6. Summary

Print to PDF

This chapter illustrated that the design of a *baseline* fraud detection system can be achieved using simple preprocessing strategies, and standard machine learning classifiers. In particular, we managed to obtain fraud detection performances that are well above those of a random classifier.

The chapter however only scratched the surface of how to approach a fraud detection problem. As we will see, a large number of more advanced techniques can be used to improve the performances. Performances can be addressed in terms of fraud detection accuracies, but also in terms of computational requirements (memory/execution times). The latter is in practice important during training, as fraud detection systems must deal with large amounts of data (much higher than those used in this baseline example) and also during inference for real-time or near real-time processing. One must generally carefully considers tradeoffs between accuracy and computational requirements.

The advanced chapters will cover in detail the possible avenues that can be explored to improve the proposed baseline approach.

Prior to that, the focus of the next two chapters will more specifically address the experimental methodology, that is, what performance measures should be used, and how these can be estimated. These issues are foundational for finding an objective way to compare the performances of different fraud detection systems and identifying the best performing one.

Previous
5. Real-world data

Next
1. Introduction