SUCCESS STORY

Major food & beverage manufacturer

Industry
- Food & beverage

Solution
- OpenText™ Insight Predict

Results

⚙ Achieved an **estimated recall of 75 percent**

📄 **Found relevant documents faster,** easier and cheaper using CAL

👍 **Met tight 30-day production deadline**

Major food & beverage manufacturer serves up high-efficiency review

Corporation almost doubles recall with continuous active learning (CAL) approach from OpenText Insight Predict

CAL achieves on average a 2:1 efficiency ratio compared to 10:1 rates for keyword search and other TAR systems.

A leading food and beverage manufacturer faced a potential class action related to allegations of deceptive food package labeling and ingredients. The client and its law firm needed to review approximately 175,000 documents and produce the relevant documents to the plaintiff within 30 days to determine whether the litigation would be certified as a class action.

Even when inefficiencies and coding mistakes initially hindered the review process, OpenText™ Insight Predict corrected for inefficiencies and significantly outperformed other search and review methods with a continuous active learning (CAL) approach.

Continuous active learning: CAL continuously integrates the judgments made by the review team and then uses those judgments to constantly re-rank the entire document collection. Unlike earlier generation TAR systems based on simple passive learning (SPL) and simple active learning (SAL), no control set is needed, thus eliminating the time and costs associated with a senior lawyer reviewing thousands of documents to build a control set, train and then test the results. CAL also accommodates rolling uploads without requiring a re-training of the system, thus saving even more time and money compared to other TAR tools.

High efficiency: CAL can achieve, on average, a 2:1 efficiency ratio. In the perfect world of a simulation, a lawyer would have to review two documents to find one relevant document during the course of the review. Efficiency is another way to look at precision; an efficiency rate of 2:1 equates to 50 percent precision. Keyword search, and even other TAR systems based on active learning, typically achieve an average of 10:1 efficiency rates.

Even when inefficiencies and coding mistakes initially hindered the review process, OpenText Insight Predict corrected for inefficiencies and significantly outperformed other methods.

Contextual diversity: The integration of contextual diversity into CAL, which is unique to Insight Predict, ensures that legal teams do not miss pockets of potentially relevant documents. The contextual diversity algorithm identifies documents based on how significant and how different they are from the ones already seen, and then selects exemplar documents that are the most representative of those unseen topics for human review.

Before integrating a CAL approach, the law firm for the food and beverage manufacturer reviewed the first 8,500 documents that had been prioritized using key custodian, date and search terms. During the course of its priority review for production, however, the firm faced a number of obstacles.

Because the attorneys had not used CAL on previous cases and were thus unfamiliar with how it works, they determined that they needed to review each and every one of the 8,500 documents based on the belief that they could use this set to "seed" the system for further review later on.

Also, the coding itself was conducted by a review team that had not been properly trained on the subject matter and issues related to the case. In addition to miscoding nonrelevant documents as relevant, the reviewers also missed significant pockets of unreviewed information. Further, the client did not implement recommended QC process to check the attorneys' coding calls. Unfortunately, that decision meant that the coding errors were not identified until the second phase of the review.

During the second phase, an additional 165,000 documents were loaded into the Insight Predict system. In just five days, an additional 3,640 documents were identified as relevant. When the batches started consistently containing few relevant documents, the team took another random sample for an updated recall estimate. This sample showed the team had only reached an estimated recall of 40 percent— substantially below its goals for a final production. It was at this point that the review team and OpenText analysts and data science experts discovered the coding discrepancies that had resulted from lack of reviewer training. The team "paused" for retraining, and then the third phase of the review commenced.
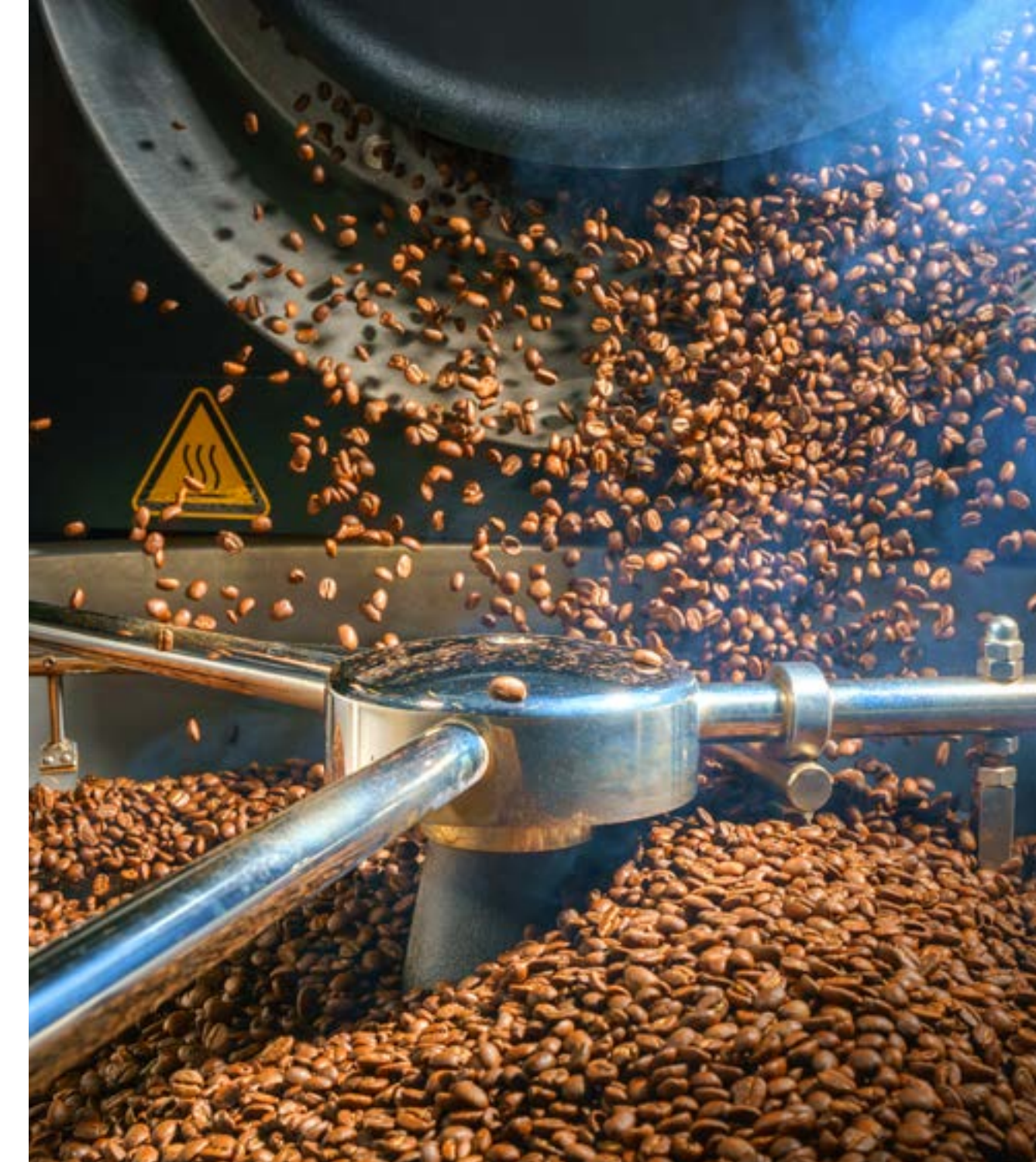
OpenText simulations illustrated the Insight Predict algorithm can quickly self-correct after large-scale coding revisions, and a review does not need a large seed set coded by a subject matter expert to jumpstart the machine learning. In fact, the simulations showed that the review would have been slightly more efficient had the team started off by coding a 100 document random sample instead of the targeted 8,500 document set.

Back on track, the review team found an additional 5,400 responsive documents. At the point in the review when the Insight Predict batch richness, or precision, was less than 10 percent for two consecutive days, the team determined it should stop the review. A final sample indicated 89.4 percent recall.

By using Insight Predict to prioritize a document review, review teams found more relevant documents more quickly, with less effort and at substantially lower cost. Even with the inefficiencies introduced by the client's workflow and inadequate reviewer training, Insight Predict produced exceptional results. The teams reached the 80 percent recall level after reviewing only about 36,000 documents with Insight Predict, whereas with a standard linear review, they would have had to review over 130,000 documents to reach that same recall level.

The client's use of CAL to accelerate the review resulted in the following:

- Comparatively high efficiency rate of 3.1:1 at 75 percent recall despite inefficient workflow in the first two phases of the review.

- Estimated recall of 89.4 percent for responsive production.

- All production timelines met.