WHITE PAPER

iMerit

DATA SERVICES & ANNOTATION FOR COMMERCE

IT'S ALMOST SECOND NATURE

Shopping at a mass market retail chain such as Walmart, Target, or Amazon, comes with the often-tacit assumption that the retailer is over time learning quite a bit about you, your likes, dislikes, and behavior. We are (most of us) creatures of habit, and those habits form patterns that savvy businesses exploit through timely – and targeted – offers. At its simplest? Buy a razor, for example, and you can expect to see a coupon with your receipt, for razor blades – or shaving cream, or even skin moisturizers. We know intuitively that it wasn't a human being generating the offer – at least not directly. No matter. Most shoppers don't understand the mechanics of how a store discount coupon is generated, how a Machine Learning algorithm ties a specific purchase to a logically adjacent marketing offer. Most just shrug and go on with their day – after setting aside the coupon for the next time they shop.

But if a shopper searches an E-Commerce site for work pants and is taken to a product page for plants and shrubbery? Shrugs instantly give way to annoyance (at best) over the inability of the search engine to provide at least a relevant result. Most people are reasonably forgiving if the result is at least in the ballpark, if not a strong match to their query. But fail the relevancy test and alarm bells go off. Some shoppers might pick up the phone and demand a human worker who can get them what they want. But others will just leave the site in frustration and look elsewhere. And therein lies a multi-billion-dollar risk that leading traditional and E-Commerce retailers are loath to accept.

Roughly a quarter century after Amazon started selling books online, and 20 years after Walmart launched Walmart.com, even some mom and pop retailers have begun to embrace E-Commerce as a core business. That has placed a commensurately higher premium on Artificial Intelligence engines that can mimic the capabilities of everyone from a floor salesperson to a customer service representative. The "intelligence" of E-Commerce platforms has become a critical factor for the global retail community, not surprising given what's at stake: potentially significant variations in the projected lifetime value of customers – and the cost (primarily in marketing dollars) of acquiring and retaining those consumers.

The average Walmart shopper who just goes to the chain's brick and mortar locations, for example, spends about US $1,400 a year. Walmart.com only customers

Source: The Walmart Museum

spend far less, roughly $200 a year. But those who shop both in store and online – where the mass marketing behemoth captures (and analyzes) a far richer set of historical data – spend upwards of $2,500 annually. That explains why industry analysts forecast that global E-Commerce and traditional retailer expenditures for AI-driven solutions – expected to exceed $3 billion in 2020 – will roughly triple in the next two years, and grow six-fold, to more than $19 billion by 2025. That's an awful lot of data categorization, attribute mining and algorithm coding – and hundreds of thousands of human worker hours devoted to data annotation and analysis.

TRAINING AI FOR BOTH THE SALES FLOOR AND BACK OFFICE

Retail investment in Machine Learning AI algorithms falls into two general buckets: the customer experience (a combination of an E-Commerce site's search engine and customer service bots) and the operations functions (stocking, fraud detection, and other so-called back office processes) that in both the physical and virtual worlds have begun to rely increasingly on machines to augment, if not entirely replace, humans.

Much of the industry's customer experience investment in Machine Learning and AI aims to enhance the ability of E-Commerce search engines to deliver unerringly relevant search results and product recommendations. There's also a significant focus on sentiment analysis of customer reviews and comments, data mining, and related initiatives.

Much of the spending in AI-driven operational initiatives, conversely, has been for brick and mortar retail. A growing number of major big box retailers already rely on advanced sensor and camera-equipped robots to roam the aisles of their physical stores to run inventory checks. Robotic arms outfitted with sophisticated sensors and cameras have been trained to examine the shelves, top to bottom, detect missing (and misallocated) merchandise, and navigate the stores aisles, much as an autonomous vehicle might navigate the streets of a small town.

The algorithms that operate those robots, though, have to match the capabilities of the inventory checkers they are replacing. Not only do they need to identify empty spaces on a product shelf; in a supermarket, for example, they have to understand the difference between a can of corn and a can of peas, mistakes in price tags, and even mis-stocked products – such as cans of dog food placed in the canned fish section.

Data annotation in this scenario leverages a combination of Computer Vision techniques – the use of semantic segmentation and bounding boxes to identify and categorize individual products – sometimes Natural Language Processing to evaluate signage, and even on occasion Geospatial Intelligence for device navigation within the store and shelf locations. One such example: a combined discipline project employed drones inside a major mass market retailer to identify empty shelves in need of restocking, the specific products (based on shelf labeling) and even mistakes in product pricing.

Human-in-the-loop expertise (and process) is particularly crucial to successful coding in the early, algorithm training stages of Machine Learning. The human touch gradually declines as the Machine Learning curve begins to bend higher. As the algorithm gains “intelligence” the role of data analysts transitions from helping to train the algorithm to correcting its mistakes.

It’s common for an initial, pre-production algorithm release to generate no better than 5-10 percent accuracy, before developers feed it a high fiber training data diet of millions of data points generated by data analysts. As the algorithm continues to learn from categorized data it will eventually achieve an accuracy rate in the high 90s – with anything north of 95 percent the benchmark for commercial success.

While the penalty for failure is not quite the same life-and-death reckoning that autonomous vehicle developers face, that impact on E-Commerce and even brick-and-mortar retail adds up quickly in two metrics: an increased cost of customer acquisition that follow bad in-store or online shopping experiences, and the cost of customer retention when retailers are forced to compensate for those failures in customer experience.

Given the pressure to outsell the competition, though, E-Commerce and other retail businesses often take an algorithm into commercial use before they have finished their work. That places an even greater premium on steady and reliable course corrections in algorithm development based on data analysis and annotation.

The value of human data annotation for retail Machine Learning projects can also be seen in the application of cross disciplinary skills in Natural Language Processing and Computer Vision annotation – and even institutional knowledge developed in other industry sectors, such as Autonomous Vehicles and Geospatial Intelligence. Think of an E-Commerce product page with a combination of text and images – or an autonomous stock inventory robot employing AV algorithms and geospatial positioning to navigate a retailer’s brick and mortar stores.

THE 20 BEST EXAMPLES OF USING ARTIFICIAL INTELLIGENCE FOR RETAIL EXPERIENCES

1. Lowes Uses Robots To Locate Items
2. Walgreens Uses AI To Track Flu Spread
3. Sephora Makes It Easy to Find Makeup
4. North Face Helps Customers Find The Perfect Coat
5. Neiman Marcus Uses AI For Visual Search
6. Taco Bell Helps Customers Order Tacos On The Go
7. Macy's Adds AI To In-Store Experience
8. Walmart Deploys Robots To Scan Shelves
9. ThredUp Uses AI To Remember Customer Preferences
10. Amazon Eliminates Cashiers With AI
11. Uniqlo Can Read Minds With AI
12. West Elm Connects Style And Products
13. Sam's Club Makes Warehouse Shopping Simple
14. Olay Uses AI To Personalize Skincare
15. Kroger App Customizes Product Recommendations
16. H&M Uses AI To Keep Popular Items Stocked
17. Zara Streamlines Order Pickup With Robots
18. Starbucks Bot Makes It Easy To Order Coffee
19. American Eagle Creates Fitting Rooms Of The Future
20. Rebecca Minkoff Designs AI-Powered Smart Store

Source: Forbes

THE EARLY STAGES OF MACHINE LEARNING IN RETAIL

The early stages of a retail data annotation project emphasize the production of algorithm training materials – data points, confirmed by data analysis, that can be fed to an algorithm to help it learn by example. Depending on the algorithm and its design, data annotation can focus on examples of customer intent as they type search queries, categorizing products through a multi-level retail taxonomy, or even pattern recognition on suspicious exchanges between an affiliate vendor and his or her customer.

At a certain point – weeks, months, or years, depending on the complexity and ambition of the project – the process transitions from algorithm training to accuracy and quality control in the form of correcting results. The goal is refining the algorithm by smoothing the "rough edges" in judgement.

In a recent search engine algorithm development project for a major American big box retailer, for example, the data analysts helping to train it reviewed roughly 7,000 different product category taxonomies to produce training data with correct product search results. Each taxonomy represented a multi-level product search path; for example, Electronics -> Television -> Samsung -> 55-inch 4K LED. By the time the algorithm was able to generate search results on its own, the algorithm could narrow down those 7,000 taxonomies to three or four potential answers, with data analysts now focused on refining an algorithm that -- if one of the answers was correct – was at least searching in the right neighborhood.

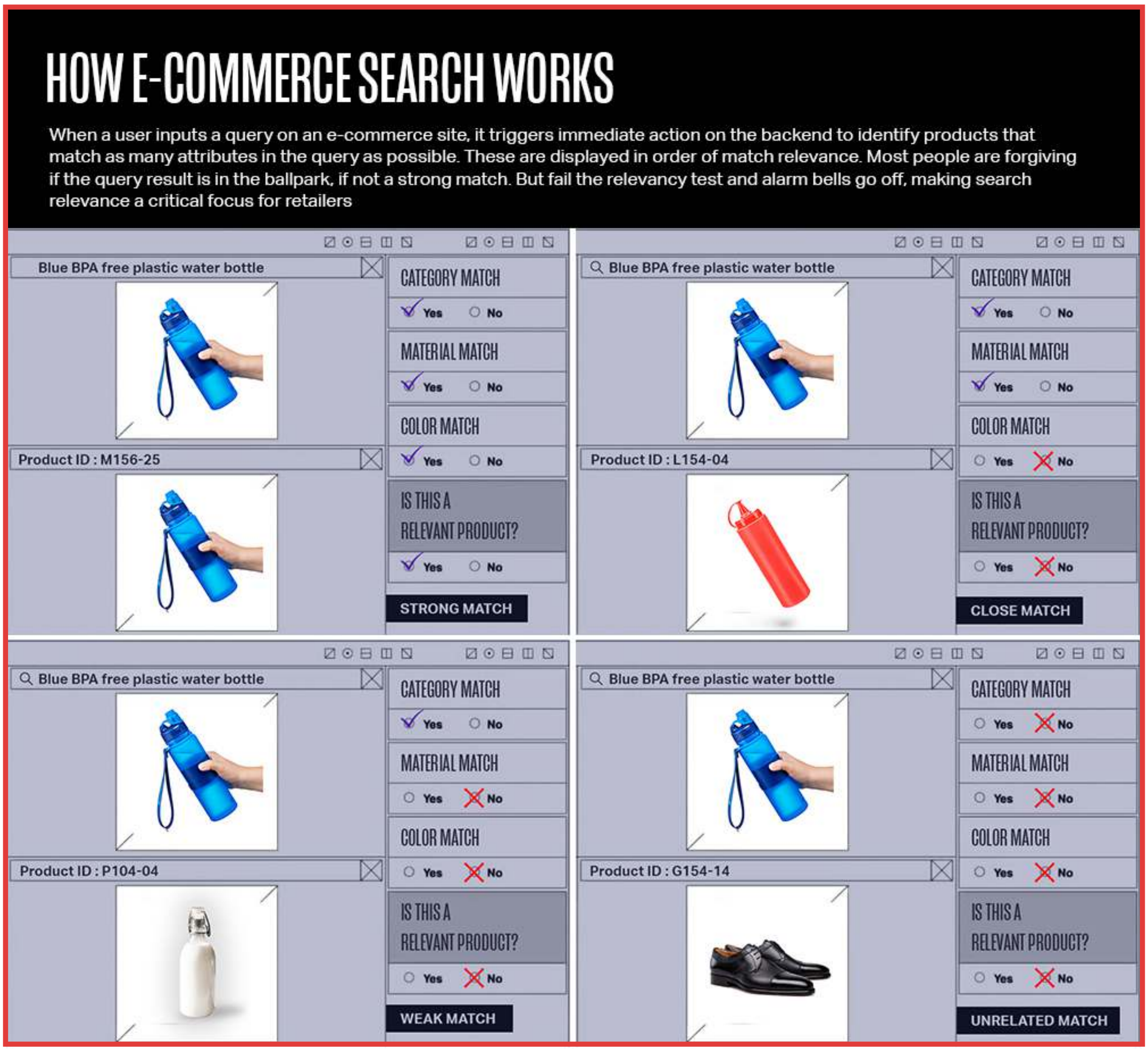
THE HUMAN FACTOR IN MODERN DAY RETAILING

Evaluating each taxonomy, identifying errors, and correcting them, requires the application of human judgement and understanding. Before an Artificial Intelligence engine can manage the millions of product search requests a major retail site typically takes in every month, someone has to show it the correct path(s) and the logic behind them.

Optimizing a large E-Commerce search engine typically involves a review of tens of millions – sometimes hundreds of millions – of data points, with each one representing a specific SKU, and thousands of product taxonomies (the path that took the search engine from request to result). The training process typically employs a four-level rubric that applies a value to each search result. Consider a shopper's search for a BPA-free blue plastic water bottle. What follows is a possible request-to-result outcome that could result from such a web search: The Level 1 result – A "strong match" would be…a blue BPA-free plastic water bottle. The Level 2 result – A "close match" could be a red BPA-free plastic water bottle, AKA an attribute mismatch. The Level 3 result – A "weak or related match" might be a red ceramic water bottle. While BPA free it's not plastic, nor blue. And finally the Level 4 result – A completely "unrelated match": A commemorative Elvis velvet portrait; cute, but not what the shopper was looking for.

Considering that search queries are generated by people, it's not surprising that human judgement is an essential factor in evaluating the results, albeit with a consistent set of evaluation criteria. At a fundamental level, evaluating an acceptable result requires a yes or no answer to a simple threshold question: would a shopper reasonably expect to be happy with the search result? If the query, for example, cites a man's winter coat, the results can be open to a fair amount of interpretation by a data analyst. Again, it's a simple question. Would a reasonable shopper consider a three-quarter length leather jacket to be a winter coat? A well-constructed data analysis project, with project leaders who fully understand the potential for ambiguity, will often employ a two-stage, multiple-analyst structure; two analysts examine every result. If they agree, the result is categorized. If they don't, a third data analyst adjudicates.

By the time the algorithm has been commercially deployed (and many retailers will actually launch an algorithm as a work in progress), the data annotation has transitioned from providing training materials to identifying errors and "hard wiring" corrections that data scientists will use to refine the algorithm's training and perhaps even the model itself.



Source: iMerit

TANTALIZING POSSIBILITIES FOR ENHANCING E-COMMERCE

A cross-pollination approach is increasingly redefining opportunities for E-Commerce algorithm development – and in the process is placing a premium on the ability of data annotation to draw on similar cross sector skills and experiences. One telling example: in August of 2018, a little more than a year after agreeing to purchase the Whole Foods grocery chain, Amazon expanded the capabilities of its Alexa voice-recognition assistant to enable Whole Foods customers to order groceries through voice command. "Alexa, I need milk" became a new addition to the automated home lexicon.

In taking that one single step, the Alexa development team had to significantly extend the platform's Natural Language Processing capabilities to embrace a new destination for an existing use case: parsing strings of spoken words to determine intent, entity and channel. In many respects, the process is no different than having a user say, "Alexa, play Harry Potter and the Prisoner of Azkaban on Hulu," where "play" is the user intent, "Harry Potter and the Prisoner of Azkaban" is the entity, and "Hulu" is the channel. It's therefore quite easy to imagine how the data scientists at Amazon were able to teach Alexa, in short order, to understand, "Alexa, buy milk at Whole Foods."

Other home assistant platforms, such as Siri and Google Assistant will inevitably enable similar capabilities – and perhaps extend them within other E-Commerce platforms. It's perhaps a half step from there to imagine the E-Commerce platforms themselves enabling voice recognition as a site navigation and ordering tool. Given the massive number of data points – Walmart had a reported 70 million distinct SKUs on Walmart.com, and Amazon more than 350 million – developing the requisite voice recognition algorithms for such an endeavor will be a multi-year effort involving hundreds of thousands of data analyst hours. But the work will draw on years of expertise, on the part of both algorithm developers and data annotation providers, for those companies that have already established themselves in those adjacent sectors.

Particularly intriguing are the possibilities to enhance E-Commerce revenues through both voice and visual augmentation – creating a virtual store based on augmented reality engines. In one very realistic scenario, shoppers will be able to take a virtual walk through a supermarket, department store, general merchandiser, (or any other retailer that is capable of making the financial investment in new technology), fill their virtual shopping carts with products on the shelves – either by virtually grabbing them or speaking commands – and then heading to their own personal cashier.

Implementing that type of service will require Machine Learning algorithms that draw on expertise in Natural Language Processing, Computer Vision, and even somewhat related sectors, such as Autonomous Vehicle development and Geospatial data analysis. And it will place a premium on working with data annotation providers with cross-disciplinary expertise in each of those sectors.

DIGITAL FORTUNE TELLERS AND PREDICTIVE MODELING

In a similar vein, forward thinking data scientists (and leading-edge data annotation providers) are also looking at ways to expand the use of predictive modeling for E-Commerce, and even brick and mortar retail. Predictive modeling holds great promise both for the sales floor (physical and online) and back office operations that inevitably make the connection back to sales and marketing decisions through inventory management.

Just as airlines and hotels use predictive modeling to manage – and price – perishable inventory in the form of plane seats and hotel rooms, major retailers are looking at ways to enhance their inventory planning, and merchandising decisions, including product placement within a physical or virtual sales floor, through a better understanding of what consumers are likely to buy (and when) based on past behavior.

Since that latter specialty requires access to consumer data, there are privacy issues to address – but experts in the space point out that many predictive operations can work with anonymized data, so long as there's a unique (albeit anonymous) ID that can be used to tie purchases to a single individual.

At the heart of the process is the notion that past purchases can be used to predict future wants. If a consumer has just purchased Product A + Product B, and taking into account time of year, intervals, and other environmental factors, a logic equation should determine the likelihood of the same person shopping for Product C and/or Product D within a predetermined window. The process extends well beyond the intuitive conclusion that someone who just bought a new razer will eventually need more razor blades; the next-level question becomes, if someone just bought a razor, and then razor blades, what else are they about to buy?

Understanding those relationships generates hugely powerful – and lucrative – data that can be applied to everything from merchandising programs, seasonal inventory decisions, product placement, loyalty programs, and other programs that speak directly to the preeminent Siamese Twins of retail: cost of customer acquisition and retention on one side, and lifetime customer value on the other. Any tool that can reduce the former and increase the latter will find a warm reception in the executive suite of any retailer with the scale and resources to put it to work.

The data scientists developing those predictive algorithms require a massive amount of annotated training data in order to complete their work. The Machine Learning algorithms need to extend beyond standard pattern recognition to acquire an understanding of the causal, or at least correlational relationships between data points.

There's an argument to be made that establishing a pure causal relationship between a past purchase and future one is less important than confirming a correla tional pattern; so long as purchasing Product C consistently follows the purchases of Products A and B within an identifiable window, retailers can plan accordingly.

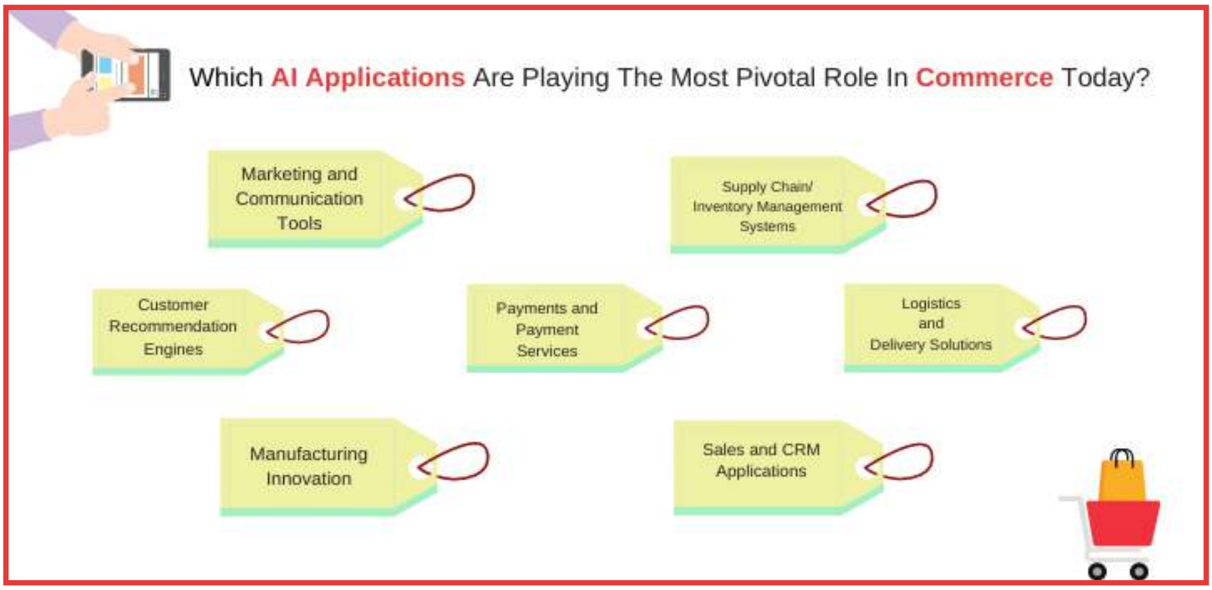
The possibilities for predictive merchandising become even more intriguing with data mining based on the personally identifiable information (PII) held by major retailers and E-Commerce sites. That opens the door to data analysis based on families rather than just individuals. Knowing who bought the cereal, after all, becomes a more critical predictive factor if it means that someone in the family is going to buy another box of cereal in the days to come.

THE HUMAN FACTOR IN A WORLD OF BINARY DECISIONS.

Virtually all of the Machine Learning algorithms deployed for retail are designed to interact with or otherwise benefit human shoppers. Unlike an algorithm for a drone or an autonomous vehicle, which is tasked primarily with navigating around, above, or alongside other inanimate objects, an E-Commerce algorithm, such as an E-Commerce search engine, by its very nature, is designed to interact with human shoppers. It needs to understand product taxonomies, product descriptions and categorization as a human would.

For that reason alone, human-in-the-loop data analysts have proven to be vital in the sector and, in fact, offer tremendous advantages where E-Commerce data annotation is concerned. Automated annotation programs can examine a scan of a festive t-shirt, for example, and recognize it as an article of clothing. But asking a computer program to understand seasonal and cultural subtleties without having a human guide to that conclusion is an exercise fraught with danger.

That's why the leaders in global retailing, and the technology labs that serve them, rely on human data analysts for leading-edge Machine Learning projects. And given the flexible borders between different Machine Learning specialties, for many projects it's an even greater advantage to bring cross-disciplinary experience and expertise.



Source: iMerit

CAT AND MOUSE BETWEEN HUMANS AND MACHINES

Algorithm development is in part the art and science of building the equivalent of human pattern recognition into a software program. That applies to everything from understanding the meaning – and intent – behind E-Commerce engine search queries to spotting suspicious user behavior. Particularly when the suspicious behavior is an attempt to game a system. With an aggregate of billions of dollars at stake every year, fraud detection has become a key feature for sales enabling platforms such as eBay, Amazon, Etsy and the like – where independent seller and buyer are supposed to play by the rules, but don't always behave.

The heart of the scam is a coded, sometimes overt but more often subtle, communication between buyer and seller to hijack the actual sale off the platform – and in the process deprive the platform operator its share of the transaction. A human observer would likely catch on quickly to a platform scam that could be as simple as a seller telling the buyer, "here's my phone number. Call me directly and I can give you a better deal." Even a rudimentary fraud detection algorithm, built through Natural Language Processing coding, would likely catch that one. But the digital wink, wink, nod, nod exchanges that characterize 2020 scammers are far better disguised, and require algorithms with a similar ability to carefully parse exchanges looking for communication worth investigating.

They're designed to flag trigger words and phrases pointing to suspicious exchanges involving phone numbers, email addresses, mailing addresses and the like, all based on context. There is, after all, a difference between a seller asking for a mailing address in order to ship product processed through the platform, and a buyer asking for a seller's address so they can meet to facilitate a side deal. Sometimes it's just the seller perpetrating the scam, by running duplicate product pages for the same product. The aim? Manipulate sellers through duplicate listings that in doing so violate the terms of the E-Commerce platform.

Teaching a machine to recognize the intent behind otherwise innocuous-seeming requests or listings requires, first, an ability to spot suspicious trigger words and phrases and then to correctly parse intent, entity, and context in a single sentence. And over time it's become something of a "whack a mole" exercise as fraudulent actors do their best to fool the fraud detection algorithm – even with something as basic as a phone number. One example: camouflaging phone numbers. "We were flagging any digits exchanged (between buyer and seller) remembers Arnab Chowdhury, iMerit Senior Manager Projects, who supervised the project. "Once sellers (looking to game the system) realized that the platform was monitoring for this they began to evolve." Step one: writing out the numerals as words. When the platform caught on, they resorted to writing the numbers as a single block (sevennineonethreeonetwonineninenineseven). When the algorithm caught on to that they moved on to numbers written as the upper-case keyboard symbol above each number.

It used to be easier when gaming the system always began with, "Psst, buddy – I got a deal for ya."

BUILDING THE TEAM – PROCESS AND TALENT

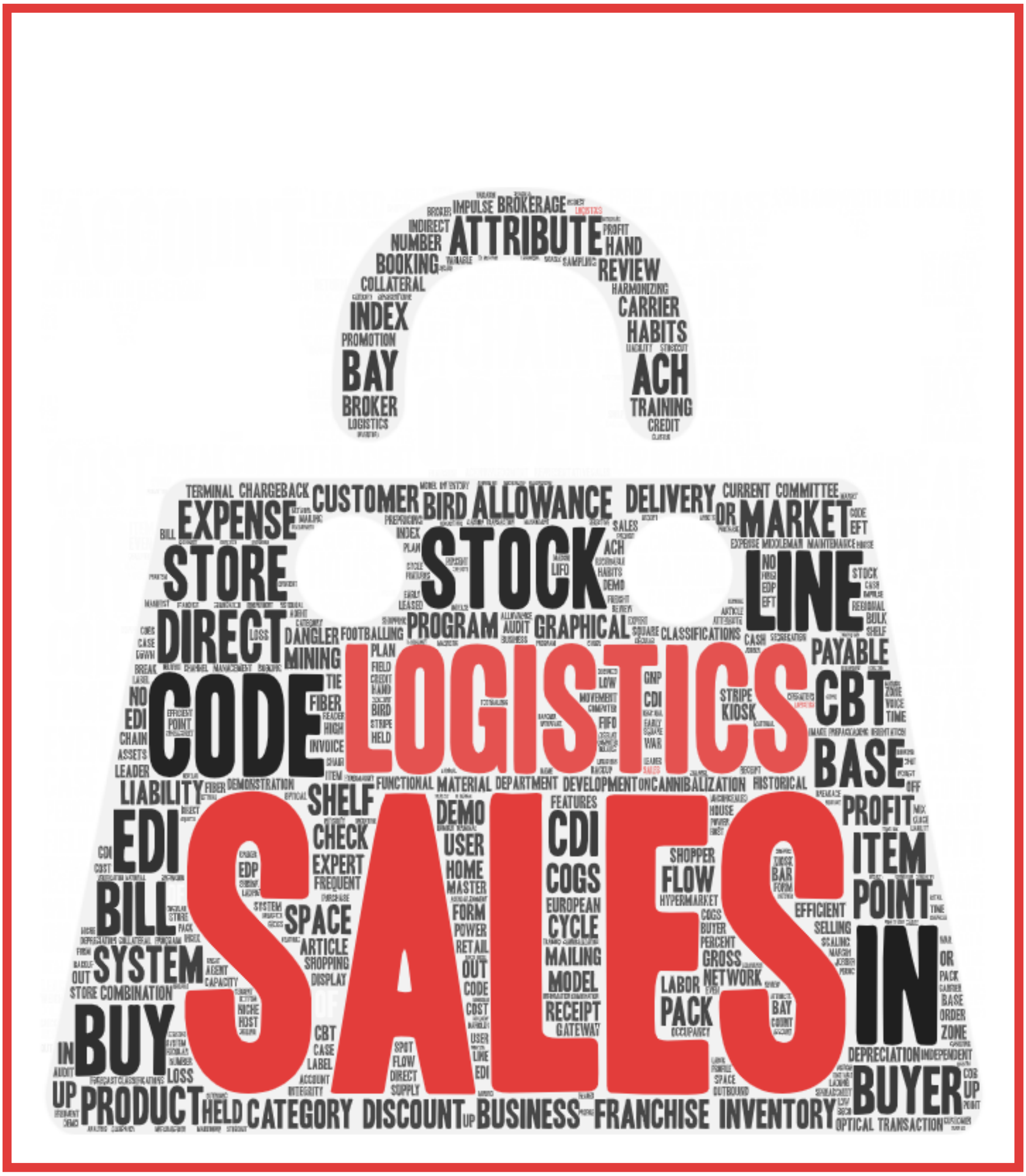
Often lost in crash projects to develop a commercially viable Machine Learning algorithm for E-Commerce is the human factor – the teams of developers, data analysts, and support staff who need to collaborate to build and train an effective logic engine. The best data analysts are smart, dogged, with a creative streak – part detective and part long-distance runner. Algorithms aside, E-Commerce is at its core a very human process, one that requires an ability to think like a shopper, to understand motivation and salesmanship, and to have above all else, empathy for the person sitting in front of a screen, searching for the perfect result. An effective retail data analyst also needs an eye for detail, the ability to spot and understand the often-subtle differences from one product to another based on multiple product attributes and descriptions.

Just as E-Commerce algorithms rely on training data and a ramp up in development to achieve a benchmark for success, the data analysts who help train them also require an initial learning and development preamble to their project work. It can take five years or longer to train some E-Commerce search engine algorithms before they hit their target accuracy, but the data annotation teams that develop the training data form and ramp up in a tiny fraction of that time. The nature and timing of that indoctrination depends on the complexity of the project, but it can typically require 2-4 weeks of classroom instruction and work floor training to get data analysts sufficiently up to speed.

In the case of iMerit, the company's Learning and Development team, solutions architect, and production delivery manager will sit down with a client for an initial kickoff and "train the trainer" orientation. The process includes a review of the data the team will analyze, a discussion of data relevance, accuracy metrics, and process, including the adapting client documents into usable training materials for the data analysts. In subsequent weeks the client and team will review initial results to confirm accuracy and quality metrics before proceeding to full production. If there is a snag, the Learning and Development team will investigate to determine if there's been human error on the part of the data analysts that requires additional training, a flaw in the training materials that needs to be corrected, or perhaps a miscommunication between client and team.

Often, the work will mix a combination of tasks over the life of the project – everything from data mining, to mapping data against taxonomies and even "hard wiring" correct search results into a database. The Learning and Development team will play a central role in, for example, helping to delineate the differences between edge case outliers and basic, common queries. All of that needs to be communicated to the team of data analysts assigned to the project, with an initial focus on building to an acceptable level of accuracy on data categorization and other aspects of the algorithm training materials. The first couple of weeks of actual work on the project represent something of a shakedown cruise managed through an initial QA process and feedback loop designed to both confirm a successful transition from the classroom to the work floor and identify any flaws in either the materials, or how they have been absorbed.

If the team of data analysts aren't generating the expected results the key question becomes: are the mistakes a function of human (data analyst) error, or something more systematic, such as a flaw in the training materials, or perhaps miscommunication requiring a reset between the client and the team? By the time the data analysts have finished this phase, usually three to four weeks in, they will be able to accurately analyze data points with an accuracy rate at or above the commonly accepted benchmark of 95 percent, and annotate data points as rapidly as one every 17-18 seconds for search relevance projects, and a minute or so each for data categorization data points – extraordinarily fast even if not quite the speed of what will eventually become a fully functioning algorithm.



Source: iMerit

THE COMING AI REVOLUTION IN RETAIL AND CONSUMER PRODUCTS: INTELLIGENT AUTOMATION IS TRANSFORMING BOTH INDUSTRIES IN UNEXPECTED WAYS

THINK BIG: TAKE A STRATEGIC APPROACH WHEN EVALUATING INTELLIGENT AUTOMATION

- Break down vertical silos and collaborate horizontally across organizational functions to assess and prioritize processes.
- Analyze every aspect of your enterprise, whether you're examining trends, consumption data and consumer feedback, or scaling that data down to the pulse of a neighborhood.
- Look for ways intelligent automation can improve brand experience throughout the customer journey and create competitive value for your brand in the long term.
- Assess both upstream and downstream processes across the integrated supply chain when evaluating intelligent automation capabilities.
- Choose your platform carefully to ensure it has measurable, scalable automation components.

START SMALL: STREAMLINE EXPANSION WITH AN AUTOMATION CENTER OF EXCELLENCE

- Establish an automation Center of Excellence that provides structure and governance to the development and use of information automation assets and processes - a key success factor.
- Assess where you are investing today, build on that investment and plan future execution. For example, if your goal is to improve inventory productivity and you are already using some components of intelligent automation for demand forecasting, consider extending this capability to inventory management.
- Continuously validate implementation priorities. Communicate regularly to fully educate executives and employees on capabilities and implications across the value chain. Because intelligent systems learn at a tremendous speed, you need to monitor them closely and train them properly to avoid biases or other potential issues.

WORK DIFFERENTLY: RETHINK THE WAY WORK GETS DONE IN THE DIGITAL AGE

- Envision the end results. Be open to uncovering new capabilities and iteratively evaluate automated tasks and activities for opportunities to redesign processes using intelligent automation capabilities.
- Partner with intelligent automation experts with design thinking expertise to guide and host regular work sessions with executives and employees, and work with customers to co-create the experience.
- Think beyond basic automation to capitalize on the potential of intelligent automation. This capability isn't just about removing human activities from business processes, but rather shifting to a culture of speed, agility and innovation.

Source: IBM Institute of Business Value, in association with the National Retail Federation

DATA SERVICES & ANNOTATION FOR COMMERCE

imerit.net