



VALUE OF EXPERIENCE

Machine learning to support a tax investigation

CHALLENGE

Navigant was engaged by a global Law Firm to collaborate on the data review element of a tax investigation, to determine whether their client should move forward with litigation. The client had suspected potential tax obligation concerns.

Navigant's involvement included the email collection from 8 people over a period of a year, which resulted in a dataset of half a million documents after data had been processed and de-duplicated. The legal team recognised early on, that they were likely looking for a small number of communications, many of which could be inconsistent and colloquial in nature. Therefore, using key words to determine the relevant review population would fail to capture the nuances being discussed.

SOLUTION

Navigant deployed multiple analytical tools to help the legal team identify and review the relevant population of documents that didn't require the use of keywords.

Data was loaded into a visual analytical tool to allow the lawyers to begin their investigation. Within a short period, large amounts of documents were eliminated from the review. By looking at only those who were sending communications within the data set, it became apparent that many of the top 10 communicators were not from the emails collected, but rather included people whose roles were not applicable to the investigation. After a review of all communication domains and entities, the legal team were able to identify 252,940 records to be eliminated.

With 246,277 documents remaining, a limited timeframe, and still no keywords, a large population of documents required reviewing. Plus, the legal team had also obtained 70 emails from the client that contained relevant communications for the investigation. Using Navigant's proprietary predictive coding tool, NAVPredict, we were able to build a predictive coding model analysing those 70 documents.

The initial sample pulled identified how many estimated relevant documents were in the remaining population, or "richness", returning only 7 relevant documents. Extrapolating that to the overall remaining population meant that we were expected only 1,000 relevant documents to be produced.



RESULTS

The Law Firm's review took two weeks to complete. The review resulted in the client uncovering 95% of the expected relevant information by only reviewing 4,784 documents. Many of the documents found through this process would not have been captured by keywords. The results of the review completely changed the outcome of the investigation.

While leveraging Navigant's technology was crucial to the success of this investigation, what our client valued most was working in partnership with Navigant – given they were working with a large data set and a small internal team. In utilising analytics (NavPredict) in the early case assessment it allowed 95% of responsive documents to be identified despite the extremely low responsive rate (0.45%) in a large 246,277 document data set.



For more information please go to navigant.com/DataManagement