

Project Requirements – Angewandtes Programmierprojekt

Von: Laura Garcia Petershof, Matrikelnummer: 7216187

Hintergrund: Der vorliegenden Datensatz „Framingham Heart Study Dataset“ umfasst 4240 Patient:innen-Daten mit jeweils 15 Eigenschaften und der Angabe einer Zielvariable. Diese Zielvariable beschreibt, ob der/die jeweilige Patient:in ein Risiko für eine Herzerkrankung in den nächsten 10 Jahren besitzt.

Die 15 Eigenschaften umfassen zum einen demographische Eigenschaften, wie Geschlecht und Alter, das Rauchverhalten der Patient:innen als auch medizinische Eigenschaften, wie der Blutdruck, einzunehmende Medikamente und die Herzrate.

Ziel: Das Ziel dieses Projektes ist es, ein Machine-Learning-Modell zu entwickeln, welches eine Vorhersage über das Risiko einer Herzerkrankung in den nächsten 10 Jahren trifft. Dieses Modell soll mit der Hilfe von überwachten maschinellen Lernverfahren entwickelt werden. Hierbei sollen diverse Verfahren getestet und das Modul mit der besten Performance ausgewählt werden.

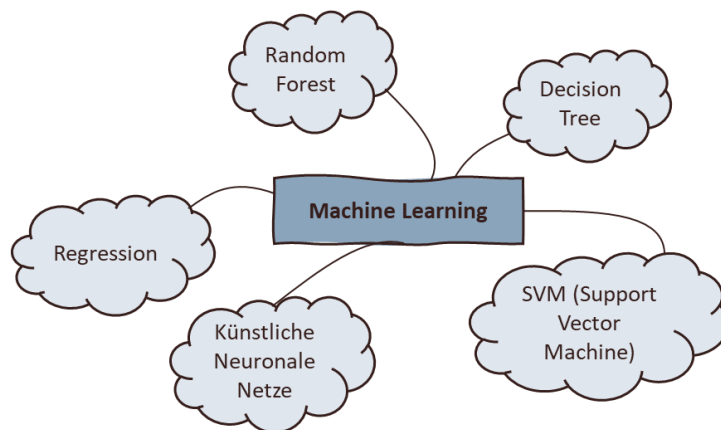


Abbildung 1: Cluster mit potenziellen überwachten maschinellen Lernverfahren (eigene Darstellung).

Vorgehensweise:

1. Problemverständnis:

Das RKI gibt an, dass eine der führenden Todesursachen in Deutschland eine Herz-Kreislauf-Erkrankung darstellt. Diese verursachen ungefähr 40 % aller Sterbefälle. Gleichzeitig führen Herz-Kreislauf-Erkrankungen zu einer Minderung der Lebensqualität und erzeugen hohe Kosten für die gesetzlichen Krankenkassen [1].

Mögliche Ursachen für eine Herzerkrankung sind beispielsweise das zunehmende Alter, ein ungesunder Lebensstil (rauchen und mangelnde körperliche Bewegung) und Bluthochdruck [1]. Aufgrund dessen ist es von hoher Bedeutung das Risiko von Patient:innen für eine Herz-Kreislauf-Erkrankung möglichst frühzeitig zu identifizieren, sodass diese vermieden werden kann. Somit kann die Lebensqualität der Patient:innen erhalten und die Krankheitskosten minimiert werden.

2. Datenerhebung:

Kaggle Datensatz: „Framingham Heart Study Dataset“ → Open Data

<https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset>

Die Daten entstammen der Framingham Heart Studie, die bereits seit 1948 in der dritten Generation durchgeführt wird. Die Studie dient der Identifikation von Ursachen und Risiken

von koronaren Herzerkrankungen. Die Studie zählt zu einer der bedeutsamsten Studien in den USA. Die Ergebnisse dieser Studie gehören mittlerweile zum medizinischen Standard [2].

3. Datenvorbereitung:

- Korrekte Beschriftung der Daten.
- Datenreinigung, Neuformatierung und Normalisierung.

4. Datenformatierung:

- Daten müssen zunächst in das gewünschte Format konvertiert werden.

5. Modelentwicklung:

Auswahl und Erprobung eines überwachten Lernverfahrens mit der besten Performance und Genauigkeit:

- Random Forest
- Decision Tree
- Regression
- KNN
- Support Vector Machine (SVM)

a. Training

Der ausgewählte ML-Algorithmus soll mit den Daten aus dem Datensatz trainiert werden.

b. Bewertung und Verbesserung

Im Anschluss soll der ML-Algorithmus auf seine Genauigkeit und Performance untersucht werden. Bei mangelhaften Ergebnissen sollen diese Parameter mittels einer Anpassung der Hyperparameter verbessert werden oder ein anderes maschinelles Lernverfahren ausgewählt werden.

➔ **Ergebnis:** ML-Modell zur Vorhersage des Risikos für eine Herzerkrankung in den nächsten 10 Jahren.

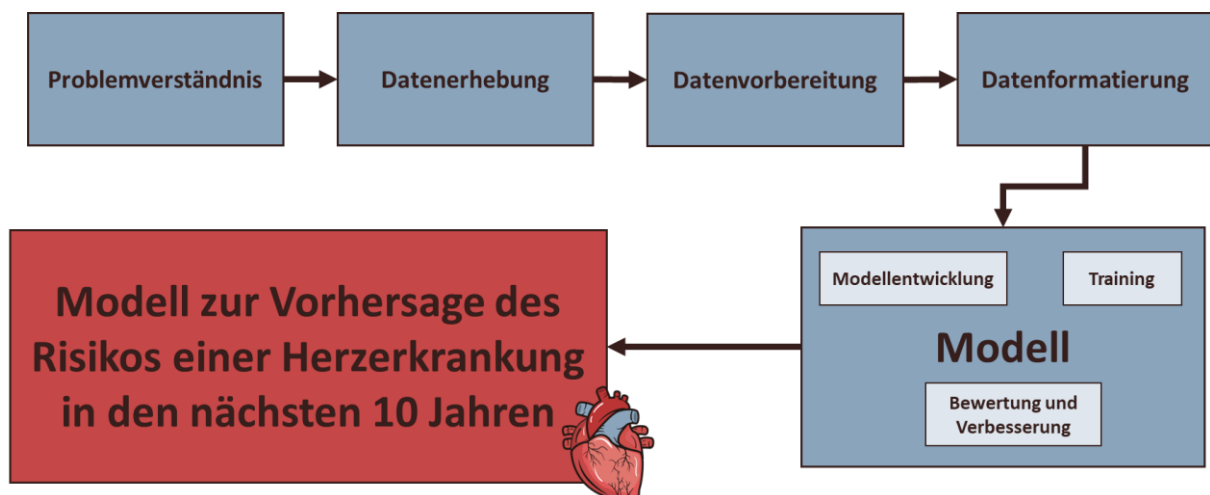


Abbildung 2: Darstellung des Projekt-Prozesses (in Anlehnung an [3]).

Ausblick: Im Anschluss an dieses Projekt ist es vorstellbar, dass das entwickelte Modell in eine Applikation eingebettet wird. Diese soll den Patient:innen eine Vorhersage über ihr Risiko, in den nächsten 10 Jahren an einer Herzerkrankung zu erkranken, geben.

Quellen

- [1] Robert-Koch-Institut (o. D.): Herz-Kreislauf-Erkrankungen, RKI - Themenschwerpunkt Herz-Kreislauf-Erkrankungen, [online]
https://www.rki.de/DE/Content/Gesundheitsmonitoring/Themen/Chronische_Erkrankungen/HKK/HKK_node.html [abgerufen am 27.10.2022].
- [2] Wikipedia-Autoren (2004): Framingham-Herz-Studie, [online]
<https://de.wikipedia.org/wiki/Framingham-Herz-Studie> [abgerufen am 27.10.2022].
- [3] Pajankar, Ashwin/Aditya Joshi (2022): *Hands-on Machine Learning with Python: Implement Neural Network Solutions with Scikit-learn and PyTorch*, 1st ed., Apress.