**Data Science for Business Chapter 2 Journal**

Chase Rawson

College of Professional Studies, University of New Hampshire

DAT 620: Data Analytics in Business Intelligence

Dr. Tony Sulpizio

September 14, 2023

**[From Business Problems to Data Mining Tasks](https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch02.html" \l "sect_data_mining_tasks)**

* Each data-driven business decision-making problem is unique, shaped by specific goals, constraints, and contexts. These uniqueness factors extend to the personalities and preferences involved in the decision-making process.
* Despite numerous data mining algorithms, they fundamentally address a few distinct types of tasks. It is essential to clearly define and understand these task types before embarking on a data mining project.

**[Supervised Versus Unsupervised Methods](https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch02.html" \l "twodot2_supervised_versus_unsupervised_m)**

* Tasks like classification, regression, and causal modeling are typically addressed using supervised methods. Some tasks, like similarity matching, link prediction, and data reduction, can be either supervised or unsupervised.
* Unsupervised methods, including clustering, co-occurrence grouping, and profiling, are generally used for tasks where labeled data is unavailable. Fundamental data mining principles underlie all these techniques.
* Deciding whether to approach a problem with supervised or unsupervised methods is a vital early decision. In supervised tasks, precise definitions of target variables are essential.

**[Data Mining and Its Results](https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch02.html" \l "twodot3_data_mining_and_its_results)**

* Data mining results should inform the process but should remain separate from it.

**[The Data Mining Process](https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch02.html" \l "twodot4_the_data_mining_process)**

* Data mining involves both science and technology but also requires creative problem-solving skills.
* The CRISP process includes stages like business understanding, data understanding, data preparation, modeling, evaluation, and deployment. It provides a structured approach to data mining projects.

**[Implications for Managing the Data Science Team](https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch02.html" \l "twodot5_implications_for_managing_the_da)**

* Data mining projects are often treated and managed as engineering projects, which is understandable when they are initiated by software departments, with data generated by a large software system and analytics results fed back into it.
* The CRISP cycle is based around exploration; it iterates on approaches and strategy rather than on software designs. Outcomes are far less certain, and the results of a given step may change the fundamental understanding of the problem.

**[Other Analytics Techniques and Technologies](https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch02.html" \l "other_analytics_techniques_and_techs)**

* Data science encompasses six related analytic techniques: statistics, database querying, data warehousing, regression analysis, machine learning and data modeling, and answering business questions with these techniques.

**[Summary](https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch02.html" \l "twodot7_summary)**

* Data mining is both a science and an art, with a well-defined process that enhances the likelihood of success.
* Canonical task types in data science, such as classification, regression, and clustering, serve different purposes and involve specific solution techniques.
* Data mining complements other technologies like statistical hypothesis testing and database querying.
* The data mining process serves as a valuable framework for analyzing and evaluating data science projects, helping to assess their viability and conceptual soundness.

Chapter 2 of “Data Science for Business” by Foster Provost and Tom Fawcett introduces the distinction between supervised and unsupervised methods, showcasing how each can be harnessed to address specific business goals. As defined, supervised learning is a “machine learning approach that’s defined by its use of labeled datasets. These datasets are designed to train or “supervise” algorithms into classifying data or predicting outcomes accurately.” (Delua, 2021). Meanwhile, unsupervised learning “uses machine learning algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns in data without the need for human intervention (hence, they are “unsupervised”).” (Delua, 2021). For instance, supervised methods like classification and regression can be employed for predictive tasks like customer churn prediction. At the same time, unsupervised techniques like clustering are instrumental in segmenting customer populations for targeted marketing strategies. Understanding the right fit for each task type is crucial for success.

The chapter also delves into the data mining process, which acts as a systematic framework for tackling complex business issues in a process called the Cross Industry Standard Process for Data Mining (CRISP-DM). This process, including stages such as data understanding, preparation, modeling, evaluation, and deployment, ensures that the insights gained from data analysis can be effectively integrated into the business decision-making process. Some see CRISP-DM as a flexible way to perform data mining, while others see it as a rigid approach “in part because of its reporting requirements are excessive for most projects.” (Hotz, 2023). Businesses can harness the full potential of their data assets by treating data science projects as iterative explorations and emphasizing the creative aspects alongside engineering practices.

**References**

Delua, J. (2021, March 12). *Supervised vs. Unsupervised Learning: What’s the Difference?* IBM Blog. <https://www.ibm.com/blog/supervised-vs-unsupervised-learning/>

Hotz, N. (2023, January 19). *What is CRISP DM?*. Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/>

Provost, F., & Fawcett, T. (2013, August). *Data Science for Business*. O’Reilly Online Learning. <https://learning.oreilly.com/library/view/data-science-for/9781449374273/ch01.html#_summary>