

October 2015

Lawyering in the Information Age: Leveraging Analytics to Be a Better Attorney

By Bennett B. Borden and Amy Ramsey Marcos

Information is the critical raw material lawyers leverage and the most fundamental product they deliver to their clients. Understanding the client's position or circumstance is the foundation of lawyers' work, and the facts that underlie those positions and circumstances are overwhelmingly contained within electronic information. In the information age, knowing how to glean useful information quickly from large amounts of disparate data to reach more accurate conclusions is what distinguishes truly great lawyers. This is not rocket science; it is data science.

Data analytics is fundamentally about identifying patterns in information faster and more accurately. These patterns in data usually represent patterns in human behavior, and thus, can be used to piece together what happened, or, in some cases, what is likely to happen. Analytics can be used to look back in time in what is referred to as descriptive analytics, or can look forward in time with predictive analytics. In both temporal contexts, analytics is changing the way we approach our everyday lives and the way we practice law.

Both public and private entities are using analytics to achieve their objectives and drive engagement. Retailers like Amazon conduct data analytics on consumer behavior to make product recommendations. Netflix predicts what movies we might like, and CVS predicts with coupons might entice us to buy. Customer segmentation, which involves grouping consumers based on behavior, demographics and psychographics, is at the heart of many companies' engagement efforts. Customer segmentation allows businesses, such as hotels and airlines, to offer differential pricing models to optimize sales and maximize profits.

In the public sphere, colleges are using information available about prospective students to decide whether to admit them and, when admitted, which classes, resources and programs they will have access to. Municipalities are using analytics to deliver public services such as power distribution, trash collection, street maintenance and the like. Each of these analytics applications is based upon patterns in human conduct. The algorithms used to understand these patterns at the consumer and citizen level are the same algorithms lawyers are employing to understand the behavior of investigative subjects.

Like entities in the public and private sectors, lawyers leveraging analytics are gaining new insights for and about their clients faster and more accurately. The most well known form of predictive analytics in the legal space is predictive coding, also called computer- or technology-assisted review ("CAR" or "TAR" respectively). Predictive coding involves manually classifying a sample of data. The characteristics of the sample data are then used to build a mathematical model that predicts the classification of the remainder of the data in the larger population. In other words, the patterns within the data are used to build a model to predict the classification of new data.

The first known federal case to validate predictive coding in the context of ediscovery was *Da Silva Moore v. Publicis Groupe*, decided in February 2012 by Magistrate Judge Andrew Peck. In his decision, Judge Peck endorsed computer-assisted review in ediscovery as "an acceptable way to search for relevant [electronically stored information] in appropriate cases."¹ Judge Peck defined "computer-assisted coding" as tools that use "sophisticated algorithms to enable the computer to determine relevance, based on interaction with (i.e., training by) a human reviewer." Human reviewers code a "seed set of documents" and "when the system's predictions and the [human] reviewer's coding sufficiently coincide, the system has learned enough to make confident predictions for the remaining documents."² *Da Silva Moore* was merely the first of many predictive coding cases.³

Predictive coding has saved lawyers countless hours and significant cost over manual linear review. But this isn't the real power of using predictive analytics in the ediscovery space. The point of ediscovery isn't to get through documents, it is to get to the facts within those documents, and this is where the true strategic advantage of analytics comes into play.

¹ *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182, 183 (S.D.N.Y. 2012), *aff'd*, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).

² *Id.* at 183-84.

³ See, e.g., *Rio Tinto PLC v. Vale S.A.*, 306 F.R.D. 125, 127, 129 (S.D.N.Y. 2015) (discussing the accepted practice of using TAR for document review and noting that TAR should not be held to a higher standard than keywords or manual review); *Progressive Cas. Ins. v. Delaney*, No. 2:11-cv-00678-LRH-PAL, 2014 WL 3563467 at *9 (D. Nev. July 18, 2014) (discussing that the court would not have prevented the parties from agreeing to use predictive coding for ediscovery had it been included in the ESI protocol); *Dynamo Holdings P'ship v. Comm'r*, 143 T.C. 183, 194 (2014) (sanctioning the use of predictive coding to produce relevant information).

Knowing what happened and why before an adversarial party provides significant strategic advantage. For example, take a corporate client being sued by a former employee in a whistleblower qui tam action alleging a violation of the False Claim Act.⁴ The suit represents a significant threat to the company's reputational and financial interests. We targeted the data most likely to illuminate the facts and applied advanced analytics to 675,000 documents, and within four days, we knew with certainty that the allegations were completely baseless. All of this was accomplished before the answer to the Complaint was filed.

Armed with a firm grasp of the facts, we asked plaintiff's counsel to meet. Prior to the meeting, we voluntarily produced 12,500 documents that described the parties' positions precisely. We then met with plaintiff's counsel and walked them through the evidence, laying out all the facts. The case settled within days for what amounted to nuisance value based on a retaliation claim—without any discovery, and at a small fraction of the cost budgeted for the litigation. This example illustrates that the real power of advanced analytics is the *strategic advantage* that comes with counsel getting to an answer quickly and accurately.

The era of big data and advanced analytics offers both challenges and opportunities for lawyers looking to understand what happened and why. Big data is more data, and more data means the potential of gaining a more accurate and fulsome understanding of what occurred. Conversely, without the proper skills, more data can obscure visibility for those lawyers who are not able to see the forest for the trees. Analytics allow lawyers to search vast repositories of electronically stored information to draw out the facts that are critical to the question at hand and arrive at more accurate conclusions quickly and effectively. This capability goes beyond classifying information as part of a traditional document review.

Only a short step away from direct litigation is using advanced analytics for investigations, either in response to a regulatory inquiry or for purely internal purposes. Corporate clients can be faced with circumstances where determining the existence of a problem, and the scope of the potential problem, is critically important. Without a clear understanding of the facts, management cannot move forward confidently, but with a clear understanding of what occurred and what didn't, they can move forward with certainty.

For example, a manufacturing company decided to outsource its product safety function. A director of the safety division being outsourced demanded four times the severance pay being offered and threatened to report alleged safety violations to a regulator if his demand was not met within forty-eight hours. We pulled 275,000 electronic documents from the division and applied analytics to determine whether the alleged violations had any merit. In less than twenty-four hours we were able to prove definitively that they didn't. The company was able

to move forward with certainty because we were able to quickly and accurately ascertain the truth.

Understanding the facts with certainty is critical for litigation and investigations. But there are many other legal circumstances where this is also true, such as merger or acquisition due diligence. Merger or acquisition agreements typically include standard representations and warranties that the information disclosed by the target is correct and has a reasonable basis under Generally Accepted Accounting Principles. If the acquiring party can prove that the target failed to disclose material information or that the information was incorrect, the agreement often contains an indemnity provision such that the acquirer receives a purchase price adjustment, sometimes for many millions of dollars. These indemnity provisions are rarely acted on, however, because they require a claim to be filed within thirty to ninety days, and it is rare that an acquirer can prove up a claim in that short of a period. Data analytics is changing that.

In several engagements we have used our fact development strategies based upon advanced analytics to obtain millions of dollars in indemnity claims for our clients. As soon as the merger or acquisition is completed, we analyze the target's information systems and test the accuracy and validity of its disclosures. Within a matter of days we know whether and to what extent the disclosures were inaccurate. The resulting indemnity claims are based not upon guesses but upon information from their own systems and communications. It is very difficult to defend against words out of your own mouth.

As demonstrated in the examples above, analytics can be used in several aspects of lawyering to create a strategic advantage. The advantages created by using analytics cannot be understated when the results are compared to a manual review of information. When a lawyer using analytics goes up against a lawyer who fails to use analytics, the results are inevitably skewed in favor of the lawyer that can reach the answer quickly and more accurately.

All of these examples pertained to situations where we were looking back in time to describe what happened. Wouldn't it be better to try to look into the future and predict when something bad will happen and act to prevent it? Well, it turns out analytics can do that, too. Analytics is being used in the commercial space to understand patterns in consumer behavior. Information is gathered regarding what we buy, what we do, and our demographic and psychographic characteristics. Millions or even billions of these data points are gathered and a model is built upon the resulting patterns within it. These models are astonishingly accurate in predicting what we are going to buy, do and even think.

Using the same analytics, we have built algorithms that can identify signs of misconduct as they are happening rather than waiting for the results of misconduct to manifest themselves in situations that damage the company's reputation or finances (or both). In building these algorithms we took large sets of data (primarily email, chat and texts) that had been the subject of litigation

4 The examples in this article are based on real cases and real clients.

or regulatory investigations, and the relevant documents identified. Using these documents, we built models using the most advanced work in data science regarding text, social network and sentiment analysis. We worked with experts in intelligence and law enforcement as well as social scientists and criminal psychologists. It turns out that just as there are patterns in all human conduct, there are patterns in *misconduct* as well. The resulting models we built are astoundingly accurate in detecting and preventing misconduct and are changing compliance from a largely reactive to a more proactive endeavor. We believe such proactive compliance systems will become the norm.

For example, JP Morgan Chase is testing a program that it plans to deploy company wide in 2016 where it is analyzing email, chat and phone transcripts to identify collusion or the hiding of intention.⁵ Lawyers that understand how these analytical tools work can help their clients use them to promote corporate compliance. Lawyers that do not understand these predictive analytics are of no use to clients until the corporate misconduct has been revealed through other, usually more public, means. Clients will appreciate the lawyer that can help keep them out of trouble proactively in addition to defending them as the need arises more than the lawyer that can only assist them reactively once the damage has been done.

However, there is a danger to these kinds of proactive monitoring systems. The algorithms we built are focused on finding those who are violating policies or laws. But these algorithms can be tailored to look for any pattern in human conduct, including the opinions we hold, how we vote, what kinds of marriages or families we have. Thus, there is a newly developing area of ethics called data ethics that we as legal and Information Governance professionals need to be aware of because these very powerful analytics can bring immense insights that give us strategic and business advantage. But they can also be misused.

For example, predictive analytics are being used by several major police departments to try to identify individuals who are most likely to commit crime in the future.⁶ The departments use data on prior criminal history, social networks and affiliations, and many other data points to try to predict who might be involved in crime, especially violent crime. The departments then focus resources to try to intervene to prevent any such crime. Some of these efforts are laudatory, such as providing drug or mental health counseling and job services. But other efforts are more disconcerting, such as increasing police monitoring or even public shaming. In one instance in Kansas City, the police during a community meeting put up mug shots of those in the community it thought were the most likely to commit crime even though these individuals had been accused of no actual crimes. Thus, while we laud the use of analytics by the government, what it does with the results of such analytics can be deeply troubling. It is our responsibilities as legal practitioners to understand the implications of this emerging field of data science and its social, legal and ethical implications.

Information is key to every aspect of a lawyer's job, from defending a company in the face of litigation to proactively optimizing regulatory compliance. Lawyers who think they can continue to practice law without analytics are putting themselves and their clients at a disadvantage. The variety and complexity of information available to lawyers will only continue to grow, which makes it increasingly important to leverage data efficiently and effectively. Lawyers who can leverage analytics to reach answers quickly and more accurately are leading their clients to better outcomes and adding the kind of value clients deserve. Lawyers who refuse to understand and embrace analytics are like the dinosaurs the day before the meteor hit – they are already extinct, they just don't know it yet.

⁵ See Hugh Son, *JPMorgan Algorithm Knows You're a Rogue Employee Before You Do*, BLOOMBERG (Apr. 8, 2015), <http://www.bloomberg.com/news/articles/2015-04-08/jpmorgan-algorithm-knows-you-re-a-rogue-employee-before-you-do>

⁶ John Eligon and Timothy Williams, *Police Program Aims to Pinpoint Those Most Likely to Commit Crimes*, N.Y. TIMES (Sept. 24, 2015), http://www.nytimes.com/2015/09/25/us/police-program-aims-to-pinpoint-those-most-likely-to-commit-crimes.html?_r=0

Bennett B. Borden is a Partner at Drinker Biddle & Reath LLP in Washington DC. He is the firm's Chief Data Scientist and Chair of the Information Governance group. He is also Chair of the Cloud Computing Committee, and Vice Chair of the Electronic Discovery and Digital Evidence and the Internet of Things Committees of the Science and Technology Division of the ABA. He is also Founder and Chair of the Information Governance Initiative.

Amy Ramsey Marcos is an Associate at Drinker Biddle & Reath LLP in Washington DC. Amy assists clients with identifying their business, litigation, and regulatory obligations to optimize information governance practices. She has broad experience developing records retention schedules and building change management plans to successfully implement information governance policies. As part of the Secretariat of the Pharmaceutical Records and Information Management Organization (PRIMO), Amy has worked closely with Member companies to develop solutions, strategies and tools relevant to managing records and promoting effective information management, including facilitating the development of a Model Records Retention Schedule focused on the records of biopharmaceutical companies.



Drinker Biddle®

www.drinkerbiddle.com

CALIFORNIA | DELAWARE | ILLINOIS | NEW JERSEY | NEW YORK | PENNSYLVANIA | WASHINGTON DC | WISCONSIN

© 2015 Drinker Biddle & Reath LLP. All rights reserved. A Delaware limited liability partnership. Promotional Materials 1092015. One Logan Square, Ste. 2000, Philadelphia, PA 19103-6996 (215) 988-2700 office (215) 988-2727 fax
Jonathan I. Epstein and Andrew B. Joseph, Partners in Charge of the Princeton and Florham Park, N.J., offices, respectively. This Drinker Biddle & Reath LLP communication is intended to inform our clients and friends of developments in the law and to provide information of general interest. It is not intended to constitute advice regarding any client's legal problems and should not be relied upon as such.