

Föreläsning 9

Övning

Boston

- Dataset Boston house-price *housing.csv*
- Undersök om det saknas värden och hantera i sådana fall de. Fundera över vilka metoder som passar för olika kolumner
- Undersök outliers med metoderna lärt i dagens lektion.
- Utför feature scaling med de olika metoderna vi har lärt idag.

King County Housing

- Testa antingen själv eller följ bloggen och gör Exploratory data analysis med King County Housing datan <https://medium.com/@allisonkelly42/exploratory-data-analysis-with-the-king-county-housing-data-c6d3f7eadc57>

Pipeline

- Vill du lära hur man skapar Machine Learning Pipelines? Det handlar om att skapa en standardiserad preprocessing som all nya data ska gå igenom för ett specifikt predikteringsproblem. Följ bloggen och koden för att lära <https://www.analyticsvidhya.com/blog/2020/01/build-your-first-machine-learning-pipeline-using-scikit-learn/>

Länkar

- Handle NULL values <https://towardsdatascience.com/7-ways-to-handle-missing-values-in-machine-learning-1a6326adf79e>
- Handle NULL values <https://www.analyticsvidhya.com/blog/2021/10/handling-missing-value/>
- Feature scaling <https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35>
- Scale data med outliers <https://machinelearningmastery.com/robust-scaler-transforms-for-machine-learning/>
- Feature Engineering for Machine Learning <https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114>

Data Camp

Supervised learning with scikit-learn <https://app.datacamp.com/learn/courses/supervised-learning-with-scikit-learn>