



CASE STUDY

Using TAR 2.0 to Expedite Multi-Language Review

How Insight Predict's Unique Capabilities Cut Review by Two Thirds

“ *The case illustrates two key capabilities that distinguish TAR 2.0 platforms such as Insight Predict from earlier TAR tools.* ”

Client Snapshot: Global Corporation

- Major shareholder class action
- Tight deadline to review dual-language doc set
- Process finds critical docs in spreadsheet files
- Review time cut by two-thirds



In a major shareholder class action alleging violations of federal securities laws, the defendant's legal team was under a tight deadline to review a collection of mixed English and Spanish documents. By prioritizing documents using Insight Predict—Catalyst's engine for technology assisted review—the team was able to cut the review by two-thirds, saving time and money.

The case illustrates two key capabilities that distinguish TAR 2.0 platforms such as Insight Predict from earlier TAR tools. One is that judgmental seeds, selected by the trial team, were used to train the system, rather than the random seeds used by earlier tools. The other is that Insight Predict's unique contextual diversity sampling helped discover critical information "hidden" in spreadsheet files.

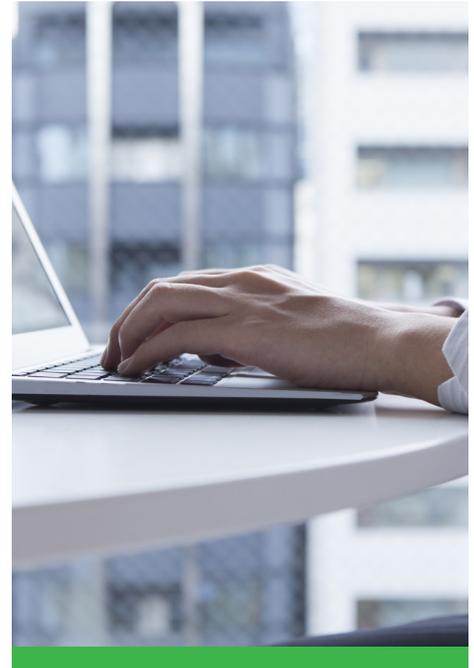
The Challenge: Quickly Review Multi-Language Documents

Our client, an international provider of products and services, was sued by shareholders in a federal class action alleging violations of the Securities and Exchange Act. Its legal team faced an imminent deadline to produce responsive documents from a collection of 66,000 files. The collection included emails, documents and spreadsheets in both English and Spanish. To save time and money, the team wanted to minimize the number of documents that would require eyes-on review.

The Solution: Prioritize Documents for Review Using Insight Predict

The team decided that the best approach would be to rank documents for review using Insight Predict. The attorneys had already reviewed a few hundred emails that had been hand-picked for them by two key custodians, so we started the ranking process using those as seeds to initially train the Predict engine.

For the two languages, we created separate rankings. Then, as Predict continuously ranked the documents, on-demand batches were sent to the review team from the top of the most-recent ranking round. Because Predict is fully integrated with the Insight Discovery platform, we set up rules to send the batched documents automatically.



By prioritizing documents using Insight Predict—Catalyst's engine for technology assisted review—the team was able to cut the review by two-thirds, saving time and money.

An initial random sample of the collection indicated it was only 10% rich. But after training predict with the initial seeds, the richness of the batches being automatically fed to the review team jumped by a factor of four, to 40%.

By using Predict to prioritize all the responsive documents at the front, the team was able to stop the review after reviewing only one third of the documents. At that point, they had achieved 91% recall—much better than the 80% recall expected from full human review. Even though they put their eyes on every document they produced, they were still able to cut their review by two thirds.

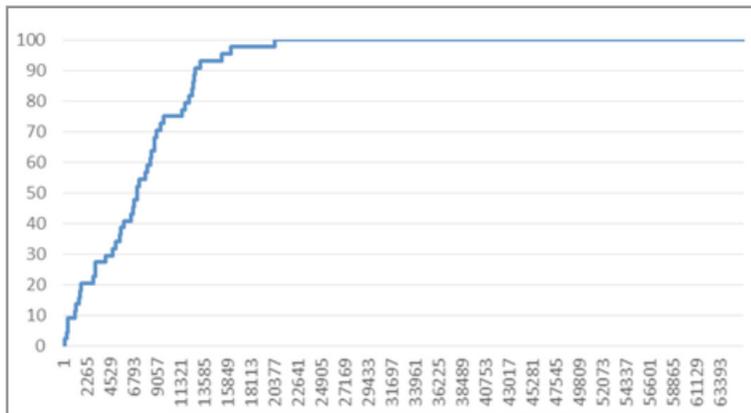
Uncovering a ‘Hidden’ Trove

A potential pitfall of judgmental sampling is bias. When lawyers hand-pick the documents used to train the TAR system, there is the risk that the system will overlook relevant documents they did not know to train it for.

Insight Predict overcomes this potential bias through a unique form of fail-safe called “contextual diversity sampling.” Predict is the only TAR engine that uses it. As Predict runs, it is constantly looking for the next biggest pocket of documents that have no reviewer judgments associated with them. From those pockets, it automatically finds the best example documents and feeds them directly to the review team for tagging.



“When lawyers hand-pick the documents used to train the TAR system, there is the risk that the system will overlook relevant documents they did not know to train it for.”



Yield curve after contextual diversity (textual docs only)

In this case, contextual diversity sampling revealed a significant pocket of several hundred financial spreadsheets that were unlike any of the



documents that were in the first set that they looked at. As soon as the review team moved from the manual seeds to the automated samples that included contextual diversity, this pocket of documents was found and the performance of the Predict ranking increased significantly.

The Bottom Line

Even with a dual-language document set, Insight Predict was able to sort every responsive text document to the top third of the ranked list for review. For the corporation's legal team, that meant two-thirds fewer documents to review. Insight Predict significantly reduced both the time and the cost of the review and enabled the team to meet their tight deadline for production.



Learn More
catalystsecure.com



Follow Us
[@catalystsecure](https://twitter.com/catalystsecure)