A Logistic Regression Qpproach to CoIL Challenge 2000

Corey Arnouts[1], Adam Douglas[1], Jason Givens-Doyle[1], & Michael Silva[1]

[1] MS in Data Science Students CUNY School of Professional Studies

Author Note

Correspondence concerning this article should be addressed to Corey Arnouts, 119 W 31st St., New York, NY 10001. E-mail: Corey.Arnouts@spsmail.cuny.edu

Abstract

A logistic regression based solution to the CoIL Challenge 2000 is described. The challenge consists of correctly identifying potential customers for an insurance product, and describing their characteristics.

Keywords: CoIL Challenge, Logistic Regression

Word count: X

A Logistic Regression Qpproach to CoIL Challenge 2000

Introduction

Businesses use data science to extract insights from data. It has many practical business applications. Identifying households to include in a marketing campaign is one application. One example using real world data is the Computational Intelligence and Learning (CoIL) Challenge. The CoIL Challenge competition was held from March 17 to May 8 in 2000. The challenge is to:

1. Identify potential customers for an insurance policy; and

2. Provide a description of this customer base.

In total 147 participants registered and 43 submitted solutions (Putten, Ruiter, & Someren, 2000). In this paper we set out to complete the first part of the COIL Challenge. **SUMARISE FINDINGS?**

Literature Review

Researchers used a variety of approaches in formulating their submissions including: Artificial Neural Networks (Kim & Street, 2000), Naïve Bayes (Elkan, 2000).

Methodology

Experimentation and Results

Discussion and Conclusions

The maximum number of policyowners that could be found was 238, the winning model selected 121 policy owners. Random selection results in 42 policy owners. The

standard benchmark tests result in 94 (k-nearest neighbor), 102 (naïve bayes), 105 (neural networks)and 118 (linear!) policy owners. (Putten et al., 2000)

References

Elkan, C. (2000). CoIL challenge 2000 entry. Retrieved from

http://liacs.leidenuniv.nl/~puttenpwhvander/library/cc2000/ELKANP~1.pdf

Kim, Y., & Street, W. N. (2000). CoIL challenge 2000: Choosing and explaining likely

caravan insurance customers. Retrieved from

http://liacs.leidenuniv.nl/~puttenpwhvander/library/cc2000/STREET~1.pdf

Putten, P., Ruiter, M., & Someren, M. (2000). CoIL challenge 2000 tasks and results:

Predicting and explaining caravan policy ownership.

Appendices

- Supplemental tables and/or figures.

R statistical programming code.