

Group Project Instructions

Learning Outcomes

- Use R perform data mining/visualization and analytics. (CLO 5)
- Describe a modeling strategy for a given data set. (CLOs 2, 3)
- Evaluate the performance of modeling strategy. (CLOs 1, 4, 6)
- Present your modeling strategy and prediction to stakeholders. (CLO 6)

Part A: Description

Churn (loss of customers to competition) is a problem for telecom companies because it is more expensive to acquire a new customer than to keep your existing one from leaving. This assignment is about enabling churn reduction using analytics.

The Business Pain:

Most telecom companies suffer from voluntary churn. Churn rate has strong impact on the life time value of the customer because it affects the length of service and the future revenue of the company. For example if a company has 25% churn rate then the average customer lifetime is 4 years; similarly a company with a churn rate of 50%, has an average customer lifetime of 2 years. It is estimated that 75 percent of the 17 to 20 million subscribers signing up with a new wireless carrier every year are coming from another wireless provider, which means they are churners. Telecom companies spend hundreds of dollars to acquire a new customer and when that customer leaves, the company not only loses the future revenue from that customer but also the resources spend to acquire that customer. Churn erodes profitability.

Steps that have been adopted by telecom companies so far:

Telecom companies have used two approaches to address churn - (a) Untargeted approach and (b) Targeted approach. The untargeted approach relies on superior product and mass advertising to increase brand loyalty and thus retain customers. The targeted approach relies on identifying customers who are likely to churn, and provide suitable intervention to encourage them to stay.

Role of predictive modeling:

In the targeted approach the company tries to identify in advance customers who are likely to churn. The company then targets those customers with special programs or incentives. This approach can bring in huge loss for a company, if churn predictions are inaccurate, because then firms are wasting incentive money on customers who would have stayed anyway.

Part B: Assignment

Objective:

ABC Wireless Inc. has hired you to help them with the customers' churn issue. In this project, you will be working as a part of a team to use historical data from ABC Wireless Inc. to build a model that can predict/identify their customers who are likely to churn. You are free to use any modeling technique/approach from what we have discussed in our course. Also, you are free to include any set of variable available in their dataset for your model.

Project Deliverables:

The following items need to be delivered

1. Project report:

This is your end of project delivery document. It is a document that summarizes different aspects of the project work. It includes the following sections:

- 1 Project Goal
- 2 Overview of data, including data exploration analysis
- 3 Details of your modeling strategy (i.e. what technique and why)
- 4 Estimation of model's performance
- 5 Insights and conclusions

You can include snapshots of your R code and the outputs in the report (recommended). You have to submit a single document per group in PDF format. The first page of the document should include a table with a list of the names of the group participants and a very brief summary of contribution of each team member.

2. R codes and script:

Submit the commented R script file that contains your work.

3. Prediction's File:

The reason to build any model is to be able to use it! In your project, once you have constructed your model, you need to use your model to predict the likelihood of churn for a list of customers that I will provide to you. You will be submitting a file containing your prediction results in terms of the probability of churn for each of those customers. I will use this to judge the accuracy of

your model (I, actually know whether those customers have actually churned or not). Appendix A shows how you can save your R workspace in a file.

4. Presentation Slides:

Finally, you need to prepare a few presentation slides sharing your insights from your project with the board of directors of the ABC Wireless Inc. This should mostly focus on the high level insights as opposed to technical details. As a part of your submission, you should include a narrated PowerPoint slide. You can easily do this by including your voice recordings in the presentations. Click [here](#) for more information and step by step guides for composing narrated presentations.

Submit items 1-4 on the course portal as a single zipped file. One Submission per group is sufficient. The zip file should be named Group_X.zip where X is your group number e.g. Group_10.zip.

Data

The training data ('**Churn_Training.csv**') is available to download from the course portal along with this assignment. There are 19 predictors, mostly numeric:

1. state (categorical),
2. account_length,
3. area_code,
4. international_plan (yes/no),
5. voice_mail_plan (yes/no),
6. number_vmail_messages,
7. total_day_minutes,
8. total_day_calls,
9. total_day_charge,
10. total_eve_minutes,
11. total_eve_calls,
12. total_eve_charge,
13. total_night_minutes,
14. total_night_calls,
15. total_night_charge,

16. total_intl_minutes,
17. total_intl_calls,
18. total_intl_charge
19. number_customer_service_calls.

The variable that you need to predict (target variable) is churn which takes two values 'no' and 'yes'.

Appendix A:

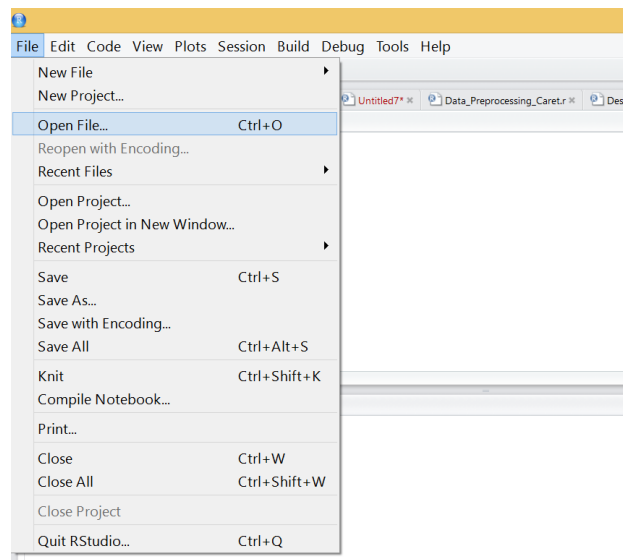
This is a brief visual guide for the steps involved in the project. For the 'churn' variable the order of factors are 'yes' and 'no'. This means the output of your prediction probabilities would corresponds to probability of 'no' (i.e. second factor level) by default. We can change the order of the factor levels using the following line:

```
churnTrain$churn= factor(churnTrain$churn,levels(churnTrain$churn)[c(2,1)])
```

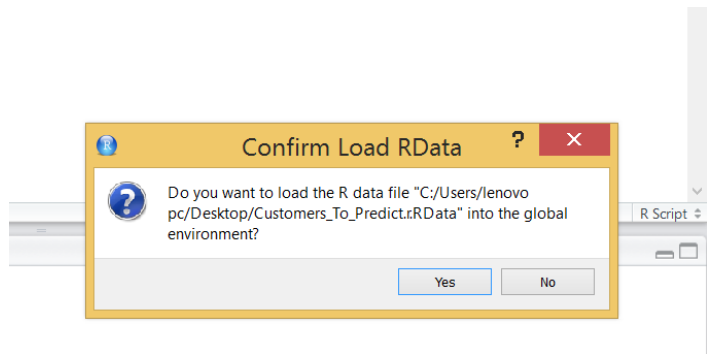
Now you are ready to build your model using the 'churnTrain' dataframe. Now let us assume that you built a model to predict 'churn' using for example logistic regression (other models as can be used well) and the name of the model is 'Model_ABC_Wireless'.

Now we want to do the predictions. Download the file containing the list of consumers that we need to predict their future churn, **Customers_To_Predict.RData**, from the course portal (under group project section). You can download this file by right clicking on the file and save it on your laptop (e.g. desktop to easily find it afterwards).

Now load the data into R studio. Go to the file menu and click on open file:



Click on yes when you are shown the following dialog box:

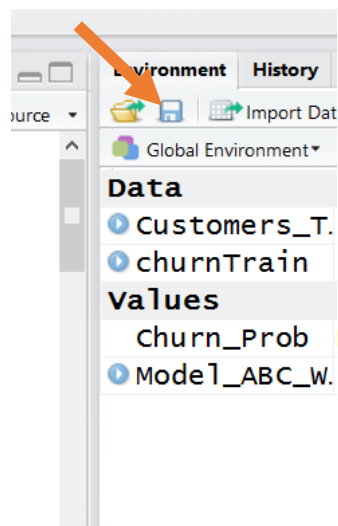


Now you should have a new dataframe called 'Customers_To_Predict' in your dataset. This dataset has 1600 records and all the variables but the target variable, 'churn', which you are trying to predict.

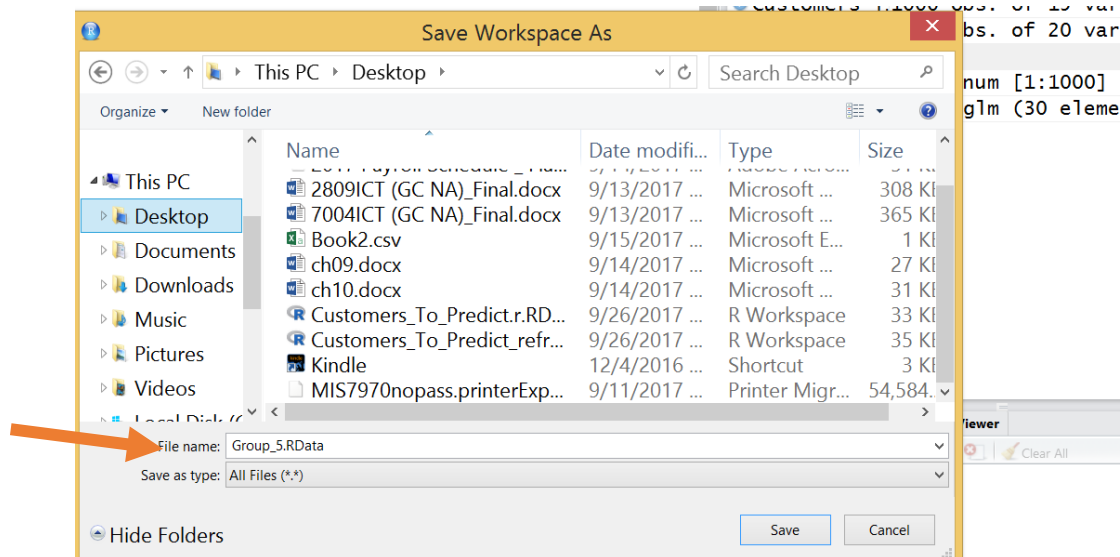
Now, use the appropriate function to do the prediction. You should provide the probability of churn for each customer and save it in a variable called 'Churn_Prob'. In case of logistic regression for example, presuming that you have a model called 'Model_ABC_Wireless', the predict function would be like this

```
Churn_Prob<-predict(Model_ABC_Wireless, newdata=Customers_To_Predict, type = 'response')
```

This create a variable called, 'Churn_Prob' which contains the probabilities of churn for all 1600 customers listed in the Customers_To_Predict datafarm. At this point you are almost done. Just save your R environment into file called Group_x.RData where x is your group number e.g. Group_7.RDadat. See the steps below:



Then name the file and save it on your laptop.



Submit this file together with your R script and the Project Document, all in a single zip file.