READING MATERIALS – EDUCATIONAL PURPOSES ONLY

BIG DATA ANALYTICS

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION TO BIG DATA ANALYTICS

1.1 AN INTRODUCTION TO BIG DATA ANALYTICS [1]

Many of you would have probably heard about Big data Analytics. Have you ever wondered what it is all about and how it can help us? Big data analytics can be defined as a process of examining large and varied data sets. We use advanced analytics techniques against the large data to uncover the hidden patterns, unknown correlations, market trends, customer preferences and other useful information. This helps the organizations to make informed decisions.

To understand Big Data Analytics you have to first understand What Analytics is?

1.1.1 What is Analytics? [1, 3]

Analytics is an encompassing and multidimensional field. It uses mathematics, statistics, predictive modeling and machine-learning techniques to find meaningful patterns and knowledge in recorded data.

Analytics is a term that is often used interchangeably with data science, data mining, knowledge discovery, and others. The distinction between all those is not clear cut. All of these terms essentially refer to extracting useful business patterns or mathematical decision models from a preprocessed data set. Different underlying techniques can be used for this purpose, stemming from a variety of different disciplines, such as:

- Statistics (e.g., linear and logistic regression)
- Machine learning (e.g., decision trees)
- Biology (e.g., neural networks, genetic algorithms, swarm intelligence)
- Kernel methods (e.g., support vector machines)

Basically, a distinction can be made between predictive and descriptive analytics. In predictive analytics, a target variable is typically available, which can either be categorical (e.g., churn or not, fraud or not) or continuous (e.g., customer lifetime value, loss given default). In descriptive analytics, no such target variable is available. Common examples here are association rules, sequence rules, and clustering. Figure 1.1 provides an example of a decision tree in a classification predictive analytics setting for predicting churn.

More than ever before, analytical models steer the strategic risk decisions of companies. For example, in a bank setting, the minimum equity and provisions a financial institution

holds are directly determined by, among other things, credit risk analytics, market risk analytics, operational risk analytics, fraud analytics, and insurance risk analytics. In this setting, analytical model errors directly affect profitability, solvency, shareholder value, the macroeconomic, and society as a whole. Hence, it is of the utmost importance that analytical models are developed in the most optimal way, taking into account various requirements that will be discussed in what follows.

Customer	Age	Recency	Frequency	Monetary	Churn
John	35	5	6	100	Yes
Sophie	18	10	2	150	No
Victor	38	28	8	20	No
Laura	44	12	4	280	Yes

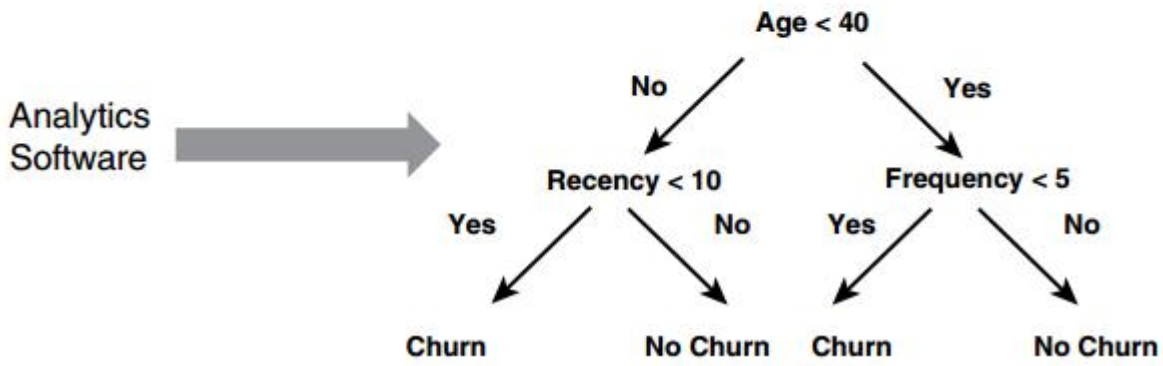

Figure 1.1 Example of Classification Predictive Analytics [3]

1.1.2 What is Big Data Analytics? [1]

As said before Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. With today's technology, it's possible to analyze your data and get answers from it immediately. Big Data Analytics helps you to understand your organization better. With the use of Big data analytics, one can make the informed decisions without blindly relying on guesses.

And it can help answer the following types of questions:

What actually happened?

How or why did it happen?
What's happening now?
What is likely to happen next?

1.1.3 History and Evolution of Big Data Analytics [1]

The concept of big data has been around for years; most organizations now understand that if they capture all the data that streams into their businesses, they can apply analytics and get significant value from it. But even in the 1950s, decades before anyone uttered the term "big data," businesses were using basic analytics essentially numbers in a spreadsheet that were manually examined to uncover insights and trends.

The new benefits that big data analytics brings to the table, however, are speed and efficiency. Whereas a few years ago a business would have gathered information, run analytics and unearthed information that could be used for future decisions, today that business can identify insights for immediate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn't have before.

1.1.4 Values of Big Data Analytics [1]

Figure 1.2 Big Data Analytics

Big data analytics helps organizations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations, higher profits and happier customers. Here are the most important values of Big Data,

1. *Cost reduction*: Big data technologies such as Hadoop and cloud-based analytics bring significant cost advantages when it comes to storing large amounts of data – plus they can identify more efficient ways of doing business.

2. *Faster, better decision making*: With the speed of Hadoop and in-memory analytics, combined with the ability to analyze new sources of data, businesses are able to analyze information immediately – and make decisions based on what they've learned.
3. *New products and services*: With the ability to gauge customer needs and satisfaction through analytics comes the power to give customers what they want. Davenport points out that with big data analytics, more companies are creating new products to meet customers' needs.

1.1.5 Uses of Big Data analytics across different industries[1]

Banking
Large amounts of information will be streaming in into banks, managing all this data and getting proper insights would be possible only with big data analytics. This is important to understand customers and boost their satisfaction, and also to minimize risk and fraud.

Government
When government agencies are able to harness and apply analytics to their big data, they gain significant ground when it comes to managing utilities, running agencies, dealing with traffic congestion or preventing crime.

Health Care
Patient records, Treatment plans, Prescription information. When it comes to health care, everything needs to be done quickly, accurately – and, in some cases, with enough transparency to satisfy stringent industry regulations. When big data is managed effectively, health care providers can uncover hidden insights that improve patient care.

Education
Educators armed with data-driven insight can make a significant impact on school systems, students, and curriculums. By analyzing big data, they can identify at-risk students, make sure students are making adequate progress, and can implement a better system for evaluation and support of teachers and principals.

Manufacturing
Armed with insight that big data can provide, manufacturers can boost quality and output while minimizing waste – processes that are key in today's highly competitive market. More and more manufacturers are working in an analytics-based culture, which means they can solve problems faster and make more agile business decisions.