

Machine Learning in der Berufsunfähigkeitsversicherung?

Eine Analyse von Risikofaktoren

Felix Müller und Antonia Schöning

30.11.2021



NÜRNBERGER
VERSICHERUNG

Data Science Challenge

Unser Team (2020 ein Team der NÜRNBERGER Versicherung)



Christian Eckert



Daniela Giesinger



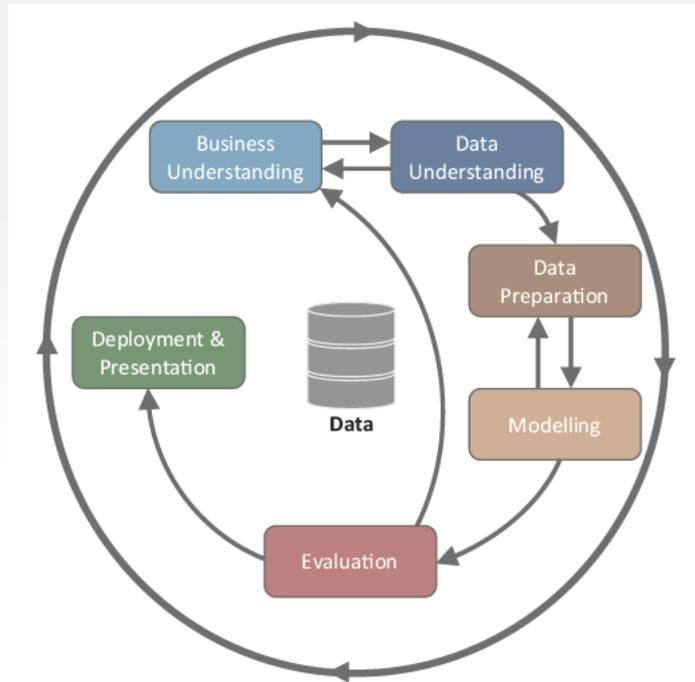
Felix Müller



Antonia Schöning

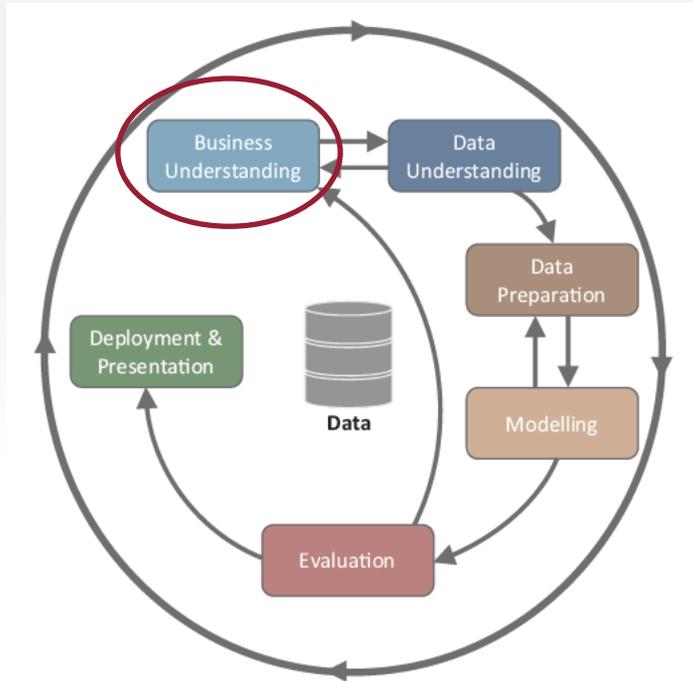
Unser Vorgehen

CRISP-DM = Cross Industry Standard Process for Data Mining



Quelle: Schnattinger (2020)

Business Understanding



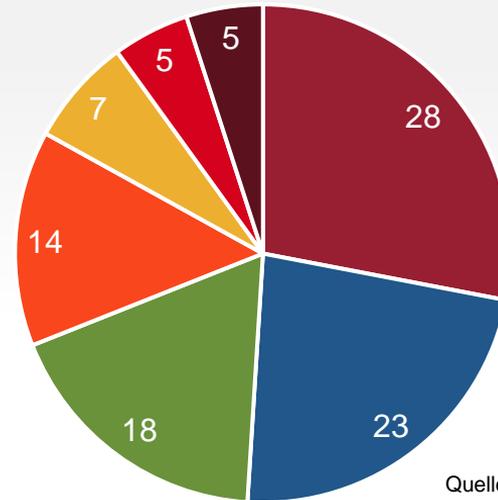
Quelle: Schnattinger (2020)

Business Understanding

Unsere Motivation

Krankheiten, die zur Anerkennung von BU führen

- Psychische Krankheiten und Verhaltensstörungen
- Krankheiten des Muskel-Skelett-Systems und Bindegewebes
- Bösartige Neubildungen
- Sonstige
- Krankheiten des Kreislaufsystems
- Krankheiten des Nervensystems
- Unfälle



Quelle: Franke und Bornberg (2021)

Business Understanding

Unsere Motivation

Psychische Krankheiten sind ein sehr häufiger Grund für Berufsunfähigkeit

Forschungsfrage.

Welche Rahmenbedingungen begünstigen psychische Krankheiten, die dann zu einer Berufsunfähigkeit führen?

D. h. was sind hier relevante (und auch: was sind hier nicht relevante) Einflussfaktoren auf die Berufsunfähigkeit und welcher Zusammenhang besteht zur Berufsunfähigkeit?

Business Understanding

Unsere Motivation

Warum interessiert das Versicherungsunternehmen?

Tarifierung

Einflussfaktoren/Risikofaktoren sollten Einfluss auf die Prämie haben

Antragsprozess

Beschränkung auf wesentliche und relevante Einflussfaktoren für optimale Customer-Experience beim Antragsprozess

Präventiv-
maßnahmen

Präventivmaßnahmen, die die Einflussfaktoren positiv beeinflussen, sind besonders erfolgsversprechend

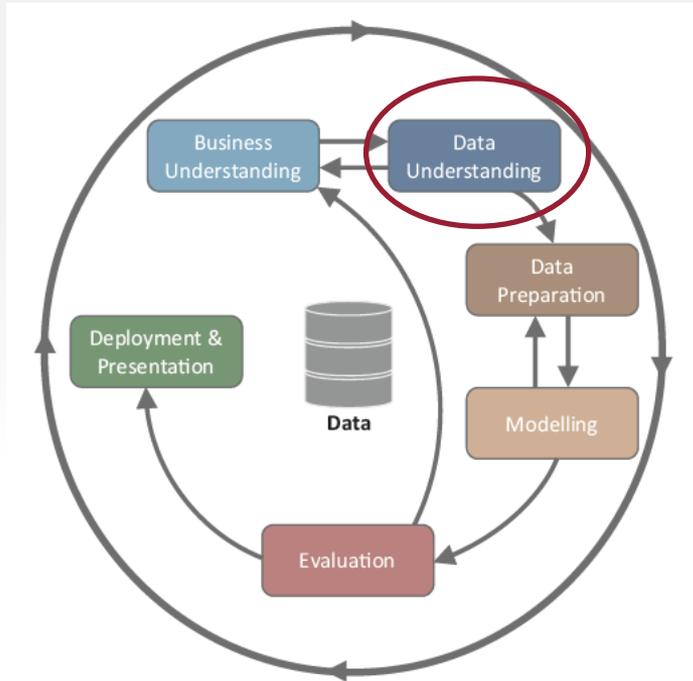
Business Understanding

Unsere Idee

Was ist dabei neu?

Wir wollen bewusst externe Daten beziehen, um auch Faktoren berücksichtigen zu können, die von VU ggf. bislang nicht erhoben wurden.

Data Understanding



Quelle: Schnattinger (2020)

Data Understanding

Welche Daten nutzen wir?



National Health and Nutrition Examination Survey

Wir nutzen Daten des National Health and Nutrition Examination Survey (NHANES).

- NHANES stellt ein für die USA repräsentatives Sample im Zeitraum von 1999 bis 2018 dar.
- Alle zwei Jahre werden äußerst umfangreiche Informationen über den Gesundheitszustand (insbesondere Psyche), die Lebensumstände und das Arbeitsverhältnis (inkl. Berufsunfähigkeit) von rund 5.000 Personen erhoben.

Data Understanding

Wie können wir damit unsere Forschungsfrage beantworten?

Wir untersuchen, wie die **abhängige Variable** „**Berufsunfähigkeit**“ („Limitations keeping you from working“) durch in den Daten vorhandene **unabhängige Variable** erklärt werden kann.

Data Understanding

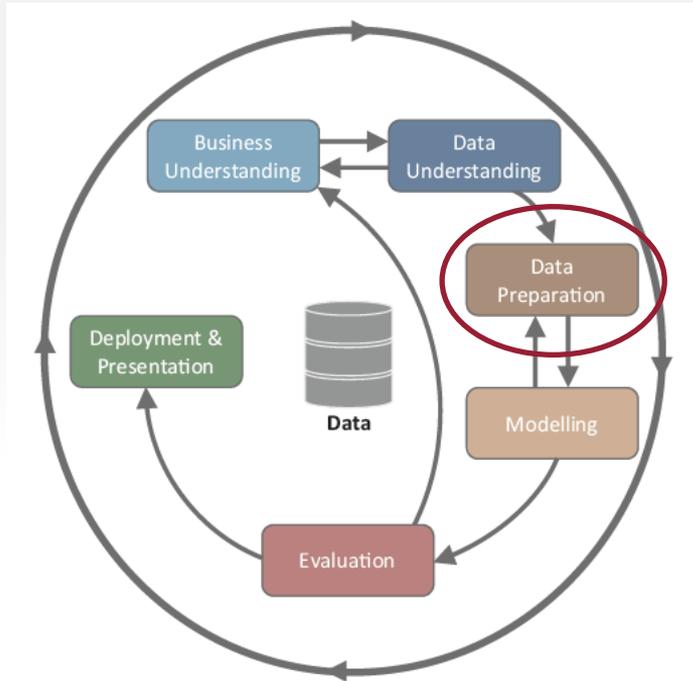
Wie können wir damit unsere Forschungsfrage beantworten?

Allerdings besitzen unsere Ergebnisse eine begrenzte Aussagekraft.

- Angaben werden von den Befragten gemacht und stellen eine Selbsteinschätzung dar.
- Es muss keine ärztlich festgestellte Berufsunfähigkeit vorliegen, wie sie für eine deutsche Versicherung notwendig wäre.

Unsere Ergebnisse geben daher nur Hinweise auf mögliche Risikofaktoren, sind aber nicht uneingeschränkt übertragbar!

Data Preparation



Quelle: Schnattinger (2020)

Data Preparation

Aufbereitung unserer Daten

U. a.

Umgang mit fehlenden Werten

One-Hot-Kodierung für kategoriale Variable

Min-Max-Normalisierung

Korrelationsanalyse



31 Variable und 21.555 Datensätze

Data Preparation

Aufbereitung unserer Daten – Umgang mit fehlenden Daten

Beobachtung.

In unserem Sample weisen einige Variable einen hohen Anteil fehlender Werte auf.

Maßnahmen.

- Ergänzen der fehlenden Daten durch logisches Schließen
- Ergänzen der fehlenden Daten durch Schätzungen

- Datensätze entfernen

Möglichst wenige Datensätze entfernen, um eine breite Datenbasis zu erhalten!

- Variable löschen

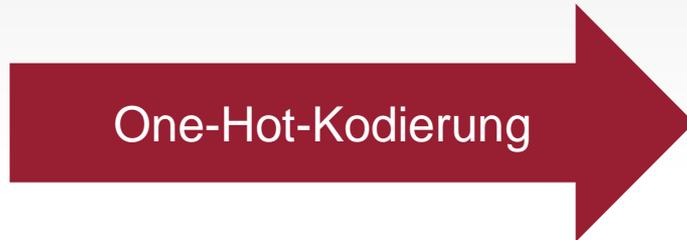
Möglichst wenig Variable löschen, um die Modellkomplexität nicht zu stark zu reduzieren!

Data Preparation

Aufbereitung unserer Daten – One-Hot-Kodierung

Beobachtung.

In unserem Sample sind auch kategoriale Variable, die wir für unsere Analysen nutzbar machen wollen.

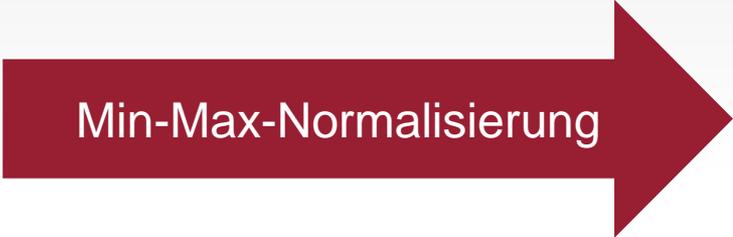


Data Preparation

Aufbereitung unserer Daten – Min-Max-Normalisierung

Beobachtung.

Die Wertebereiche unserer Variablen sind sehr heterogen und Algorithmen reagieren darauf teilweise sehr empfindlich.



Min-Max-Normalisierung

Data Preparation

Aufbereitung unserer Daten – Korrelationsanalyse

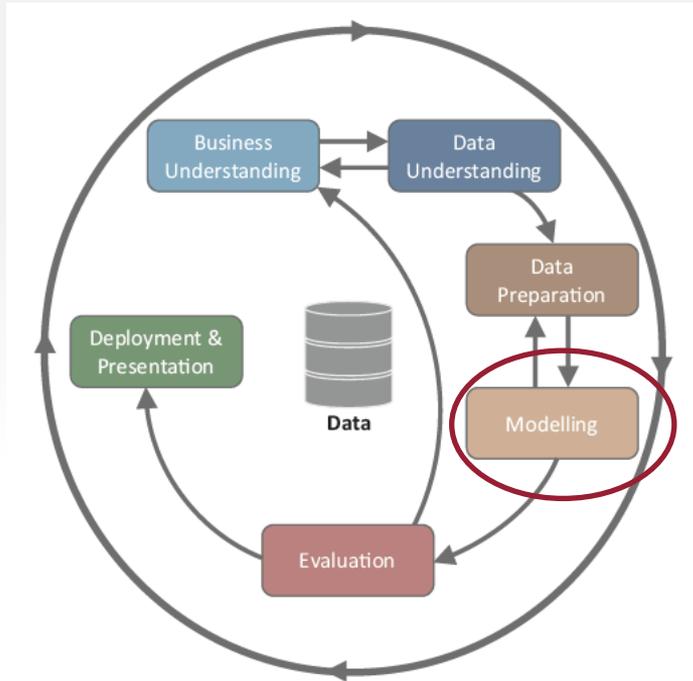
Motivation.

Weisen zwei oder mehr erklärende Variable eine sehr starke Korrelation auf, führt das zu Problemen in der Modellierung und Evaluation. (Multikollinearität)

Maßnahmen.

- Analyse der Korrelationen zwischen den erklärenden Variablen
- Entferne Variable, um die Multikollinearität zu reduzieren

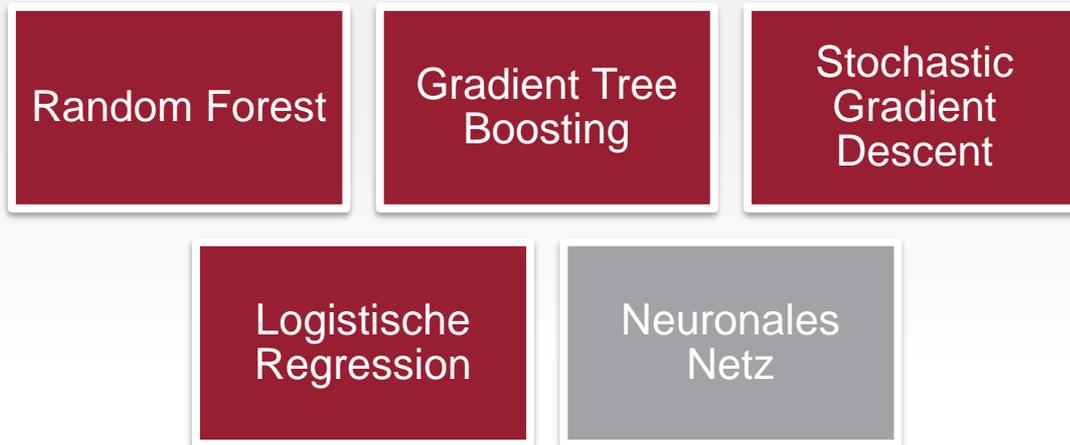
Modeling



Quelle: Schnattinger (2020)

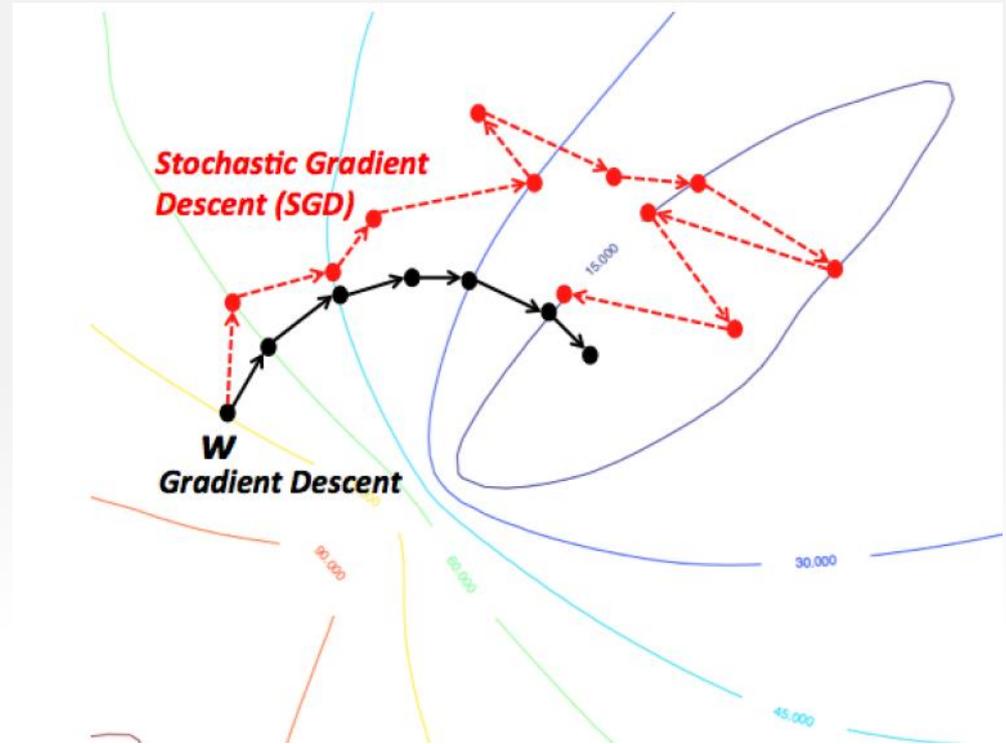
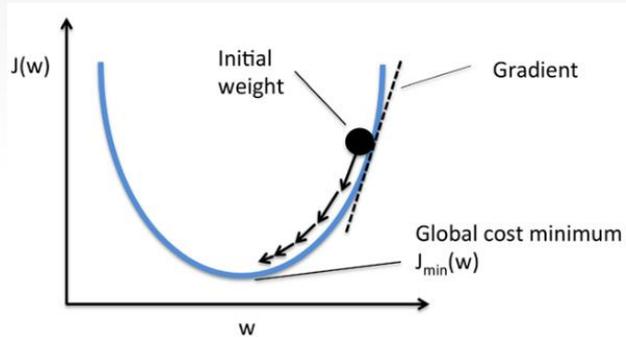
Modeling

Angewendete Modelle



Modeling

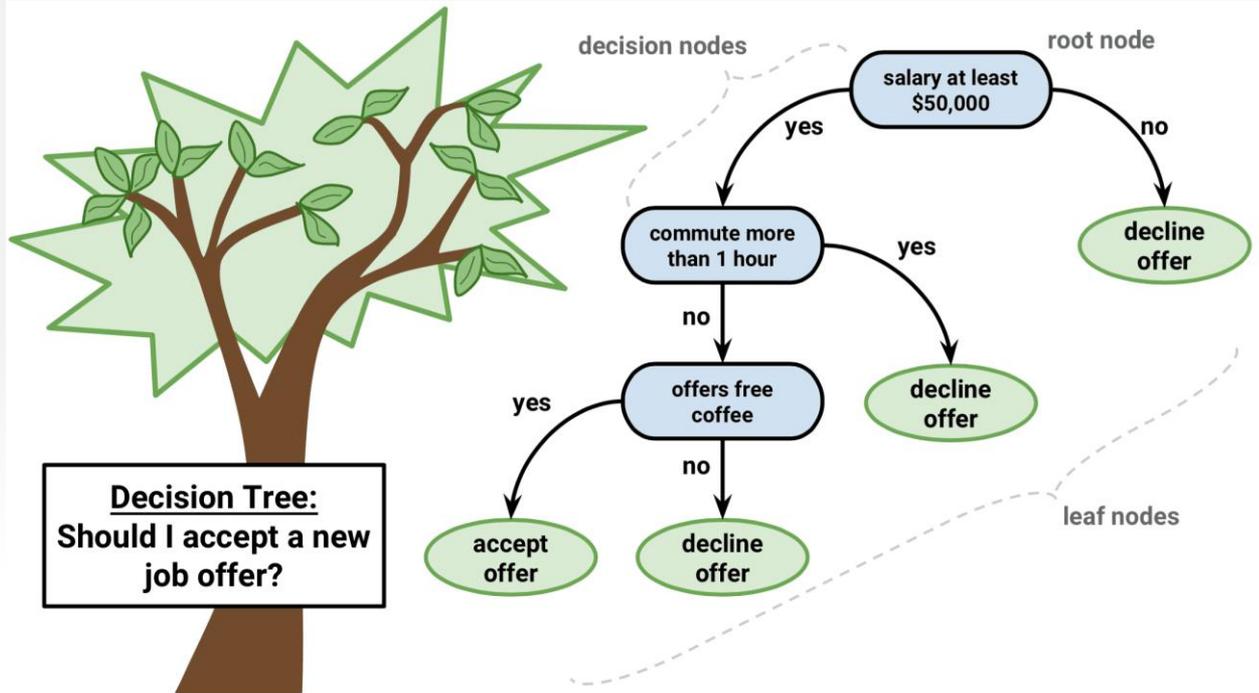
Stochastic Gradient Descent



Quelle: [scikit-learn: Batch gradient descent versus stochastic gradient descent - 2020 \(bogotobogo.com\)](https://scikit-learn.org/stable/auto_examples/optimization/plot_batch_vs_stochastic.html)

Modeling

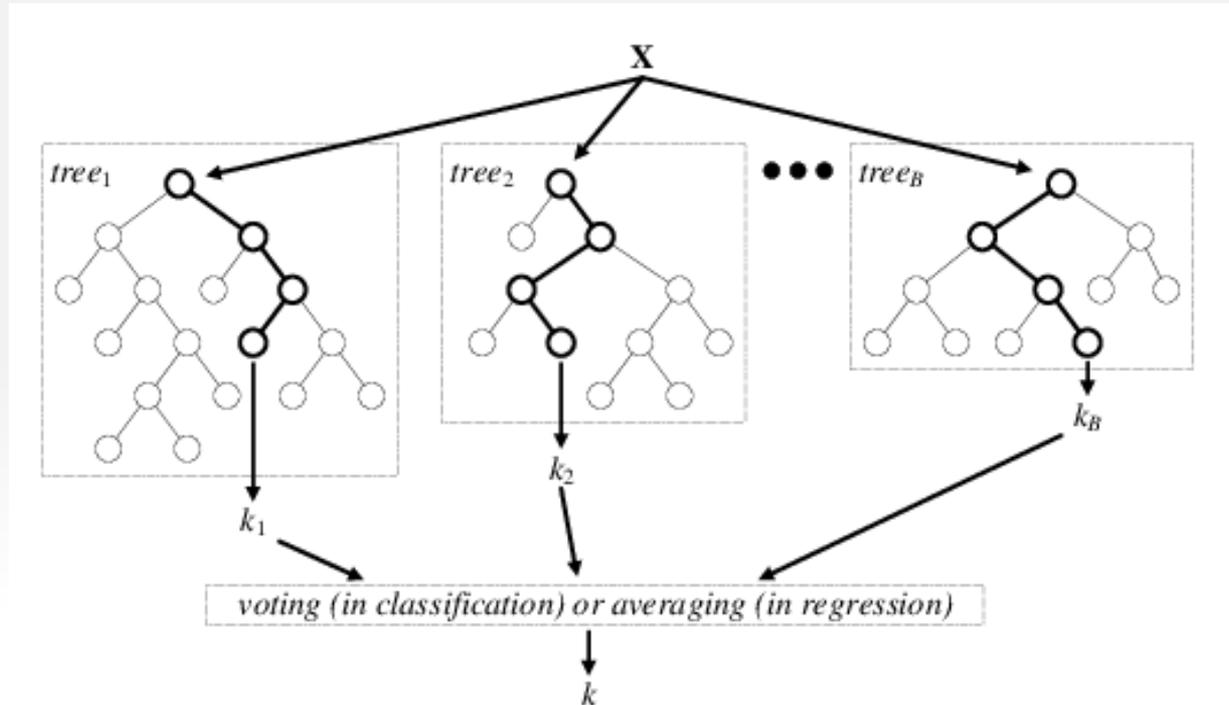
Decision Tree



Quelle: <https://www.datacamp.com/community/tutorials/decision-trees-R>

Modeling

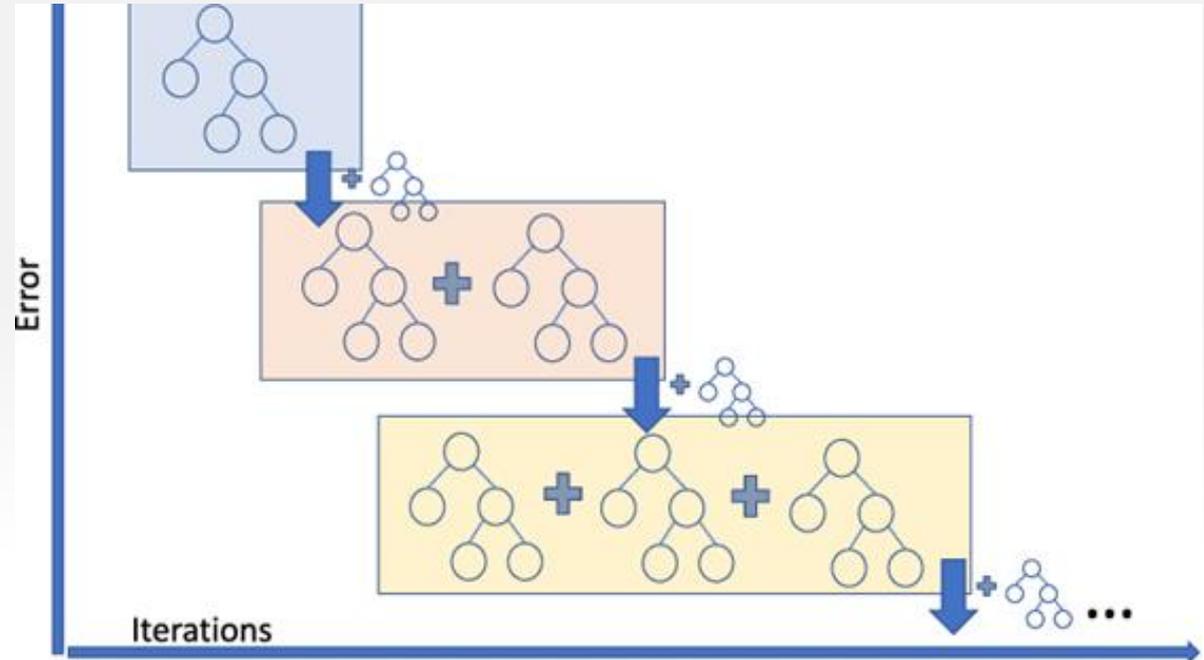
Random Forest



Quelle: https://www.researchgate.net/figure/Architecture-of-the-random-forest-model_fig1_301638643

Modeling

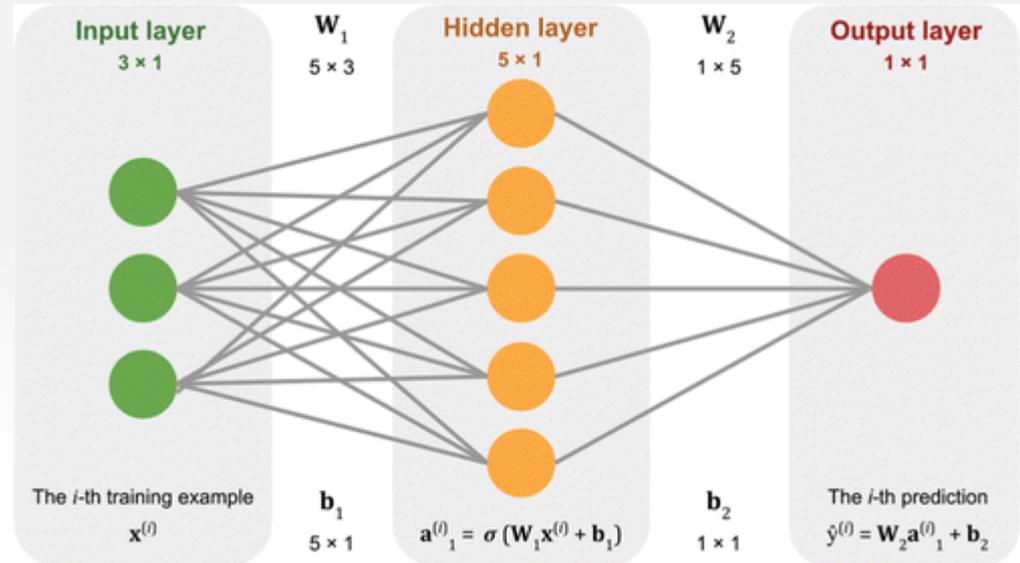
Gradient
Boosting



Quelle: <https://medium.com/analytics-vidhya/what-is-gradient-boosting-how-is-it-different-from-ada-boost-2d5ff5767cb2>

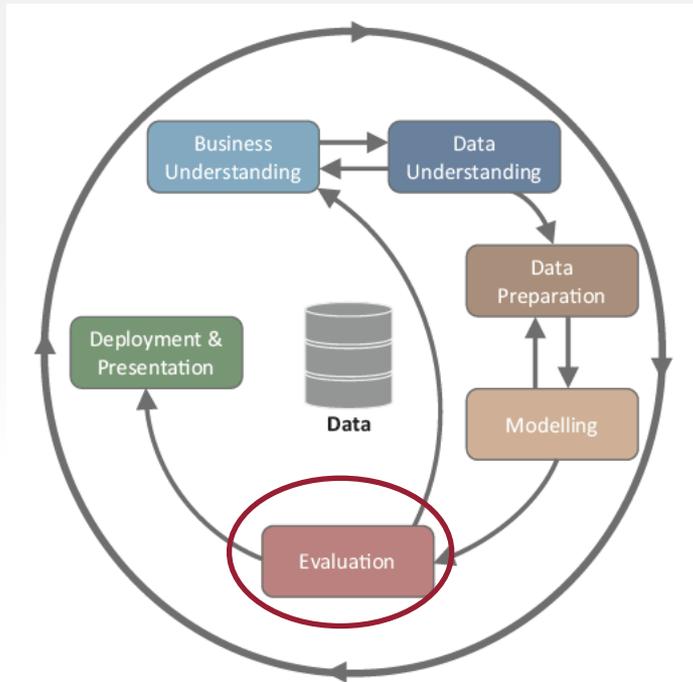
Modeling

Neuronales
Netz



Quelle: [Neural networks – SEG Wiki](#)

Evaluation



Quelle: Schnattinger (2020)

Modell-Evaluation

Überblick

- Da das Merkmal BU zwei Merkmalswerte hat (BU, nicht BU), treten bei Klassifizierungen vier verschiedene Fälle auf:

Korrekt negativ	Falsch positiv
Falsch negativ	Korrekt positiv

Modell-Evaluation

Überblick

Confusion Matrix

		Vorhersage	
		Negativ	Positiv
Tatsächliche Werte	Negativ	Korrekt negativ	Falsch positiv
	Positiv	Falsch negativ	Korrekt positiv

Modell-Evaluation

Überblick

Confusion Matrix

		Vorhersage	
		Negativ	Positiv
Tatsächliche Werte	Negativ	Korrekt negativ	Falsch positiv
	Positiv	Falsch negativ	Korrekt positiv

Klassische Kennzahlen, um die Prognosefähigkeit des Modells zu analysieren:

- Treffergenauigkeit
- Genauigkeit
- Sensitivität
- Ausfallrate

Modell-Evaluation

Überblick

Confusion Matrix

		Vorhersage		
		Negativ	Positiv	
Tatsächliche Werte	Negativ	Korrekt negativ	Falsch positiv	Ausfallrate
	Positiv	Falsch negativ	Korrekt positiv	Sensitivität
		Genauigkeit		Treffergenauigkeit

Klassische Kennzahlen, um die Prognosefähigkeit des Modells zu analysieren:

- Treffergenauigkeit
- Genauigkeit
- Sensitivität
- Ausfallrate

Modell-Evaluation

Kennzahlen

F-beta Maß (in unserem Modell: $\beta = 10$)

Formel 1

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{Genauigkeit} \cdot \text{Sensitivität}}{(\beta^2 \cdot \text{Genauigkeit}) + \text{Sensitivität}}$$

Modell-Evaluation

Benchmark

- Für einen Benchmark wird ein naiver Schätzer betrachtet
- Dieser Schätzer prognostiziert als Ergebnis immer „Gesund“
- Im Test-Sample befinden sich 6.467 Datensätze
 - davon sind 5.626 gesund
 - 841 ‚BU‘
- Der Naive Schätzer erreicht eine Treffergenauigkeit (Accuracy) von $\frac{5.626}{6.467} \sim 87$ Prozent

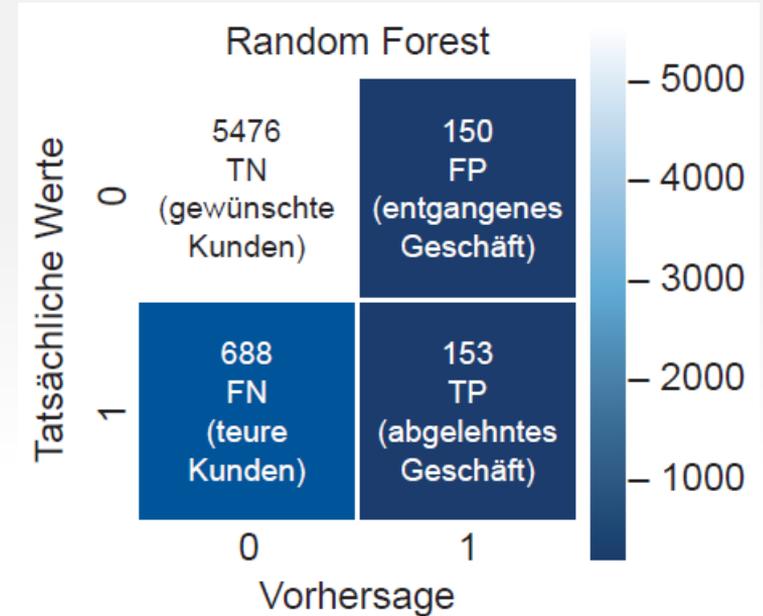
Modell-Evaluation

Confusion Matrix – Random Forest

Treffergenauigkeit

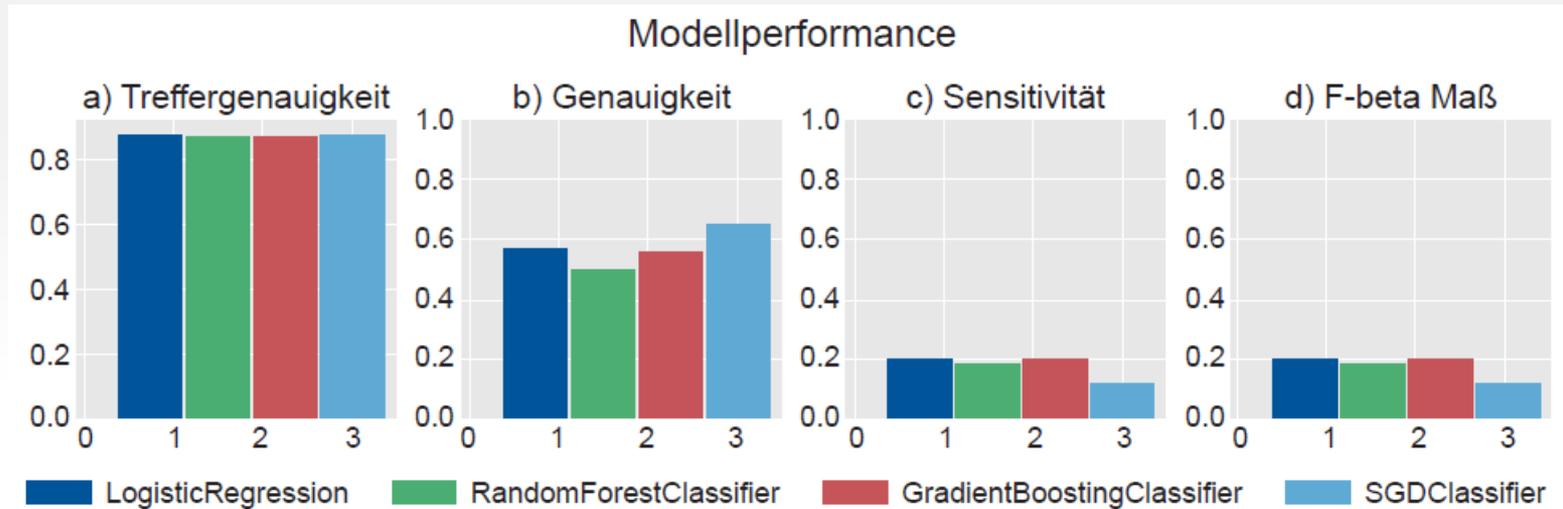
$$= \frac{153 + 5.476}{6.467}$$

~ 87 Prozent



Modell-Evaluation

Vergleich der verschiedenen Modelle



Wie können wir die Ergebnisse verbessern?

Idee: Resampling

- **Ausgangslage: Unbalancierte Daten**

- ~ 87 Prozent sind nicht BU

- ~ 13 Prozent sind BU

-  BU wird schlecht ‚gelernt‘

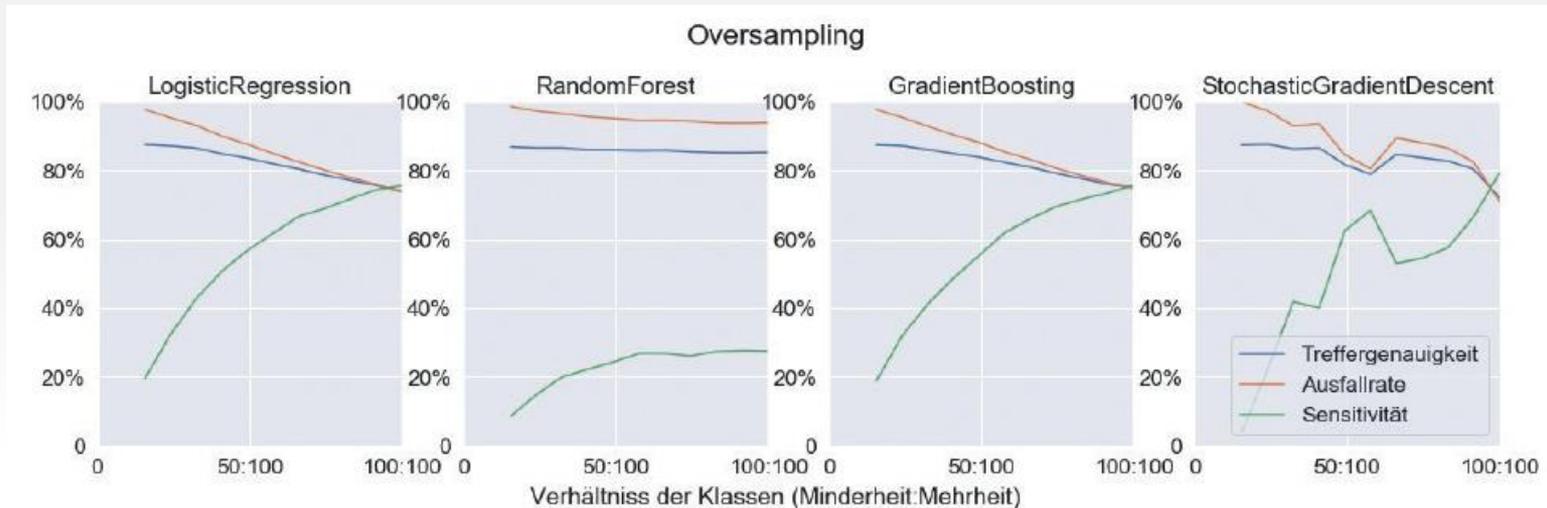
- **Idee: Klassengrößen annähern**

- Oversampling der kleineren Klasse

- Downsampling der größeren Klasse

Wie können wir die Ergebnisse verbessern?

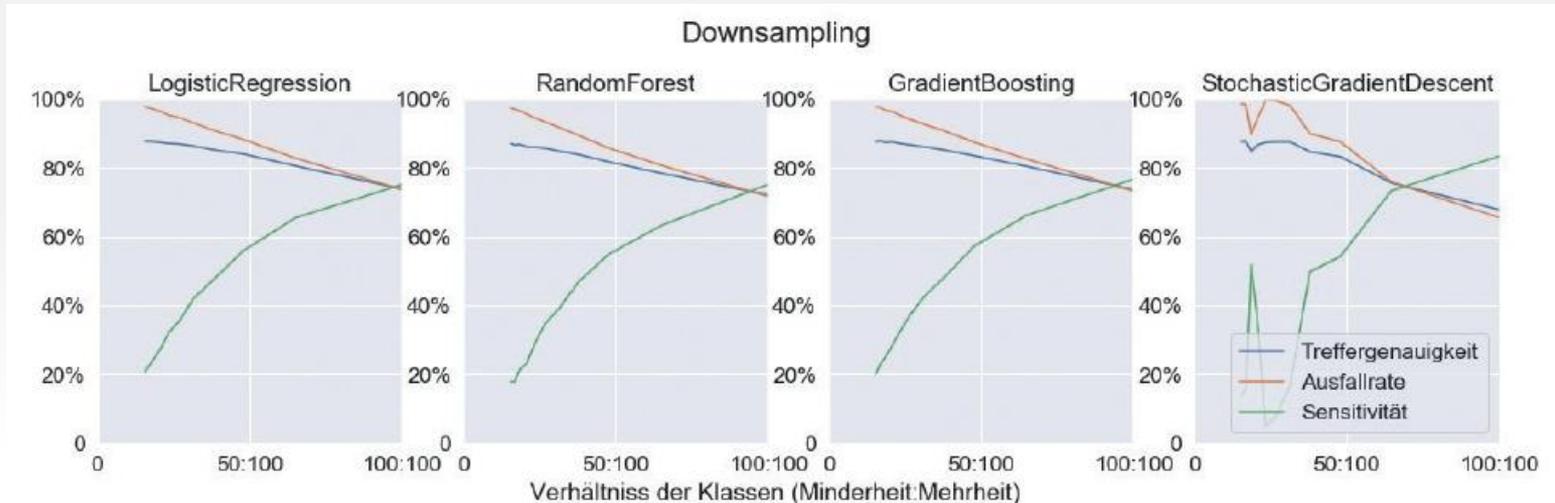
Oversampling



Auf der x-Achse sind verschiedene Verhältnisse der kleineren Klasse (BU-Fälle) zu der größeren Klasse (Gesunde) dargestellt bis zu einem Verhältnis von 100:100.

Wie können wir die Ergebnisse verbessern?

Downsampling



Auf der x-Achse sind verschiedene Verhältnisse der kleineren Klasse (BU-Fälle) zu der größeren Klasse (Gesunde) dargestellt bis zu einem Verhältnis von 100:100.

Wie können wir die Ergebnisse verbessern?

Downsampling

Nach dem Downsampling (Gradient Boosting)

Im fiktiven Antragsprozess:

- von 6467 Anträgen
- würden 2148 (33,2 %) abgelehnt (FP + TP),
- 4319 angenommen und
- 196 davon berufsunfähig werden

-> **BU-Quote von 4,5%**

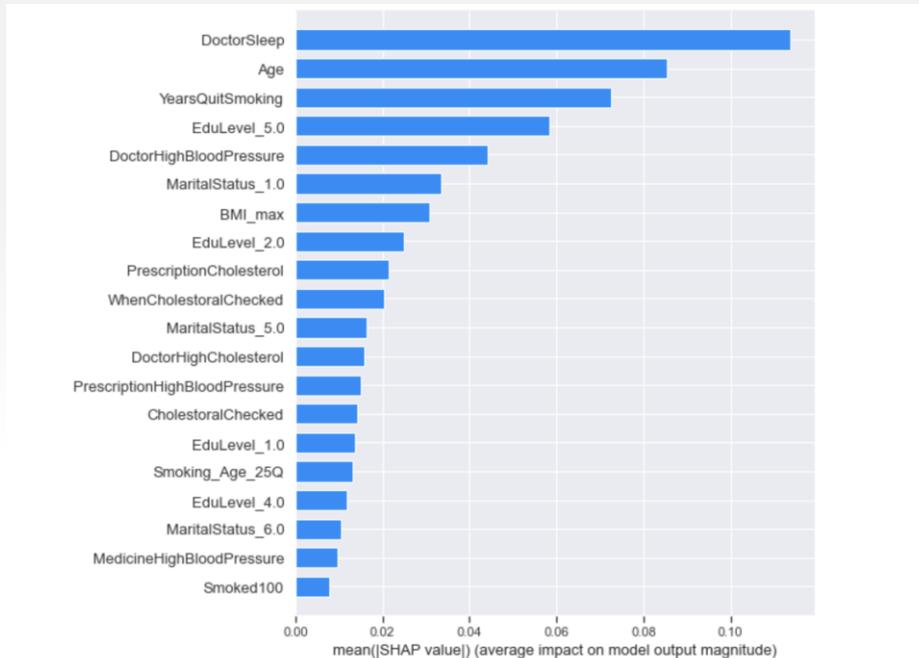
Erklärbarkeit der Modelle

Größte Einflussfaktoren – hier: Permutation Importance (Gradient Boosting)

Weight	Feature
0.0305 ± 0.0058	DoctorSleep
0.0121 ± 0.0054	Age
0.0057 ± 0.0042	YearsQuitSmoking
0.0055 ± 0.0018	Sleep hours
0.0054 ± 0.0035	MaritalStatus_1.0
0.0040 ± 0.0050	BMI_max
0.0029 ± 0.0029	MaritalStatus_5.0
0.0023 ± 0.0009	MedicineHighBloodPressure
0.0016 ± 0.0016	PeopleInHousehold
0.0012 ± 0.0010	EduLevel_4.0
0.0007 ± 0.0016	Smoking_Age_25Q
0.0006 ± 0.0033	EduLevel_5.0
0.0004 ± 0.0022	DoctorHighCholesterol
0.0002 ± 0.0003	MaritalStatus_2.0
0.0002 ± 0.0007	Smoked100
0.0001 ± 0.0006	Smoking_Age_75Q
0 ± 0.0000	MaritalStatus_4.0
-0.0001 ± 0.0004	MaritalStatus_3.0
-0.0001 ± 0.0005	Smoking_Age_100Q
-0.0001 ± 0.0042	DoctorHighBloodPressure
-0.0003 ± 0.0003	Smoking_Age_50Q
-0.0003 ± 0.0009	EduLevel_3.0

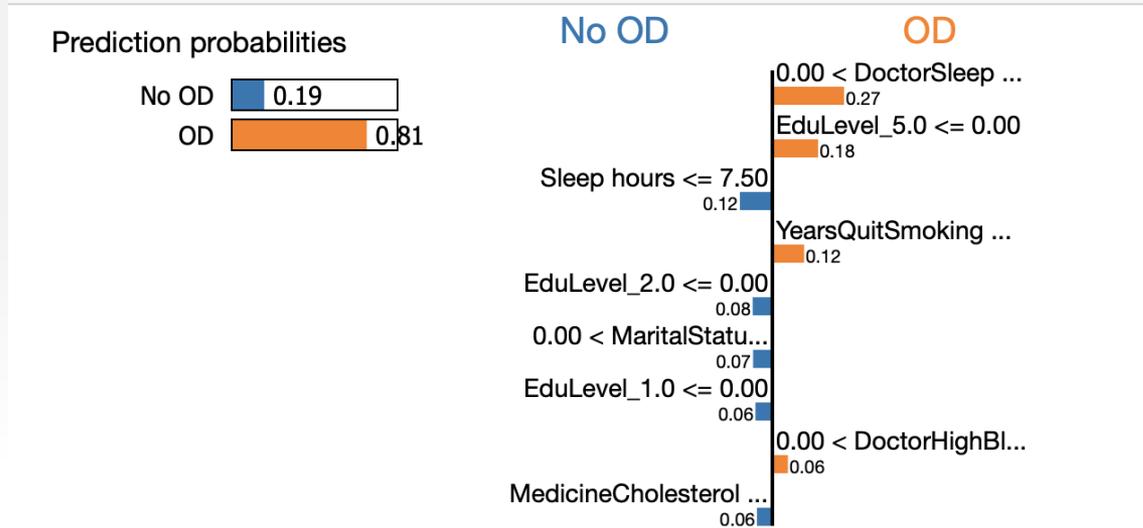
Erklärbarkeit der Modelle

Größte Einflussfaktoren – hier: SHAP-Werte (Gradient Boosting)



Erklärbarkeit der Modelle

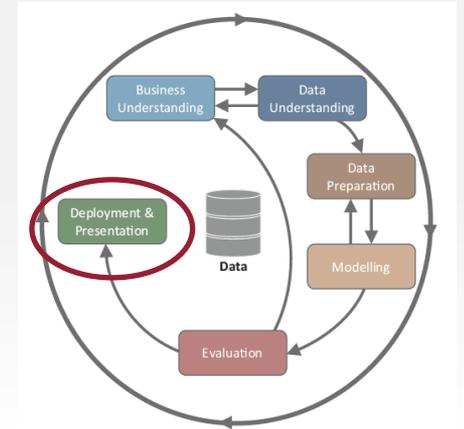
LIME (Local interpretable model-agnostic explanations)



Fazit

Anwendungsfall BU

- Laut unserer Analyse sind relevante Einflussfaktoren auf Berufsunfähigkeit (in diesem speziellen US-Datensatz) u. a.: BMI, Alter, Schlafprobleme
- Aber: Ungewisse Kausalitäten
- U. a. mögliche Berücksichtigung der Erkenntnisse bei Präventivmaßnahmen



Quelle: Schnattinger (2020)

Fazit

Methodik

- Datenbeschaffung und Datenaufbereitung stellen meist die Hauptarbeit dar
- Resampling-Verfahren können bei unausgewogenen Datensätzen zu einer deutlichen Verbesserung der Ergebnisse führen
- Ganzheitliche Evaluation der Modelle unter Berücksichtigung der Erklärbarkeit sehr bedeutend

Danke!

