10 R packages to win Kaggle competitions

Xavier Conort, DataRobot

*Contact author: xavier@datarobot.com

Keywords: Kaggle, Machine Learning

R is rapidly growing in popularity among statisticians, data scientists and actuaries. An actuary by training, I became an *R* enthusiast myself 3 years ago when I discovered that *R* offered me a powerful platform for statistical and actuarial analysis. The main draw for me was the large palette of Machine Learning algorithms to tackle predictive modeling problems outside my comfort zone -- ranging from churn prediction to essay scoring, sales forecasts, flight arrival prediction, recommendation engine and credit scoring. Machine Learning helped me significantly reduce the modeling effort compared to traditional statistical parametric techniques, work on dirty data with limited domain insight and extract value from unstructured or large dimensionality datasets. Thanks to *R*, I placed in the top 10 of more than 15 Kaggle competitions and won several of them. 10 *R* packages were the key ingredients of my Kaggle solutions. In this talk, I will cover why I found those 10 packages particularly powerful and how I used them to build winning solutions.