**BANK TERM DEPOSIT PREDICTION**

**INFERENCE**

**Data exploratory**

•Provided the detailed analysis of each and every attribute of the data set using plots.

•Found the anamoly in housing loan column with value xxxyy which is handled in data preprocessing.

**Data preprocessing**

•Dropping NA Values: From the given dataset, it is seen that 8 records have empty values. Since it is negligible (< 1%) these entries can be deleted.

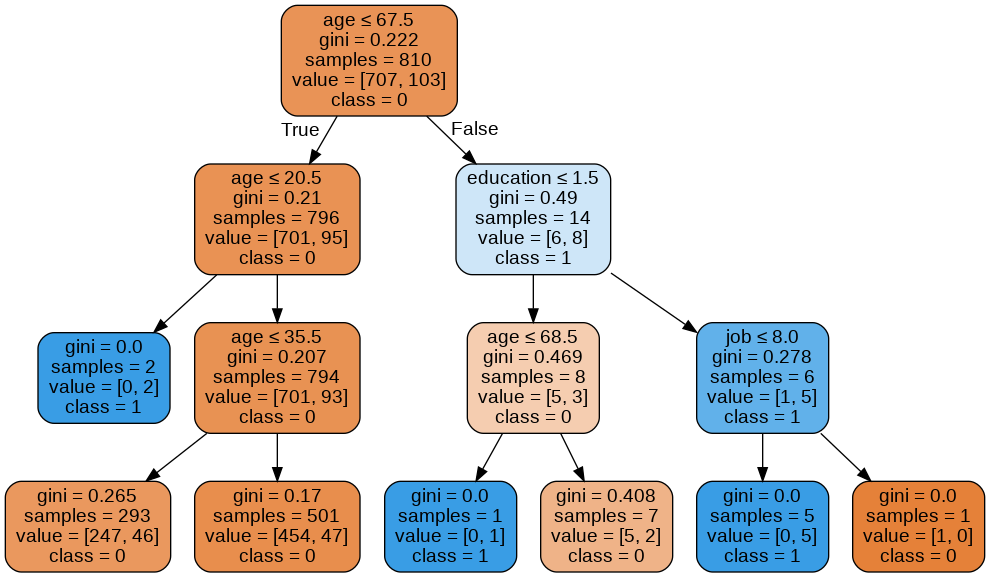
•Replacing with max occurence of the field: The anamoly in housing column with value "xxxyy" replaced with maximum occurence value("yes") of the field.

**Data selection, Model creation and testing**

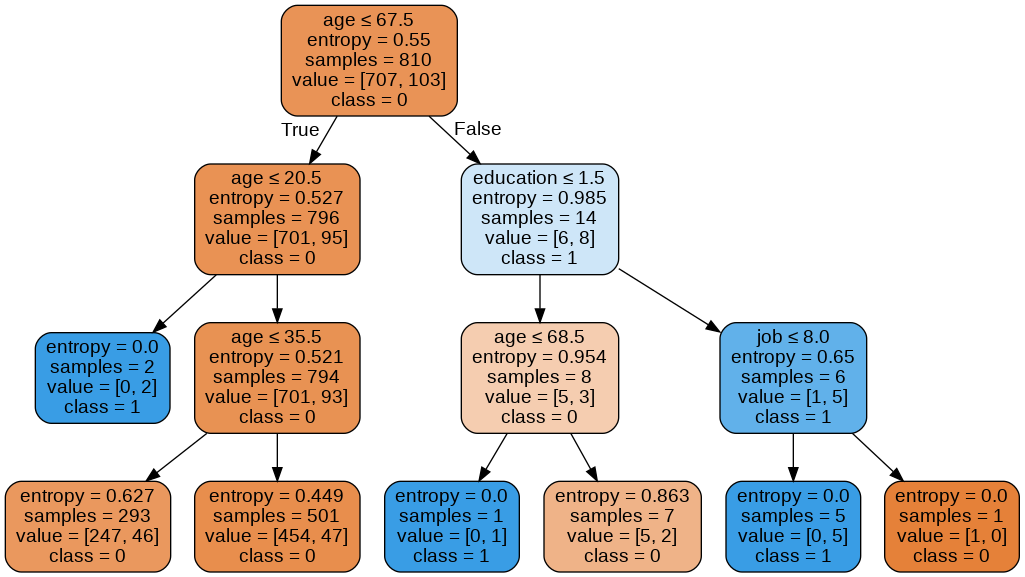
•From the given data set, 80% data is used for training purpose and remaining 20% is used for testing purpose.

•Verified the accuracy with both entropy and gini index classifier criterion.

**Decision tree classifier using Gini**

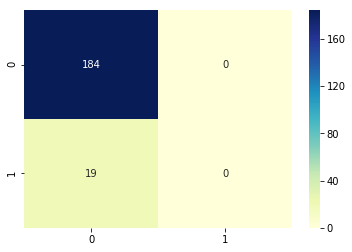


**Decision tree classifier using Entropy**



**Evaluate the model performance**

•Evaluated the model performance with confusion matrix and classification report.



**Cross Validation**

•Used Cross validate function (sklearn.model\_selection.cross\_validate) since 1) It allows specifying multiple metrics for evaluation. 2) It returns a dict containing fit-times, score-times (and optionally training scores as well as fitted estimators) in addition to the test score.

•AUC score obtained by using cross validate function for the given data set is 0.607.

