Machine Learning: Making the Data Deluge Work For You

By Bob Krantz, CEO of Mindseye, and Jeff Fehrman, Chief Strategy Officer of Mindseye

One of the overarching issues in eDiscovery today is the problem of data. There's so much of it, with more being created every second, making it harder and harder to identify what we are looking for. Technological innovations are constantly evolving to address the issue and they go by several names, including artificial intelligence, big data, and predictive analytics, to name just a few. Unlike computer systems that follow static program instructions, machine learning uses algorithms that consume large quantities of data, learn from it, and make predictions or suggestions about the data. So how exactly is machine learning being used today and what can it do for eDiscovery?

Machine learning is everywhere, even when you can't see it

Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. Sometimes confused with data mining, which focuses on extracting data for use in a specific domain, machine learning uses data to improve the program's own understanding.

It's important to note that machine learning isn't just about automation — its purpose is not simply to process more information faster. Machine learning is transformative; it enables us to generate predictive models, uncover obscured patterns, and deliver insights in ways that human users simply cannot.

The best implementations of machine learning are invisible. When the onus is on the user, sorting through and filtering data is slow, cumbersome, and inefficient. Machine learning, when done right, provides a new user experience, one that complements cognitive thinking and transforms users into super-users.

Machine learning doesn't replace humans — it complements them

Machine learning sounds futuristic — and, to some people, even ominous — but, in fact, it's already being used and implemented today in many facets of our everyday life. Spam filters, Amazon's recommendations, dating websites, and Pandora's customized playlists are all examples of machine learning. There are a few high-profile examples of machine learning applications — one recent example is ***the AP using robots to write stories*** — but while these stories may capture the headlines, they don't point to a "Terminator" style takeover. In the AP example, it has been reported that no jobs were lost due to the introduction of robot journalism and quality has improved as the automatically generated stories contain far fewer errors than those written by humans. For the most part, the machines correlate data, build predictive models, and deliver insights, but they don't actually make decisions nor are they executing on the insight or recommendations provided — that still requires human involvement.

Machine learning in the enterprise

When it comes to the vast volumes of data being created and captured on a continual basis, most organizations view it as something to be "dealt with" — it's a storage and management issue. To take advantage of the transformative power of machine learning, enterprises have to adopt a different mindset; they need to view the data as something to be leveraged for strategic and competitive advantage. This could mean using machine learning as part of their core offering; but it could also mean taking advantage of machine learning to improve internal operations and administrative functions.

In the legal sphere, there are already examples of machine learning helping to transform the ***contract review*** and ***lawyer selection*** processes. eDiscovery is another area where machine learning is being utilized, but much of the focus to date has been confined to the document review phase. We believe the full power of machine learning should be exercised as far left on the Electronic Discovery Reference Model (EDRM) as possible.

The promise for eDiscovery

The fundamental problem eDiscovery has today is one of "findability." ***How do you find the tiny number of relevant documents among the vast volumes of enterprise data?*** Machine learning provides a way to prioritize what you want, leveraging users' actions to teach and improve the system. Again, the point is not to replace legal professionals with machines — the machines are simply using their algorithmic and processing power to present the user with better, more relevant results. The legal professional is still needed to make decisions that a machine simply cannot. But because the machine is capable

of looking at information from many more perspectives than humans can, the overall process is more efficient and cost-effective.

But in order for machine learning to transform users into super-users in the eDiscovery arena, it has to do three things.

- **Organize** — Trained computers can review and model — and classify — data much more efficiently and comprehensively than humans can. Using classification to organize results helps limit the "distractions" that take away from the key objective: finding relevant documents for your case.
- **Expand** — Machine learning can suggest similar areas of interest to help uncover patterns that may not present themselves clearly when you look at documents on a record-by-record basis. Systems should learn from user behavior to guide them to other documents of interest (think Amazon's "if you liked this, you might like that" recommendations).
- **Enlighten** — Discovery is often about learning something you didn't previously have knowledge of. Insight is meant to drive action. Machine learning should help guide you effectively in developing strategy and achieving objectives.

Make your data work for you

Enterprise data growth isn't going to slow its pace any time soon. In fact, it continues to accelerate. eDiscovery — which is about identifying, collecting, and producing all the documents that are relevant to a particular case or investigation — is only going to get increasingly more challenging in the face of this growth. Machine learning provides the ability to prioritize what you want and use human expertise to teach — and improve — the system. Technology will continue to evolve, and conducting an investigation or an eDiscovery matter may one day become as simple as shopping on Amazon thanks to advancements in machine learning.

MINDSEYE HELPS YOU GET TO THE INFORMATION YOU NEED TO MAKE EFFECTIVE DECISIONS

The Mindseye platform is a powerful and comprehensive discovery solution that can provide you with the visibility you need to:

- *Uncover valuable information about your data*
- *Quickly identify and cull junk data and uncover relevant business information through a comprehensive, single view of your results*
- *Integrate your eDiscovery process with your information governance strategy*
- *Refine and improve your information governance strategy and develop data classifications*

For more information about Mindseye or to request a demo, please visit ***www.mindseyesolutions.com.***

Bob is an innovative problem solver and, as CEO of Mindseye, is responsible for the vision and leadership of the company. Over the past 12 years, he has played a key role in helping organizations address their growing concerns related to information. Bob has experience in all aspects of information discovery and, serving in both sales and consulting roles, he has assisted companies in cases ranging from bet-the-company antitrust litigation and SEC investigations to internal and employment suits. He provides clients with innovative ways of establishing workflows, maximizing technology investments, and minimizing data volume by focusing on finding the key information that drives decisions that matter. In 2008, Bob co-founded Mindseye with two colleagues to create a technology company that provides companies with solutions to address growing data volumes by connecting the people who have knowledge with the information needed to make effective decisions. Under Bob's leadership, Mindseye has seen steady growth and is poised to change how companies view and manage their information. Contact Bob at bkrantz@mindseyesolutions.com or (888) 770-3876.

Since joining Mindseye, Jeff has helped the team design and launch Mindseye's appliance and mobile offerings and established key technology partnerships. Jeff is on the Board of Governors for the Organization of Legal Professionals (OLP) and is an active member in the Sedona Conference and AIIM. He frequently speaks on innovations and obstacles facing corporations and law firms today. He has developed classes approved by the Continuing Legal Education (CLE) boards in the states of California, New York, Pennsylvania, and Virginia and has trained hundreds of attorneys in the technical aspects of digital forensics and electronic discovery. Jeff is co-founder of the EDD Blog Online, one of the electronic discovery industry's premier resources for news and related content, with over 4,000 articles to date, and co-author of Focused Discovery: Perspectives on the Future of eDiscovery and Information Governance. Contact Jeff at ***jfehrman@mindseyesolutions.com*** *or (571) 483-0639.*