

Clustering

futuresbydesign.net

Clustering ist eine grundlegende Fähigkeit, die ein Unternehmen erlernen sollte, um das volle Potenzial seiner Daten zu nutzen

Grundlagen des Clusterings und weshalb dieses Verfahren von großer Bedeutung für Unternehmen sein kann

Unternehmen, die ihre Daten für eine effiziente Entscheidungsfindung nutzen, werden ihr Geschäft viel schneller ausbauen als ihre Konkurrenten. Ob Ihr Unternehmen seine Daten in Excel-Tabellen oder in einem vollständigen Datenbanksystem speichert, spielt dabei keine Rolle. Was zählt, ist Folgendes: *Nutzen Sie das volle Potenzial Ihrer Daten?*

Jetzt fragen Sie sich vielleicht: "Mit welcher Analyse-Methode kann mein Unternehmen beginnen, um mehr Erkenntnisse aus Daten zu gewinnen?" Unabhängig davon, welche Art von Unternehmen Sie führen, wäre die Antwort: "**Clustering.**"

Es ist die Technik, die hinter Amazons Erfolg bei Online-Verkäufen steht. Auf diese Weise finden Facebook und Google heraus, welche Werbung für Sie in Frage kommt, wenn Sie ihre Website besuchen. Für viele kleinere Unternehmen ist es ein zentrales Instrument zur Kundenanalyse und hat ihnen geholfen, ihre Marketing- und Verkaufsstrategien und manchmal sogar ihr Produktangebot radikal zu verbessern. Vor allem aber ist es eine intuitive Technik.

In diesem Artikel werden Sie erfahren, was Clustering ist. Wir werden untersuchen, wie und warum erfolgreiche Unternehmen wie Netflix und Amazon Clustering im Kern ihres Geschäfts einsetzen. Noch interessanter ist, dass wir sehen werden, wie sogar ein kleines Restaurant Clustering nutzen kann, um intelligentere Entscheidungen zu treffen und sein Geschäft auszubauen. Und schließlich werden Sie erfahren, dass der Einstieg in das Clustering auch für die kleinsten Unternehmen leicht zu bewerkstelligen ist.

Was ist Clustering?

Bei Clustering-Methoden werden die Datenpunkte in Ihren Daten auf der Grundlage ihrer Ähnlichkeit in Gruppen zusammengefasst. Mit diesem Verfahren wird eine Struktur in Daten gefunden, die von sich aus nicht viel Struktur aufweisen. Das bedeutet, dass wir es auf fast alle Daten anwenden können, die in einem kleinen oder großen Unternehmen bereits vorhanden sind. Zu Beginn wissen wir vielleicht nicht genau, nach welchen Mustern wir suchen. Das ist



nicht schlimm. Genau dabei wird uns das Clustering helfen. Nachdem Sie "Cluster" erstellt haben, können Sie interessante Muster in den Daten Ihres Unternehmens weiter untersuchen. Sie können die Ergebnisse des Clustering nutzen, um wertvolle, gezielte Fragen zu den aufgedeckten Mustern und Gruppen zu stellen. Im nächsten Kapitel erfahren Sie am Beispiel eines Restaurants mehr über die Funktionsweise des Clustering.

Clustering in einem kleinen Restaurant

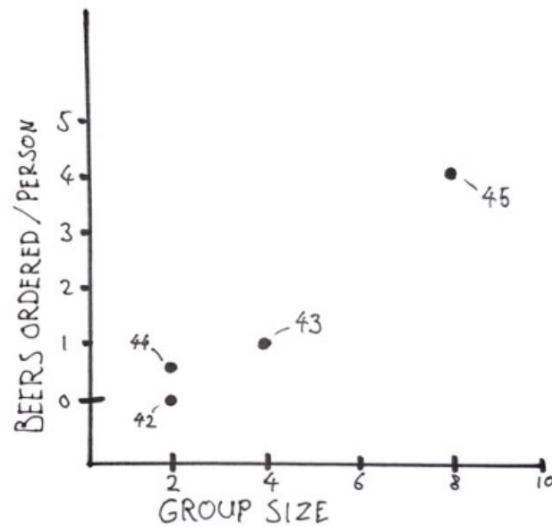
Betrachten wir als anschauliches Beispiel ein gemütliches kleines Lokal namens "Restaurant La Estadística". Wie die meisten Restaurants verwendet La Estadística ein Bestellsystem, um dem Koch mitzuteilen, was er zubereiten soll. Außerdem druckt es am Ende des Besuchs eine Quittung für die Gäste aus. Das System speichert diese Daten in einer Tabelle. Diese Tabelle könnte etwa so aussehen:

PROPERTIES

GUEST NUMBER	Group SIZE	DISHES ORDERED PER PERSON	DRINKS ORDERED PER PERSON	BILL AMOUNT PER PERSON
...
42	2	2	2 ½ RED WINE	€ 26
43	4	2 ½	1 BEER 1 WHITE WINE	€ 28
44	2	1	2 WHITE WINE ½ BEER	€ 20
45	8	3	4 BEER ½ SODA	€ 45
...

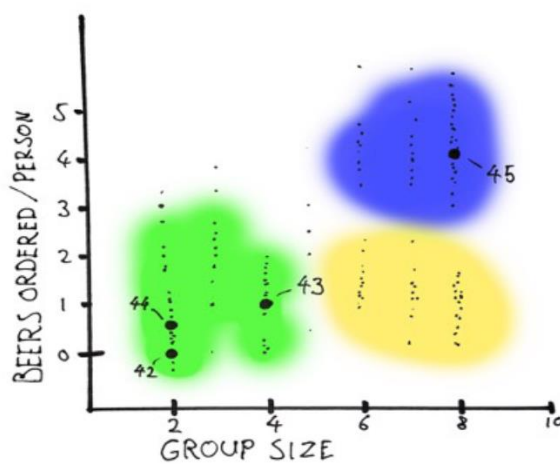
DATA POINTS →

Wir sehen jede Gruppe von Gästen als eine Zeile in der Tabelle dargestellt. Wir können daher jede Zeile als einen Datenpunkt betrachten. Jeder Datenpunkt besitzt bestimmte Eigenschaften. Wir können nun manuell ein Clustering erstellen, indem wir alle Datenpunkte anhand ihrer Eigenschaften in einer Abbildung einzeichnen. Wir wählen die *Gruppengröße* und die *Anzahl der Biere*.



Wir können sofort sehen, dass der Datenpunkt mit 8 Gästen sich von den anderen drei Datenpunkten unterscheidet, wenn wir uns ansehen, wie viel Biere sie pro Person bestellt haben. Jetzt haben wir nur vier der Datenpunkte im Bestellsystem eingezeichnet.

Das Restaurant verfügt natürlich über weitaus mehr Daten. Zeichnen wir auch die restlichen Datenpunkte ein und fügen außerdem Farben hinzu. Vielleicht erkennen Sie nun ein Muster.



Können Sie die drei Gruppen erkennen, die unser Clustering unterscheidet?

Diese Gruppen werden Cluster genannt. Cluster sind Gruppen von Datenpunkten, die nahe beieinanderliegen. Das bedeutet, dass ein Cluster in diesem Fall jene Gäste enthält, die einander sehr ähnlich sind.



Eine kurze Anmerkung zu Clustering-Algorithmen

Wir konnten unsere Datenpunkte auf Papier einzeichnen, um ein einfaches Clustering zu erstellen. Wenn wir jedoch mehr als zwei Eigenschaften in einem Clustering berücksichtigen wollen, stoßen wir schnell auf Probleme, weil es schwierig ist, mehr als drei Achsen zu visualisieren. Hier kommen Clustering-Algorithmen ins Spiel. Man muss nicht genau wissen, wie sie funktionieren, aber es ist nützlich, eine allgemeine Vorstellung davon zu haben, was sie tun.

Ein **Algorithmus** ist ein Programm, das einige Datenpunkte aufnimmt und als Ergebnis eine Antwort liefert. Der Algorithmus für die Addition zweier Zahlen nimmt beispielsweise zwei Zahlen als Eingabe, z.B. 1 und 2, und erzeugt eine Antwort, nämlich 3.

Ein **Clustering-Algorithmus** erzeugt eine Antwort auf das Problem:

Ordne jeden Datenpunkt in den Eingabedaten einem Cluster zu. Ähnliche Datenpunkte sollten im selben Cluster landen, aber Datenpunkte, die sehr unterschiedlich sind, müssen in verschiedenen Clustern landen.

Ein Clustering-Algorithmus wird eine weitaus genauere Antwort liefern als das manuelle Clustering, welches wir oben vorgenommen haben. Außerdem können sie Hunderte von Eigenschaften gleichzeitig berücksichtigen, anstatt nur zwei. Aber die Idee bleibt dieselbe. Wir ordnen immer noch ähnliche Datenpunkte verschiedenen Gruppen zu.

Wie Sie mit Clustering einen Mehrwert für Ihr Unternehmen generieren

Zurück zu unserem Beispiel. Mit dem Clustering kann der Chef von „La Estadística“ das blaue Cluster (siehe oben) leicht filtern. Er kann schnell feststellen, dass diese Gruppe im Durchschnitt mehr pro Person zahlt als andere Gruppen. Er stellt zudem fest, dass sie oft mehr Bier und mehr Fleisch bestellen als andere Gäste.

Der Chefkoch beschließt, ein Gruppenmenü speziell für diese Gäste zu kreieren. Er gibt einige gute Fleischempfehlungen, kombiniert mit einem Rabatt auf spezielles Bier, wenn die Gäste mindestens drei Gerichte bestellen. Das Ergebnis: Die Gäste sind viel zufriedener mit dem Essen und dem Erlebnis und werden zu Stammkunden. Außerdem verbringen die Kellner aufgrund der standardisierten Speisekarte weniger Zeit mit dem Einsammeln der Bestellungen von großen Gruppen.

Der Restaurantbesitzer beginnt auch, die Clusterbildung zu nutzen, um vorherzusagen, wie viel Fleisch er auf Vorrat kaufen sollte, je nachdem, wie viele Reservierungen er hat. Er weiß, dass größere Gruppen höchstwahrscheinlich viel Fleisch bestellen werden, so dass das Restaurant sich effizienter auf diese Gruppen vorbereiten kann.

Schließlich nutzt das Restaurant die Clusteranalyse zur Analyse der demografischen Daten seiner Kunden. Mithilfe der Clusteranalyse kann effektiv analysiert werden, welche Arten von Kunden den größten Wert für das Restaurant haben. Anhand dieser Informationen können die richtigen Personen z.B. mittels Facebook-Werbung angesprochen werden.



Clustering-Anwendungen überall

Das Restaurant “La Estadística” ist nur eines von vielen Beispielen dafür, wie Clustering das Wachstum eines Unternehmens fördern kann. Damit Sie sich ein Bild von den Möglichkeiten machen können, finden Sie hier eine Liste von Anwendungen aus verschiedenen Branchen. Ein Webshop wie bol.com kann die Clusteranalyse nutzen, um ähnliche Online-Besucher zu gruppieren und wirksame Empfehlungen für Produkte zu geben, die ihnen gefallen könnten.

- *Netflix* nutzt die Clusteranalyse, um sowohl Nutzer als auch Filme zu gruppieren und eine nützliche Auswahl an Filmen zu zeigen, die Ihnen gefallen könnten, basierend auf Ihrem Nutzerprofil, Ihren Genre-Vorlieben und den Vorlieben vergleichbarer Nutzer.
- *Fluggesellschaften* nutzen die Clusteranalyse für die Kundenbetreuung. Es ist möglich, Text in einem numerischen Format darzustellen, und damit kann die Fluggesellschaft automatisch eine große Anzahl von Fragen und Beschwerden zusammenfassen. Diese Beschwerdekategorien können dann überwacht und auf strukturierte Weise gelöst werden. Das ist auch der Grund, warum Chatbots oft sehr generische Antworten geben.
- *Wohnungsbaugesellschaften* nutzen das Clustering, um ihre Kunden zu segmentieren und Ausreißer vorherzusagen, die ihre Miete nicht pünktlich zahlen.
- *Transportunternehmen* nutzen die Clusteranalyse für geografische Daten, um die optimalen zentralen Standorte für den Bau von Verteilzentren für ihre Waren zu ermitteln.
- *Anwaltskanzleien* nutzen Clustering, um ähnliche Rechtsdokumente zusammenzufassen und Fälle zu finden, die dem Fall, den sie gerade vorbereiten, ähneln.

Womit soll man anfangen?

Wahrscheinlich denken Sie nun bereits darüber nach, wie Sie Ihre Daten clustern können. Ich gebe Ihnen ein paar Tipps für den Anfang: Überlegen Sie zunächst, welche Daten Ihnen überhaupt vorliegen. Sie benötigen Daten in einem digitalen Format. Überlegen Sie, welche Eigenschaften bereits mit jedem Datenpunkt verknüpft sind. Überlegen Sie, welche Muster Sie erwarten, wenn Sie die Datenpunkte in eine Abbildung einzeichnen würden.

Welche Eigenschaften könnten eine Gruppe von einer anderen Gruppe unterscheiden?

Überlegen Sie nun, was Sie mit diesen Informationen anfangen könnten. Könnten Sie vielleicht Ihre Entscheidungsfindung erleichtern oder einen Mehrwert für Kunden generieren? Wie bei allen Analyse-Techniken sollten Sie auch beim Clustering nicht erwarten, dass Sie alles, was Sie brauchen, auf einem Silbertablett serviert bekommen. Sobald Sie ein Clustering vorliegen haben, liegt es an Ihnen, es zu interpretieren. Sie müssen jedes Cluster daraufhin untersuchen, warum es sich so sehr von den anderen unterscheidet und was diese Gruppe für Sie interessant macht. Dies ist der letzte wichtige Schritt, um aus dem Clustering und Ihren Daten einen Nutzen zu ziehen.



Schlussfolgerung

Unternehmen, die noch nicht damit begonnen haben, ihre Daten zur Entscheidungsfindung zu nutzen, haben vermutlich mehr Schwierigkeiten damit, die Bedürfnisse ihrer Kunden gut zu erfüllen. Clustering ist ein guter erster Schritt, um mit der Nutzung Ihrer Daten zu beginnen und dem Team Ihres Unternehmens nützliche Erkenntnisse zur Verfügung zu stellen. Überschätzen Sie nicht den Aufwand, der für den Einstieg in das Clustering erforderlich ist und unterschätzen Sie nicht den Wert, den es für Ihr Unternehmen haben kann.

