

## Gradient Boosting vs. Deep Learning. Möglichkeiten des Einsatzes Künstlicher Intelligenz im Banking

### Key Facts

- › Steigender Wettbewerbsdruck für Finanzinstitute durch digitale und individuellere Angebote
- › Kundenabwanderung als manifestes Problem
- › Aktuell geringe Analyse- und Reaktionsmöglichkeiten
- › Neue Analyse und Handlungsmöglichkeiten durch Machine Learning
- › Nachweis des Vorteils von Gradient Boosting gegenüber Deep Neural Networks im Falle des Churn Case durch COREai

### Report

Das Retailbanking befindet sich im Umbruch. Neue Player im Markt sowie die Änderung aufsichtlicher Rahmenbedingungen (bspw. PSD 2) bedeuten eine marktbedingte sowie regulatorische Öffnung des Marktes und stellen klassische Banken vor neue Herausforderungen. Auch das Kundenverhalten hat sich entscheidend verändert. Die Bereitschaft der Kunden, das Konto zu wechseln, hat sich in den letzten vier Jahren verdoppelt, wobei die absolute Anzahl an Konten im deutschen Markt über diesen Zeitraum hinweg konstant blieb. Das traditionelle Instrumentarium der Finanzinstitute bietet aktuell kaum Möglichkeiten, um auf die Effekte dieses Wandels zu reagieren.

Für die beschriebenen Herausforderungen stellt Künstliche Intelligenz diverse Lösungsmodelle zur Verfügung, von denen im Folgenden zwei entlang eines konkreten Use Cases besprochen werden sollen. Fokus dieses Blogposts ist der sogenannte „Churn Case“, bei dem anhand von Kundenstamm- und Transaktionsdaten abwanderungswillige Kunden sowie nach Möglichkeit die damit in Zusammenhang stehenden Gründe identifiziert werden sollen, um Maßnahmen für die nachhaltige Kundenbindung ergreifen zu können.

### Reaktionsmöglichkeiten der Banken

Im Zuge der zuvor beschriebenen Herausforderungen bieten sich den Banken verschiedene Reaktionsmöglichkeiten, auf einen abnehmenden Kundenstamm zu reagieren.

---

<sup>1</sup> Bezeichnung von Anwendungsfällen zur Ermittlung abwanderungswilliger Kunden

**a) Akquise von Neukunden**

Der Fokus der Geschäftsstrategie liegt auf der Gewinnung neuer Kunden. Dem Verlust bestehender Kunden wird nicht entgegengewirkt.

**b) Ansprache von bereits gekündigten Kunden**

Die zweite Variante legt den Fokus auf die Rückgewinnung von Kunden, die bereits gekündigt haben, sich aber aufgrund rechtlich gesetzter Fristen noch in einem Geschäftsverhältnis mit der Bank befinden. Dies bedeutet, ehemalige Kunden mit attraktiven Neukundenangeboten zu incentivieren, der Bank treu zu bleiben.

**c) Identifikation abwanderungswilliger Kunden**

Im Rahmen der dritten Variante werden potentielle Kündigungen sowie die zugrundeliegende Motivation antizipiert und im Vorfeld durch bedarfsgerechte Betreuung adressiert. Grundlage hierfür bildet ein Modell, das eine Identifikation unzufriedener Kunden anhand der verfügbaren Datengrundlage ermöglicht.

In der Praxis wird in den meisten Fällen eine Kombination der genannten Optionen gewählt. Dies steht allerdings in keinem Widerspruch dazu, dass für Option c) Aufwand zur Schaffung einer Informationsgrundlage und Erlangung von Handlungsfähigkeit betrieben werden muss, welcher den Fokus dieses Blogposts darstellt.

**Herausforderungen**

Laut Expertenschätzungen ist die Neukundenakquise etwa fünfmal so teuer wie das Halten eines Bestandskunden. Um die Option, abwanderungswillige Kunden zu identifizieren allerdings zu verwirklichen, stehen Banken vor zwei zentralen Herausforderungen.

- 1) Korrekte und zeitgerechte Prognose, d.h. Antizipation des Kündigungswunsches
- 2) Korrekte Ermittlung der Beweggründe für einen Wechsel des Kontoanbieters

Häufig erlangen die Banken die Information über eine Kündigung erst, wenn diese bereits vorliegt. Wissen über die Beweggründe verliert zu einem späteren Zeitpunkt an Wert, da eine Teilmenge der Handlungsoptionen nicht mehr verwirklichbar ist. Beispiele für Wechselgründe des Kunden können neben den naheliegenden, in den Medien diskutierten Gründen wie der Wechsel zu einem günstigeren Anbieter, z.B. Direktbanken, mangelnde Erreichbarkeit der Filialen, schlechte Erfahrungen mit der

Kundenberatung, zu wenige Geldautomaten im näheren Umkreis oder ein unzureichendes Produktportfolio sein.



**Privatkundenmarkt und insbesondere die traditionellen Institute leiden unter der steigenden Zahl von Kunden, die die Bank wechseln**

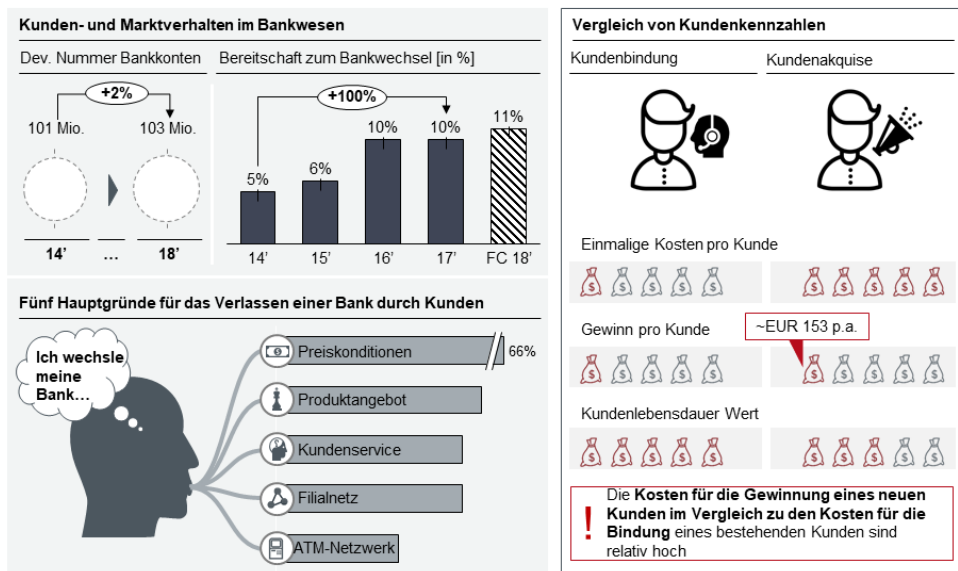


Abbildung 1: Kundenbewegungen im deutschen Retail-Banking

**Neue Handlungsmöglichkeiten mit Hilfe von KI**

Modelle des Machine Learning können dazu genutzt werden, die oft in Banken schon aus regulatorischen Gründen vorhandene Datengrundlage auszunutzen, um ein Prognosemodell zur Kundenabwanderung zu trainieren. Klassifikationsmodelle werden mit zunehmendem Erfolg dazu eingesetzt, auch hochdimensionale Muster (Muster über eine große Menge von Variablen), die oft zu komplex für menschliche Analysekapazitäten sind, zu identifizieren und den Zusammenhang mit beispielsweise einem Kündigungsereignis zu erkennen.

Die beschriebene Problemstellung wurde auch von der Banco Santander als entscheidend identifiziert. So veröffentlichte die spanische Bank bereits im Jahr 2016 auf Kaggle eine öffentliche Ausschreibung für Datenanalysten weltweit. Der zur Verfügung gestellte, vollständig anonymisierte Datensatz umfasste 76.000 Kunden, von denen etwa 3.000 ihr Konto kündigten. 370 Variablen charakterisierten die Kunden, wobei aus dem Datensatz, aufgrund der Anonymisierung, nur für die Banco Santander eine Rückführung auf die realen Umstände möglich war. Ziel der Ausschreibung war die möglichst präzise Vorhersage der Kündigung eines Kunden, wobei die korrekte

Vorhersage von kündigenden Kunden stärker bei der Bewertung der Lösung ins Gewicht fiel als die korrekte Prognose eines in der Bank verbleibenden Kunden. Der Datensatz wurde im Zeitraum von zwei Monaten durch mehr als 5.000 Datenanalyse-Teams analysiert.

### Vergleich Deep Learning mit Gradient Boosting

Auch das COREai Team analysierten die ausgeschriebenen Daten. Als Grundlage für die spätere Diskussion wird die Analysestrategie im Folgenden näher beleuchtet.

#### 1. Vorbereitung: Datenbereinigung und Modellierung

Die Vorbereitung einer jeden Analyse sieht die Datenbereinigung sowie –modellierung vor. In einem ersten Schritt wurde, unter Nutzung eines Random Forest Algorithmus, die Vielfalt der Variablen reduziert, um solche mit geringen Informationsgehalten zu eliminieren. Weitere Datenbereinigungen, wie sie üblicherweise in Datenanalyseprojekten anfallen würden, waren nicht notwendig, da die Daten bereits in einem verwertbaren Zustand vorlagen. Auf der bereinigten Datengrundlage wurden die zwei Modelle Gradient Boosting mit Entscheidungsbäumen und Deep Learning getestet und ihre Performance verglichen. Um das Verständnis für die Funktionsweise der Algorithmen zu verbessern, wird im Folgenden eine grobe Erläuterung der Modelle gegeben.

#### 2. Gradient Boosting- und Deep Learning- Modelle im Detail

Deep Neural Networks sind sogenannte Neuronen (nicht-lineare Rechenoperationen) angeordnet in einer geschichteten Architektur, bei denen die Information Schicht für Schicht abstrahiert wird, wobei für die Klassifikation wichtige Features aus den Daten extrahiert werden. Durch diese schrittweise Abstraktion wird es möglich, auch sehr hochdimensionale Information zu verarbeiten. Dank ihrer flexiblen Architektur sind diese Modelle in der Lage, verschiedene Quellen von Informationen in einem Modell zu nutzen und auch hochkomplexe Muster zu erkennen.

Das Gradient Boosting ist eine Strategie, bei der mehrere für einen Teilbereich des Lösungsraums performante Einzelmodelle zu einem über den gesamten Lösungsraum performanten Gesamtmodell gewichtet zusammengeschlossen werden. Ziel hierbei ist, die Zusammenarbeit im Ensemble auszunutzen, wobei die Einzelmodelle so erstellt werden, dass sie die Defizite der bereits trainierten Teilmodelle ausgleichen. Als Einzelmodell werden meist Entscheidungsbäume eingesetzt, die gewichtet ins Gesamtergebnis eingehen.

**Gradient Boosting ist ein Metamodell, das eine Reihe von schwachen Klassifikatoren kombiniert, um einen starken Klassifikator zu erstellen**

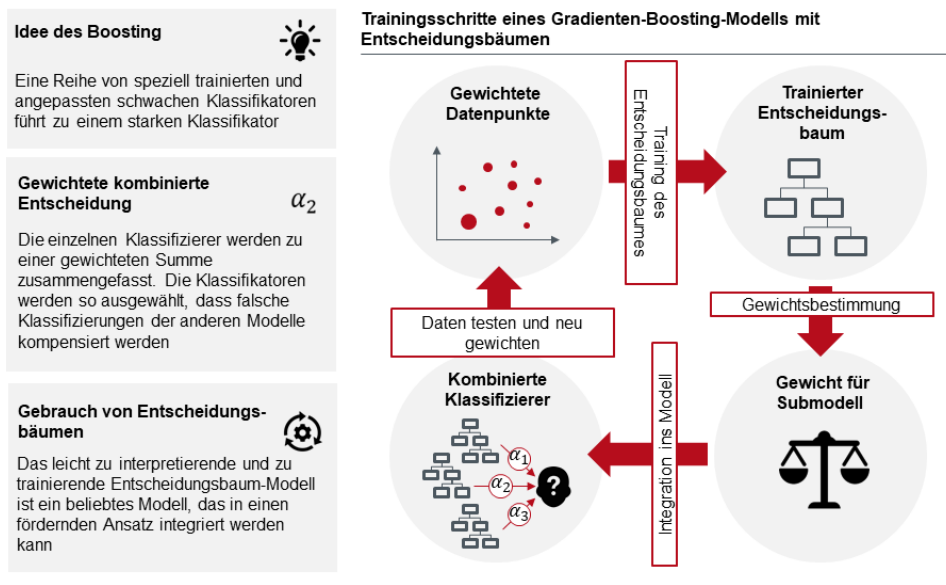


Abbildung 2: Gradient Boosting im Modell

**3. Ergebnis 1: Vorhersagegenauigkeit und Performance**

Die beiden vorgestellten Modelle wurden hinsichtlich ihrer Vorhersagegüte mittels der Missklassifikationsrate - dem Anteil der falsch prognostizierten Datenpunkte, respektive Kunden - getestet, wobei eine korrekte Klassifikation eines kündigenden Kunden höher als die eines verbleibenden Kunden gewichtet wurde. Um eine bessere intuitive Interpretierbarkeit der Modelle in diesem Artikel zu gewährleisten, werden die Ergebnisse in ungewichteter Form angegeben. In den Testläufen kam das Deep Neural Network in 23,0% der Fälle zu einer fälschlichen Prognose. Das Gradient Boosting-Modell mit Entscheidungsbäumen kam nur in 3,3% der Fälle zum falschen Ergebnis.

Neben der besseren Vorhersagegenauigkeit überzeugte das Gradient Boosting Modell mit kürzeren Lern- und Vorhersagezeiten. Im betrachteten Testszenario führte dieser Umstand aufgrund des recht kleinen Datensatzes zu keinem ausschlaggebenden Selektionsgrund, doch fällt bei größeren Datensätzen dieser Aspekt zunehmend ins Gewicht (Stichwort Big Data).

#### 4. Ergebnis 2: Interpretierbarkeit

In Anbetracht dessen, dass im Kündigungsfall neben der richtigen Prognose auch das richtige Angebot gewählt werden muss, stellt die Interpretierbarkeit von Modellen einen nicht zu vernachlässigenden Vorteil dar. Die Stärke von Deep Learning Modellen, ihrer in Tiefe und Breite skalierbaren Architektur, führt bedauerlicherweise im Gegenzug zu Abstrichen hinsichtlich der Interpretierbarkeit – trotz deutlicher Fortschritte auf diesem Themengebiet. Gradient Boosting hingegen bietet aufgrund der guten Interpretierbarkeit sowohl der individuellen Entscheidungsbäume als auch des gewichteten Ensembles eine gute Interpretierbarkeit des Gesamtmodells, welche sich gerade im Anforderungsumfeld des Bankings als Vorteil herausstellt. Durch Nutzung eines interpretierbaren Modells kann möglicherweise neben der Prognose von Kündigungen ein Rückschluss auf die Gründe der Kündigung gezogen werden. Auch wenn wir in unseren Test dies gerne am Datensatz der Banco Santander getan hätten, konnten aufgrund der Anonymisierung keine derartigen Erkenntnisse erlangt werden.

#### Fazit

Gradient Boosting hat sich für den beschriebenen Datensatz und Use Case als die vielversprechendere Lösungsoption gegenüber Deep Learning erwiesen. Neben den besseren Analyseergebnissen punktet das Gradient Boosting mit besserer Interpretierbarkeit, welche dem Anwender eine verbesserte Möglichkeit bietet, die Entscheidungsfindung des Algorithmus nachzuvollziehen und somit die Handlungsfähigkeit im Kündigungsfall begünstigt.

Eine allgemeine Aussage zur Dominanz einer der beiden Algorithmen kann jedoch nicht getroffen werden, da diese von den in den Daten enthaltenen Mustern abhängig ist. Beispielsweise sind Deep Neural Networks auf dem Gebiet der Bild-, Sprach- und Schrifterkennung weiterhin unangefochten. Es gilt nun im Einzelfall festzustellen, welche Informationen in den vorhandenen Daten tatsächlich enthalten sind und somit ob, wie schon bei der Banco Santander, eine neue Art des Kundenservice ermöglicht werden kann.

**Quellen**

<https://www.wiwi-treff.de/Fuehrung-and-Strategie/Kundenmanagement/Beziehungspflege-zahlt-sich-aus-Neukundengewinnung-fuenfmal-so-teuer/Artikel-4977>

Hadden J., Tiwari A., Roy R., Ruta D. (2005) "Computer assisted customer churn management: State-of-the-art and future trends", *Computers & Operations Research* 34, pp. 2902-2917

<https://www.wiwo.de/unternehmen/banken/studie-warum-die-deutschen-ihre-bank-wechseln/11478036.html>

[www.kaggle.com](http://www.kaggle.com)

<https://www.kaggle.com/c/santander-customer-satisfaction>

<http://konto-report.de/girokonto/#girokonto>

<https://de.statista.com/statistik/daten/studie/187567/umfrage/anzahl-der-girokonten-bei-deutschen-kreditinstituten-seit-2002/>

<https://yougov.de/news/2017/04/05/das-wechselkarussell-im-bankensektor-nimmt-fahrt-a/>

<https://de.statista.com/statistik/daten/studie/652802/umfrage/gruende-von-bankkunden-fuer-einen-wechsel-des-girokontos/>

<http://www.tagesspiegel.de/advertorials/ots/a-t-kearney-deutsche-privatbanken-bleiben-im-europaeischen-wettbewerb-weiter-zurueck/20010868.html>