

Fachpräsentation zum Kurs

Model Engineering (DLMDWME01)

Tutor: Markus Pak

Vorgelegt von:

Iskra Fodor, Matr.-Nr. 32112241

Potsdam, den 20.10.2024

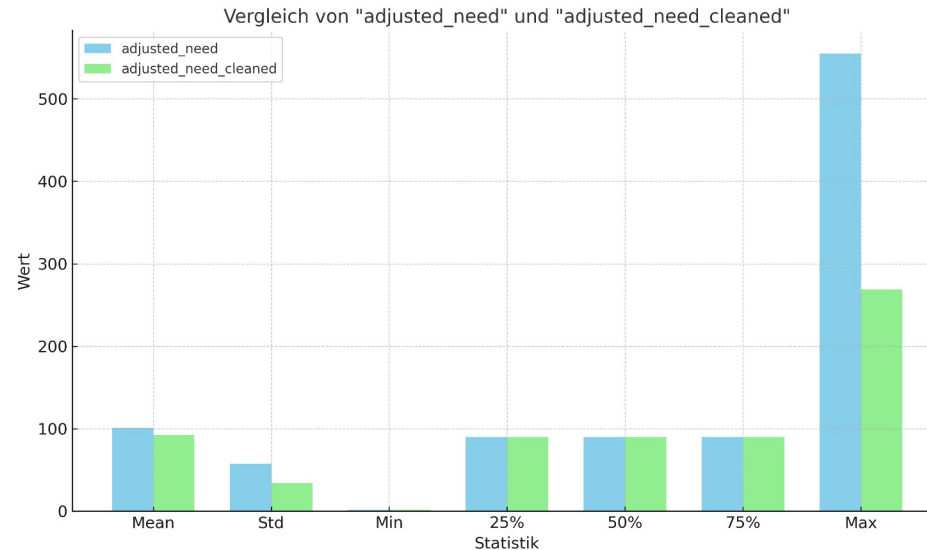
- Ziel
- Datengrundlage
- Datenvorbereitung
- Modellentwicklung
- Modellbewertung –
fortgeschrittenes Modell
- Ergebnisse & Fazit
- Benutzeroberfläche
- Ausblick und Empfehlungen

Ziel des Projekts ist die Entwicklung eines Vorhersagemodells für den Berliner Rotkreuz-Rettungsdienst, das auf Tagesbasis die Anzahl der benötigten Einsatzfahrenden im Bereitschaftsdienst effizient abschätzt.

- Datenquelle: sickness_table.csv
- Anzahl der Datensätze: 1152
- Hauptmerkmale: calls_per_duty, month_sin, month_cos, year, day_of_week
- Zielvariable: adjusted_need, eine berechnete Variable zur Bestimmung des tatsächlichen Bedarfs an Einsatzfahrenden.

- Feature Engineering: Zeitbasierte Features (z.B. zyklische Transformationen für Monate und Wochentage) sowie gleitende Durchschnitte wurden erstellt, um saisonale Muster besser abzubilden.
- Erkennung und Behandlung von Ausreißern: `adjusted_need` wurde mithilfe des Z-Scores bereinigt, um die Prognose zu verbessern:
 - `adjusted_need` enthält die ursprünglichen Werte der Zielvariable, der Maximalwert von 555 ist deutlich höher als der Median von 90.

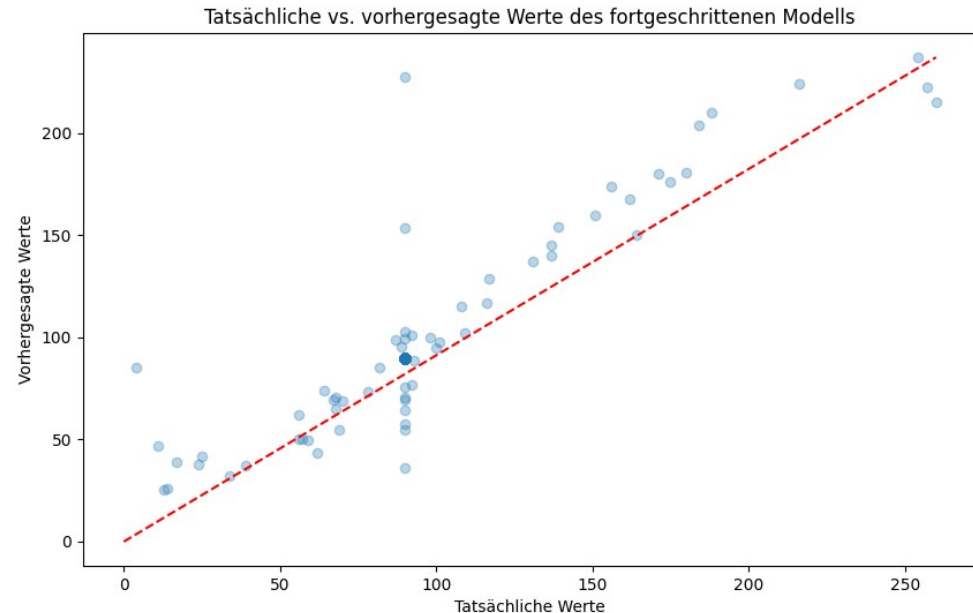
- `adjusted_need_cleaned` enthält keine Ausreißer mehr, was zu einer Reduzierung des Maximums auf 269 und einer geringeren Standardabweichung führt.



- Baseline-Modell: DummyRegressor (auf Basis des Medians) zur Bestimmung der Ausgangsleistung.
- Leistungsmetriken:
 - MAE: 11.28 (durchschnittliche Abweichung von ca. 11 Einheiten).
 - RMSE: 31.06 (weist auf größere Fehler hin).
 - R^2 : -0.01 (keine Varianz erklärt).

- RandomForestRegressor mit optimierten Hyperparametern.
- Benutzerdefinierte Verlustfunktion: Bestraft Untervorhersagen dreimal stärker als Übervorhersagen.
- Modellleistung:
 - MAE: 4.47 (durchschnittliche Abweichung von ca. 4 Einsatzfahrenden).
 - RMSE: 14.19 (einige größere Fehler vorhanden).
 - R^2 : 0.79 (zeigt eine moderate Anpassung an die Daten).
 - Custom Loss: 1862.52 (besonders für die spezifischen Anforderungen des Rettungsdienstes angepasst).

- Die roten gestrichelten Linie repräsentiert den idealen Verlauf (perfekte Vorhersage)
- Die meisten Punkte liegen nahe roten Linie --> hohe Vorhersagegenauigkeit
- Einige Ausreißer deuten auf größere Vorhersagefehler hin, die das Modell nicht vollständig erfasst hat. Dennoch bietet es eine robuste Schätzung für die Mehrzahl der Datenpunkte.



- Das Random Forest Modell liefert für den spezifischen Anwendungsfall des Rettungsdienstes gute Ergebnisse.
- Durch die Optimierung konnte die Leistung gegenüber dem einfachen Baseline-Modell deutlich gesteigert werden.
- Die benutzerdefinierte Verlustfunktion, welche die Untervorhersagen strenger bestraft, half dabei, den potenziellen Personalmangel zu minimieren.
- Das Modell eignet sich somit als wertvolles Werkzeug zur Planung und Vorhersage des Personaleinsatzes im Rettungsdienst.

Eine beispielhafte GUI wurde mit folgenden Funktionalitäten implementiert:

- Datenverwaltung: Benutzer können täglich Anruferdaten eingeben und speichern.
- Daten sichern: Eingaben werden in die CSV-Datei gespeichert.
- Vorhersage erstellen: Die GUI generiert auf Basis des fortgeschrittenen Modells eine Prognose für den nächsten Monat, sofern genügend Daten vorliegen.
- Visualisierung: Die Vorhersagen werden sowohl textuell als auch grafisch angezeigt.

- Mögliche Verbesserungen: Weitere Hyperparameter-Optimierungen oder das Hinzufügen externer Datenquellen.
- Langfristiges Ziel: Kontinuierliche Verbesserung des Modells durch neue Daten und Feedbackschleifen.

Fragen?

Vielen Dank für Ihre Aufmerksamkeit!