Summary

* Tabular data is data organized in rows and columns, such as data in CSV files or relational database tables.
* Structured data is sometimes used as an alternative term for tabular data, but it is a broader concept, including JSON formatted data.
* Tabular data makes up a small portion of all the digital data in the world, but it has an enormous effect on our lives.
* Compared to other types of data (e.g., images, video, text, audio), the type of data that most jobs revolve around is tabular data, so learning how to efficiently apply machine learning/deep learning to tabular data is a useful skill that many people can apply to their jobs.
* In this book, we simply refer to machine learning approaches excluding neural networks (going from linear regressions to gradient boosting methods) as classical machine learning or just machine learning to distinguish them from deep learning.
* Compared to deep learning with other types of data (e.g., images, video, text, audio), deep learning with tabular data gets little attention from academic researchers (lack of good business data).
* Conventional wisdom is to use a gradient boosting approach like XGBoost with tabular data.
* There’s a lively debate in social media about whether or not there is a place for deep learning in solving problems involving tabular data. In this book, we don’t pick a side in this debate. Instead, we try to objectively describe why you would use machine learning or deep learning for a given tabular data problem and the best practices to use for each approach.
* Tabular data has some unique characteristics not shared by other types of data, such as images, video, or text. These characteristics include a lack of large open-source datasets that represent the kinds of tabular datasets that you would see in real-world business problems.
* Generative AI, especially LLMs, significantly affects how AI is perceived, diffused across individuals and organizations, and utilized. LLMs can help automate various tasks related to tabular data analysis and modelling, especially when related to textual inputs and outputs