Law

Big data: legal firms play 'Moneyball'

Is the hunt for data-driven justice a gimmick or a powerful tool to give lawyers an advantage and predict court outcomes?

BARNEY THOMPSON - LONDON

For the lawyers who ply their trade in the US District Court for the Northern District of California, the ability to read the mind of Judge Richard Seeborg would be extremely useful.

Until a few years ago, trying to predict how he or any other judge might rule depended on encountering them enough times in court, or exchanging tips with colleagues on which arguments the judge had recently found persuasive.

These days, however, Judge Seeborg and his colleagues produce a rich seam of data that is being mined by a group of companies threatening to upturn the way the legal profession operates. Just like a professional baseball or tennis player, Judge Seeborg, who hears a full range of criminal and civil cases, has his personal statistics that are now being rigorously collected and scrutinised.

Anyone hoping to bring a class-action lawsuit should know, for instance, that he has been presented with such cases on 37 occasions, allowing 51 per cent of them to proceed in full. By comparing him to the 670 other US district judges, lawyers can see where he ranks on granting such cases, how long they take to complete, how likely he is to make a summary judgment before trial or whether lawyers should expect to be forced to plead the case to a jury.

Under the common law system in the US, Britain and several other countries, the validity of legal arguments depends heavily on precedent. So a Californian lawyer might like to know, for instance, that Judge Seeborg's favourite US Supreme Court decision is Ashcroft v Iqbal, a 2009 case concerning the responsibility of a senior official for the actions of a subordinate, which he has cited on 423 occasions. He also likes Balistreri v Pacifica Police Dept, a 1990 appeals court decision stemming from a domestic violence case, which he has mentioned 312 times.

In Philip K Dick's short story *Minority Report*, a trio of "precogs" plugged into a machine are used to foretell all crimes so potential felons could be arrested before they were able to strike. In real life, a growing number of legal experts and computer scientists are developing tools they believe will give lawyers an edge in lawsuits and trials.

Having made an impact in patent cases these legal analytics companies are now expanding into a broad range of areas of commercial law.

"This is not about replacing judges," says Daniel Lewis, co-founder of Ravel Law, a San Francisco lawtech company that built the database of judicial behaviour. "It is about showing how they make decisions, what they find persuasive and the patterns of how they rule."

Judge Seeborg is aware he is being monitored by Ravel but declined to comment for this article.

Mr Lewis says the capabilities of legal analytics are rapidly evolving. "A lot of lawyers think of their case as a special snowflake — that the facts are unique, that nothing like it has ever happened before. And it's just not true."

The tech revolution has not only seen an explosion in innovation, it has also prompted a surge in legal disputes over intellectual property.

Some of these lawsuits were the result of competing claims among companies over who invented what first, but by the mid-2000s a new breed of third-party entities had arrived — buying up patents and using the threat of litigation to extract a quick payment from their targets.

To some, they are "patent trolls", the shakedown artists of IP; to others, they hold businesses to account for plagiarising the work of genuine innovators. Either way, the amount of IP litigation ballooned and there are now thousands of patent litigation cases in the US every year.

Until recently, a large portion of these cases were heard in the eastern
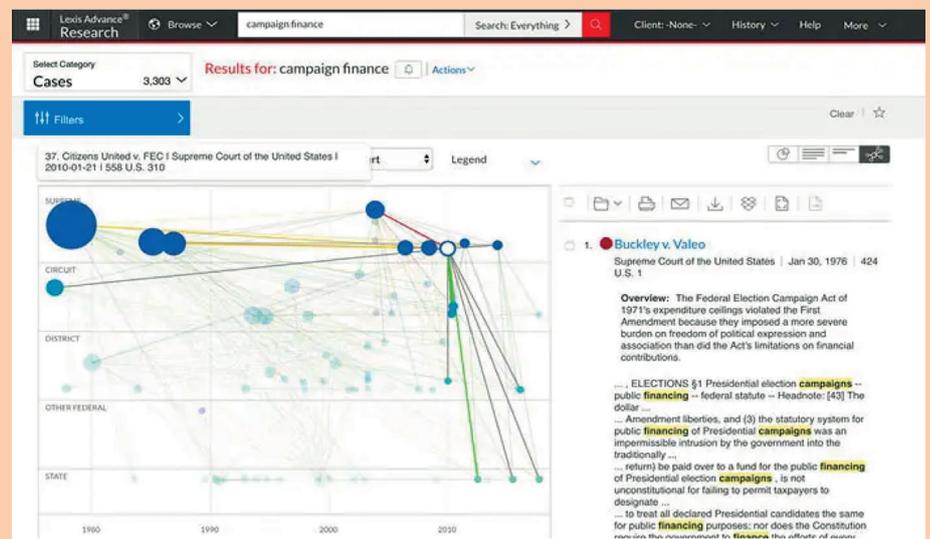
Programs such as Ravel's 'Context' can build up a search history of cases and outcomes

district of Texas, which became the number one venue for patent litigation thanks to its speed in processing claims, the probability a case would proceed to trial, and the friendliness of its juries to patent holders suing for infringement — which in turn meant bigger payouts.

According to PwC figures, between 2013 and 2017 median damages in US patent claims were $1.9m when judges made awards but $10.2m when a jury did. Since claimants were able to "venue shop" for the most favourable court, they queued up to sue in Texas.

However, in May 2017 the Supreme Court in effect ended venue shopping for patent cases; instead, litigation would have to take place in the state where the defending business was based. Suddenly it was vital to know the records of judges, lawyers and courts in jurisdictions such as Delaware, California — Judge Seeborg's domain — or New Jersey.

In an office above a nail bar in the Silicon Valley town of Menlo Park, the legal data business Lex Machina amasses as many rulings from US courts as it can get its hands on. Josh Becker, chairman, claims that three-quarters of the top 100 US law firms are Lex Machina clients.

$10.2m

Median damages in US patent claims in 2013-17 when awarded by a jury, compared with $1.9m when awarded by a judge

"These days everyone has analytics in baseball," he says, a reference to Moneyball, the book by Michael Lewis that inspired the sporting mania for using often obscure performance stats. "Soon it will be the same with the law. And once everybody is doing that, it will be about how well you are using that data."

The idea behind Lex Machina — which, like Ravel, is now owned by the data company LexisNexis — was to enable companies and their lawyers to assess their chances of winning a case as soon as the notice to sue arrives. The sort of information that might be analysed includes how many times the opposing lawyer has filed certain
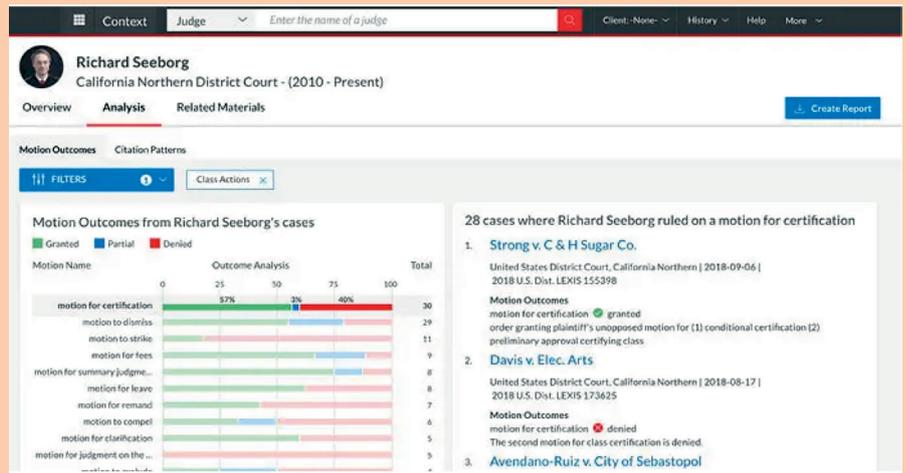
Case history: Californian District Court Judge Richard Seeborg

types of lawsuit, in which court, with what success rate, who they have represented, and which attorneys they have faced. Once a judge has been assigned to the case, legal research companies can provide statistics on his or her record as well.

IP litigation was the spark for Lex Machina, which emerged as a Stanford Law School start-up, but once it had access to LexisNexis's broader database, it expanded into several other "high-volume" areas of the law, from employment and tax to product liability, medical negligence, insurance and bankruptcy. Other legal analytics companies followed a similar path.

To get ahead of the competition, US commercial lawyers set up alerts on Pacer, the electronic court records system, for when a new case has been filed against a company in their field. Once an alert pops up, says Christian Mammen, a San Francisco partner at law firm Hogan Lovells, the in-house legal team of the affected business will start getting calls within minutes from lawyers offering to come to the company's defence.

"Three hours later they'll be getting full pitches," says Mr Mammen, who has been litigating IP and tech cases since the dotcom boom in the late 1990s. That means offering advice on venue, strategy and personnel, backed up by the data. "You have to be ready with a compelling argument as to why you can handle their case best."

As the potential use of these analytical tools spreads, more companies have emerged to meet demand. Among them is Premonition, based in New York, which shows the litigation history of judges, lawyers and law firms, including win/loss

rates for trials that are benchmarked to competitors, the success rates of different types of motion in individual courts and a database of who sues and gets sued most often.

Bloomberg Law's Litigation Analytics and Los Angeles-based Gavelytics have similar functions, while Casetext and Judicata both offer deep-dive analysis of the legal documents most relevant to the case a lawyer is fighting, such as similar briefs filed by other firms, relevant case history and judges' citations, often down to the most cited paragraph.

Blue J Legal, a Toronto-based business, mines Canadian court rulings on tax and employment. After getting clients to answer questions on their individual circumstances, weighting each factor depending on the case in hand, its software produces a list of similar cases, relevant citations and a stark assessment in percentage terms of the client's chances of winning or losing.

The march toward data-driven justice has one big flaw, however: most of the data are missing.

The vast majority of civil litigation, possibly 90 per cent, is dropped or settled out of court, which means the documents from the case are never made public. The reason is financial — "it is easy to start a lawsuit but expensive to continue it", says Mr Mammen.

Furthermore, while court documents are relatively easy to accumulate in the US, in other jurisdictions, including Britain, finding case law is much harder; the UK's main online database, Bailii, has improved public access over the past two decades but suffers

from gaps in historical documents and struggles to keep up with new ones.

In legal areas where data are hard to come by, the law firm Herbert Smith Freehills has created what it says is a "mini-hive mind" of real-life lawyers to assess a dispute. Called Decision Analysis, the project attributes risk weightings to each stage of a case — assessing if a business failed in its duties to a customer, for instance, and which factors might determine damages.

Lawyers think of their case as a special snowflake — that nothing like it has ever happened. And it's just not true

Daniel Lewis, co-founder of Ravel Law

"In 'slips and crashes' tribunals, say, you have a rich pool of data points," says Donny Surtani, a commercial disputes partner at HSF and former financial professional. "In commercial litigation it is hard to find five similar cases, let alone 5,000. But clients are increasingly asking for probability-based terms [to express outcomes] . . . So this is trying to capture the legal team's judgment and help a client better understand the risks of litigation."

For all the hype about analytics, Pablo Arredondo, a fellow at the Stanford Center for Legal Informatics and co-founder of Casetext, says there are limits to what data can show.

"The judge analytics demonstrations I have seen to date oscillate between the blindingly obvious and the statistically irrelevant," he says. Knowing where to file a lawsuit or estimating how long a case will run for is undoubtedly important in developing your strategy, he adds, but some case histories — such as Ashcroft v Iqbal — are so commonly cited as to be useless in profiling an individual judge.

Yet proponents of legal analytics insist that it is only a matter of time before there will be massive data sets to cover wide areas of the law. They make a bold claim that this will lead to a better justice system. From showing which cases are a waste of time and money to pursue, to exposing which judges are mavericks and outliers, supporters say the insights from using data can improve the way that the legal system works.

"As we expand our data set, we hope the justice ministry and the relevant regulators will look at how justice is applied [across the country] and where the inconsistencies are," says Edward Bird, chief revenue officer at Solomonic, a UK company formed by a group of commercial lawyers and data scientists that aims to replicate some of the US analytics models.

"The top judges in the UK commercial courts are extremely consistent but in lower-value cases [in courts around the country], you shouldn't be in a situation where you get a very different result depending on who you're in front of."

As with the malfunctioning pre-crime system in *Minority Report,* there are dangers with putting too much emphasis on what the data analytics suggest.

"The big challenge is, if eventually we do get to the point where the numbers do the predicting, do you start to short-circuit the [legal] process?" asks Mr Bird. "What does that do to access to justice for a claimant who is told, 'your chances are no more than this number and you won't get your day in court'? That's a big ethical question."

'FantasySCOTUS':
Geeks try to call Supreme Court's market moving cases

The mixture of sport and legal data analysis in the US extends even to the highest court in the land. In the law school equivalent of a virtual sports league, FantasySCOTUS (short for the Supreme Court of the United States) invites law geeks to predict the outcome of some of America's most important legal battles.

Points are awarded for the right forecast, for guessing the split among the nine justices and for predicting who was in the majority and who dissented.

The online game was set up as a hobby for the type of law school nerds who reject the usual fantasy sports leagues. However, there is a serious side, given the capacity for Supreme Court decisions on controversial business cases to move markets.

In 2013, the Court ruled on whether Myriad Genetics could be the exclusive provider of tests used to screen patients for cancer, and whether human genes could be patented — the underlying legal point.

The decision was a complicated compromise but initially the market thought Myriad had won and the stock traded up. A few hours later, however, it became clear the company had in fact lost exclusivity on its tests and the stock sank 20 per cent in two days.

Out of about 75 Supreme Court cases a year, about five are market moving, estimates Dan Katz, co-founder of LexPredict (which set up FantasySCOTUS) and a professor at Chicago-Kent College of Law.

As the Myriad case showed, "this isn't a well understood area — it's not like earnings announcements or central banks setting rates," he says. "You've got to know the law and what it means for the market."