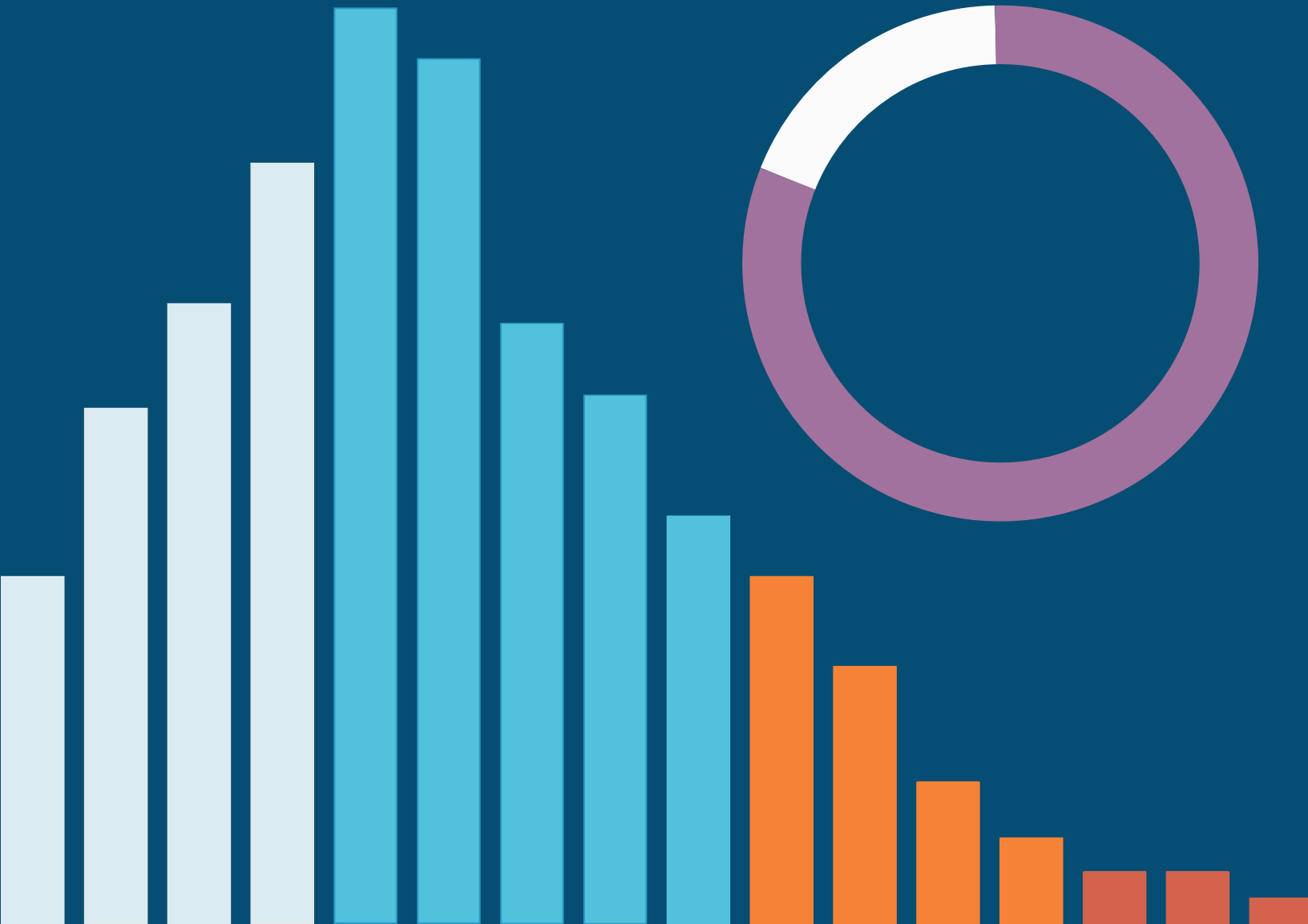
everlaw

Tactical Review with Predictive Coding



everlaw

Tactical Review with Predictive Coding

Linear review used to be standard in ediscovery when the volume of an entire document collection tended to be smaller. Before electronic discovery, it was industry practice to have human eyes review every document, and this practice carried over into electronic document review. Linear review soon became impractical when modern data formats and communication technologies made the volume of information in a body of documents grow exponentially. Nowadays, a document corpus could number in the hundreds of thousands, if not millions, of documents. Linear review is no longer practical for many cases involving electronically stored information.

Predictive coding is frequently suggested as a way to empower attorneys to focus their time on information that is relevant to the claims and defenses of their case, from the investigation stage of a lawsuit to responding to discovery requests. What is sometimes mysterious to those receiving these suggestions, however, is how predictive coding actually works or how it can be used effectively in a case. This white paper will provide an overview of predictive coding, a description of how it works in Everlaw, judicial opinions on predictive coding, and ways to tactically use predictive coding in litigation.

Predictive coding has transformed ediscovery, reducing the amount of data that need to be reviewed by upwards of 80% in some matters.

What is Predictive Coding?

Predictive coding is the application in ediscovery of machine learning which provides computers with the ability to learn from human input to make educated guesses beyond that input for which humans have not provided explicit programming. Predictive coding has transformed ediscovery, reducing the amount of data that need to be reviewed by upwards of 80% in some matters.

Predictive coding may seem foreign, but many of the services that consumers use every day incorporate this technology. For example, any time someone gives a song a thumbs-up on Pandora, that person receives more (and more accurate) recommendations for additional music as the system learns the types of music most relevant to their tastes. Movie and TV recommendations on Netflix or Amazon are also driven by machine learning, as their platforms learn from prior viewing choices to suggest better, more relevant content. Even Google search results are informed by a machine learning algorithm that examines and learns from past search and browsing behavior.

In most cases, the machine learning process is very similar:

1. The machine learning system catalogs the features of each object in the corpus (e.g., songs, movies, products, websites, etc.). For a song, the features might include the key, tempo, lyrics, artist(s), etc.
2. A human classifies some objects as relevant or irrelevant. In the Pandora example, this is giving songs a thumbs up or thumbs down.
3. The machine learning system analyzes the human input to determine which features affect relevance, and how.
4. There are at least a dozen different algorithmic approaches to this particular analysis. The machine learning system uses this information to classify the remaining objects as relevant or irrelevant.

In ediscovery, this machine learning approach is deployed in predictive coding systems—also known as Technology Assisted Review, or TAR—that generally fall into one of two categories:

1. Simple Passive Learning (or TAR 1.0): A subject matter expert classifies some documents to be used for training, which the system then uses to test the reliability of the predictions as more documents are classified. Once the performance is acceptable, the prediction model is rolled out to all remaining documents.
2. Continuous Active Learning (or TAR 2.0): All review decisions automatically train the system, and the system continually updates the predictions as new human classifications are made.

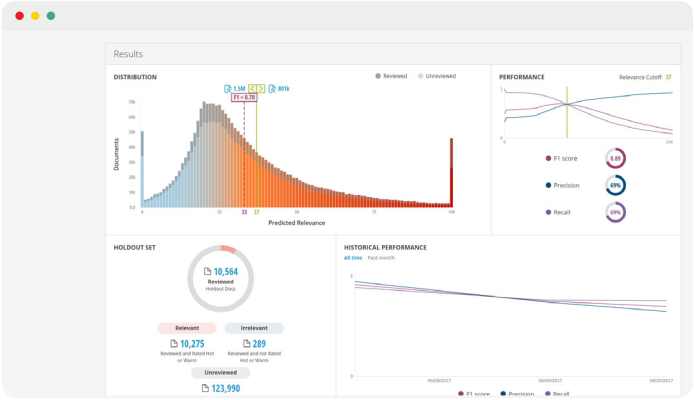
TAR 2.0 has many advantages over TAR 1.0. TAR 1.0 requires experts to do the initial training, and it is less effective, because it can't learn from subsequent decisions. TAR 1.0 also cannot handle rolling productions without having to start over. Finally, TAR 1.0 doesn't work well when the proportion of relevant documents is low.

As with any specialized domain, there is some jargon with which it helps to be familiar:

RICHNESS OR PREVALENCE The percentage of documents in the dataset that are relevant. If there are 100 documents in total, and only 10 are relevant, the prevalence is 10%.

RECALL The percentage of relevant documents retrieved. If there are 10 relevant documents in the dataset, and the system correctly identifies 6 of them, the recall would be 60%. This is a measure of the completeness of the results.

PRECISION The percentage of retrieved documents that are relevant. If the system correctly identifies 6 relevant documents, but incorrectly identifies another 54 as relevant, then precision would be 10%. This is a measure of the purity of the results.



The predictions generated by a predictive coding model can be enormously powerful. They can be used to avoid the manual review of documents that are almost certain to be uninteresting, saving potentially thousands or even millions of dollars in review time.

Precision and recall are used universally as closely related measures of a predictive coding model's performance. There is an inherent tradeoff between the two: classifying everything as relevant will maximize recall, but at the cost of minimizing precision because the "relevant" set would include many irrelevant documents as well. Similarly, classifying only one document as relevant might maximize precision (assuming this one document is indeed relevant), but it would minimize recall if there were in fact more than one relevant document. The ideal predictive coding model seeks to optimize this tradeoff, although there are some situations where one metric is more important than the other (e.g., recall, in the hunt for a "smoking gun" document). The predictions generated by a predictive coding model can be enormously powerful. They can be used

to avoid the manual review of documents that are almost certain to be uninteresting, saving potentially thousands or even millions of dollars in review time.

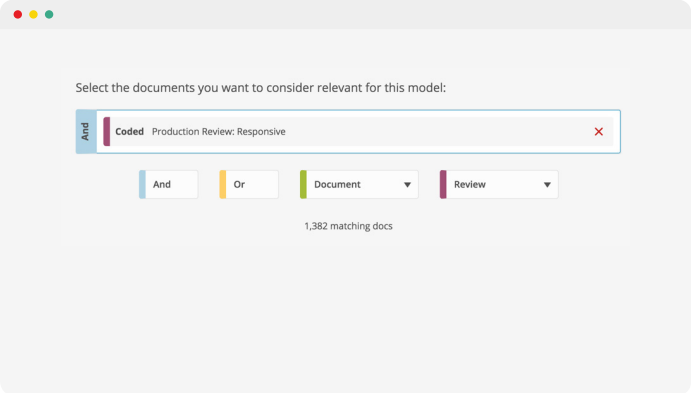
They can be used to prioritize documents for relevance review, making the review process much more efficient, or for other kinds of review, such as identifying potentially privileged documents, or reducing the chance of a costly or damaging clawback. They can also be used for quality control in a review (i.e., by looking for misalignment between the human and machine classifications), making the review process much more accurate. Given these benefits, it is not surprising that the vast majority of corporate counsel are using predictive coding on their cases,[1] nor that courts are routinely approving the use of predictive coding in the cases before them.[2]

How Does Predictive Coding Work in the Everlaw Platform?

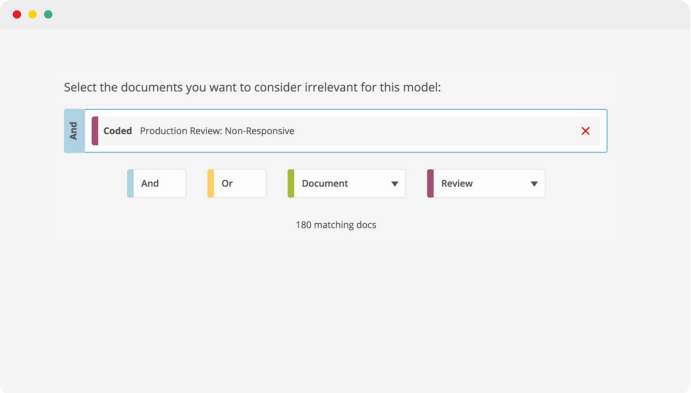
Predictive coding on Everlaw continuously learns from reviewer decisions, enabling reviewers to use their standard workflow—assigning ratings, codes, and attributes to reviewed documents—in order to teach the system how to find more relevant documents on their behalf.

Everlaw enables both novices and power users to easily create a predictive model via the wizard-driven process, ensuring that its powerful capabilities can be applied to a variety of use cases, including surfacing the most relevant predictions or helping to ensure quality control in a collaborative document review. The creation of a predictive coding model uses the same visual, color-coded logical operators and search interface seen elsewhere on the platform. This makes it easy to benefit from the power of artificial intelligence without paying more for the functionality or relying on a certified expert to build and manage the models. For example, to identify data responsive to all discovery requests, a user could create the following prediction model:

Relevant records are any coded Responsive:



Irrelevant records are any coded Non-responsive:



Predictive coding on the Everlaw platform is based on the Continuous Active Learning approach, and scores documents on a 0–100 scale based on the ratings of reviewers, with scores closer to 100 indicating a higher likelihood of relevance. A subset of documents is automatically set aside as a "holdout set" in the predictive coding

1 Per Norton Rose Fulbright's 2016 Litigation Trends Annual Survey, 71% of corporate counsel reported that they are using predictive coding on at least a minority of their cases.

2 See, e.g., Magistrate Judge Andrew Peck, of New York's Southern District, who declared that "the case law has developed to the point that it is now black letter law that where the producing party wants to utilize TAR for document review, courts will permit it." *Rio Tinto PLC v. Vale SA*, 306 F.R.D. 125 (S.D.N.Y. 2015).

system to compare prediction scores against actual user ratings and measure the performance of the prediction engine.

The holdout set comprises 5% of the total case data created from random sampling of 5% of the data uploaded to the case. This data is not used to train prediction models, in order to eliminate possible bias in evaluating a model's performance.

In addition to the traditional precision and recall metrics, Everlaw also displays the F1 score, a weighted average of the precision and recall metrics that is used to gauge the overall performance of a model. Given the aforementioned tradeoff between precision and recall, the F1 is useful because it takes both into account. Indeed, the predictive coding system's default sets the relevance boundary to the score that results in the highest F1—the score that has the best balance of precision and recall. This creates a more defensible model, designed to hold up to legal scrutiny, with more insight and statistical confidence to determine when to stop review.

A new prediction model automatically starts generating predictions once about 200 documents have been reviewed (or 5% of the case, whichever is smaller), with at least 50 satisfying the "relevant" criteria and 50 satisfying the "irrelevant" criteria for the given model. Because the system gives every document a score from 0 to 100 (with 100 being the most likely to be relevant), judgment about what prediction cutoff to use to define relevance/irrelevance is deferred to the user. This is in contrast to binary systems, which essentially give a thumbs up/thumbs down prediction classification to documents.

As a continuous active learning system, Everlaw's predictive coding engine will learn from any documents classified per the definitions of relevant or irrelevant in the model. The review team should therefore take steps to review a diverse sampling of the data to avoid any bias in the training documents, such as reviewing specific communications and not looking at others.

What Are the Best Tactical Uses of Predictive Coding?

Attorneys can use predictive coding in multiple stages of litigation to expedite the review of data, catch mistakes before they become costly, and potentially avoid the manual review of certain documents altogether. Based upon the needs of the case, specific prediction models can be created for individual requests for production, privilege calls, or defenses. The following are some of the best tactical uses of predictive coding over the life of a case.

Rule 11 Reasonable Inquiries

While attorneys do not need to plead facts, Rule 11 requires attorneys to know facts after conducting a reasonable investigation that are enough to make it reasonable to bring a case to the point of seeking discovery.[3] As such, lawyers should ask themselves at the start of a case: have we made a reasonable investigation?[4]

In the era of "big data," using predictive coding can help lawyers prepare a complaint as part of their reasonable inquiry under Rule 11. The factual contentions in a complaint must have evidentiary support after a reasonable inquiry by the attorneys who filed the lawsuit.[5]

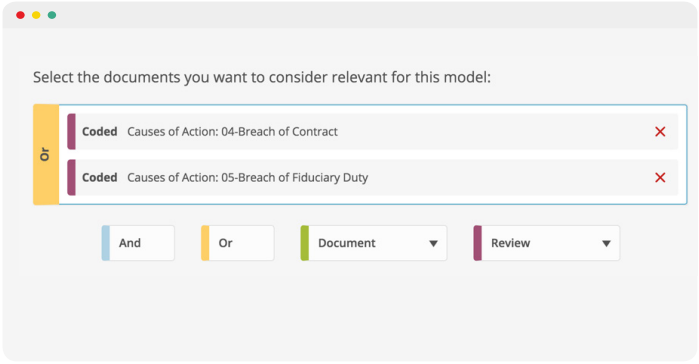
---

3 *Frantz v. United States Powerlifting Fed'n,* 836 F.2d 1063 (7th Cir. 1987).
4 *Bradgate Assocs. v. Fellows, Read & Assocs.,* 999 F.2d 745 (3d Cir. 1993).
5 USCS Fed Rules Civ Proc R 11(b)(3).

Predictive coding can help attorneys ensure the complaint is well grounded in fact by creating prediction models to identify facts supporting the allegations in the complaint. This does not mean starting full document review prior to filing the complaint, but the known data that forms the basis for the allegations can be the template for a prediction model. This model can then be used to identify other records that can support the causes of action in the complaint.

Consider a case where there are claims of breach of contract and breach of fiduciary duty in the development of a commercial property. The key facts involve the elimination of a property management fee that was to begin after the construction was completed. To support the plaintiff's claim that the elimination of the fee was a breach of contract and fiduciary duty between the business partners, the attorneys would need to find the different versions of the contracts, communications regarding the management fee, and a contract without the management fee.

Searches would first be done to identify the contracts and known communications between the business partners. These documents can be coded to denote the causes of action they support. A prediction model can then be created based on the issue coding to identify similar documents that could support the prospective causes of action in the draft complaint. The relevance criteria for that model might look like this:



Predictive coding can help attorneys conduct a reasonable inquiry into their data before filing a lawsuit. This initial analysis helps to determine the key facts supporting the allegations in the case, yielding complaints that are more well-grounded in fact. Furthermore, the predictive coding model can learn from this early attorney review to identify other possibly relevant information that may be useful in later stages of the litigation.

Rule 26(a) Initial Disclosures

Predictive coding can be helpful in quickly identifying initial disclosures under Rule 26(a), because the system can learn from work product from Rule 11 analysis to identify other possible documents that support a party's claims or defenses.

A party must identify all ESI within its possession, custody, or control to support any of its claims or defenses in a lawsuit without waiting for a discovery request.[6] Initial disclosures are due either at or within 14 days of their Rule 26(f) conference, but a different time can be stipulated by the parties or by court order.[7]

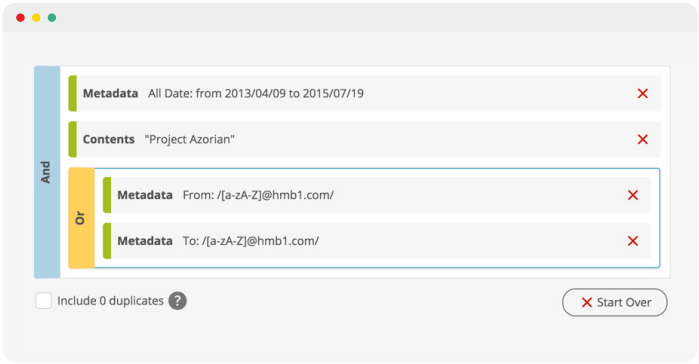
---

6 Fed. R. Civ. P. 26(a)(1)(A)(ii).
7 Fed. R. Civ. P. 26(a)(1)(C).

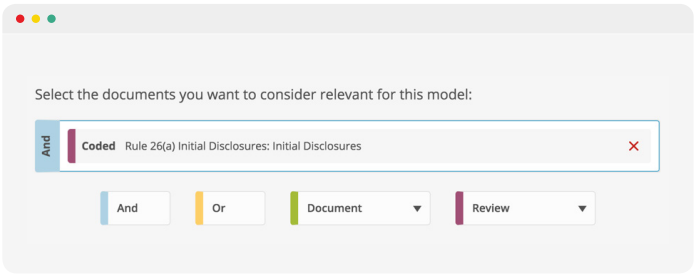
The Rule 26(f) conference must be "as soon as practicable" and at least 21 days before a scheduling conference is held or a scheduling order is due pursuant to Federal Rule of Civil Procedure Rule 16(b).[8] The Rule 16(b) scheduling order must be issued within 90 days after any defendant has been served with the complaint or 60 days after a defendant has appeared, whichever is earlier, or if there is good cause for delay.[9] As such, initial disclosures are due at the latest 83 days after the service of the complaint or 53 days after a defendant has appeared, whichever is earlier.[10]

Considering the legal analysis of the data that must be done to comply with the default initial disclosure deadlines, predictive coding can help identify data for Rule 26(a) initial disclosures that supports a party's claims or defenses with prediction models based upon the allegations in the complaint.

Keyword searches can be created focused on the scope of the date range, between specific individuals, and known relevant subject matter. For example, in a commercial construction dispute, the broadest search term could be the name of the project. Moreover, most lawsuits have an incident that is the basis for the litigation, such as an accident or failure to perform contractual obligations. Those events can be the basis for a search in Everlaw, which might look like this:

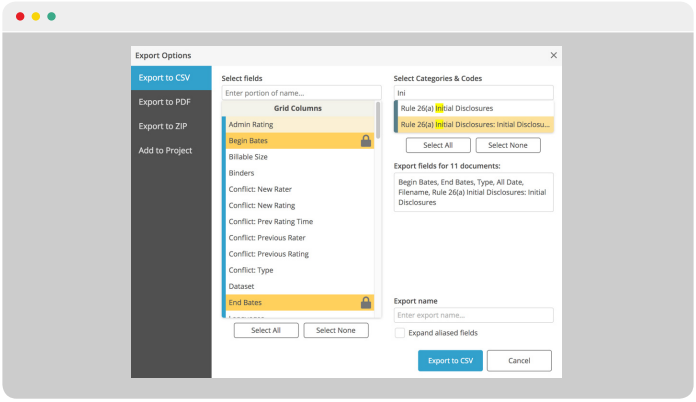


The documents that are the result of this search can be reviewed, and those review decisions can then be used to train a prediction model designed to identify potential initial disclosures. Documents that support a party's claims or defenses can be issue coded as "Initial Disclosures" and fed into a prediction model with the following relevance criteria:



8 Fed. R. Civ. P. 26(f)(1).
9 Fed. R. Civ. P. 16(b)(2).
10 90 days – 21 days = 69 days + 14 days = 83 days after service of complaint OR 60 days after a defendant has appeared – 21 days = 39 days +14 days = 53 days after a defendant has appeared.

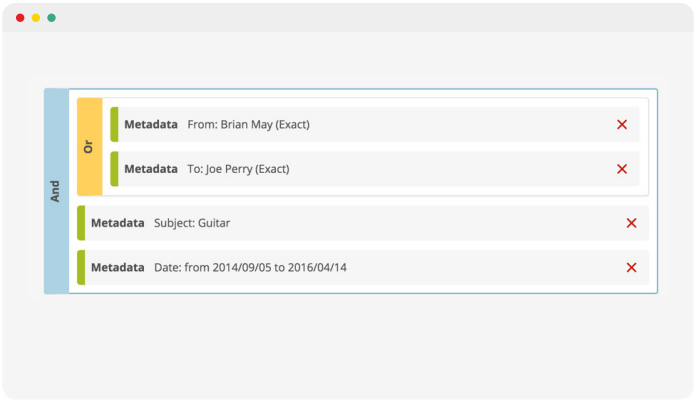
Once the model has had enough review to give predictions, unreviewed documents above a certain prediction threshold—a limit on inclusiveness between 0 and 100—can be reviewed to determine whether they, too, should be included in the Rule 26(a) initial disclosures. Once all of the documents to include in the initial disclosures are identified, it's easy to generate a report out of Everlaw:



Responding to Discovery Requests

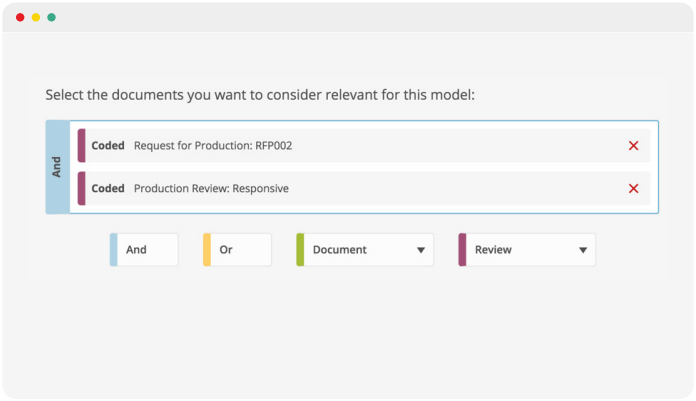
Documents responsive to Requests for Production can be identified with a combination of search terms and predictive coding models. The first step is to create issue coding for categories of the opposing party's requests for production, or for specific requests if the number of discovery requests is low. The coding protocol would then call for each document to be coded as to its responsiveness to each request category.

Initial review would focus on documents responsive to search terms likely to yield responsive documents. A project manager can create a search for each request for production, which can then be assigned to attorneys for review. One such search may look like this:



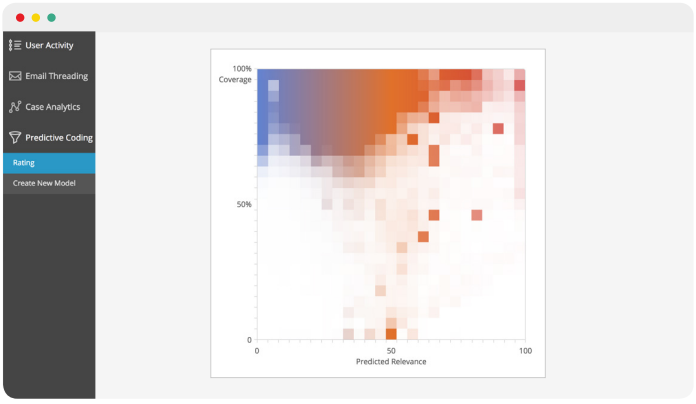
Attorneys can be assigned search hits for review by the project manager. Issue codes can be created to denote which request the hits are responsive to. This can allow for the creation of an index of responses, if necessary.

A predictive coding model can then be created to use the initial review as training input, helping to predict which unreviewed documents are likely to be responsive. The model's relevance criteria would look like this:



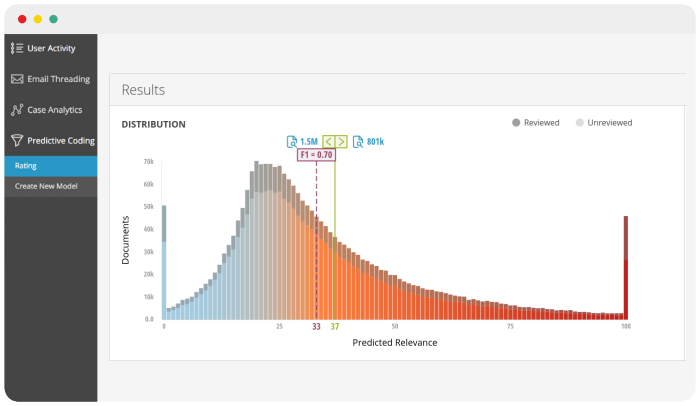
As with any review with search terms, "trust but verify" is part of standard operating procedure. The reviewers should sample hits without the prediction engine to see if there are other responsive records.

The project manager can also use the coverage chart to determine if there are clusters of documents that are not like the documents that have already been reviewed. Reviewing these documents is an excellent way to improve the accuracy of the prediction model. In this chart, those documents are the ones generally clustered towards the bottom and middle:



Once these pockets of idiosyncratic documents have been addressed, and confidence in the prediction scores is high, a sample of unreviewed data can be assigned to review attorneys to validate whether the data is relevant or irrelevant to the case. If the data shows potentially responsive information, the project manager can assign the data for review. If irrelevant, the data can be bulk issue coded and culled from the review sets, so time is not spent on reviewing irrelevant information.

The distribution chart provides good information about where the most and least promising documents are likely to reside:



Supplemental Discovery Productions

If litigants learn their production or initial disclosures are incomplete or incorrect, they have a duty to supplement products pursuant to Rule 26(e).[11] Parties may therefore realize the need for a supplemental production while conducting review.

In a case where a producing party had to make a supplemental production, predictive coding was used to focus review. The first production consisted of 6,000 records. Additional collection increased the review database by 130,000 records.

The first production was used to train a prediction model for what was relevant in the lawsuit. This was a commercial construction case where the producing party over-collected ESI. The other irrelevant construction projects and clients were identified by search terms and issue-coded to be irrelevant and rated Cold.

The prediction model identified 37,000 records out of 130,000 as potentially responsive, based on the other documents already produced. Search terms devised from two requests for production were applied with the prediction model. With only four days of work, a supplemental production of 1,000 records was made to the requesting party, saving many months and tens of thousands of dollars in unnecessary review.[12]

Does a Party Need Prior Approval to Use Predictive Coding?

Litigants are strongly encouraged to have a meaningful Rule 26(f) conference with the opposing party. The first meet and confer should define the scope of discovery, which includes identifying the relevant time period, relevant individuals to the lawsuit for data preservation, sources of relevant information, and terms of art for searching the

11 Fed. R. Civ. P. 26(e)(1)(A).

12 Assuming that one attorney could review two records a minute, thus 120 records an hour, thus 960 records in an eight-hour day, it would have taken one lawyer approximately 135 days to complete review. If that lawyer billed $150 an hour, the project would have taken around 1,080 hours, thus $162,000. While these numbers can vary based upon billing rates and time to finish the project, completing document review in four days is significantly less expensive that over 160 days of billing.

data. Rather than committing to specific "search terms" that are a binding agreement, a good strategy would be for the parties to discuss "search concepts," as there are many ways concepts can be communicated in language.

There are many cases where litigants have asked the Court for permission to use predictive coding in a case.[13] This is effectively asking a Court to give an advisory opinion, because there is no controversy for the Court to resolve, such as a claim a production was inadequate.

In Rio Tinto Plc v. Vale S.A., where review costs were estimated to be $450,000, the parties sought judicial approval to use predictive coding. Judge Ronald Buch stated the following on the issue of using predictive coding:[14]

> *"And although it is a proper role of the Court to supervise the discovery process and intervene when it is abused by the parties, the Court is not normally in the business of dictating to parties the process that they should use when responding to discovery. If our focus were on paper discovery, we would not (for example) be dictating to a party the manner in which it should review documents for responsiveness or privilege, such as whether that review should be done by a paralegal, a junior attorney, or a senior attorney. Yet that is, in essence, what the parties are asking the Court to consider–whether document review should be done by humans or with the assistance of computers. Respondent fears an incomplete response to his discovery. If respondent believes that the ultimate discovery response is incomplete and can support that belief, he can file another motion to compel at that time. Nonetheless, because we have not previously addressed the issue of computer-assisted review tools, we will address it here."*

The central issue in producing discovery is whether the production itself is adequate. In this light, there is no legitimate reason to hold predictive coding to a different standard than manual review or keywords.[15] Asking a Court to allow the use of predictive coding is akin to asking for permission to use a specific legal research tool. Moreover, as the Court in the Rio Tinto opinion explained, holding predictive coding to a higher standard "discourages parties from using TAR for fear of spending more in motion practice than the savings from using TAR for review."[16]

The standard for producing ESI is governed by reasonableness and not one of perfection.[17] Courts recognize the producing party is in the best position to determine how to search for responsive ESI.[18] The issue for producing parties is whether they have "conducted a search reasonably calculated to uncover all relevant documents."[19]

---

13 See, *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182, 184 (S.D.N.Y. 2012); *United States v. ExxonMobil Pipeline Co.*, No. 4:13-cv-00355 KGB, 2014 U.S. Dist. LEXIS 81607, at *6-7 (E.D. Ark. June 9, 2014).

14 *Dynamo Holdings v. Comm'r*, 2014 U.S. Tax Ct. LEXIS 40, at 10-11 (Docket Nos. 2685-11, 8393-12. Filed September 17, 2014).

15 *Rio Tinto Plc v. Vale S.A.*, 2015 U.S. Dist. LEXIS 24996, at *10 (S.D.N.Y. Mar. 2, 2015).

16 *Id.*

17 *Chen-Oster v. Goldman, Sachs & Co.*, 285 F.R.D. 294, 306 (S.D.N.Y. 2012), citing The Sedona Conference, The Sedona Conference Database Principles: Addressing the Preservation and Production of Databases and Database Information in Civil Litigation, March 2011 Public Comment Version, at 32.

18 *Hyles v. New York City* 2016 U.S. Dist. LEXIS 100390, at *7-8, (S.D.N.Y. Aug. 1, 2016) citing Sedona Principle 6.

19 *Mullen v. United States Army Crim. Investigation Command*, 2012 U.S. Dist. LEXIS 93977, 12-15 (E.D. Va. July 6, 2012), citing *Rein v. United States PTO*, 553 F.3d 353, 362-63 (4th Cir. Va. 2009).

This raises the issue: what is a production that is not reasonable? A producing party should be concerned their production is not reasonable, for example, if keyword searches show over 101,000 documents with hits, but result in a production of only 109 emails.[20] Moreover, if a production consists of only 25 email messages, the producing party should be prepared to explain their sources of ESI and searches for responsive ESI.[21]

If a production is inadequate, the requesting party should first attempt to meet and confer over the production pursuant to Federal Rule of Civil Procedure Rule 37(a)(1). If that attempt to resolve the dispute fails, the next course of action is to bring a motion to compel pursuant to Rule 37(a)(3)(B)(iv). Nowhere in the Federal Rules of Civil Procedure does it say a party using predictive coding must seek permission from the opposing party before conducting document review.

Conclusion

Predictive coding can be used tactically when conducting a reasonable inquiry for Rule 11 compliance, leveraged to identify Rule 26(a) initial disclosures, and in responding to discovery requests. Producing parties are in the best position to determine how they should search for electronically stored information. While it is a good best practice to agree upon the scope of discovery, key players, subject matter, and other objective criteria for searching for ESI, the issue for every production is whether the producing party did a reasonable search. Tactically using predictive coding can empower attorneys to meet their production responsibilities and save time conducting review.

20 *Blumenthal Distrib., Inc. v. Herman Miller, Inc.*, 2016 U.S. Dist. LEXIS 184932, at *41 (C.D. Cal. July 12, 2016).
21 *Am. Home Assur. Co. v. Greater Omaha Packing Co.*, 2013 U.S. Dist. LEXIS 129638, at *17 (D. Neb. Sept. 11, 2013).