

Medical Data Science - Übungsabgabe

Krankheitsbild: Schwere Pneumonie
Schwerpunkt: Klinische Scores und ihre Validierung

Simon Bosse 613202
Anna Göing 606171
Jonas Jakob 640960
Vyvy Tran Ngoc 623420

7. Mai 2025

Inhaltsverzeichnis

1. Einleitung	2
2. Setup	2
2.1. Datenbankeinrichtung	2
2.2. Erste Exploration des Datensatzes	2
2.3. CITI-Kurs und MIMIC-IV Zugang	3
A. Anhang	4
A.1. CITI-Zertifikate aller Gruppenmitglieder	4

1. Einleitung

Die schwere Pneumonie ist eine Entzündung des Lungengewebes, die zu respiratorischer Insuffizienz führen kann und häufig eine mechanische Beatmung erfordert. Sie ist eine der häufigsten Ursachen für die Aufnahme auf die Intensivstation und war vor allem während der Corona-Pandemie ein oft diskutiertes Gesprächsthema in den Nachrichten, insbesondere in Bezug auf die Triage.

Im Rahmen der Übung könnte es interessant sein, im Hinblick auf die Priorisierung von Patienten und die grundsätzliche Entscheidung einer Aufnahme auf die Intensivstation, die verschiedenen Scoring-Systeme für Pneumonie zu betrachten und somit die Methoden für die Datenaufbereitung, sowie auch die darauf angewendete Berechnung der Scores zu betrachten.

- **Github Repository:** medicaldata (öffentlich zugänglich)

2. Setup

2.1. Datenbankeinrichtung

Die Software von Docker wurde für das entsprechende System heruntergeladen und installiert, sowie die Version mit dem `docker --version` Befehl getestet und ein PostgreSQL-Container konnte mithilfe der bereitgestellten `docker-compose.yml` gestartet werden. Das github repository wurde heruntergeladen und die benötigten Dateien `create.sql` und `load_gz.sql` zusammen mit dem Demo-MIMIC Datensatz in den Container geladen, so dass die Indizierung direkt vor Ort stattfinden konnte. Somit war der Zugriff auf den Datensatz eingerichtet.

Die Installation wurde mit dem Ausführen der folgenden Abfragen verifiziert:

- Anzahl der Patienten: 100 Patienten
- Anzahl der verschiedenen Diagnosen: 1472 Diagnosen

2.2. Erste Exploration des Datensatzes

Die Recherche nach ICD-9 und/oder ICD-10 Codes führte zu der folgenden Liste im Zusammenhang mit Pneumonie:

- **ICD-9:** Diseases of the respiratory system 460-519; Pneumonia und Influenza 480 - 488; Pneumonia 480 - 486
- **ICD-10:** Grippe und Pneumonie J09 - J18; Pneumonie J13 - J18

Die Exploration des Datensatzes ergab, dass exakt 8 verschiedene ICD-9 Codes für Pneumonie vorkommen, davon am häufigsten war 486 *Pneumonia, organism unspecified*, sowie zweimal 4846 *Pneumonia in aspergillosis*, während die anderen Codes nur jeweils einmal

vorkamen. Diese verschiedenen ICD-9 Codes beziehen sich auf exakt 11 Patienten. Die ICD-10 Codes gaben exakt zwei Ergebnisse für genau zwei Patienten. Während der Ausführung mussten die vorgegebenen SQL-Schemata jeweils um die dazugehörige Tabelle ergänzt werden, um die korrekte Relation aufzurufen. Beispielsweise *FROM mimic_hosp.diagnoses_icd* anstatt direkt *diagnoses_icd* Liste der ICD-Codes nach der Datenexploration im SQL-Listenschema:

- **ICD-9:** '4809', '48020', '48241', '48242', '4249', '4829', '4846', '486'
- **ICD-10:** 'J15211', 'J159'

Diese Listen werden während der folgenden Übungsaufgaben benutzt werden, um auf die spezifischen Patientenfälle mit Pneumonie zuzugreifen.

2.3. CITI-Kurs und MIMIC-IV Zugang

Das erste Modul des Kurses dreht sich um den *Belmont Report*, eine US-amerikanische Richtlinie zu ethischen Prinzipien für die Forschung am Menschen. Die drei zentralen ethischen Prinzipien sind hierbei: (1) Achtung der Menschenrechte, (2) Benefizienz (Risikenminimierung und Vorteilsmaximierung) und (3) Gerechtigkeit. Das zweite Modul erläutert die Entstehungsgeschichte dieser und anderer ethischer Richtlinien für die Forschung am Menschen.

Darauffolgend wird auf IRBs (institutional review boards) eingegangen, deren Funktionsweise und Aufbau erklärt, und in welchen Fällen ein IRB eingesetzt werden muss. Das vierte Modul behandelt *record-based research*, es wird unter anderem erläutert, dass auch hier abgewogen werden muss zwischen persönlichen Interessen (wie Datenschutz und Vertraulichkeit) und dem Lösen von Forschungsfragen.

Im Folgenden werden Arbeitsweisen und Vorteile der Gen(om)forschung vorgestellt, aber auch damit verbundene ethische Probleme angesprochen. In Modul 6 werden verschiedene vulnerable Gruppen vorgestellt und ethische Implikationen (auch mit Hinsicht auf die Prinzipien des *Belmont Report*) diskutiert. Darauffolgend werden dann Regularien wie der HIPAA-Standard vorgestellt, die sensible Gesundheitsdaten schützen sollen.

Im achten Modul werden verschiedene Arten von Interessenskonflikten vorgestellt (z.B. monetäre) und wie diese reguliert werden können (z.B. COI committees).

Im folgenden ist der aktuelle Status der Absolvierung des CITI-Kurses und des Zugangs zum Datensatz zum Zeitpunkt der Abgabe dokumentiert:

	CITI-Kurs	MIMIC-IV Zugang	Datum Zugang
Simon Bosse	Absolviert	Vorhanden	29.04.2025
Anna Göing	Nicht absolviert	Nein	-
Jonas Jakob	Nicht absolviert	Nein	-
Vyvy Tran Ngoc	Absolviert	Abgelehnt	-

A. Anhang

A.1. CITI-Zertifikate aller Gruppenmitglieder

