Most Important Documents Get Looked at First:
Using Predictive Coding to Prioritize & Expedite Review

Executive Summary

In a recent pro bono case, a global 50 law firm representing the plaintiff faced the challenge of reviewing a produced set of about 31,000 documents quickly to prepare for depositions with limited review resources. Counsel began to review the documents linearly but realized that the deadline could not be met without a more expeditious prioritization approach. Because all produced documents contained keywords, the firm asked Consilio to provide an alternative prioritization approach that was fast, efficient and effective at finding the most relevant documents.

Consilio leveraged its predictive coding solution to stack-rank documents based on likelihood of relevance to each of 12 independent issues. This approach enabled counsel to review the highest-value documents first, speeding the deposition prep work and bypassing review of thousands of marginal or irrelevant docs. Then, after depositions were completed, Consilio re-ran the predictive coding workflows for quality assurance and to educate the attorneys about what the process would have looked like had predictive coding been used from the outset. That exercise showed that predictive coding could have saved, conservatively, 200 review hours (or 70%) as compared to linear review, and it would have yielded more accurate results earlier in the process.

Background

A law firm serving as plaintiff's counsel in a pro bono litigation matter faced a common dilemma: with a short deposition deadline looming, it had to review multiple document productions from defense counsel. The attorneys began this endeavor with linear review (document by document, without benefit of prioritization, categorization, or other technological assistance) of all produced documents. However, after reviewing roughly half of the collection, the attorneys found that 60 percent of the documents were not relevant to any of the 12 issue tags. Thus, more than half of the documents being reviewed by the firm were irrelevant to the case ("noise"), compounding the firm's challenge of finding the most important documents under tight deadlines.

> **"** *That exercise showed that predictive coding saved, conservatively, 200 review hours (or 70%) as compared to linear review...* **"**

Phase 1: Prepare for Depositions by Getting to the Key Documents

With linear review unable to keep pace with rapidly approaching deposition dates, counsel sought expert advice on a more expeditious method. The firm asked Consilio to recommend a solution that would achieve three goals:

1. Find the produced documents most meaningful for depositions that had not yet been reviewed,
2. Keep control of costs, and
3. Meet the existing deposition schedule

To meet the firm's goals, Consilio conceived a strategy to leverage predictive coding to quickly find documents in the not-yet-reviewed pile that were likely relevant to any of the 12 issue tags. First, Consilio used the documents that counsel had already reviewed as the seed set to train the computer models. In this step, Consilio optimized 12 independent predictive computer models: one for each issue tag. Next, Consilio used these 12 computer models to score each of the approximately 13,650 documents in the not-yet-reviewed pile on an issue-by-issue basis, which yielded 12 stack-rankings, each based on the document's likelihood to be relevant to each issue. Then, Consilio conferred with counsel on the generated document scores and collectively agreed to assign the 1,200 highest-scored documents to counsel to review. Upon review of these 1,200 documents most likely to be relevant to one or more issues, counsel agreed that these documents exemplified characteristics of the most relevant documents and that the application of predictive coding in this manner met the firm's goals.

In this case, predictive coding workflows and technology helped attorneys look at the documents most likely to be relevant to the core issues of the case quickly (within three days of engaging Consilio) and with minimal wasted time or effort. Further, this solution stack-ranked the documents by issue tag, so counsel could review the documents most relevant to specific witnesses.

Consilio

Phase 2: Ensure Review Quality and Understand How Predictive Coding Could Have Helped From the Beginning

After the first wave of depositions was completed, the firm had more time and wanted to ensure that it had uncovered all relevant documents that were not identified in time for the depositions. The firm also wanted to perform a quality-control sweep of the already-reviewed documents to assess the solution's appropriateness as a quality control tool. In addition, the firm's attorneys were interested in understanding what the workflow would have looked like had Consilio's predictive coding solution been used from the outset in lieu of linear review.

In this second phase, Consilio ran a predictive coding workflow from scratch as though partial linear review had not taken place. Consilio first took a random sample of the entire corpus—about 1,400 documents—to create the control set for the predictive coding software. Consilio used previously applied coding for those of the 1,400 documents that had been reviewed, and counsel reviewed those docs which had not previously been reviewed. After optimizing the computer models with the control set, Consilio selected 25 "disagreement documents" for each issue tag, or 300 total disagreement documents. These documents were ones that the attorney coded relevant but the computer believed to be irrelevant and vice versa. Counsel re-reviewed these disagreement documents and decided to overturn the original coding on 25 percent of these documents. Consilio fed these reversals back into the predictive coding software and re-optimized the models to generate final scores for each issue tag. Through this process, Consilio identified an additional 1,600 documents in the corpus likely relevant to one or more issue tags that counsel had not yet reviewed because they were not yet reviewed in the preceding linear review, and the Phase 1 deposition readiness review didn't review deeply enough. After review, counsel noted that many of these documents were also useful to support legal arguments.

The results of this phase emphasize how predictive coding software delivers the consistency that is typically missing from linear review. The volume, time, and complexity challenges of reviewing 31,000 documents concurrently for 12 issues—creating more than 360,000 decision points that threaten to boggle even the most adept human mind—were minimized by the technology, which dramatically lowered the number of decision points the attorneys needed to make. The fact that one-quarter of the disagreement documents were overturned on re-review by the same attorney reveals how human review is subject to inconsistency. By reviewing only 1,400 documents, the predictive coding software could have helped the attorneys eliminate considerable amounts of irrelevant document noise that was littering defense counsel's production.

Recommendations and Conclusions

As this case study demonstrates, predictive coding software lends itself to a number of applications, including classification and quality control – and in this case, document prioritization. The Consilio solution allows clients to stratify resources and focus attention on the documents that are most likely to be relevant soonest, which is critical to attorneys who are working to beat tight deadlines and who have limited review bandwidth. In this matter, had counsel used predictive coding from the outset, counsel would have significantly limited the scope of review. With only one round of optimized training of the computer model and one round of disagreement review, counsel would have only needed to review 30 percent of the full corpus – including training documents, disagreement documents and likely relevant documents – eliminating 70 percent of the corpus from consideration. Even with this modest-sized document set, predictive coding could have saved 200 hours of attorney review time, presuming a review rate of 100 documents per hour, and with larger data sets, these savings would have been exponentially larger.

Comparison of Review Hours Needed

Bar chart comparing review hours: "Without Predictive Coding" (~31,000) and "With Predictive Coding" (~11,500). Y-axis marked 10,000; 20,000; 30,000; 40,000.

Consilio™