Prologue to *The Master Algorithm*

Pedro Domingos

You may not know it, but machine learning is all around you. When you type a query into a search engine, it's how the engine figures out which results to show you (and which ads, as well). When you read your email, you don't see most of the spam, because machine learning filtered it out. Go to Amazon.com to buy a book or Netflix to watch a video, and a machine learning system helpfully recommends some you might like. Facebook uses machine learning to decide which updates to show you, and Twitter does the same for tweets. Whenever you use a computer, chances are machine learning is involved somewhere.

Traditionally, the only way to get a computer to do something—from adding two numbers to flying an airplane—was to write down an algorithm explaining how, in painstaking detail. But machine learning algorithms, also known as learners, are different: they figure it out on their own, by making inferences from data. And the more data they have, the better they get. Now we don't have to program computers; they program themselves.

It's not just in cyberspace, either: your whole day, from the moment you wake up to the moment you fall asleep, is suffused with machine learning.

Your clock radio goes off at 7:00 a.m. It's playing a song you haven't heard before, but you really like it. Courtesy of Pandora, it's been learning your tastes in music, like your own personal radio jock. Perhaps the song itself was produced with the help of machine learning. You eat breakfast and read the morning paper. It came off the printing press a few hours earlier, the printing process carefully adjusted to avoid streaking using a learning algorithm. The temperature in your house is just right, and your electricity bill noticeably down, since you installed a Nest learning thermostat.

As you drive to work, your car continually adjusts fuel injection and exhaust recirculation to get the best gas mileage. You use Inrix, a traffic prediction system, to shorten your rush-hour commute, not to mention lowering your stress level. At work, machine learning helps you combat information overload. You use a data cube to summarize masses of data, look at it from every angle, and drill down on the most important bits. You have a decision to make: will layout A or B bring more business to your Web site? A Web learning system tries both out and reports back. You need to check out a potential supplier's Web site, but it's in a foreign language. No problem: Google automatically translates it for you. Your email conveniently sorts itself into folders, leaving only the most important messages in the inbox. Your word processor checks your grammar and spelling. You find a flight for an upcoming trip, but hold off on buying the ticket because Bing Travel predicts its price will go down soon. Without realizing it, you accomplish a lot more, hour by hour, than you would without the help of machine learning.

During a break you check on your mutual funds. Most of them use learning algorithms to help pick stocks, and one of them is completely run by a learning system. At lunchtime you walk down the street, smartphone in hand, looking for a place to eat. Yelp's learning system helps you find it. Your cell phone is chock full of learning algorithms. They're hard at work correcting your typos, understanding your spoken commands, reducing transmission errors, recognizing bar codes, and much else. Your phone can even anticipate what you're going to do next, and advise you accordingly. For example, as you're finishing lunch it discreetly alerts you that your afternoon meeting with an out-of-town visitor will have to start late, because her flight has been delayed.

Night has fallen by the time you get off work. Machine learning helps keep you safe as you walk to your car, monitoring the video feed from the surveillance camera in the parking lot and alerting off-site security staff if it detects suspicious activity. On your way home you stop at the supermarket, where you walk down aisles that were laid out with the help of learning algorithms: which goods to stock, which end-of-aisle displays to set up, whether to put the salsa in the sauce section or next to the tortilla chips. You pay with a credit card. A learning algorithm decided to send you the offer for that card, and approved your application. Another one continually looks for suspicious transactions, and alerts you if it thinks your card number was stolen. A third one tries to estimate how happy you are with this card. If you're a good customer but seem dissatisfied, you get a sweetened offer before you switch to another one.

You get home and walk to the mailbox. You have a letter from a friend, routed to you by a learning algorithm that can read handwritten addresses. There's also the usual junk, selected for you by other learning algorithms (oh, well). You stop for a moment to take in the cool night air. Crime in your city is noticeably down since the police started using statistical learning to predict where crimes are most likely to occur and concentrating beat officers there. You eat dinner with your family. The mayor is in the news. You voted for him because he personally called you on Election Day, after a learning algorithm pinpointed you as a key undecided voter. After dinner, you watch the ball game. Both teams selected their players with the help of statistical learning. Or perhaps you play games on your Xbox with your kids, and Kinect's learning algorithm figures out where you are and what you're doing. Before going to sleep, you take your medicine, which was designed and tested with the help of yet more learning algorithms. Your doctor, too, may have used machine learning to help diagnose you, from interpreting X-rays to figuring out an unusual set of symptoms.

Machine learning plays a part in every stage of your life. If you studied online for the SAT college admission exam, a learning algorithm graded your practice essays. And if you applied to business school and took the GMAT exam recently, one of your essay graders was a learning system. Perhaps when you applied for your job, a learning algorithm picked your resume from the virtual pile and told your prospective employer: here's a strong candidate; take a look. Your latest raise may have come courtesy of another learning algorithm. If you're looking to buy a house, Zillow.com will estimate what each one you're considering is worth. When you've settled on one, you apply for a home loan, and a learning algorithm studies your application and recommends accepting it (or not). Perhaps most important, if you've used an online dating service, machine learning may even have helped you find the love of your life.

Society is changing, one learning algorithm at a time. Machine learning is remaking science, technology, business, politics, and war. Satellites, DNA sequencers, and particle accelerators probe nature in ever finer detail, and learning algorithms turn the torrents of data into new scientific knowledge. Companies know their customers like never before. The

candidate with the best voter models wins, like Obama against Romney. Unmanned vehicles pilot themselves across land, sea, and air. No one programmed your tastes into the Amazon recommendation system; it figured them out on its own, by generalizing from your past purchases. Google's self-driving car taught itself how to stay on the road; no engineer wrote an algorithm instructing it, step by step, how to get from A to B. No one knows how to program a car to drive, and no one needs to, because a car equipped with a learning algorithm picks it up by observing what the driver does.

Machine learning is something new under the sun: a technology that builds itself. Ever since our remote ancestors started sharpening stones into tools, artifacts have been designed by humans only, whether they're then hand-built or mass-produced. But learning algorithms are artifacts that design other artifacts. "Computers are useless," said Picasso. "They can only give you answers." Computers aren't supposed to be creative; they're supposed to do what you tell them to. If what you tell them to do is be creative, you get machine learning. A learning algorithm is like a master craftsman: every one of its productions is different, and exquisitely matched to the customer's needs. But instead of turning stone into masonry or gold into jewelry, learners turn data into algorithms. And the more data they have, the more intricate the algorithms can be.

Humans are the species that adapts the world to itself instead of adapting itself to the world. Machine learning is the newest chapter in this million-year saga: with it, the world senses what you want and changes accordingly, without you having to lift a finger. Like a magic forest, your surroundings—virtual today, physical tomorrow—rearrange themselves as you move through them. The path you picked out between the trees and bushes grows into a road. Signs pointing the way spring up in the places you got lost.

These seemingly magical technologies work because, at its core, machine learning is about prediction: predicting what we want, the results of our actions, how to achieve our goals, how the world will change. Once upon a time we relied on shamans and soothsayers for this, but they were much too fallible. Science's predictions are more trustworthy, but they are limited to what we can systematically observe and tractably model. Big data and machine learning greatly expand that scope. Some everyday things can be predicted by the unaided mind, from catching a ball to carrying on a conversation. Some things, try as we might, are just unpredictable. For the vast middle ground between the two, there's machine learning.

Paradoxically, even as they open new windows on nature and human behavior, learning algorithms themselves have remained shrouded in mystery. Hardly a day goes by without a story in the media involving machine learning, whether it's Apple's launch of the Siri personal assistant, IBM's Watson beating the human *Jeopardy!* champion, Target finding out a teenager is pregnant before her parents do, or the N.S.A. looking for dots to connect. But in each case the learning algorithm driving the story is a black box. Even books on big data skirt around what really happens when the computer swallows all those terabytes and magically comes up with new insights. At best, we're left with the impression that learning algorithms just find correlations between pairs of events, such as googling "flu medicine" and having the flu. But finding correlations is to machine learning no more than bricks are to houses, and people don't live in bricks.

When a new technology is as pervasive and game-changing as machine learning, it's not wise to let it remain a black box. Opacity opens the door to error and misuse. Amazon's algorithm, more than any one person, determines what books are read in the world today. The N.S.A.'s algorithms decide whether you're a potential terrorist. Climate models decide what's a

safe level of carbon dioxide in the atmosphere. Stock-picking models drive the economy more than most of us do. You can't control what you don't understand, and that's why you need to understand machine learning—as a citizen, a professional, and a human being engaged in the pursuit of happiness.

This book's first goal is to let you in on the secrets of machine learning. Only engineers and mechanics need to know how a car's engine works, but every driver needs to know that turning the steering wheel changes the car's direction and stepping on the brake brings it to a stop. Few people today know what the corresponding elements of a learner even are, let alone how to use them. The psychologist Don Norman coined the term "conceptual model" to refer to the rough knowledge of a technology we need to have in order to use it effectively. This book provides you with a conceptual model of machine learning.

Not all learning algorithms work the same, and the differences have consequences. Take Amazon's and Netflix's recommenders, for example. If each were guiding you through a physical bookstore, trying to determine what's "right for you," Amazon would be more likely to walk you over to shelves you've frequented previously; Netflix would take you to unfamiliar and seemingly odd sections of the store but lead you to stuff you ended up loving. In this book we'll see the different kinds of algorithms that companies like Amazon and Netflix use. Netflix's algorithm has a deeper (even if still quite limited) understanding of your tastes than Amazon's, but ironically that doesn't mean Amazon would be better off using it. Netflix's business model depends on driving demand into the long tail of obscure movies and TV shows, which cost it little, and away from the blockbusters, which your subscription isn't enough to pay for. Amazon has no such problem; although it's well placed to take advantage of the long tail, it's equally happy to sell you more expensive popular items, which also simplify its logistics. And we, as customers, are more willing to take a chance on an odd item if we have a subscription than if we have to pay for it separately.

Hundreds of new learning algorithms are invented every year, but they're all based on the same few basic ideas. These are what this book is about, and they're all you really need to know to understand how machine learning is changing the world. Far from esoteric, and quite aside even from their use in computers, they are answers to questions that matter to all of us: How do we learn? Is there a better way? What can we predict? Can we trust what we've learned? Rival schools of thought within machine learning have very different answers to these questions. The main ones are five in number, and we'll devote a chapter to each. Symbolists view learning as the inverse of deduction, and take ideas from philosophy, psychology, and logic. Connectionists reverse engineer the brain, and are inspired by neuroscience and physics. Evolutionaries simulate evolution on the computer, and draw on genetics and evolutionary biology. Bayesians believe learning is a form of probabilistic inference, and have their roots in statistics. Analogizers learn by extrapolating from similarity judgments, and are influenced by psychology and mathematical optimization. Driven by the goal of building learning machines, we'll tour a good chunk of the intellectual history of the last hundred years, and see it in a new light.

Each of the five tribes of machine learning has its own "master algorithm," a general-purpose learner that you can in principle use to discover knowledge from data in any domain. The symbolists' master algorithm is inverse deduction, the connectionists' is backpropagation, the evolutionaries' is genetic programming, the Bayesians' is Bayesian inference, and the analogizers' is the support vector machine. In practice, however, each of these algorithms is good for some things but not others. What we really want is a single algorithm combining the key features of all of them: the Master Algorithm. For some this is an unattainable dream, but for

many of us in machine learning, it's what puts a twinkle in our eye and keeps us working late into the night.

If it exists, the Master Algorithm can derive all knowledge in the world—past, present and future—from data. Inventing it would be one of the greatest advances in the history of science. It would speed up the progress of knowledge across the board, and change the world in ways that we can barely begin to imagine. The Master Algorithm is to machine learning what the Standard Model is to particle physics or the Central Dogma to molecular biology: a unified theory that makes sense of everything we know to date, and lays the foundation for decades or centuries of future progress. The Master Algorithm is our gateway to solving some of the hardest problems we face, from building domestic robots to curing cancer.

Take cancer. Curing it is hard because cancer is not one disease, but many. Tumors can be triggered by a dizzying array of causes, and they mutate as they metastasize. The surest way to kill a tumor is to sequence its genome, figure out which drugs will work against it—without harming you, given *your* genome and medical history—and perhaps even design a new drug specifically for your case. No doctor can master all the knowledge required for this. Sounds like a perfect job for machine learning: in effect, it's a more complicated and challenging version of the searches that Amazon and Netflix do every day, except it's looking for the right treatment for you instead of the right book or movie. Unfortunately, while today's learning algorithms can diagnose many diseases with superhuman accuracy, curing cancer is well beyond their ken. If we succeed in our quest for the Master Algorithm, it will no longer be.

The second goal of this book is thus to enable *you* to invent the Master Algorithm. You'd think this would require heavy-duty mathematics and severe theoretical work. On the contrary, what it requires is stepping back from the mathematical arcana to see the overarching pattern of learning phenomena; and for this the layman, approaching the forest from a distance, is in some ways better placed than the specialist, already deeply immersed in the study of particular trees. Once we have the conceptual solution, we can fill in the mathematical details; but that is not for this book, and not the most important part. Thus, as we visit each tribe, our goal is to gather its piece of the puzzle and understand where it fits, mindful that none of the blind men can see the whole elephant. In particular, we'll see what each tribe can contribute to curing cancer, and also what it's missing. Then, step by step, we'll assemble all the pieces into the solution—or rather, *a* solution that is not yet the Master Algorithm, but is the closest anyone has come, and hopefully makes a good launch pad for your imagination. And we'll preview the use of this algorithm as a weapon in the fight against cancer. As you read the book, feel free to skim or skip any parts you find troublesome; it's the big picture that matters, and you'll probably get more out of those parts if you revisit them after the puzzle is assembled.

I've been a machine learning researcher for more than twenty years. My interest in it was sparked by a book with an odd title I saw in a bookstore when I was a senior in college: *Artificial Intelligence*. It had only a short chapter on machine learning, but on reading it I immediately became convinced that learning was the key to solving A.I., and that the state of the art was so primitive that maybe I could contribute something. Shelving plans for an M.B.A., I entered the Ph.D. program at the University of California, Irvine. Machine learning was then a small, obscure field, and U.C.I. had one of the few sizable research groups anywhere. Some of my classmates dropped out because they didn't see much of a future in it, but I persisted. To me nothing could have more impact than teaching computers to learn: if we could do that, we would get a leg up on every other problem. By the time I graduated five years later, the data mining explosion was underway, and so was my path to this book. My doctoral dissertation unified

symbolic and analogical learning. I've spent much of the last ten years unifying symbolism and Bayesianism, and more recently those two with connectionism. It's time to go the next step and attempt a synthesis of all five paradigms.

I had a number of different but overlapping audiences in mind when writing this book.

If you're curious what all the hubbub surrounding big data and machine learning is about, and suspect that there's something deeper going on than what you see in the papers, you're right! This book is your guide to the revolution.

If your main interest is in the business uses of machine learning, this book can help you in at least six ways: to become a savvier consumer of analytics; to make the most of your data scientists; to avoid the pitfalls that kill so many data mining projects; to discover what you can automate without the expense of hand-coded software; to reduce the rigidity of your information systems; and to anticipate some of the new technology that's coming your way. I've seen too much time and money wasted trying to solve a problem with the wrong learning algorithm, or misinterpreting what the algorithm said. It doesn't take much to avoid these fiascoes. In fact, all it takes is to read this book.

If you're a citizen or policymaker concerned with the social and political issues raised by big data and machine learning, this book will give you a primer on the technology—what it is, where it's taking us, what it does and doesn't make possible—without boring you with all the ins and outs. From privacy to the future of work and the ethics of roboticized warfare, we'll see where the real issues are, and how to think about them.

If you're a scientist or engineer, machine learning is a powerful armory that you don't want to be without. The old, tried-and-true statistical tools don't get you far in the age of big (or even medium) data. You need machine learning's nonlinear chops to accurately model most phenomena, and it brings with it a new scientific worldview. The expression "paradigm shift" is used too casually these days, but I believe it's not an exaggeration to say that that's what this book describes.

If you're a machine learning expert, you're already familiar with much of what the book covers, but you'll also find in it many fresh ideas, historical nuggets, and useful examples and analogies. Most of all, I hope the book will provide a new perspective on machine learning, and maybe even start you thinking in new directions. Low-hanging fruit is all around us, and it behooves us to pick it, but we also shouldn't lose sight of the bigger rewards that lie just beyond. (Apropos of which, I hope you'll forgive my poetic license in using the term "Master Algorithm" to refer to a general-purpose learner.)

If you're a student of any age—high-schooler wondering what to major in in college, undergraduate deciding whether to go into research, seasoned professional considering a career change—my hope is that this book will spark in you an interest in this fascinating field. The world has a dire shortage of machine-learning experts, and if you decide to join us you can look forward to not only exciting times and material rewards, but also a unique opportunity to serve society. And if you're already studying machine learning, I hope the book will help you get the lay of the land; if in your travels you chance upon the Master Algorithm, that alone makes it worth writing.

Last but not least, if you have an appetite for wonder, machine learning is an intellectual feast, and you're invited—R.S.V.P.!