

5. Summary

Model selection consists in selecting the model that is expected to provide the best prediction performances on future data. For a fraud detection system, the best model can be defined as the model that has the highest expected fraud detection performances on the next block of transactions.

The estimation of model performances on future data is obtained by a validation procedure. This chapter covered different types of validation procedures and highlighted the benefits of the **prequential validation strategy** for estimating the fraud detection performances of a prediction model. Prequential validation allows to provide **accurate estimates of the fraud detection performances on future transactions, together with confidence intervals**.

Validation procedures are however computationally intensive tasks. They require to repeat the training procedures many times in order to assess the performances of prediction models with different hyperparameters and using different sets of data. The computation time of the validation procedure can become a bottleneck when models must be regularly updated.

Print to PDF ►

A key challenge for model selection consists in efficiently exploring the space of model hyperparameters in order to best address the trade-off between fraud detection performances and computation times. This chapter covered random search as a possible strategy to more efficiently explore the space of model hyperparameters. The next chapter will present alternative strategies that can address this trade-off by reducing the size of the dataset, in particular using undersampling strategies (Chapter 6).

◀ Previous
4. Real-world data

Next ▶
1. Introduction

By [Machine Learning Group \(Université Libre de Bruxelles - ULB\)](#).

Code released under a [GNU GPL v3.0 license](#). Prose and pictures released under a [CC BY-SA 4.0 license](#).