

Crowdsourcing Predictive Analytics

2012 CANE Spring Meeting

Boston, MA

April 4, 2013



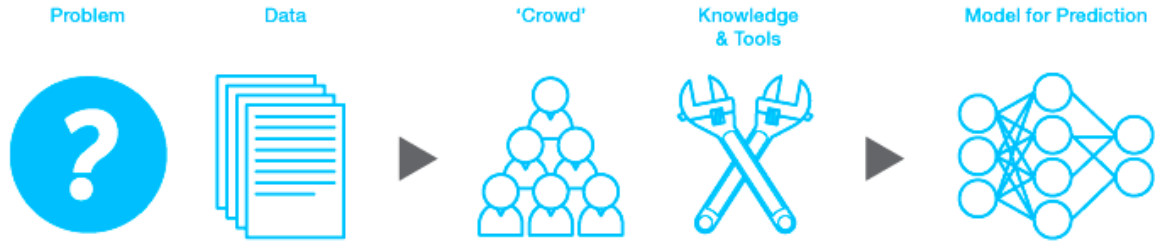
Allstate[®]
You're in good hands.



Crowdsource or Not?

• Running the contest

- Designing the question
- Structuring the competition
- Data
 - Developing training and test data sets
 - Appropriately protecting privacy
 - Protecting confidential information, trade secrets
- Running the competition



• What are you trying to accomplish?

- Assessing Distributed Knowledge
- Crowd Funding
- Crowd Labor
- Open Innovation
- Crowd Creativity

• How will you use the result?

• Does this approach fit with who you are?





kaggle™



Allstate®
You're in good hands.



Claim Prediction Challenge

Finished

Wednesday, July 13, 2011

Wednesday, October 12, 2011

\$10,000 • 105 teams



Predicting liability for injury from car accidents

Many factors contribute to the frequency and severity of car accidents including how, where, and under what conditions people drive, as well as what they are driving. The goal of the Claim Prediction Challenge was to predict bodily injury liability, based solely on the characteristics of the insured vehicle.

202 players in 107 teams + \$10k in prize money = 271% improvement to prediction



DATA PRIVACY/COMPETITION STRUCTURE

Players were given a data set that included three year's worth of coded data about what cars people were driving (i.e. code names of cars vs. the real makes and models), 26 coded variables for different vehicle characteristics, and the dollar amount of bodily injury liability for each vehicle. Using these three years of data to train their models, players submitted injury claim predictions for two subsequent years worth of vehicle data.

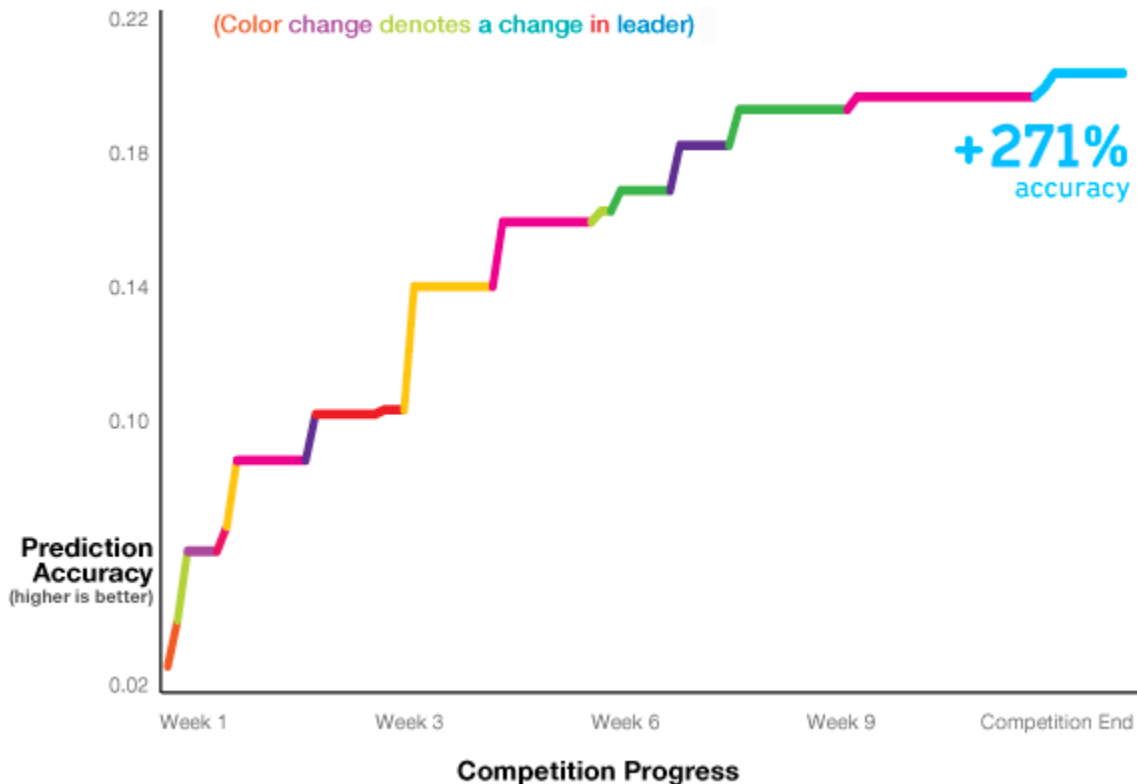
| Model Year | Blind Make | Blind Subcat |
|------------|------------|--------------|
| 2001 | P | P.24.0 |
| 2001 | P | P.24.0 |
| 2003 | Y | Y.29.0 |
| 2003 | Y | Y.29.0 |
| 2006 | AU | AU.14.1 |
| 2006 | AU | AU.14.1 |
| 2005 | P | P.18.0 |

The make/model data was obscured to keep the original private.



COMPETITION RESULTS

202 players, competing as 107 teams, submitted 1290 entries to the \$10,000 prize competition over the course of three months. The winning entry was 340% more accurate than the sponsor's existing method for predicting claims based on vehicle characteristics. Although the competitors developed their algorithms based on coded data, the sponsor now has predictive insight on exactly which characteristics of a vehicle translate into increased risk of bodily injury insurance claims, and can apply that insight to its product and pricing strategies.





kaggle™



Allstate®
You're in good hands.

Will I Stay or Will I Go? limited

Predict which of our current customers will stay insured with us for an entire policy term.

An important part of succeeding as an insurance company is having a good understanding of which of the company's current customers will be with the company into the future. Every customer comes with a different risk profile and it is critical to plan appropriately for that future risk. The goal of this competition is to predict which current customers will still be with the company in 6 months, given many of the customer's characteristics.

Started: 9:09 pm, Wednesday 12 September 2012 UTC
Ends: 12:00 am, Wednesday 12 December 2012 UTC
(90 total days)