Using Predictive Coding for HSR Second Requests

By Robert Keeling

Under the Hart-Scott-Rodino Antitrust Improvements Act of 1976, transactions of a sufficient size require the parties to file a notification to the Federal Trade Commission (FTC) and the antitrust division of the U.S. Department of Justice (DOJ). (Hart-Scott-Rodino Antitrust Improvements Act of 1976, 15 U.S.C. § 18a (2015).) If either the FTC or the antitrust division suspect the transaction will limit competition, it may request additional information, known as an HSR second request.

These second requests often require the parties to review massive volumes of data in a short period to satisfy the federal review. Here, the high-speed, high-volume, power of machine learning, referred to as predictive coding or technology-assisted review in the legal arena, offers a cost-effective and accurate tool to cull through repositories of a company and find documents to respond to a second request. (See, e.g., Tracy Greer, Senior Litigation Counsel E-Discovery, Technology-Assisted Review and Other Discovery Initiatives at the Antitrust division, U.S. DEP’T OF JUSTICE, 5 (Mar. 26, 2014).

Over the last several years, the antitrust division largely has embraced predictive coding, recently describing it as “working effectively for both the division and the producing party in the majority of investigations.” Indeed, the antitrust division’s latest Model Second Request, released in December 2016, includes instructions for Technology Assisted Review, or TAR.

While those instructions offer a useful starting point for practitioners, practitioners should consider other guidance before using predictive coding. Most notably, in a white paper recently submitted for the American Bar Association’s Spring Meeting, the Division’s Senior Counsel Tracy Greer expressed a “concern” over parties reviewing for responsiveness during the course of their privilege review after the application of TAR. Ms. Greer stated that the antitrust division “will not agree to a party conducting essentially a second responsiveness review.” (Tracy Greer, Senior Litigation Counsel E-Discovery, Avoiding E-Discovery Accidents & Responding to Inevitable Emergencies: A Perspective from the Antitrust division, ABA Spring Meeting (March 2017).)

In other words, after implementing a predictive model, the division will not allow counsel to make its
own responsiveness determinations on those documents determined by the model to be likely responsive during the course of a privilege review.

The impact of this guidance could be far reaching. As predictive coding has helped mitigate the ever-increasing costs for responding to second requests, agency guidance — from both DOJ and the FTC — has grown more specific. Practitioners therefore should consider these new agency pronouncements prior to deciding whether to employ predictive coding when producing documents in response to second requests.

**Predictive Coding Helps Reduce Costs**

The burden of producing documents under a second request has risen rapidly over the past decade. A 2014 survey of parties responding to second requests between 2011 and 2013 reported a range of costs between $2 million and $9 million to comply, accounting for up to 5.5 percent of the value of the transaction. (Peter Boberg & Andrew Dick, Findings from the Second Request Compliance Burden Survey, XIV THE THRESHOLD 3, at 33 (Summer 2014).)

The median review time took nearly six months from the issuance of the second request until the investigation was closed or the agency took action, and companies spent a median of nearly $1 million per month to comply with second requests.

Two factors are frequently identified as driving the costs of answering a second request: (1) the quantity of information sought by the government, and (2) the increasing amount of data created and maintained by companies.

First, a federal agency’s market analysis of the transaction is fact-intensive and the government may require “substantial volumes of documents and quantitative data.” (Deborah Platt Majoras, Chairman, Federal Trade Commission, Reforms to the Merger Review Process, FED. TRADE COMM’N, at 2 (Feb. 16, 2006).) Thus, the scope of discovery under second requests is frequently more open-ended and less definite than in litigation. Respondents to the 2014 survey reported that as many as 18 separate product markets and up to eight geographic markets may be included in the scope of a second request. Parties searched for documents in up to 24 repositories to comply with federal requests.

Significant time is spent not only to identify responsive documents, but also to allow the federal agency to review the production and make a determination. Delays in clearing the merger or acquisition impose costs on the parties — from opportunity costs of the time spent by the parties’ employees in gathering the data, to costs associated with the delayed realization of the merger efficiencies and synergies, to the diminution in value of the assets associated with the transaction. (See Comments of the Section on Antitrust Law of the American Bar Association in Response to the Antitrust Modernization Commission’s Request for Public Comment Regarding the Hart-Scott-Rodino Second request Process, ABA (Dec. 5, 2005), available at http://www.americanbar.org/content/dam/aba/administrative/antitrust_law/at-comments/2005/12-05/hsr_2nd_request_comm.authcheckdam.pdf.)

The 2014 survey found that the median time to respond to a second request consumed 550 hours from in-house counsel and 500 hours from non-legal in-house management.

Given the substantial burden second requests place on parties, any reduction in compliance costs may provide significant benefits to both merging parties and consumers. Parties therefore increasingly turn to predictive coding as a means to cull through the corpus of a company’s data and narrow the amount of documents requiring “eyes-on” review by both the company’s counsel and the DOJ’s attorneys. (See Greer, supra.)

While there is no comprehensive analysis of the costs and time saved by parties that utilize predictive coding, anecdotal evidence suggests users gain significant benefits. One case study reported using predictive coding to review 880 GB of data from two countries in two languages in response to a second request by the FTC. (Second Requests: Complex Cases, Demanding Deadlines, Impressive Results, KROLL ONTRACK (2015), available at http://www.ediscovery.com/cms/pdf/CST_HSR_krollontrack2015.pdf). Predictive coding returned responsive documents with a 90 percent accuracy rate and reduced the final number of documents produced to under 500,000. Predictive coding also eliminated an estimated 1,200 hours in document review and saved the company almost $1.2 million.


Part way through the review, it became clear that it would not be possible to complete it within the schedule, so predictive coding was brought in to finish the process. Using predictive coding, the evaluation of the remaining 35 percent of the collection cost only 5 percent of the total amount and took only 10 percent of the total project time. It was estimated that, had the parties used predictive coding from the start, the review may have been completed in less than half of the time.

**Feds Embraced Predictive Coding, Issued Guidance**

As more parties have turned to predictive coding to respond to second requests, the antitrust division likewise has encouraged its use, reporting that predictive coding has allowed parties to save substantial time and money. For example, the division has reported negotiating the use of technology-assisted review in approximately a dozen instances. (Greer, supra.) It has found that the use of this technology has produced smaller, more responsive document productions that offer “substantial benefit” to division staff and cost less time and money from producing parties. Nevertheless, the division and other regulators continue to impose requirements on the use of predictive coding in second requests that may actually discourage the use of predictive coding in this context.

In March 2014, the division published a paper discussing best practices for using predictive coding. The paper outlined four key considerations for a party contemplating use of predictive coding.
First, the 2014 protocol encouraged parties to negotiate with the division what process would be used to collect the information that will be reviewed by predictive coding. The division discouraged using search terms to narrow the collection and preferred that parties instead limit date ranges, de-duplicate the collection, eliminate certain domain names from email collections, or eliminate some file types. Federal authorities also encouraged parties using predictive coding to identify any categories of documents that may be technically responsive to the second request but are not essential to resolving the antitrust questions involved in the investigation. Production of such documents may be deferred or foregone entirely.

Second, the division asked in the 2014 protocol that certain documents be removed from the collection to be reviewed by predictive coding because it believed technology-assisted review to not yet be an appropriate tool to capture relevant information (1) not associated with an identifiable custodian (e.g., websites, corporate intranets), and (2) from spreadsheets or non-text-based information (e.g., pictures, drawings, audio, or video files). The division indicated that it was not convinced that predictive coding worked effectively with foreign and mixed-language documents.

Third, the antitrust division indicated that a TAR workflow should be discussed and reviewed with the division. The division expected that the collection would be de-duplicated prior to applying TAR, but discouraged other analytical tools such as consolidating email threads. Federal authorities wanted to know who is responsible for training and quality control of the collection and the means by which that person would identify the “seed set.” The division was especially interested in seeing the “overturn rate” or the frequency with which the subject-matter expert reviewing the results disagrees with the responsiveness determination of the predictive model. Federal authorities have thus far required a confidence level of 90 percent, but recognized that this is much higher than the likely confidence level for manual review.

Fourth, the division asked that producing parties provide a statistically significant sample of nonresponsive documents to verify that a significant number of responsive documents have not been excluded from the review process. Generally, federal authorities request multiple samples and a representative will select one or more sets to review. This often occurs in the offices of the producing party’s counsel and the division generally will agree to complete its review in one or two days.

Increased Requirements on Producing Parties

Further agency guidance may raise concerns for practitioners going forward. The FTC revised its “Model Second Request” to include instructions on the use of predictive coding. It contemplates significant FTC oversight over the protocols used to set and verify how the software identifies responsive documents. Parties must describe to the FTC their methodology including:

(i) “how the software was utilized to identify responsive documents,”
(ii) “the process the Company utilized to identify and validate the seed set documents,”
(iii) “the total number of documents reviewed manually,”
(iv) “the total number of documents determined non-responsive without manual review,”
(v) “the process the Company used to determine and validate the accuracy of the automatic determinations of responsiveness,”
(vi) “how the Company handled exceptions,” and
(vii) “if the Company’s documents include some foreign language documents, whether reviewed manually or by some technology-assisted method.”

It also calls for providing the FTC with “all statistical analyses utilized or generated by the Company or its agents related to the precision, recall, accuracy, validation, or quality of its document production.” (Model Request for Additional Information and Documentary Material (Second Request), FTC Premerger Notification Office (rev’d. Aug. 2015.).)

While the FTC guidelines are limited to M&A transactions under its jurisdiction, they may be persuasive authority to other government agencies considering allowing the use of predictive coding to respond to regulatory inquiries. Nonetheless, the FTC’s provisions are still novel and may impact a responding party’s decision to use predictive coding or reduce some of the benefits the party may gain.

Although the level of transparency suggested may help regulators be more confident in the production, such oversight is not typical in a manual review and may create additional risks for the responding parties. (Jennifer Kennedy Park & Scott Reents, Use of Predictive Coding in Regulatory Enforcement Proceedings, BNA (Aug. 7, 2012.).)

For example, the iterative rounds of attorney coding associated with certain forms of predictive coding may be longer if statistical measures must be shared with the requesting regulator. Further, the regulator may opt to expand its requests for documents based on a review of nonresponsive documents in the seed set. (Id.) Micro-managing the use of predictive coding provokes further discussion about the extent to which such a detailed review is helpful.

More significantly, the antitrust division recently offered new guidance in a white paper to the ABA that practitioners should consider prior to employing predictive coding. As noted above, Tracy Greer, at the antitrust division, has emphasized that the division is concerned about the use of a second responsiveness review after the technology-assisted review process has been completed. (See Greer, supra.)

Greer explained that the division finds predictive coding more attractive because it relies on “more knowledgeable reviewers” making responsiveness determinations, and thus the results of the predictive coding process are better and more consistent than manual review. The paper begins with the proviso that it represents Ms. Greer’s personal views and not those of the division, but it states that “the Division will not agree to a party conducting essentially a second responsiveness review of the production during the privilege review process.” (See id. at 6.)

The white paper therefore suggests that – absent a modification from DOJ – parties using predictive coding when responding to second requests may be prohibited from removing documents that are not responsive (i.e., irrelevant) to the second request merely because such documents scored above the predictive model’s “cut-off line” – the line between documents likely to be respon-
sive versus likely to be nonresponsive (according to the predictive model).

More specifically, if attorneys encounter obviously non-responsive documents in the course of a privilege review – such as employee’s medical records or pictures of the employee’s children – the division’s policy suggests that the human attorneys will not be allowed to exercise their own independent judgment to mark such documents as non-responsive. Accordingly, the division’s policy suggests that such non-responsive documents must be produced, even though they are plainly outside the scope of the second request.

The white paper further seems to contemplate the use of more traditional predictive coding applications (such as Relativity Assisted Review or Equivio) where “expert” attorneys “train” the system by coding exemplar documents. The algorithm subsequently scores the larger document population based on the review-coding provided by the attorney expert. The white paper’s pronouncements against further responsiveness review are directed at this larger document population that is separate from the initial training.

But numerous predictive coding vendors have introduced new applications employing what is known as active learning in machine learning or “continuous active learning,” as it’s referred to in the legal domain. Using continuous active learning, attorneys responding to a second request allow the predictive model to identify new responsive documents for training and review, providing an iterative and continuous review until an acceptable volume of responsive documents are identified.

Accordingly, there is no distinction between the training and responsiveness review because the training is the responsiveness review. The white paper does not address how the division will treat companies who choose to employ active learning when responding to the second request.

**Conclusion**

When responding to second requests, practitioners should consider whether to employ search terms or predictive coding/TAR (or a combination of the two) to reduce the volume of data needing review and to identify potentially responsive documents. Predictive coding offers significant benefits in the litigation context, and similar time and cost-saving benefits can be achieved by using predictive coding to respond to HSR second requests. Even regulators have touted the benefits of predictive coding.

As the use of predictive coding has increased, however, the antitrust division of the U.S. Department of Justice and the Federal Trade Commission have become increasingly specific in their guidelines for using predictive coding to respond to HSR second requests. Before choosing to employ predictive coding when responding to second requests, practitioners should consider recent guidance from the FTC and DOJ. In particular, practitioners should consider the recent guidance from the DOJ that the division will not agree to a party conducting a responsiveness review as part of its privilege review process if that party has employed predictive coding/TAR.

Attorneys familiar with predictive coding know that it does not deliver perfect results. Predictive coding is a powerful tool in the review process, but it typically is not a complete replacement for human review. Practitioners should therefore consider whether the requirements that the FTC and DOJ recently have placed upon companies using predictive coding are enough to forego technology-assisted review and its benefits in responding to HSR second requests.

********

This article has been prepared for informational purposes only and does not constitute legal advice. This information is not intended to create, and receipt of it does not constitute, a lawyer-client relationship. Readers should not act upon this without seeking advice from professional advisers. The content therein does not reflect the views of the firm.