Abaixo segue a sugestão de dois livros que apresentam desde conceitos básicos até avançados sobre Machine Learning:

[**Introduction to Machine Learning** de **Ethem Alpaydin**](https://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf)

[**Foundations of Machine Learning** de **Mohri Mehryar**](https://pdfs.semanticscholar.org/e923/9469aba4bccf3e36d1c27894721e8dbefc44.pdf)

A primeira fonte que você pode buscar mais exemplos sobre pré-processamento de bases de dados é a própria documentação do sklearn. Veja abaixo:

[Preprocessing data](https://scikit-learn.org/stable/modules/preprocessing.html)

Livro **Python Machine Learning** de **Sebastian Raschka**: o capítulo 4 do livro é sobre a construção de boas bases de dados utilizando técnicas de pré-processamento

Livro **Machine Learning with Python** de **Chris Albon**: os capítulos iniciais do livro apresentam uma variedade grande de técnicas de pré-processamento

Capítulo **Data Quality** de **Tamraparni Dasu** e **Theodore Johnson** da série de livros **Wiley Series in Probability and Statistics**: ótimo complemento depois da aprendizagem das técnicas de pré-processamento

Caso você queira aprender como o Naive Bayes é utilizado para classificação de textos, veja meu outro curso **Mineração de Emoção em Textos com Python e NLTK**: é mostrada a teoria deste algoritmo aplicada em textos e é criado um classificador para detectar emoção em frases (nos recursos dessa aula você pode conferir mais sobre o curso)

Livro **Thoughtful Machine Learning** de **Matthew Kirk**: o capítulo 4 apresenta mais sobre a teoria do Naive Bayes e mostra como construir um classificador de spam

Livro **Data Algorithms** de **Mahmoud Parsian**: o capítulo 14 apresenta explicações sobre a teoria do Naive Bayes

Resources for this lecture

* [Curso Mineração de Emoção em Textos com Python e NLTK](https://www.udemy.com/mineracao-de-emocao-em-textos-com-python-e-nltk/?couponCode=MESITEN)

Livro **Data Mining with Decision Trees: Theory and Applications (Machine Perception and Artificial Intelligence)**de **Oded Z. Maimon**: livro de fácil compreensão e focado somente em árvores de decisão

Livro **C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)** de **J. Ross Quinlan**: um dos livros mais clássicos sobre o assunto

Livro **Decision Tress and Random Forests: A Visual Introduction For Begginers: A Simple Guide to Machine Learning With Decision Trees** de **Chris Smith** e **Mark Koning**: um dos livros mais fáceis e didáticos sobre o assunto

Livro **Machine Learning With Random Forests And Decision Trees: A Visual Guide For Beginners** de **Scott Hartshorn**: outro livro também muito didático

Artigo **Mineração de Dados com Árvores de Decisão** de **Jones Granatyr**, **Fábio Spak**, **Fabrício Enembreck** e **Otto Robert Lessing**: artigo que fizemos para a revista Devmedia, que mostra a teoria a implementação prática de árvores de decisão no software Weka