

Chapter 1. Introduction: Data- Analytic Thinking

The Ubiquity of Data Opportunities

- With vast amounts of data now available, companies in almost every industry are focused on exploiting data for competitive advantage.
- Widest applications of data-mining techniques are marketing for tasks such as targeted marketing, online advertising, and recommendations for cross-selling. Data mining is used for general customer relationship management to analyze customer behavior in order to manage attrition and maximize expected customer value. The finance industry uses data mining for credit scoring and trading, and operations via fraud detection and workforce management.
- There is a fundamental structure to data-analytic thinking, and basic principles that should be understood. There are also particular areas where intuition, creativity, common sense, and domain knowledge must be brought to bear.
- Data science = data mining
- At a high level, data science is a set of fundamental principles that guide the extraction of knowledge from data. Data mining is the extraction of knowledge from data, via technologies that incorporate these principles.

note

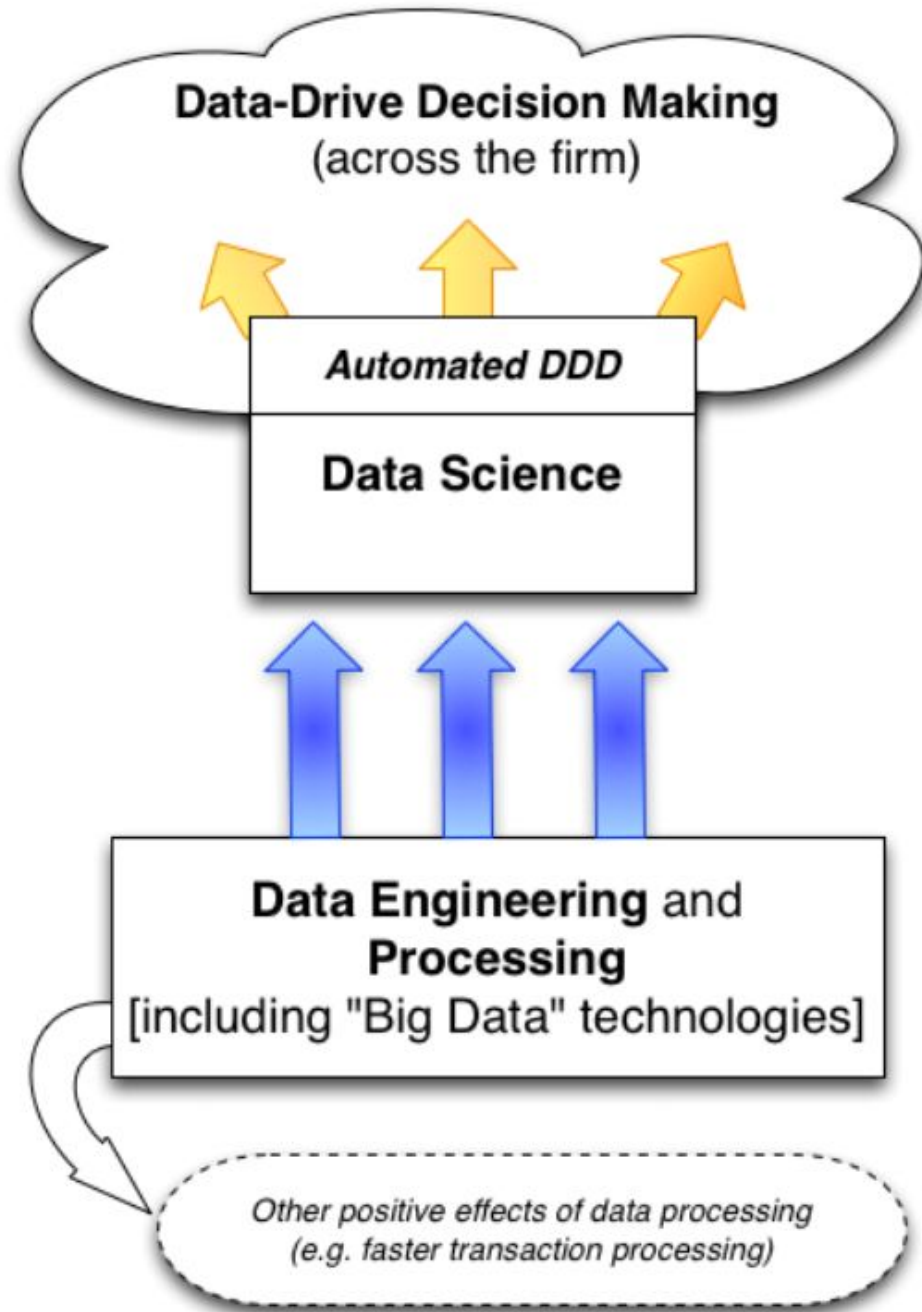
- Important to understand data science.
- Data-analytic thinking enables you to evaluate proposals for data mining projects.
- You should be able to assess the proposal systematically and decide whether it is sound or flawed.
- You should be able to spot obvious flaws, unrealistic assumptions, and missing pieces.

Example: Hurricane Frances

- Wal-Mart Stores decided that the situation offered a great opportunity for one of their newest data-driven weapons...predictive technology.
- Linda M. Dillman, Wal-Mart's chief information officer, pressed her staff to come up with forecasts based on what had happened when Hurricane Charley struck several weeks earlier.
- she felt that the company could 'start predicting what's going to happen, instead of waiting for it to happen.
- She can identify unusual local demands for products.
- They can understand which foods are more popular before and during a hurricane.

Example: Predicting Customer Churn

- MegaTelCo. They are having a major problem with customer retention in their wireless business.
- Your task is to devise a precise, step-by-step plan for how the data science team should use MegaTelCo's vast data resources to decide which customers should be offered the special retention deal prior to the expiration of their contracts.
- Think carefully about what data you might use and how they would be used.
- Specifically, how should MegaTelCo choose a set of customers to receive their offer in order to best reduce churn for a particular incentive budget?



Science, Engineering, and Data-Driven Decision Making

- Data science involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data.
- Data-driven decision-making (DDD) refers to the practice of basing decisions on the analysis of data, rather than purely on intuition. For example, a marketer could select advertisements based purely on her long experience in the field and her eye for what will work. Or, she could base her selection on the analysis of data regarding.
- DDD is not an all-or-nothing practice.

The benefits of data-driven decision-making

- Economist Erik Brynjolfsson and his colleagues from MIT and Penn's Wharton School developed a measure of DDD that rates firms as to how strongly they use data to make decisions across the company.
- DDD also is correlated with higher return on assets, return on equity, asset utilization, and market value, and the relationship seems to be causal.
- The sort of decisions.
- (1) decisions for which "discoveries" need to be made within data, and (2) decisions that repeat, especially at massive scale, and so decision-making can benefit from even small increases in decision-making accuracy based on data analysis.

2012 Target

- Target cares about consumers' shopping habits, what drives them, and what can influence them.
- But, consumers tend to have inertia in their habits and getting them to change is very difficult.
- Target knew, however, that the arrival of a new baby in a family is one point where people do change their shopping habits significantly "As soon as we get them buying diapers from us, they're going to start buying everything else too."
- Since most birth records are public, retailers obtain information on births and send out special offers to the new parents.
- They were interested in whether they could predict that people are expecting a baby. Target analyzed historical data on customers who later were revealed to have been pregnant .
- For example, pregnant mothers often change their diets, their pregnant wardrobes, their vitamin regimens, and so on.
- Importantly, in both the Walmart and the Target examples, the data analysis was not testing a simple hypothesis. Instead, the data were explored with the hope that something useful would be discovered.

2DDD problem

- MegaTelCo has hundreds of millions of customers, each a candidate for defection.
- If we can improve our ability to estimate, for a given customer, how profitable it would be for us to focus on her, we can potentially reap large benefits by applying this ability to the millions of customers in the population.
- Increasingly, business decisions are being made automatically by computer systems- automatic decision-making .
- The finance and telecommunications industries were early adopters, they allowed the aggregation and modeling of data at a large scale, as well as the application of the resultant models to decision –making.
- In the 1990s, automated decision-making changed the banking and consumer credit industries dramatically.

Data Processing and “Big Data”

- There is a lot of data processing that is not data .
- Many data processing skills, systems, and technologies often are mistakenly cast as data science.
- "Big data" technologies. Big data essentially means datasets that are too large for traditional data processing systems, and therefore require new processing technologies. Big data technologies are used for many tasks, including data engineering. Big data technologies are actually used for implementing data mining techniques. Big data is used for data processing in support of the data mining techniques and other data science activities.
- Prasanna Tambe examined the extent to which big data technologies seem to help firms . He finds that, after controlling for various possible confounding factors, using big data technologies is associated with significant additional productivity growth.

From Big Data 1.0 to Big Data 2.0

- In Web 1.0, businesses busied themselves with getting the basic internet technologies in place, so that they could establish a web presence, build electronic commerce capability, and improve the efficiency of their operations.
- Web 2.0, where new systems and companies began taking advantage of the interactive nature of the Web.
- We should expect a Big Data 2.0 phase to follow Big Data 1.0. Once firms have become capable of processing massive data in a flexible fashion, they should begin asking: "What can I now do that I couldn't do before, or do better than I could do before?"
- Example Amazon: incorporating the consumer's "voice" early on, in the rating of products, in product reviews (and deeper, in the rating of product reviews).

Data and Data Science Capability as a Strategic Asset

- Data, and the capability to extract useful knowledge from data, should be regarded as key strategic assets.
- Previously, in the 1980s, data science had transformed the business betrortee of consumer credit. Modeling the probability of default had changed the industry from personal assessment of the likelihood of default to strategies of massive scale and market share, which brought along concomitant economies of scale.
- (Richard Fairbanks and Nigel Morris) realized that information technology was powerful enough that they could do more sophisticated predictive modeling using.
- Signet Bank's management was convinced that modeling profitability, not just default probability, was the right strategy, but they did not have appropriate data.

What could Signet bank do?

- They brought into play a fundamental strategy of data science; acquire the necessary data at the cost.
- They should think about whether and how much they are willing to invest. The data-analytic thinker needs to consider whether she expects the data to have sufficient value to justify the investment.
- Losses continued for a few years. Because the firm viewed these losses as investments in data, they persisted despite complaints from stakeholders. Eventually, Signet's credit card operation turned around and became so profitable.
- They proceeded to apply data science principles throughout the not just customer acquisition but retention as well.

Martens and Provost 2011

- The bank built models from data to decide whom to target with offers for different products.
- Detailed data on customers' individual (anonymized) transactions improve performance substantially over just using.
- Banks with bigger data assets may have 'an important strategic advantage over their smaller competitors
- The net result will be either increased adoption of the bank's products, decreased cost of customer acquisition, or both.
- The idea of data as a strategic asset is certainly not limited to Capital One, nor even to the banking industry.
- Amazon was able to gather data early on online customers: consumers find value in the rankings and recommendations that Amazon provides. Amazon therefore can retain customers easily, and can even charge a premium.

Data-Analytic Thinking

- Understanding the fundamental concepts, and having frameworks for organizing data-analytic thinking not only will allow one to interact competently, but will help to envision opportunities for improving data-driven decision-making, or to see data-oriented competitive threats.
- For examples, if a consultant presents a proposal to mine a data asset to improve your business, you should be able to assess whether the proposal makes sense. If a competitor announces a new data partnership, you should recognize when it may put you at a strategic disadvantage.
- Is this reasonable? With an understanding of the fundamentals of data science you should be able to devise a few probing questions to determine whether their valuation arguments are plausible.

Data Mining and Data Science, Revisited

- Fundamental concept: Extracting useful knowledge from data to solve business problems can be treated systematically by following a process with reasonably well-defined stages. The Cross Industry Standard Process for Data Mining, abbreviated CRISP-DM (CRISP-DM Project, 2000).
- Fundamental concept: From a large mass of data, information technology can be used to find informative descriptive attributes of entities interest.
- Alternatively, the analyst could apply information technology to automatically discover informative attributes-essentially doing large-scale automated experimentation.
- Fundamental concept: Formulating data mining solutions and evaluating the results involves thinking carefully about the context in which they be used.

Summary

- Data should be thought of as a business asset, and once we are thinking in this direction we start to ask whether (and how much) we should invest in data.
- There is convincing evidence that data-driven decision-making and big data technologies substantially improve business performance.