USING DATA ANALYTICS TOOLS TO SUPPLEMENT TRADITIONAL RESEARCH AND ANALYSIS IN FORECASTING CASE OUTCOMES

Mark K. Osbeck[*]

In the past several years, some of the most significant technological advances in legal research have involved non-traditional research tools. For example, Bloomberg Law, Lexis Advance, and WestlawNext now provide much better access to business and financial information. Similarly, the most significant technological advances in the next several years may take place not in the traditional domain of legal research (i.e., in finding primary and secondary sources), but rather in the complementary domain of case forecasting.

Prediction has always played a vital role in the practice of law. Suppose, for example, that the police arrest you and charge you with a crime. You definitely need a lawyer—but not necessarily a great trial lawyer. Instead, since the vast majority of criminal cases result in plea agreements, what you need most is a lawyer skilled at negotiating such agreements, who can help you decide whether you should accept the prosecutor's deal or take your chances at trial. And this requires your lawyer to make a prediction as to the likelihood of prevailing at trial, should you reject the prosecutor's offer.

Predictive analysis is no less important in the civil arena. To properly evaluate settlement prospects, a lawyer must be able to assess the rough odds of winning at trial, and the potential exposure should the case proceed to trial. The same is true with respect to the desirability of initiating lawsuits: it is generally wise to litigate only if the expected recovery exceeds the expected costs of litigation.

The traditional analysis lawyers use to predict case outcomes relies heavily on legal research. For it is primarily through

analyzing legal research results (e.g., case precedents) and applying them to the facts of particular disputes that lawyers are able to forecast the likely outcome of those disputes. This requires the lawyer to closely analyze the applicable elements and defenses of a claim, as well as the likely applicability of each, based upon a comparison of the facts of the dispute to the facts of the applicable precedents. For generations of lawyers, this analytical method has formed the backbone of predictive analysis.

Of course, experience plays a significant role as well. An experienced lawyer often has an intuitive sense as to the likely outcome of a case, even before looking closely at the applicable law. The lawyer can then balance this intuitive sense against the traditional element-focused analysis to predict likely case outcomes.

Traditionally, lawyers have memorialized the results of their research and analysis in formal office (i.e., research) memoranda. Lawyers have typically organized these memoranda around the elements and possible defenses of one or more causes of action that potentially apply. And while in recent years the use of formal office memoranda has declined somewhat, giving ground to less expensive alternatives, such as informal email memoranda and oral research reports, the underlying, element-focused predictive analysis lawyers use to evaluate likely case outcomes has not changed.

Unfortunately, this type of element-focused analysis is far from perfect. As any experienced lawyer can attest, it is a rough tool, even in the best of circumstances. There are a number of reasons for this, most of which are inherent in the nature of litigation. First, the factual predicate upon which a lawyer bases such an analysis depends primarily upon the accuracy of the client's story, at least at the preliminary stages of a dispute, when legal memoranda are widely used to assess the viability of potential lawsuits. Second, the law itself is frequently uncertain when applied to the facts of a particular dispute. A legal rule that seems relatively clear within the factual context of a particular precedent may not readily lend itself to application in a different factual context. Furthermore, lawyers cannot compare cases on their facts without determining which facts are legally relevant, and this sometimes requires considerable judgment.

Other factors confound the traditional predictive analysis as well. Individual judges have predilections (based upon their

political views, judicial philosophies, etc.) that can influence their decision-making. Courts, moreover, change in composition over time, which can undermine the reliability of older precedents. And while the lawyer drafting a predictive office memorandum can try to take these factors into account, there typically is little meaningful information to rely on in assessing how differences between judges might affect the possible outcome.

Case-specific factors can also skew the traditional predictive analysis. These include the equities of a given case, and the likeability (or lack thereof) of the particular parties and their lawyers. Similarly, the jury's likely assessment of the credibility of the parties and the witnesses adds another level of complexity to predicting case outcomes.

All of these factors are well recognized, and it comes as no surprise to any experienced lawyer that the traditional predictive analysis—that is, a precedent-focused analysis of the potentially applicable elements and defenses—will not always produce accurate forecasts of case outcomes. And for that reason, a seasoned lawyer's own experience in similar cases is often a helpful supplement. But personal experience has obvious limits as well, and it is of little help to less experienced lawyers when they counsel clients. Thus, forecasting legal outcomes often feels a bit like gambling, and as a result, advising clients on how to proceed with matters such as plea agreements can be quite daunting.

Fortunately, there may be some help on the horizon, as companies are now developing legal research tools that employ the power of data analytics to aid case forecasting. These tools hold significant promise as a supplement to the traditional element-focused predictive analysis. Instead of having to rely solely on their own experience to balance the results of the traditional element-focused analysis, lawyers may soon be able to rely on software products that mine data about past cases, and then run the data through algorithms to detect patterns. Those patterns can then inform predictions about likely case outcomes, based upon similarities between the facts, the courts, the individual judges, etc.

The large commercial online research services already offer some rather basic versions of these tools. WestlawNext, for example, has a tool called *Case Evaluator*, which provides averages and ranges for verdicts concerning a variety of different case types. It also allows the user to filter the results by jurisdic-

tion, damages, company, industry, and key terms. Lexis Advance has a similar tool called the *LexisNexis Verdict & Settlement Analyzer*.

In addition, some newer, specialized companies are developing more-sophisticated data analytics tools to drive case forecasts. For example, a group of computer scientists and law professors at Stanford University have created a company called Lex Machina that provides a sophisticated case-forecasting product to law firms and corporations in the area of intellectual property. It mines data from court filings, the United States International Trade Commission, and the United States Patent and Trademark Office, and then uses sophisticated algorithms to detect patterns and predict outcomes.

Probably the most significant challenge to using data analytics in this way is the difficulty of obtaining access to the necessary raw data, given that only some of the information is publically available. Lawyers have long used jury verdict reporters to assess potential recoveries, so that data can easily be mined. Likewise, court filings are available from databases such as Pacer. In addition, agency records and other governmental records are widely available. But a problem arises in gaining access to reliable *settlement* data. And since most cases settle, and most settlements are confidential, analyzing only the data currently available to the public yields incomplete information regarding likely case outcomes.

Much of this needed settlement information is *privately* available, however. Insurance companies, for example, have information about the settlements they pay out, as do corporations that are involved in litigation. And if these companies were willing to make this information publicly available—which would presumably improve the efficiency of the settlement process for all concerned—data analytics tools could provide much more thorough forecasts. But whether and when that will happen remains unclear.

In summary, the use of data analytics to predict legal outcomes has some hurdles to clear before it becomes a conventional tool. And it is unlikely it will ever fully supplant the traditional predictive analysis. But it does have the potential to become a valuable legal practice aid in the not-too-distant future.