



## CASE STUDY

# Predict Proves Effective Even With High Richness Collection Finds 94% of the Relevant Documents Despite Review Criteria Changes

“ The team achieved a remarkable review efficiency and looked at very few nonresponsive documents. ”

### Client Snapshot: Global Corporation

- Federal investigation into alleged price fixing
- 2 million documents to review in just four months
- Highly responsive collection with changing review criteria
- Client saved \$550,000 by using our technology and review teams



## Background

Our client, a global corporation, was hit with a federal investigation into alleged price fixing.<sup>1</sup> The regulators believed they would find the evidence in the documents. The request to produce was broad, even for this three-letter agency. Our client would have to review over 2 million documents. And the deadline to respond was short, just four months to get the job done.

This wasn't a case of finding a needle in a haystack. Rather, a wide range of documents were responsive. A sample of the initial collection suggested that as many as 45% of the documents would be responsive. One option was to produce everything but the client had a lot of confidential and proprietary information that it didn't want in the hands of competitors or the public. The assignment was to produce responsive documents but only responsive documents.

## Making Review Efficient

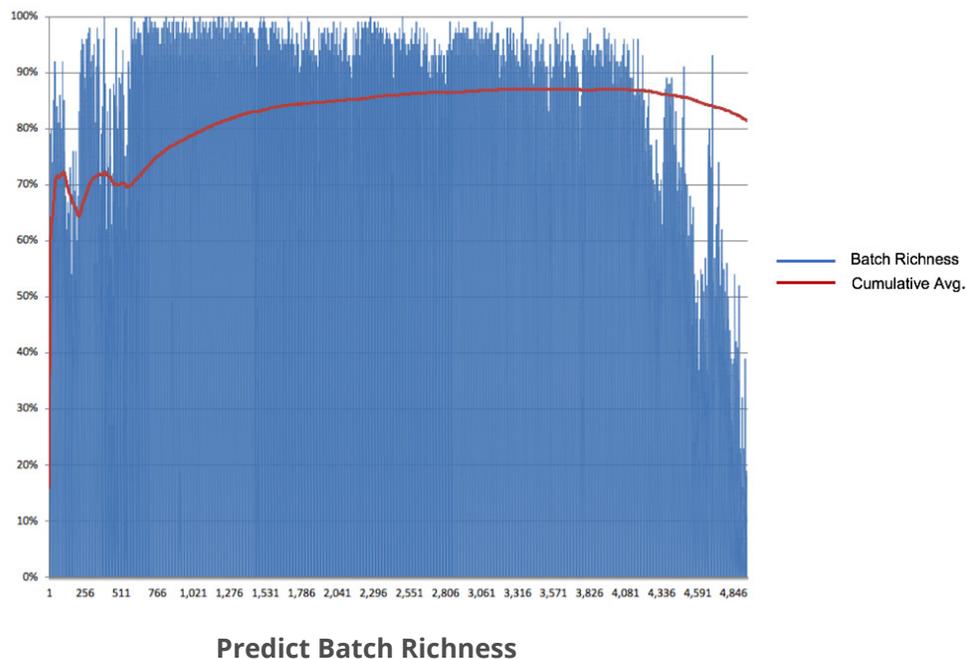
Our goal was simple. Use Insight Predict, our TAR 2.0 continuous active learning (CAL) algorithm, to find the relevant documents as quickly and efficiently as possible. We matched our engine with the Catalyst Review Team, a bunch of well-trained Predict ninjas who are versed in getting the most out of our software.

We started using CAL from the beginning, no waiting around for a senior lawyer "subject matter expert" (SME) to wave hands of the document under the guise of training. Rather, the team got started right away using the responsive documents already identified for initial training.

The measure of a predictive review is how quickly the algorithm can surface relevant documents. Like a bloodhound born to track, Insight Predict picked up the trail almost immediately.

Here is a chart showing the percentage of relevant documents Predict found on a batch by batch basis.

There were almost 5,000 batches in this review. Each blue line represented the percentage of relevant documents in the batch. In the early stages, the number reached 80% to 90%. That meant the reviewer



1. Party and claim facts have been changed to preserve client confidentiality. Our goal is to show the power of Insight Predict and not to comment about the specifics of any matter.



found 80 to 90 responsive documents out of 100 in their batch. It also meant the reviewer saw only a few nonresponsive documents, which was our ultimate goal. Make the reviewers as efficient as possible, to keep review costs as low as possible.

## Review Efficiency

This is an important but seldom-discussed topic. How many nonresponsive documents does the reviewer see for each responsive one?

When keywords are used to cull documents, reviewers typically have to look at as many as nine nonresponsive documents for each responsive one. That means they are wasting their time for about 90% of their review efforts. With Insight Predict, our statistics show that the ratio is much narrower, about 2 to 1, which means the team finishes faster and bills less. We call this "review efficiency," and it is an important ratio to consider when looking at e-discovery alternatives.

In this case, the team achieved a review efficiency of 1.33 to 1, which is pretty remarkable. That means that the review team looked at very few nonresponsive documents over the course of the project. Taking the math up two decimals, the average team member only had to look at 133 documents to find 100 responsive ones. Not a lot of time wasted.

That's why the chart showed batches that quickly approached 100% responsive. Interestingly, you will see a responsive rate dip early on in the project. That was because the team had to finish a couple key custodians first and when they started running out of relevant documents, the numbers dipped. When we opened the review back up to the entire collection, the responsive rate jumped as well. Some batches were 100% responsive.

You can also see that batch richness dropped at the end. This is to be expected. With CAL, the goal is to keep reviewing until you stop seeing responsive documents. Once the responsive rate drops substantially (say to a tenth of the high water mark), that is a signal that it is time to stop. In this case the team kept reviewing batches to make sure they were nearing the end.

## Adding Documents at the End

A few spikes at the end of the process occurred because the team collected some additional documents and added them to the review. With Predict, rolling collections are not a problem. The added documents simply join the ranking and are promoted accordingly. That is what happened here. With TAR 1.0, in contrast, you have to start the training over again.

## Changing Review Criteria

There was another wrinkle in the review process, although it is not uncommon. At a couple of points along the way, team leaders refined their view of responsiveness. This is a natural process as you learn more about your documents and about your case. Many call it relevance drift. The simple fact is that you know more about your needs at the end of the process than at the beginning. It is one of



the biggest weaknesses of the old TAR 1.0 process. If all your training is done at the beginning, how do you account for what gets learned as you go along?

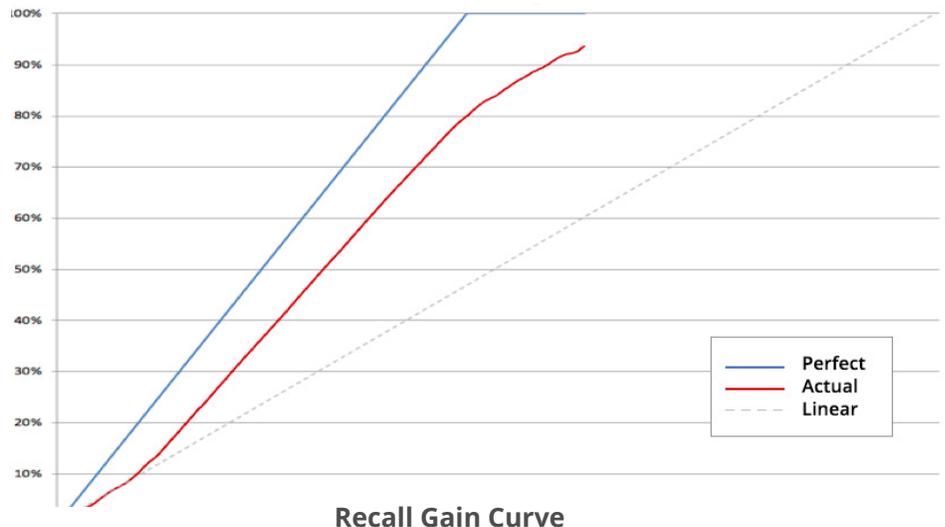
Catalyst's TAR 2.0 algorithm is noise tolerant, which means that coder inconsistencies and even changes in direction do not adversely affect it. With CAL, every ranking starts fresh, with no memory of the previous one. That way, if you were to retag tens of thousands of documents, the next ranking would take it in stride. The same is true as the team refines its search objectives.

That happened several times in this case as understanding increased. Yet, we could see no adverse impact on performance.

## The Results

The team achieved 94% recall in this review, far greater than that required by the courts. They did so having reviewed only 60% of the collection. You can see how efficient Predict was in the adjacent chart.

This is a yield or gain curve. The x-axis shows the percentage of documents reviewed. The y-axis shows the percentage of responsive documents found.



The diagonal dashed line shows the expected progress of a linear review. In a linear review, if you look at 20% of the documents, you will expect to see, on average, 20% of the relevant ones. At 50%, 50%. At 100%, 100%.

The blue line shows a perfect ranking. It represents how the review would go if the algorithm pushed all of the responsive documents to the front and the reviewers never had to look at a single nonresponsive document. Of course, this never happens.

The red line shows how this review went. Because richness was high (45%) you can't expect the typical straight up rise you might see with lower richness (we have many examples including our TREC published work). Rather, the proper measure is how it compares to the alternative: linear review.

So, for example, in this case the team found 80% of the documents after reviewing just 45% of the population. To get to 80% in a linear review, you would have to review 80% of the documents, or an additional 370,000 documents. At an average of \$1.50 per document reviewed (and QC'd), the client saved \$550,000 dollars using a predictive review.

No algorithm can achieve perfection. The question to ask is "How close did we come? In this case, the answer



is pretty close. As we mentioned before, the team achieved a review efficiency of 1.33 to 1. That translates to viewing about 60% of the population to get better than 90% recall, which you can see in the chart. The review was about as efficient as it can be.

TAR 1.0 systems generally can't handle low richness collections, requiring users to cull the population using keywords or otherwise until richness rises to 15% or higher. TAR 2.0 systems excel at low richness, which makes them a safer bet for overall e-discovery. This case study also showed us that TAR 2.0, at least in the form of Insight Predict, handles high richness collections as well.

The Catalyst review ninjas met their deadline with room to spare. Review costs were a lot lower than they would have been with linear review or keyword search.



**Learn More**  
[catalystsecure.com](https://catalystsecure.com)



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