

Versuch 02: Anpassung & Erweiterung der Features

Datengrundlage: complete_data.csv

Im Rahmen des zweiten Versuchs wurden die Features erweitert und angepasst, um die Vorhersagequalität der Modelle zu verbessern.

Die Features wurden dabei in vier Hauptgruppen unterteilt:

- Grundlegende statistische Merkmale
- Variabilitäts- und Streuungsmerkmale
- dynamische Merkmale zur Prozesscharakterisierung
- signalverarbeitungstechnische Merkmale.

Diese Anpassungen sollen die Fähigkeit der Modelle verbessern, zwischen guten und fehlerhaften Schweißungen zu unterscheiden und präzise Vorhersagen der Zugfestigkeit zu ermöglichen. Die optimalen Parameter der Modelle aus Versuch 01 wurden übernommen, um den Einfluss der neuen Merkmale effizient bewerten zu können.

Grundlegende Statistische Merkmale

Folgende Feature wurden in diesem Versuch berücksichtigt:

- **P_mean, F_mean, D_mean:** Der Durchschnittswert der jeweiligen Messgrößen. Dies gibt an, auf welchem typischen Leistungs-, Kraft- oder Distanzniveau das System arbeitet.
- **P_max, F_max, D_max:** Das Maximum zeigt den höchsten Wert während eines Zyklus, was wichtig für Spitzenlasten und Überlastungen ist.

Variabilitäts- und Streuungsmerkmale

Diese Features helfen, Unterschiede in der Schweißqualität zu erkennen.

- **P_std, F_std, D_std:** Die Standardabweichung zeigt die Streuung der Werte innerhalb eines Schweißzyklus. Eine hohe Abweichung kann auf Instabilitäten im Prozess hinweisen.
- **P_IQR, F_IQR, D_IQR:** Der Interquartilsabstand (IQR) misst die Differenz zwischen dem oberen (75%) und unteren (25%) Quartil. Dies ist robust gegenüber Ausreißern und zeigt die Kernvariabilität.
- **P_25, F_25, D_25:** Das erste Quartil gibt an, wo 25% der Werte unterhalb liegen, also ein Maß für den unteren Bereich der Verteilung.
- **P_50, F_50, D_50:** Der Median (50%) zeigt den zentralen Wert der Verteilung.
- **P_75, F_75, D_75:** Das dritte Quartil gibt an, wo 75% der Werte unterhalb liegen, was den oberen Bereich der Verteilung beschreibt.

Dynamische Merkmale

Diese Features messen Schwankungen und Unterschiede während eines Zyklus.

- **P_range, F_range, D_range:** Der Wertebereich zwischen Minimum und Maximum gibt die gesamte Spanne der Schwankungen an.
- **P_rel_change, F_rel_change, D_rel_change:** Die relative Veränderung ist ein normalisierter Wert, der die Schwankungen in Relation zur mittleren Größe setzt. Dadurch können Modelle unabhängig von absoluten Werten lernen.

Signalverarbeitungstechnische Merkmale

Diese Merkmale helfen, die Gesamtenergie und Belastung eines Schweißprozesses zu bestimmen.

- **P_integral, F_integral, D_integral:** Die Summe der jeweiligen Werte entspricht der Energie oder Gesamtlast während des Prozesses. Dies ist besonders relevant, da der Energieeintrag direkt die Schweißqualität beeinflussen kann.

Ergebnisse

Als Hyperparameter wurden die optimalen Werte aus Versuch01 ausgewählt:

| Model | MAE |
|-----------------------|--------------------|
| LinearRegression | 150.98136925703616 |
| Ridge | 150.73270250171518 |
| Lasso | 153.9188857661624 |
| ElasticNet | 150.51441856814324 |
| DecisionTreeRegressor | 112.92660628330759 |
| GradientBoosting | 82.55055407120436 |
| KNeighborsRegressor | 142.88996804555242 |
| MLPRegressor | 9111.165939859122 |

Fazit:

- GradientBoosting-Modell zeigte deutliche MAE-Verbesserungen.
- Andere Modelle, wie RandomForest und SVR, zeigten Instabilitäten und brachen ab.
- Merkmale wie Min, Range und Integral sind wenig aussagekräftig (Da immer 0)
- Einige Modelle, insbesondere neuronale Netzwerke, konvergierten selbst bei hoher Iterationsanzahl nicht.
- RandomForest und SVR stürzten bei größeren Datensätzen ab.