Car Insurance Claims Classification Report

Hamed Vaheb

hamed.vaheb.001@student.uni.lu

January 12, 2023

Abstract

This report describes the machine learning project which aims at analysis and prediction of the car insurance claims dataset. Based on historical data of the clients, the variable of interest is the outcome of insurance, which indicates whether a customer has claimed his loan or not. The `aiinsurance` R package is developed to make this work reproducible, accessible, and equipped with advanced features, such as a pipeline that performs all steps in a single command, and an interactive app to display the performance of the models. Using package's functions, two machine learning models, i.e., logistic regression and random forest is implemented to predict the outcomes using the historical data. The results indicate promising predictions with various evaluation metrics (e.g., accuracy, precision, recall) of over 80 percent. Besides, prediction, informative insights from the dataset and models have been drawn. For instance, by dint of extracting feature importance from the random forest model, the parts of the dataset that play effective role in the prediction become evidenced.

1 Introduction

Rapid advances in artificial intelligence (AI) and machine learning are creating products and services with the potential not only to change the environment in which actuaries operate but also to provide new opportunities within actuarial science [1].

The use of statistical learning models has been a common practice in actuarial science since the 1980s. It was not long after since the field adopted classical models, such as linear models and generalized linear models (GLMs). While actuaries use GLMs frequently in practice, it was in the past few years that the use of AI and machine learning, and hence more modern models garnered significant attention in the field [2].

In this work, both a classical model and a modern model is used and compared for predicting clients' claims. The former is logistic regression, which is an example of GLMs, accommodated to classification setting, i.e., for predicting classes that in our case are "outcome" of insurance that indicates whether a customer has claimed his loan or not

The remainder of this work is organized as the following: In ...

[3]

2 Preliminary Concepts

2.1 Random Forest

Recent models learn nonlinear transformations and interactions between variables from the data without manually specifying them. This is performed implicitly with tree-based models, e.g., random forest.

5 Headings: first level

LaTeX command can be used to reference other section. See Section 5. However, you can also use **bookdown** extensions mechanism for this.

5.1 Headings: second level

You can use equation in blocks

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j|y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}$$

But also inline i.e $z = x + y$

5.1.1 Headings: third level

Another paragraph.

6 Examples of citations, figures, tables, references

You can insert references. Here is some text (**kour2014real?**; **kour2014fast?**) and see (**hadash2018estimate?**).

The documentation for `natbib` may be found at

You can use custom blocks with LaTeX support from **rmarkdown** to create environment.

http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf%7D

Of note is the command `\citet`, which produces citations appropriate for use in inline text.

You can insert LaTeX environment directly too.

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated...

https://www.ctan.org/pkg/booktabs

6.1 Figures

You can insert figure using LaTeX directly.

See Figure 1. Here is how you add footnotes. [^Sample of the first footnote.]

But you can also do that using R.

```
plot(mtcars$mpg)
```

You can use **bookdown** to allow references for Tables and Figures.

6.2 Tables

Below we can see how to use tables.

See awesome Table~1 which is written directly in LaTeX in source Rmd file.

You can also use R code for that.

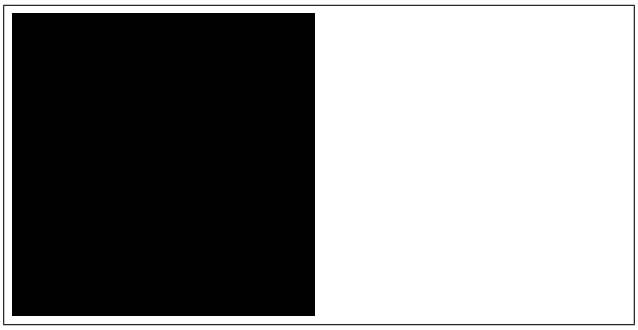

Figure 1: Sample figure caption.

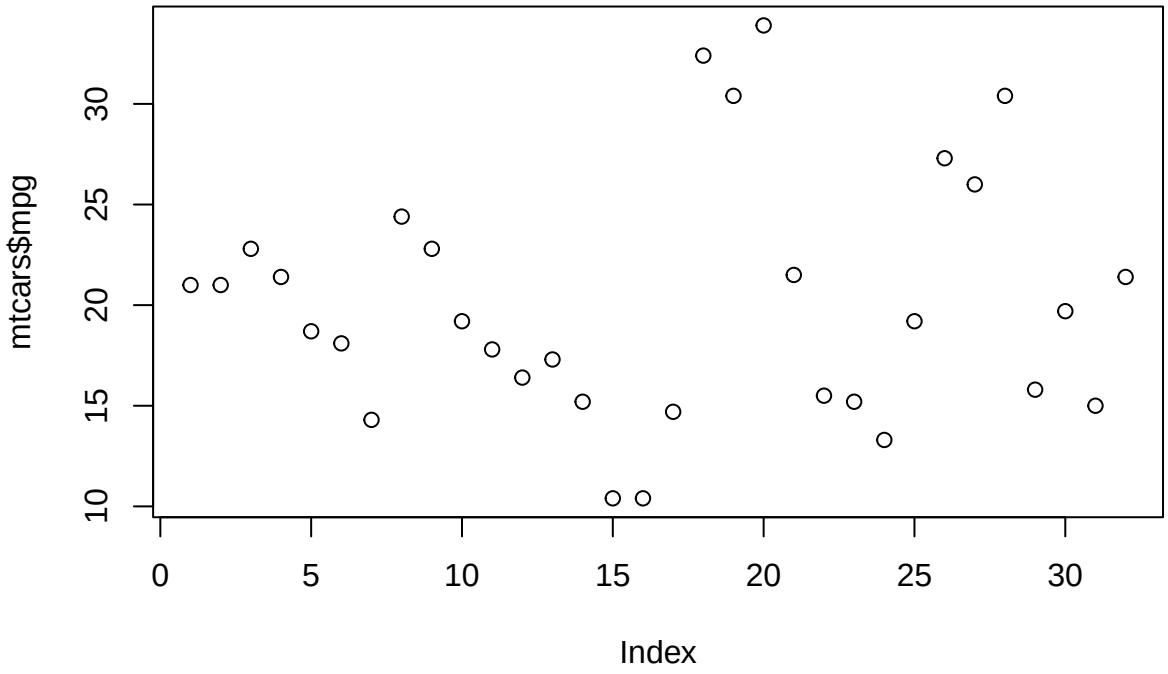

Figure 2: Another sample figure

Table 1: Sample table title

Part		
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

```
knitr::kable(head(mtcars), caption = "Head of mtcars table")
```

Table 2: Head of mtcars table

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

6.3 Lists

- Item 1
- Item 2
- Item 3

References

[1] Ronald Richman. Ai in actuarial science – a review of recent advances – part 1. *Annals of Actuarial Science*, 15(2):207–229, 2021.

[2] Christopher Blier-Wong, Hélène Cossette, Luc Lamontagne, and Etienne Marceau. Machine learning in p&c insurance: A review for pricing and reserving. *Risks*, 9(1), 2021.

[3] Hamed Vaheb. AI Insurance Package. Available at http://github.com/berserkhmdvhb/aiinsurance.