

Summer 2020 Project

ONLINE BOOK PROJECT ON PREDICTIVE ANALYTICS AND DATA SCIENCE

Supervisors

- Daniel Linders:
 - o Assistant professor in Actuarial Science, University of Illinois.
 - o www.daniellinders.com
- Haoen Cui:
 - o Associate Data Scientist at Uptake, Chicago.
 - o haoen-cui.github.io

Introduction

Computer programming, predictive analytics, and data science are among the hot skillsets that the industry is chasing after. As a result, students need to become proficient in these domains to build a competitive edge. Though there are many existing courses and books on statistical learning, applying such academic knowledge to practice is seldomly straightforward. For example, data is noisy, incomplete, or even missing, which requires extra work than merely fitting a textbook model. Moreover, unstructured data, such as text and image, has shown its potential in many critical applications; hence new creative solutions are required to work with these special data types.

The “*online book project on predictive analytics*” is a larger-scale project that aims to develop a toolbox for practical issues related to data science and predictive analytics as it pertains to actuarial and financial applications. It should, in the first place, serve as an online database of supplementary material for the course “ASRM 499/552: Predictive analytics.” However, we hope to develop it so that it is useful for a broader community. The online book will have 3 main parts:

- Part 1: Theory.
 - o A practical primer on statistical learning.
 - o Simple examples to illustrate deep ideas.
- Part 2: Applications. Case studies to...
 - o illustrate different statistical learning techniques;
 - o illustrate how to solve practical problems in actuarial science and finance using statistical learning.
- Part 3: Beyond numeric data (the following is not an exhaustive list)
 - o Categorical and ordinal variables;
 - o Text data and natural language processing;
 - o Image recognition.

Project description

In this project, you are asked to contribute a chapter (or multiple) for the online book. Since this project is still in its infancy, you can either start with a chapter we suggest or develop your own ideas. A chapter should have a concrete question to be answered and a practical solution for this question along with necessary theoretical justifications. We emphasize precise and elegant academic writing and encourage creativity, as you find appropriate. Below are a few example topics:

- Generalized Linear Models: applications in finance and insurance.
 - o This CAS monograph may be helpful:
<<https://www.casact.org/pubs/monographs/index.cfm?fa=goldburd-Khare-Tevet-monograph05>>
- K Nearest Neighbor: theory and applications.
 - o Use kNN to illustrate bias-variance tradeoff
 - o Interesting regression / classification applications
- On the use/abuse of p-values.
- Gauss process regression for life insurance pricing.

Requirements

- Enthusiastic, creative, and persistent.
- Able to work independently and reach out for help when needed.
- Proficient in R and RMarkdown.
- Strong and creative academic writing skills.
- Nice to have: Knowledge on Git, GitHub, and R Shiny.

How to apply?

- Send the following information to dlinders@illinois.edu before **May 26**.
 - o Name, email, major and past experience in the IRisk Lab.
 - o CV or resume.
 - o Motivation letter of max 1 page, which also serves as a writing test.
 - o Feel free to deliver other material you feel could help your application.
- A decision is expected before May 29 and we will kick off the project shortly after.

Why apply?

- Improve your coding skills.
- Improve your writing skills.
- Your name will be listed as author for your chapter on the website, unless you prefer otherwise.
- You the option, but not the obligation, to register for ASRM 390/490.