

5 »Das Datenexperiment«

In diesem Kapitel präsentiere ich ein »Datenexperiment« von Earlybird und Earlybird Digital. Von der ersten Idee bis zur Durchführung und zur Präsentation der Resultate erstreckte es sich von Ende 2016 bis Sommer 2017. Wie ich im vorigen Kapitel theoretisch andeutete, erhalten Daten ihren Wert in einem mehrstufigen Prozess der Datafizierung. Dieser umfasst die Idee der Potenzialität von Daten (ausgedrückt in Narrativen und Datenmetaphern) sowie verschiedene Transformationsvorgänge, in welchen soziale Relationen erst erzeugt und dann entfernt oder unsichtbar gemacht werden. Schließlich werden neue Relationen mit Hilfe algorithmischer Vergleiche abgeleitet. Das Datenexperiment verdeutlicht in verdichteter Form drei Probleme, mit denen sich Earlybird im Kontext ihrer Digitalisierungsstrategie beschäftigte: Das Problem der fehlenden Daten, das Problem, dass sie bisher keinen zuverlässigen, datenbasierten Mechanismus besaßen, um Member und Memberkategorien zusammenzubringen und daran anschließend das Problem, wie aus bestehenden Daten neue Relationen zwischen Membern und Dingen abzuleiten waren.

5.1 Fehlende Daten

Das erste Problem bestand darin, überhaupt Daten über die eigenen Member¹ zu generieren. »Soviel wie möglich über die Member [zu] wissen«, war das Motto von Earlybirds digitaler Transformation: Member müssen dazu gebracht werden, sich in den zur Verfügung gestellten Infrastrukturen zu bewegen und dabei Spuren zu hinterlassen. Das ist aber gar nicht so einfach: Denn Member benutzten die App nicht, ohne durch teure Maßnahmen dazu motiviert zu werden. Wie Earlybird herausfinden musste, teilen sie der

¹ Ich verwende hier anstelle von Kunden den emischen Begriff der Member, wie er von Earlybird verwendet wird.

Firma auch nicht ohne Weiteres ihre Interessen mit, wenn sie direkt danach gefragt werden.

Weshalb sind Personendaten aber überhaupt wichtig? Wie bereits geschen, kommt die Potenzialität von Personendaten in Narrativen und Metaphern zum Ausdruck. Mit dem Datenexperiment versuchten Nik und Simon konkreter aufzuzeigen, »was mit Daten möglich ist«. Sie präsentierten gemeinsam einen Satz »schöner Daten«, in welchem Member freigiebig ihre Wünsche mitteilten. Dieser sollte sowohl die Fantasie der Member als auch jene von Earlybird beflügeln. Es tauchte die Frage auf, wie man an weitere solche Daten komme beziehungsweise was die Firma den Membern bieten müsste, damit sie sich mitteilen.

Der Wettbewerb

Im Sommer 2016 führte Earlybird einen Wettbewerb für einen Businesspartner – ein Reisebüro für junge Leute – durch. Die Wettbewerbsteilnehmerinnen wurden aufgefordert, aus einem Dropdown-Menü zwischen drei Wunschdestinationen zu wählen und kurze Kommentare in Freitextfeldern zu hinterlassen. Darin sollten sie erklären, weshalb sie gerne dort hingehen und was sie dort tun würden. Den Gewinnerinnen winkten diverse Preise. Ungefähr 3000 Personen nahmen teil und das Reisebüro war zufrieden. Earlybird Digital sah aber noch mehr als bloß eine gelungene Marketingaktion in diesem Wettbewerb. Er lieferte etwas, das bisher eher schwierig zu kriegen gewesen war: persönliche Daten.

Earlybird beauftragte Simon, den Senior Data Scientist von Earlybird Digital, damit, herauszufinden, was »mit diesen Daten möglich ist«. Bei Earlybird war schon länger bekannt, dass Wettbewerbe »engagement« generieren. Womöglich ließen sich damit auch persönliche Daten von ihren ansonsten eher wenig mitteilsamen Membern generieren.

Zuvor hatte Earlybird bereits versucht, etwas über die Interessen ihrer Member zu erfahren, indem sie diese in teuren Kampagnen dazu aufforderten, ihre Memberprofile auf der Earlybird Webseite auszufüllen. Die Member sollten verschiedene Interessenskategorien wie Nachtleben, Shopping oder andere Konsumpräferenzen anklicken. Unglücklicherweise für Earlybird ignorierten die allermeisten Member diese Aufforderung. Vor diesem Hintergrund erschien der Erfolg des Wettbewerbs überraschend. Im Vergleich zu vergangenen Bemühungen, an Daten ihrer eher unkooperativen Member zu kommen, zeigte dieser Wettbewerb, dass die Member unter den richtigen

Umständen durchaus bereit wären, etwas über sich mitzuteilen – selbst dann, wenn sie in keiner Weise dazu gezwungen werden, schließlich wäre eine Teilnahme auch unter alleiniger Angabe einer Emailadresse möglich gewesen.

Bevor Earlybird damit beginnen konnte, eigene Member hinsichtlich ihrer Interessen zu kategorisieren, mussten diese erst in das »Earlybird-Universum« eintreten. Nominal werden Jugendliche zwar zum Member, wenn sie ein Bankkonto eröffnen. Ohne sich aber auf der Webseite oder in der App einzuloggen, bleiben sie in Bezug auf ihre Interessen für Earlybird unsichtbar. Der Wettbewerb stellte für Earlybird eine Lösung für das Problem dar, ansonsten ungerührte und unsichtbare Personen in »engagierte« Member zu transformieren. Der Wettbewerb war Evidenz, dass Member unter den richtigen Umständen durchaus bereit sind, zu kooperieren und etwas über sich mitzuteilen: Die Member teilen persönliche Daten. Im Gegenzug erhalten sie eine Chance, lukrative Preise zu gewinnen. Um an Daten heranzukommen, musste Earlybird affektive Relationen zwischen sich, ihren Angeboten (bzw. den Angeboten ihrer Partnerunternehmen) und ihren Membern erzeugen.

Wie sich herausstellte, erzeugte der Wettbewerb nicht nur Daten, sondern sogar »schöne Daten«. Viele der Teilnehmerinnen nahmen ihre Aufgabe äußerst ernst und lieferten detaillierte Texte, weshalb sie gerne verschiedene Wunschdestinationen besuchen würden. Diese »natural language data« versprachen in den Augen von Earlybirds Datenteam wertvolle »insights« für Marketingkampagnen.

5.2 Von Sinnüberschüssen zu verrechenbaren Daten

Das zweite Problem bei Earlybird bestand darin, eine kategoriale Verbindung zwischen vorhandenen Membersegmenten und Membern zu ziehen. Das sei mit den schönen Daten möglich, stellte Simon in Aussicht. Wie ich beschreiben werde, setzt das jedoch voraus, den in den Daten identifizierten Sinnüberschuss weitgehend zu ignorieren. Erst dann werden die verschiedenen Member vergleichbar. Hier werden die Verhaltensweisen der Nutzerinnen – i.e. Wettbewerbsantworten – zu »warenförmigen« Personendaten gemacht, indem sie von den Sinnüberschüssen, mit denen die Teilnehmerinnen sie ausgestattet haben, befreit werden. Daten werden produktiv, indem sie dazu gebracht werden, Muster und Regelmäßigkeiten zu offenbaren.

Ein Kategorisierungsexperiment: Von Freitextantworten zu Segmenten

Nach internen Diskussionen darüber, was mit den Daten möglich sei, kam Simon eine »Glühweinidee«, wie Beni es nannte. Beni und Simon hatten sich schon öfter über Earlybirds Kundensegmente unterhalten, die vor einiger Zeit von einem externen Unternehmen erstellt wurden, aber nie richtig zur Anwendung kamen. Sie verfügten über keine günstige und konsistente Methode, um Kunden und Segmente miteinander zu verlinken. Dieses Problem wollte Simon mit Hilfe der »schönen Daten« und einer Idee für algorithmische Kategorisierung lösen.

Simon und sein Team transformierten die Freitextantworten und die Beschreibungen der Jugendmilieus zu Vektoren in einem multidimensionalen Raum: Jede Teilnehmerin und alle fünf Jugendmilieus wurden als Punkte in einem hoch abstrakten Raum repräsentiert, so dass sie in Bezug auf ihre Ähnlichkeit – gemessen am Winkel ihrer jeweiligen Vektoren – verglichen werden konnten (siehe dazu genauer Kapitel 7).

Um festzustellen, wie genau diese Resultate der algorithmischen Zuordnung waren, ließ Earlybird drei Expertinnen im Bereich Jugendmarketing einen »Goldstandard« erstellen. Diese drei Expertinnen – ich war einer davon – ordneten den Jugendmilieus unabhängig voneinander 600 Textantworten zu – basierend auf »Intuition« und »Marketing Expertise«.

Die Zuordnungen des Goldstandards ließen sich dann mit den algorithmischen Kategorisierungen vergleichen. Dieser Vergleich macht nicht nur eine Aussage darüber, wie gut oder wie genau der Algorithmus kategorisieren kann. Er bringt auch manuelle und algorithmische Kategorien in einen Vergleichszusammenhang: Der Erfolg des Algorithmus wird daran gemessen, wie gut er die Urteile der Expertinnen replizieren kann. Gleichzeitig erhalten die schwierigen und in hohem Maß uneindeutigen Expertenurteile den Anschein technischer Neutralität und Objektivität, welche sie in der Praxis nicht haben. Das Experiment galt als Erfolg: Der Algorithmus produzierte knapp 40 Prozent Übereinstimmung mit den Expertenurteilen.

5.3 Neue Relationen generieren

Das dritte Problem ist jenes der Anwendung. Die schönen Daten galten als Indiz für potenzielle Relationen zwischen Membern und Dingen oder Membern und Partnerunternehmen. Die Frage, wie Member einer bestimmten Kategorie adressiert werden, überließ Simon aber weitgehend »dem Marketing«.

Für die federführenden Datenwissenschaftler war das Experiment ein Erfolg: Es resultierten computerwissenschaftliche Vorträge und ein Paper in einer einschlägigen Fachzeitschrift. Bei Earlybird setzt sich die Erkenntnis durch, dass Wettbewerbe eine wichtige, datengenerierende Ressource sind. Die vorgeschlagene Kategorisierung wurde in der beschriebenen Form aber nicht weiterverwendet. Earlybird Digital adaptierte den Classifier zu einem »freebie«, um potenzielle Kunden für Data-Science-Dienstleistungen zu gewinnen. Diese sollten den Kundinnen erlauben, mehr über ihre eigenen Nutzerinnen oder Kundinnen zu erfahren.

Der Classifier als Generator von Relationen

Meine Feldforschung bei Earlybird endete, kurz bevor Simon und sein Team ihren algorithmischen Classifier an einer akademischen Konferenz präsentierten. In informellen Interviews stellte sich heraus, dass der Classifier vor allem als computerwissenschaftliche Forschung erfolgreich war. 2019 veröffentlichten Simon und einer seiner Mitarbeiter dazu ein Paper in einer Fachzeitschrift. In Simons Vorstellung hätten die kategorialen Zuordnungen der Member ermöglichen sollen, dass das Marketingteam für die verschiedenen Memberkategorien spezifische Werbekampagnen durchführen können.² Wie mir Nik berichtete, stellte sich jedoch heraus, dass zwischen den Jugendmilieus zu geringe Differenzen bestanden, um diese interpretieren zu können. Die Textbeschreibungen der Jugendmilieus waren »Kacke«, wie er mir erklärt, da sie »nichts ausgesagt« hätten.

Der Classifier war aber nicht tot, sondern generierte unter der Bezeichnung »smart customer segmentation« neue Relationen: Nik nutzte den neuen Classifier als »freebie« für die Kundenacquirierung. Kann er einer potenziellen Kundin den »Mehrwert der Datenanalyse« vermitteln – indem er demonstriert, wie sie dank seiner Analyse neue Beziehungen zu Kundinnen knüpfen kann –, ist die Kundin möglicherweise bereit, eine längerfristige Geschäftsbeziehung mit Earlybird Digital einzugehen.

Der im Rahmen des Datenexperiments entwickelte Classifier versuchte, die Lücke zwischen Membern und den von Earlybird angebotenen »Deals« über eine »algorithmische Zuordnung« von Marketingsegmenten zu schließen. Obwohl das einigermaßen gut funktionierte, wurde der auf Segmenten basierende Classifier nicht in Betrieb genommen, sondern in abgespeckter Form ohne Segmente als »freebie« verwendet, um weitere Aufträge einzuholen.

In den nächsten drei Kapiteln formuliere ich nun auf der Grundlage meiner Feldbeobachtungen und mithilfe verschiedener theoretischer Bezugspunkte das Konzept »Momente der Datafizierung« (Kapitel 6 bis 8). Um die Lücke zwischen den kategorisierten Nutzerinnen und den Dingen konzeptuell zu schließen, lege ich dar, wie automatisierte Empfehlungssysteme die Relation zwischen Nutzerinnen und neuen Dingen knüpfen (Kapitel 9 im Allgemeinen und Kapitel 10 im Detail).

2 Schon vor dem Experiment verwendete das Marketingteam eine eigene »Taxonomie« von Deals beziehungsweise Interessenskategorien. Die von Simon verwendeten Memberkategorien – so meine Vermutung – waren beim Marketingteam gar nicht anschlussfähig.