FINGERTIPS

Python Case Study

Telecom Churn Analysis

Evaluation Parameters:

Total Marks – 200

Part – 1
Data cleaning – 50
EDA – 50

Part – 2 (Modelling)
Accuracy of classification model- 75
Clustering – 25

Project Description:

In this particular project, we are using a dataset that contains information like State, account length, international plan, Total day cells, Total day charge and using that to predict whether the customer will remain a customer in the future.

However, before you go ahead and make a prediction, it is advised that you first pre-process the data, since it may contain some irregularities and noise.

In addition, try various tricks and techniques in order to gain the best accuracy in your predictions.

Data Details:

- **State:** Self Explanatory **(**string)
- **Account length**: Self Explanatory **(**integer)
- **Area code:** Self Explanatory **(**integer)
- **International plan:** Self Explanatory **(**string)
- **Voice mail plan:** Self Explanatory **(**string)
- **Number vmail messages:** Self Explanatory **(**integer)
- **Total day minutes:** Self Explanatory **(**double)
- **Total day calls:** Self Explanatory (integer)
- **Total day charge:** Self Explanatory (double)
- **Total eve minutes:** Self Explanatory (double)
- **Total eve calls:** Self Explanatory **(**integer)
- **Total eve charge:** Self Explanatory (double)
- **Total night minutes:** Self Explanatory **(**double)
- **Total night calls:** Self Explanatory **(**integer)
- **Total night charge:** Self Explanatory **(**double)
- **Total intl minutes:** Self Explanatory **(**double)
- **Total intl calls:** Self Explanatory **(**integer)
- **Total intl charge:** Self Explanatory **(**double)
- **Customer service calls:** Self Explanatory **(**integer)
- **Churn:** Self Explanatory **(**string)

Part-1: Data Exploration and Pre-processing

1) load the given dataset
2) print all the column names
3) describe the data
4) find all the Null values
5) plot the customers who have international plans
6) plot the customers who have Voice mail plan
7) Plot the total day calls
8) Plot the total day charge
9) Display pie chart for value count in Churn column
10) Display a scatter plot between total day calls and total day charges
11) Display a scatter plot between total day calls and total night calls
12) Display a boxplot of Total day minutes with respect to Churn

13) Display a boxplot of Total day charge with respect to Churn

Part-2: Working with models

1) Perform encoding on churn
2) Perform encoding on International Plan
3) Perform encoding on voice mail plan using sklearn
4) Check the correlation among all the columns
5) Create features and target data. Only select features data that are highly correlated with target data.
6) Scale the target data (churn)
7) Check the shape of both training data and testing data
8) Apply Logistic regression
9) Display confusion matrix
10) Perform Hyper parameter tuning
11) Create a model
12) Check the model score of both training and testing data
13) Perform cross validation technique with SVM Classifier
14) Perform hyperparameter tuning with different classifier models
15) Perform k-means clustering on dataset and divide it into four clusters
16) Apply PCA give n components value to 3 show we only get 3 columns after applying PCA