

Data Preparation & Feature Engineering

Leitfragen:

- Welche Daten stehen zur Verfügung? Welche externen Datenquellen wurden hinzugefügt?
- Welche Schritte wurden zur Datenbereinigung unternommen (Fehlwerte, Duplikate, Outlier)?
- Welche Feature Engineering-Techniken wurden angewendet (z. B. Aggregationen, zeitliche Merkmale, Encoding, Scaling / Normalization)?
- Welche Features haben sich als besonders wichtig erwiesen?
- Gibt es saisonale, geografische oder zeitliche Muster in den Daten?
- Welche Korrelationen bestehen zwischen den vorhandenen Features und der Zielvariable?
- Gab es Ungleichgewichte in den Daten (z. B. bestimmte Stadtteile mit mehr Daten als andere)? Wie wurde damit umgegangen?
- Gab es herausfordernde Datenpunkte (z. B. fehlerhafte Sensorwerte, unplausible Werte)? Wie wurde damit umgegangen?

Model Engineering

Leitfragen:

- Welche Algorithmen wurden getestet, und warum?
- Wurde Ensemble Learning (z. B. Bagging, Boosting, Stacking) genutzt? Falls ja, mit welchem Effekt?
- Welche Techniken wurden für die Datenaufteilung (Train/Validation/Test) genutzt?
- Wurde eine Zeitreihen-Split-Strategie verwendet (falls Zeitabhängigkeit besteht)?
- Wie wurde Overfitting vermieden? (Dropout, Regularisierung, Frühabbruch)
- Wurden unterschiedliche Hyperparameter ausprobiert? Welche Methoden wurden genutzt (Grid Search, Random Search, Bayesian Optimization)?
- Hat die Integration externer Daten die Modellleistung verbessert? Falls ja, wie?
- Wie unterscheiden sich die Modelle in ihrer Komplexität und Interpretierbarkeit?

- Welche Modelle sind robuster gegenüber Datenveränderungen oder Ausreißern?

Model Evaluation & Improvement

Leitfragen:

- Welche Evaluationsmetriken wurden genutzt (z. B. RMSE, MAE, R^2 , F1-Score, Precision/Recall)? Warum sind sie geeignet?
- Wie unterscheiden sich die Modelle hinsichtlich der wichtigsten Metriken?
- Gibt es Unterschiede in der Modellperformance über verschiedene Zeiträume oder Regionen hinweg?
- Welche Arten von Fehlern macht das Modell am häufigsten? Gibt es bestimmte Datenbereiche mit schlechterer Vorhersagequalität?
- Wie verhalten sich die Modelle auf unbekannten Daten (z. B. Testset oder Cross-Validation)?
- Welche Techniken wurden zur Verbesserung genutzt (z. B. Hyperparametertuning, Ensemble Learning)?
- Welche Features oder Methoden haben die Modellleistung am meisten beeinflusst?
- Wurden Modelle speziell für unterschiedliche Segmente der Daten optimiert (z. B. Cluster-basierte Modelle)?
- Gab es alternative Modellansätze, die sich als überlegen herausgestellt haben? Falls ja, warum?
- Gibt es Korrelationen zwischen bestimmten Features und Vorhersagefehlern?
- Wie hoch ist die Latenzzeit der Modelle? Könnte dies im produktiven Betrieb ein Problem sein?

Hypothetisches Deployment, Monitoring & Betrieb

Leitfragen:

- Wie könnte das Modell deployed werden (Batch vs. Echtzeit)?
- Welche Metriken sollten zur Laufzeit überwacht werden?
- Wie könnte Model Drift erkannt und behandelt werden?
- Welche ethischen oder technischen Herausforderungen könnten auftreten?