Bank Churners

IST 707
Applied Machine Learning
Renjini Rajan
Amanda Norwood

This Photo by Unknown Author is licensed under CC BY-NC-ND

Introduction

Time is money! Our dataset focuses on current and past credit card customers

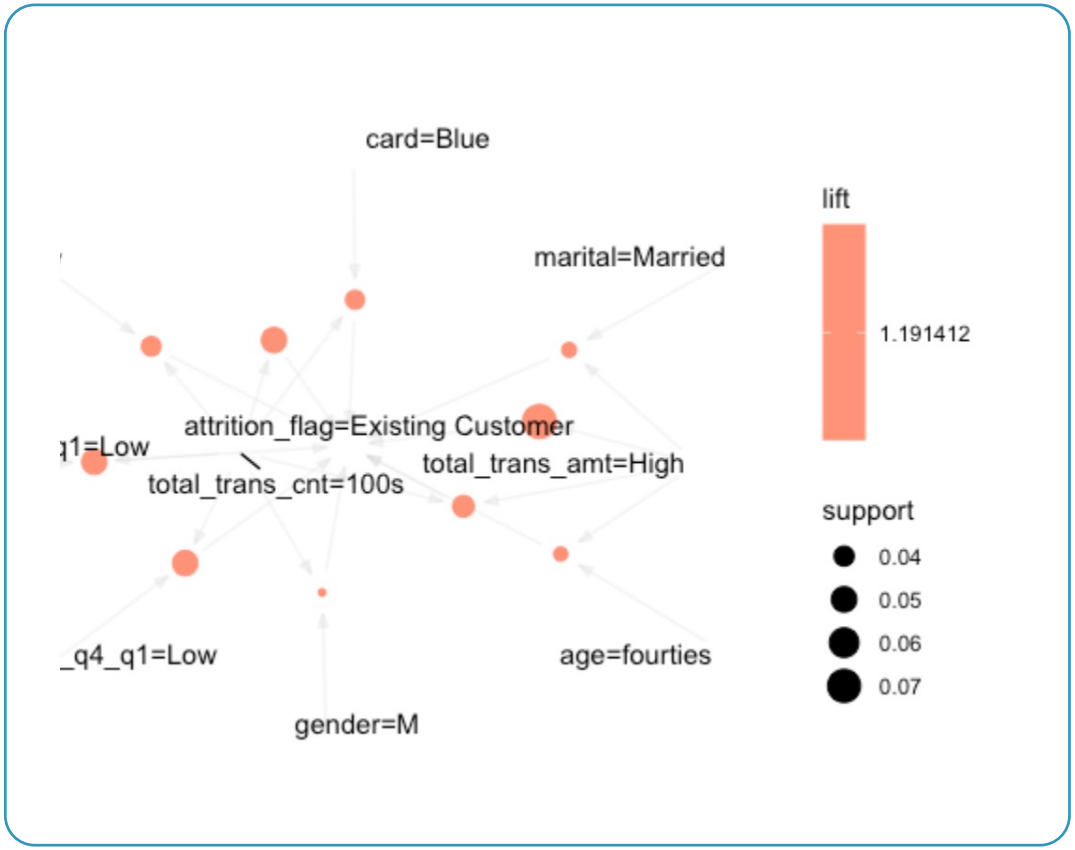
The dataset contains demographic information regarding the customers and their credit card and banking behavior

Goals

1. Identify key factors that influence customer churn
2. Evaluate effectiveness of retention strategies
3. Develop actionable insights

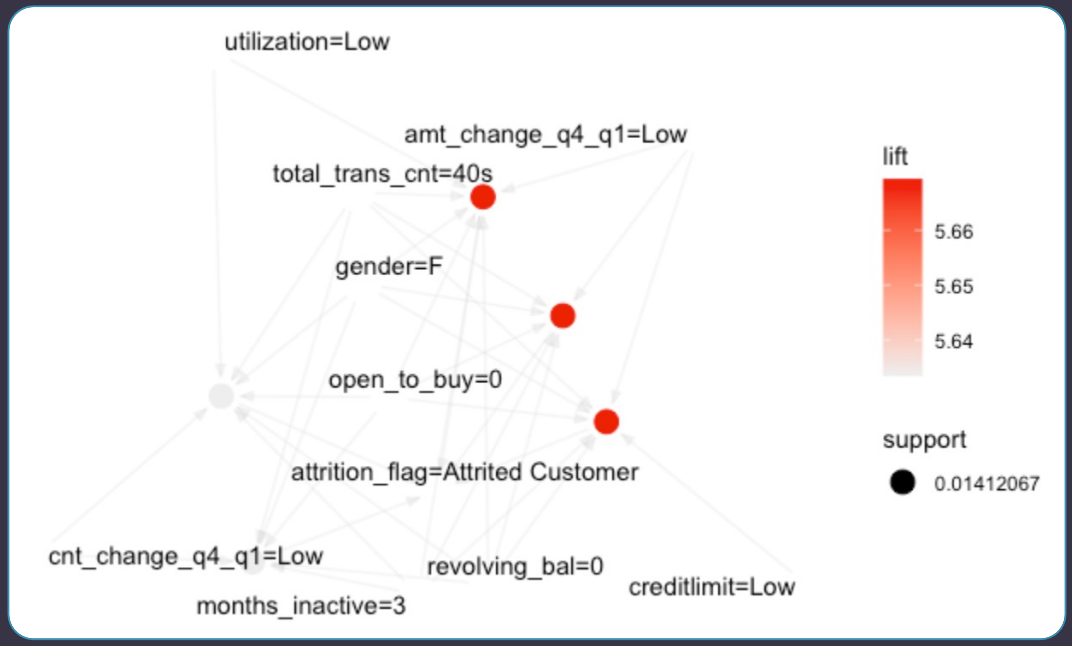
APRIORI: EXISTING CUSTOMERS

- Customers tend to stay with the bank when:
  - Transaction amounts are high
  - Transaction counts are higher (100s)
  - They have a combination of both high amounts and counts
  - Customer age in the 40s
  - Generally male

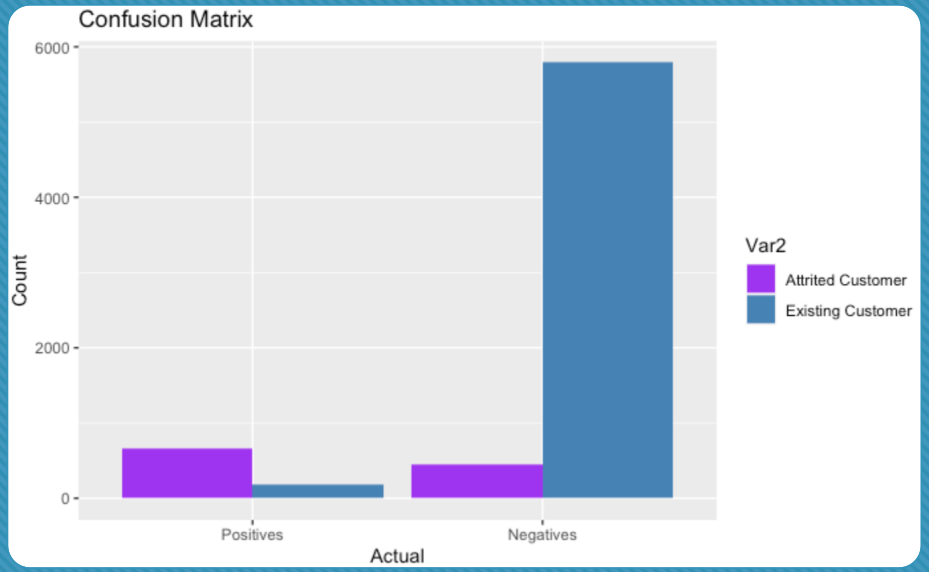


APRIORI: PAST CUSTOMERS

- Customers have less number of contacts/interactions from the bank
- Customers have transaction counts in the 40s
- Customers that are relatively new
- Show no utilization for at least 3 consecutive months

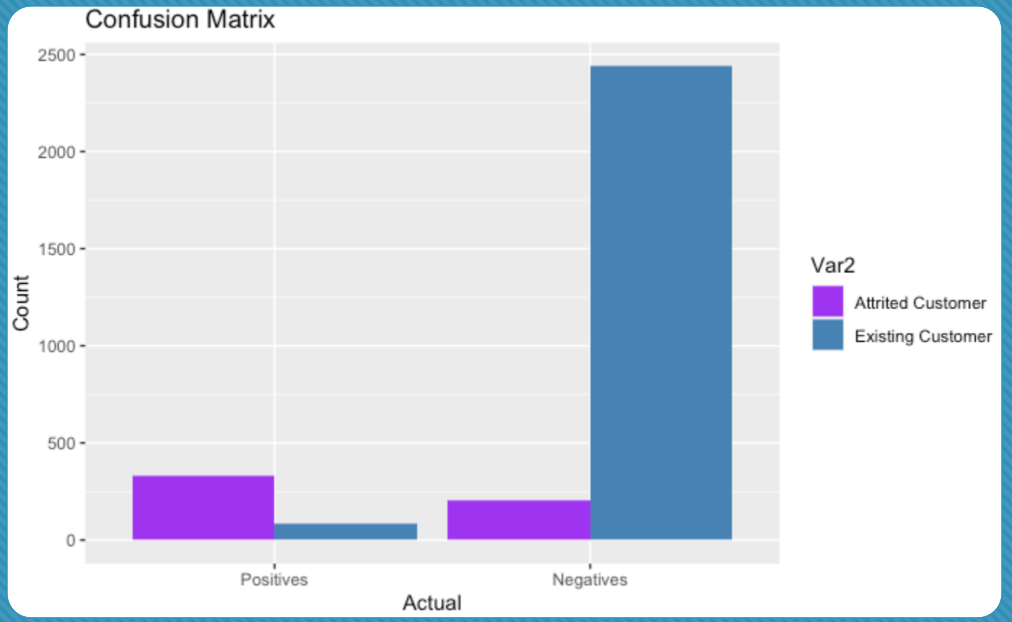


```
Confusion Matrix and Statistics

                  Reference
Prediction          Attrited Customer Existing Customer
  Attrited Customer               668               186
  Existing Customer               442              5792

               Accuracy : 0.9114
                 95% CI : (0.9045, 0.9179)
    No Information Rate : 0.8434
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6298

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.60180
            Specificity : 0.96889
         Pos Pred Value : 0.78220
         Neg Pred Value : 0.92910
             Prevalence : 0.15660
         Detection Rate : 0.09424
   Detection Prevalence : 0.12049
      Balanced Accuracy : 0.78534

       'Positive' Class : Attrited Customer
```



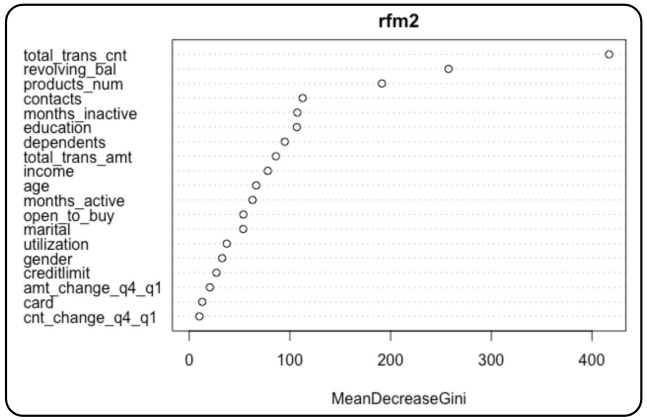
Naïve Bayes Method Train

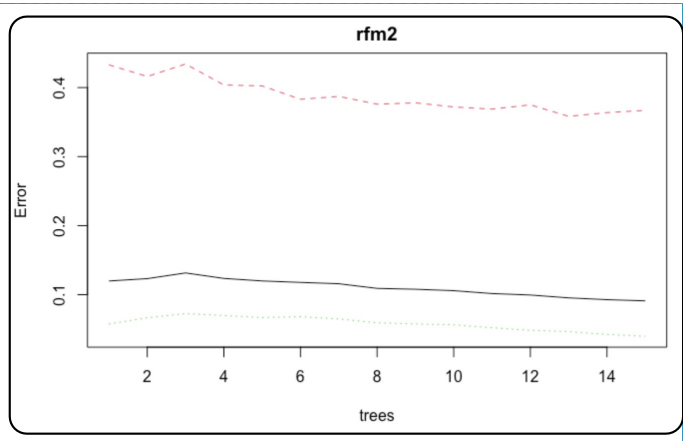
```
Confusion Matrix and Statistics

                 Reference
Prediction          Attrited Customer Existing Customer
  Attrited Customer               311                83
  Existing Customer               206              2439

               Accuracy : 0.9049
                 95% CI : (0.8939, 0.9151)
    No Information Rate : 0.8299
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.628

 Mcnemar's Test P-Value : 7.153e-13

            Sensitivity : 0.6015
            Specificity : 0.9671
         Pos Pred Value : 0.7893
         Neg Pred Value : 0.9221
             Prevalence : 0.1701
         Detection Rate : 0.1023
   Detection Prevalence : 0.1296
      Balanced Accuracy : 0.7843

       'Positive' Class : Attrited Customer
```



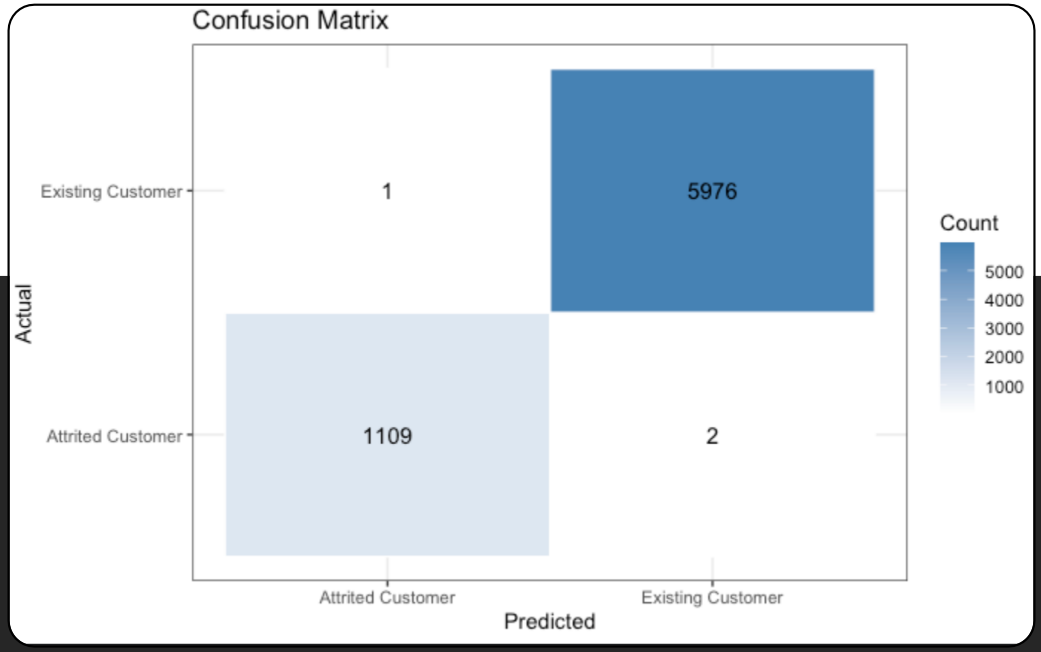
Naive Bayes Method Test

```
Confusion Matrix and Statistics

                 Reference
Prediction          Attrited Customer Existing Customer
  Attrited Customer              1109                1
  Existing Customer                 2             5976

               Accuracy : 0.9996
                 95% CI : (0.9988, 0.9999)
    No Information Rate : 0.8433
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9984

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9982
            Specificity : 0.9998
         Pos Pred Value : 0.9991
         Neg Pred Value : 0.9997
             Prevalence : 0.1567
         Detection Rate : 0.1565
   Detection Prevalence : 0.1566
      Balanced Accuracy : 0.9990

       'Positive' Class : Attrited Customer
```

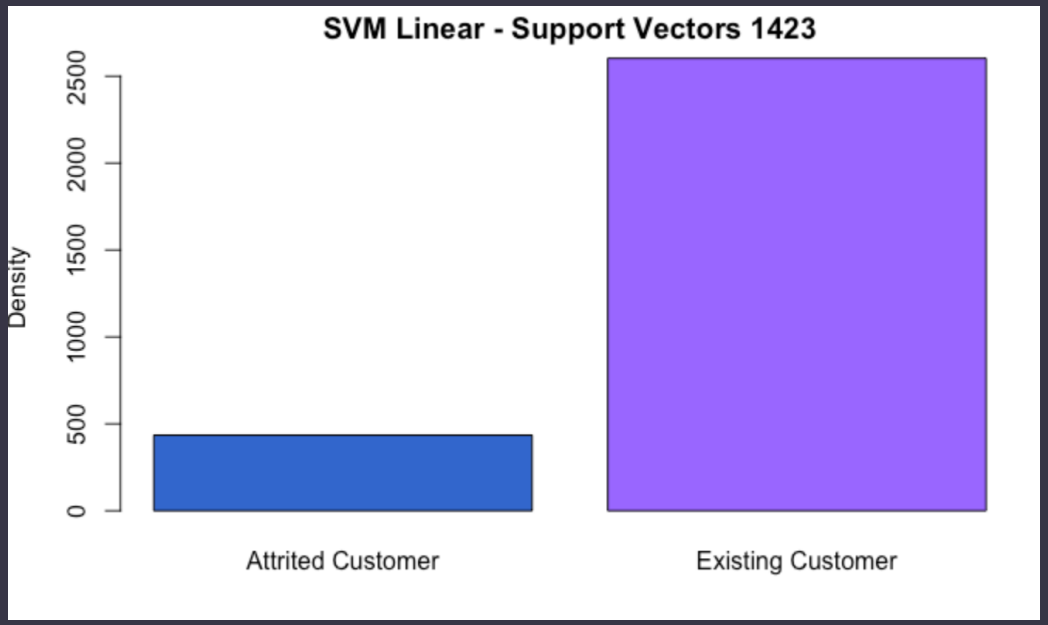




Random Forest

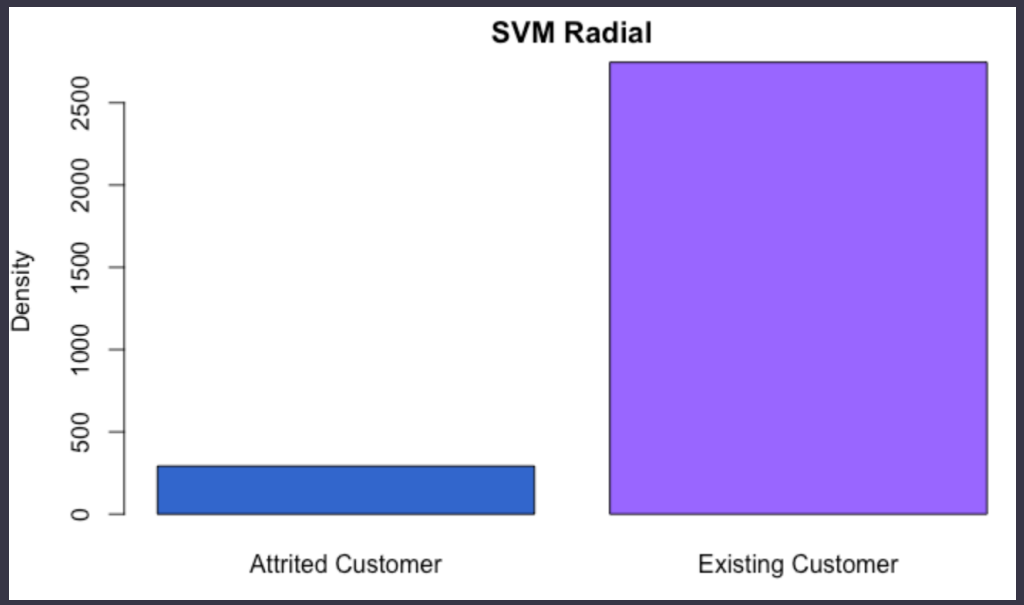


SVM: Linear



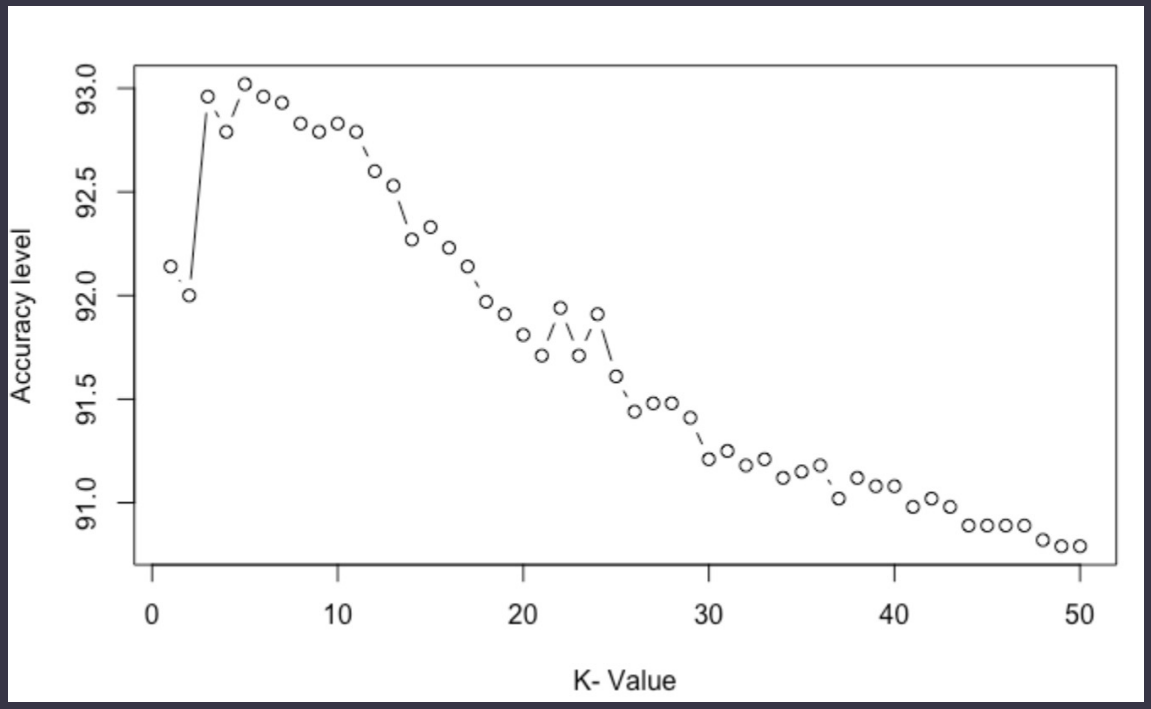

```
Confusion Matrix and Statistics

                  Reference
Prediction          Attrited Customer Existing Customer
  Attrited Customer               346               153
  Existing Customer                89              2450

               Accuracy : 0.9203
                 95% CI : (0.9101, 0.9297)
    No Information Rate : 0.8568
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6941

 Mcnemar's Test P-Value : 5.126e-05

            Sensitivity : 0.7954
            Specificity : 0.9412
         Pos Pred Value : 0.6934
         Neg Pred Value : 0.9649
             Prevalence : 0.1432
         Detection Rate : 0.1139
   Detection Prevalence : 0.1643
      Balanced Accuracy : 0.8683

       'Positive' Class : Attrited Customer
```

SVM: Radial




```
Confusion Matrix and Statistics

                  Reference
Prediction          Attrited Customer Existing Customer
  Attrited Customer               247               252
  Existing Customer                45              2494

               Accuracy : 0.9022
                 95% CI : (0.8911, 0.9126)
    No Information Rate : 0.9039
    P-Value [Acc > NIR] : 0.6353

                  Kappa : 0.5727

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.84589
            Specificity : 0.90823
         Pos Pred Value : 0.49499
         Neg Pred Value : 0.98228
             Prevalence : 0.09612
         Detection Rate : 0.08130
   Detection Prevalence : 0.16425
      Balanced Accuracy : 0.87706

       'Positive' Class : Attrited Customer
```

kNN




```
Confusion Matrix and Statistics

      testknn_labels
knn.84    1    2
     1  170    6
     2  308 2555

               Accuracy : 0.8967
                 95% CI : (0.8853, 0.9073)
    No Information Rate : 0.8427
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4755

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.35565
            Specificity : 0.99766
         Pos Pred Value : 0.96591
         Neg Pred Value : 0.89242
             Prevalence : 0.15729
         Detection Rate : 0.05594
   Detection Prevalence : 0.05791
      Balanced Accuracy : 0.67665

       'Positive' Class : 1
```

Accuracy Comparision between Models

Model	Accuracy	Comments
Naïve Bayes	91.14	
SVM Linear	92.03	
SVM Polynomial	83.57	Cost Tuning did not produce much changes
SVM Radial	90.22	
kNN	93.02	K=5 (From k=1 thru 50 and 84-85)
Random Forest	99.96	15 Trees

Conclusion and Recommendation

- Random Forest provided a higher level of accuracy.
- The Variables of Importance were
    a) Transaction Count
    b) Transactions Amount
    c) Contacts
    d) Revolving Balances
- These are consistent with APRIORI Rules.
- The Marketing Team can focus on the above data points and can reach out to the customers when they see a decline/change in the above variables

Additional Food for thought:

- Data collection can be enhanced by sourcing additional details like
    a) Spending Patterns
    b) Credit Score
    c) Usage of Rewards (if offered/used)/competitor data
- Data is mainly focused on customer base and sometimes external factors such as Inflation, Pandemic etc. can influence decision making process.