Canopy

CASE STUDY

Canopy Reduces Review Population by 89%, Saving Response Team 2,000 Hours

OVERVIEW

A U.S.-based online retailer experienced a business email compromise resulting in 186,479 compromised documents.

Their incident response team first tried tackling this project using traditional ediscovery methods, which pulled approximately 90,000 documents — nearly half of the original data set — into the review population.

The project lead suspected this number was too high. They didn't want to waste the team's budget and limited response time reviewing potentially thousands of unnecessary documents.

SOLUTION

The incident response team enlisted Canopy's Data Breach Response software. Its advanced machine learning models automatically detected and validated over 70 types of PII.

Then, the application's image and document classification tools helped guide the response team's strategy and focus their review efforts.

RESULTS

Canopy narrowed the review scope by eliminating 95% of the original data set, accurately flagging just 10,106 documents (89% fewer than ediscovery methods). This saved the response team nearly 2,000 paid hours in document review.

canopyco.io | contact@canopyco.io

v20230712

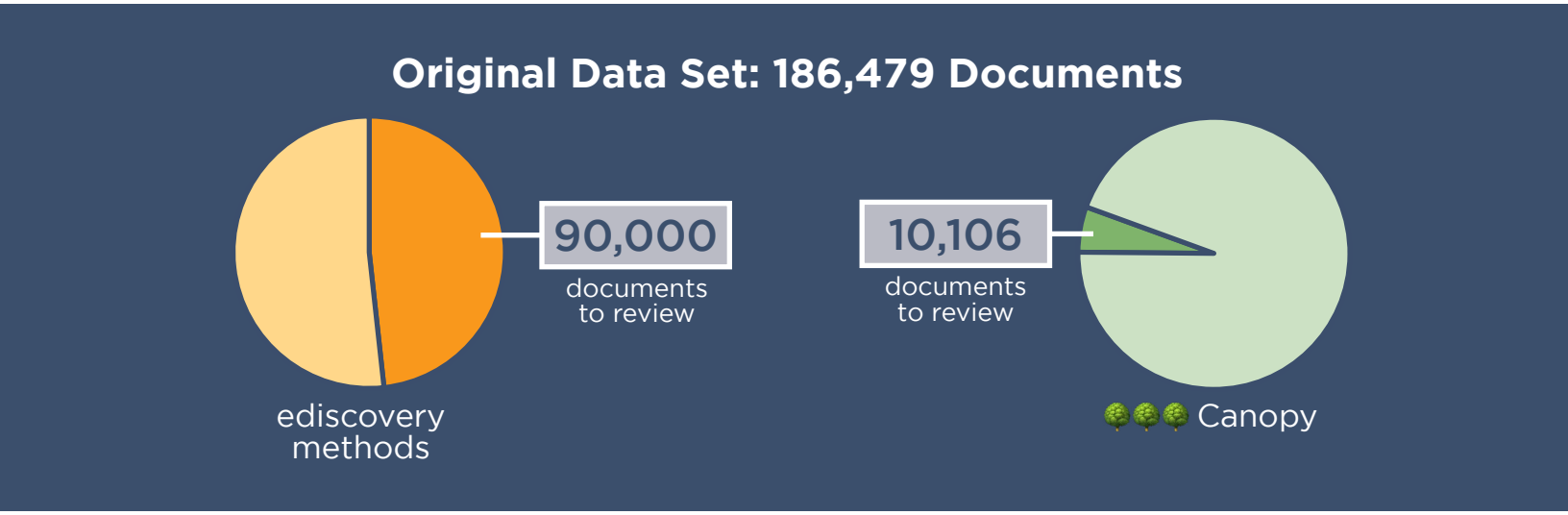
Canopy

Over 180,000 Documents — Where Do We Start?

When a business email compromise (BEC) occurs, response teams need to search the compromised documents for personally identifiable information (PII) and protected health information (PHI), then pull those suspected of containing PII/PHI into review for further analysis. To keep review costs low and enable the client to meet legally-mandated breach notification deadlines, the goal is to narrow the review team's focus by eliminating as many documents as possible during this initial sweep.

Canopy's partner, a leading review management company, faced this challenge when one of its customers, a midsize online retailer based in the United States, experienced a BEC. The accounting manager's inbox was breached, resulting in 186,479 compromised documents.

The response team initially swept these documents for PII using traditional ediscovery tools and techniques. As often happens with cyber incidents, these were overinclusive and produced a high rate of suspected false positives. When 48% of the initial data set (approximately 90,000 documents) was pulled into the review population, the project lead suspected that something wasn't right. The flagged number was too high, and they were not willing to dedicate more time and money to reviewing thousands of unnecessary documents.



Canopy Is the Better Way to Data Mine

The project lead enlisted Canopy's Data Breach Response software. Its advanced machine learning models immediately detected and validated over 70 types of potentially-reportable PII elements in the documents.

One major improvement was in the accuracy of social security number (SSN) detection. Unlike traditional ediscovery tools that rely on methods like regular expressions and pattern matching, Canopy compared potential SSNs with the Social Security Administration's guidelines. This ensured that only documents with valid SSNs were pulled into the review population, ignoring dummy SSNs like 123-45-6789.

Canopy

Canopy then calculated how many elements of each PII type were present, along with the total number of documents containing that sensitive information. These insights and more were immediately visible and available as filters, allowing the project lead to assess the data with custom queries.

Further, Canopy's image and document classification tools provided insight beyond the detection of concrete PII elements. The application's machine learning models grouped visually similar images (e.g. JPG, PNG) and documents (e.g. PDF, Word), and the interface allowed the project lead to quickly scan thumbnails to identify what was reviewable versus what could be ignored.

Armed with all of this information, the response team was able to create an efficient, tailored review strategy.

Ediscovery Is Not a Data Breach Response Solution

Speed is critical when responding to a BEC. Canopy's Data Breach Response software was able to narrow the review team's focus to just 10,106 documents — 89% fewer than were initially flagged via ediscovery methods.

Without Canopy, this response team would have wasted significant time needlessly reviewing 79,894 documents that did not contain PII. Assuming the typical hourly-paid reviewer averages 90 seconds per document, they would have paid for 1,997 more hours than were necessary.

By eliminating the false positives pulled in via ediscovery techniques, Canopy's advanced PII detection and machine learning-powered tools enabled them to complete this project and meet notification requirements much faster.

Notify Faster

79,894 documents
incorrectly flagged via ediscovery methods

90 seconds
average time to review each document

~2,000 hours saved
with Canopy