

HUMBOLDT-UNIVERSITÄT ZU BERLIN  
MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT  
INSTITUT FÜR INFORMATIK

# **Analyse von Visitentexten mittels maschinellern Lernen zur Optimierung der Leitlinienadhärenz auf einer Intensivstation**

Bachelorarbeit

zur Erlangung des akademischen Grades  
Bachelor of Science (B. Sc.)

eingereicht von: Martin Hoffmann

geboren am: 06.09.1998

geboren in: Berlin

Gutachter: Prof. Dr. med. Dr. rer. nat. Felix Balzer  
Dr. med. Fridtjof Schiefenhövel

eingereicht am: .....

verteidigt am: .....

## **Zusammenfassung**

[1]

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>3</b>
1.1	Datenerfassung auf Intensivstationen . . . . .	3
1.1.1	medizinische Scores . . . . .	3
1.1.2	Ziel der Arbeit . . . . .	3
1.2	Maschinelles Lernen . . . . .	3
1.2.1	Supervised Learning . . . . .	4
1.2.2	Regression vs Klassifizierung . . . . .	4
<b>2</b>	<b>Übersicht über die Daten</b>	<b>6</b>
2.1	Datenerfassung an der Charité . . . . .	6
2.2	Übersicht über die vorliegenden Daten . . . . .	6
2.2.1	Exemplarische Vorstellung eines Patienten . . . . .	8
2.2.2	Generierung von Schlüssel-Wert-Paaren . . . . .	8
2.2.3	Genauigkeit der erfassten Daten . . . . .	8
<b>3</b>	<b>Vorgehensweise</b>	<b>10</b>
<b>4</b>	<b>Fazit</b>	<b>11</b>
4.1	Auswertung der Ergebnisse . . . . .	11
4.2	Bewertung der Leitlinienadhärenz auf den Intensivstationen der Charité . . . . .	11
4.3	Blick in die Zukunft . . . . .	11

# Kapitel 1

## Einführung

### 1.1 Datenerfassung auf Intensivstationen

blabla.

#### 1.1.1 medizinische Scores

Was sind medizinische Scores?

#### 1.1.2 Ziel der Arbeit

Das Ziel der vorliegenden Arbeit ist es, mit Hilfe von maschinellem Lernen ein statistisches Modell zu entwickeln, um anhand von Freitexten medizinische Scores möglichst akkurat vorherzusagen.

### 1.2 Maschinelles Lernen

Der Begriff Maschinelles Lernen bezeichnet einen modernen Ansatz in der Arbeit an künstlicher Intelligenz. (cite).

hat sich im Zeitalter von Big Data und leistungsstarken Rechnern zu einem der Hauptforschungspunkte der Informatik entwickelt (cite).

(et al) definiert den Vorgang maschinellen Lernens folgendermaßen:

A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .

Im Allgemeinen werden also Algorithmen, die aus großen Datensätzen „lernen“ und damit Vorhersagen über unbekannte Daten machen, als maschinelles Lernen bezeichnet.

### 1.2.1 Supervised Learning

Neben vielen weiteren Methoden stellt Supervised Learning einen zentralen Ansatz im Gebiet des maschinellen Lernen dar.

### 1.2.2 Regression vs Klassifizierung

Probleme aus dem Bereich des supervised machine learnings lassen sich im Allgemeinen in eine von zwei Kategorien einordnen:

Klassifizierung bezeichnet den Prozess, bei dem ein Datensatz einer oder mehreren Klassen aus einer endlichen Liste möglicher Klassen zugeordnet wird.

Dieser Ansatz findet beispielsweise bei der automatischen Kategorisierung von E-Mails (Spam oder nicht Spam) oder bei der Erkennung von handschriftlichen Texten (welches Symbol aus einem gegebenen Alphabet ist dargestellt?) Anwendung (CITE). Da es sich bei den betrachteten medizinischen Scores um diskrete, ganzzahlige Werte aus einem endlichen Wertebereich handelt, liegt auch hier die Anwendung eines Klassifizierungs-Verfahrens nahe.

Würde man aber die verschiedenen möglichen Werte eines Scores als separate und voneinander unabhängige Klassen betrachten, so ginge eine wichtige Information über deren Anordnung verloren. Im mathematischen Sinne stellen alle hier betrachteten medizinische Scores Totalordnungen dar. Sie erfüllen also die Anforderungen der Reflexivität, Antisymmetrie, Transitivität und Totalität. Bezeichne  $M$  die Menge aller möglichen Werte eines beliebigen medizinischen Scores. Folglich gilt für alle  $a, b, c \in M$ :

$$\begin{array}{ll}
a \leq a & \text{(Reflexivität)} \\
a \leq b \wedge b \leq a \Rightarrow a = b & \text{(Antisymmetrie)} \\
a \leq b \wedge b \leq c \Rightarrow a \leq c & \text{(Transitivität)} \\
a \leq b \vee b \leq a & \text{(Totalität)}
\end{array}$$

Damit lassen sich die verschiedenen Scores vergleichen und in ein Verhältnis setzen. So ist ein RASS-Wert<sup>1</sup> von -4 (Tief sediert) beispielsweise deutlich näher an -3 (mäßig sediert) als an +1 (unruhig). Bei gängigen Verfahren zur Klassifizierung ginge diese Information verloren, da bei Kenngrößen zur Bewertung solcher Modelle nur betrachtet werden kann, ob ein gegebener Eingabetext der richtigen Kategorie (dem richtigen Score) zugeordnet wurde oder nicht. (Da viele der Scores bei der Vergabe zumindest teilweise auch auf dem persönlichen Ermessen des behandelnden Arztes/der behandelnden Ärztin beruht wäre selbst für menschliche Experten eine genaue Zuordnung eines Textes zu einer Punktzahl problematisch.)

Bei dieser Arbeit habe ich mich demnach dafür entschieden, die Vergabe von Scores anhand von Eingabetexten als klassisches Regressionsproblem zu betrachten, und die Ausgaben der Modelle im Zweifelsfall auf den nächstmöglichen ganzzahligen Wert zu runden. Dieser Ansatz fand auch bei früheren Arbeiten, die sich mit ähnlichen Fragestellungen befassten, Anwendung (CITE). Die Abweichung des vorhergesagten Werts eines Modells von dem nächst möglichen Wert stellt somit sogar einen rudimentären Ansatz zur Bewertung der Konfidenz bei der Vergabe einzelner Werte dar.

---

<sup>1</sup>Richmond Agitation-Sedation Scale

# Kapitel 2

## Übersicht über die Daten

### 2.1 Datenerfassung an der Charité

information über copra (fridtjof ausquetschen). bla bla bla

Die daten wurden vor dem Export pseudonymisiert. Nach dem Export lagen die Daten in Form mehrerer Datendateien vor. `patienten.csv` enthält Metainformationen über die betrachteten Patienten. Separat gibt `delir.csv` für jeden Patienten an, ob für diesen während seines Aufenthalts ein Delir diagnostiziert wurde.

### 2.2 Übersicht über die vorliegenden Daten

Der mir für diese Arbeit zur Verfügung stehende Datensatz enthält Informationen über insgesamt 1357 Patienten-Aufenthalte auf den Intensivstationen der Charité Berlin, die allesamt im Zeitraum von Juni 2019 bis Juni 2020 stattfanden. Jeder Aufenthalt wird über eine numerische, inkrementell aufsteigende Nummer<sup>1</sup> eindeutig identifiziert. Hierbei gilt es zu beachten, dass ein Patient bzw. eine Patientin, der/die im Laufe seiner/ihrer Behandlung mehrere Male auf eine Intensivstation verlegt wird, für jeden separaten Aufenthalt eine neue Identifikationsnummer erhält. Für jeden Patient/für jede Patientin

---

<sup>1</sup>in den Datensätzen als `n.ID` bezeichnet

liegen Informationen über das Geschlecht, BMI<sup>2</sup>, das Alter zum Zeitpunkt der Aufnahme und ob während des Aufenthalts ein Delir<sup>3</sup> diagnostiziert wurde, vor. Weiterhin ist die Dauer des Aufenthalts auf der Intensivstation vermerkt, sowie ob der Patient/die Patientin während seines/ihres Aufenthalts verstorben ist. Die Mortalität während des Aufenthalts lag bei etwa 17% ( $n = 225$ ). Die beobachteten Aufenthalte hatten eine Dauer zwischen wenigen Stunden und mehreren Monaten. Die mittlere Aufenthaltsdauer während des Beobachtungszeitraums liegt bei x Tagen, der Median beträgt y Tage. Etwas mehr als ein Viertel der Patienten ( $n = 375$ ) wurden für eine Woche oder länger auf der Intensivstation behandelt.

Für jeden Patient/jede Patientin werden während der Dauer seines/ihres Aufenthalts medizinische Scores erfasst sowie Visitentexte geschrieben. Die Visitentexte werden digital erfasst und liegen im Unicode-Zeichensatz vor, folgen allerdings im Allgemeinen keinem einheitlichen Muster. Es handelt sich also um Freitexte, und es liegt im Ermessen der behandelnden Ärzte bzw. Pflegekräfte, einen aussagekräftigen Text zu formulieren. Ebenso können sich Faktoren wie Zeitdruck oder Stress auf Umfang und Genauigkeit der eingetragenen Texte auswirken(CITE).

Bei den erfassten Scores handelt es sich um medizinische Bewertungssysteme, bei denen die Verfassung des betrachteten Patienten anhand klar definierter Regeln anhand einer Punktzahl bewertet wird(CITE). Häufig beziehen sich die Scores dabei nur auf einen kleinen Teil der Gesamtverfassung des Patienten, beispielsweise auf den Grad der Sedierung oder auf Schmerzempfinden, sofern dieses nicht vom Patienten selber mitgeteilt werden kann. Tabelle 1 gibt einen Überblick über alle Scores und Freitexte, die in dem gegebenen Datensatz erfasst wurden. Bei der VarID handelt es sich um eine Zahl, mit der jede Art von auf der Intensivstation erfasstem Wert bzw. eingetragendem Text repräsentiert und eindeutig identifiziert wird.<sup>4</sup>

---

<sup>2</sup>Body-Mass-Index

<sup>3</sup>F05.\* gemäß ICD-10

<sup>4</sup>Fridtjof fragen ob das stimmt



### **2.2.1 Exemplarische Vorstellung eines Patienten**

Aufenthalt des Patienten hier detailliert beschreiben, und seinen Scatterplot einfügen (aber noch ohne Pfeile).

### **2.2.2 Generierung von Schlüssel-Wert-Paaren**

Eine besondere Herausforderung stellte dar, dass die verschiedenen Werte und Texte zu unterschiedlichen Zeiten und unabhängig voneinander eingetragen werden. Die Informationen über Zeitpunkt und Art der Eintragung sowie der eigentliche Wert liegen jeweils als Tripel in Form mehrerer Log-Dateien vor. Bei den Visitentexten entspricht der eingetragene Text dem Wert der Eintragung.

Um ein Machine-Learning Modell aus dem Bereich des supervised learnings zu trainieren ist eine hohe Anzahl von Trainingspaaren notwendig. Jedes Wertepaar enthält einen Text sowie einen medizinischen Score, der den in dem Text angegebenen Informationen über die Verfassung des Patient/die Patientin möglichst genau entspricht. Zusammen bilden diese Paare die Grundlage der Modelle, selbstständig noch nicht gesehene Texte bewerten zu können. Aufgrund der zeitlichen Unterschiede zwischen den Eintragungen von Texten und Scores erwies sich allerdings ebenjene Zuordnung zueinander als nicht trivial.

### **2.2.3 Genauigkeit der erfassten Daten**

Hier beschreiben, dass viele Texte nicht wirklich den Werten entsprechen. Das mache es schwerer, Modelle zu trainieren, und muss berücksichtigt werden. Gründe:

1. Zeitdruck, Stress bei Ärzten, kann vorkommen dass sie einfach Wert vom letzten mal kopieren
2. Werte können sich innerhalb von Minuten ändern (z.B. RASS von 0 auf -5)
3. Weitere Gründe?

Diese Gründe sind aber nicht weiter Beobachtungsgegenstand der vorliegenden Arbeit. Dennoch muss das bei der Konzeption, Entwicklung und Bewertung beachtet werden, weil sie ohne weitere Maßnahmen möglicherweise eine obere Schranke für die Performance der Modelle darstellen.

## Kapitel 3

### Vorgehensweise

# Kapitel 4

## Fazit

### 4.1 Auswertung der Ergebnisse

...

### 4.2 Bewertung der Leitlinienadhärenz auf den Intensivstationen der Charité

...

### 4.3 Blick in die Zukunft

...

# Literaturverzeichnis