

Data Mining Thanks to Twitter

By

ELIZABETH DWOSKIN

Updated Oct. 6, 2013 11:31 p.m. ET

In its IPO filing Thursday, [Twitter](#) Inc. disclosed how much the microblogging platform earned from a lesser-known side business: \$47.5 million came from selling off its data to a fast-growing group of companies that analyze the data for insights into news events and trends.

That is a small amount compared with the revenue generated from advertising, but Twitter's data business has rippled across the economy. The site's constant stream of experiences, opinions and sentiments has spawned a vast commercial ecosystem, serving up putative insights to product developers, Hollywood studios, major retailers and—potentially most profitably—hedge funds and other investors. Backed by millions of dollars in venture capital, hundreds of "social listening" firms have emerged.

"The economic impact of [Twitter](#) is massively bigger than whatever lies in their [IPO] filing," says Rob Bailey, chief executive of Data Sift, a London-based social-data analytics company.

Social-data firms spot trends that it would take a long time for humans to see on their own. The United Nations is using algorithms derived from Twitter to pinpoint hot spots of social unrest. [DirecTV](#) [DTV -0.13%](#) uses Twitter data as an early-warning system to spot power outages based on customer complaints. Human-resources departments analyze the data to evaluate job candidates.

One opportunity may be on Wall Street. Dataminr Inc., which has raised \$30 million in funding this year alone, has made serving traders its core business. It uses algorithms similar to those employed by hedge funds to conduct high-speed trading. Five minutes before news of last week's Capitol Hill shooting broke on TV, Dataminr subscribers got an alert to take action—giving them a leg up on the news that led the S&P Index to drop by 20 points five minutes later.

Dataminr's algorithms spotted just a handful of early eyewitness shooting reports posted on Twitter and instantaneously deduced from their timing, along with the Tweeters' influence and physical location that something urgent was happening. Other trading algorithms use "sentiment scores" to track the mood about certain stocks, driving decisions about whether to be bullish or bearish on firms.

These social-data companies are subsets of a burgeoning data industry that collects, disseminates and analyzes all the information that firms amass and people leave online. While estimates of the market value of the social-data industry are hard to come by, one research firm, IDC, estimates that the entire "big data" market has grown seven times as quickly as the information technology sector as a whole. It may be valued at \$16.9 billion in two years.

Most of the social companies pull data from a range of sources, including [Yahoo Inc.](#) [YHOO +1.14%](#) 's Tumblr, [Facebook Inc.](#) [FB +0.30%](#) 's Instagram, and the review site [Yelp Inc.](#) [YELP +1.04%](#) But Twitter, because of its public real-time data stream, is the industry giant. Even the Library of Congress is now cataloging tweets. Twitter's social network rival Facebook, on the other hand, has largely eschewed getting involved in the data business. In part, that is because Facebook's product isn't set up for public conversations, but rather for private sharing among friends. But with an eye toward Twitter's success in leading public conversations about real-time events, Facebook has recently partnered with the major media companies to provide aggregated information about consumer conversations on certain topics.

Each social-data firm boasts proprietary data-mining tools that go beyond basic keyword searches. Some can zoom in on a subset of people—say, women in a certain ZIP Code—and monitor phrases that show emotion. Then they can create a heat map or a sentiment score that measures how that subset feels about a topic. They have trained natural language processing algorithms to look at slang and broken grammar and to highlight tweets that indicate urgency because of words like "BREAKING."

"We don't just count the volume of these trends. That's naïve," says Nova Spivak, CEO of the Los Angeles-based firm Bottlenose. Rather, his firm looks at the momentum of trends.

"Businesses have spent years understanding the past," says Mr. Spivak. "Twitter has found a way to measure the present."

Twitter data aren't a perfect crystal ball. The service's most active U.S. users aren't entirely representative of the American population. According to the Pew Internet Project, though three-quarters of American adults are sharing information on social-networking sites, Twitter users, in particular, skew young—30% are under the age of 30. College graduates and wealthier Americans also have higher rates of Twitter use, but that gap is small, and narrowing.

Companies say it is still too early to determine the long-term value of Twitter data. One risk that Twitter raised in its IPO prospectus is that the company's growth could be undermined by spammers and fake Twitter accounts. Another risk: Twitter accounts can be hacked and used to put out incorrect information.

A Twitter spokesman declined to comment.

Companies are still struggling to understand how the Twitter conversation relates to the real world. Most aren't yet abandoning focus groups or customer-satisfaction surveys for social media monitoring, says Darrell Jursa, of Fleishmann Hiller, a public-relations firm specializing in crisis management. Some "80% to 90% of conversations still take place offline," he says. "You need to put the social data in context with everything else."

Moreover, companies that tap the data are increasingly trying to figure out whether a Twitter "crisis" actually resonates with all Americans or stays in an echo chamber.

Restaurant-chain Five Guys used Washington, D.C.-based New Brand Analytics to compare hamburger-bun quality in different parts of the country, leading the company to flag issues in some areas. But the firm warned the chain not to overreact to complaints, says Five Guys spokeswoman Molly Catalano. While some individual complaints were nasty, overall Five Guys was popular among Twitter users, New Brand found.

Twitter is still figuring out how to extract the most value from its data business, which grew 53% over the past year, according to papers the company filed with the Securities and Exchange Commission. The company has clamped down on the amount of data that outside firms are allowed to pull from its system free of charge.

Many smaller analytics startups are now turning to four companies that Twitter has dubbed "certified data resellers." These brokers, Gnip, Data Sift, Topsy and the Japanese firm NTT Data, account for the bulk of Twitter's data revenue. Last year, they paid Twitter monthly fees of about \$35.6 million.

Twitter's exponential growth has meant its influence extends well beyond marketing and crisis PR. Nonprofits, human-resource managers and politicians have found Twitter data useful, too. The Red Cross began analyzing Twitter data during 2012's Hurricane Sandy, says spokeswoman Anne Marie Borrego, using social-media-generated heat maps to direct aid. Twitter itself analyzed the site's chatter to predict the outcome of last year's U.S. presidential election.

But as Twitter plans to go public, some smaller firms worry that Twitter could further limit access to its valuable data. "Free," says Kristin MuhIner, CEO of New Brand Analytics, "isn't interesting to investors."

<http://online.wsj.com/news/articles/SB20001424052702304441404579118531954483974>