

Versuch 04: Anpassung & Erweiterung der Features und Ausdünnung der Modelle:

Anpassung und Erweiterung der verwendeten Features vorgenommen, um die Effizienz und Vorhersagegenauigkeit ausgewählter Modelle

- Ridge
- Lasso
- ElasticNet
- SVR mit linearem Kernel

zu verbessern. Dabei wurden Merkmale wie das Minimum, der Wertebereich (Range) und die relative Veränderung entfernt, da sie zuvor als wenig aussagekräftig identifiziert wurden. Eine Signifikanzanalyse durch F-Regression (ANOVA F-Test) und Lasso ermöglichte die gezielte Auswahl relevanter Features. Ziel des Versuchs war es, die Modelle durch eine fokussierte Auswahl stabiler und signifikanter Merkmale effizienter zu gestalten.

Signifikanzanalyse mittels F-Regression (ANOVA F-Test) & Lasso:

Feature	LASSO-Koeffizient	F-Value	p-Value (ANOVA)
P_mean	0,78	15,23	0,0001
F_max	0,62	12,89	0,0003
D_std	0,41	9,56	0,0021
P_IQR	0,35	8,75	0,0045
F_75	0,29	6,92	0,0098
D_50	0,21	6,15	0,0152
P_max	0,18	4,89	0,0376
F_mean	0,12	4,12	0,0511
D_max	0,05	2,75	0,0983
P_50	≈ 0,00	1,32	0,2175
F_std	≈ 0,00	1,1	0,2784
D_IQR	≈ 0,00	0,87	0,3512
P_75	≈ 0,00	0,64	0,4147
D_25	≈ 0,00	0,59	0,4671
F_50	≈ 0,00	0,42	0,5198
D_mean	≈ 0,00	0,31	0,6142
F_IQR	≈ 0,00	0,27	0,6783
P_25	≈ 0,00	0,19	0,7419

Anova:

- $p < 0.01 \rightarrow$ Sehr signifikant (stark relevant)
- $p < 0.05$ $p < 0.05 \rightarrow$ Signifikant (relevant für das Modell)
- $p < 0.1$ $p < 0.1 \rightarrow$ Mäßig signifikant (könnte hilfreich sein)
- $p > 0.1$ $p > 0.1 \rightarrow$ Weniger signifikant (evtl. nicht notwendig)

Lasso:

- Stark signifikante Merkmale: P_mean, F_max, P_IQR, D_std
- Mittelmäßig signifikante Merkmale: F_75, D_50, P_max, F_mean
- Weniger signifikante Merkmale: D_max, P_50
- Nicht signifikante Merkmale: Alle anderen

Vorgehen:

Die 4 genannten Modelle werden mit folgenden statstischen Merkmalen trainiert:

- 'P_mean': group['P'].mean(),
- 'F_max': group['F'].max(),
- 'D_std': group['D'].std(),
- 'P_IQR': group['P'].quantile(0.75) - group['P'].quantile(0.25),
- 'F_75': group['F'].quantile(0.75),
- 'D_50': group['D'].median(),
- 'P_max': group['P'].max(),
- 'F_mean': group['F'].mean()

Model	Best Params	MAE
Ridge	{'alpha': np.float64(0.0020235896477251557)}	224.37695568892678
RidgeCV	{'cv': 5}	227.47146407935816
Lasso	{'alpha': np.float64(0.028117686979742307)}	224.37075762640302
ElasticNet	{'l1_ratio': np.float64(1.0), 'alpha': np.float64(0.05689866029018299)}	224.37276404054523
SVR	{'gamma': 'scale', 'epsilon': np.float64(0.001623776739188721), 'degree': np.int64(3), 'C': np.float64(12.067926406393289)}	271.17378272310486

