

EINFÜHRUNG IN DAS MASCHINELLE LERNEN – L2

KLASSIFIKATION MIT XGBOOST

Bekommen in L2

Abgabe in L3

Benutze das bereinigte Ames Housing Dataset, mit passendem Encoding für kategoriale Features.

Installiere die xgboost Bibliothek.

Benutze train-test-split mit 10%, bzw. 20%. Wie ändern sich die Ergebnisse?

Implementiere xgboost als Multiclass-Classifer. Welche sollten die Klassen sein? Was ist am sinnvollsten und wichtigsten zu vorhersagen? Denke an [stratification](#). Versuche mit oder ohne.

Zeichne den ersten und den letzten Baum, und die Feature Importance. Siehe [hier](#) für die Dokumentation. Präsentiere die [Metriken](#) precision, accuracy, recall, f1. Interpretiere die Ergebnisse. Zeichne die [Konfusionsmatrix](#).

Benutze [k-fold cross validation](#). Benutze k=5 für train test split mit 20%, und k=10 für train test split mit 10%. Interpretiere das Ergebnis.

Vergleiche die Ergebnisse für verschiedene Datenvorbereitungsstrategien.

Experimentiere mit Hyperparameter vom Modell. Ändere händisch Hyperparameter des Modells und untersuche, wie sich die Ergebnisse ändern.

Beschreibe alle Ergebnisse im Notebook, versuche, Erklärungen zu finden.

Bonus: Dasselbe auch für lightgbm. Vergleiche die entstandenen Bäume und Ergebnisse für beide Methoden.