BUSINESS

Zestimates are way off, data crunchers compete to improve Zillow's model

More than 200 teams are working on beating the Zestimate algorithm.

By Mike Rosenberg Seattle Times | **JUNE 1, 2017 — 3:57PM**

If you have a house or you have looked for one, you have probably checked its Zestimate — Zillow's best guess as what a house is worth today.

The number might help you determine whether you should sell your house, or if the home you are trying to buy is a rip-off or a bargain. Or you might just use it to see what your friend or neighbor's home is worth.

But how accurate is that number?

Long a source of complaints from real estate agents, Zillow's online valuations are far from perfect, the company acknowledged. For example, instance, according to its own data, the Zestimate on the typical single-family house sold in Seattle today is off by an average of nearly $40,000.

More than 200 teams have started working on new home value algorithms in the 48 hours since the Seattle-based company started a contest to improve its Zestimates model — with a top prize of $1 million.

(http://stmedia.startribune.com/images/ows_1496344798933

PATRICK KEHOE - NEW YORK TIMES

Stan Humphries, the chief analytics officer at Zillow, said his team has been able to drop the home value predictor's error rate.

To win, teams will use Zillow data to create an algorithm that better predicts sales prices for about 110 million U.S. homes listed on the site. Each team's members and their scores are publicly available on the Kaggle platform, which is hosting the contest: The leading teams' models have improved slightly just from making tweaks over a 24-hour period, but still lag behind Zillow's predictions. One hundred finalists will be chosen next January, and the winners will be announced in January 2019.

Zillow said its Zestimates have improved significantly since launching 11 years ago and now have a median error rate of 5.6 percent. So half of homes end up selling within 5 to 6 percent of what the Zestimate said it was worth.

But that means half are sold for a price that's not so close to what the Zestimate predicted. In some cases, the error can be huge.

Most notably, Zillow CEO Spencer Rascoff sold his Seattle home last year for $1.05 million — 40 percent less than what the Zestimate said it was worth. Then a couple of months later, he bought a house in Los Angeles for $20 million — or $1.6 million more than the Zestimate.

The error rate varies by location, but even the average error can amount to a lot of money in a pricey market. The median single-family house in Seattle now goes for $722,000, and the Zestimate error rate here is 5.4 percent. That means the Zestimate on a Seattle house selling today will be off by an average of about $39,000.

Zillow acknowledges that real estate agents are still slightly better at predicting sales prices than its computers, but said the gap between man and machine is narrowing.

Data from Redfin show that listing prices put together by agents are usually only about 1 percent different from the final sales price in the Seattle region.

(There is a theory, however, that some agents deliberately list homes for less than they think they are worth to draw in more buyers and drive up competition).

Zillow said it has the best online home value estimations, though other companies with similar tools, like Seattle-based Redfin, disagree.

Zillow chief analytics officer Stan Humphries, who created the Zestimate, said its team of 15 data scientists and machine learning engineers have been able to drop the home value predictor's error rate significantly, from 14 percent when it launched in 2006.

They now run about 7.5 million statistical models nightly to update the Zestimates, using things like comparable home sales, local assessed values, home size and features, and hundreds of other data points. Lately, they've used machine learning to analyze home pictures from listings to determine how fancy various parts of the home are, taking the analysis beyond data like size and listing descriptions.

Humphries said the way to make Zestimates better is by using as much hyperlocal data as possible, which lends itself better to a nationwide contest than to further research and tweaks from its Seattle-based team.

"At this point we think that future gains will come from widening the circle of ideas to the global data scientist community," he said.

"I believe we'll be able to drive [the error rate] lower, I think a lot lower," he said. "Neither a human nor a computer will get to zero. People will disagree with human and computer opinions, and that's OK. But we think more opinions make consumers more comfortable."

The Zestimate is just one of many ways to get a sense of your home's value. Sites like Redfin and homes.com have their own algorithms, while local banks and real estate agencies can give you quotes almost instantly.

It's also easy now to look up home sales data from comparable properties online — something that, just a decade ago, was information kept only by real estate professionals.

But the Zestimate is probably still the most-used computer estimate. It's the most trafficked real estate information site in the United States, with 166 million average monthly users.

The algorithm has also been the subject of lawsuits. The company was sued recently by suburban Chicago homebuilders who argue that the numbers are being illegally passed off as official appraisals, which are typically done by licensed professionals who visit homes for sale. Zillow said the home value should be used as a starting point — just one piece of data to consider — and not an official appraisal.