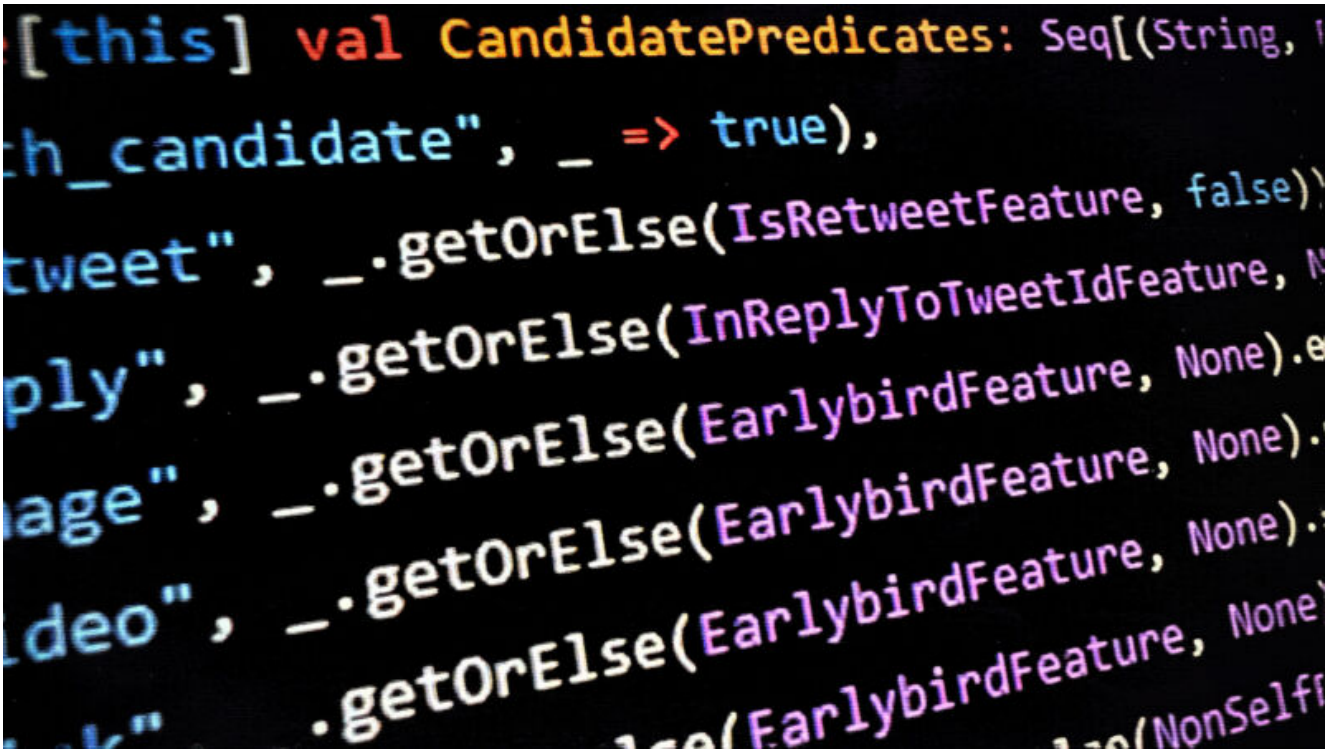
Twitter posts the code it claims determines which tweets people see, and why



Enlarge / Twitter has posted what it states is the code used by its algorithm to recommend tweets to its users.

reader comments

80 with

Share this story

Twitter has made good on one of CEO Elon Musk’s many promises, posting on a Friday afternoon what it claims is the code for its tweet recommendation algorithm on GitHub.

The code, posted under a GNU Affero General Public License v3.0, contains numerous insights as to what factors make a tweet more or less likely to show up in users’ timelines.

In a blog post accompanying the code release, Twitter's engineering team (under no particular byline) notes that the system for determining which "top Tweets that ultimately show up on your device's For You timeline" is "composed of many interconnected services and jobs." Each time a Twitter home screen is refreshed, Twitter pulls "the best 1,500 Tweets from a pool of hundreds of millions," the post states.

The largest source of those tweets are "In-Network Sources," or users someone follows. The top tweets from that pile are ranked on the likelihood of a user's engagement with that tweet's author; the more likely, the more their tweets show up in For You. For the "Out-of-Network Sources," those not followed by the user, Twitter says it considers tweets that attracted engagement from people users follow and tweets liked by those who like tweets similar to a user.

Already, those who have looked through the code have spotted considerations that raise many more questions. Many have posted them, naturally, on Twitter itself.

> *Twitter just released source code for "the algorithm"*
>
> *Oh, what file is this? Predicates for tweets on the home timeline?*
>
> *Oh what is that 2nd image? pic.twitter.com/UE3dU8e3Os*
>
> *— Ólafur Waage (@olafurw) March 31, 2023*

Ólafur Waage, a senior software developer at Norwegian software consulting service TurtleSec, noted that inside "HomeTweetTypePredicates.scala," some of the seeming considerations for a tweet to be a candidate for the "For You" section are:

- author_is_elon
- author_is_power_user
- author_is_democrat

- author_is_republican

Elsewhere in the code, a code comment presumably left by a Twitter engineer clarifies that those identification values are “used purely for metrics collection.” The comment reads as follows:

Advertisement

> *These author ID lists are used purely for metrics collection. We track how often we are serving Tweets from these authors and how often their tweets are being impressed by users. This helps us validate in our A/B experimentation platform that we do not ship changes that negatively impacts one group over others.*

The names of the objects in question such as “DDGStatsDemocratsFeature” or “DDGStatsElonFeature” seem to support this interpretation, but it may not be possible to confirm that with the available code. It’s interesting that Twitter is checking and collating these variables, however. During a Twitter Spaces audio session, a Twitter engineer noted that the Democrat and Republican labels were used for metrics. Musk, who claimed he was unaware of the labels before today, suggested they should not be there.

Other things considered about a tweet include whether it’s less than 30 minutes old, if it has pictures, and whether it’s from a “power user,” which some believe means a “legacy” verified account.

> *Most of the recommendation algorithm will be made open source today. The rest will follow.*
>
> *Acid test is that independent third parties should be able to determine, with reasonable accuracy, what will probably be shown to users.*
>
> *No doubt, many embarrassing issues will be…*

> *https://t.co/41U4oexIev*
>
> *— Elon Musk (@elonmusk) March 31, 2023*

Musk tweeted alongside the company's blog post that the recommendation algorithm, claiming that the "acid test" will be if "independent third parties" can "determine, with reasonable accuracy, what will probably be shown to users."

Twitter's posting of its algorithm code comes just days after the social network's broader source code was discovered on GitHub, potentially having been there for months, according to The New York Times. Twitter then obtained a subpoena forcing GitHub to reveal the GitHub poster's information.

A report from Platformer earlier this week suggested that Twitter utilized a secret list of 35 top Twitter users, including President Biden, LeBron James, Ben Shapiro, and Musk. Evidence of that list's implementation, reportedly spurred partly from Musk's dissatisfaction with his own engagement, has not been found so far in Twitter's posted code base.

Most notably, the code arrives just hours before "legacy verified" users—those given a blue checkmark to indicate authenticity or notability before Musk's purchase of the service—are to be un-verified in favor of paying Twitter Blue subscribers. While some users connected to governments and large organizations may apply for checkmarks of other colors, only Twitter Blue subscribers, at $8 per month, will receive "prioritized ranking in conversations," among other features.

All of those changes happen to arrive on April 1, or April Fool's Day.

https://arstechnica.com/?p=1928420