

****Abschnitt 1: Vorhersagemodelle (Prognosemodelle)****

1. ****Datenaufteilung (Daten Splitting):**** In diesem Schritt wurden die Daten mithilfe von `train_test_split` in Trainings- und Testsets aufgeteilt. Dieser Schritt ist entscheidend bei der Entwicklung eines Prognosemodells. Die Technik der `MinMaxScaler` wurde verwendet, um die Merkmale in einen bestimmten Bereich zu transformieren. Dies hilft, Probleme aufgrund unterschiedlicher Wertebereiche der Merkmale zu vermeiden.

2. ****Statistische Modelle:**** In diesem Abschnitt wurden zwei statistische Modelle zur Vorhersage der Anzahl der ausgeliehenen Fahrräder vorgestellt:

- ****Lineare Regression:**** Es wurde erläutert, wie ein lineares Regressionsmodell aufgebaut und zur Vorhersage der Anzahl der ausgeliehenen Fahrräder verwendet wird. Leistungsmaße wie MSE, MAE, R-Quadrat und RMSE wurden zur Bewertung der Modellleistung präsentiert.

- ****RandomForest-Modell:**** Es wurde erläutert, wie ein RandomForest-Modell aufgebaut und zur Vorhersage der Anzahl der ausgeliehenen Fahrräder verwendet wird. Hier wurden ebenfalls die oben genannten Leistungsmaße zur Bewertung des Modells vorgestellt.

****Abschnitt 1.1.2: Optimierung der Parameter****

1. ****Rastersuche (Grid Search):**** Es wurde erklärt, wie eine Rastersuche zur Optimierung der Parameter des linearen Regressionsmodells durchgeführt wird. Dies beinhaltet das Ausprobieren verschiedener möglicher Parameterwerte und das Messen der Modellleistung mit verschiedenen Metriken. Dies hilft, die besten Parameterkonfigurationen für das Modell zu finden.

2. ****Zufallssuche (Randomized Search):**** Dieser Abschnitt erläutert, wie eine zufällige Suche zur Optimierung der Parameter des RandomForest-Modells verwendet wird. Dies beinhaltet das Ausprobieren einer zufälligen Auswahl von Parameterwerten und das Messen der Modellleistung. Dies kann nützlich sein, wenn Sie eine große Menge möglicher Parameter haben und wenig Zeit zur Modelloptimierung zur Verfügung steht.

****Interessanter Abschnitt: Experiment zur Zieltransformation****

Es wurde erläutert, wie die Zielvariable (Anzahl der ausgeliehenen Fahrräder) mithilfe der Funktion `np.log1p` in einen Logarithmus transformiert wurde. Die möglichen Gründe für diese Transformation wurden erläutert:

- Verbesserung der Verteilung: Es wurde darauf eingegangen, wie die ursprüngliche Zielvariable möglicherweise nicht normal verteilt ist und wie die Transformation in einen Logarithmus dazu beitragen kann, die Verteilung der Daten zu normalisieren.
- Reduzierung großer Ausreißer: Es wurde erklärt, wie die Transformation in einen Logarithmus dazu beitragen kann, große Ausreißerwerte zu reduzieren, wodurch die Daten stabiler werden und der Einfluss von Ausreißern verringert wird.
- Linearisierung der Beziehung: Es wurde erläutert, wie die Transformation des Ziels in einen Logarithmus die Beziehung zwischen den Variablen linearer machen kann, was dem Modell beim Erlernen der Beziehung zwischen ihnen hilft.

Für die präsentierten Ergebnisse wurden sie im Detail für beide Modelle (lineare Regression und RandomForest) vor und nach der Zieltransformation (Logarithmus) dargestellt. Es wurden verschiedene Metriken wie RMSLE, MSE, MAE, R-Quadrat und Maximaler Fehler verwendet, um die Modellleistung zu bewerten.

****Erklärung:** Warum hat sich der R-Quadrat-Wert für das lineare Modell nicht geändert?

Es wurde erklärt, wie der R-Quadrat-Wert funktioniert und warum er sich durch die Transformation in einen Logarithmus nicht wesentlich verändert hat. Der Fokus lag darauf, dass der R-Quadrat-Wert das Verhältnis der Varianz im abhängigen Variable, das vom Modell erklärt wird, misst. Da die Transformation in einen Logarithmus diese Varianz nicht beeinflusst, hat sich der R-Quadrat-Wert für das lineare Modell nicht wesentlich verändert.

****Abschnitt 6: Ergebnisse****

Die endgültigen Ergebnisse für die beiden besten Modelle, die optimierte lineare Regression und das optimierte RandomForest-Modell, wurden präsentiert. Die Leistung dieser Modelle wurde mit verschiedenen Metriken wie RMSE, MAE, R-Quadrat und Maximaler Fehler gemessen.

Die Ergebnisse ohne Zieltransformation (Logarithmus) lauteten wie folgt:

****Lineare Regression:****

- Wurzel des mittleren quadratischen Fehlers (RMSE): 901,1443
- Mittlerer absoluter Fehler (MAE): 660,3402
- R-Quadrat (Bestimmtheitsmaß): 0,3187
- Maximaler Fehler: 4113,6991

****RandomForest:****

- Wurzel des mittleren quadratischen Fehlers (RMSE): 393,3547

- Mittlerer absoluter Fehler (MAE): 241,6334
- R-Quadrat (Bestimmtheitsmaß): 0,8702
- Maximaler Fehler: 2627,8388

Die Ergebnisse mit Zieltransformation (Logarithmus) lauteten wie folgt:

****Lineare Regression:****

- Wurzel des mittleren quadratischen logarithmischen Fehlers (RMSLE): 0,1359
- Mittlerer quadratischer Fehler (MSE): 0,8568
- Mittlerer absoluter Fehler (MAE): 0,7196
- R-Quadrat (Bestimmtheitsmaß): 0,4752
- Maximaler Fehler: 3,2866

****RandomForest:****

- Wurzel des mittleren quadratischen logarithmischen Fehlers (RMSLE): 0,0372
- Mittlerer quadratischer Fehler (MSE): 0,0477
- Mittlerer absoluter Fehler (MAE): 0,1390
- R-Quadrat (Bestimmtheitsmaß): 0,9707
- Maximaler Fehler: 1,6422

****Vergleich der Ergebnisse:****

Beim Vergleich der Ergebnisse zwischen den beiden Modellen (lineare Regression und RandomForest) und zwischen den Ergebnissen mit und ohne Zieltransformation (Logarithmus)

wurde Folgendes festgestellt:

1. ****Lineare Regression ohne Zieltransformation vs. Lineare Regression mit Zieltransformation (Logarithmus):****

- RMSE: 901,1443 vs. 0,9352
- MAE: 660,3402 vs. 0,7337
- R-Quadrat: 0,3187 vs. 0,4808
- Maximaler Fehler: 4113,6991 vs. 3,2866

Es ist zu erkennen, dass die Zieltransformation in einen Logarithmus zu einer signifikanten Verbesserung der Leistung des linearen Regressionsmodells auf den meisten Metriken geführt hat. Die RMSE- und MAE-Werte wurden erheblich reduziert, und das R-Quadrat hat sich verbessert. Der einzige Aspekt, der sich verschlechtert hat, ist der maximale Fehler, der aufgrund der Transformation zugenommen hat.

2. ****RandomForest ohne Zieltransformation vs. RandomForest mit Zieltransformation (Logarithmus):****

- RMSE: 393,3547 vs. 0,3379
- MAE: 241,6334 vs. 0,2524
- R-Quadrat: 0,8702 vs. 0,9322
- Maximaler Fehler: 2627,8388 vs. 1,6422

In diesem Fall hat die Zieltransformation in einen Logarithmus ebenfalls zu einer erheblichen Verbesserung der Leistung des RandomForest-Modells auf den meisten Metriken geführt. Die RMSE- und MAE-Werte wurden erheblich reduziert, und das R-Quadrat hat sich verbessert.

****Zusammenfassung der Ergebnisse:****

Die präsentierten Ergebnisse zeigen, dass das RandomForest-Modell im Vergleich zum linearen Modell insgesamt eine bessere Leistung erzielt. Es wurde betont, wie wichtig es ist, das richtige Modell basierend auf dem Projektumfeld, den Genauigkeitsanforderungen und den Eigenschaften der Daten auszuwählen.

****Schlussfolgerungen und Empfehlungen:****

Aufgrund des Vergleichs zwischen den beiden Modellen (lineare Regression und RandomForest) und zwischen den Ergebnissen mit und ohne Zieltransformation (Logarithmus) kann festgestellt werden, dass die Zieltransformation zu einer erheblichen Verbesserung der Modellleistung geführt hat. Die Wahl des geeigneten Modells sollte basierend auf dem Projektumfeld und den Genauigkeitsanforderungen getroffen werden.

Zusätzlich können weitere Verbesserungen in Betracht gezogen werden, wie die Optimierung der Datenmerkmale, die Anpassung der Modellparameter und die Verwendung von Deep Learning-Techniken, insbesondere bei großen Datensätzen. Eine mögliche Erweiterung dieser Analysen könnte auch den Vergleich mit anderen Modellen umfassen, falls erforderlich.