

Andreas Wolik

Predictive Modeling

Bestimmung von Hospitalisierungswahrscheinlichkeiten mithilfe von Data-Mining-Verfahren

Die Kenntnis bevorstehender Patientenkarrerien und die Identifikation zukünftiger Hochkostenfälle sind für ein optimales Versorgungsmanagement entscheidend. Ziel des hier vorgestellten Predictive-Modeling-Ansatzes ist die Vorhersage derjenigen Versicherten mit der höchsten Wahrscheinlichkeit für einen Krankenhausaufenthalt. Die Ergebnisse bestehender Versichertenklassifikationssysteme werden dabei als Information im Modell verwendet. Dabei werden Hospitalisierungswahrscheinlichkeiten zum einen mit Entscheidungsbaumverfahren und zum anderen mit Neuronalen Netzen prognostiziert. Beide Methoden werden hinsichtlich ihrer Prognosegüte miteinander verglichen.

Einführung

Die Veränderungen durch das Wettbewerbsstärkungsgesetz (GKV-WSG) haben zur Folge, dass die Kassen noch intensiver als in der Vergangenheit Konzepte zur Kostenreduktion nutzen müssen. Die Fallkosten im Krankenhausbereich sind erheblich und haben damit eine vorrangige Relevanz für die Konzeption eines Kostensenkungsprogramms. Ein Weg zur Kostenreduktion und gleichzeitigen Steigerungen der Versorgungsqualität liegt in der Steuerung von Patientenkarrerien und der dadurch erreichbaren Vermeidung von Krankenhausaufenthalten.

Bisherige Ansätze sahen breit angelegte Programme für Versichertengruppen mit einem fest definierten Krankheitsbild, wie beispielsweise Diabetes-Mellitus, vor. Doch ist nicht jeder Diabetiker im gleichen Maße gefährdet, zu entgleisen und hospitalisiert werden zu müssen. Daher werden die in Frage kommenden Versicherten nach ihrem Risiko, hospitalisiert zu werden, differenziert. Dazu ist es notwendig, die Wahrscheinlichkeit der Hospitalisierung zu ermitteln, um sich auf die Gruppe mit den höchsten Hospitalisierungswahrscheinlichkeiten konzentrieren zu können.

Diese Personengruppe kann daraufhin durch ein Case-Management beziehungsweise Versorgungsmanagement eine optimierte Betreuung erhalten, wodurch die Wahrscheinlichkeit einer Hospitalisierung vermindert werden kann.

Die Kassen stehen vor dem Problem, den in Frage kommenden Personenkreis möglichst genau identifizieren zu müssen, um dadurch den optimalen Nutzen bei gegebener Ressource (Geld und Personal) zu erreichen. Eine Lösung besteht darin, durch ein Prognosemodell möglichst solche Personengruppen zu identifizieren, die in besonderer Weise für ein Betreuungs-/Vorsorgeprogramm geeignet sind und die eine möglichst geringe Quote von Fehlklassifizierungen aufweisen. Mithilfe von Predictive-Modeling-Methoden können solche Aufgaben gelöst werden.

Unter dem Begriff Predictive Modeling werden Methoden zur Bestimmung von Eintrittswahrscheinlichkeiten zukünftiger Ereignisse oder Zustände zusammengefasst. Die Basis für eine Prognose bilden in den meisten Anwendungen Informationen zu einzelnen Individuen. Diese Informationen unterteilen sich in soziodemografische Daten und Informationen, die sich aus dem Verhalten oder den Morbiditätsinformationen der betrachteten Individuen ablesen lassen.

Die Anwendungsbereiche des Predictive Modeling sind nicht allein auf die Prognose zukünftiger Ereignisse beschränkt. Auch die Compliance einzelner Versicherter kann prognostiziert werden. Das Wissen um die Compliance kann dazu genutzt werden, die Art des Ansprechkanals und die Intensität des Kontaktes zu bestimmen. Für einen Diabetiker könnte beispielsweise die Steigerung seiner niedrigen Compliance dadurch erreicht werden, dass er durch eine Schulung eine höhere Sensibilität für seine Ernährung erlangt. Hierdurch könnte eine Unterzuckerung vermieden und damit eine Hospitalisierung wegen eines entgleisten Diabetes-Mellitus unwahrscheinlicher werden.

Versichertenklassifikationssoftware als ein Predictive Modeling zukünftiger Kosten

Prognosesysteme sind in den letzten Jahren fester Bestandteil im deutschen Gesundheitswesen geworden. Sowohl in der morbiditätsorientierten ambulanten Vergütung als auch im Kontext des morbiditätsorientierten Risikostrukturausgleichs ist es notwendig, auf Basis von Morbiditätsinformationen der Versicherten die Kosten mithilfe eines Zuschlagsansatzes zu prognostizieren und damit eine morbiditätsorientierte Vergütung respektive Zuweisung aus dem RSA abzubilden.

Die Versichertenklassifikationssoftware von DxCG/Urix hat sich in diesem Zusammenhang in Deutschland etabliert. Die im IGES-Gutachten empfohlene Versichertenklassifikationssoftware der Firma DxCG (Reschke et al. 2004) stellt für ein Prognosesystem eine geeignete Datengrundlage zur Verfügung, sodass mithilfe eines Zuschlagsansatzes und einer Regression die Kosten der einzelnen Versicherten mit einem prospektiven Ansatz prognostiziert werden können.

Im Rahmen der Einführung des morbiditätsorientierten RSA entwickelt das Bundesversicherungsamt (BVA) eine Versichertenklassifikationssoftware. Diese basiert auf der Software von DxCG und bildet die Restriktionen der 80 Krankheiten, die vom BVA definiert wurden, als Systematik des morbiditätsorientierten RSA für das deutsche GKV-System ab. Damit ist das BVA in der Lage, die Zuschläge für eine Verteilung innerhalb des GKV-Systems zu ermitteln.

Das Versichertenklassifikationssystem von DxCG basiert auf den Informationen zu Alter, Geschlecht und Morbidität in Form von Diagnosen und/oder Arzneimittelverordnungen. Eine alleinige Berücksichtigung der Kosten der Vergangenheit eines Versicherten zur Prognose seiner zukünftigen Kosten respektive Morbidität oder Hospitalisierungswahrscheinlichkeit (Prior Cost, siehe Ash et al. 2001) ist ein nicht zielführender Ansatz. Grund hierfür ist ein Phänomen, das in der Literatur auch als „regression to the mean“ bezeichnet wird (Welch 1985). Darunter

versteht man den Effekt, dass sich sowohl die Kosten besonders teurer als auch die besonders günstiger Versicherter mit der Zeit zu einem Mittelwert hin entwickeln: Die (alleinigen) Kosten eines Versicherten sind also ein schlechter Prädiktor zukünftiger Kosten oder Krankheitszustände. Insbesondere hätte eine Beschränkung auf bestehende Hochkostenfälle zur Folge, dass die Gruppe derer, die noch nicht im Hochkostensegment angekommen sind, außer Betracht gelassen würde.

Neben der Grundfunktionalität der DxCG-Software, die Versicherten auf der Grundlage ihres Alters, Geschlechts und Morbiditätsinformationen zu klassifizieren, beinhaltet die Software in einem Zusatzmodul die Möglichkeit, eine Hospitalisierungswahrscheinlichkeit (Likelihood-Of-Hospitalisation – LOH) der einzelnen Versicherten zu prognostizieren. Auf der Basis des LOH-Modells ist es möglich, die Versicherten nach ihrer Hospitalisierungswahrscheinlichkeit absteigend zu sortieren und die Gruppe mit dem höchsten LOH-Wert als Gruppe mit der höchsten Hospitalisierungswahrscheinlichkeit gezielt anzusprechen.

Für eine vorausschauende Patientenbetreuung ist es jedoch nicht ausreichend, allein die Eintrittswahrscheinlichkeit eines relevanten Ereignisses zu kennen, sondern auch die beeinflussbaren Größen identifizieren zu können, um für diese entsprechende Maßnahmen abzuleiten. Insbesondere für die Entwicklung eines geeigneten Interventionsprogramms ist es hilfreich, das Zusammenspiel der einzelnen Faktoren und ihre jeweilige Relevanz identifizieren zu können. Die Regeln, die zur Angabe eines LOH-Wertes führen, sind innerhalb der Software nicht ersichtlich. Das Prognosesystem stellt sich somit für den Endanwender weitestgehend als Blackbox dar. Zudem ist das Modul der LOH-Bestimmung noch nicht auf das deutsche System angepasst worden und generiert wegen der geringen Hospitalisierungsquote in den USA zu optimistische LOH-Werte.

Nachteilig am Grundkonzept des LOH-Moduls der DxCG-Software ist, dass diese auf einer fest definierten Datenschnittstelle von

Eingabeparametern basiert. Eine Erweiterung der Eingabeparameter ist innerhalb der Software nicht definierbar und stellt somit eine starke Einschränkung dar. So ist es beispielsweise im LOH-Modul nicht möglich, soziodemografische Informationen wie den Bildungsstand eines Versicherten ebenfalls zu berücksichtigen.

Im Weiteren wird gezeigt, wie mithilfe der DxCG-Software – aber ohne Zuhilfenahme des LOH-Moduls – ein Prognosemodell zur Bestimmung einer Hospitalisierungswahrscheinlichkeit generiert werden kann. Die gewählte Vorgehensweise ermöglicht es, weitere Einflussgrößen frei zu definieren und als neue Eingabeparameter zu den Modellierungsdaten hinzufügen zu können.

In der Praxis ist es durchaus üblich, in einem Prognosesystem die Prognosewerte eines vorgelagerten Verfahrens als Prädiktor für eine nachgelagerte Modellierung zu nutzen. Somit hat man stets die Möglichkeit, die Ergebnisse einer Prognosesoftware zu verbessern, indem man mehrstufig vorgeht. Im Fall der Software DxCG würde dies bedeuten, dass der LOH-Wert der einzelnen Versicherten als neue Inputgröße innerhalb einer nachgelagerten Prognosemodellgenerierung mit einer anderen Prognosesoftware genutzt werden könnte.

Verfahren zur Generierung einer Hospitalisierungswahrscheinlichkeit

Eine Methodengruppe zur Generierung von Prognosemodellen umfasst die Verfahren des Data-Mining. Data-Mining-Algorithmen sind Verfahren, die für die Anwendung auf sehr große und komplexe Datenbestände hin optimiert wurden und daher geeignete Verfahren für ein Predictive Modeling im Kontext der Hospitalisierung darstellen (siehe Hand, Manila und Smyth 2001, Perry et al. 2004).

Ein Verfahren aus der Gruppe der Data-Mining-Verfahren sind Entscheidungsbäume. Diese bieten neben der Möglichkeit einer Prognosemodellgenerierung ein transparentes Regelwerk, das die Einflussfaktoren klar ablesbar macht und damit eine adäquate Basis zur Identifizierung

relevanter Einflussfaktoren für eine Hospitalisierung sein kann. Diese können dann auch Aufgreifkriterien eines Interventionskonzeptes sein. Entscheidungsbaumverfahren stellen genauso wie die regelbasierten Prognosemodelle, auf Basis von empirisch Boole'schen Regeln (Elsner et al. 2004), das Regelwerk durch Wenn-Dann-Regeln dar.

Der Unterschied dieser beiden Methodiken liegt in der Art und Weise der Regelgenerierung. Während regelbasierte Prognosemodelle durch die Nutzung von empirisch Boole'schen Regeln eine Abbildung von Expertenwissen sind und damit händisch definiert werden müssen, werden die Regeln der Entscheidungsbaumverfahren datengetrieben und damit autonom durch einen Algorithmus entwickelt. Somit ist es möglich, bisher unbekannte oder indirekte Zusammenhänge, die in den Daten verborgen sind und vorher noch nicht bekannt waren, als Regel aufzuzeigen.

Ein Vorteil der datengetriebenen Modellierung besteht darin, dass eine komplexe Wartung des Regelsystems entfällt. Eine Anpassung an neue Daten und Bedingungen erfolgt beim datengetriebenen Ansatz durch die Neugenerierung eines Prognosemodells mit all seinen Regeln.

Eine andere Verfahrensgruppe des Data-Mining generiert zwar keine transparenten Regelwerke als Resultat eines Prognosemodells, kann aber in einigen Fällen eine bessere Modellqualität generieren. Diese Verfahren sind dazu in der Lage, nichtlineare Zusammenhänge darstellen zu können. So entstehen Regeln, die nicht mehr durch klare Wenn-Dann-Regeln abbildbar sind. Kandidaten dieser Methodenklasse sind insbesondere die Neuronalen Netze (Lackes und Mack 2000), aber auch die logistische Regression. Letztere wird insbesondere zur Prognose von Eintrittswahrscheinlichkeiten angewandt (Bartfay, Mackillop und Pater 2006). Der gebräuchlichste Typ von Neuronalen Netzen sind die „Feedforward Multilayer Perceptrons“ (MLPs) (Hand, Mannila und Smyth 2001), welche dem biologischen Prinzip der Reizauswertung nachempfunden sind. Das Zusammenspiel der einzelnen durch den Algorithmus

des Neuronalen Netzes als relevant erachteten Attribute ist jedoch nur sehr schwer oder gar nicht nachvollziehbar. Ein Vergleich von logistischer Regression und Neuronalen Netzen basierend auf klinischen Daten findet sich in Behzad et al. (2005).

Wegen des dargestellten Vorteils der Entscheidungsbäume, ihre Regeln und Einflussfaktoren ablesen zu können, werden im Weiteren besonders die Entscheidungsbaumverfahren diskutiert. Um aber die Vorteile der nichtlinearen Verfahren, potenziell höhere Prognosegüte zu generieren, bewerten zu können, werden die Prognosegüten der Entscheidungsbaumverfahren mit denen der Neuronalen Netze verglichen. Zuvor werden die gängigsten in Frage kommenden Entscheidungsbaumverfahren skizziert.

Verfahren zur Generierung von Prognosemodellen

Entscheidungsbaumverfahren sind in vielen Softwareprodukten zur Datenanalyse anzutreffende Methoden und zur Generierung von Prognosemodellen geeignet. Die gängigsten Verfahren sind:

- CHAID – CHI-squared Automatic Interaction Detection (Biggs, De Ville und Suen 1991)
- CART – Classification And Regression Tree (Breiman et al.1984)
- C4.5/C5 – Nachfolger von ID3 (Quinlan 1993, Quinlan 1996)
- QUEST – Quick, Unbiased, Efficient, Statistical Tree (Loh und Shih 1997)

Das in den präsentierten Analysen genutzte Verfahren, der Microsoft Decision-Tree, ist in die Klasse der ID3-Verfahren einzuordnen. Daher besitzt das Verfahren wie C4.5 und C5 die Möglichkeit der Bildung nichtbinärer Verzweigungen und generiert damit eine kompakte Regelstruktur, die leicht interpretierbar ist. Des Weiteren wird auch bei diesem Algorithmus die Entropie zur Bewertung des Informationsgewinns genutzt (Tang und Mac Lennan 2005). Eine Eigenschaft des Entropiemaßes ist es, dass Attribute mit einer höheren Anzahl an Ausprägungen

tendenziell bevorzugt werden. Bei der Aufbereitung der Daten sollte man daher darauf achten, keine Attribute wie beispielsweise Alter und Geschlecht als RSA-Klasse zusammenzufassen, sondern getrennt vorzuhalten. Ein Algorithmus, der dieser Tendenz entgegenwirkt, ist der QUEST-Algorithmus, der allerdings den Nachteil besitzt, nur binäre Bäume bilden zu können.

Vorgehensweise zur Generierung eines Prognosemodells

Definition einer Datenbasis

Das Prognosemodell wird so aufgebaut, dass auf Basis der Versicherteninformationen des Jahres 2005 eine Hospitalisierung im Jahr 2006 prognostiziert werden soll. Die Prognose erfolgte auf einen Zeitraum von zwölf Monaten, um genügend Zeit für eine potenzielle Intervention bieten zu können. Insbesondere können durch die Betrachtung eines kompletten Kalenderjahres saisonale Effekte ausgeschlossen werden, die die Stabilität eines Prognosemodells beeinträchtigen können.

Das LOH-Modell von DxCG beruht – im Gegensatz zu dem hier gewählten Ansatz – auf einem Prognosezeitraum von lediglich sechs Monaten. Das wirft insbesondere die Frage auf, ob dieser Zeitraum, neben dem Nachteil saisonaler Schwankungen im Kalenderjahr, als Interventionszeitraum ausreichend ist. Der relativ geringe Prognosezeitraum von sechs Monaten ist auch deshalb problematisch, da die Diagnoseinformationen und Abrechnungsdaten erst mit einem gewissen Verzug vorliegen und der Interventionszeitraum dadurch zusätzlich geschmälert wird.

Die genutzte Datenbasis ist eine Stichprobe basierend auf dem Datenbestand der BARMER. Eine Stichprobenziehung erfolgte aus Gründen der Praktikabilität, um die Programmlaufzeiten der Verfahren zu verkürzen. Die Stichprobenziehung auf der Grundgesamtheit aller BARMER-Versicherten erfolgte analog zu dem Vorgehen der Generierung einer RSA-Stichprobe (VdAK 2007). Es wurden die Versicherten, die sowohl im Jahr 2005 als auch im Jahr 2006 mindestens einen Versichertentag besaßen,

betrachtet. Aus dieser Grundgesamtheit wurde dann eine Geburtstagsstichprobe gezogen. Die verwendeten Geburtstage entsprechen damit den Vorgaben des RSA (VdAK 2007). Die sich daraus ergebende Datenbasis beträgt 480.509 Versicherte.

Neben dem Alter, Geschlecht und der Information, ob bei dem Versicherten eine RSA-relevante Erwerbsunfähigkeit vorliegt, wurden die Diagnosen und Arzneimittelverordnungen des Jahres 2005 vorgehalten. Diese Daten sind auch die Eingabeparameter für eine Versichertenklassifikation mit der DxCG-Software. Diese Informationen wurden um folgende Versicherteninformationen des Jahres 2005 erweitert:

- Häufigkeiten der Arztbesuche nach Arztgruppen
- Identifikation der Fachgruppe des Hauptarztes (hier wird die Fachgruppe ausgewiesen, die die meisten Verordnungen durchführt)
- Ausweisung der Häufigkeit der diagnostischen Hilfsmittel Röntgen, MRT und CT

Die Anwendung von Data-Mining-Verfahren macht es notwendig, die Daten so vorzuhalten, dass alle Informationen eines Versicherten in einer einzigen Zeile vorliegen. Insbesondere bei kategorialen Attributen mit sehr vielen Ausprägungen ist ein Vorverarbeitungsprozess der Daten notwendig. Die Diagnosen können entweder innerhalb eines einzigen Attributes mit 15.000 Ausprägungen oder durch eine Aufteilung in 15.000 verschiedene Attribute (Spalten) für eine Modellierung genutzt werden.

Beide Ansätze sind wegen technischer Restriktionen oder Beschränkungen durch die genutzten Algorithmen nicht durchführbar. Daher wurde so vorgegangen, dass die Ausprägungen innerhalb eines Vorverarbeitungsprozesses gruppiert respektive zusammengefasst wurden. Ein Lösungsansatz besteht darin, sowohl die ICD-Ziffern als auch die Arzneimittelverordnungen zu gruppieren und diese dadurch in Gruppen zu bündeln, die in sich homogen sind. Dadurch ergibt sich eine handhabbare Menge von Ausprägungen.

Diese Aufgabe wird in dem vorgestellten Ansatz durch die Nutzung der Versichertenklassifikationssoftware von DxCG umgesetzt. Die durch die DxCG-Software gebildeten CCs für die Diagnosen und RxGroups für die Medikamentenverordnungen stellen einen notwendigen Vorverarbeitungsschritt zur Anwendung der gewählten Data-Mining-Algorithmen dar.

Zuerst wurden alle für eine Versichertenklassifikation (Grouping) erforderlichen Daten zusammengetragen. Diese bestanden neben Alter und Geschlecht aus den Versicherteninformationen der ambulanten und stationären Diagnosen sowie den Arzneimittelverordnungen des Jahres 2005. Das Grouping wurde aus technischen Gründen in zwei Phasen durchgeführt, zum einen mit den Informationen aus dem ambulanten Bereich und zum anderen mit denen aus dem stationären Bereich. Die Informationen der CCs wurden zu ambulant-stationären CCs zusammengefasst (logische Oder-Verknüpfung). Ebenso wurde ein Grouping der Arzneimittelverordnungen durchgeführt und RxGroups generiert.

Damit standen zu jedem Versicherten die folgenden Informationen zur Verfügung:

- 184 CCambstat (ambulante und stationäre CCs)
- 164 RxGroups (Apotheken-Informationen auf Basis der Pharmazentralnummern)
- Alter
- Geschlecht
- Information über Erwerbsunfähigkeit nach Alter und Geschlechtsklassen
- Abgeleitete Informationen aus den Versicherteninformationen (Anzahl der Arztbesuche, ...)

Die Anwendung eines Modells zur Prognose der Hospitalisierungswahrscheinlichkeit besteht in der Unterstützung eines Versorgungsmanagements. Dieses richtet sich aber ausschließlich an nicht akute oder durch

Unfall hervorgerufene Hospitalisierungen. Daher wurde es als sinnvoll erachtet, Versicherte, die im Jahr 2006 hospitalisiert wurden, aber typische akute oder durch Unfall erzeugte Krankenhaus-DRGs besitzen, aus der Modellierung herauszunehmen. Ferner wurden auch Versicherte mit einer Hospitalisierung durch eine Geburt im Jahr 2006 herausgefiltert.

Folgende DRGs des Jahres 2006 dienten als Ausschlussfilter:

- Entbindungen (O60 und O02)
- Polytrauma mit Eingriffen an Hüftgelenk, Femur, Extremität und Wirbelsäule (W02)
- Verletzung von Schulter, Arm, Ellenbogen, Knie, Bein und Sprunggelenk (I78)
- Komplexe Eingriffe am Kniegelenk (Kreuzbandriss) (I30)
- Verbrennungen (Y03)
- Neurologische Komplexbehandlungen des akuten Schlaganfalls (B39)
- Augenverletzung (C01)
- Herzinfarkt (F60)

Dadurch vermindert sich die genutzte Stichprobe von 480.509 auf 477.503 Versicherte und die Hospitalisierungsquote von 14,94 Prozent auf 14,41 Prozent.

Bildung von Trainings- und Validierungsmengen

Prognosemodelle werden mit Vergangenheitsinformationen auf einer Trainingsmenge generiert. Die Validierung ihrer Güte erfolgt ebenfalls auf Vergangenheitsinformationen, die aber nicht im Modellierungsprozess genutzt wurden. Haben sich die Modelle in der Validierung als stabil erwiesen, können die Prognosemodelle auf gegenwärtige oder zukünftige Versicherteninformationen angewandt werden.

Die Data-Mining-Algorithmen bauen ihre jeweiligen Prognosemodelle so auf, dass sich eine möglichst geringe Fehlerquote ergibt. Dies kann aber zur Folge haben, dass die Algorithmen die Informationen auswendig

lernen. Das Ergebnis könnte also sein, dass die auf einer Stichprobe generierten Modelle eine hervorragende Prognosegenauigkeit besitzen, aber bei der Anwendung auf neue, unbekannte Daten versagen. Dieser Effekt wird auch Overfitting genannt und ist sowohl bei den Entscheidungsbaumalgorithmen als auch bei den Neuronalen Netzen anzutreffen.

Die Algorithmen gehen dieses Problem auf unterschiedliche Art und Weise an. Die Entscheidungsbaumverfahren entwickeln dadurch eine starke Generalisierungsfähigkeit, indem sie die Komplexität, das bedeutet die Ebenentiefe, reduzieren. Dies geschieht entweder dadurch, dass die Bäume nicht weiter aufgespannt werden (beispielsweise CHAID) oder der Baum in einer zweiten Phase beschnitten wird, auch Pruning genannt (C4.5).

Bei den Neuronalen Netzen gehen die Algorithmen das Problem so an, dass das generierte Modell an einer abgetrennten Datenmenge auf seine Generalisierbarkeit geprüft wird. Diese Methode wird auch Stopped Training genannt, bei der abwechselnd Lernschritte und Testläufe mit disjunkten Mengen aus Trainings- und Testmustern durchgeführt werden, anstelle einer vollständig getrennten Lern- und Testphase wie bei den Entscheidungsbaumalgorithmen. Die Prognosegüte der Modelle beider Data-Mining-Verfahren wird daher immer an einer vom Modellierungsprozess unbeteiligten Menge validiert, um ein eventuelles Overfitting identifizieren und eine unabhängige Bewertung der Modellgüte vornehmen zu können.

Die Umsetzung der Prognosemodell-Generierung erfolgte daher nicht auf der gesamten Datenbasis, sondern auf einer 70-prozentigen Zufallsstichprobe (Trainingsmenge). Die übrigen 30 Prozent dienten der Evaluierung des generierten Modells (Evaluierungsmenge). Um dies zu erreichen, wurde die Versichertenpopulation in einer großen Tabelle vorgehalten und eine gleich verteilte Zufallszahl angefügt. Durch diese war es möglich, eine reproduzierbare Zufallsstichprobe zu ziehen.

Tabelle 1: Verteilung der Hospitalisierungen in der Stichprobe

| Hospitalisierung ohne Akutfälle und Geburten | absolut | Prozent |
|--|---------|---------|
| Gesamt | | |
| nein | 408.700 | 85,59 |
| ja | 68.803 | 14,41 |
| Summe | 477.503 | |
| Training | | |
| nein | 285.711 | 85,60 |
| ja | 48.053 | 14,40 |
| Summe | 333.764 | |
| Evaluierung | | |
| nein | 122.989 | 85,56 |
| ja | 20.750 | 14,44 |
| Summe | 143.739 | |

Modifizierung der Trainingsmenge

Es hat sich bereits in einigen Vorprojekten gezeigt, dass es bei einigen Algorithmen vorteilhaft ist, Prognosemodelle nicht mit der Originalverteilung zu trainieren, sondern das Modell auf einer geschichteten Stichprobe, die eine deutlich erhöhte Ereignisquote besitzt, zu trainieren. Diese Vorgehensweise ist insbesondere bei den Neuralen Netzen vorteilhaft. Dort wird die Trainingsmenge oft auf eine 50-Prozent-Eintrittswahrscheinlichkeit geschichtet. Auch bei den Entscheidungsbaumalgorithmen hat sich diese Schichtung schon oftmals als vorteilhaft erwiesen. Daher wurde in diesem Projekt nach einigen Versuchen ebenfalls eine Schichtung nahe 50 zu 50 gewählt (vergleiche Tabelle 2), was auch *Balanced Train* genannt wird (Berry und Linoff 2004: 68).

Tabelle 2: Verteilung der Hospitalisierung in der geschichteten Stichprobe

| Hospitalisierung ohne Akutfälle und Geburten | absolut | Prozent |
|--|---------|---------|
| Balanced Train | | |
| nein | 60.000 | 55,53 |
| ja | 48.053 | 44,47 |
| Summe | 108.053 | |

Ergebnisse

Auf der beschriebenen Datenbasis der Trainings- und Validierungsmenge wurden mit den beiden Data-Mining-Algorithmen Microsoft Decision-Tree und Microsoft Neural Network, die im Lieferumfang des Microsoft SQL-Server 2005 Enterprise Edition enthalten sind, Prognosemodelle generiert (Tang und Mac Lennan 2005).

Die Ergebnisdarstellung erfolgt sowohl durch Lift-Charts (oft auch Gains-Charts genannt) als auch durch die Darstellung von Trefferquoten der Prognosen mit den höchsten Hospitalisierungswahrscheinlichkeiten.

Da die Trefferquote die relative Häufigkeit der Hospitalisierungen (A-priori-Wahrscheinlichkeit) in der Grundgesamtheit nicht berücksichtigt, ist eine Bewertung der Modellgüte nicht allein auf Basis der Trefferquote möglich. Lift-Charts eignen sich besonders dazu, die Modellgüten zweier Modelle miteinander zu vergleichen und vor allem die Güte des Modells im Vergleich zu einer reinen Zufallsauswahl bewerten zu können. Der Lift-Chart (siehe Abbildung 2) zeichnet den kumulierten Anteil der mittels eines Modells korrekt erkannten Objekte (Kurve oberhalb der Winkelhalbierenden) im Vergleich zur Zufallsauswahl (Winkelhalbierende) ab.

Da man in der praktischen Anwendung eines Prognosemodells jedoch nicht alle, sondern nur die Gruppe mit den höchsten Hospitalisierungswahrscheinlichkeiten ansprechen möchte, konzentriert man sich beispielsweise auf das oberste eine Prozent der Versicherten mit der höchsten Hospitalisierungswahrscheinlichkeit. Dabei handelt es sich um die vom Prognosemodell mit der höchsten Konfidenz bewerteten Prognosen für eine Hospitalisierung. Die Quoten, die man bei der Berücksichtigung der am wahrscheinlichsten zu hospitalisierenden Versicherten erreicht, werden in den Abbildungen der Trefferquoten dargestellt.

Modell 1: Entscheidungsbaum 1

Auf Basis der Trainingsmenge wurde mithilfe des Entscheidungsbaumalgorithmus ein Prognosemodell entwickelt. Der generierte Entscheidungsbaum wurde in Abbildung 3 (Auszug) in der Gesamtübersicht mit einer Tiefe von sechs Ebenen dargestellt. Die Trefferquote des Modells von Entscheidungsbaum 1 weist für die Top-1-Prozent der Versicherten mit der höchsten Hospitalisierungswahrscheinlichkeit eine Trefferquote von etwa 47 Prozent auf. Interessant ist, dass die Trefferquote relativ langsam fällt (siehe Abbildung 1).

Im Lift-Chart in Abbildung 2 lässt sich die Modellqualität des Entscheidungsbaumes ablesen. Ein Lift-Chart zeigt die Modellqualität, durch die Darstellung der Fähigkeit des Modells die in der Validierungsmenge vorhandenen Hospitalisierungen möglichst genau der Gruppe der Versicherten mit den höchsten Hospitalisierungswahrscheinlichkeiten zuzuordnen. In der Evaluierungsmenge befinden sich 14,44 Prozent Hospitalisierungen. Die bestmögliche Modellgüte ist gegeben, wenn bei einer Ansprache von 14,44 Prozent der Versicherten 100 Prozent der Hospitalisierungen getroffen werden. Daher ist ein Modell umso besser, je dichter es sich diesem Bestwert annähert.

Abbildung 1: Trefferquoten Entscheidungsbaum 1

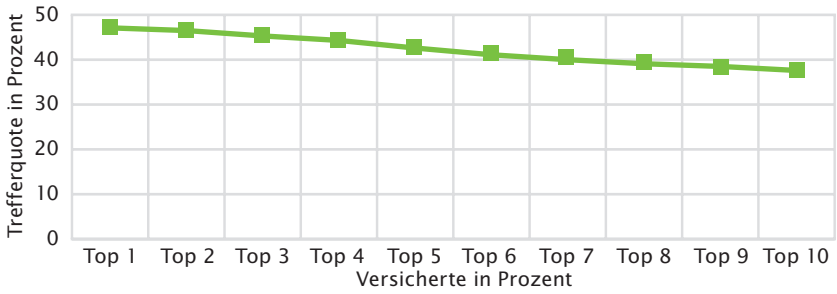


Abbildung 2: Lift-Chart Entscheidungsbaum 1

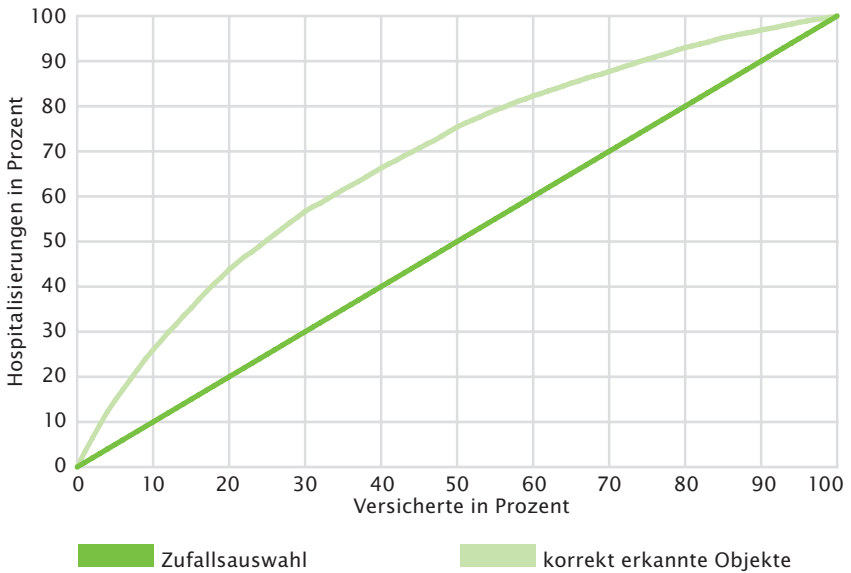
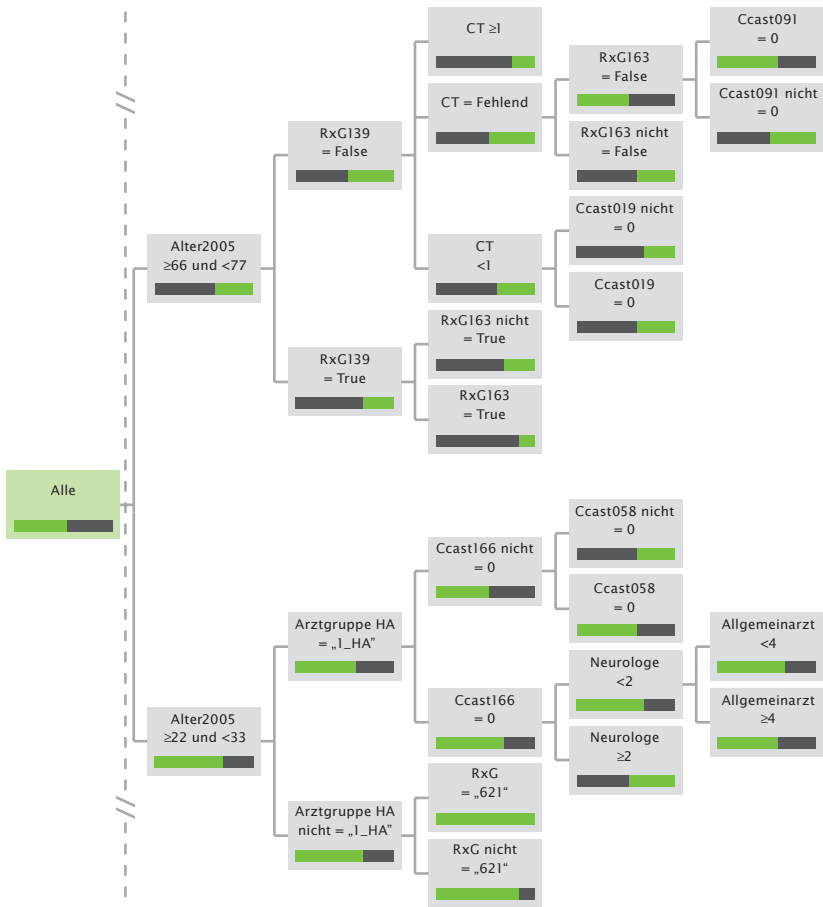


Abbildung 3: Ausschnitt aus einem Entscheidungsbaum (Baum 1)



Da die Attribute nicht nur binäre Ja-/Nein-Ausprägungen besitzen (wie etwa bei den CCs und RxGroups), sondern darüber hinaus auch metrische Attribute wie Alter und Anzahl der Arztbesuche vorliegen, werden durch den Algorithmus auch Intervalle gebildet und als Trennungskriterium in das Regelwerk des Baumes eingefügt.

Aus dem in Abbildung 3 (Auszug) dargestellten Entscheidungsbaum ergibt sich beispielsweise folgende Regel:

WENN Alter ≥ 66 und < 77 UND Ulcer/GERD (PPI) UND Ungruppierte ATCs UND Radiologie > 0 und ≤ 2 DANN Hospitalisierung im Kalenderjahr

Ein Entscheidungsbaum wird so aufgebaut, dass die Attribute, die eine besonders hohe Relevanz besitzen, näher zur Baumwurzel hin zu finden sind und die weniger relevanten Attribute näher Richtung Blätter vertreten sind. Einzelne Attribute können natürlich auf unterschiedlicher Bauebene in den einzelnen Pfaden auftauchen.

Tabelle 3: Attribute des Baumes 1 nach Relevanz sortiert (Übersetzungen des Autors)

| Rang | Attribut | Beschreibung |
|------|---------------|---|
| 1 | Alter | Alter der Versicherten im Jahr 2005 |
| 2 | RxG139 | Ulcer/GERD (PPI) |
| 3 | Röntgen | Anzahl der Röntgenuntersuchungen |
| 4 | Arztgruppe HA | Arztgruppe des „Hauptarzes“ |
| 5 | CT | Anzahl der CT-Untersuchungen |
| 6 | Allgemeinarzt | Anzahl der Besuche |
| 7 | Neurologe | Anzahl der Besuche |
| 8 | CC166 | Hauptsymptome, Abnormalitäten |
| 9 | Internist | Anzahl der Besuche |
| 10 | RxG163 | ungruppierte ATCs |
| 11 | CC019 | Diabetes-Mellitus mit keiner oder unspezifischer Komplikation |
| 12 | Kinderarzt | Anzahl der Untersuchungen |
| 13 | CC179 | postoperativer Zustand/Nachsorge/... |
| 14 | CC091 | Hypertension |
| 15 | CC036 | andere gastrointestinale Störungen |
| 16 | RGS | Regionalstruktur der BARMER, die dem Versicherten zugeordnet wird (PLZ-Gruppen) |
| 17 | CC058 | Depression |

Tabelle 3 zeigt die Attribute des Entscheidungsbaumes 1 nach ihrer Relevanz innerhalb des Regelmodells aufgelistet. Das Attribut, das vom Algorithmus als differenzierend identifiziert wurde, ist erwartungsgemäß das Alter der Versicherten. Danach folgt die RxG139 (Ulcer/GERD (PPI) Proton Pump Inhibitors).

Modell 2: Neuronales Netz 1

Auf derselben Trainingsmenge, die für die Generierung des Entscheidungsbaumes genutzt wurde, ist ein Neuronales Netz trainiert worden. Die Trefferquote stieg dadurch deutlich um fast zehn Prozentpunkte auf etwa 57 Prozent an (siehe Abbildung 4). Die Prognosegüte, im Lift-Chart ablesbar, stieg insbesondere im wichtigen unteren Versichertenanteil an (siehe Abbildung 5).

Abbildung 4: Trefferquote Neuronales Netz 1

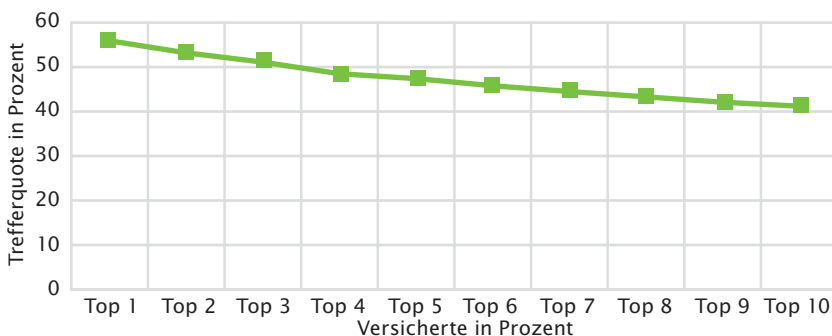
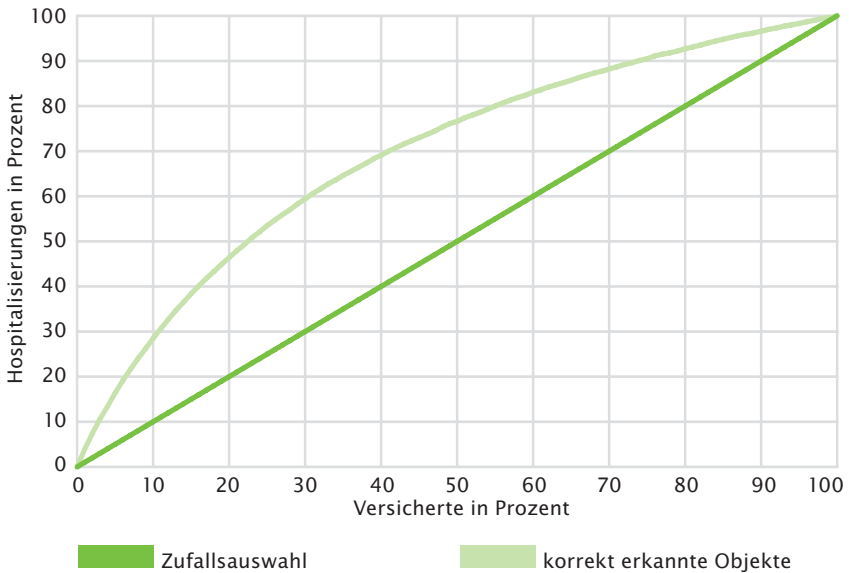
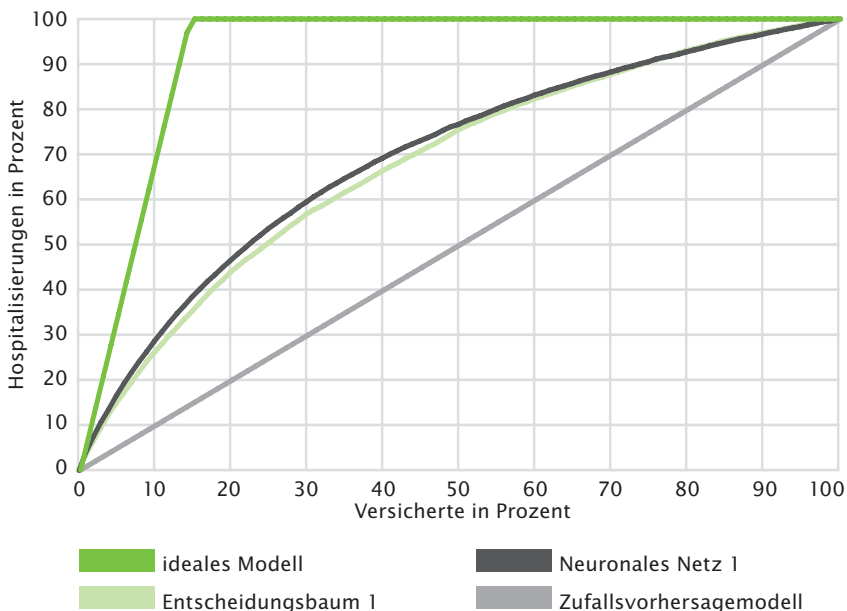


Abbildung 5: Lift-Chart Neuronales Netz 1

Die Modellgüten der beiden Prognosemodelle (Entscheidungsbaum und Neuronales Netz) lassen sich am besten innerhalb eines Lift-Charts miteinander vergleichen. Ein Vergleich der beiden Modelle ist in Abbildung 6 dargestellt und zeigt, dass das Neuronale Netz bis zu einem Versichertenanteil von etwa 60 Prozent eine höhere Prognosegüte besitzt.

Abbildung 6: Lift-Chart, Vergleich Entscheidungsbaum 1 gegenüber Neuronalem Netz 1



Weitere Einschränkung auf die Versichertenpopulation

In einer weiteren Analyse wurde die Prognosemodellgenerierung auf eine Versichertenpopulation eingeschränkt, die im Jahr 2005 nachfolgend aufgelistete Diagnosen besaß. Diese Diagnosen wurden vonseiten eines Versorgungsmanagements als relevant und als Indikator für Chroniker erachtet und ergänzen die in der ersten Modellierung gesetzten Filter:

- Diabetes-Mellitus-Typ-2 (E11 bis E14)
- Hypertonie (I10 bis I13 und I15)
- ischämische Krankheiten (I20 bis I25 und I27)
- Herzinsuffizienz (I50)
- bestimmte zerebrovaskuläre Krankheiten (I63 bis I66)

- Krankheiten der Arterien (I70 bis I74)
- chronische Krankheiten der unteren Atemwege (J44 bis J47)
- Rückenschmerzen (M54), unspezifischer Rückenschmerz

Tabelle 4: Verteilung der Hospitalisierungen ohne Akutfälle und Geburten unter Berücksichtigung der Einschlussdiagnosen

| Hospitalisierung | absolut | Prozent |
|--------------------------|---------|---------|
| Trainingsmenge | | |
| nein | 124.786 | 78,9 |
| ja | 33.332 | 21,1 |
| Summe | 158.118 | |
| Evaluierungsmenge | | |
| nein | 17.899 | 78,5 |
| ja | 4.891 | 21,5 |
| Summe | 22.790 | |
| Balanced Train | | |
| nein | 35.000 | 51,2 |
| ja | 33.332 | 48,8 |
| Summe | 68.332 | |

Modell 3: Entscheidungsbaum 2

Der generierte Entscheidungsbaum besitzt nun zwar eine Tiefe von acht Ebenen, zeigt sich aber deutlich kompakter und stellt damit eine geringere Anzahl an Regeln im Vergleich zu Modell 1 dar. Tabelle 5 zeigt die in das Modell eingeflossenen Attribute mit ihrer Relevanz, wobei sich ihre Anzahl um eins verringert hat. Einige Attribute wie beispielsweise die Depression sind nicht mehr aufgeführt. Neu hingegen ist die Diagnose CC084, welche bereits den siebten Rang belegt, obwohl sie im Auszug des Entscheidungsbaums 1 kein Bestandteil der Regelmenge ist. Eine Einschränkung der Datenbasis generiert somit nicht eine einfache Teilmenge eines Basismodells, sondern sucht immer die im jeweiligen Kontext stärksten Regeln.

Tabelle 5: Attribute des Baumes 2 nach Relevanz sortiert
(Übersetzungen des Autors)

| Rang | Attribut | Beschreibung |
|------|---------------|--|
| 1 | Alter | Alter der Versicherten im Jahr 2005 |
| 2 | RxG139 | Ulcer/GERD (PPI) |
| 3 | CC036 | andere gastrointestinale Störungen |
| 4 | Röntgen | Anzahl der Untersuchungen |
| 5 | CC166 | Hauptsymptome, Abnormalitäten |
| 6 | CC179 | postoperativer Zustand/Nachsorge/... |
| 7 | CC084 | koronare Atherosklerose/ andere chronische ischämische Herzerkrankungen |
| 8 | Chirurg | Anzahl der Untersuchungen |
| 9 | Radiologe | Anzahl der Untersuchungen |
| 10 | CC019 | Diabetes mit keiner oder unspezifischer Komplikation |
| 11 | CC167 | Begleitsymptome/Nebensymptome |
| 12 | Allgemeinarzt | Anzahl der Untersuchungen |
| 13 | Internist | Anzahl der Untersuchungen |
| 14 | Laborarzt | Anzahl der Untersuchungen |
| 15 | RxG163 | ungruppierte ATCs |
| 16 | Orthopäde | Anzahl der Besuche |

Die Prognosequalität und die Trefferquote sind aus Abbildung 7 und Abbildung 8 ablesbar. Die Verbesserung der Trefferquote um etwa sieben Prozentpunkte auf 54 Prozent im Vergleich zu der Trefferquote aus dem Entscheidungsbaum resultiert in erster Linie aus der höheren Hospitalisierungsquote und der damit verbundenen höheren A-priori-Wahrscheinlichkeit in der Evaluierungsmenge mit einer Steigerung um ebenfalls sieben Prozentpunkte. Die Modellqualität hingegen wird durch die Verbesserung der Versicherten Auswahl durch das Modell im Vergleich zur zufälligen Auswahl ermittelt. Es zeigt sich in den Lift-Charts, dass der Entscheidungsbaum 1 eine höhere Modellqualität besitzt als der alternative Entscheidungsbaum 2, da dessen Lift-Chart flacher verläuft.

Abbildung 7: Trefferquote Entscheidungsbaum 2

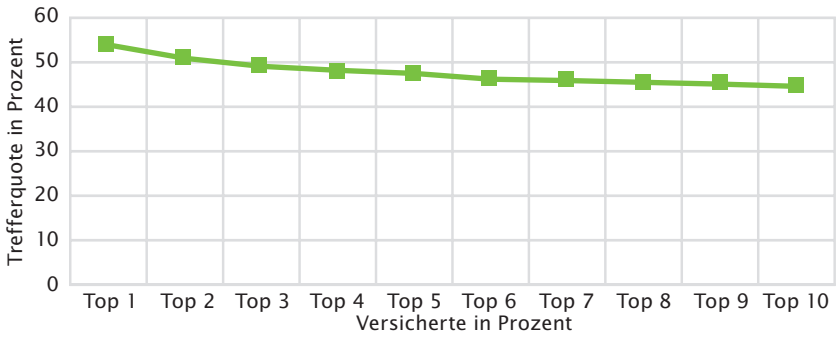
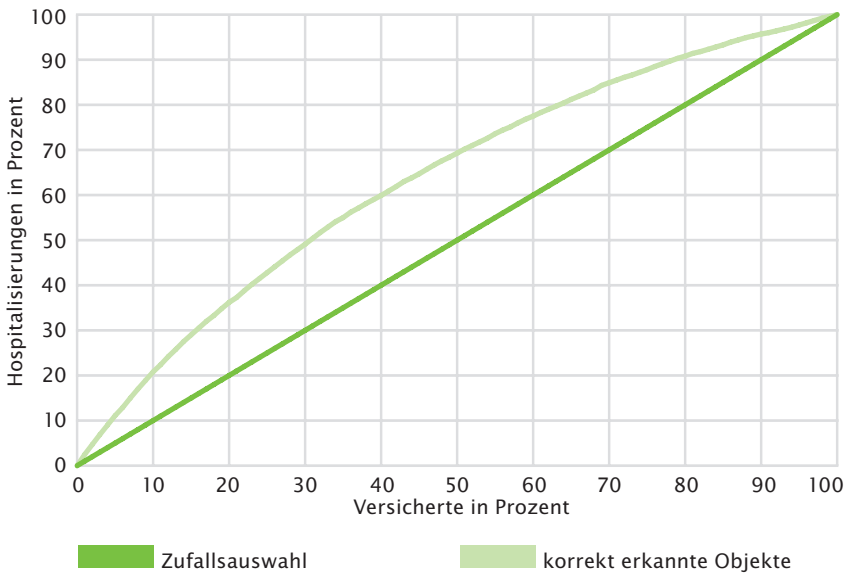


Abbildung 8: Lift-Chart Entscheidungsbaum 2



Dies ist insbesondere dadurch zu begründen, dass die Filterung auf Chroniker-Diagnosen bei den Regeln, die eine Nicht-Hospitalisierung prognostizieren, eine höhere Fehlerrate bewirkt.

Modell 4: Neuronales Netz 2

Auch im Kontext der Diagnoseeinschränkungen wurde ein Prognosemodell auf Basis eines Neuronales Netzes gebildet. Ein Vergleich der Prognosequalität des Entscheidungsbaumes mit dem Neuronales Netz zeigt, dass die Neuronales Netze auf dieser Datenbasis eine deutliche Verbesserung erreichen konnten. Die Trefferquote konnte um 13 Prozentpunkte auf 67 Prozent angehoben werden (siehe Abbildung 10). Ferner zeigte auch der Lift-Chart eine deutliche Verbesserung gegenüber dem Entscheidungsbaum 2 (siehe Abbildung 9).

Abbildung 9: Lift-Chart, Vergleich Entscheidungsbaum 2 gegenüber Neuronalem Netz 2

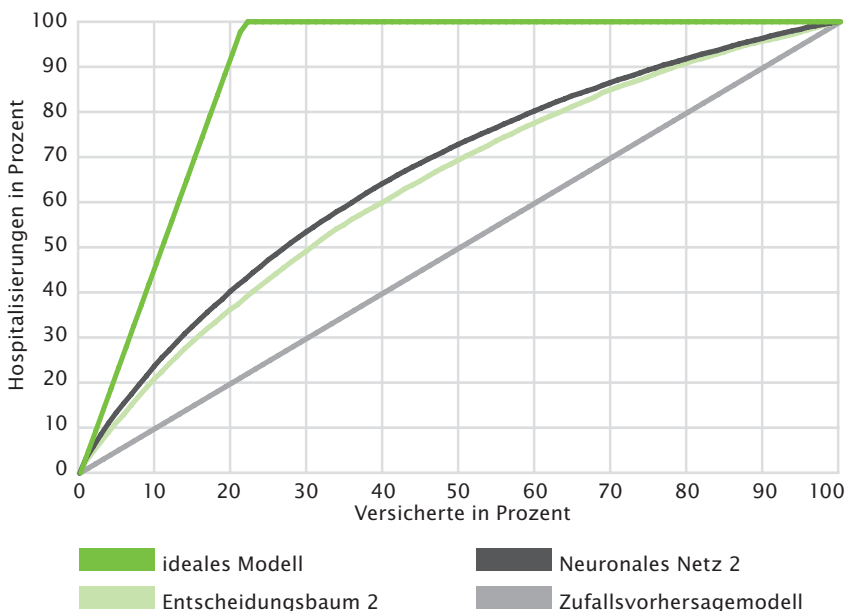


Abbildung 10: Trefferquote Neuronales Netz 2

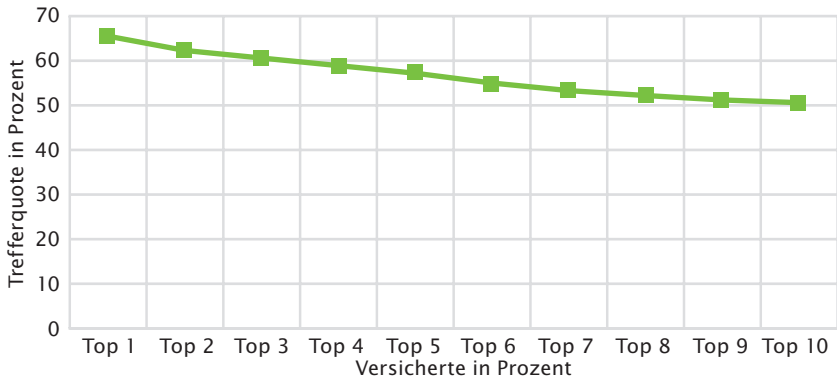
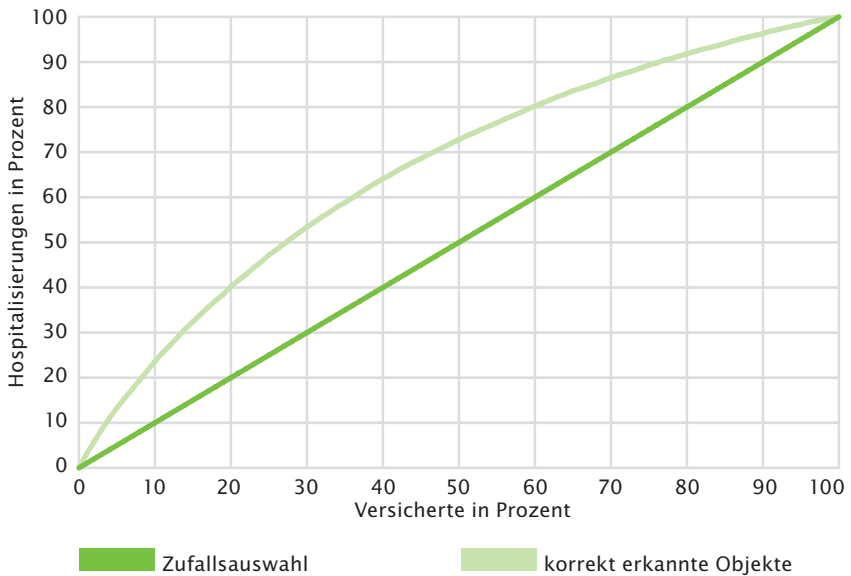


Abbildung 11: Lift-Chart Neuronales Netz 2



Fazit und Ausblick

Die dargestellte Vorgehensweise zur Prognosemodellgenerierung zeigt, dass es mit Verfahren des Data-Mining und auf Basis einer Versichertenklassifikationssoftware möglich ist, gute Prognosemodelle zu generieren. Die höchste Trefferquote der vorgestellten Modelle liegt bei 67 Prozent für ein Prozent der Versichertenpopulation mit der höchsten Hospitalisierungswahrscheinlichkeit. Durch die Nutzung dieses Prognosemodells ist ein mehr als dreifach besseres Ergebnis im Vergleich zu einer Auswahl auf der Basis allgemeiner Diagnosefilter (hier den Chroniker-Diagnosen) erzielt worden. Die Evaluierung der Modelle hat gezeigt, dass diese stabil sind und daher auf Nachfolgejahre angewandt werden können. Dadurch können die Versicherten in Zukunft mit einer hohen Kosteneffizienz angesprochen und einer Steuerung zugewiesen werden.

Die generierten Prognosemodelle sind von den Restriktionen, wie sie durch fest implementierte Module zur Prognose von Hospitalisierungswahrscheinlichkeiten gesetzt werden, weit weniger betroffen. So ist es von sehr großem Vorteil, die Inputgrößen der Modellierung selber festlegen zu können und auch selbst die Zeithorizonte des Betrachtungs- und Prognosezeitraumes bestimmen zu können. Damit ist man insbesondere in der Lage, Informationen zu berücksichtigen, die in dem derzeit noch auf amerikanische Verhältnisse hin entwickelten LOH-Prognosemodul der DxCG-Software nicht implementiert wurden.

So hat sich insbesondere in den durch Entscheidungsbaumverfahren generierten Modellen gezeigt, dass die Kriterien der Anzahl an Arztbesuchen und Diagnostiken einen relevanten Einfluss auf die entstandenen Regelwerke gefunden haben. Dies bestätigt, dass die Erweiterbarkeit der Attribute der Prognosebasis sinnvoll ist. Insbesondere zeigten die durch die Entscheidungsbäume generierten Modelle die gewünschte Regeltransparenz.

Zur Verbesserung der Trefferquoten hat es sich als günstig erwiesen, neben der Modellierung mit linearen Verfahren wie den Entscheidungsbäumen auch nichtlineare Verfahren wie die Neuronalen Netze zu nutzen. Den Nachteil, dass nichtlineare Verfahren intransparente Regelstrukturen bilden, kann man entweder dadurch kompensieren, dass man den Prognosewert (LOH-Wert) als zusätzlichen Input für ein Entscheidungsbaumverfahren nutzt oder aber Verfahren wählt, die trotz eines nichtlinearen Ansatzes die Regeltransparenz optimieren. Die Firma Prudential Systems hat beispielsweise eine Software entwickelt, die in der Lage ist, einen Entscheidungsbaum mit nichtlinearen Verzweigungen zu bilden und darzustellen.

Es ist zu prüfen, wie die Abbildung medizinischer Kenntnisse die Datenbasis weiter anreichern kann. So erscheint es sinnvoll, bei bestimmten Medikamenten die Verordnungsfrequenz und die DDD in geeigneter Weise zu klassifizieren und diese als Inputgrößen zu berücksichtigen.

Bei der Umsetzung innerhalb eines Versorgungsmanagements sollte man einen geschlossenen Regelkreis aufbauen, in dem die Ergebnisse eines Versichertenkontaktes zurück in die Datenbasis geschrieben werden. Ein Beispiel dafür wäre, die Informationen über Behandlungsabbrüche oder den Grad an Kooperationsbereitschaft als neues Versichertenkennzeichen in die Datenbasis zurückzuspielen. Dieser im analytischen CRM auch Closed Loop genannte Ansatz bewirkt, dass sich die Modellgüte stetig verbessern kann. Dies resultiert daraus, dass das Feedback aus dem Versichertenkontakt in die nächste Modellierung einfließen kann und somit eine noch stärkere Konzentration der Kräfte auf eine Erfolg versprechende Versichertengruppe stattfindet. Dies korrespondiert vor allem mit dem Ziel, eine Steigerung der Erfolgsquote durch die Verbesserung der Versichertencompliance zu erreichen.

Literatur

- Ash, A., Y. Zhao, E. Randall und M. S. Kramer (2001): Finding Future High-cost Cases: Comparing Prior Cost Versus Diagnostic-based Methods. In: Health Service Research. Jahrgang 36. Nummer 6. Teil II. S. 194-206.
- Bartfay, E., W. J. Mackillop und J. L. Pater (2006): Comparing the predictive value of neural network models to logistic regression models on the risk of death for small cell lung cancer patients. In: European journal of cancer care. Jahrgang 15. Nummer 2. S. 115-124.
- Behzad, E., K. Mohammad, H. E. Ardebili, M. Ghodsi und E. Ketabchi (2005): Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. www.biomedcentral.com/1472-6947/5/3 (27.06.2008).
- Berry, M. J. A., und G. S. Linoff (2004): Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 2nd Edition. Wiley.
- Biggs, D., B. De Ville und E. Suen (1991): A method of choosing multiway partitions for classification and decision trees. In: Journal of Applied Statistics. Jahrgang 18. Nummer 1. S. 49-62.
- Breimann, L., J. H. Friedman, R. A. Olshen und C. J. Stone (1984): Classification and regression trees. Belmont, CA: Wadsworth.
- Elsner, C., D. W. Plocher, S. Roze und C. Mazzi (2004): „Predictive Modeling“ – Werkzeug für effizientes Disease Management. In: Deutsches Ärzteblatt. Jahrgang 101. Heft 24. A-1724/B-1434/C-1381.
- Hand, D., H. Mannila und P. Smyth (2001): Principles of Data Mining. MIT Press. Cambridge, MA.
- Kudyba, S., G. B. Hamar und W. M. Gandy (2006): Utilising neural network-applications to enhance efficiency in the healthcare industry: predicting populations of future chronic illness. In: IJBIDM. Jahrgang 1. Nummer 4. S. 371-383.
- Lackes, R., und D. Mack (2000): Neuronale Netze in der Unternehmensplanung: Grundlagen, Entscheidungsunterstützung/Projektierung. München: Vahlen.

- Loh, W. Y., und Y. S. Shih (1997): Split selection methods for classification trees. In: *Statistica Sinica*. Nummer 7. S. 815-840.
- Magidson, J. (1992): Chi-squared analysis of a scalable dependent variable. In: *Proceedings of the 1992 Annual Meeting of the American Statistical Association, Educational Statistics Section*.
- Perry T. H., T. Tucker, L.R. Hudson, W. Gandy, A. L. Neftzger und G. B. Hamar (2004): *The Application of Data Mining Techniques in Health Plan Population Management: A Disease Management Approach*. In: S. Kudyba (Hrsg.): *Managing Data Mining. Advice from Experts*. Idea Group Publishing. S. 135-153.
- Quinlan J. R. (1996): Improved use of continuous attributes in c4.5. In: *Journal of Artificial Intelligence Research*. Nummer 4. S. 77-90.
- Quinlan, J. R. (1993): *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.
- Reschke, P., S. Sehlen, G. Schiffhorst, W. F. Schröder, K. W. Lauterbach und J. Wasem (2004): *Klassifikationsmodelle für Versicherte im Risikostrukturausgleich. Endbericht. Untersuchung zur Auswahl geeigneter Gruppenbildungen, Gewichtungsfaktoren und Klassifikationsmerkmale für einen direkt morbiditätsorientierten Risikostrukturausgleich in der gesetzlichen Krankenversicherung. Im Auftrag des Bundesministeriums für Gesundheit und soziale Sicherung*. Berlin, Köln und Essen.
- Tang, Z. H., und J. Mac Lennan (2005): *Data Mining with SQL Server 2005*. Wiley.
- Verband der Angestellten-Krankenkassen e.V. – VdAK Vereinbarung der Spitzenverbände nach § 267 Absatz 7 Nummer 1 und 2 SGB V in der Fassung vom 3. Juli 2007.
- Welch, W.P. (1985): Regression toward the Mean in Medical Care Costs: Implications for Biased Selection in Health Maintenance Organizations. In: *Medical Care*. Jahrgang 23. Nummer 11. S. 1234-1241.