

# Leistungsanalyse in der privaten Krankenversicherung: Methoden und Ergebnisse

Alexander Krauskopf, DAV, CERA  
Katrin Höllering, MSc  
Alexandra Pflumm, DAV



## Einleitung

Diese Arbeit zeigt die Einsatzmöglichkeiten des Machine Learning zum Zwecke der Leistungsanalyse im Bereich der deutschen Krankenversicherung. Dabei besteht das Ziel darin, die Höhe künftiger Leistungen für chronische Krankheiten möglichst gut vorherzusagen und drauf aufbauend in verschiedenen Bereichen Steuerungsmaßnahmen zu ergreifen.

Einführend geben wir einen Überblick über das deutsche Gesundheitssystem, im zweiten Abschnitt wird der gewählte Algorithmus erläutert, anschließend die Aufbereitung des Datensatzes mit der Kalibrierung des gewählten Algorithmus am Datensatz und abschließend die Auswertung der Ergebnisse.

## Übersicht über das deutsche Gesundheitswesen

Die Krankenversicherung bildet in Deutschland eine der Fünf Säulen des Sozialversicherungssystems. Das Prinzip des Sozialversicherungssystems ist der Schutz der Bürger vor existenzgefährdenden Risiken. Daraus resultiert, dass die Gesundheitsversorgung zum Recht der gesamten Bevölkerung gehört, welches sich in der Pflicht zur Krankenversicherung manifestiert.

Das deutsche Gesundheitssystem ist ein duales Krankenversicherungssystem und setzt sich aus der gesetzlichen Krankenversicherung (GKV) und der privaten Krankenversicherung (PKV) zusammen. Grundsätzlich ist jede in Deutschland lebende Person verpflichtet, sich in einer der aktuell ca. 110 gesetzlichen Krankenkassen zu versichern. Unter bestimmten Voraussetzungen (z.B., wenn das Einkommen die Jahresarbeitsentgeltgrenze überschreitet, für Selbstständige, für Beamte) endet die Versicherungspflicht. Diese Personen haben die Möglichkeit sich freiwillig in der GKV zu versichern oder auf einen substitutiven Vollversicherungstarif in der PKV auszuweichen.

Gesetzliche Krankenkassen erfüllen die Aufgaben, die ihnen vom Gesetz übertragen werden. Das Sozialversicherungsgesetz (SGB) regelt z.B. den Leistungskatalog und kann bei Bedarf von der Regierung geändert werden (z.B. um die Gesamtkosten für Schadensfälle zu senken). Finanziert wird die GKV als Umlagesystem, die Ausgaben im laufenden Jahr werden demnach durch Einnahmen desselben Jahres finanziert. Die Einnahmen

resultieren i. w. aus den Beiträgen der Versicherten; zusätzlich gibt es einen Steuerzuschuss. Jedes Mitglied zahlt einen Beitrag, der einem Prozentsatz des Einkommens entspricht (im Jahr 2020 durchschnittlich 15,7% pro Monat) und unabhängig vom individuellen Risiko ist. Davon wird die Hälfte vom Arbeitgeber bzw. der deutschen Rentenversicherung übernommen. Familienmitglieder ohne eigenes Einkommen sind über die Familienversicherung über ein beitragspflichtiges Familienmitglied mitversichert, ohne eigene Beiträge entrichten zu müssen. Alle Einnahmen fließen in einen Gesundheitsfond, aus dem die Krankenkassen gemäß ihrer Risikostruktur die Finanzmittel beziehen, die sie für die Leistungszahlungen der versicherten Behandlungen (einschließlich stationärer, ambulanter, zahnärztlicher Behandlung und Krankengeld) benötigen. Die Erstattung erfolgt nach dem Sachleistungsprinzip direkt zwischen den Krankenkassen und den Ärzten.

Im Gegensatz zur GKV erstreckt sich das Leistungsspektrum in der PKV auf verschiedene Tarife mit unterschiedlich zusammengesetzten Leistungsvarianten. Angefangen beim Basis-Schutz (der mit der GKV vergleichbar ist) bis zu einem Top-Schutz (der nahezu alle Behandlungen abdeckt), mit Preisabstufungen und verschiedenen Höhen eines Selbstbehaltes. Die Beiträge werden lebenslang kalkuliert und berücksichtigen das individuelle Risiko des Versicherten (insb. Alter und Gesundheitszustand) zu Beginn des Versicherungsverhältnisses. Die zu zahlenden Beiträge sind über die gesamte Lebenszeit des Versicherungsnehmers gleich (abgesehen von Beitragsanpassungen aufgrund medizinischer Inflation, Sterblichkeitsverbesserungen oder Änderungen des Rechnungszinses). Um einen konstanten Beitrag zu gewährleisten, muss in den jüngeren Jahren des Versicherten eine Altersreserve aufgebaut werden, die zur Finanzierung der höheren Leistungen im Alter verwendet wird.

Die PKV ist gegenüber der GKV flexibler im Hinblick auf die Erstattung neuer Behandlungsmethoden oder neuer Medikamente. Der Abrechnungsprozess verläuft anders als in der GKV nach dem Kostenerstattungsprinzip. Die Behandlungs- und Medikamentenkosten werden dem Patienten in Rechnung gestellt, der die Kosten zunächst selbst trägt und je nach Tarifwahl ganz oder anteilig vom Versicherer erstattet bekommt. Daher besteht normalerweise keine direkte Beziehung zwischen dem Arzt und dem PKV-Unternehmen.

Ungefähr 90% der deutschen Bevölkerung ist Mitglied in der GKV und die übrigen 10% sind privat krankenversichert, von denen ca. 50% Beamte sind.

Die Leistungen der GKV sind in einigen Bereichen begrenzt (z. B. Erstattung von medizinischen Hilfsmitteln oder Zahnersatz). Aus diesem Grund haben fast 30% der gesetzlich krankenversicherten Personen eine private Zusatzversicherung zur Leistungsverbesserung abgeschlossen.

In der folgenden Analyse werden nur die Vollkostentarife der PKV betrachtet.

## Motivation für die Leistungsanalyse in der PKV

Eine Besonderheit der PKV ist die gesetzlich geregelte lebenslange Versicherungsgarantie, d.h. der Versicherer kann den Vertrag nicht kündigen. Um dieser langfristigen Verpflichtung nachkommen zu können, ist eine Beitragsanpassungsklausel im Versicherungsvertragsgesetz und Versicherungsaufsichtsgesetz verankert. Dadurch haben die PKV-Unternehmen die Pflicht, ihre Beiträge jährlich zu überprüfen und ggf. anzupassen (insb. im Hinblick auf die Leistungsentwicklung, Sterbewahrscheinlichkeiten und den Rechnungszins).

Simulationen der Deutschen Aktuarvereinigung (DAV) (1) haben gezeigt, dass die medizinische Inflation in den kommenden Jahrzehnten voraussichtlich der Haupttreiber für die Erhöhung der Beiträge im PKV-System sein wird. Somit ist das Kostenmanagement (insb. im Hinblick auf chronische Krankheiten) für die PKV im Hinblick auf die Kundenbindung, Reputation und die Wettbewerbsfähigkeit von großer Bedeutung.

## Ausgewählter Algorithmus des Machine Learnings

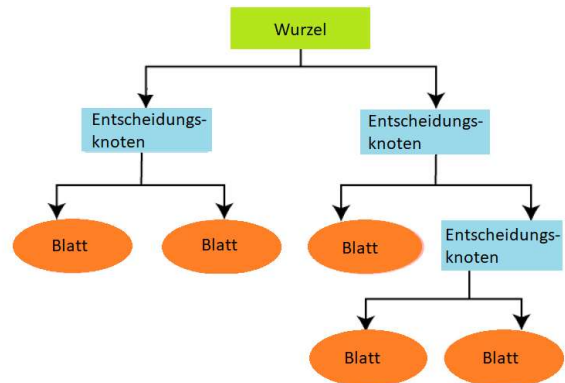
In diesem Abschnitt wird der Algorithmus zur Erstellung eines optimalen Modells dargestellt. Innerhalb des Algorithmus existieren verschiedene Einstellungsoptionen anhand derer das Modell kalibriert werden kann. Die Wahl dieser Optionen sowie die Bewertungsgrößen sind die maßgeblichen Variablen zur Entwicklung eines adäquaten Modells. Auf die Vor- und Nachteile bei der Wahl wird ebenfalls eingegangen.

## Entscheidungsbaum

Zu Beginn stellt sich die Frage was ein optimaler Algorithmus ist. Wir haben uns in diesem Fall für einen Entscheidungsbaum entschieden. Ein Entscheidungsbaum ist ein geordneter und gerichteter Baum, der der Darstellung von Entscheidungsregeln dient. Startpunkt des Algorithmus ist der Wurzelknoten. In ihm und jeden folgenden Knoten wird ein Attribut abgefragt, das eine Entscheidung über die Auswahl des folgenden Knoten trifft. Dies wird so lange fortgesetzt, bis ein Blatt (das Ende einer Verzweigung) erreicht wird. Ziel ist es, ein Modell zu erstellen, das den Wert einer Zielvariablen vorhersagt, indem einfache Entscheidungsregeln aus den Ausprägungen des Attributs abgeleitet werden.

Entscheidungsbäume können sowohl für Regressions- als auch für Klassifikationsprobleme angewendet werden. Für ein Regressionsproblem wird ein kontinuierlicher Wert (z.B. die Wahrscheinlichkeit eines möglichen Ergebnisses) und für ein Klassifizierungsproblemeine natürliche Zahl (z.B. 0, 1, 2, ..., k) berechnet.

ABBILDUNG 1: ENTSCHEIDUNGSBAUM



Vorteile des Entscheidungsbaums:

- Einfach zu verstehen und zu interpretieren.
- Gut visualisierbar.
- Erfordert wenig Datenvorbereitung.
- Aufwand wächst logarithmisch zur Anzahl der Datenpunkte, die zum Trainieren des Baums verwendet werden.
- Verwendung von sowohl numerischen als auch kategorischen Daten möglich.
- Validierung des Modells mithilfe statistischer Tests möglich.

Nachteile des Entscheidungsbaums:

- Overfitting: komplexe Entscheidungsbäume ohne zielführende Verallgemeinerung.
- Kleine Abweichungen in den Daten können dazu führen, dass ein völlig anderer Baum generiert wird.
- Lokal optimale Entscheidungen in den Knoten garantieren nicht, dass der global optimale Entscheidungsbaum zurückgegeben wird.
- Nicht geeignet bei unbalancierten Daten (d.h. wenn es zu einer Klasse deutlich mehr Daten gibt als in den übrigen) ohne vorherige Anpassung des Datensatzes.

Da die in unserem Fall verwendeten Daten sehr unausgeglichen sind, ist ein „einfacher“ Entscheidungsbaum keine gute Wahl für die Leistungsanalyse. Um die Vorteile des Entscheidungsbaumes zu nutzen, betrachten wir eine Variation, bei der eine Aneinanderreihung mehrerer Bäume genutzt wird.

## XGBoost<sup>1,2</sup>

XGBoost steht für „extreme gradient boosting“. Das Paket XGBoost beinhaltet den Algorithmus zur Erstellung eines Gradient Boosting-Entscheidungsbaumes. Der Algorithmus besteht aus einer zusammengesetzten Lernmethode. Boosting ist hierzu eine weit verbreitete Technik.

Beim Boosting werden neue Modelle hinzugefügt, um die Fehler bestehender Modelle zu korrigieren. Jeder Baum lernt von seinen Vorgängern und aktualisiert die Residualfehler. Modelle werden nacheinander hinzugefügt, bis keine weiteren Verbesserungen mehr vorgenommen werden können. Die einzelnen Bäume beim Boosting sind schwache Modelle mit hohem Bias und die Vorhersagekraft ist nur inkrementell besser als zufälliges Raten. Die Zusammensetzung der einzelnen schwachen Modelle ergibt ein starkes Modell, welches sowohl den Bias und als auch die Varianz reduziert.

Durch Boosting wird die Varianz bei jedem Iterationsschritt verringert.

Eine hohe Anzahl an Entscheidungsbäumen könnte zu Overfitting führen. Deshalb ist es wichtig die maximale Anzahl der Bäume nicht zu groß zu wählen.

Der Boosting-Vorgang beinhaltet drei einfache Schritte:

1. Es wird ein initiales Modell  $F_0$  anhand der originalen Trainingsdaten  $(x_i, y_i)$  erstellt, um die Zielvariable  $y$  zu schätzen.

$$F_0(x) = \underset{p}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, p(x_i))$$

wobei  $L(y, F(x))$  eine differenzierbare Verlustfunktion ist (z.B.  $L(y, F(x)) = \|F(x) - y\|$  oder  $L(y, F(x)) =$

$\|F(x) - y\|^2$ ) und  $p$  für den Entscheidungsbaum steht, der die beschreibenden Variablen  $x_i$  der Zielvariable  $y_i$  zuordnet. Es wird die Summe über alle Beobachtungen gebildet. Der Gradient wird verwendet, um das Minimierungsproblem zu berechnen.

Das Modell führt zu dem Residuum  $r_0 = y - F_0$ .

2. Für jeden Iterationsschritt  $m=1, \dots, M$  gilt:

- Berechne die Residuen  $r_{m-1}$ .
  - Passe ein Modell  $f_m$  an die Residuen  $r_{m-1}$  des vorherigen Modells an.
  - Aktualisiere das Modell:  $F_m(x) = F_{m-1}(x) + \eta * f_m(x)$ , wobei  $\eta$  eine gegebene Lernrate ist.
3. Gib  $F_M(x)$  aus.

Dieser Ansatz wird als Gradienten-Boosting bezeichnet, da ein Gradientenabstiegs-Algorithmus verwendet wird, um die Verlustfunktion beim Hinzufügen neuer Modelle zu minimieren.

Dabei wurde die Funktion `xgb.train` des Pakets XGBoost in R genutzt:

```
xgboost_model <- xgb.train(      params =
xgb_params_typ2, data = training_data, nrounds = n, watchlist =
watchlist_typ2, eta = eta_param, max.depth = md)
```

mit den folgenden Eingangsdaten:

- `params`: Hier können Parameter wie die Anzahl der Klassifikations-Kategorien festgelegt werden.
- `data`: Die Trainingsdaten.
- `nrounds` (default: 100): Die Maximale Anzahl an Iterationsschritten.
- `watchlist`: Definiert die Ausgabe der Funktion.
- `eta` (default: 0.3, range: [0,1]): Steuert die Lernrate, d.h. die Rate, mit der das Modell Muster aus Daten lernt. Nach jeder Runde werden die Merkmalsgewichte verkleinert, um das beste Optimum zu erreichen. Ein niedrigeres `eta` führt zu einer langsameren Berechnung (in Bezug auf die Leistung).
- `max_depth` (default: 6, range: [0, inf]): Steuert die Tiefe des Baums. Je größer die Tiefe, desto komplexer das Modell, mit größerem Risiko des Overfitting. Es gibt keinen Standardwert für `max_depth`. Größere Datensätze erfordern tiefere Bäume, um die Muster aus Daten zu erkennen.

Zur Vorhersage der Testdaten kann die Funktion „`predict`“ verwendet werden. Sie gibt die Wahrscheinlichkeiten aller Testdaten für jede Klasse zurück.

Einige Vorteile des XGBoost sind:

1 XGBoost Algorithm: XGBoost in Machine Learning. Siehe [analyticsvidhya.com](https://analyticsvidhya.com).

2 XGBoost in R : A Complete Tutorial Using XGBoost in R. Siehe [analyticsvidhya.com](https://analyticsvidhya.com).

- Schnelle Berechnung.
- Hohe Genauigkeit.
- Einfach zu verwenden.
- Kann unausgeglichene Datensätze verarbeiten (d.h. die Zielvariable enthält in einer Klasse mehr Beobachtungen als in einer anderen).
- Kann mit großen Datenmengen umgehen.

Der Nachteil des XGBoost:

- Nicht so einfach zu interpretieren wie ein einfacher Entscheidungsbaum.
- Overfitting ist möglich, wenn die Parameter nicht richtig eingestellt<sup>3</sup> sind.

Es gibt verschiedene Möglichkeiten, die Qualität des angepassten Modells zu messen. Hier sind die ROC-Kurve (Receiver Operating Characteristic) und die Fläche unter der Kurve (AUC, area under the curve) zu nennen, die im Folgenden dargestellt werden.

## ROC Kurve

Die Receiver Operating Characteristic (ROC) Kurve ist eine Qualitätsmessung für Klassifizierungsprobleme. Sie wird in einem Diagramm dargestellt und veranschaulicht die Diagnosefähigkeit eines binären Klassifikators, wenn sein Schwellenwert variiert wird.

Die ROC Kurve wird erstellt, indem die True Positiv Rate (TPR) gegen die False Positive Rate (FPR) im Bezug zu verschiedenen Schwellenwerten aufgezeichnet wird. Dabei ist:

$$TPR = \frac{TP}{TP + FN}$$

und

$$FPR = \frac{FP}{FP + TN}$$

mit:

- TP (True Positives) steht für alle richtig klassifizierte Fälle mit der gewählten Eigenschaft.
- FP (False Positives) steht für alle falsch klassifizierte Fälle mit der gewählten Eigenschaft.
- TN (True Negatives) steht für alle richtig klassifizierte Fälle ohne die gewählte Eigenschaft.
- FN (False Negatives) steht für alle falsch klassifizierte Fälle ohne die gewählte Eigenschaft.

Durch Verringern des Schwellenwerts werden mehr Fälle als positiv klassifiziert, wodurch sowohl False Positives als auch True Positives erhöht werden.

<sup>3</sup> Um Overfitting zu vermeiden, sollten max\_depth und Eta nicht zu groß sein. Das Modell ist sehr wahrscheinlich

ABBILDUNG 2: BEISPIEL FÜR ROC KURVEN

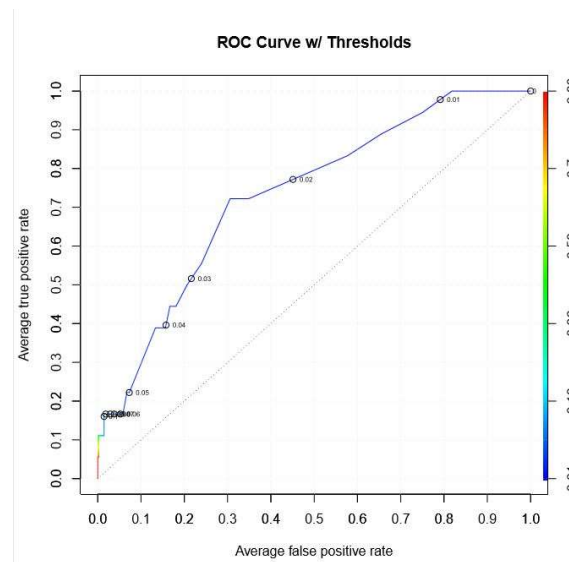
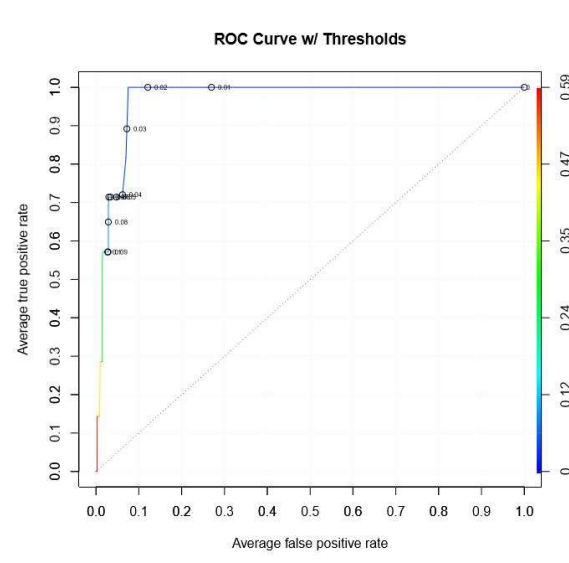


Abbildung 3 zeigt zwei ROC Kurven. Anhand der Kurven ist zu sehen, dass das Modell der linken Grafik eine bessere Schätzung abliefern als das Modell der rechten Grafik, da die ROC Kurve weiter in der linken oberen Ecke verläuft, d. h. man erhält eine hohe Anzahl an True Positives bei gleichzeitig geringer Anzahl an False Positives

## AUC

**DIE „AREA UNDER THE CURVE“ (AUC) MISST DEN GESAMTEN BEREICH UNTERHALB DER ROC-KURVE VON (0,0) BIS (1,1).**

Das perfekte Modell führt zu AUC = 1 und hat eine ROC-Kurve, die durch die obere linke Ecke verläuft (ungefähr 100%

overfitted, wenn es die Trainingsdaten gut schätzt, aber die Testdaten nicht.

True Positive Rate und 0% False Positive Rate; die ROC-Kurve in der linken Grafik in Abbildung 3 führt zu AUC nahe 1).

## Fallstudie

### DATENSATZ UND DESSEN AUFARBEITUNG

Der verwendete Datensatz wurde von einem deutschen privaten Krankenversicherungsunternehmen zur Verfügung gestellt und enthält für eine Vielzahl von Tarifen die Informationen Alter, Geschlecht und die Historie aller Diagnosen von 2010 bis 2018. Die Diagnosen sind als ICD-10-GM-Codes enthalten.

Wie in der Beschreibung des Algorithmus erwähnt, sollte darauf geachtet werden, dass das Modell nicht zu detailliert aufgebaut wird (Vermeidung von Overfitting). Dieser Punkt findet schon in der Vorbereitung des Datensatzes Anwendung. Die ICD-10-GM-Codes enthalten mehr als 16.000 Diagnosen, und sind tief aufgeschlüsselt. Eine dermaßen detaillierte Betrachtung erscheint als nicht zielführend, auf Grund dessen wurde ein Mapping der Codes auf die ersten 3 Stellen vorgenommen. Mit dieser Vereinfachung gehen 245 beschreibende Variablen in das Modell ein, die aus 241 ICD-Codes und Informationen zu Geschlecht, Alter und Tarif bestehen. Die Analyse des Datensatzes ist beispielhaft auf Diabetes Typ 2 als chronische Zielerkrankung aufgebaut. Die Zielvariable ist eine kategorische Variable, die eine Diabetesdiagnose anzeigt (0/1). Da für die Auswertung eine ausreichende Datenhistorie benötigt wird, werden alle Verträge ohne dreijährige Historie entfernt. Der ursprüngliche Datensatz enthält 72.371 Versicherte aus 79 verschiedenen Tarifen mit 1.884.505 Diagnosen (vor der Zuordnung zu den ersten drei Ziffern von ICD-10-GM). Es wurde eine Teilmenge des Datensatzes mit Vollkostentarifen verwendet, die 21.101 Personen und 576.742 Diagnosen nach der Validierung umfasst. Für jeden Versicherungsnehmer im Datensatz wird eine einzelne Zeile erstellt. Beispielhaft ist dies in Abbildung 4 dargestellt.

**ABBILDUNG 3: AUSZUG DER DATENMATRIX FÜR XG BOOST**

Sex	Diabetes	age_diab_2018	current_tariff	A00	A15	A20	...
1	0	60	xxxx	0	0	0	...
1	1	61	yyyy	2	0	0	...
...	...	...	...	...	...	...	...

Anmerkungen:

Sex: 0 = Frau, 1 = Mann.

Diabetes: 0 = kein Diabetes, 1 = Diabetes.

age\_diab\_2018:

Entweder das Alter, in dem die Diabetesdiagnose zum ersten Mal gestellt wurde, oder das Alter im Prognosejahr 2018, für Versicherte ohne Diabetesdiagnose.

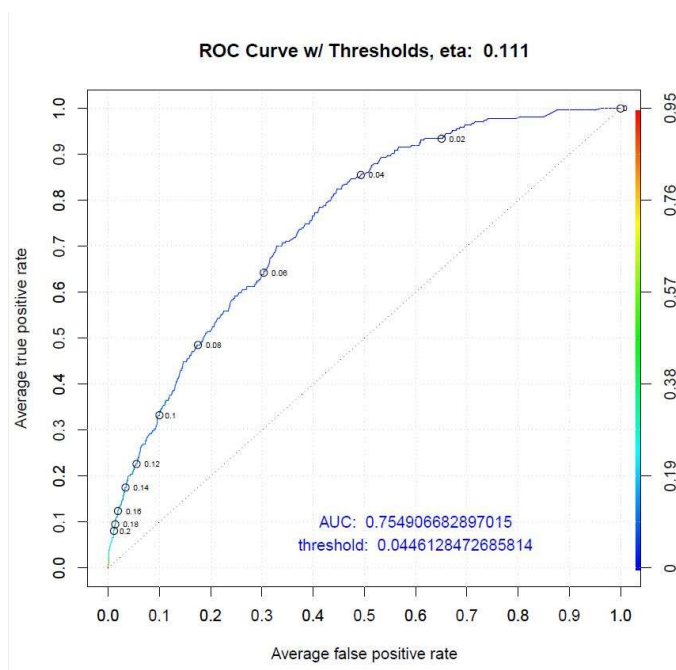
current\_tariff: Tarif im Prognosejahr (Tarifwechsel zwischen Vertragsabschluss und Projektion werden nicht beachtet).

A00, ...: Anzahl der gestellten Diagnosen pro Versicherungsnehmer für jeden ICD-10-GM-Code innerhalb der letzten drei Jahre. Wenn die Diagnose in mehreren Abrechnungsnummern gestellt wurde, wird hochgezählt.

Der Datensatz enthält 1.135 identifizierte Diabetiker von 21.101 Versicherten und wurde zufällig in Trainingsdaten (drei Viertel der Gesamtdaten) und Testdaten (ein Viertel der Gesamtdaten) unterteilt. Die Trainingsdaten enthalten 861 identifizierte Diabetiker und die Testdaten 274 identifizierte Diabetiker.

### ERGEBNISSE UND QUALITÄT DER ERGEBNISSE

Der XGBoost-Algorithmus als Machine Learning Methode verhilft bei der Erstellung eines Modells zur Vorhersage von Diabetikern. Dabei werden folgende Vorschriften beachtet, um Overfitting zu vermeiden und Prinzipien des Machine Learnings genutzt, um z.B. eine Rangordnung der Diagnosen (siehe Abbildung 6) aus den Trainingsdaten zu erlernen und diese in der Schätzung der Wahrscheinlichkeit der Zieldiagnose pro Versicherungsnehmer in den Testdaten anzuwenden. Diese Wahrscheinlichkeiten können mit Hilfe der oben eingeführten ROC-Kurve und der dazugehörigen AUC zur Kalibrierung verwendet werden. Die ROC-Kurve gibt Auskunft darüber, wie viele Versicherte zu jedem möglichen Schwellenwerten der errechneten Wahrscheinlichkeiten korrekt als Diabetiker (Verlauf der Kurve in der oberen linken Ecke) oder falsch (Verlauf der Kurve in der unteren rechten Ecke) geschätzt wurden. Die Kurve, die im Datensatz nach Kalibrierung zu einem angemessenen Modell führt, ist in Abbildung 5 zu sehen. In diesem Fall verläuft die Kurve leicht nach oben links gekrümmt. Dies ist keine perfekte Form, jedoch wesentlich besser als die Diagonale, die nur zufälliges Raten darstellen würde und führt von allen getesteten Parameterkombinationen zum besten Ergebnis. Die genauen Einstellungen wurden mit dem Flächenmaß unter der ROC-Kurve (AUC) bewertet. Die Parameterkombination mit der höchsten AUC qualifizierte sich automatisch als beste Einstellung, da ein AUC von 1 eine Trefferquote von 100% TP und 0% FP darstellt. Der XGBoost Algorithmus erstellt die Entscheidungsbäume iterativ und errechnet zu jedem Iterationsschritt ("nrounds") den Residualfehler. Die Zahl der hintereinander gelagerten Entscheidungsbäume konnte mit einer Minimumfunktion über das Residuum gewählt werden. Auch bei dem Umfang der Daten blieb das Minimum unter 100. Der default-Wert für die maximale Tiefe ist 6 und sollte nicht zu groß sein, um Overfitting zu vermeiden. Nach dem Testen der AUC-Ergebnisse wurde der Wert 5 gewählt. Es besteht auch die Gefahr von Overfitting bei großem eta. Alle Werte von 0,001 bis zum default- Wert von 0,3 wurden getestet. Der beste Wert für AUC führt zu einem eta von 0,111.

**ABBILDUNG 4: : ROC KURVE FÜR DIE TARIF-KATEGORIE**


Die Summe der Einzelwahrscheinlichkeiten spiegelt die erwartete Gesamtzahl der Diabetiker wider. Das Modell schätzt eine erwartete Anzahl von 268,5 neuen Diabetes Diagnosen im Testdatensatz, der tatsächlich 274 Diabetes Diagnosen enthält.

In der folgenden Abbildung 6 werden die Diagnosen gezeigt, die als die 15 relevantesten Treiber für das erstellte Modell identifiziert wurden (abnehmende Relevanz):

**ABBILDUNG 5 : IMPORTANCE MATRIX DER DIAGNOSEN FÜR DIE TARIF-KATEGORIE 2**

Position	Feature	Diagnoses/Erklärung
1	age_diab_2018	Alter bei Diabetes-Erstdiagnose oder zum Start der Prognose
2	I48	Herz- und Kreislauferkrankungen
3	E65	Folgen der Überernährung
4	G35	Demyelinisierende Krankheit des Zentralnervensystems
5	I10	Sekundäre Hypertonie
6	R70	Abnormer Befund der Blutchemie
7	H49	Refraktionsfehler
8	K00	Zahnerkrankungen
9	B35	Mykose
10	K70	Leberkrankheiten
11	current_tariff	Tarif zu Prognosebeginn
12	M05	Arthropathien
13	E70	Ernährungs- und Stoffwechselstörungen
14	K55	Hämorrhoiden
15	E50	Vitamin A-Mangel

<sup>4</sup> Trasino, S.E. et al. (1 December 2014). Vitamin A deficiency causes hyperglycemia and loss of pancreatic  $\beta$ -cell mass. Journal of Biological Chemistry. Auszug aus <https://www.jbc.org/content/early/2014/12/01/jbc.M114.616763.full.pdf+html> vom 09.01.2021.

Diabetes ist ein gutes Analysebeispiel, da ein umfangreiches Allgemeinwissen zu Komorbiditäten existiert. Die Ränge der Treiber- Diagnosen aus dem Modell können mit bekannten Differentialdiagnosen und Komorbiditäten zu Diabetes validiert werden und bestätigen die Sinnhaftigkeit der vom Modell quantifizierten Diagnosen.

Es ist eine allgemein akzeptierte Annahme, dass Diabetes hauptsächlich im mittleren oder höheren Alter auftritt. Aus diesem Grund ist es nicht überraschend, dass das Alter im Modell als Haupttreiber für Diabetes identifiziert wird. Darüber hinaus gibt es verschiedene Krankheiten, die signifikant mit Diabetes korrelieren.

Beispielsweise können Herzerkrankungen aus Diabetes resultieren. Sie können bei einer späten Diabetesdiagnose möglicherweise auch vor ihr erkannt werden und sind ebenfalls unabhängig von Diabetes denkbar.

Die Diagnosen E65 und E70 beziehen sich auf falsche Ernährung oder Störungen in der Verstoffwechslung der Ernährung und korreliert mit Diabetes Typ 2.

Studien<sup>4,5</sup> zeigen, dass Vitamin A-Mangel (E50) eine verminderte Insulinproduktion nach sich zieht und somit im direkten Zusammenhang mit Diabetes Typ 2 steht.

Der vom Versicherungsnehmer gewählte Tarif (current\_tariff) gibt Auskunft darüber, in welcher sozio-ökonomischen Situation er sich befindet. Die finanziellen Mittel der Versicherungsnehmer beeinflussen ihre Gesundheitsversorgung in hohem Maß.

Einige der anderen aufgeführten Diagnosen sind bekannte Komorbiditäten von Diabetes. Einige sind medizinisch mit einer

<sup>5</sup> Health and Science (23 April 2018). Can a vitamin A deficiency promote diabetes? Auszug aus [https://www.healthandscience.eu/index.php?option=com\\_content&view=article&id=1494:kann-ein-vitamin-a-mangel-diabetes-beguenstigen&catid=20&lang=de&Itemid=316](https://www.healthandscience.eu/index.php?option=com_content&view=article&id=1494:kann-ein-vitamin-a-mangel-diabetes-beguenstigen&catid=20&lang=de&Itemid=316) vom 9.01.2021.

Korrelation belegt (wie z.B. B356 oder M057). Andere haben allgemein eine hohe Diagnosefrequenz oder stehen im Zusammenhang mit dem Alter (wie z.B. G35, H49, K00 oder K55) und erhalten infolgedessen einen hohen Stellenwert, stehen jedoch in keinem signifikant bekannten Zusammenhang mit Diabetes.

In der ROC Kurve (Abbildung 5) wird der perfekte Schwellenwert entweder als der Wert identifiziert, für den die ROC-Kurve den größten Abstand zur Diagonale oder den kleinsten Abstand zur oberen linken Ecke aufweist. Diese Ansätze führen zu den Schwellenwerten von 4,461% bzw. 5,662%. Wenn diese als optimaler Schwellenwert gewählt werden, folgen fast 2.000 FP's. Dieses Resultat erscheint jedoch ebenfalls nicht völlig abwegig. Es kann auf eine unvollständige Historie der Diagnosen zurückzuführen sein, da Rechnungen deren Leistungsbetrag unter der Grenze des Selbstbehaltes liegen, häufig nicht bei den Versicherungsunternehmen eingereicht werden. Es könnte auch darauf hindeuten, dass in Zukunft (nach dem zur Verfügung stehenden Testzeitraum) möglicherweise eine Diabetesdiagnose gestellt wird.

Aus praktischer Sicht kann es sinnvoll sein, den Schwellenwert abhängig von den Größen der unternehmensindividuellen Anwendung zu wählen. Dafür kann die ROC-Kurve ebenfalls als Confusion Matrix (Abbildung 7) dargestellt werden. In ihr wird die TPR und die TFR in Bezug auf den Schwellenwert dargestellt.

ABBILDUNG 6 : ALLGEMEINE CONFUSION MATRIX

Tatsächlicher Wert	Geschätzt		
	Nicht-Diabetiker	Diabetiker	Zeilensumme
Nicht-Diabetiker	TN = tatsächliche_nicht-Diabetiker - FP	FP = FPR * tatsächliche_nicht-Diabetiker	tatsächliche_nicht-Diabetiker
Diabetiker	FN = tatsächliche_Diabetiker - TP	TP = TPR * tatsächliche_Diabetiker	tatsächliche_Diabetiker
Spaltensumme	geschätze_nicht-Diabetiker = TN + FN	geschätze_Diabetiker = FP + TP	Gesamtsumme

Abhängig von den Schwellenwerten ergeben sich für den Datensatz die folgenden Verhältnisse der TPR und FPR zu den Schwellenwerten:

6 Eine schlechte medikamentöse Diabetes-Einstellung vergünstigt pathologische Hautveränderungen. Siehe: [https://www.diabetesde.org/ueber\\_diabetes/begleiterkrankungen\\_bei\\_diabetes/hautbeschwerden](https://www.diabetesde.org/ueber_diabetes/begleiterkrankungen_bei_diabetes/hautbeschwerden).  
 7 Aus Untersuchungen der Universität Erlangen geht hervor, dass Diabetes ein Risikofaktor für Arthrose ist. Dieser Einfluss

ABBILDUNG 7 : TREFFERQUOTEN IM DATENSET

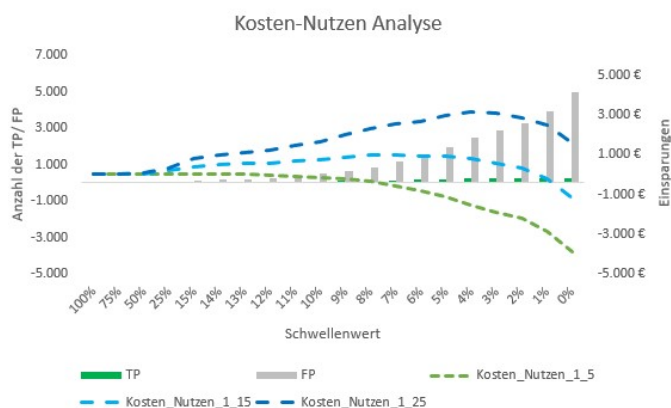
Schwellenwert	TP	FP	TN	FN	TPR	FPR
100 %	0	0	5.001	274	0%	0%
75 %	2	0	5.001	272	0%	1%
50 %	3	5	4.996	271	0%	1%
25 %	15	28	4.973	259	1%	5%
15 %	38	129	4.872	236	3%	14%
10 %	91	501	4.500	183	10%	33%
5 %	205	1.934	3.067	69	39%	75%
4 %	234	2.466	2.535	40	49%	85%
3 %	251	2.914	2.087	23	58%	92%
2 %	256	3.254	1.747	18	65%	93%
1 %	268	3.953	1.048	6	79%	98%
0 %	274	5.001	0	0	100%	100%

Der Testdatensatz enthält 274 Diabetiker. In Abbildung 8 ist zu sehen, dass die absolute Anzahl der FP's umso größer ist, je höher die Zahl der TP ist. Nun stellt sich die Frage welches Verhältnis von TP und FP und damit die Wahl des Schwellenwerts für den Kontext der speziellen Anwendung sinnvoll ist.

Eine denkbare Anwendung wäre das Anbieten von Präventivmaßnahmen. Bei der Betrachtung von Diabetes könnten Ernährungsberatungstrainings oder die Subventionierung von Fitnessstudio-Mitgliedschaften als solche in Erwägung gezogen werden. Dabei bleibt das Verhältnis aus Kosten zu Nutzen abzuwägen. Als Nutzen sind die „gesparten“ Leistungsausgaben und als Kosten sind die Aufwände für Präventivmaßnahmen anzusehen. Angenommen die Einsparungen wären 100-mal höher als die Kosten für die Präventivmaßnahmen, dann sollte die Maßnahme dem gesamten Portfolio angeboten werden, da die Einsparungen für die echten Diabetiker ausreichend hoch sind, dass Kosten für die Präventivmaßnahme nicht ins Gewicht fallen. In der Realität liegt das Kosten-Nutzen-Verhältnis jedoch vermutlich unter 1:100, sodass die optimale Anzahl von Teilnehmern für die Präventivmaßnahmen berechnet werden muss. Abbildung 9 veranschaulicht die Analyse für unterschiedliche Kosten-Nutzen-Verhältnisse (1:5 / 1:15 / 1:25).

wurde auch unabhängig von Alter und dem Body Maß Index (BMI) festgestellt. Siehe: <https://www.bvou.net/warum-leiden-diabetiker-so-offt-an-arthrose/> and <https://www.aerztezeitung.de/Medizin/Typ-2-Diabetes-erhoeht-Risiko-fuer-Arthrose-282599.html>.

ABBILDUNG 8 : KOSTEN-NUTZEN-ANALYSE



Diese Beispiele zeigen, dass die Vorhersage der zukünftigen Diabetesdiagnosen (TP möglichst exakt und FP möglichst klein) umso besser sein muss, je niedriger das Kosten-Nutzen-Verhältnis ist.

Wenn der mit den Kosten verbundene Nutzen gering ist, müssen die Kosten gesenkt werden. Dies bedeutet im Umkehrschluss, dass entweder die Anzahl der fälschlicherweise als Diabetiker geschätzten Personen möglichst gering sein muss oder die Präventivkosten zu senken sind, sich dabei jedoch die Sinnhaftigkeit der Präventivmaßnahme stellt.

Wenn der Nutzen im Verhältnis zu den Kosten steigt, muss das Verhältnis zwischen TPs und FPs optimaler werden. Die Anzahl der TPs muss erhöht werden.

Der Schwellenwert kann in Abhängigkeit vom Verhältnis zwischen den Vorteilen für TPs und den unnötigen Kosten für FPs gewählt werden.

## Fazit

Eine gute Vorhersage relevanter Krankheiten bietet den Unternehmen Optionen zur Steuerung ihrer Leistungsausgaben. Chronische Krankheiten spielen eine entscheidende Rolle bei der Entwicklung zukünftiger Leistungsausgaben, weshalb sich die Analyse als Beispiel auf Diabetes Typ-2 konzentriert.

Die Analyse zeigt, dass Algorithmen für Machine Learning geeignet sind, um die zukünftige Entwicklung spezifischer

chronischer Krankheiten abzuschätzen. Das Hauptaugenmerk liegt dabei auf der Aufbereitung der Eingabedaten, was für die Aufstellung eines geeigneten Modells wichtig ist. Der Entscheidungsbaum-Algorithmus bietet die Möglichkeit, die Ergebnisse zu erklären, und die Importance Matrix ermöglicht einen Einblick in die Treiber für eine Diabetesdiagnose. Dies ist sehr wertvoll und kann bei der Verwaltung zukünftiger Leistungsausgaben hilfreich sein.

Aufgrund der im Bundesdatenschutzgesetz (DSGVO) geregelten Datenschutzbestimmungen dürfen die PKV-Unternehmen diese Analyse nicht verwenden, um gezielt einzelnen Versicherungsnehmer die scheinbar drohende Diagnose und daraufhin geforderten und abgestimmten Maßnahmen mitzuteilen. Die Unternehmen könnten jedoch anhand der Analyse herausfinden, welche Versicherungsnehmer von dem Risiko einer chronischen Krankheit betroffen sind und könnten daraus Teilportfolios erstellen und verwalten (z.B. bei bestimmten bestehenden Krankheiten).

Der Algorithmus liefert eine erwartete Anzahl von Diabetikern im nächsten Jahr und damit eine gute Schätzung der anstehenden Leistungsausgaben. Es besteht auch die Möglichkeit, vorbeugende Maßnahmen im Hinblick auf einen Kosten-Nutzen-Ansatz zu bewerten. Außerdem ist das Modell hinsichtlich der Beschreibungs- und Zielvariablen sehr flexibel, so dass Anwendungen für andere chronische Krankheiten oder die Verwendung einer Vorgeschichte verschreibungspflichtiger Medikamente (anstelle der Historie der Diagnosen) möglich sind.



## References

1. javapoint.com. [Online]  
<https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>.
2. (DAV), German Association of Actuaries.  
Auswirkungen demografischer Effekte auf die Krankenversicherung (Ergebnispaper). 2018.
3. 3. [Online] 12 2020.  
[https://www.healthandscience.eu/index.php?option=com\\_content&view=article&id=1494:kann-ein-vitamin-a-mangel-diabetes-beguenstigen&catid=20&lang=de&Itemid=316](https://www.healthandscience.eu/index.php?option=com_content&view=article&id=1494:kann-ein-vitamin-a-mangel-diabetes-beguenstigen&catid=20&lang=de&Itemid=316).
4. Lund, Pernille. healthandscience. healthandscience. [Online] [Cited: 12 30, 2020.]  
[https://www.healthandscience.eu/index.php?option=com\\_content&view=article&id=1494:kann-ein-vitamin-a-mangel-diabetes-beguenstigen&catid=20&lang=de&Itemid=316](https://www.healthandscience.eu/index.php?option=com_content&view=article&id=1494:kann-ein-vitamin-a-mangel-diabetes-beguenstigen&catid=20&lang=de&Itemid=316).
5. Journal of Biological Chemistry. [Online] [Cited: 12 30, 2020.]  
<https://www.jbc.org/content/early/2014/12/01/jbc.M114.616763.full.pdf+html>.
6. deutsche Diabetes Hilfe. [Online] [Cited: 12 30, 2020.]  
[https://www.diabetesde.org/ueber\\_diabetes/begleiterkrankungen\\_bei\\_diabetes/hautbeschwerden](https://www.diabetesde.org/ueber_diabetes/begleiterkrankungen_bei_diabetes/hautbeschwerden).
7. Berufsverband für Orthopädie und Unfallchirurgie. [Online] [Cited: 12 30, 2020.]  
<https://www.bvou.net/warum-leiden-diabetiker-so-oft-an-arthrose/>.
8. XGBoost in R : A Complete Tutorial Using XGBoost in R (analyticsvidhya.com).
9. XGBoost Algorithm: XGBoost in Machine Learning (analyticsvidhya.com).
10. ÄrzteZeitung. [Online] Springer Medizin, 8 14, 2013. [Cited: 1 7, 2021.]  
<https://www.aerztezeitung.de/Medizin/Typ-2-Diabetes-erhoeht-Risiko-fuer-Arthrose-282599.html>.



Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

[de.milliman.com](https://de.milliman.com)

#### **KONTAKT**

**Alexander Krauskopf**  
[alexander.krauskopf@milliman.com](mailto:alexander.krauskopf@milliman.com)

**Katrin Höllering**  
[katrin.hoellering@milliman.com](mailto:katrin.hoellering@milliman.com)

**Alexandra Pflumm**  
[alexandra.pflumm@milliman.com](mailto:alexandra.pflumm@milliman.com)