Jaclyn Coate, Josh Eysenbach, Branum Stephan

Abstract

A basic logistic regression analysis to predict a bank customer’s subscription and additional competing complex models for best prediction results.

Subscriptions

Table of Contents

[1. Introduction 4](#_Toc37267915)

[Banking institutions are one of the largest in the world. They complete millions of transactions daily in person, digitally, and over the phone. 4](#_Toc37267916)

[2. Data description 4](#_Toc37267917)

[The data we used was provided by a. 4](#_Toc37267918)

[Variable Name 4](#_Toc37267919)

[Variable Description 4](#_Toc37267920)

[3. Exploratory data analysis (EDA) 4](#_Toc37267921)

[Upon first look we reviewed the 4](#_Toc37267922)

[Next we decided to drop logically irrelevant variables. 4](#_Toc37267923)

[Our next round of variables that were removed based on usability were 4](#_Toc37267924)

[The next few steps in our EDA were to ensure things like a clean data set, outlier detection and multicollinearity. We did a ‘zero’ value check on variables where that does not make sense 4](#_Toc37267925)

[Lastly, we checked for signs of linearity between our remaining two categorical variables and our dependent variable 4](#_Toc37267926)

[Our EDA complete we begin our model building for a simple logistic regression. 5](#_Toc37267927)

[4. Objective 1: Simple logistic regression model 5](#_Toc37267928)

[A. Problem Statement and Approach Review 5](#_Toc37267929)

[B. Model Selection 5](#_Toc37267930)

[i. Type of Selection 5](#_Toc37267931)

[ii. Checking Assumptions 5](#_Toc37267932)

[C. Parameter Interpretation 5](#_Toc37267933)

[i. Interpretation 5](#_Toc37267934)

[ii. Confidence Intervals 5](#_Toc37267935)

[D. Conclusion 5](#_Toc37267936)

[5. Objective 2: competing models 5](#_Toc37267937)

[A. Model Selection 5](#_Toc37267938)

[i. Make sure it is clear how many models were created to compete against the one in Objective 1. Make note of any tuning parameters that were used and how you came up with them (knn and random forest). 5](#_Toc37267939)

[B. Main Analysis Content 5](#_Toc37267940)

[i. Overall report of the error metrics on a test set or a CV run. Also, if the two best models have error rates of .05 and .045, can we really say that one model is outperforming the other? For the ambition, McNear’s test could be helpful in answering that. 5](#_Toc37267941)

[C. Conclusion / Discussion 5](#_Toc37267942)

[i. The conclusion should reprise the question and conclusions of objective 2 with recommendations of the final model, what could be done to help analysis and model building in the future and any insight as to why one method outshined all the rest if that is indeed the case. If they are all similar why did you go with your final model? 5](#_Toc37267943)

[6. Appendix 6](#_Toc37267944)

[A. Figure 1 6](#_Toc37267945)

[B. Figure 2 6](#_Toc37267946)

[C. Figure 3 6](#_Toc37267947)

[D. Figure 4 6](#_Toc37267948)

[E. Figure 5 6](#_Toc37267949)

[F. Figure 6 6](#_Toc37267950)

[G. Figure 7 6](#_Toc37267951)

[H. Figure 8 6](#_Toc37267952)

[I. Figure 9 6](#_Toc37267953)

[J. Figure 10 6](#_Toc37267954)

[K. Figure 11 6](#_Toc37267955)

[L. Figure 12 6](#_Toc37267956)

[M. Figure 13 6](#_Toc37267957)

[N. Figure 14 6](#_Toc37267958)

[O. Figure 15 6](#_Toc37267959)

[P. Figure 16 6](#_Toc37267960)

[Q. Figure 17 6](#_Toc37267961)

[R. Figure 18 6](#_Toc37267962)

[S. Figure 19 6](#_Toc37267963)

[T. Figure 20 6](#_Toc37267964)

[U. Figure 21 6](#_Toc37267965)

[V. Figure 22 6](#_Toc37267966)

[W. Figure 23 6](#_Toc37267967)

[X. Figure 24 6](#_Toc37267968)

[Y. Figure 25 6](#_Toc37267969)

[Z. Figure 26 6](#_Toc37267970)

[AA. Figure 27 6](#_Toc37267971)

[BB. Figure 28 6](#_Toc37267972)

[CC. Figure 29 6](#_Toc37267973)

[DD. Figure 30 6](#_Toc37267974)

Subscriptions

# Introduction

## Banking institutions are one of the largest in the world. They complete millions of transactions daily in person, digitally, and over the phone.

# Data description

## The data we used was provided by a.

|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Exploratory data analysis (EDA)

## Upon first look we reviewed the

## Next we decided to drop logically irrelevant variables.

## Our next round of variables that were removed based on usability were

## The next few steps in our EDA were to ensure things like a clean data set, outlier detection and multicollinearity. We did a ‘zero’ value check on variables where that does not make sense

## Lastly, we checked for signs of linearity between our remaining two categorical variables and our dependent variable

## Our EDA complete we begin our model building for a simple logistic regression.

# Objective 1: Simple logistic regression model

## Problem Statement and Approach Review

## Model Selection

### Type of Selection

#### Any or all: LASSO, Ridge, Elastic Net, Stepwise, Forward, Backward, Manual/Intuition

### Checking Assumptions

#### Lack of Fit Test

#### Influential Point Analysis

#### Residual Plots **(optional)**

## Parameter Interpretation

### Interpretation

### Confidence Intervals

## Conclusion

# Objective 2: competing models

## Model Selection

### Make sure it is clear how many models were created to compete against the one in Objective 1. Make note of any tuning parameters that were used and how you came up with them (knn and random forest).

## Main Analysis Content

### Overall report of the error metrics on a test set or a CV run. Also, if the two best models have error rates of .05 and .045, can we really say that one model is outperforming the other? For the ambition, McNear’s test could be helpful in answering that.

## Conclusion / Discussion

### The conclusion should reprise the question and conclusions of objective 2 with recommendations of the final model, what could be done to help analysis and model building in the future and any insight as to why one method outshined all the rest if that is indeed the case. If they are all similar why did you go with your final model?

# Appendix

## Figure 1

## Figure 2

## Figure 3

## Figure 4

## Figure 5

## Figure 6

## Figure 7

## Figure 8

## Figure 9

## Figure 10

## Figure 11

## Figure 12

## Figure 13

## Figure 14

## Figure 15

## Figure 16

## Figure 17

## Figure 18

## Figure 19

## Figure 20

## Figure 21

## Figure 22

## Figure 23

## Figure 24

## Figure 25

## Figure 26

## Figure 27

## Figure 28

## Figure 29

## Figure 30