

Data Science Aufgabe 4: Kaggle Competition

Nun ist es an der Zeit, euer Wissen praktisch anzuwenden und an einer realen Datenanalyse-Herausforderung teilzunehmen. In dieser Aufgabe werdet ihr an der "Titanic Spaceship Competition" (<https://www.kaggle.com/c/spaceship-titanic>) auf Kaggle teilzunehmen. Diese Aufgabe bietet eine großartige Gelegenheit, eure Fähigkeiten in der Datenexploration, Feature-Engineering, Modellauswahl und -evaluation sowie im Hyperparameter-Tuning zu vertiefen.

Wichtige Hinweise für die Aufgabe:

1. **Datenexploration und -reinigung:** Untersucht das Datenset, um fehlende Werte zu identifizieren und geeignete Strategien zur Datenbereinigung anzuwenden.
2. **Feature Engineering:** Entwickelt neue Features aus den vorhandenen Daten, um zusätzliche Informationen zu extrahieren und die Modellgenauigkeit zu verbessern.
3. **Modellauswahl und -bewertung:** Experimentiert mit verschiedenen Machine-Learning-Modellen wie Entscheidungsbäumen, Random Forests oder logistischer Regression. Nutzt Cross-Validation, um die Leistung eurer Modelle zu bewerten.
4. **Hyperparameter-Tuning:** Optimiert die Parameter eurer Modelle mittels Grid Search oder Random Search, um die bestmöglichen Ergebnisse zu erzielen.

Abgabe der Aufgabe:

- Erstellt ein Jupyter-Notebook, das eure gesamte Analyse und eure Ergebnisse dokumentiert.
- Reicht eure Vorhersagen und euer Jupyter-Notebook auf der Kaggle-Plattform ein, um an der Competition teilzunehmen.

Teamorganisation:

Organisiert euch in den bereits bestehenden Teams auf Kaggle.

Viel Erfolg und viel Spaß beim Lernen und Entdecken!

PS: Falls ihr euch nicht auf Kaggle anmelden möchtet meldet euch bitte direkt bei mir.