Learning Implementation
  + 3 Algorithms is implemented
    - Logistic Regression with 26%
    - Random Forest Classifier with 73%
    - XG Boost Classifier with 75%
* Test Data
  + The Nan value is filled with previous value using fillna function.
  + Train dataset columns are encoded as the model is trained on encoded columns of train set
  + Random Forest Classifier is used for predictions.