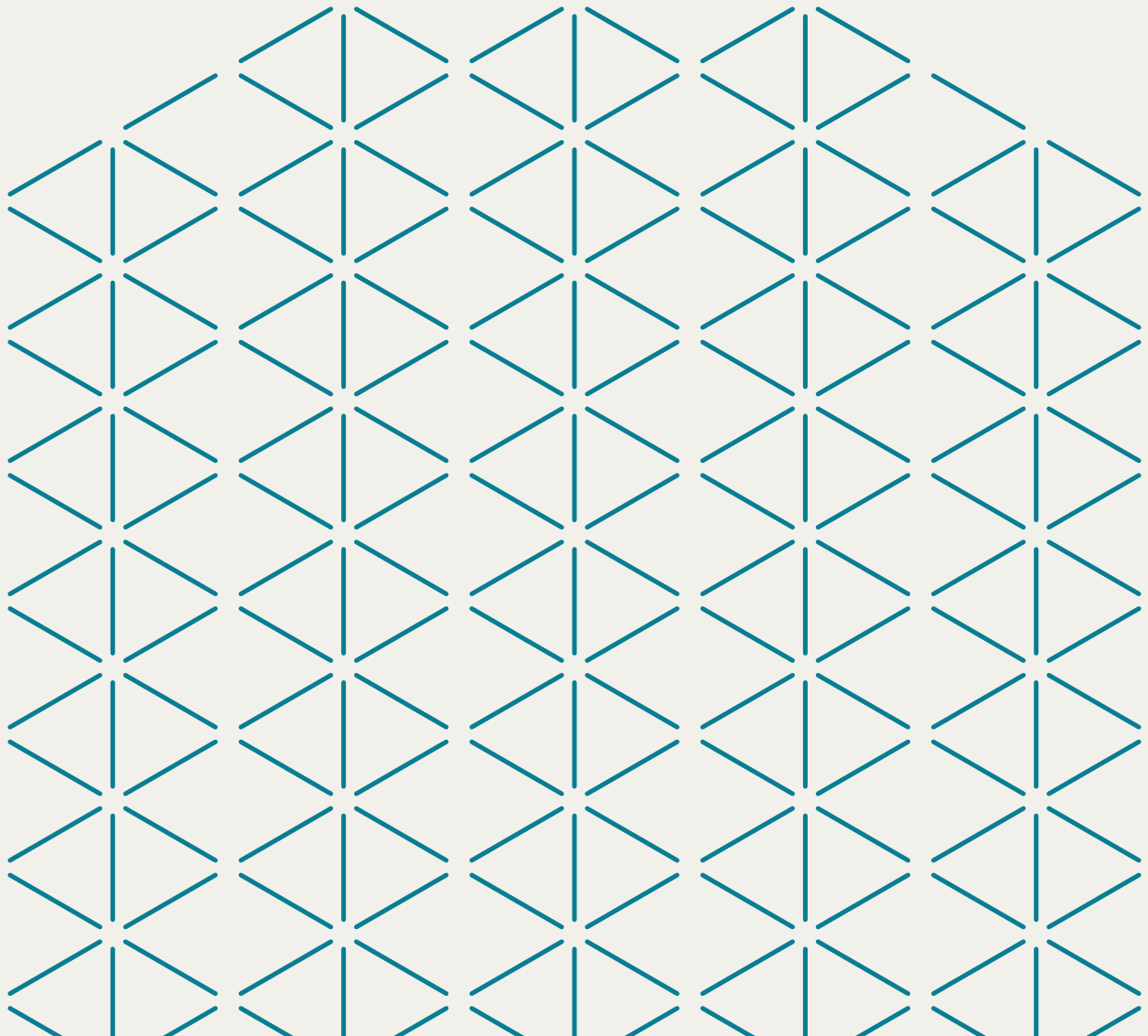


Leveraging Machine Learning During Document Review



HOW-TO-GUIDE

Leveraging Machine Learning During Document Review

Introduction

So you have been given massive data sets to parse through for your cases and are still doing review manually? And the AI solutions you have tried using are not seamlessly integrated and hard to use? It can be an excruciating and slow process to ensure your team accurately finds all of the key pieces of evidence quickly.

As complex, overwhelming, and costly as the ediscovery process can be, many of the obstacles are caused by the use of outdated or inadequate tools. The need for quality tools to help with streamlining review workflows has only become more apparent with the influx of new regulations surrounding data privacy and the existing issues associated with the current state of data management. In particular, the unprecedented increase in the volume of ESI, or electronically stored information, involved in modern litigation and internal investigations has not lessened the pressure to find relevant information under tight deadlines. It will only become more challenging for legal professionals if they don't develop new strategies and leverage modern technology, such as machine learning, which has emerged as a viable solution to this problem.

But what exactly is machine learning?

What Is Machine Learning?

Machine learning is a type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed. Instead, the computer uses your input and/or the contents of data to make educated guesses beyond that input. There are two designations of machine learning: supervised and unsupervised.

Supervised machine learning relies on labeled input and output training data, whereas **unsupervised machine** learning processes unlabeled or raw data. With supervised machine learning, the model requires actively training the models (i.e., reviewers coding documents as relevant or not relevant) to improve their predictions, whereas unsupervised machine learning does not. Predictive Coding is an example of supervised machine learning, and Clustering is an example of unsupervised machine learning.

Machine Learning Within Ediscovery

Machine learning has the potential to transform ediscovery, [reducing the amount of data that needs to be reviewed by upward of 80%](#) in some cases. It has also transformed machine translation, machine transcription, OCR, or optical character recognition, and other technologies employed in ediscovery, among other things.

Machine learning has had an especially strong impact during review.

Review tools have gone through an evolution over the last few years. The most effective solutions feature intuitive user interfaces for discovery, timesaving capabilities like Clustering and Predictive Coding that can reveal the hidden details of documents at scale, and tools that enable automated workflows that enhance and expedite the review process.

Leveraging Machine Learning During Review

Automated processes built with machine learning improve predictability and efficiency by reducing the likelihood of human error, which causes delays and dependence on outside vendors to prioritize case-specific requests. With automation, legal professionals can get things done reliably, on time, and without introducing timeline uncertainties for client requests.

In particular, there are four specific steps during the review stage where, when utilized properly, supervised and unsupervised machine learning has a tremendous impact:

1. Leveraging Clustering to view the whole document corpus in visualization during Early Case Assessment (ECA).
2. Using Predictive Coding to identify relevant documents and data in other review categories to dramatically reduce document sets during review.
3. Combining Clustering with Predictive Coding overlays to better identify and understand meaningful trends.
4. Leveraging Clustering coding and rating overlays to mitigate risk of human error during the quality control process.

What is Predictive Coding?

Within the world of ediscovery, Predictive Coding, also known as TAR, or technology-assisted review, [automates document review](#). A type of supervised machine learning technology, Predictive Coding enables computers to learn from human input and make educated guesses when classifying documents. Legal teams across industries use Predictive Coding to identify responsive or relevant documents and data in other review categories, including privilege, issue codes, and other vital classifications, in order to dramatically reduce document sets.

What is Clustering?

In layman's terms, Clustering is a machine learning tool that helps detect trends within data sets. The tool visualizes and organizes documents, within a given set, that are conceptually similar, and it does this without requiring users to train the model.

Clustering is considered an example of unsupervised machine learning because the tool uses an unsupervised machine learning algorithm that identifies and analyzes metadata (e.g., author, subject, title, and email sender/recipient) and words across an entire data set to determine conceptual similarity among digital files and documents. Clustering also utilizes what is known as a density-based clustering algorithm, which allows users to visualize trends within data sets more easily than with traditional K-means clustering algorithms.

Step 1: Use Clustering for ECA

It's easier to sift through and review document sets when items are organized and cataloged. Grouping conceptually similar documents and files expedites the review process by allowing you to quickly identify any consequential documents and files, saving your team resources, time, and money.

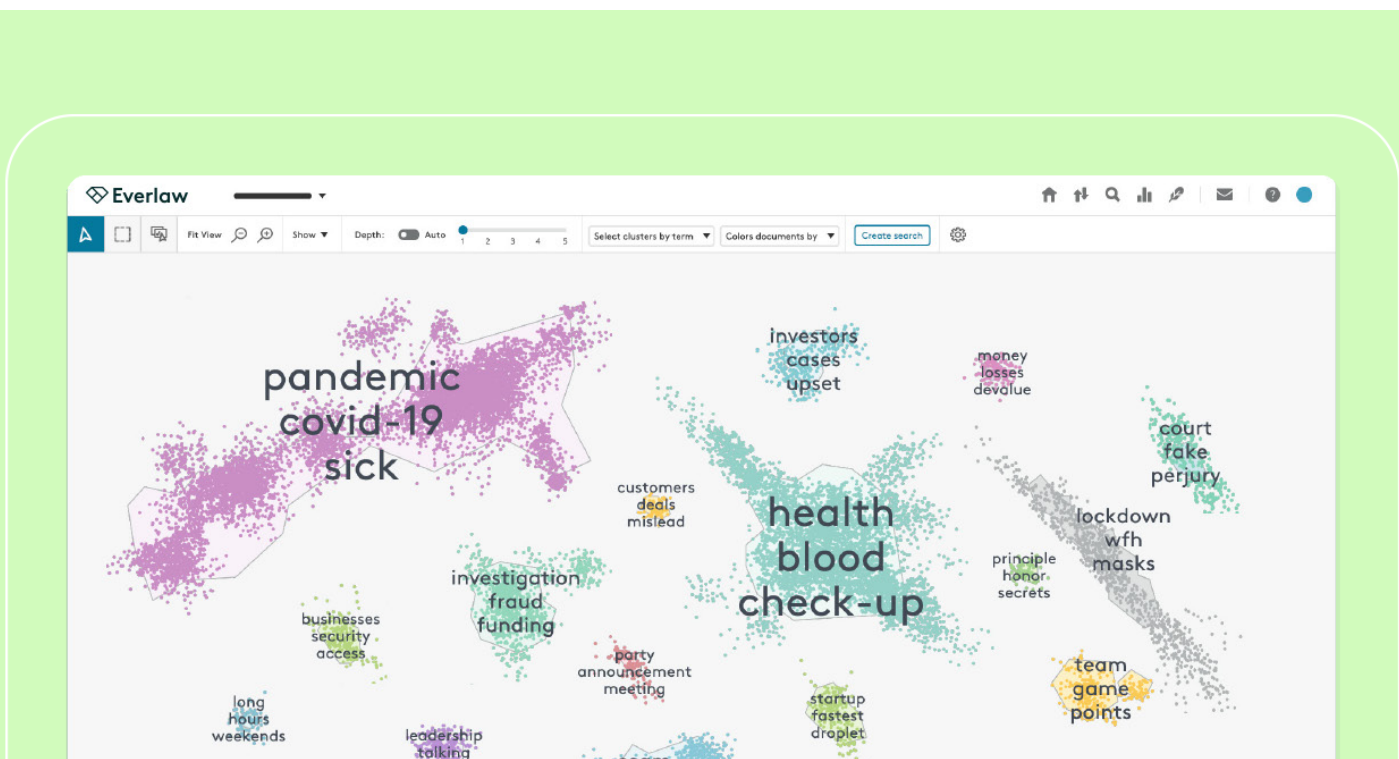
Instead of reviewing documents on the individual level or creating search reports to understand your data, Clustering allows you to move from a planetary survey of all the evidence in a case down to a blade-of-grass view – a single document – in a unified view. However, knowing where to begin can be overwhelming when a legal team has a million documents to review. Analytic tools like Clustering can dramatically streamline the process.

Clustering During ECA With Everlaw

Spotting and identifying which documents and files are relevant to a particular case can help establish the timeline of events and lay the foundation for the legal strategy. But often, there are mountains of data (duplicate documents and redundant information) that can lengthen the initial review. [Ediscovery software that utilizes automation technology and machine learning](#) can expedite this process.

For example, Everlaw Clustering uses unsupervised machine learning to quickly pinpoint conceptually similar files and presents them via an intuitive graphic display, making it easier to uncover valuable insights into your data set without manually building a search during Early Case Assessment.

Also, documents can be easily filtered and sorted, enabling Everlaw users to explore their data set in order to find documents that potentially have evidentiary value, discover new search terms, and make critical decisions early on (i.e., prioritizing and organizing documents for review).



Step 2: Use Predictive Coding for Document Review

Predictive Coding is a powerful tool that can help facilitate an efficient review, especially considering the increasing sizes of data sets involved in even routine cases. Though the technology is no longer a novelty, techniques have matured over the years. Some are understandably intimidated by the jargon and technicality involved, but it's a surprisingly straightforward process.

Predictive Coding During Document Review With Everlaw

Legal teams can utilize Predictive Coding to find responsive ESI during the review phase of a case, [shaping and altering the discovery process](#) for law firms, lawyers, clients, and the courts. Here's how it works:

- / After quickly [setting up a model for their case](#), reviewers “train” the software with a seed set (i.e., a sample of documents retrieved from the larger group of documents needing review). Training consists of coding each document as relevant or not relevant to the case.
- / The software, using AI, learns from the reviewers' coding decisions (and content of the reviewed documents) to generate prediction scores for all documents in the set. These prediction scores indicate how likely reviewers will find documents to be relevant based on prior coding decisions.
- / As document review continues, the software continually learns from reviewer decisions and improves the accuracy of the prediction scores.

The predictions generated by Predictive Coding can be extremely powerful. Prediction scores can help reduce the amount of manually reviewed documents (saving thousands or even millions of dollars in review time), prioritize documents for review to find important documents sooner, and assist with quality control to check documents prior to production.



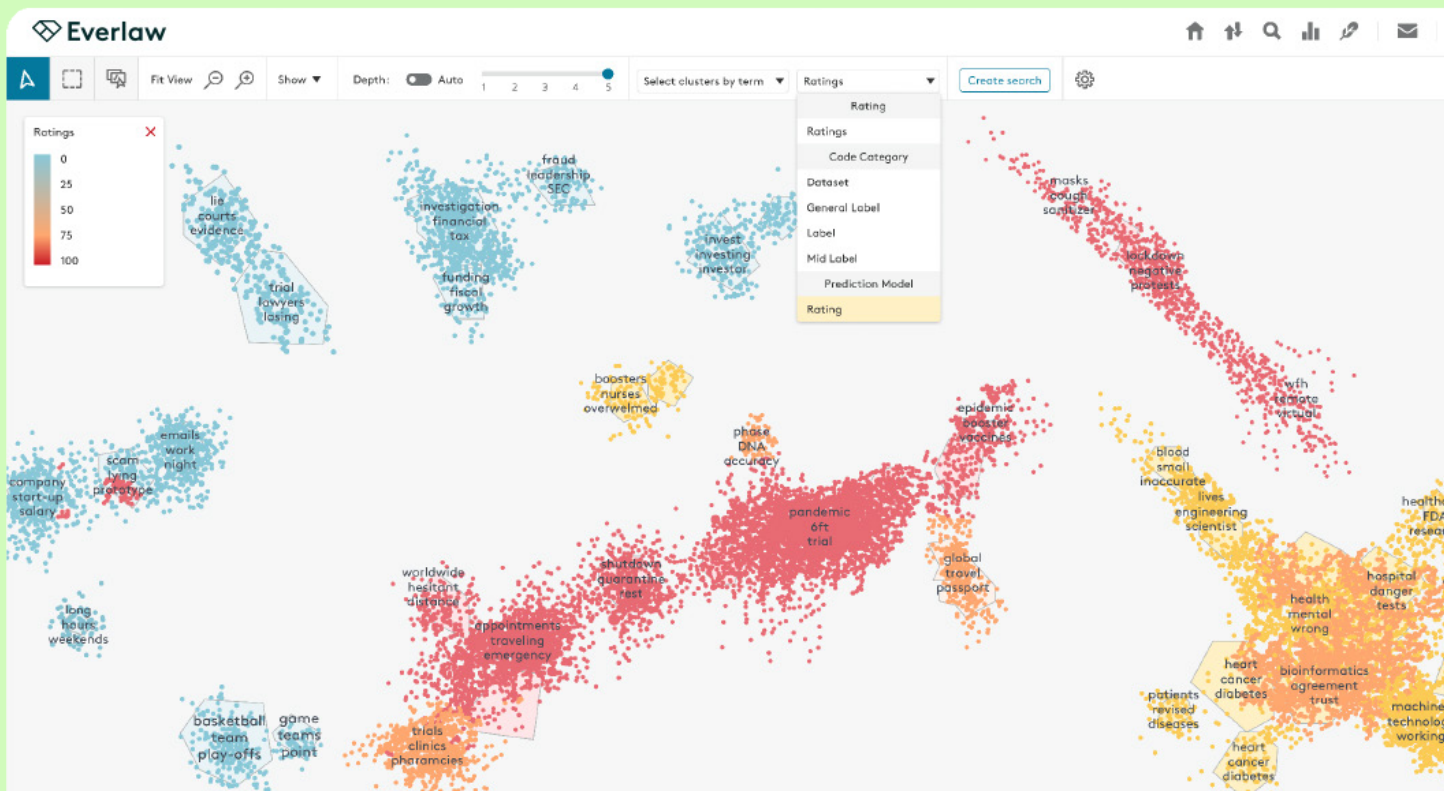
Step 3: Use Clustering With Predictive Coding Overlays for Document Review

One of the interesting aspects of machine learning is that both supervised and unsupervised tools can work together in perfect concert. For example, a user can leverage Clustering to group related unlabeled documents together and leverage Predictive Coding to identify those groups of documents that are highly predicted to be relevant. This pairing is particularly useful when it's difficult to extract digital files and documents within large data sets.

Identifying Relevant Concepts With Everlaw

Once you have set up Predictive Coding on your project and have started the review process, it's time to identify what concepts are most relevant by prediction score. Here's how it works on the Everlaw platform:

- / Using the [color coding overlay](#), users select their prediction model from the drop-down list.
- / Users can zoom and pan across the page to see which clusters of documents might be more likely to be rated hot.
- / Once documents are selected, users can open the selection in a results table and share the documents with their team to prioritize those documents for review.



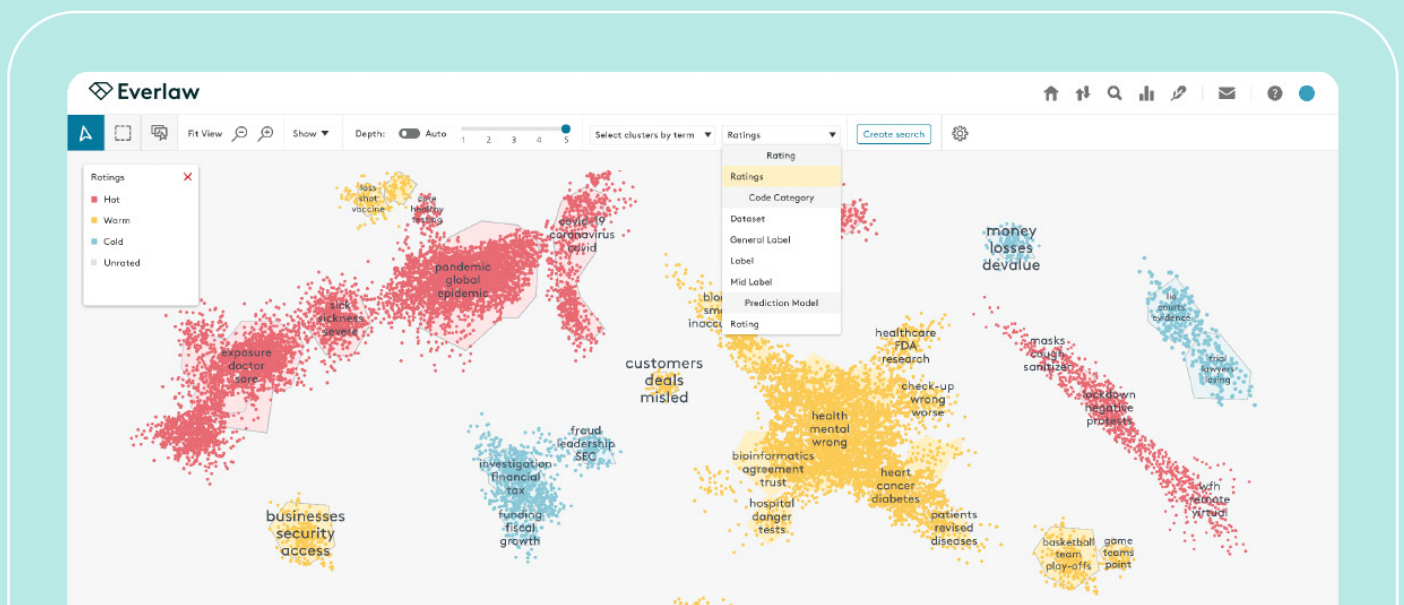
Step 4: Use Clustering With Coding and Rating Overlays for Quality Control

People make mistakes, and so do legal professionals. That's why conducting quality control is a vital step during the review process. For example, an attorney responsible for assessing the quality of review decisions made by their team needs to make sure nothing slipped through the cracks. That means double-checking that all relevant documents were reviewed, coded, and rated correctly. Ensuring that all documents are categorized accurately, especially before the production is exported, can reduce the risk that privileged, confidential, or simply nonresponsive information is produced.

Conducting Quality Control With Clustering on Everlaw

Documents can be clustered based on critical subject matter to verify coding decisions and ensure that relevant documents in hot clusters were identified during the review. With Everlaw, you can utilize the [coding overlay](#) to identify potentially uncoded or incorrectly coded documents. Here's how it works:

- / Users select the coding category of interest from the drop-down list.
- / To see whether conceptually similar documents are coded differently, users can identify outlier documents by color.
- / Users can also view uncoded documents, which are light gray. This might be helpful to see if these documents slipped through the cracks during review.
- / Users can select a subsection of a cluster or documents across multiple clusters by using “document selection mode” in the toolbar (or the “d” shortcut on your keyboard).
- / After clicking and dragging to select the section of interest, users can open a results table of those documents to see why the outliers are coded differently.



Transforming Your Approach to Ediscovery With Everlaw Clustering

Everlaw Clustering is seamlessly integrated with the Everlaw platform to help legal teams accelerate finding key pieces of evidence, mitigate the risk of human error, and confidently navigate ediscovery at terabyte scale. Clustering also complements Everlaw Predictive Coding's supervised learning for more powerful AI workflows.

More specifically, here's what Everlaw Clustering enables legal teams to do:

- / See clusters dynamically separate and merge based on zoom level through dynamic zoom.
- / Overlay predictive coding models and use the prediction scores to find hot documents.
- / Overlay ratings and codes to prioritize certain document sets or validate review decisions.
- / Recluster at any given moment.
- / View the most common terms found in clusters and any transcribed A/V files.
- / Filter visualization to only display a specific search.
- / Access similar documents in a cluster through the context panel in the document review window.
- / Open documents directly into Data Visualizer.

To learn more about Everlaw Clustering and Predictive Coding technology, request a demo [here](#).

"I routinely use Everlaw to conduct investigations and analyze document productions. The platform simply makes searching across a large volume of documents so much easier than anything else I've come across. The ability to create tags has completely changed how I conduct investigations because it's enabled me to drastically increase my level of organization."

**Associate
AmLaw 100
Law Firm**

Contact us

US Headquarters
2101 Webster St.
Suite 1500
Oakland, CA 94612
1 844 everlaw

London Office
32-38 Scrutton Street
London, EC2A 4RQ
0800 068 9249