

Case 2

Selecting the Best Classification Model Using R

1. Churn Data Introduction

Churn rate is also sometimes called attrition rate. It is one of two primary factors that determine the steady-state level of customers a business will support. In its broadest sense, churn rate is a measure of the number of individuals or items moving into or out of a collection over a specific period of time. This data set contains a total of 3333 mobile subscriber mobile plans. There are a 17 attributes that might affect churn. The classification goal is to predict if the customer will churn(y) or not (n) as well as to identify business rules that can help minimize customer attrition.

1.1. Predictor Variables

- Account Length: length of time in days the customer is using the plan.
- Int'l Plan: plan has an international promo.
- VMail Plan: plan has a voicemail booster.
- VMail Message: number of voice mail messages received
- Day Mins: number of day minutes called (6am – 6pm)
- Day Calls: number of calls made
- Day Charge: total cost of day calls in USD
- Eve Mins: number of eve minutes called (6pm-12 midnight)
- Eve Calls: number of eve calls made
- Eve Charge: total cost of eve calls in USD
- Night Mins: number of night minutes called (12 midnight-6am)
- Night Calls: number of night calls made
- Night Charge: total cost of night calls in USD
- Intl Mins: number of international minutes called
- Intl Calls: number of international calls
- Intl Charge: total cost of international calls
- CustServ Calls: number of calls to call center for service support

2. Modeling

For each of the models here in this section, create the specified model and utilize 10-fold cross validation to fill in the requested information about the model.

2.1. Modeling a Decision Tree

Create a decision tree for the Churn Dataset using the J48 command. Summarize the needed information as follows:

- 2.1.1. Accuracy: _____
- 2.1.2. Confusion Matrix: _____



	Predicted False.	Predicted True.
Actual False.		
Actual True.		

2.1.3. True Positive Rate of Churn=True Class: _____

2.1.4. Precision of Churn=True Class: _____

2.1.5. ROC Area of Churn=True Class: _____

2.2. Creating a Rule Based Classifier

Create a rule based classifier for the Churn Dataset using the JRip command. Summarize the needed information as follows:

2.2.1. Accuracy: _____

2.2.2. Confusion Matrix:

	Predicted False.	Predicted True.
Actual False.		
Actual True.		

2.2.3. True Positive Rate of Churn=True Class: _____

2.2.4. Precision of Churn=True Class: _____

2.2.5. ROC Area of Churn=True Class: _____

2.3. Creating an ANN

Create an ANN classifier for the Churn Dataset using the MLP command. Summarize the needed information as follows:

2.3.1. Accuracy: _____

2.3.2. Confusion Matrix:

	Predicted False.	Predicted True.
Actual False.		
Actual True.		

2.3.3. True Positive Rate of Churn=True Class: _____

2.3.4. Precision of Churn=True Class: _____

2.3.5. ROC Area of Churn=True Class: _____

2.4. Creating an Adaboost Learner with Rule Classifiers

Create an Adaboost Classifier with Rule Based Classifiers for the Churn Dataset using the AdaboostM1 + JRip command. Summarize the needed information as follows:

2.4.1. Accuracy: _____

2.4.2. Confusion Matrix:

	Predicted False.	Predicted True.
Actual False.		



Actual True.		
--------------	--	--

2.4.3. True Positive Rate of Churn=True Class: _____

2.4.4. Precision of Churn=True Class: _____

2.4.5. ROC Area of Churn=True Class: _____

2.5. Creating a Random Forest Model

Create a Random Forest Classifier for the Churn Dataset using the RF command. Summarize the needed information as follows:

2.5.1. Accuracy: _____

2.5.2. Confusion Matrix:

	Predicted False.	Predicted True.
Actual False.		
Actual True.		

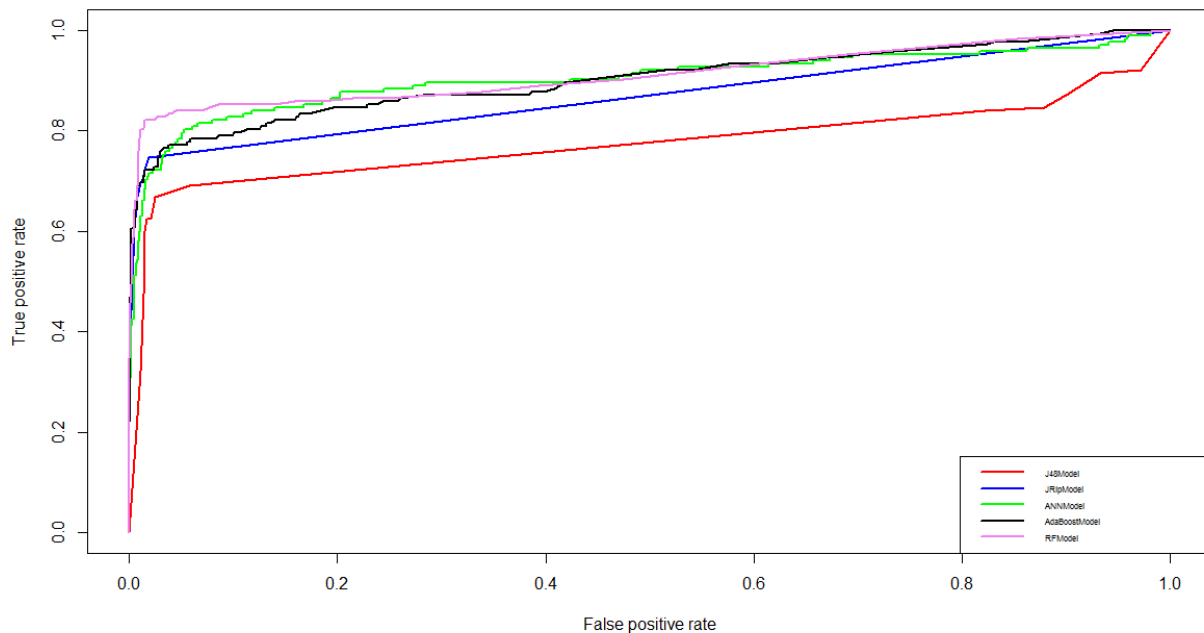
2.5.3. True Positive Rate of Churn=True Class: _____

2.5.4. Precision of Churn=True Class: _____

2.5.5. ROC Area of Churn=True Class: _____

3. ROC Curves

Create an ROC curve for the 5 models using Training and Testing Data. Utilize 67%-33% mix of the data. Choose the Churn = True class. For replicable results, please utilize set.seed(123). The ROC curve should look something like this:



While model will you choose? _____



NATIONAL ENGINEERING CENTER

University of the Philippines
Diliman, Quezon City

