

Enter a KDD Cup or Kaggle Competition

You don't need to be an expert!

Follow-Up

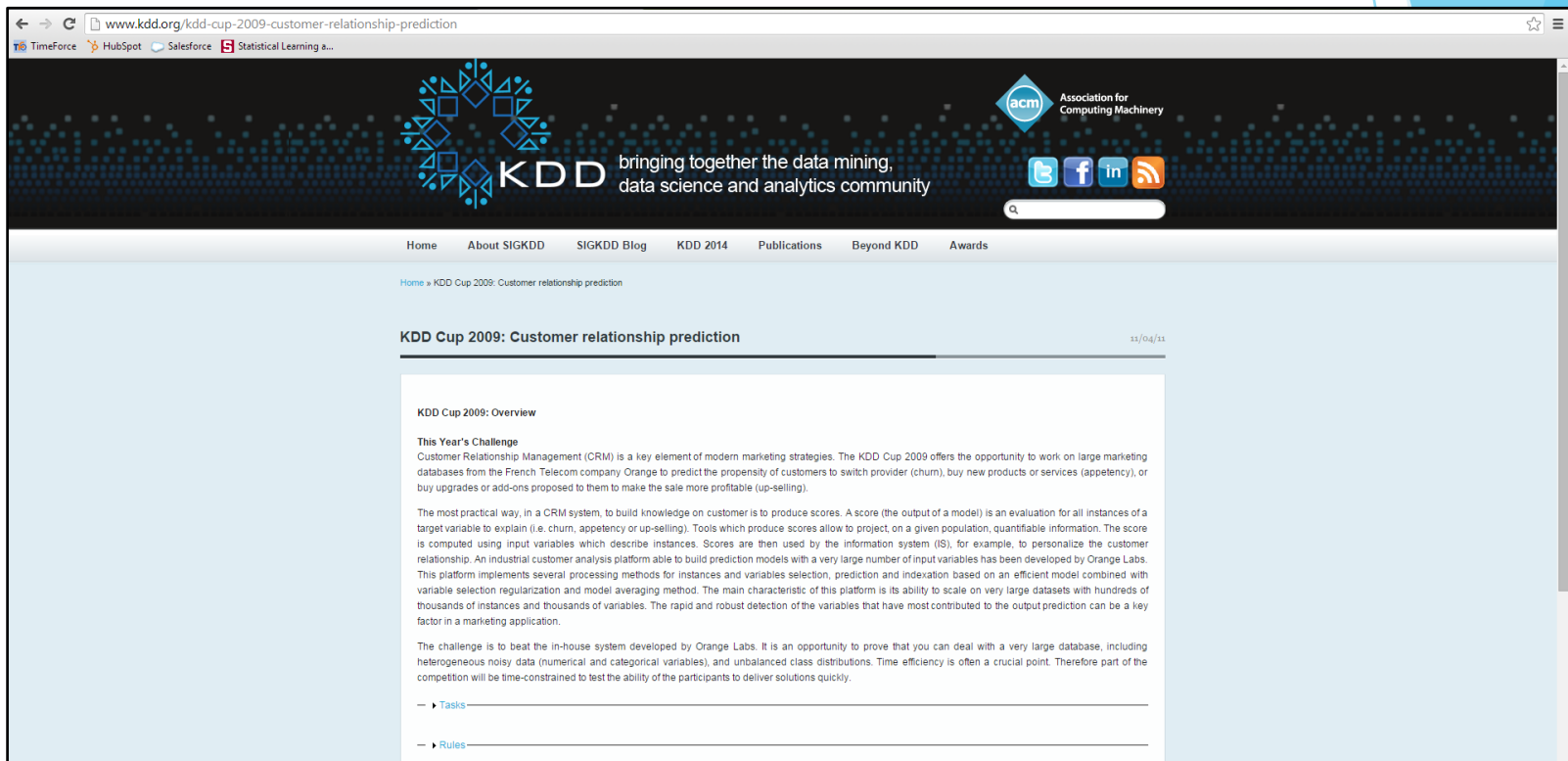
- ▶ Recording
- ▶ Datasets
- ▶ SPM trial download
- ▶ Step-by-step tutorial

- ▶ 3 Ways to Improve your Regression: Webinar
 - ▶ February 10th @ 10AM

- ▶ Contact us to provide up-to-date email address
 - ▶ support@salford-systems.com

KDD Cup 2009

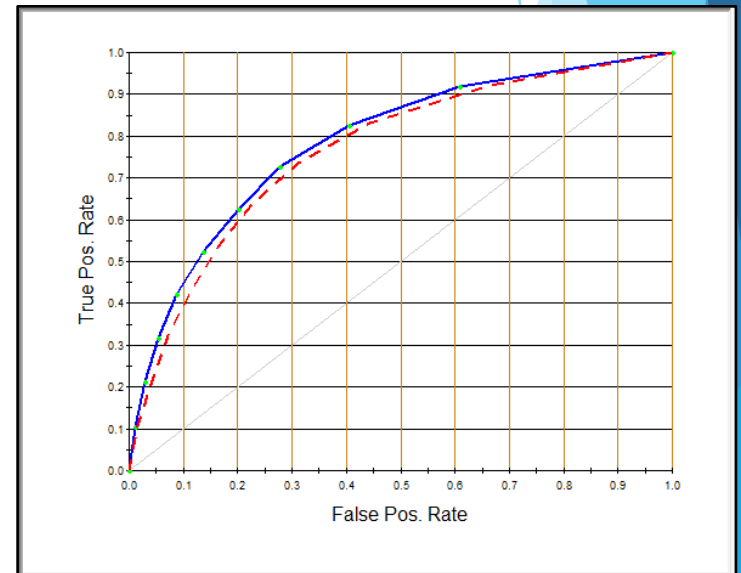
▶ Customer relationship prediction



The screenshot shows a web browser window with the URL www.kdd.org/kdd-cup-2009-customer-relationship-prediction. The page features a dark header with the KDD logo and the tagline "bringing together the data mining, data science and analytics community". The ACM logo is also present. Below the header is a navigation menu with links for Home, About SIGKDD, SIGKDD Blog, KDD 2014, Publications, Beyond KDD, and Awards. The main content area displays the title "KDD Cup 2009: Customer relationship prediction" and a date "11/04/11". The "Overview" section includes a "This Year's Challenge" heading and a paragraph describing the challenge: "Customer Relationship Management (CRM) is a key element of modern marketing strategies. The KDD Cup 2009 offers the opportunity to work on large marketing databases from the French Telecom company Orange to predict the propensity of customers to switch provider (churn), buy new products or services (appetency), or buy upgrades or add-ons proposed to them to make the sale more profitable (up-selling)." It also mentions that the challenge is to beat the in-house system developed by Orange Labs. At the bottom of the overview section, there are links for "Tasks" and "Rules".

KDD Cup 2009

- ▶ 3 binary classification tasks
 - ▶ Appetency: customer buys products or services
 - ▶ Churn: customer switches providers
 - ▶ Upselling: customer buys offered upgrades
- ▶ 15,000 predictors
 - ▶ ex) demographic, geographic, behavioral
- ▶ Best average area under ROC wins

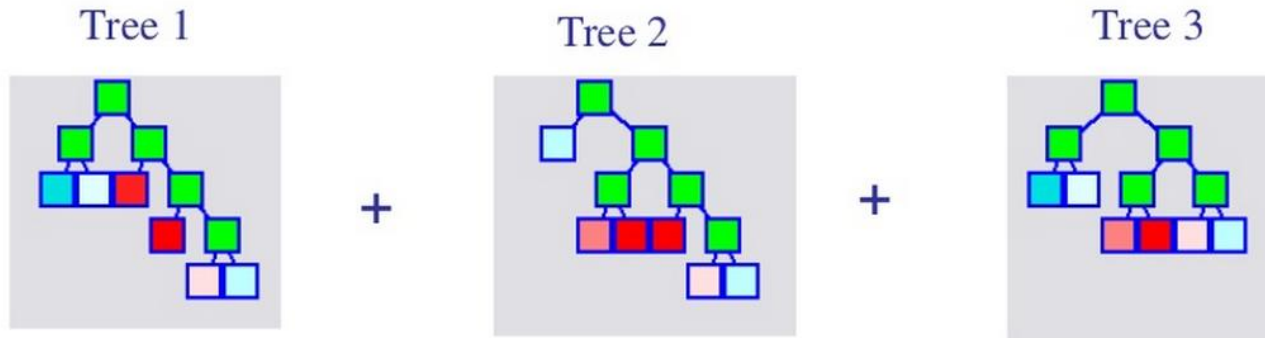


Challenges

- ▶ Large database
 - ▶ 50,000 x 15,000
- ▶ Numerical *and* categorical variables
- ▶ Missing data
- ▶ Unbalanced class distributions
 - ▶ Many more customers NOT doing these things

Approach

- ▶ TreeNet Gradient Boosting
 - ▶ Sequential error-correcting CART trees
 - ▶ Fast and accurate



- ▶ Variable selection
 - ▶ Shaving off unimportant predictors
 - ▶ Every decimal place counts!

Results

Rank	Team	Appetency	Churn	Upselling	Score
1	IBM Research	0.8830	0.7611	0.9038	0.8493
-	You!	0.9048	0.7346	0.9059	0.8484
2	ID Analytics, Inc.	0.8724	0.7565	0.9056	0.8448
3	Old dogs with new tricks	0.8740	0.7541	0.9050	0.8443
4	Crusaders	0.8688	0.7569	0.9034	0.8430
5	Financial Engineering Group, Inc. Japan	0.8732	0.7498	0.9057	0.8429

- ▶ Unable to compare to true target values
- ▶ Results can vary based on optimal selection criterion, random number seed, etc.

Follow-Up

- ▶ Recording
- ▶ Datasets
- ▶ SPM trial download
- ▶ Step-by-step tutorial

- ▶ 3 Ways to Improve your Regression: Webinar
 - ▶ February 10th @ 10AM

- ▶ Contact us to provide up-to-date email address
 - ▶ support@salford-systems.com