

# Projektbeschreibung für die Vorlesung "Maschinelles Lernen"

---

## Aufgabenstellung

Erstellung eines *Jupyter-Notebooks* mit der Anwendung von Verfahren des Maschinellen Lernens auf den im Text-File *KHK\_Klassifikation.csv* vorgefundenen medizinischen Datensatz. Der Datensatz enthält mehrere klinische Parameter und eine (Risiko-)Klassifikation. Ziel ist es, eine möglichst gute Vorhersage für die (Risiko-)Klassifikation per ML-Modell zu erreichen.

## Aufgabenpunkte

- Das Jupyter-Notebook soll grob so ähnlich aufgebaut sein, wie die als Beispiel in der Vorlesung gezeigten Jupyter-Notebooks zur Regression und Klassifikation.
- Erstellung eines "*durchgehenden*" Jupyter-Files (mit Erklärungen zwischen den Code- bzw. "Analyse"-Blöcken)
- Laden der Daten, explorative Analyse der Daten (Umfang, Klassen, Klasseneinteilung, ...)
- Probieren Sie eventuell eine geeignete Dimensionsreduktion aus, z.B. eine PCA, t-SNE, UMAP .... Visualisieren Sie auf dem ins 2-Dimensionale projizierten Datensatz die Daten mit ihrer zugehörigen Klassifikation. Läßt sich aus dieser Visualisierung bereits ein Hinweis auf eine potentiell gute Separierbarkeit der Klassen ablesen?
- Anwendung verschiedener vorgestellter Klassifikationsverfahren, wie kNN, SVM, logistische Regression, Entscheidungsbäume, Ensembles von Entscheidungsbäumen und Neuronalen Netzen, um von den klinischen Parametern (*Features*) auf die Ziel-Variable *Risikoklasse* zu schließen.
- Beurteilen Sie die Ergebnisse Ihrer Klassifikationen.
- Manche Verfahren bieten die Möglichkeit die Bedeutung der einzelnen *Feature* bezüglich der Klassifikation abzuschätzen. Versuchen Sie mit dem/den geeigneten Verfahren eine solche Abschätzung. Welche *Feature* scheinen für die Klassifikation die bedeutenderen zu sein?
- Eventuell läßt sich durch *Feature-Engineering*, wie z.B. der Auswahl von prädiktiveren Features bzw. der Transformation von Features (beispielsweise die ersten x Hauptkomponenten der PCA), die Klassifikationsleistung verbessern. Unternehmen Sie einen Versuch mit einem der Klassifikationsverfahren auf einem veränderten *Feature-Set*!
- Vergessen Sie auch nicht Quellen anzugeben, falls Sie bestimmte andere Quellen heranziehen oder auf besondere (nicht-standard) Module/Packages zurückgreifen, z.B. für besondere grafische Darstellungen.

## Abgabe

Datum der Abgabe: 29.04.2025

Die Abgabe sollte ein Zip-Archiv sein, das das Jupyter-File und die von Ihnen verwendeten Dateien, falls Sie auf separate Python-Skripte oder sonstige für das Jupyter-Notebook notwendige Ressourcen zugegriffen haben, sowie das gerenderte Endergebnis enthält.