

# Customer Lifetime Value

Der Kundenlebenswert als zentrale Kennzahl für die unternehmensseitige Aussteuerung der Kundenbeziehung

Dr. Steffen Wagner



Der Customer Lifetime Value (kurz: CLV) oder auch "Kundenlebenswert" ist der gesamte Wert, den ein Kunde während seines Kundendaseins ("Kundenleben") für ein Unternehmen generiert. Diese enorm wichtige Kennzahl wird in vielen Unternehmen abteilungsübergreifend genutzt, um Ressourcen umsatzverursachungsgerecht zu verteilen und Gewinne zu optimieren. Für die Modellierung des CLVs werden nicht nur historisch und aktuell generierte Umsätze in Betracht gezogen. Der schwierige Teil besteht darin, auch zukünftige Umsätze auf Basis von Prognosen einfließen zu lassen.

## Zielstellung des Customer Lifetime Value

### Differenzierung von Kunden

Welcher Kunde ist für mich rentabel und welcher kostet mich unterm Strich Geld? Diese interessante Frage datengetrieben zu beantworten stellt das wesentliche Ziel des CLVs dar. Dabei geht es ausdrücklich nicht allein um den jetzigen Zeitpunkt oder die zurückliegende Kundenbeziehung, sondern auch die zukünftige. Häufig werden Annahmen über die Rentabilität von Kunden getroffen, die dem Bauchgefühl entspringen. Eine solche Annahme ist beispielsweise, dass Kunden, die besonders häufig bestellen, auch besonders rentabel sind. Das ist bei weitem nicht immer der Fall. Der CLV hilft Ihnen, solche Trugschlüsse zu vermeiden und Informationen zu erhalten, die auf tatsächlichen Daten und Fakten basieren.

### Gezielte Aussteuerung des Customer Relation Management

Sobald Sie wissen, welche Kunden für Sie wie rentabel sind, können Sie Ihr Customer Relation Management entsprechend aussteuern. Unrentable Kunden können zum Beispiel von der Ansprache oder von bestimmten Services ausgeschlossen werden, um die entsprechenden Ressourcen an anderer Stelle gewinnbringender einzusetzen. Sie können sich ganz den Kunden widmen, die Ihnen unterm Strich auch einen Gewinn generieren.

### Umsatzpotential ausschöpfen

Service ist teuer. Doch Service ist ein ausschlaggebender Punkt für die Zufriedenheit unserer Kunden. Und viele unserer bereits rentablen Kunden haben das Maximum ihrer Umsatzrentabilität noch gar nicht erreicht. Je besser der Service und je besser Ihr Verständnis über den Kunden, desto wahrscheinlicher nähern Sie sich diesem Maximum. Verwandeln Sie Ihren Service in eine gewinnbringende Investition.

**Letztendlich profitieren Sie also doppelt. Einerseits ersparen Sie sich unrentable Investitionen, andererseits können Sie diese Ressourcen nutzen um in besonders rentable Kunden zu investieren.**

Doch so offensichtlich der Nutzen des CLV ist, so herausfordernd kann die konkrete Entwicklung und Umsetzung sein. Das vorliegende Paper gibt einen Überblick über Grundlagen und die vielfältigen Einsatzmöglichkeiten des CLV. In diesem Rahmen werden gängige Methoden mit ihren jeweiligen Stärken sowie Strategien zur Beurteilung der resultierenden CLV-Prognosen skizziert. Abschließend gibt ein exemplarischer Projektplan einen Einblick in den typischen Ablauf eines CLV-Projekts.

### CLV: Was steckt dahinter?

Der CLV bietet die Möglichkeit, den Wert eines Kunden datengetrieben zu ermitteln und in Form einer einzigen Zahl zu quantifizieren. Er ist hervorragend geeignet, um verschiedene CRM-Maßnahmen oder Maßnahmen zur Neukundengewinnung zu beurteilen und zu steuern. Ironischerweise schrecken trotzdem viele Unternehmen vor der Nutzung des CLVs zurück, weil sie das Konstrukt für zu akademisch halten. Dabei ist der CLV – eine pragmatische Operationalisierung vorausgesetzt – eine grundsätzliche Kenngröße, die hilft, die besten Kunden zu identifizieren.

In der Theorie umfasst der CLV eine Vielzahl potenzieller Dimensionen, wie Abbildung 1 verdeutlicht. Ganz zu Recht ist für Unternehmen schwer vorstellbar, wie all diese Dimensionen erfasst und zu einem CLV verrechnet werden sollen. Dies ist jedoch auch überhaupt nicht notwendig. Die Empfehlung geht hier viel mehr hin zu einer pragmatischen Umsetzung auf Basis der Bestellhistorie. Diese Daten sind in der Regel leicht zugänglich. Ganz im Sinne der 80-20-Regel lässt sich so mit überschaubarem Aufwand bereits ein großer Mehrwert erzeugen. Dies schließt nicht aus, das Modell zu einem späteren Zeitpunkt um weitere Dimensionen zu erweitern, z.B. um das Referenzpotenzial, d.h. die Weiterempfehlungsbereitschaft eines Kunden sowie den Wert der empfohlenen Kunden.

### Retrospektiver vs. prospektiver CLV

Bei Bestandskunden wird häufig zwischen retrospektivem und prospektivem CLV unterschieden. Der sogenannte retrospektive CLV lässt sich auf Basis der vorhandenen Daten leicht berechnen und fällt in den Bereich des internen Rechnungswesens. Auch wenn er interessante Informationen liefern kann, bietet er keinen Blick in die Zukunft. Oft ändern Kunden ihr Verhalten mit der Zeit und es kann nicht angenommen werden, dass der retrospektive CLV sich auch gut als Prognose eignet.

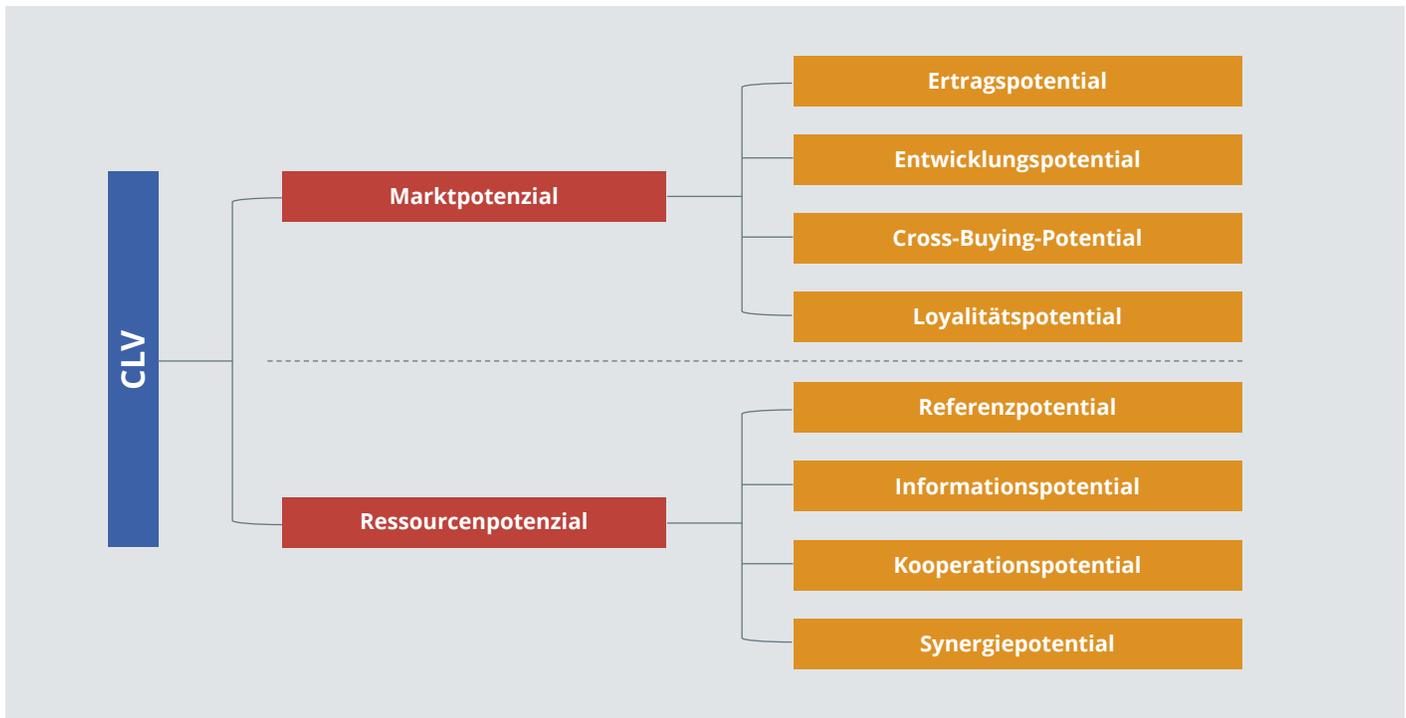


Abbildung 1: Facetten des CLV (Darstellung in Anlehnung an Torsten Tomczak/Elisabeth Rudolf-Sipötz (2009): Bestimmungsfaktoren des Kundenwertes: Ergebnisse einer branchenübergreifenden Studie S. 132, in: Bernd Günter/Sabrina Helm: Kundenwert, Grundlagen - Innovative Konzepte - Praktische Umsetzungen, Springer, S. 127-155).

Hilfreicher ist hier der prospektive CLV, bei dem mit Hilfe statistischer Verfahren das zukünftige Kundenverhalten vorhergesagt wird. Dabei werden selbstverständlich auch die Metriken berücksichtigt, die in den retrospektiven CLV einfließen. Jedoch können hier auch Tendenzen und Änderungen im Kundenverhalten vorhergesagt werden.

### Shop-Modell vs. Abo-Modell

Je nach Geschäftsmodell unterscheidet sich die konkrete Umsetzung des CLV. Beim transaktionsorientierten Shop-Modell (z.B. Autovermietung, Online-Fashionshop) ist von Interesse, wie viel Nutzen ein Kunde dem Unternehmen in einem bestimmten Zeitraum, z.B. den kommenden 365 Tagen, bringen wird. Dies lässt sich beispielsweise über den erwarteten Wert für Deckungsbeitrag, Umsatz oder Anzahl der Bestellungen in diesem Zeitraum operationalisieren. Handelt es sich hingegen um Abonnement-Kunden (z.B. Laufzeitvertrag im Mobilfunk), ist die verbleibende Dauer der Kundenbeziehung von Interesse. In diesem Fall möchte man frühzeitig erkennen, ob ein Kunde akut abwanderungsgefährdet ist oder ob die prognostizierte Kundenzugehörigkeit sehr hoch ist. Eine hohe erwartete Zugehörigkeit entspricht hier einem hohen CLV.

### Umsatz vs. Deckungsbeitrag

Wird der monetäre Kundenwert berechnet, eignen sich grundsätzlich sowohl Umsätze als auch Deckungsbeiträge als Prognoseziel. Deckungsbeiträge stehen mitunter im CRM-System nicht zur Verfügung, sodass in diesen Fällen ein Ausweichen auf Umsätze nötig ist. Sind jedoch Informationen über Deckungsbeiträge verfügbar, empfiehlt es sich, diese als Prognoseziel zu wählen. Der Grund ist einleuchtend: Der Deckungsbeitrag spiegelt den Wert des Kunden für das Unternehmen sehr viel genauer wider, da hierbei die Kosten bereits berücksichtigt sind. Ein hoher Umsatz kann auch mit einem sehr geringen Deckungsbeitrag einhergehen, wenn z.B. ein Kunde regelmäßig hochpreisige, aber reduzierte Artikel kauft. Ein anderer Kunde mit moderatem Umsatz hingegen kann durchaus wertvoller sein, wenn sein Umsatz durch Artikel mit sehr hohem Deckungsbeitrag zustande kommt.

Stehen weder Umsätze noch Deckungsbeiträge in zuverlässiger Genauigkeit zur Verfügung und unterscheiden sich die Werte der Bestellungen bzw. gekauften Artikel nicht besonders stark voneinander, bietet sich auch ein Ausweichen auf die bloße Anzahl von Bestellungen bzw. Artikeln an.

## Customer Lifetime Value: Breite Einsatzmöglichkeiten

Unternehmen investieren große Mengen finanzieller Ressourcen in die Neukundengewinnung und das Kundenbindungsmanagement. Werden diese Ressourcen an falscher Stelle eingesetzt, führt das nicht nur zu überflüssigen Ausgaben, sondern im schlimmsten Fall sogar zur Akquise unrentabler Kunden. Der CLV hilft dabei, die Ressourcen gezielter und gewinnbringender einzusetzen.

Zum Beispiel wird bei der Beurteilung von Maßnahmen zur Neukundengewinnung im Online-Marketing häufig nur betrachtet, wie viele Neukunden ein Kanal liefert. Dadurch ergibt sich jedoch schnell ein Fokus auf Kanäle, die viele, aber unrentable Kunden liefern. Im schlimmsten Fall belohnt diese Betrachtung einen unseriösen Affiliate-Partner, der aggressiv Gutscheine einsetzt und Kunden liefert, die nur einmalig bestellen und einen Großteil retournieren. Ein eigentlich viel wertvollerer Kanal, der seriöse, treue Kunden liefert, erhält hingegen weniger Budget. Entscheidend für die Budgetverteilung ist also nicht, wie viele Neukunden über einen Kanal akquiriert werden, sondern wie wertvoll diese Kunden sind. Der CLV quantifiziert diesen Kundenwert, sodass sich die Rentabilität verschiedener Kanäle viel genauer beurteilen lässt und das Budget sinnvoller und gewinnbringender investiert werden kann.

Sind die Kunden erst einmal gewonnen, sind die Einsatzmöglichkeiten des CLV noch lange nicht erschöpft. Auch im Kundenbindungsmanagement ergeben sich vielfältige Einsatzmöglichkeiten: So können beispielsweise Gutscheine oder kostenlose Extras eingesetzt werden, um besonders wertvolle Kunden zum Kauf oder zu einer Vertragsverlängerung zu motivieren. Diese Anreize sollen jedoch nicht nach dem "Gießkannenprinzip" an alle Kunden verteilt, sondern gezielt eingesetzt werden: So ist es zum Beispiel insbesondere bei rentablen Kunden sinnvoll, kostenlose Extras anzubieten oder sich bei Beschwerden kulant zu zeigen. Auf den ersten Blick ist jedoch nicht immer ersichtlich, wie rentabel ein Kunde eigentlich ist. Auch hier schafft der CLV Abhilfe, da er den Wert eines Kunden genau quantifiziert.

Nicht zuletzt generiert die Berechnung des CLV Insights darüber, wie verschiedene Metriken mit dem Kundenwert zusammenhängen. Dadurch lassen sich bestehende Annahmen überprüfen, aber auch bisher unbekannte Effekte aufdecken, aus denen sich häufig weitere CRM-Maßnahmen ableiten lassen.

## Daten

Als Datengrundlage für die Berechnung des CLV werden als Minimum Daten zur Bestellhistorie der Kun-

den inkl. Anzahlen/Preisen/Deckungsbeiträgen der bestellten Produkte/Dienstleistungen, Rabatte und Retouren bzw. Beschwerden benötigt. Diese Daten lassen sich zusätzlich mit weiteren Kundeneigenschaften anreichern, z.B. Alter, Geschlecht oder Region. Gibt es im Analysezeitraum grundlegende Veränderungen, z.B. eine starke Umsatzsteigerung aufgrund einer effektiven Werbekampagne, so sollte diese Information ebenfalls in die Analyse mit einbezogen werden.

## Individualisierung nach Geschäftsmodell

Auf Basis der genannten Daten kann zwar bereits ein CLV berechnet werden, wofür auch zahlreiche SaaS-Lösungen angeboten werden. Das Potenzial des CLV lässt sich jedoch erst dann voll ausschöpfen, wenn weitere Metriken hinzugefügt werden, die stark vom jeweiligen Geschäftsmodell abhängen. Ein maßgeschneiderter Algorithmus ermöglicht die Berücksichtigung genau jener Daten, die für Ihr Geschäftsmodell spezifisch sind. So sind für eine Autovermietung ganz andere Metriken relevant als für einen Online-Fashionshop oder einen Telefonanbieter. Je nach Produkt oder Dienstleistung könnte es sich dabei um Eigenschaften/Kategorien der gekauften Produkte, Dauer und Intensität der Nutzung, gebuchte Zusatzleistungen, Zeitpunkt der Buchung bzw. Nutzung (z.B. Wochentag, Weihnachtszeit, ...) oder Zahlungsmethoden handeln.

Die Berücksichtigung solcher spezifischer Metriken macht die CLV-Berechnung nicht nur genauer, sondern bietet auch größeres Potenzial für Insights über die Wirkungszusammenhänge zwischen Metriken und Kundenwert.

## Online-Tracking-Daten

Den meisten Unternehmen steht eine große Menge an Online-Tracking-Daten ihrer Kunden zur Verfügung. Daher taucht häufig die Frage auf, ob diese ebenfalls in die CLV-Berechnung miteinfließen können. Ob die Berücksichtigung dieser Daten einen Mehrwert bietet, hängt von verschiedenen Faktoren ab.

Insbesondere bei Geschäftsmodellen mit niedriger Bestellfrequenz sind Trackingdaten eine hilfreiche Ergänzung, um hochfrequenter Informationen zu erhalten. Doch auch bei höheren Bestellfrequenzen können Informationen über zusätzliche Kontakte ohne Bestellung zusätzlichen Nutzen bringen.

Allerdings muss zunächst geprüft werden, wie gut sich die Daten überhaupt den einzelnen Kunden zuordnen lassen. Häufig sind Kunden nicht eingeloggt, wenn sie die Unternehmenswebsite besuchen, ohne etwas zu kaufen. Dann hängt es stark von den Cookie-Einstellungen im Browser des Kunden ab, ob sich die Bewegungsdaten auf der Website überhaupt mit dem

Kunden-Account in Verbindung bringen lassen. Einfacher wird die Situation, wenn Kunden vorwiegend über eine Smartphone- bzw. Tablet-App auf das Unternehmensportal zugreifen, da man in Apps üblicherweise mit dem Benutzernamen eingeloggt ist. Um zu entscheiden, ob die Nutzung der Tracking-Daten sinnvoll ist, sollte ihre Qualität und Plausibilität im Vorfeld genau geprüft werden.

## Prüfung der Datenqualität

Die Prognoseergebnisse können nur so gut sein wie die Daten, die zur Berechnung genutzt werden. Daher sollte die Datenqualität aller verwendeten Daten genau untersucht werden. Hierbei erfolgt beispielsweise eine Prüfung auf Abweichung von der Dokumentation, fehlende/unstimmige Werte, Konsistenzprobleme, Ausreißer, unerwartete Peaks oder Einbrüche im zeitlichen Verlauf der Buchungen. Mögliche Fragen sind hier:

- Gibt es unplausible Werte (z.B. negative Aufenthaltsdauer bei Hotels, sehr hohes Alter)?
- Sind in den Daten noch Testkunden bzw. interne Accounts enthalten?
- Tauchen in den Daten einzelne Kunden (z.B. Firmenkunden) mit extrem hohen Umsätzen auf, welche die Analysen verzerren können?

Werden hierbei Ungereimtheiten aufgedeckt, bedeutet das jedoch nicht, dass das Projekt zum Scheitern verurteilt ist. Häufig reicht es bereits aus, einzelne extreme Beobachtungen auszuschließen bzw. auf einen weniger extremen Wert festzusetzen, den Analysezeitraum einzuschränken oder notfalls einzelne Metriken von der Analyse auszuschließen. Falls diese Maßnahmen nicht helfen, ist es im Zweifelsfall immer noch besser, eine Verzögerung durch eine erneute, korrigierte Datenlieferung in Kauf zu nehmen, als aufgrund fehlerhafter Daten unbrauchbare Ergebnisse zu erhalten.

## Auswahl des Analysezeitraums

Es überrascht nicht, dass der Zeitraum der Daten ausreichend groß sein muss, um eine zufriedenstellende Prognosegüte zu erreichen. Doch auch wenn in der Bezeichnung "CLV" das Wort "Lifetime" enthalten ist, muss in der Regel nicht die gesamte Kundenhistorie für die Analyse berücksichtigt werden. Im ungünstigen Fall kann ein zu langer Zeitraum sogar von Nachteil sein: Mitunter können die Information aus weit zurückliegenden Bestellungen die Prognose sogar verschlechtern, da die zugrundeliegenden Mechanismen sich zwischenzeitlich zu stark verändert haben. Diesem Problem lässt sich auch durch eine zeitliche Gewichtung der Daten Rechnung tragen.

Außerdem ist der optimale Zeitraum abhängig von der typischen Bestellfrequenz der betrachteten Produkte oder Dienstleistungen. So leuchtet es ein, dass bei der Betrachtung von Reisebuchungen ein längerer Zeitraum (mehrere Jahre) berücksichtigt werden muss, während bei Bestellungen von Lebensmitteln für den täglichen Gebrauch schon Wochen oder Monate ausreichen. Zudem kann es sinnvoll sein, bei der Wahl des Zeitraums Saisonalitäten zu berücksichtigen. In der Regel variiert die Anzahl der Bestellungen über die Woche und über das Jahr hinweg systematisch, beispielsweise im Rahmen typischer Urlaubszeiten oder des Weihnachtsgeschäfts. Um unerwünschte Einflüsse dieser Schwankungen auf die Analyse zu verhindern, empfiehlt es sich, die entsprechenden Zeiträume in den Daten stets vollständig abzudecken (z.B. zwei Jahre statt nur eineinhalb).

Nicht zuletzt spielt der gewählte Prognosezeitraum eine wichtige Rolle. Als Faustregel lässt sich sagen, dass der Datenzeitraum mindestens zweimal so lang sein sollte wie der gewünschte Prognosezeitraum, besser jedoch länger. Bei der Berechnung des CLV für ein Jahr zum Beispiel sollten mindestens zwei Jahre Kundenhistorie verfügbar sein.

## Statistische Modellierung

Zur Berechnung des CLV gibt es eine Vielzahl verschiedener Verfahren, die sich im Hinblick auf Komplexität, Interpretierbarkeit und Wartungsaufwand unterscheiden. Sie lassen sich grob in zwei Gruppen einteilen: Machine Learning-Verfahren und statistische Regressionsansätze. Es gibt jedoch auch Mischformen oder Kombination von Methoden aus diesen beiden Bereichen.

Machine Learning-Algorithmen wurden in erster Linie in der Informatik zur Erkennung von Kategorien und Mustern entwickelt. Meistens handelt es sich dabei um eine Black Box, die keinen Aufschluss über Wirkungszusammenhänge zwischen verwendeten Metriken und prognostizierter Variable erlaubt. Machine Learning-Verfahren sind häufig hochgradig automatisiert, jedoch kann die "Sinnhaftigkeit" der für die Klassifizierung genutzten Regeln häufig nicht geprüft werden.

Statistische Regressionsansätze im Kontext des CLV gehen weit über die multiple lineare Regression hinaus. Vielmehr handelt es sich um die Klasse der generalisierten additiven Modelle, die ein sehr flexibles und bewährtes Spektrum an Modellen umfasst. Diese ermöglichen die Modellierung der Beziehungen zwischen einer abhängigen und nahezu beliebig vielen erklärenden Variablen. Es stehen weiterhin Techniken zur Variablenselektion, Modellierung nichtlinearer Zusammenhänge sowie zur zeitlichen Gewichtung und Ausbalancierung der Datenbasis zur Verfügung. Im Rahmen der Modellbildung können Hypothesen



Abbildung 2: Zur CLV-Berechnung verwendbare Daten.

über vermutete Einflussfaktoren getestet und Wirkungszusammenhänge identifiziert werden. Regressionsansätze öffnen also die Black Box und ermöglichen ein Verständnis der Gesetzmäßigkeiten, die einen hohen oder niedrigen CLV bestimmen.

Welches Verfahren sich für Ihr konkretes Geschäftsmodell am besten eignet, müssen Sie nicht blind im Voraus entscheiden. Im Rahmen der Entwicklung eines maßgeschneiderten Algorithmus werden auch Entscheidungen über die optimale statistische Umsetzung getroffen. Dabei werden unter anderem die spezifischen Eigenschaften Ihres Geschäftsmodell, die verfügbaren Daten, der geplante Einsatzbereich des CLV und die erreichte Prognosegüte der statistischen Modelle berücksichtigt.

### Mögliche Zielgrößen und Verfahren

Zu Beginn jeder CLV-Berechnung muss die Zielgröße der Analyse ausgewählt werden. Mögliche Zielgrößen sind zum Beispiel:

- **Zugehörigkeit:** Verbleibende Dauer der Kundenbeziehung (insbesondere bei Abo-Modellen)
- **Kaufwahrscheinlichkeit:** Wahrscheinlichkeit, dass ein Kunde im festgelegten Prognosezeitraum einen Kauf tätigt
- **Anzahl Bestellungen:** Anzahl der Bestellungen im Prognosezeitraum
- **Umsatz bzw. Deckungsbeitrag:** erwarteter Gesamtumsatz bzw. Gesamt-Deckungsbeitrag im Prognosezeitraum

Abhängig von der Zielgröße steht eine Vielzahl von Methoden aus dem Bereich der Regressionsanalyse und des Machine Learning zur Auswahl. Ziel ist es, daraus den optimalen Ansatz für Ihre spezifischen Anforderungen auszuwählen:

- Random Forest
- Random Survival Forest
- Classification Trees
- Regression Trees
- Support Vector Machines
- Support Vector Regression
- Neuronale Netze
- Logistische Regression
- Poisson-Regression
- Survival-Modelle
- Generalisierte additive Modelle
- Ridge- bzw. Lasso-Penalisierung

Dabei lassen sich die verschiedenen Ansätze teilweise auch miteinander kombinieren, um ein besseres Ergebnis zu erhalten.

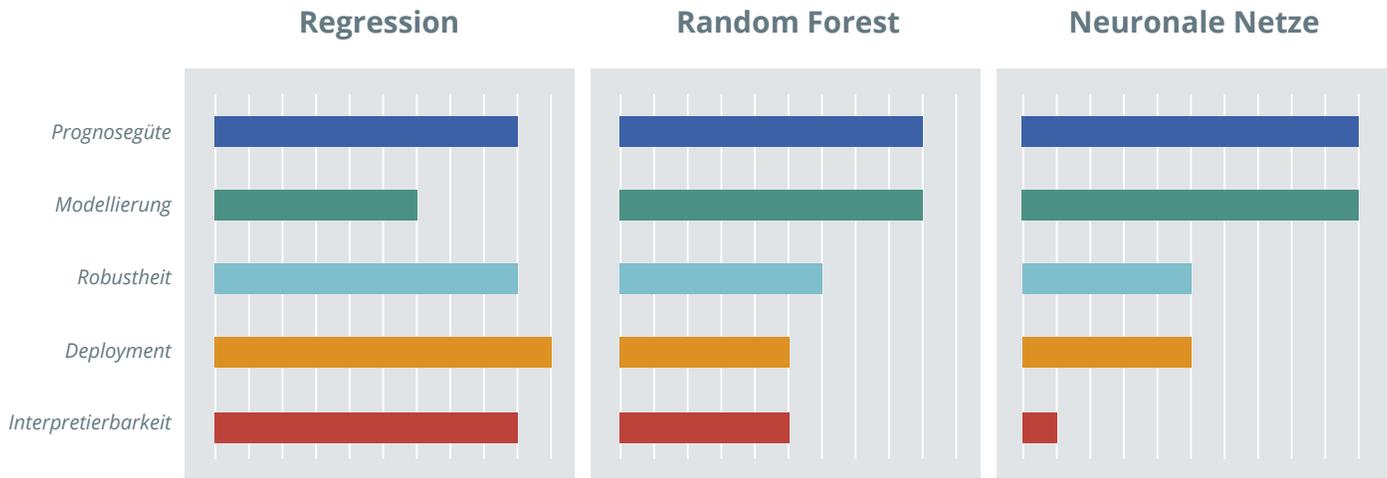


Abbildung 3: Vergleich ausgewählter statistischer und Machine Learning-Methoden.

## Vergleich der Verfahren

Jeder statistische Ansatz hat Stärken und Schwächen. Die Herausforderung ist, den Ansatz auszuwählen, der für Ihre Situation und Ihr Geschäftsmodell den meisten Nutzen liefert. Dazu lassen sich die konkreten Modellierungsergebnisse verschiedener Ansätze bezüglich ihrer Prognosegenauigkeit vergleichen. Es gibt aber auch einige generelle Merkmale, die charakteristisch für unterschiedliche Ansätze sind. Abbildung 3 visualisiert exemplarisch Stärken und Schwächen für Regressionsansätze, Random Forest und neuronale Netze.

### Prognosegüte

Eine hohe Prognosegüte ist für ein gutes Modell unverzichtbar. Neben Umfang und Qualität der Daten beeinflusst vor allem das Modell die Prognosegüte. Da bei Machine Learning-Algorithmen automatisiert entschieden wird, welche Metriken zur Prognose verwendet werden und in welcher Form, wird die in den Daten enthaltene Information in der Regel optimal genutzt. Daher können Regressionsansätze bezüglich der Prognosegüte unterlegen sein, wenn bei der Regressions-Modellierung unüberlegte Entscheidungen getroffen werden. Das kann beispielsweise der Ausschluss von Metriken ohne systematische Untersuchung der Folgen für die Prognosegüte sein oder das Auslassen einer wichtigen Wechselwirkung zwischen zwei Metriken. Dies lässt sich durch einen Austausch mit fachlichen Experten verhindern.

### Modellierung

Im Rahmen der Modellierung muss entschieden werden, welche Metriken in das Modell eingehen und

in welcher Form (z.B. linear vs. nichtlinear, einzeln oder in Interaktion, ...). Machine Learning-Algorithmen sind in der Regel hochgradig automatisiert, sodass in die Modellierung verhältnismäßig wenig Zeit investiert werden muss. Bei Regressionsansätzen hingegen muss "von Hand" geprüft werden, ob eine bestimmte Veränderung das Modell verbessert oder nicht. Dies ist Stärke und Schwäche zugleich: Zum einen muss mehr Zeit und Expertise investiert werden, zum anderen kann domänenspezifisches Wissen gezielt umgesetzt werden.

### Robustheit

Eine Voraussetzung für eine gute Prognose ist, dass das Modell dazu in der Lage ist, die Variationen in den vorhandenen Daten abzubilden. Doch damit ist es nicht getan: Benötigt wird ein robustes Modell, das auch zukünftig auf neuen Daten eine stabile Prognose produziert. Beispielsweise dürfen extreme Ausprägungen auf einzelnen Metriken nicht zu stark verzerrten, unrealistischen Prognosen führen. Sowohl für Regressions- als auch für Machine Learning-Ansätze stehen Methoden zur Verfügung, um die Robustheit zu prüfen und sicherzustellen.

### Deployment und Wartung

Nach dem Abschluss der Modellentwicklung zur CLV-Prognose folgt als nächster Schritt in der Regel das Deployment, um die Prognosen in Echtzeit im unternehmensinternen CRM-System zu erstellen. Steht im Kunden-Backend eine Statistiksoftware zur Verfügung, unterscheidet sich der Aufwand für die Implementierung nicht wesentlich zwischen den Ansätzen. Soll die Prognose jedoch in einer anderen Programmiersprache

(z.B. Java, SQL) umgesetzt werden, ist dies für Regressionsansätze mit deutlich weniger Aufwand verbunden. Hier müssen lediglich Modell-Koeffizienten, welche in Form einer Mappingtabelle hinterlegt werden können, in eine simple Formel eingesetzt werden. Dies lässt sich leicht in jeder beliebigen Sprache umsetzen.

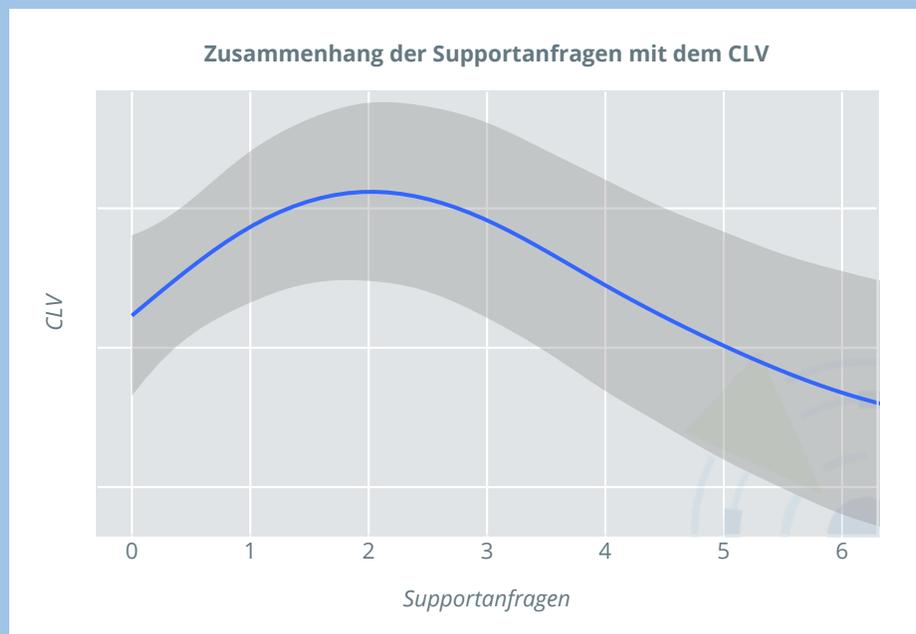
### Interpretierbarkeit der Ergebnisse

Bei Machine Learning-Algorithmen handelt es sich sehr häufig um Black-Box-Ansätze, die zwar die Zusam-

menhänge zwischen erklärenden Variablen und der Zielgröße identifizieren, deren Parameter sich aber inhaltlich nicht (z.B. neuronale Netze) oder nur sehr schwer (z.B. Random Forest) interpretieren lassen. Regressionsmodelle hingegen haben den Vorteil, dass sich die identifizierten Wirkungszusammenhänge in Form direkt interpretierbarer Koeffizienten und der dazugehörigen Information zur statistischen Verlässlichkeit angeben und überprüfen lassen.

## Beispiel: Interpretierbarkeit der Ergebnisse

Die folgende Abbildung zeigt exemplarisch den Zusammenhang zwischen der Anzahl von Supportanfragen und dem CLV in einem Regressionsmodell mit Berücksichtigung nichtlinearer Effekte. Der Zusammenhang ist kontraintuitiv: Eigentlich würde man erwarten, dass Supportanfragen ein Zeichen für Unzufriedenheit mit dem Produkt sind, sodass mit steigender Anzahl von Anfragen der CLV generell sinkt. In diesem Fall zeigt sich jedoch ein anderes Bild. Im Bereich von null bis zwei Supportanfragen steigt der CLV zunächst an, erst danach sinkt er ab. Dies zeigt, dass eine moderate Anzahl von Supportanfragen im Falle des betrachteten Produkts förderlich für den CLV ist. Vermutlich handelt es sich um ein komplizierteres Produkt, welches viele Kunden erst nach Hilfestellung durch den Support zu ihrer eigenen Zufriedenheit nutzen können. Aus dieser Erkenntnis lassen sich konkrete Maßnahmen ableiten, um die Kundenzufriedenheit und somit den CLV zu erhöhen: Bei anfänglichen Schwierigkeiten sollten Kunden dazu ermutigt werden, den Support zu kontaktieren. Auch kann gut geschultes Support-Personal in ausreichender Anzahl sicherstellen, dass Anfragen zeitnah bearbeitet werden können – mit dem Ergebnis zufriedener, loyaler Kunden.



## Beurteilung der Modellgüte

### Modellanpassung vs. Prognosegüte

Bei der Modellentwicklung wird das Modell naturgemäß auf die zugrundeliegenden Daten geeicht. Dadurch sind sehr flexible Modelle unvermeidlich im Vorteil, denn sie können sich besonders stark an die vorliegenden Daten anpassen. Entscheidend ist jedoch nicht eine optimale Anpassung an die bereits bekannten Daten, sondern eine möglichst zuverlässige Prognose für die Zukunft. Dementsprechend sind nicht etwa möglichst flexible Modelle zu favorisieren, sondern robustere Modelle, die sich nicht an jede kleine (möglicherweise zufällige) Unregelmäßigkeit in den Daten anpassen, sondern die grundlegenden Gesetzmäßigkeiten abbilden. Um zu beurteilen, ob ein auf bekannten Daten entwickeltes Modell auch zuverlässige Prognosen liefert, lässt sich die "Out-of-sample"-Güte betrachten. Hierzu bleibt der letzte Zeitraum in den Daten (z.B. die letzten vier Wochen oder das letzte Jahr) zunächst unangetastet. Man bezeichnet diesen Zeitraum auch als "Testdaten". Das Modell wird auf den übrigen Daten, den Trainingsdaten, geeicht. Dann werden mit dem resultierenden Modell die Werte für den Testzeitraum prognostiziert und mit den tatsächlichen Werten verglichen. Auf diese Weise lässt sich zuverlässig einschätzen, wie genau die ermittelten Prognosen sind.

### Gütemaße

Zur Beurteilung statistischer Modelle existiert eine Vielzahl von Gütemaßen, abhängig von der genauen Modellklasse. Ein Maß ist das sogenannte  $R^2$ , welches zwischen 0 und 1 liegt und angibt, welcher Anteil der Variation der tatsächlichen Werte durch das Modell abgebildet werden kann. Außerdem kann die mittlere absolute, quadrierte oder prozentuale Abweichung der prognostizierten von den tatsächlichen Werten berechnet werden.

Auf welches Gütemaß bei der Beurteilung das Hauptaugenmerk gelegt werden sollte, hängt unter anderem vom konkreten Anwendungsfall ab: Sollen beispielsweise Akquisekosten und monetärer Kundenwert gegeneinander aufgerechnet werden, so ist besonders wichtig, dass die Kundenwerte im Mittel den korrekten Wert treffen, statt diesen zu über- oder unterschätzen. In diesem Fall wäre eine Betrachtung der Abweichung am sinnvollsten. Im Kundenbindungsmanagement hingegen wird häufig eine Einteilung der Kunden in Kategorien (A/B/C) benötigt. An dieser Stelle ist daher besonders wichtig, dass die Überdeckung der prognostizierten Kategorien mit den tatsächlichen Kategorien möglichst hoch ist. Die wahren Werte sind daher weniger relevant als die korrekte Reihung der Kunden nach ihrem CLV.

### Wann ist die Prognose "gut"?

Mitunter ist es nicht einfach, zu beurteilen, ob ein konkreter Wert für ein bestimmtes Gütemaß "gut" oder "schlecht" ist. Um eine Vorstellung davon zu bekommen, ob die erreichte Modellgüte im Kontext der verfügbaren Daten und Metriken zufriedenstellend ist, bietet es sich daher an, ein sogenanntes "Null-Modell" zum Vergleich zu berechnen. Dies kann ein sehr simples Modell (z.B. Deckungsbeitrag des letzten Jahres als Prognose für das kommende Jahr) oder eine bisher verwendete Methode zur CLV-Berechnung sein. Alle betrachteten Gütemaße lassen sich dann auch für das Null-Modell berechnen und mit denen des anderen Modells vergleichen. Dadurch lässt sich der Nutzen der komplexeren statistischen Modellierung präziser beurteilen.

Die benötigte Prognosegenauigkeit ist auch abhängig vom Einsatzbereich des CLV. So werden bei der Aussteuerung von Maßnahmen, welche sich auf einzelne Kunden beziehen, ausreichend genaue Prognosen benötigt, um im Einzelfall sicher entscheiden zu können. Werden jedoch ganze Kohorten analysiert – wie beim Vergleich von Kanälen zur Neukundengewinnung – sind die Anforderungen an die Prognosegüte deutlich geringer.

Weiteren Aufschluss über Stärken oder Schwächen des Modells gibt die Betrachtung der Gütemaße differenziert nach Gruppen, z.B. Kundengruppen aus einer Kundensegmentierung, Regionen oder klassiert nach Dauer der Kundenbeziehung, Gesamtumsatz oder Anzahl der Bestellungen.

### Fazit

INWT entwickelt für Sie einen maßgeschneiderten Algorithmus zur Prognose des CLV und identifiziert den optimalen Modellierungsansatz für Ihr Geschäftsmodell und Ihre Daten. Mit Hilfe des CLV lässt sich der Wert Ihrer Kunden quantifizieren. Darauf basierend können Strategien zur Neukundengewinnung zuverlässig beurteilt und Maßnahmen zur Kundenbindung gezielt angesteuert werden, um beispielsweise abwanderungsgefährdete oder besonders rentable Kunden an das Unternehmen zu binden. Zusätzlich zu RFM-Metriken lassen sich dabei auch Kundencharakteristika, geschäftsmodellspezifische Merkmale sowie Tracking-Daten berücksichtigen. Auf diese Weise können Investitionen in Kundenakquise und Kundenbindungsmanagement viel zielgerichteter angesteuert und überflüssige Investitionen vermieden werden.

Als Ergebnis erhalten Sie nicht nur den berechneten CLV zum Zeitpunkt der Projektdurchführung, sondern auch die Werkzeuge, um zukünftig in Echtzeit den CLV – auch für Neukunden – auf Basis aktualisierter Daten in Ihrem CRM-System zu berechnen.

## Ihr Ansprechpartner

---



### Dr. Steffen Wagner

Steffen ist Mitgründer von INWT. Er beschäftigt sich schwerpunktmäßig mit den Themen Predictive Analytics, Online Marketing und Customer Relationship Management. Der promovierte Physiker gibt als Dozent im 'Joint Masters Program of Statistics' in Berlin Einblicke in seine Tätigkeit als Data Scientist.

### Kontakt

- **Tel.:** +49 30 1208231-58
- **E-Mail:** [steffen.wagner@inwt-statistics.de](mailto:steffen.wagner@inwt-statistics.de)

 **INWT Statistics GmbH**  
Hauptstraße 8  
Meisenbach Höfe, Aufgang 3a  
10827 Berlin

 +49 30 1208231-0

 [info@inwt-statistics.de](mailto:info@inwt-statistics.de)

 [www.inwt-statistics.de](http://www.inwt-statistics.de)

