

Versuch 3: Modelltraining und Hyperparameter-Optimierung ohne 0-Werte

Datengrundlage: complete_data.csv

Features: Mean, Min, Max von P,D und F

Für die Vorhersage der Zugfestigkeit wurden folgende maschinelle Lernmodelle evaluiert:

- Lineare Modelle:
 - Linear Regression
 - Ridge
 - Lasso
 - ElasticNet
- Baumbasierte Modelle:
 - DecisionTreeRegressor
 - RandomForest
 - GradientBoosting
- Nichtlineare Modelle:
 - Support Vector Regression (SVR)
 - KNeighborsRegressor
 - MLPRegressor

Um optimale Modellparameter zu finden, wurde RandomizedSearchCV mit Kreuzvalidierung eingesetzt. Dabei wurden verschiedene Hyperparameter getestet, um Überanpassung zu vermeiden und die Generalisierungsfähigkeit der Modelle zu verbessern.

Evaluation der Modelle

Die Modelle wurden anhand folgender Metriken evaluiert:

- Mean Absolute Error (MAE)
- Visualisierung der Vorhersagen durch Soll-Ist-Vergleiche

Die Ergebnisse wurden gespeichert und zur weiteren Analyse in einer CSV-Datei festgehalten.

Ergebnisse

Model	MAE
LinearRegression	271,52646
Ridge	271,5266
Lasso	271,5265
ElasticNet	271,52695
DecisionTreeRegressor	114,82627
RandomForest	99,999263
GradientBoosting	102,01773
SVR	290,10232
KNeighborsRegressor	137,43968
MLPRegressor	181,5113

Vergleich zu Versuch 01 (Alle Rohdaten wurden verwendet):

Modell	MAE (bereinigt)	MAE (Rohdaten)	Änderung
LinearRegression	271,53	270,75	+0,30%
Ridge	271,53	270,75	+0,30%
Lasso	271,53	270,75	+0,30%
ElasticNet	271,53	270,84	+0,25%
DecisionTreeRegressor	114,83	110,07	+4,1
RandomForest	99,99	95,77	+4,2%
GradientBoosting	102,02	96,67	+5,2%
SVR (linear)	290,1	288,88	0,40%
KNeighborsRegressor	137,44	127,79	+7%
MLPRegressor	181,51	170,16	+6,3%

Fazit:

- Besonders KNeighborsRegressor und MLPRegressor zeigen eine deutliche Verschlechterung nach der Bereinigung.
- Lineare Modelle bleiben stabil, egal ob bereinigt oder nicht
- Dies deutet darauf hin, dass diese Modelle stark von der Rohdatenvielfalt profitieren und durch das Entfernen bestimmter Daten an Vorhersagekraft verlieren.
- SVR bleibt weiterhin das schlechteste Modell, unabhängig von der Datenbereinigung.



