Consilio
Global Leader in Legal Consulting and Services

Success with Predictive Coding:
Eliminating 89% of the Review with Predictive Coding

Company Profile
Large Global Firm

Matter Summary
Opposing counsel requested broad set of search terms resulting in an exceptionally high volume of documents for review.

Services Utilized
• Discovery Consulting

• Document Review

• eDiscovery Analytics

Client Challenge

In a civil litigation matter, a large global law firm collected and processed more than 1 million documents on behalf of the defendant. After processing the data, the firm and opposing counsel agreed to isolate relevant documents in the corpus using keyword searches, but they could not agree on the terms to apply. While defense counsel proposed a narrow set of search terms to limit the amount of review needed, opposing counsel insisted on a broader set that would result in a tenfold increase in documents needing review and a dramatic increase in the defendant's cost. Defense counsel believed opposing counsel's proposed keyword list to be exceptionally broad, requiring review of tens of thousands of documents irrelevant to the issues of the case. Recognizing the need for assistance to manage the situation, defense counsel called upon Consilio for advice on how to proceed.

Consilio Response

After engaging Consilio, our client saw an opportunity to compromise: agree to opposing counsel's broader keyword list if counsel would agree to using predictive coding to cull the set of documents prior to review. Because the attorney team at our client had little experience with predictive coding, Consilio provided consultative guidance, offering extensive insight into the technology, the workflow processes, the quality-assurance routines and the proper semantic verbiage specific to the technology that so many in the industry find confusing. This conveyed knowledge enabled our client to confidently propose the use of predictive coding to opposing counsel.

As opposing counsel were similarly inexperienced in the use of predictive coding, they were reluctant to agree to use this tool. Consilio's experts participated in conferences between both parties, bridging the partial understanding of each and helping our client successfully leverage its adversary's requirements to its own advantage. Ultimately, with the average cost to perform first-level document review at roughly $1.25 per document, the client saved at least $209,000 in variable costs as well as days or weeks under Consilio's guidance, the parties agreed to a detailed, defensible workflow that outlined recall thresholds, the allowable confidence level, the format for reporting results and the quality-assurance protocols. Consilio's thoughtful participation ensured that both sides became comfortable with every issue important to the predictive coding protocol, including communication, timing and sharing of results.

After obtaining the parties' consent to proceed, Consilio commenced the predictive coding workflow with the goal of winnowing down the near 200,000 keyword-culled document corpus to the likely responsive set of documents with a target recall of 90% ±5%. Moreover, because minimizing review cost was paramount but speed to production was also important, the firm desired as high a precision as possible in as little time as possible.

Training Round 1: Perform Initial Training of the Software with Randomly Drawn Documents

After doing a randomly drawn prevalence test to right-size the control set, Consilio assigned a random sample of 2,057 documents to a senior associate who was experienced with matter and the issues to code the documents for responsiveness. Of those documents, the senior associate found that 140 were responsive, meeting opposing counsel's margin of error threshold for responsiveness. After training the computer models on these coding decisions, Consilio was able to immediately set aside 162,000

documents as likely nonresponsive, leaving a pool of only 33,000 likely responsive documents—an immediate avoidance of 83% of the keyword-culled documents.

Further, the control set now informed the attorneys that the prevalence of responsive documents in the corpus was around 7%, which confirmed our client's suspicion that opposing counsel's expanded keyword list was unfocused and merely served to bloat the defendant's expense in the litigation.

Round 2: Refine the Predictive Coding Software's Precision and Reduce the Consideration Pool

These documents were "biased" because they were selected from the narrow keyword list our client originally proposed. Surprisingly, after training the computer model with these 7,000 biased documents, the computer model was barely more refined: the number of potentially responsive documents shrunk by 2,000, from 33,000 to 31,000—a slim improvement to the model's precision. Because of this unexpected result, Consilio began to suspect that our client's coding decisions on those 7,000 biased documents were inaccurate. Consilio's expert project managers discussed the problem with the senior associate and uncovered that he had been coding families without reviewing each document on its face value for responsiveness, which explained why the modeled algorithms made little progress. The common practice of coding family documents without reviewing them on their face value content—which is often done for efficiency in eDiscovery—confuses the software, which relies on concepts within each document as the primary determinants of responsiveness.

Round 3: Limit the Pool and Create a Parallel Workstream

To reduce the size of the consideration pool of documents even further, Consilio recognized it needed more responsive exemplar documents. This time, the senior associate reviewed an additional, randomly drawn 857 documents. After coding, these documents were fed into the software to advance its training, resulting in the elimination of another 6,000 documents from the consideration pool. Now, our client only had to review approximately 25,000 documents, one eighth of the original pool of 200,000.

At this point, Consilio advised our client of another way to accelerate the review toward production: adding a parallel workstream. Consilio recommended that the firm's junior attorneys begin the second-level review of documents in the corpus identified as very likely responsive (with responsive probability scores above 80% likelihood) to further expedite the review toward production. The client accepted Consilio's recommendation and was able to produce documents one day sooner than it would have been able to otherwise, alleviating the stress of meeting a production deadline.

Round 4: Resolve Discrepancies to Yield Final, Defensible Review Set

In the final stage of training, Consilio was able to further reduce the consideration pool. This time, instead of doing another random grab of documents or a biased draw of documents, Consilio recommended a "disagreement reversal" and asked the senior associate to re-review 108 of the documents where his responsiveness decision conflicted most significantly with the computer's probability score. For example, one of the documents asked to be re-reviewed had a computer-generated probability of less than 1% of being responsive, yet the senior associate had tagged the document as responsive.

Although the first round of training had achieved the recall target of 90% ±5%, Consilio saw opportunities to improve upon the first round's relatively low 36% precision score. Moreover, our client wanted to refine the algorithm to continue shrinking the "consideration pool" (i.e., documents the software had not identified as likely nonresponsive) further to decrease costs and review time. In the second training iteration, Consilio leveraged coding decisions from about 7,000 documents that our client began reviewing prior to the agreement to use predictive coding.

Upon re-review, the senior associate overturned 72% of his own coding decisions. This relatively high reversal rate reflected the senior associate's decision to code documents as a family without looking at each document. This inconsistency explained why the second iterative training round made only marginal improvements to the precision score: the variability of the human review, which was based on subjective

factors, paled in comparison to the consistency of computer review, which was based on mathematical formulas. Because of this round of disagreement reversals, the consideration pool of documents for review shrunk to a very manageable 21,500 documents, in which 90% ±5% of the expected 15,000 responsive documents in the corpus would be found. At this point, Consilio advised our client that any continued training of the predictive coding software would see diminishing returns.

Given that our client was reporting excellent results with the parallel workstream set in motion after the third training round, our client agreed that we should stop further model training and proceed to final scoring of all documents in the corpus.

Services Offered

eDiscovery Consulting, Services & Analytics

Document Review

Risk Management & Compliance

Data Forensics & Investigations

Law Department Management

*For a full list of services, visit: consilio.com*

Results Achieved

As this case study reveals, Consilio's project management skills and technical expertise enabled our client to minimize the resources needed to create a defensible and efficient workflow, setting aside 89% of the original document corpus from consideration using predictive coding. More importantly, with Consilio's guidance, our client was able to curtail its work by looking at a very limited pool of randomly drawn documents—fewer than 3,000 —and leveraging the prior coding decisions of attorneys who had already manually reviewed a subset of the documents based on keyword hits.

Had our client proceeded on its own and agreed to its adversary's proposal, the lawyers would have looked at all 200,000 documents, delaying second-level review and incurring cost and delays without improving confidence in the review's quality. With the average cost to perform first-level document review at roughly $1.25 per document, the client saved at least $209,000 in variable costs as well as days or weeks.

Our client also recognized that self-service predictive coding software may offer the promise of some of the advantages of machine learning, but its usage presupposes proficiency and comfort with predictive coding and statistics. Without a predictive coding expert guiding the way, parties are likely to face one of two undesirable outcomes: either (1) failing to reach agreement to leverage technology, spending excessive amounts of time and money doing unnecessary review or (2) agreeing to a protocol that turns out to be unworkable, wasting money, causing frustration for both sides and leading to further costly disputes and sanctions. Engaging a predictive coding expert early in discovery can help translate the tools and jargon of an unfamiliar technology into a common language that serves both parties' legitimate goals in discovery.

Global Reach, Local Expertise
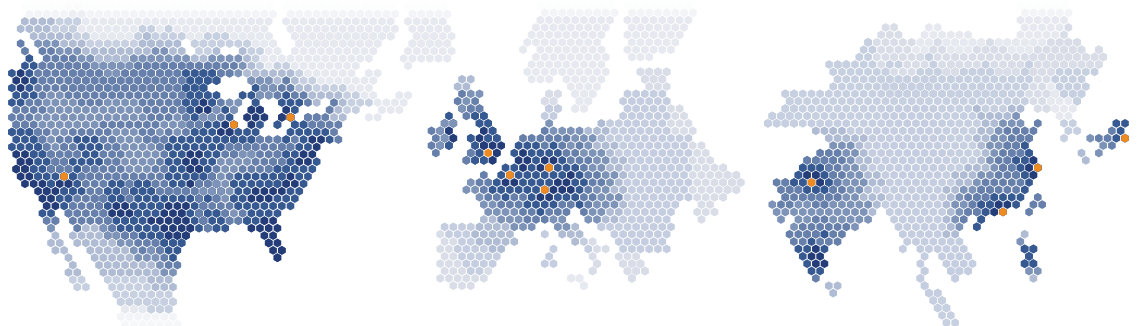
1,700+
Employees
Worldwide

20+
Countries Worldwide
with Staffed Review
Professionals

31+
Languages Covered
in Review Projects
Globally

13+
Data Centers
Around the World

Ready to Learn More?
Visit consilio.com or email info@consilio.com