

4 Preprocessing and pipelines

0%

This chapter will introduce the notion of pipelines and how scikit-learn allows for transformers and estimators to be chained together and used as a single unit. Pre-processing techniques will be then be introduced as a way to enhance model performance and pipelines will be the glue that ties together concepts in the prior chapters.

▶ Preprocessing data	50 xp
</> Exploring categorical features	100 xp
</> Creating dummy variables	100 xp
</> Regression with categorical features	100 xp
▶ Handling missing data	50 xp
</> Dropping missing data	100 xp
</> Imputing missing data in a ML Pipeline I	100 xp
</> Imputing missing data in a ML Pipeline II	100 xp
▶ Centering and scaling	50 xp
</> Centering and scaling your data	100 xp
</> Centering and scaling in a pipeline	100 xp
</> Bringing it all together I: Pipeline for classification	100 xp
</> Bringing it all together II: Pipeline for regression	100 xp
▶ Final thoughts	50 xp

HIDE CHAPTER DETAILS