



MORAE  
GLOBAL

Igniting Business Performance

# Predictive Coding in a Regulatory Investigation: A Case Study

---

## The Problem

A large financial services institution faced an investigation where regulators from the United States and Europe wanted to look at personnel responsible for having developed an index fund's strategy and its subsequent modifications for other jurisdictions (as the product was developed for two different European countries and later modified for the U.S.). The investigation also involved analyzing data from staff responsible for marketing and selling as well as trading behavior that was linked to the index's fluctuations.

Data from the matter spanned across three jurisdictions and required the analysis of a wide range of custodians' data, amounting to more than 5,000 GB. The initial volume to be analyzed was over 10 million documents hosted on multiple client databases, so the company needed to leverage technology to efficiently identify relevant information while concurrently making all efforts to limit costs.

## Context and Description

The bank had established an index fund managing a number of assets based on an internally developed trading strategy and marketed and sold it as shares to clients. The investigation emerged from concerns that the methods of calculating value created by the fund's developers were inconsistent with marketing materials. There were additional concerns pertaining to specific misconduct related to traders who had foreknowledge of the fund's strategy.

The challenges consisted of identifying key facts while meeting the regulator's deadlines for production.

## Challenges

- Work had to be completed within strict deadlines and to a high degree of accuracy to avoid regulatory penalty.
- The total amount of data consisted of more than 10 million documents.
- The data resided in multiple client databases, so it had to be mapped, identified and centralized prior to reducing and prioritizing.
- The data was housed on different types of media (e.g., email, structured data, chats), and the activities needed to be reconstructed chronologically irrespective of type.

## Solution

The best way to efficiently identify relevant information while reducing costs was to develop a defensible predictive model, leveraging technology-assisted review to identify key documents for investigative and production purposes, with minimal linear review.

Subject matter experts reviewed the sample and seed sets of the data population to build the predictive model by harnessing targeted Boolean searches, relationship analyzers and concept and lexicon searches. The model was backed with a unified strategy: to rapidly identify critical documents while defensibly excluding nonrelevant documents. In addition to these targeted searches, supplemental analytics tools such as near-duping technologies were layered over the predictive model.

The model was run over the entire data set, and the results yielded approximately 150,000 documents with the highest probability of relevance to use as the base review population.

## Outcome

- The review only took four weeks and utilized only a fraction of the resources that would have been required by a linear process.
- The company met all regulator requests, despite the regulators constantly expanding the scope of the inquiry.
- Only 150,000 documents were eyes-on reviewed, rather than the 3.4 million documents expected from a model based on search terms – a 95 percent reduction.
- As a direct result of the deployment of predictive coding and other analytical tools, the company achieved cost savings of almost \$3 million.

## Long-term Benefits

- The decision to drive the review through a predictive model and advanced analytical tools drastically shortened the timeline and delivered tremendous cost savings.
- The predictive model not only streamlined review but proved to be more effective in its analysis, with a relevancy rate three to four times higher compared to similar projects where review was based solely on search terms.
- Key information not previously known to counsel – such as undisclosed fees associated with the fund – was discovered, strengthening their ability to make more informed decisions about the investigation.
- The analysis illustrated for counsel a marked difference in the index calculation methods in practice compared to those stated in public offering documents.
- Conflicts of interest among bank staff were identified.
- These critical discoveries, as well as insights from across 30 issues within the matter, were memorialized in one of nine fact development communications, highlighting more than 11,000 “Priority and Relevant” documents.