Anticipating Predictive Analytics' Potential Uses In Litigation

By **Richard Finkelman and Karl Schliep** (February 9, 2021, 3:23 PM EST)

Without knowing it, we all use predictive analytics in the technology we use in our everyday lives.

We use recommendation engines on Netflix and Amazon to decide what to watch and what to buy. Professional sports predictions and analytics are commonplace enough that the odds the next play wins the game are displayed on screen before the play.

Election news coverage is boring without prediction. Anyone who has seen Steve Kornacki and his interactive board during election night has seen a great of example predictive analytics in election coverage, as well as great theater.

Richard Finkelman

This article focuses on the use of these and other technologies in litigation and offers predictions of its future use in the legal industry.

In the not too distant future, a trial attorney will be in court arguing their client's case in front of a jury selected with the help of advanced analytics. These analytics will have predicted the likelihood of who will be the jury foreperson and how each juror will ultimately vote. The jury will have been presented information about any predictive analytics involved with evidence.

Karl Schliep

The judge will have held a science day prior to trial starting, allowing each side to explain in layman terms the artificial intelligence-powered products or services in dispute.

The client will have been briefed about the calculated probability of winning or losing the case, winning or losing specific motions, as will the likelihood that submitting a motion at all will change the judge's sentiment about the case.

The expert witness's trial exhibits will include a schematic that explains how predictive analytics was used to calculate their damage analysis. Appellate work will use analytics to support claims on alternative outcomes.

These events are all coming to a courthouse near you soon and the lawyers and their clients that understand the technologies, the analytics and their uses will have clear advantage over those who do not.

Predictive analytics in litigation is currently most commonly used during document review and production. Many are already familiar with how predictive coding is used to predict responsiveness of documents during document review. Similarly, technology assisted review has been used for more than a decade and has increasingly come to be an accepted practice for large document reviews.

However, predictive analytics will have important ramifications and implications for other litigation activities.

Today, predictive analytics can already be used to help negotiate custodians and keywords, can be used to analyze and perform fact development from document productions, and can be used in predicting the outcome of motions by type, jurisdiction and judge.

Increasingly, analytics are being used by expert witnesses, particularly those involved in assessing economic damages. Perhaps most interesting is the emerging use of predictive analytics in settlement prediction and probable case outcomes.

Predictive analytics was born from the marriage of statistics and machine learning.

Statistics are defined as the practice or science of collecting and analyzing numerical data in large quantities. Machine learning is a field of artificial intelligence and is defined by the ability to develop computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.

Programming computers used to involve coding specific instructions for the software to execute. Machine learning involves teaching computers how to recognize patterns in information without specific instructions.

Historically, the industry has focused litigation technology and technology advances on activities related directly to discovery, particularly those that involved the processing, review and production of documents. Over the last decade, electronically stored information, or ESI, and the protocols surrounding these activities have become commonly known to lawyers and judges.

Today, most attorneys are comfortable with the initial discovery work that goes along with negotiating ESI protocols and filing agreed upon stipulations that don't require the courts to intervene or rule on ESI disputes. Courts have also been clear that they do not want to be involved in ESI discussions and negotiations between parties.

Agreed-upon ESI stipulations are now common and discovery disputes on keywords or custodians are rarely fought over in court. Practically, this has led to an acceptance of the now long-held beliefs that negotiations on ESI should focus on limiting custodians and agreeing on keywords for filtering and document review. However, these agreed-upon guidelines are antiquated and in need of fresh thinking.

The activities involved in prediscovery work, especially those involving attorneys learning about their client's business, the issues involved in the case, who was involved in important activities and what the organization looked like during the time period of the matter, are ripe for rethinking based on techniques that help create better filtering results and more rational identification of organizational structures and custodians.

Information gain is one use of predictive technology that can measurably increase the filtering of potential responsiveness of documents and data.

Information gain is a measure of how informative a word or phrase is relative to other words and phrases. "Racquet" or "bat" are more informative than "game" or "league" since they tell you the type of game being played.

Similarly, if you type "apple" into Amazon.com, you get two sets of search results, apples from Amazon Fresh and Apple Inc. accessories such as AirPods and Apple watches. If you keep scrolling you may even find souvenirs from New York City, the Big Apple. In this example, the words "computer" or "fruit" in a phrase or sentence can enhance the information gain for the word "apple."

Data scientists have more complex and technical definitions for information gain, which include mathematical concepts and formulas for decision trees, random forests and entropy. The reduction of entropy or impurity in a dataset is accomplished with specific formulas and algorithms used in training datasets.

For attorneys, the most important things to understand about the data science work is that it can be explained to a court and the results can be used to reduce the litigation costs of discovery.

Similarly, organizational charts are used in determining possible custodians during ESI negotiations and are often the first production request from opposing counsel. While they are a helpful start, these charts rarely tell the real story of who worked together on specific projects or assignments.

A company's controller may be a more important to a case than the chief financial officer, and the regional compliance head for Latin America may be more important than the chief compliance officer. Scoring documents by importance of words and phrases and analyzing who communicated more frequently with others allows you to build organizational charts that might not look anything like the published charts from a company.

This information is important because it can be used to properly identify the most important custodians prior to discovery requests. It is also important because it can be used on opposing party productions as well. This allows attorneys and their legal team to better understand the actual groups or people involved in activities related to a case. Often, these people are not easily identified without robust and costly discovery.

Many of the technologies that comprise today's use of predictive analytics have been made available by advances in the artificial intelligence technologies already mentioned and in advances using natural language processing.

Natural language processing is another field of AI and it uses linguistics and computer science to process and analyze large amounts of language data. This technology is more easily understood in the context of emails and techniques that provide sentiment analysis of communications between and among a group of people.

Nonkeyword predictive analytics, such as the ability to predict the likelihood that a document is privileged, can be accomplished by scoring the information, phrases and writing styles of the language in a document. The results yield faster reviews that require fewer attorneys. Privilege logs can be statically optimized to certain thresholds and used to negotiate clawback agreements.

The current process for privilege reviews is antiquated and in need of repair. This process includes keyword searching for law firm domains, attorney names and the term "privileged and confidential." The results are used to create a subset of documents for privilege review. This document set often contains a majority of presumptively privileged documents that are not privileged and need to be culled out and potentially produced.

This activity is generally done by reviewing every document in the presumptively privileged document set. But predictive analytics can help legal teams review subsets of documents that are scored from most likely privileged to least likely privileged. Machine learning can be used to rescore and predict the next likely set of privileged documents.

Reviewing documents in this way does not require eyes on every document. Eventually the predicted probability that documents are privileged becomes obvious to reviewers and the need to review the entire set becomes unnecessary.

Many people assume that the use of data science techniques end when discovery ends. Historically, the focus of technology innovation was applied to discovery because of the massive datasets and high costs of the underlying activities that occur in discovery. Increasingly, we see the use of data science techniques also being used post-discovery.

Fact development and expert work are two examples where data science and advanced analytics are emerging as important and meaningful tools.

Predictive analysis can be applied to understanding and analyzing opposing and third-party productions, the same way it can be used to understand your client's information. Document analytics, communication sentiment and the unofficial organization charts are helpful for deposing witnesses, filing motions and educating your expert witnesses.

Experts and the work they do also increasingly rely on underlying predictive analytics and the conclusions they infer from them. Data scientists and their work are now part of an expert's arsenal for their conclusions, reports and testimony.

We see a day in the near future when data scientists and statisticians are called upon as fact witnesses to explain the algorithms and models that were used to create the particular analytics relevant to a damage calculation or class certification.

One of the more exciting technology advances we see in litigation are analytics and models that predict settlements and probable case resolution outcomes. Liability issues need to be discussed, and clients will need to accept that prediction is not a guarantee of anything. But the power of predicting the future value of a case is an exciting development for attorneys, their clients and mediators.

We would be remiss if we didn't note the evolution of AI and emerging litigation involving its uses. Cases involving the AI of autonomous vehicles, the use of predictive analytics in the labor and employment practices of companies, and the actions of independent AI programs colluding with each other and price-fixing are just the beginning of this trend.

AI litigation will certainly challenge all of us as society comes to terms with privacy and ethics and intent and damages from the AI technologies that are increasingly embedded in the technology of our daily lives.

Lastly, we note an emergence of master's of law programs in litigation management, and science and law, as bellwethers of the growing recognition that in the future, we'll need attorneys who will understand the business of litigation as well as the science of predictive analytics and computational law.

While we don't see a future where trial lawyers are wearing khaki pants like Kornacki, jumping around a courtroom, writing with their fingers on electronic displays in front of a jury, we do see predictive analytics becoming a common feature embedded in many litigation activities, including those inside a courtroom.

Richard Finkelman is a managing director and Karl Schliep is a managing consultant at Berkeley Research Group LLC.