How big does a case need to be before it makes sense to load it into a TAR system? Is 10,000 documents enough? How about 100,000?

Client Snapshot: International Law Firm
- SEC investigation
- Just 16,000 documents in the set
- 1 week to review and fixed-fee budget
- Case shows advantages of TAR 2.0
The question has persisted since technology assisted review got its start. How big does a case need to be before it makes sense to load it into a TAR system? Is 10,000 documents enough? How about 100,000? In this case, it was just 16,000 documents, but TAR enabled the company to cut its review by 75% and get it done in under a week.

The Problem: A Tight Deadline and a Fixed Budget

Our client was a major law firm representing a company in an SEC investigation on a fixed-fee basis. It had a relatively small document set to review—just 16,000 documents. But the deadline was tight, the review had to be done quickly, and the company did not want to use contract attorneys, preferring instead to have the firm’s own lawyers take the laboring oar.

With So Small a Set, Would TAR Make Sense?

Mindful of its tight deadline, the firm wanted to use Insight Predict, our TAR 2.0 predictive ranking engine, to find the relevant documents more quickly. But it was concerned that using TAR with so few documents would be costly and ineffective.

With a first-generation TAR 1.0 system, the concern would have been justified. With TAR 1.0, they would need a senior lawyer to serve as subject matter expert to start the training process. The SME would first have to review and tag 500 or more randomly selected documents as a control set to use for measuring training efforts. Then the SME would have to do multiple rounds of training before the review could even begin.

This might require review of 1,600 or more documents before the system stabilized. Depending on the case, the training could require review of many more documents, easily as many as 3,000.

After that, the SME would have to tag an additional sample, perhaps another 500 documents, just to test whether the training was complete. All told, the SME might have to review 3,000 to 4,000 documents just to train the TAR 1.0 classifier. Only then could the review begin.

That’s a lot of work for a case with a small number of documents. It is why many e-discovery professionals advise that TAR should only be used for larger cases. Some suggest the threshold is 100,000 documents before you can justify the SME’s training efforts and expense.

“Many e-discovery professionals advise that TAR should only be used for larger cases where you can justify the SME’s training efforts and expense.”
Insight Predict Requires No Control Set or SME

With Insight Predict, the size of the collection is not a factor. With Predict’s Continuous Active Learning protocol, there is no minimum threshold because there is no need to create a control set. Predict ranks all of the documents all of the time. Predict does not require an SME to do the initial training. With no need for an SME, the review team can get started right away. Review is training and training is review.

In this case, the CAL process was quick and simple to implement. The team initially found 67 relevant documents through keyword search. We used these as training seeds for the initial ranking. Then the team started reviewing documents.

It turned out that about 11% of the documents were relevant, for a total of about 1,800 relevant documents. As the team reviewed documents, Predict continuously learned from their tagging and presented increasingly relevant batches to the reviewers. Relevance in the batches quickly rose to as high as 70%. Review efficiency increased seven-fold as a result.

In just days, batch relevance dropped to single digits as the team depleted the relevant population. The team stopped review and moved to a systematic sample to determine what level of recall was obtained. The sample involved 800 documents for a confidence
level of 98% and a 4% margin of error. By this point, the team had reviewed just 3,900 documents.

The sample showed the team had found 96% of the relevant documents after reviewing only 25% of the population. The remaining documents could be safely discarded cutting out 75% of the review time and costs.

**Conclusion: Insight Predict is Effective Even for Small Collections**

Does TAR work for smaller cases? With Insight Predict and Continuous Active Learning, the size of the collection does not matter. In this case, the team used Predict to finish their review in less than a week. They easily met the SEC’s deadline while also keeping well within their fixed-fee budget.