THOMSON REUTERS

Separating Real News from Fake News on Twitter in 40 milliseconds

"We have dozens and dozens of examples where Reuters Tracer discovered ground-breaking events ahead of major news organizations."

— Khalid Al-Kofahi, Head, Corporate Research & Development, Thomson Reuters

Thomson Reuters is the world's leading source of news and information for the financial and risk, legal, tax and accounting, and media markets.

Challenge

Today, Twitter is a key news source, with witnesses sharing firsthand experiences as events unfold. For journalists and businesses, it is a challenge to not only rapidly make sense of all the tweets, but also separate real news from fake news and opinions. Thomson Reuters, which works to help journalists and industry professionals find trusted answers among all the noise, turned to machine learning and advanced analytics to solve this challenge. "We set out to build a platform that could understand with great speed if an event was newsworthy, while maintaining our commitment to accuracy," said Khalid Al-Kofahi, head, Corporate Research & Development at Thomson Reuters.

Solution

The company refers to its intelligent news service, Reuters Tracer, as a "bot journalist in training." The solution processes about 13 million tweets daily, capturing events as they happen and determining: is an event true, is it newsworthy, and what is the scope and impact of that event. If a tweet is an opinion, Reuters Tracer can determine whether it comes from a recognized expert, and is therefore of news value. In delivering its results, it provides journalists and businesses with a "newsworthiness score" for each event that rates its assessed level of accuracy and credibility.

"To assist in evaluating the veracity of an event, we rely on hundreds of features and have trained the platform to look at the history and diversity of sources, the language used in tweets, propagation patterns, and much more, just as an investigative journalist would do," said Sameena Shah, director of Research, who was the lead scientist on Reuters Tracer.

Implementation

Thomson Reuters powers Reuters Tracer with Cloudera's modern platform for machine learning and advanced analytics to achieve the speed and accuracy it needed in analyzing tweets. "Cloudera provides us with state-of-the-art technology to help us analyze data, synthesize text, and extract value and meaning from data to deliver the insights that our customers are looking for," said Al-Kofahi. "The whole application is very fast. It takes less than 40 milliseconds to capture and detect events."

cloudera®

cloudera®

Key Highlights

Industry
- Media

Location
- Headquarters: New York, New York, USA

Business Application Supported
- Social media events monitoring

Impact
- Uncovered ground-breaking events ahead of major news organizations
- Captures and detects news events across millions of tweets in 40 milliseconds
- Frees journalists to focus on higher level reporting

Data Sources
- Twitter
- Database of government and nongovernment entities

Solution
- Modern Data Platform: Cloudera Enterprise
- Workloads: Data Science & Engineering
- Components: Apache Spark

Big Data Scale
- 13 million tweets daily

Results

Reuters Tracer helps journalists and businesses keep pace with a rapidly changing news landscape. "We are in the business of building information-based solutions for our professional customers in the financial, legal, tax, and accounting industries, and for Reuters, one of the leading news organizations," said Al-Kofahi. "With Reuters Tracer, we can alert our customers when market-moving events happen as they are reported, without delays. We have dozens and dozens of examples where Reuters Tracer discovered ground-breaking events ahead of major news organizations. Additionally, because we help journalists discover events, they can focus on higher value-add work as opposed to just reporting on events."

About Cloudera

Cloudera delivers the modern platform for machine learning and advanced analytics built on the latest open source technologies. The world's leading organizations trust Cloudera to help solve their most challenging business problems by efficiently capturing, storing, processing and analyzing vast amounts of data. Learn more at cloudera.com.