

Classification Activity

For Churn Modelling.csv dataset, perform the following

1) Data Cleaning

- a) Analyse missing Values in the data. If you found any, explore Imputation techniques which can be used to impute them.
- b) Check for outliers in the data, see if any of the columns in the data has to be modelled as numeric or categorical. If yes, explain the reasoning and do the conversion.
- d) Figure out irrelevant rows and columns. Based on your understanding, drop them
- e) Derive and explain insights based on descriptive statistics

2) Exploratory Analysis

- a) Perform Univariate Analysis for all variables and comment on the insights
- b) Perform Bivariate analysis for all variables with attrition column and comment on the insight
- c) Write a code to perform chi square test of dependency with all categorical and target – attrition.
- d) Write a code to bin numeric columns and then perform chi square test of dependency with target – attrition.

3) Modelling

- a) Build a model based on logistic regression using stats models.
- b) Fine tune the model based by changing scaling techniques, dropping insignificant variables, fine tune the metric by implementing the concepts learnt so far.
- c).Choose a metric based on business understanding of the data.
- d) Check for ROC AUC curve, AUC curve, also draw precision recall curve