Data Science Mini-Project

GEICO provides a number of ways for its policyholders to make payments. While our service counselors can take payments over the phone, it is more cost-efficient for customers to make payments through our self-service channels such as online or through the automated phone system. We would like to use a predictive model to select people to receive a pre-emptive e-mail message designed to encourage them to pay online.

You have been tasked with identifying which customers are likely to make a service payment call in the next 5 days. The attached file contains data on customers who have had a bill due in the next 5 days and whether they made a service payment call. Construct a model that predicts the likelihood that each policyholder will make a service payment call (CALL\_FLAG=1). You may use whatever methods you see fit.

You may use any modeling package you choose to complete your analysis. Include any relevant work flow, code, or scripts you used. Summarize your data preparation, variable selection, modeling methods and fit statistics into a presentation. You should, at a minimum, describe how you selected model inputs, any data transformations preformed, how you dealt with missing or extreme values, reasons for your selected algorithm, and how your analysis addresses the business problem.

Please submit your code and presentation within 120 hours of receiving this assignment.

Thank you, and have fun!

GEICO Data Science Team

# Data Dictionary:

|  |  |
| --- | --- |
| **Var Name** | **Description** |
| DATE\_FOR | Date of Record Processing |
| RTD\_ST\_CD | Rated State of Policy |
| CustomerSegment | Segment the Customer falls into |
| Tenure | Years of Tenure with Company |
| Age | Age of Policyholder |
| MART\_STATUS | Marital Status of Policyholder |
| GENDER | Gender Of Primary Insured |
| CHANNEL1\_6M | # payments made through Channel 1 in last 6 months |
| CHANNEL2\_6M | # payments made through Channel 2 in last 6 months |
| CHANNEL3\_6M | # payments made through Channel 3 in last 6 months |
| CHANNEL4\_6M | # payments made through Channel 4 in last 6 months |
| CHANNEL5\_6M | # payments made through Channel 5 in last 6 months |
| METHOD1\_6M | # of payment made with method 1 (irrespective of channel) |
| RECENT\_PAYMENT | Payment made in last 15 days (1/0) |
| PAYMENTS\_6M | # of total payments in last 6 months |
| CHANNEL1\_3M | # payments made through Channel 1 in last 3 months |
| CHANNEL2\_3M | # payments made through Channel 2 in last 3 months |
| CHANNEL3\_3M | # payments made through Channel 3 in last 3 months |
| CHANNEL4\_3M | # payments made through Channel 4 in last 3 months |
| CHANNEL5\_3M | # payments made through Channel 5 in last 3 months |
| METHOD1\_3M | # of payment made with method 1 (irrespective of channel) |
| PAYMENTS\_3M | # of total payments in last 3 months |
| NOT\_DI\_3M | Had this customer been enrolled in automated payments in the last 3 months? 1/0 |
| NOT\_DI\_6M | Had this customer been enrolled in automated payments in the last 6 months? 1/0 |
| EVENT1\_30\_FLAG | Has this customer been sent a cancellation notice in the last 30 days? 1/0 |
| EVENT2\_90\_SUM | How many cancellation notices have been sent in the last 90 days? |
| LOGINS | How many times has this policy logged into self-service online in the last 30 days? |
| POLICYPURCHASECHANNEL | How was this policy purchased? 1/0 |
| Call\_Flag | Was there a service payment 1/0? **TARGET VARIABLE** |