

Jeff Maxey

Predictive Analytics

Theory of and Use Cases in Predictive Analytics Using R







— |

To the most amazing cooks in my life, Xie Shaobai and Si Zhinan.

—Yihui

To my supporting wife, Caroline, and our lovely newborn, Axel.

—Christophe

To my mom, who taught me the joy of life-long learning.

—Emily

— |



Contents



List of Figures



List of Tables

```
source("_common.R")
```

```
## -- Attaching core tidyverse packages -----
tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.1      v tibble     3.2.1
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.0.4
## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter()      masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()         masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```



Preface



Part I

Introduction to Predictive Analytics



O

Overview of Predictive Analytics

Introduction

This foundational chapter streamlines the material scattered in different parts of the PA e-learning modules and provides a broad and coherent introduction to predictive analytics.

Setting the scene for pretty much everything that comes later, it walks you through the main steps in building a predictive model and presents the fundamental concepts in predictive analytics that you will see repeatedly in the rest of this manual.

These concepts are universally applicable in the sense that they apply to essentially all types of predictive model, and will be illustrated in the context of specific types of model (GLMs and decision trees, in particular) in later chapters.

0.1 Basic Terminology

Predictive Analytics in a nutshell. Three main categories of predictive analytics problems:

TABLE 0.1: TABLE 1: Categories of Predictive Analytics Problems

Type	Definition	Example(s)
Descriptive	<ul style="list-style-type: none"> Focuses on what happened in the <i>past</i> and aims to “describe” or explain the observed patterns by identifying relationships between different variables in the data. 	<ul style="list-style-type: none"> If you saw an increase in the lapse rate among the policyholders of a certain line of business, what kind of policyholders had the highest tendency to lapse? What are their key characteristics? This is a question addressed by descriptive analytics.

Type	Definition	Example(s)
Predictive	<ul style="list-style-type: none"> Focuses on what will happen in the <i>future</i> and is concerned with making accurate “predictions”. 	<ul style="list-style-type: none"> For a prospective policyholder with certain characteristics, what is their predicted probability of lapse? The ability to make such a prediction will be useful for identifying future policyholders who will have a lower probability of lapse and contribute to the profitability of an insurer.
Prescriptive	<ul style="list-style-type: none"> Combination of <i>optimization and simulation</i> to quantify the impact of different “prescribed” actions in different scenarios. 	<ul style="list-style-type: none"> If we reduce the premium by a certain amount, how will this affect the lapse rate? More generally, what is the best course of action to reduce the lapse rate? Prescriptive analytics may give us some useful insight.