Machine Learning for Dummies

Table of Content

Introduction

Machine learning is having a dramatic impact on the way software is designed so that it can keep pace with business change. Machine learning is so dramatic because it helps you use data to drive business rules and logic. How is this different? With traditional software development models, programmers wrote logic based on the current state of the business and then added relevant data. However, business change has become the norm. It is virtually impossible to anticipate what changes will transform a market. The value of machine learning is that it allows you to continually learn from data and predict the future. This powerful set of algorithms and models are being used across industries to improve processes and gain insights into patterns and anomalies within data. But machine learning isn't a solitary endeavor; it's a team process that requires data scientists, data engineers, business analysts, and business leaders to collaborate. The power of machine learning requires a collaboration so the focus is on solving business problems

About This Book

Machine Learning For Dummies, IBM Limited Edition, gives you insights into what machine learning is all about and how it can impact the way you can weaponize data to gain unimaginable insights. Your data is only as good as what you do with it and how you manage it. In this book, you discover types of machine learning techniques, models, and algorithms that can help achieve results for your company. This information helps both business and technical leaders learn how to apply machine learning to anticipate and predict the future.

Foolish Assumptions

The information in this book is useful to many people, but we have to admit that we did make a few assumptions about who we think you are:

- You're already familiar with how machine learning algorithms are being used within your organization to create new software. You need to be prepared to lead your team in the right direction so that the company gains maximum value from the use of these powerful algorithms and models.
- You're planning a long-term strategy to create software that can stand the test of time. Management wants to be able to leverage all the important data about customers, employees, prospects, and business trends. Your goal is to be prepared for the future.
- You understand the huge potential value of the data that exists throughout your organization.
- You understand the benefits of machine learning and its impact on the company, and you want to make sure that your team is ready to apply this power to remain competitive as new business models emerge.
- You're a business leader who wants to apply the most important emerging technologies to be as creative and innovative as possible.

Chapter 1

Understanding Machine Learning

Machine learning, artificial intelligence (AI), and cognitive computing are dominating conversations about how emerging advanced analytics can provide businesses with a competitive advantage to the business. There is no debate that existing business leaders are facing new and unanticipated competitors. These businesses are looking at new strategies that can prepare them for the future. While a business can try different strategies, they all come back to a fundamental truth — you have to follow the data. In this chapter, we delve into what the value of machine learning can be to your business strategy. How should you think about machine learning? What can you offer the business based on advanced analytics technique that can be a game-changer?

What Is Machine Learning?

Machine learning has become one of the most important topics within development organizations that are looking for innovative ways to leverage data assets to help the business gain a new level of understanding. Why add machine learning into the mix? With the appropriate machine learning models, organizations have the ability to continually predict changes in the business so that they are best able to predict what's next. As data is constantly added, the machine learning models ensure that the solution is constantly updated. The value is straightforward: If you use the most appropriate and constantly changing data sources in the context of machine learning, you have the opportunity to predict the future. Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming. However, machine learning is not a simple process. Machine learning uses a variety of algorithms that iteratively learn from data to improve, describe data, and predict outcomes. As the algorithms ingest training data, it is then possible to produce more precise models based on that data. A machine learning model is the output generated when you train your machine learning algorithm with data. After training, when you provide a model with an input, you will be given an output. For example, a predictive algorithm will create a predictive model. Then, when you provide the predictive model with data, you will receive a prediction based on the data that trained the model. Machine learning is now essential for creating analytics models. You likely interact with machine learning applications without realizing. For example, when you visit an e-commerce site and start viewing products and reading reviews, you're likely presented with other, similar products that you may find interesting. These recommendations aren't hard coded by an army of developers. The suggestions are served to the site via a machine learning model. The model ingests your browsing history along with other shoppers' browsing and purchasing data in order to present other similar products that you may want to purchase.

Iterative learning from data

Machine learning enables models to train on data sets before being deployed. Some machine learning models are online and continuously adapt as new data is ingested. On the other hand, other models, called offline machine learning models, are derived from machine learning algorithms but, once deployed, do not change. This iterative process of online models leads to an improvement in the types of associations that are made between data elements. Due to their complexity and size, these patterns and associations could have easily been overlooked by human observation. After a model has been trained, these models can be used in real time to learn from data. In addition, complex algorithms can be automatically adjusted based on rapid changes in variables, such as sensor data, time, weather data, and customer sentiment metrics. For example, inferences can be made from a machine learning model — if the weather changes quickly, a weather predicting model can predict a tornado, and a warning siren can be triggered. The improvements in accuracy are a result of the training process and automation that is part of machine learning. Online machine learning algorithms continuously refine the models by continuously processing new data in near real time and training the system to adapt to changing patterns and associations in the data.

What's old is new again

AI and machine learning algorithms aren't new. The field of AI dates back to the 1950s. Arthur Lee Samuels, an IBM researcher, developed one of the earliest machine learning programs — a self-learning program for playing checkers. In fact, he coined the term machine learning. His approach to machine learning was explained in a paper published in the IBM Journal of Research and Development in 1959. Over the decades, AI techniques have been widely used as a method of improving the performance of underlying code. In the last few years with the focus on distributed computing models and cheaper compute and storage, there has been a surge of interest in AI and machine learning that has lead to a huge amount of money being invested in startup software companies. Today, we 6 Machine Learning For Dummies, IBM Limited Edition These materials are © 2018 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited. are seeing major advancements and commercial solutions. Why has the market become real? There are six key enablers:

- Modern processors have become increasingly powerful and increasingly dense. The density to performance ratio has improved dramatically.
- The cost of storing and managing large amounts of data has been dramatically lowered. In addition, new storage innovations have led to faster performance and the ability to analyze vastly larger data sets.
- The ability to distribute compute processing across clusters of computers has dramatically improved the ability to analyze complex data in record time.
- There are more commercial data sets available to support analytics, including weather data, social media data, and medical data sets. Many of these are available as cloud services and well-defined Application Programming Interfaces (APIs).

- Machine learning algorithms have been made available through open-source communities with large user bases. Therefore, there are more resources, frameworks, and libraries that have made development easier.
- Visualization has gotten more consumable. You don't need to be a data scientist to interpret results, making use of machine learning broader within many industries

Defining Big Data

Big data is any kind of data source that has at least one of four shared characteristics, called the four Vs:
- Extremely large Volumes of data
- The ability to move that data at a high Velocity of speed
- An ever-expanding Variety of data sources
- Veracity so that data sources truly represent truth

The accuracy of a machine learning model can increase substantially if it's trained on big data. Without enough data, you are CHAPTER 1 Understanding Machine Learning 7 These materials are © 2018 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited. trying to make decisions on small subsets of your data that might lead to misinterpreting a trend or missing a pattern that is just starting to emerge. While big data can be very useful for training machine learning models, organizations can use machine learning with just a few thousand data points. Don't underestimate the task at hand. Data must be able to be verified based on both accuracy and context. An innovative business in a fast-changing market will want to deploy a model that can make inferences in milliseconds to quickly assess the best offer for an at-risk customer to keep her happy. It is necessary to identify the right amount and types of data that can be analyzed to impact business outcomes. Big data incorporates all data, including structured, unstructured, and semi-structured data from email, social media, text streams, images, and machine sensors. Traditional Business Intelligence (BI) products weren't really designed to handle the complexities of constantly changing data sources. BI tools are typically designed to work with highly structured, well-understood data, often stored in a relational data repository. These traditional BI tools typically only analyze snapshots of data rather than the entire data set. Analytics on big data requires technology designed to gather, store, manage, and manipulate vast amounts data at the right speed and at the right time to gain the right insights. With the evolution of computing technology and the emergence of hybrid cloud architectures, it's now possible to manage immense volumes of data that previously could have only been handled by supercomputers at great expense

Big Data in Context with Machine Learning

Machine learning requires the right set of data that can be applied to a learning process. An organization does not have to have big data in order to use machine learning techniques; however, big data can help improve the accuracy of machine learning models. With big data, it is now possible to virtualize data so it can be stored in the most efficient and cost-effective manner whether on premises or in the cloud. In addition, improvements in network speed and reliability have removed other physical limitations of 8

being able to manage massive amounts of data at the acceptable speed. Add to this the impact of changes in the price and sophistication of computer memory, and with all these technology transitions, it's now possible to imagine how companies can leverage data in ways that would've been inconceivable only five years ago. No technology transition happens in isolation; change happens when there is an unsolved business problem combined with the maturation of technology. There are countless examples of important technologies that have matured enough to support the renaissance of machine learning. These maturing big data technologies include data virtualization, parallel processing, distributed file systems, in-memory databases, containerization, and micro-services. This combination of technology advances can help organizations address significant business problems. Businesses have never lacked large amounts of data. Leaders have been frustrated for decades about their inability to use the richness of data sources to gain actionable insights from their data. Armed with big data technologies and machine learning models, organizations are able to anticipate the future and be better prepared for disruption.

The Need to Understand and Trust your Data

It is not enough to simply ingest vast amounts of data. Providing accurate machine learning models requires that the source data be accurate and meaningful. In addition, these data sources are meaningful when combined with each other so that the model is accurate and trusted. You have to understand the origin of your data sources and whether they make sense when they're combined. In addition to trusting your data, it also important to perform data cleansing or tidying. Cleaning data means that you transform your data into a form that can be understood by a machine learning algorithm. For example, algorithms use numbers, but data is often in the form of words. You have to turn those words into numbers. In addition, you have to make sure those numbers are CHAPTER 1 Understanding Machine Learning 9 sensibly derived and internally consistent. You need to decide how you handle missing data and other data irregularities. Data refinement provides the foundation for building analytical models that deliver results you can trust. The process of data refinement will help to ensure that your data is timely, clean, and well understood.