FAST COMPANY.com

Where ideas and people meet

Article location:http://www.fastcompany.com/anthony-goldbloom-kaggle-whos-next
October 31, 2011
Tags: Leadership, kaggle, anthony goldbloom, Who's Next

Kaggle's Anthony Goldbloom Helps Companies Crunch Data With Crowdsourcing For Quant Geniuses

By Alina Dizik

It can take a bit more than a 30-second elevator pitch to understand Kaggle [1], a new Silicon Valley startup founded by 28-year-old Australian Anthony Goldbloom. After all, it's not selling anything, creating another social network, or peddling yet another useless app.

Instead, Kaggle is using the multitude of corporate data to build prediction models, which can be just as complicated as it sounds. Even its website doesn't provide a clear explanation of what the company does. "I'm a bit too familiar with it, so I take for granted that it's intuitive to everyone," says Goldbloom, who started the company in April 2010.

But once you understand the problems that Kaggle attempts to solve, the need for this kind of data platform becomes totally obvious.

Essentially, Kaggle taps into the heaps of corporate data that's readily available within a company to create competitions for data-prediction models. Once the competition is announced, data scientists eager to work on the problem compete to create a working model by mulling over the available information. In a way, it's crowdsourcing for quant geniuses [2]. Typically there's a monetary prize involved for the scientists and, of course, bragging rights. Kaggle now has 17,000 PhDs, data analysts, and scientists from all over the world who want to compete in building these models, Goldbloom says.

Here's one example of what Kaggle does: In 2010, an academic from Drexel University in Philadelphia signed on with Kaggle to ask data analysts to create a model predicting HIV progression. During the three-month-long competition, 109 teams of scientists used the records of 1,000 patients to create a model of the genetic blueprint. (The University provided all necessary data.) The crowdsourced model now predicts the individual severity of the virus with 77%

accuracy, compared with the 70% accuracy previously achieved using conventional methods. Solving bioinformatics problems is where his own interest lies, but Kaggle's approach is applicable to anything from accounting firms to aeronautics (Deloitte and NASA are clients).

Fast Company recently spoke with Goldbloom about Kaggle's mission, and what keeps him motivated.

Tough sell: The HIV progression competition really put Kaggle on the map. Prior to that, getting companies to fork over data was extremely difficult, but he stuck with the company, Goldbloom says. "The first nine months were really slow; it was a battle to get companies to host their data sets with us. After that case study things have started to explode. We won a contract with the Heritage Health Prize which will predict who will go to the hospital in the next year. The winning model might predict this with 90% probability, and will alert the provider system about at-risk patients," he says. Unlike the few short months of the HIV competition, analysts competing for the Heritage Health Prize will win $3 million dollars instead of $500. The new competition takes two years and will end in 2013.

Tapping in: Building data models is nothing new, but it requires resources some companies don't have. "Banks rely on statistical analysis to see who will default on a loan. Insurance companies rely on statistics to tell them who is going to crash. Google relies on stats to tell them whether the person searching for "blackberry" wants to know about the phone or the fruit," Goldbloom says. With so much data just circulating around most companies, Kaggle finds the geniuses willing to turn the data into actual knowledge.

Corporate approach: Goldbloom got the idea for Kaggle while interviewing sources for his internship at *The Economist* in 2008. "My wife will attest to the fact that I came back from that internship totally obsessed by the idea of big data," he said. Goldbloom realized that statistical modeling hasn't yet caught up with the amount of data available. More importantly, he realized that it's difficult for companies to make sense of data on their own, even if they know they need to be taking advantage of it. "Most CIOs had predictive modeling near the top of the list of things they wanted to implement. However, they were finding it a difficult product to buy: One vendor would sell a "Support Vector Machine" solution for $2 million, another a "Neural Network" solution for $1 million, and it's impossible to know which would work best on their problem. To me, predictive modeling competitions seemed like the obvious solution: Pit vendors and data scientists against each other and offer a big prize to the best approach--not dissimilar from architectural design competitions used to decide contracts for big property developments."

Bye-bye bureaucracy: Before starting Kaggle, Goldbloom worked at the Australian Treasury building economic models (he took a three-month leave to work at *The Economist*). It took 10 months before he actually left to start the company. "I felt like a jack in the box. After the Treasury, I moved to the Reserve Bank because I was restless. The bureaucracy was still pretty suffocating, so I ended up leaving."

Down time: "I like spending time with my wife but for a large part of this year she's been in Australia because of the visa. She's moving here promptly, but that's been really hard," Goldbloom says. "Without the discipline of having a wife to come home to you end up just working all the

time. I love kite surfing and mountain bike riding. It's kind of interesting, my kite surfing ability has probably deteriorated with the rate of Kaggle's success."

Building up: Being his own boss in a growing startup is difficult. "I feel like I want to go in and do all the work myself but that doesn't scale very well," he says. With six employees, he can still do most things himself, but that will change in the next few weeks when Kaggle announces its next round of funding. He's looking to grow the company to 25 people in the next year, and preparing to think long term. "I suspect that I'll be doing less and less work and instead I'll be managing and directing," he says. "The one advantage I have is a good sense of how every bit of the organization works and how the pieces fit together. I think I'm okay at everything and not spectacular at any one thing that Kaggle does."

Startup culture: At the infant stage, he and his partner, chief data scientist Jeremy Howard, relocated the company from Melbourne, Australia, to San Francisco. The transition was difficult, but being in Silicon Valley has made a huge difference, he says. "In America there is much more of a culture of joining a startup. If you look at the biggest 50 companies in Australia, only four of them are newer than 50 years old, versus 40 of them in America. It's also the mindset that everybody expects you to build a billion dollar and that kind of thing is infectious."

Being a "kid": As a twentysomething speaking to top executives is an adjustment. "I definitely remember the situation when [a CEO] said, 'My god, I've fired this other formidable company and I've hired a kid.' I never thought that people thought of me like that. I definitely dressed a little bit more formally to client meetings afterward. But I don't think that really makes a difference, to be perfectly honest. I think it's just being confident and making [companies] feel like you've dealt with their competitors."

Everybody wins: If Goldbloom is right, more data analysts toiling in the middle rungs of their organizations will rise to the top by participating in Kaggle's competitions. "We are moving towards a situation where data scientists rely on Kaggle to earn their full-time incomes," he says. "We plan to do this by having more prize money and having that prize money awarded to more than just the top three teams. For example, if all the major banks start building their credit scoring models using Kaggle competitions, they will bid up prize money to attract the best talent to their problem." (A credit score determines a customer's credit worthiness and the risk a financial institution takes on; banks compete on who has the most accurate model.) "If competition for Kaggle's top talent becomes fierce enough among banks, insurance companies, hedge funds--we hope the world's best data scientists will earn more than $50 million per year, just like the world's best hedge fund managers."

Check out our Who's Next series [3] for more profiles of the big thinkers everyone's talking about. For more leadership coverage, follow us on Twitter [4] and LinkedIn [5].

Links:
[1] http://www.kaggle.com/
[2] http://www.kaggle.com/Competitions
[3] http://www.fastcompany.com/next
[4] https://twitter.com/#!/FastCoLeaders

[5] http://www.linkedin.com/today/fastcompany.com