

40.016 The Analytics Edge

Brief Introduction to the caret package

Stefano Galelli

SUTD

Term 6, 2020

Introduction

caret stands for **C**lassification **A**nd **R**Egression **T**raining. The package contains functionalities for:

- Data splitting
- Pre-processing
- Variable selection
- Model tuning (calibration, validation, testing)
- Variable importance estimation

Note (for Python users): comparable to the Scikit-learn library (but keep in mind R and Python are two different languages!).

How to install it?

- It is hosted on CRAN, so it can be installed like any other R package.
- **Note:** it consolidates, under the same umbrella, many R packages for statistical learning, so it takes time to install it!

Useful resources

- Online guide
- Article on the Journal of Statistical Software
- Webinar
- Many, many, many articles on blogs ...

So, why don't we use it in The Analytics Edge?

caret makes many packages available at our fingertips, but:

- We first need to learn the basics of statistical learning (before experimenting with different techniques)
- It can be overwhelming (too many packages, if we don't know which one to use)

This said, caret can become a useful tool as we learn more about statistical learning and R programming. It's a great tool to use in projects, internships, and capstone!

Note: we will touch upon it briefly, so the material on caret will not be tested!