OS4106: Project II: Cell Phone Plan Cancellations
Operations Research Michael Atkinson
Naval Postgraduate School

Due date: Thursday, 19 December 2024 by 2300

Our data set comes from a cell phone company. Each observation is a cell phone account. We have some information about the account owner (e.g., age and marital status), the type of account, and phone usage. Based on this information, our task is to predict whether the account has been canceled. This is very useful for the company because if an account has a high probability of canceling, the company would want to engage and entice the account to ensure it does not cancel.

Your objective is to develop a final model to predict whether an account will cancel. Most of your grade will depend upon the process and analysis you go through to determine your final model. Part of your grade will depend upon your final model performance on a test set. The test set provided to you does not contain the response (i.e., whether account has canceled), so you will not know how well you perform on the test set until I reveal the response.

This project is an individual effort. No interactions with others about the project.

1 Rubric

- [**25 points**] Model Selection. What models were considered? How were the models analyzed? How was the final model chosen?
- [**25 points**] Validation. Did you implement comprehensive and correct validation procedures to evaluate your models and avoid overfitting?
- [**10 points**] Clarity. Does your report flow in a cohesive manner? Is it clear exactly what you did, or do I have to guess?
- [**10 points**] Ambition. Did you merely copy and paste from the class scripts and change the name of the data file? Did you attempt anything beyond the "obvious" models?
- [**10 points**] Final model uniqueness. If no one else has the same final model, you receive 10. If one other person has the same final model, you receive 5. If two or more people have the same final model, you receive 0.
- [**20 points**] Final model performance on test set. I sort all individuals based on performance on the test set. Only I have access to the actual test response (`Cancel`). The highest test performance receives the max points, the lowest receives 0 and all other scores are linearly interpolated by rank.

  - [**10 points**] Log-loss performance
  - [**10 points**] Accuracy performance

2 Data Description

Data set contains 1077 observations about cell phone customers and whether they canceled their plan. There are 29 variables

- `CustomerAge` [numeric]: Age of the owner of the cell phone account
- `Married` [categorical]: Marital status of account owner, 3 options: `single, married, divorced`
- `HouseholdSize` [numeric]: Number of individuals living in account owner's home
- `AccountAge` [numeric]: How many months the account has been active
- `PaymentMethod` [categorical] : Payment method for most recent bill, 3 options: `Automatic, Check, Credit`
- `LastNewPhone`[numeric]: How many months since the account owner activated their current phone
- `BasePlan` [categorical]: Account plan, 3 options: `deluxe, standard, economy`
- `IntlPlan` [categorical]: Whether the account includes an international plan, 2 options: `Yes, No`
- `Deal` [categorical]: Whether the account is currently on a deal (or was at the time of cancellation), 2 options: `Yes, No`
- `CustServCall` [numeric]: Number of calls to customer service in last month
- `NumVmail` [numeric]: Number of voicemails received in last month
- `NumText` [numeric]: Number of text messages received in last month
- `NumApps` [numeric]: Number of Apps the user actively used in the last month
- `DayMin` [numeric]: Number of day minutes used (in phone calls) in the last month
- `DayCall` [numeric]: Number of day phone calls made in the last month
- `DayData` [numeric]: Amount of day data used in the last month
- `DayCharge` [numeric]: Charges in the last month due to day phone usage

- `EveMin` [numeric]: Number of evening minutes used (in phone calls) in the last month
- `EveCall` [numeric]: Number of evening phone calls made in the last month
- `EveData` [numeric]: Amount of evening data used in the last month
- `EveCharge` [numeric]: Charges in the last month due to evening phone usage
- `NightMin` [numeric]: Number of night minutes used (in phone calls) in the last month
- `NightCall` [numeric]: Number of night phone calls made in the last month
- `NightData` [numeric]: Amount of night data used in the last month
- `NightCharge` [numeric]: Charges in the last month due to night phone usage
- `IntlMin` [numeric]: Number of international minutes used (in phone calls) in the last month
- `IntlCall` [numeric]: Number of international phone calls made in the last month
- `intlCharge` [numeric]: Charges in the last month due to international phone usage
- `Cancel` [categorical]: The *response*: whether the account has been canceled, 2 options: `Yes, No`

3 Submission

You must turn in 2 items

(a) Report summarizing your analysis and results. The page limit is 2. I will not read anything beyond 2 pages. The page limits includes, figures, tables, R output etc.

(b) A `CSV` file named `predictions.csv` that contains your final model predictions on the test set. This file has two columns, named `Probability` and `Label`. The number of rows in `predictions.csv` equals the number of rows in the test set. Each row of `predictions.csv` contains the prediction for the corresponding row in the test set. You must include a prediction for two quantities: the probability the plan cancels and the class label (`Yes/No`). The function `write.csv` may be useful in creating `predictions.csv`.

I provide a template on Sakai for what `predictions.csv` should look like.

You are welcome to submit supplementary material: R-scripts, figure, appendices etc. I will not look at them initially. However, if there is a grade dispute, we can look at those files.