

Mobile Analysen: Die Nutzung von Predictive Analytics ist auch per Smartphone möglich.

Komplexe Analysen: Mitarbeiter mit statistischen Kenntnissen sind das A und O bei der Einführung von Predictive Analytics in den Unternehmen.

► PREDICTIVE ANALYTICS

KNIFFELEIEN.

Wie können Unternehmen den Einsatz analytisch basierter Prognosen nutzbringender gestalten? Welche Programme, Methoden und Organisationsformen sollten dabei eingesetzt werden? Ein praktischer Leitfaden.

► Von Leonhard Geisler

Für viele ist Predictive Analytics bereits ein neuer Trendbegriff geworden, wie vorher zum Beispiel Big Data. Dabei wird Predictive Analytics oft mit Machine Learning in einen Topf geworfen, ohne dass es hiermit in der Praxis etwas zu tun haben müsste. Spricht man über die verwendeten Algorithmen, dann werden oft naturwissenschaftlich klingende Begriffe wie Naive Bayes und Random Forests genannt.

Was in der Diskussion und den Berichten zu Predictive Analytics jedoch oft fehlt, sind grundsätzlichere Fragen zum Thema: 1. Wen muss ich im Unternehmen von Predictive Analytics überzeugen, damit das Projekt ein Erfolg wird? 2. Ist unser Unternehmen reif für ein Predictive Analytics-Projekt, oder reicht die bislang manuell erstellte Planung? 3. Für welche Methoden eignen sich die Stammdaten und was müssen wir eventuell an ihnen ändern, um sie für die Erstellung einer automatisierten Vorhersage zu präparieren?

Bei meiner früheren Tätigkeit als Data Analyst bei der Primavera Life GmbH

zeichnete sich für mich bereits ab, dass derartige Fragen geklärt werden müssen, bevor überhaupt auch nur ansatzweise über die Verwendung von Machine Learning nachgedacht werden sollte.

Für mich ist der Weg zu fortgeschrittenen und komplett automatisierten Methoden der Predictive Analytics eher ein evolutionärer Prozess. In dessen Verlauf müssen etwa folgende Fragen beantwortet werden: Welche Hürden existieren in diesen Prozessen? Auf welche Weise lassen sich pragmatische Lösungsansätze finden? Und bevor schließlich der Startschuss für ein Predictive Analytics-Projekt fällt, muss geklärt werden: Warum soll die Technologie überhaupt eingesetzt werden? Reicht die bisherige Vorhersagemethode nicht aus?

Grundsätzliches Herangehen.

Um diese Fragen zu beantworten, muss zunächst einmal genau aufgelistet werden, wie Vorhersagen in einem bestimmten Unternehmen bisher erstellt wurden. Bei Primavera Life, einem Spezialherstel-

ler naturreiner ätherischer Öle für die Aromatherapie und Naturkosmetik, lief der Prozess früher so ab, dass ein rollierender Forecast für rund 2'000 Produkte quartalsweise manuell mithilfe von Tabellenkalkulationen und einer OLAP-Datenbank von Mitarbeitern erstellt wurde. Der aus dem Forecast resultierende Plan lieferte durchaus gute Vorhersagen.

Die Zielsetzung der gewünschten Predictive Analytics-Lösung bestand hauptsächlich darin, die Mitarbeiter zu entlasten, die mit dem Forecast betraut sind. Daneben spielte auch die Optimierung der Lieferkette eine Rolle. Denn eine Produktpalette mit vielen exotischen Rohstoffen und einer Wiederbeschaffungszeit von teils mehr als einem Jahr muss mit Bedacht geplant werden (siehe auch BUSINESS INTELLIGENCE MAGAZINE Nr. 1/2019, Seite 26 ff.)

Know-how-Basis.

Eine große Herausforderung stellt auch die Komplexität der neuartigen Lösungen dar. Es fehlt intern oft an dem nötigen

Know-how, um Predictive Analytics zu implementieren. Laut einer BARC-Studie¹ in der DACH-Region betrifft dieser Mangel an Expertise sowohl Unternehmen, in denen die Fachbereiche das Thema vorantreiben, als auch in solchen, in denen es den Data-Scientists obliegt. Gerade in puncto statistisch-mathematisches Verständnis, Datenaufbereitung und Tool-Kenntnisse besteht oft Nachholbedarf.

Um das notwendige Know-how bereitzustellen, kann man auf externe Partner zurückzugreifen oder interne Experten ausbilden. In einer Studie von IDG aus dem Jahr 2017 mit über 400 Teilnehmern² sind es 83 Prozent der befragten Unternehmen, die einen externen Partner als Unterstützung hinzuziehen. Und 47 Prozent der Betriebe mit weniger als 100 Mitarbeitern gaben sogar an, gleich mehrere Dienstleister zum Thema Predictive Analytics zu beschäftigen.

Bei Primavera wird internes und externes Know-how kombiniert: Zwei angestellte Datenanalysten kümmern sich einerseits um das interne Berichtswesen und stehen als Experten für hauseigene Anfragen zur Verfügung; außerdem sind sie das Bindeglied zu externen Partnern, wie zum Beispiel zu Hendricks, Rost und Cie. Diese unterstützen bei der Auswahl des für die Lösung passenden Predictive Analytics-Programms und des Vorhersagemodells.

Damit wurde sichergestellt, dass die Predictive-Modelle von geschulten Schlüsselnutzern selbstständig angewendet werden können.

Datenvorbereitung als Basis.

Doch selbst wenn das Vorgehen klar definiert ist und alle Beteiligten in den Schlüsselpositionen mitspielen, tauchen später Schwierigkeiten bei der Datenvorbereitung auf. In der IDG-Studie nennen 28

Prozent der befragten Unternehmen eine mangelnde Datenqualität als größtes Problem bei der Einführung von Predictive Analytics.

Fehler und Lücken in den Stammdaten können hierbei eine verheerende Rolle spielen und die Predictive-Modelle unberechenbar machen. Haben sich die Datenexperten im Unternehmen kein umfassendes Verständnis der Strukturen erarbeitet, stehen sie vor großen Herausforderungen, wenn sie zum Beispiel einer Zeitreihe das entsprechende Modell zuweisen möchten.

Ein unzureichendes internes Datenverständnis ist auch laut der IDG-Studie das Hauptproblem in vielen Unternehmen: 42 Prozent sehen hierin die größte Herausforderung bei Predictive Analytics-Projekten. In der Studie von BARC zeigt sich darüber hinaus, dass vor allem die Unternehmen, in denen die Fachbereiche besonders aktiv sind, noch Nachholbe-

darf in Sachen Datenaufbereitung haben: 36 Prozent der Befragten bemängeln dies.

Bei Primavera hatten wir das Glück, auf den reichen Erfahrungsschatz einer langjährigen Mitarbeiterin zurückgreifen zu können, welche zu jedem Produkt und der dazugehörigen Zeitreihe ein weit zurückreichendes Detailwissen hatte.

Da aber Unternehmen normalerweise nicht über ein solches Expertenwissen verfügen, muss am Anfang jedes Predictive Analytics-Projekts die Erarbeitung eines Datenverständnisses mittels Data Mining stehen. Dies wird genutzt, um Muster in den Zeitreihen zu erkennen und diese somit Clustern und den passenden Modellen zuweisen zu können. Aus diesem Data Mining-Prozess erwächst ein wohlgeordneter Entscheidungsbaum, welcher möglichst viele Sonderfälle abgreift und fall-spezifischen Algorithmen zuordnet.

Die Grenzen neuronaler Netze.

Und wie ist es mit dem Einsatz Neuronaler Netze? Bei Primavera entschieden wir uns dagegen. Zwar wurde ihre Verwendung anfangs in Betracht gezogen, jedoch später in Abstimmung mit den externen Experten verworfen. Gründe dafür waren unter anderem die durchschnittliche Länge der zu betrachtenden Zeitreihen, die fehlenden Erklärungsvariablen und der ungewisse Nutzen, der sehr wahrscheinlich einen größeren Implementierungsaufwand nach sich gezogen hätte. Zudem sind bei einer Modellierung auf der Basis

neuronaler Netze die Ergebnisse oft nur schlecht erklärbar.

Außerdem wurde bei Primavera davon abgesehen, die implementierte Lösung mit Maschinellem Lernen (ML) zu kombinieren. Eine Erweiterung der Automatisierung der Planung ist zwar ange-dacht, jedoch sprachen zwei Argumente gegen den ML-Einsatz:

Einerseits ist die Technologie häufig auf externe Rechenleistung in Form von zusätzlicher Hardware angewiesen. Andererseits müsste der zugrundeliegende Entscheidungsbaum zu viele Sonderfälle im Produktportfolio von Primavera Life abgreifen.

Um zunächst ein Gespür für die Eigenheiten der Daten zu bekommen, entschied sich das Management letztendlich für einen manuell überwachten Lernprozess. Dieser ermöglicht es, zusätzlich zur automatischen Modellfindung, das interne Know-How bei der Optimierung der Vorhersagen mit einzubeziehen.

Diese eher vorsichtige Herangehensweise kann für viele Unternehmen empfohlen werden. Denn wenn Predictive Analytics-Modelle ohne tieferes Datenverständnis automatisiert werden, entsteht schnell die Gefahr der Black Box.

Sicherung der Qualität.

Um die Qualität zu sichern, ist es – wie das Praxisbeispiel zeigt – sehr wichtig, dass die Modelle nach ordentlichen Gütekriterien bewertet werden. Denn nicht jede Kennzahl für Modellgüte eignet sich in der Praxis. Eine qualitativ hochwertige Prognose benötigt das ständige Weiterentwickeln geeigneter Kennzahlen zur Modellfindung.

Im Extremfall kann das Fehlen eines objektiven Rahmens sogar dazu führen, dass die gängigen Maßstäbe auch zu unrealistischen Modellen führen. Die Lösung: Durch die Implementierung der bei Statistikanwendungen bewährten Programmiersprache R in die bestehende Lizenzsoftware SPSS Modeler wurden bei Primavera zusätzliche Gütekriterien zur Verfügung gestellt, die es ermöglichen, die Modelle auch dann zu optimieren, wenn keine Ist-Daten mehr zur Modellevaluation zur Verfügung stehen.

Die Wahl der passenden Softwarelösung für Predictive Analytics stellt eine nicht zu unterschätzende Herausforderung

dar. Denn mittlerweile gibt es ein kaum noch überschaubares Sammelsurium an Werkzeugen.

Die Grundsatzdebatte über den Einsatz von Open Source-Software versus lizenzierter und vorprogrammierter Lösungen macht die Sache nicht einfacher. Lizenzierte Lösungen sind meistens anwenderfreundlicher, und sie bieten eingebaute Schnittstellen zu gängigen Datenbanken. Was dagegen spricht, ist der oft hohe Anschaffungs- oder Mietpreis.

Dennoch entschied sich Primavera für die Analytics-Software SPSS Modeler, die einen eher intuitiven, anwenderfreundlichen Charakter hat, einem Baukastensystem gleicht und zu der auch bereits Grundkenntnisse im Unternehmen vorhanden waren. Trotz des enormen Funktionsumfangs beinhaltet dieses Programm zusätzlich eine eigene Schnittstelle zur Programmiersprache R, welche als Ergänzung des vorhandenen Zeitreihenmodells verwendet wird.

Die Vorteile von Open Source.

Grundsätzlich jedoch spricht nichts gegen Open Source. Denn die Programmiersprachen wie R und Python werden von einer lebendigen Gemeinschaft aus Programmierern, Entwicklern und Anwendern getragen, ständig angepasst und erweitert.

Mehr noch: Wenn es in der Organisation Mitarbeiter gibt, die diese Programmiersprachen beherrschen, liegt es sogar nahe, eine eigene Lösung für Predictive Analytics-Aufgaben zu entwickeln. Bereits 43 Prozent der Unternehmen mit weniger als 100 Mitarbeitern setzt laut IDG auf selbstentwickelte Lösungen.

Die Vorteile dieser unternehmensspezifischen Analytics-Lösungen liegen auf der Hand: Man spart sich Lizenzkosten, schmort nicht in der Support-Warteschlange, wenn doch etwas mit der lizenzierten Lösung nicht klappt und kann Modelle besser auf die individuellen Bedürfnisse zuschneiden.

Oft hält die Lizenzsoftware auch nicht das, was sie verspricht: Der IDG-Studie zufolge bemängelt mehr als ein Viertel der Befragten, dass die bereits implementierten Lösungen im Advanced Analytics-Bereich nicht den erhofften Funktionsumfang bieten. Darüber hinaus fehle den Analytics-Lösungen in vielen Fällen die Flexibilität, die notwendig sei, um die

Detaillierte Analysen: Eine Expertin beim Drilldown in einer Übersicht für den Vertrieb.

Software auf die Bedürfnisse im eigenen Unternehmen anzupassen.

Bei Primavera wurde dieses Problem durch die Integration von R in die vorhandene Software umgangen. Diese Schnittstelle ermöglicht es zusätzlich zu den bereits vielen Funktionen des Tools, auch noch die enorme Masse kostenloser R-Pakete zu nutzen.

Tools und individuelle Präferenzen.

Zusammenfassend teile ich die Meinung vieler Kollegen aus der betrieblichen Praxis, dass es weder die Software noch der Algorithmus ist, welche Geschäftsprobleme lösen können. Data Scientists verwenden zu Recht gern vorgefertigte Algorithmen, die in ein Datenmodell eingebettet werden. Statt das Rad neu zu erfinden, kommt es darauf an, die richtigen Rahmenbedingungen zu schaffen.

Zu dieser Grundlage zählen vor allem ein ordentliches Change Management, die

für das Unternehmen passende Struktur von internem und eventuell externem Know-how, verlässliche Stammdatenpflege, ein tiefgreifendes Datenverständnis, sowie eine akribische Datenaufbereitung und gewissenhafte Modellevaluation. Vor allem in kleineren Unternehmen sollte man ein Predictive Analytics-Projekt zunächst in einem übersichtlichen Bereich beginnen – mit klarer, realistischer Zielsetzung und verständlichen Algorithmen.

In Bezug auf das Know-how für Predictive Analytics gibt es zwei Aspekte zu bedenken: Einerseits bringen die unternehmensinternen Experten einen enormen Vorteil, wenn es um das Datenverständnis und die Datenvorbereitung geht. Dies können auch Schlüsselanwender aus den Fachbereichen sein.

Andererseits sollte man, vor der Akquisition eines internen Data Scientists darüber nachdenken, ob dieser benötigt wird, um kurzfristige Arbeitserleichterungen

zu ermöglichen, oder ob das Unternehmen eine fortlaufende Unterstützung bei seinen Geschäftsentscheidungen durch Advanced Analytics möchte – was für Organisationen gilt, die stark auf Data Science setzen. Im Extremfall kann es sonst passieren, dass sich der Data Scientist durch seine Arbeit selbst überflüssig macht. In diesem Fall wäre ein externer Data Scientist die richtige Wahl gewesen.

Auch die Frage nach der geeigneten Software lässt sich nicht eindeutig beantworten. Zwar bin ich ein Verfechter von Open Source-Programmen, weil diese weitaus flexibler und umfangreicher sind als lizenzierte Produkte. Jedoch möchte ich nicht generell von letzteren abraten.

In jedem Fall wichtig sind die Überlegungen zur Wahl des Tools vor dem Projektbeginn. Zu den individuellen Präferenzen zählen unter anderem der Stand der IT-Systemintegration, das vorhandene Know-how und nicht zuletzt das Budget. ■



Foto: HENDRICKS, ROST & CIE

» Leonhard Geisler ist Data Scientist bei der Düsseldorfer Unternehmensberatung HENDRICKS, ROST & CIE. Er begeistert sich beruflich wie auch privat für Data Analytics, Machine Learning und KI.



Foto: Shutterstock/Alcedorino