* Decision Tree was used for modeling since the output from Decision Tree is easy to explain
* Used statistical significant features to predict Churn – 16 such features
  + ['Tenure', 'PreferredLoginDevice', 'CityTier', 'WarehouseToHome', 'PreferredPaymentMode', 'Gender', 'NumberOfDeviceRegistered', 'PreferedOrderCat', 'SatisfactionScore', 'MaritalStatus', 'NumberOfAddress', 'Complain', 'OrderAmountHikeFromlastYear', 'OrderCount', 'DaySinceLastOrder', 'CashbackAmount']
* Handle Missing Values
  + Impute missing values with mean for the following features
    - Tenure 264 (4.7%)
    - WarehouseToHome 251 (4.5%)
    - OrderAmountHikeFromlastYear 265 (4.7%)
    - OrderCount 258 (4.5%)
    - DaySinceLastOrder 307 (5.5%)
* Encode Categorical Features
  + ['Gender', 'PreferredLoginDevice', 'PreferredPaymentMode', 'PreferedOrderCat', 'MaritalStatus']
* Encoding –
  + Gender –
    - {'Female': 0, 'Male': 1}

* + PreferredLoginDevice –
    - {'Computer': 0, 'Mobile Phone': 1, 'Phone': 2}
  + PreferredPaymentMode –
    - {'CC': 0, 'COD': 1, 'Cash on Delivery': 2, 'Credit Card': 3, 'Debit Card': 4, 'E wallet': 5, 'UPI': 6}
  + PreferedOrderCat -
* {'Fashion': 0, 'Grocery': 1, 'Laptop & Accessory': 2, 'Mobile': 3, 'Mobile Phone': 4, 'Others': 5}
  + MaritalStatus -
    - {'Divorced': 0, 'Married': 1, 'Single': 2}

* + Train/Test Split – 80% for training, 20% for testing
    - 1126 test samples, 941 no-churn and 185 churn
  + Hyper-parameter tuning
    - Decision Criterion = ‘gini’
    - Max depth of the tree = Tried for 3, 5, 7, .., 15
  + Performance Metric:
    - Precision, recall and F1
    - Class distribution is not balanced so it is important to look at precision, recall and F1 for each class instead of overall accuracy.

**Model-1 – Very simple and easy to explain decision tree**

**Set max-depth of the decision tree to 3**

precision recall f1-score support

0 0.91 0.96 0.93 941

1 0.72 0.50 0.59 185

accuracy 0.89 1126

macro avg 0.81 0.73 0.76 1126

weighted avg 0.88 0.89 0.88 1126

We sacrifice accuracy (especially F1-score for churn class) to get a simple and easily explainable decision tree

[[905 36]

[ 92 93]]

|--- Tenure <= 1.50

| |--- Complain <= 0.50

| | |--- NumberOfAddress <= 4.50

| | | |--- class: 0

| | |--- NumberOfAddress > 4.50

| | | |--- class: 1

| |--- Complain > 0.50

| | |--- NumberOfAddress <= 3.50

| | | |--- class: 1

| | |--- NumberOfAddress > 3.50

| | | |--- class: 1

|--- Tenure > 1.50

| |--- CashbackAmount <= 124.38

| | |--- Complain <= 0.50

| | | |--- class: 0

| | |--- Complain > 0.50

| | | |--- class: 1

| |--- CashbackAmount > 124.38

| | |--- Complain <= 0.50

| | | |--- class: 0

| | |--- Complain > 0.50

| | | |--- class: 0

**Decision Rules for Detecting Churn:**

* + If Tenure <= 1.5 months and Complain = 1
  + If Tenure <= 1.5 months and Complain = 0 and NumberOfAddress > 4
  + If Tenure > 1.5 months and CashbackAmount <= 124.38 and Complain = 1

**Model-2 – Increasing Accuracy with complexity of the tree**

**Max-depth = 5**

precision recall f1-score support

0 0.91 0.98 0.94 941

1 0.84 0.49 0.62 185

accuracy 0.90 1126

macro avg 0.87 0.73 0.78 1126

weighted avg 0.90 0.90 0.89 1126

[[924 17]

[ 95 90]]

|--- Tenure <= 1.50

| |--- Complain <= 0.50

| | |--- NumberOfAddress <= 4.50

| | | |--- DaySinceLastOrder <= 1.50

| | | | |--- WarehouseToHome <= 22.50

| | | | | |--- class: 0

| | | | |--- WarehouseToHome > 22.50

| | | | | |--- class: 1

| | | |--- DaySinceLastOrder > 1.50

| | | | |--- DaySinceLastOrder <= 8.50

| | | | | |--- class: 0

| | | | |--- DaySinceLastOrder > 8.50

| | | | | |--- class: 1

| | |--- NumberOfAddress > 4.50

| | | |--- OrderAmountHikeFromlastYear <= 18.50

| | | | |--- NumberOfDeviceRegistered <= 2.50

| | | | | |--- class: 0

| | | | |--- NumberOfDeviceRegistered > 2.50

| | | | | |--- class: 1

| | | |--- OrderAmountHikeFromlastYear > 18.50

| | | | |--- SatisfactionScore <= 4.50

| | | | | |--- class: 0

| | | | |--- SatisfactionScore > 4.50

| | | | | |--- class: 1

| |--- Complain > 0.50

| | |--- NumberOfAddress <= 3.50

| | | |--- DaySinceLastOrder <= 1.50

| | | | |--- NumberOfDeviceRegistered <= 1.50

| | | | | |--- class: 0

| | | | |--- NumberOfDeviceRegistered > 1.50

| | | | | |--- class: 1

| | | |--- DaySinceLastOrder > 1.50

| | | | |--- DaySinceLastOrder <= 5.50

| | | | | |--- class: 0

| | | | |--- DaySinceLastOrder > 5.50

| | | | | |--- class: 1

| | |--- NumberOfAddress > 3.50

| | | |--- OrderAmountHikeFromlastYear <= 18.50

| | | | |--- class: 1

| | | |--- OrderAmountHikeFromlastYear > 18.50

| | | | |--- PreferredPaymentMode <= 4.50

| | | | | |--- class: 1

| | | | |--- PreferredPaymentMode > 4.50

| | | | | |--- class: 1

|--- Tenure > 1.50

| |--- CashbackAmount <= 124.38

| | |--- Complain <= 0.50

| | | |--- DaySinceLastOrder <= 0.50

| | | | |--- SatisfactionScore <= 3.50

| | | | | |--- class: 0

| | | | |--- SatisfactionScore > 3.50

| | | | | |--- class: 1

| | | |--- DaySinceLastOrder > 0.50

| | | | |--- WarehouseToHome <= 30.50

| | | | | |--- class: 0

| | | | |--- WarehouseToHome > 30.50

| | | | | |--- class: 0

| | |--- Complain > 0.50

| | | |--- SatisfactionScore <= 3.50

| | | | |--- NumberOfAddress <= 4.50

| | | | | |--- class: 0

| | | | |--- NumberOfAddress > 4.50

| | | | | |--- class: 1

| | | |--- SatisfactionScore > 3.50

| | | | |--- PreferedOrderCat <= 2.50

| | | | | |--- class: 0

| | | | |--- PreferedOrderCat > 2.50

| | | | | |--- class: 1

| |--- CashbackAmount > 124.38

| | |--- Complain <= 0.50

| | | |--- PreferedOrderCat <= 0.50

| | | | |--- SatisfactionScore <= 2.50

| | | | | |--- class: 0

| | | | |--- SatisfactionScore > 2.50

| | | | | |--- class: 0

| | | |--- PreferedOrderCat > 0.50

| | | | |--- OrderCount <= 6.50

| | | | | |--- class: 0

| | | | |--- OrderCount > 6.50

| | | | | |--- class: 0

| | |--- Complain > 0.50

| | | |--- SatisfactionScore <= 2.50

| | | | |--- NumberOfAddress <= 10.50

| | | | | |--- class: 0

| | | | |--- NumberOfAddress > 10.50

| | | | | |--- class: 0

| | | |--- SatisfactionScore > 2.50

| | | | |--- CityTier <= 2.50

| | | | | |--- class: 0

| | | | |--- CityTier > 2.50

| | | | | |--- class: 0

**Model-3 – Increasing Accuracy with complexity of the tree**

**Max-depth = 7**

precision recall f1-score support

0 0.93 0.98 0.95 941

1 0.85 0.62 0.71 185

accuracy 0.92 1126

macro avg 0.89 0.80 0.83 1126

weighted avg 0.92 0.92 0.91 1126

[[921 20]

[ 71 114]]

|--- Tenure <= 1.50

| |--- Complain <= 0.50

| | |--- NumberOfAddress <= 4.50

| | | |--- DaySinceLastOrder <= 1.50

| | | | |--- WarehouseToHome <= 22.50

| | | | | |--- NumberOfDeviceRegistered <= 5.50

| | | | | | |--- Gender <= 0.50

| | | | | | | |--- class: 0

| | | | | | |--- Gender > 0.50

| | | | | | | |--- class: 1

| | | | | |--- NumberOfDeviceRegistered > 5.50

| | | | | | |--- class: 1

| | | | |--- WarehouseToHome > 22.50

| | | | | |--- NumberOfDeviceRegistered <= 4.50

| | | | | | |--- class: 1

| | | | | |--- NumberOfDeviceRegistered > 4.50

| | | | | | |--- CityTier <= 1.50

| | | | | | | |--- class: 0

| | | | | | |--- CityTier > 1.50

| | | | | | | |--- class: 1

| | | |--- DaySinceLastOrder > 1.50

| | | | |--- DaySinceLastOrder <= 8.50

| | | | | |--- SatisfactionScore <= 2.50

| | | | | | |--- PreferedOrderCat <= 1.00

| | | | | | | |--- class: 1

| | | | | | |--- PreferedOrderCat > 1.00

| | | | | | | |--- class: 0

| | | | | |--- SatisfactionScore > 2.50

| | | | | | |--- PreferredLoginDevice <= 0.50

| | | | | | | |--- class: 0

| | | | | | |--- PreferredLoginDevice > 0.50

| | | | | | | |--- class: 0

| | | | |--- DaySinceLastOrder > 8.50

| | | | | |--- OrderCount <= 5.00

| | | | | | |--- class: 1

| | | | | |--- OrderCount > 5.00

| | | | | | |--- PreferredPaymentMode <= 3.50

| | | | | | | |--- class: 0

| | | | | | |--- PreferredPaymentMode > 3.50

| | | | | | | |--- class: 1

| | |--- NumberOfAddress > 4.50

| | | |--- OrderAmountHikeFromlastYear <= 18.50

| | | | |--- NumberOfDeviceRegistered <= 2.50

| | | | | |--- PreferredPaymentMode <= 4.50

| | | | | | |--- class: 0

| | | | | |--- PreferredPaymentMode > 4.50

| | | | | | |--- class: 1

| | | | |--- NumberOfDeviceRegistered > 2.50

| | | | | |--- MaritalStatus <= 1.50

| | | | | | |--- DaySinceLastOrder <= 3.50

| | | | | | | |--- class: 1

| | | | | | |--- DaySinceLastOrder > 3.50

| | | | | | | |--- class: 1

| | | | | |--- MaritalStatus > 1.50

| | | | | | |--- class: 1

| | | |--- OrderAmountHikeFromlastYear > 18.50

| | | | |--- SatisfactionScore <= 4.50

| | | | | |--- class: 0

| | | | |--- SatisfactionScore > 4.50

| | | | | |--- WarehouseToHome <= 7.50

| | | | | | |--- class: 0

| | | | | |--- WarehouseToHome > 7.50

| | | | | | |--- CashbackAmount <= 183.90

| | | | | | | |--- class: 1

| | | | | | |--- CashbackAmount > 183.90

| | | | | | | |--- class: 0

| |--- Complain > 0.50

| | |--- NumberOfAddress <= 3.50

| | | |--- DaySinceLastOrder <= 1.50

| | | | |--- NumberOfDeviceRegistered <= 1.50

| | | | | |--- class: 0

| | | | |--- NumberOfDeviceRegistered > 1.50

| | | | | |--- WarehouseToHome <= 8.50

| | | | | | |--- OrderAmountHikeFromlastYear <= 12.50

| | | | | | | |--- class: 1

| | | | | | |--- OrderAmountHikeFromlastYear > 12.50

| | | | | | | |--- class: 0

| | | | | |--- WarehouseToHome > 8.50

| | | | | | |--- NumberOfAddress <= 1.50

| | | | | | | |--- class: 1

| | | | | | |--- NumberOfAddress > 1.50

| | | | | | | |--- class: 1

| | | |--- DaySinceLastOrder > 1.50

| | | | |--- DaySinceLastOrder <= 5.50

| | | | | |--- MaritalStatus <= 1.50

| | | | | | |--- OrderAmountHikeFromlastYear <= 23.50

| | | | | | | |--- class: 0

| | | | | | |--- OrderAmountHikeFromlastYear > 23.50

| | | | | | | |--- class: 1

| | | | | |--- MaritalStatus > 1.50

| | | | | | |--- PreferredPaymentMode <= 4.50

| | | | | | | |--- class: 0

| | | | | | |--- PreferredPaymentMode > 4.50

| | | | | | | |--- class: 1

| | | | |--- DaySinceLastOrder > 5.50

| | | | | |--- CashbackAmount <= 228.17

| | | | | | |--- WarehouseToHome <= 8.00

| | | | | | | |--- class: 0

| | | | | | |--- WarehouseToHome > 8.00

| | | | | | | |--- class: 1

| | | | | |--- CashbackAmount > 228.17

| | | | | | |--- class: 0

| | |--- NumberOfAddress > 3.50

| | | |--- OrderAmountHikeFromlastYear <= 18.50

| | | | |--- class: 1

| | | |--- OrderAmountHikeFromlastYear > 18.50

| | | | |--- PreferredPaymentMode <= 4.50

| | | | | |--- class: 1

| | | | |--- PreferredPaymentMode > 4.50

| | | | | |--- NumberOfAddress <= 6.50

| | | | | | |--- class: 1

| | | | | |--- NumberOfAddress > 6.50

| | | | | | |--- class: 0

|--- Tenure > 1.50

| |--- CashbackAmount <= 124.38

| | |--- Complain <= 0.50

| | | |--- DaySinceLastOrder <= 0.50

| | | | |--- SatisfactionScore <= 3.50

| | | | | |--- PreferedOrderCat <= 2.50

| | | | | | |--- class: 1

| | | | | |--- PreferedOrderCat > 2.50

| | | | | | |--- CashbackAmount <= 115.59

| | | | | | | |--- class: 0

| | | | | | |--- CashbackAmount > 115.59

| | | | | | | |--- class: 0

| | | | |--- SatisfactionScore > 3.50

| | | | | |--- PreferredLoginDevice <= 1.50

| | | | | | |--- SatisfactionScore <= 4.50

| | | | | | | |--- class: 0

| | | | | | |--- SatisfactionScore > 4.50

| | | | | | | |--- class: 1

| | | | | |--- PreferredLoginDevice > 1.50

| | | | | | |--- SatisfactionScore <= 4.50

| | | | | | | |--- class: 0

| | | | | | |--- SatisfactionScore > 4.50

| | | | | | | |--- class: 0

| | | |--- DaySinceLastOrder > 0.50

| | | | |--- NumberOfDeviceRegistered <= 4.50

| | | | | |--- OrderCount <= 8.00

| | | | | | |--- CashbackAmount <= 120.60

| | | | | | | |--- class: 0

| | | | | | |--- CashbackAmount > 120.60

| | | | | | | |--- class: 0

| | | | | |--- OrderCount > 8.00

| | | | | | |--- OrderCount <= 12.00

| | | | | | | |--- class: 1

| | | | | | |--- OrderCount > 12.00

| | | | | | | |--- class: 0

| | | | |--- NumberOfDeviceRegistered > 4.50

| | | | | |--- OrderAmountHikeFromlastYear <= 13.50

| | | | | | |--- class: 1

| | | | | |--- OrderAmountHikeFromlastYear > 13.50

| | | | | | |--- class: 0

| | |--- Complain > 0.50

| | | |--- SatisfactionScore <= 3.50

| | | | |--- NumberOfAddress <= 4.50

| | | | | |--- CityTier <= 2.00

| | | | | | |--- WarehouseToHome <= 6.50

| | | | | | | |--- class: 0

| | | | | | |--- WarehouseToHome > 6.50

| | | | | | | |--- class: 0

| | | | | |--- CityTier > 2.00

| | | | | | |--- PreferredLoginDevice <= 1.50

| | | | | | | |--- class: 1

| | | | | | |--- PreferredLoginDevice > 1.50

| | | | | | | |--- class: 0

| | | | |--- NumberOfAddress > 4.50

| | | | | |--- PreferredPaymentMode <= 0.50

| | | | | | |--- class: 0

| | | | | |--- PreferredPaymentMode > 0.50

| | | | | | |--- class: 1

| | | |--- SatisfactionScore > 3.50

| | | | |--- PreferedOrderCat <= 2.50

| | | | | |--- class: 0

| | | | |--- PreferedOrderCat > 2.50

| | | | | |--- PreferredLoginDevice <= 1.50

| | | | | | |--- class: 1

| | | | | |--- PreferredLoginDevice > 1.50

| | | | | | |--- CashbackAmount <= 120.72

| | | | | | | |--- class: 0

| | | | | | |--- CashbackAmount > 120.72

| | | | | | | |--- class: 1

| |--- CashbackAmount > 124.38

| | |--- Complain <= 0.50

| | | |--- PreferedOrderCat <= 0.50

| | | | |--- SatisfactionScore <= 2.50

| | | | | |--- WarehouseToHome <= 13.50

| | | | | | |--- PreferredPaymentMode <= 2.50

| | | | | | | |--- class: 0

| | | | | | |--- PreferredPaymentMode > 2.50

| | | | | | | |--- class: 0

| | | | | |--- WarehouseToHome > 13.50

| | | | | | |--- CashbackAmount <= 207.21

| | | | | | | |--- class: 0

| | | | | | |--- CashbackAmount > 207.21

| | | | | | | |--- class: 0

| | | | |--- SatisfactionScore > 2.50

| | | | | |--- NumberOfAddress <= 5.50

| | | | | | |--- class: 0

| | | | | |--- NumberOfAddress > 5.50

| | | | | | |--- CityTier <= 1.50

| | | | | | | |--- class: 0

| | | | | | |--- CityTier > 1.50

| | | | | | | |--- class: 0

| | | |--- PreferedOrderCat > 0.50

| | | | |--- OrderCount <= 6.50

| | | | | |--- DaySinceLastOrder <= 1.50

| | | | | | |--- OrderAmountHikeFromlastYear <= 24.50

| | | | | | | |--- class: 0

| | | | | | |--- OrderAmountHikeFromlastYear > 24.50

| | | | | | | |--- class: 1

| | | | | |--- DaySinceLastOrder > 1.50

| | | | | | |--- NumberOfAddress <= 8.50

| | | | | | | |--- class: 0

| | | | | | |--- NumberOfAddress > 8.50

| | | | | | | |--- class: 0

| | | | |--- OrderCount > 6.50

| | | | | |--- DaySinceLastOrder <= 4.77

| | | | | | |--- OrderAmountHikeFromlastYear <= 15.85

| | | | | | | |--- class: 0

| | | | | | |--- OrderAmountHikeFromlastYear > 15.85

| | | | | | | |--- class: 1

| | | | | |--- DaySinceLastOrder > 4.77

| | | | | | |--- PreferredPaymentMode <= 2.50

| | | | | | | |--- class: 0

| | | | | | |--- PreferredPaymentMode > 2.50

| | | | | | | |--- class: 0

| | |--- Complain > 0.50

| | | |--- SatisfactionScore <= 2.50

| | | | |--- NumberOfAddress <= 10.50

| | | | | |--- CashbackAmount <= 127.98

| | | | | | |--- DaySinceLastOrder <= 1.00

| | | | | | | |--- class: 1

| | | | | | |--- DaySinceLastOrder > 1.00

| | | | | | | |--- class: 0

| | | | | |--- CashbackAmount > 127.98

| | | | | | |--- NumberOfAddress <= 8.50

| | | | | | | |--- class: 0

| | | | | | |--- NumberOfAddress > 8.50

| | | | | | | |--- class: 0

| | | | |--- NumberOfAddress > 10.50

| | | | | |--- DaySinceLastOrder <= 2.00

| | | | | | |--- class: 1

| | | | | |--- DaySinceLastOrder > 2.00

| | | | | | |--- class: 0

| | | |--- SatisfactionScore > 2.50

| | | | |--- CityTier <= 2.50

| | | | | |--- WarehouseToHome <= 35.50

| | | | | | |--- PreferedOrderCat <= 3.50

| | | | | | | |--- class: 0

| | | | | | |--- PreferedOrderCat > 3.50

| | | | | | | |--- class: 0

| | | | | |--- WarehouseToHome > 35.50

| | | | | | |--- class: 1

| | | | |--- CityTier > 2.50

| | | | | |--- WarehouseToHome <= 25.50

| | | | | | |--- MaritalStatus <= 1.50

| | | | | | | |--- class: 0

| | | | | | |--- MaritalStatus > 1.50

| | | | | | | |--- class: 1

| | | | | |--- WarehouseToHome > 25.50

| | | | | | |--- PreferredPaymentMode <= 3.50

| | | | | | | |--- class: 0

| | | | | | |--- PreferredPaymentMode > 3.50

| | | | | | | |--- class: 1

As we increase the depth of the decision tree, we increase accuracy with increasing complexity.

As an example, the most accurate model we get when we set max depth = 15.

precision recall f1-score support

0 0.98 0.98 0.98 941

1 0.91 0.91 0.91 185

accuracy 0.97 1126

macro avg 0.95 0.95 0.95 1126

weighted avg 0.97 0.97 0.97 1126

However, the resultant decision tree is huge and rules become complex.