DGI-Praxistage 2016

Dr. Gisela Kramer*, Dr. Frauke Schleer-van Gellecom, Alexander Häußer und Tobias Flath

Valide Prognosen mit der Predictive Analytics Suite

DOI 10.1515/iwp-2017-0013

In einem zunehmend globalisierten und digitalisierten Wirtschaftsumfeld, in dem sich relevante Informationen häufig ändern, wird objektive und fundierte Planungsund Entscheidungssicherheit für Unternehmen immer wichtiger. Während die Anzahl der Einflussfaktoren (Treiber) auf eine Planungs- oder Steuerungsgröße in global vernetzten Wertschöpfungsketten steigt, erschwert die hohe Informationsgeschwindigkeit zugleich das Herausfiltern der entscheidungsrelevanten Informationen. Aus diesem Grund haben wir – ein Team aus PwC Risk Consulting und PwC Economics – die *Predictive Analytics Suite (PAS)* entwickelt, die relevante Treiber der Plangröße identifiziert, diese in ein Prognosemodell überführt und anschließend die Treiber und die Plangröße prognostiziert. So liefert unsere Applikation nachvollziehbare, auf statistischen Methoden aufbauende Prognosen für Plangrößen wie Absatzmengen, Preise oder Herstellungskosten.

Anstatt auf eine Vielzahl von Indikatoren achten zu müssen, werden unsere Mandanten in die Lage versetzt, mittels der identifizierten Treiber die relevanten Erkenntnisse aus den Daten zu gewinnen und zur Prognose zu nutzen. Dies ermöglicht ihnen neue Informationen direkt in aktualisierte Prognosen umzusetzen und in die zukünftige Unternehmensplanung und -steuerung einzubeziehen. Die Mandanten sind damit in der Lage proaktiv zu handeln, anstelle allein auf die Entwicklungen des Marktes zu reagieren. Zudem können sie Chancen frühzeitig erkennen, sich auf zukünftige Herausforderungen vor-

bereiten und somit Risiken besser einschätzen und sogar vermeiden.

Nachfolgend wird ein kurzer Überblick über die Anwendung gegeben. Zu diesem Zweck werden ausgewählte Funktionalitäten und die zugrundeliegenden Methoden skizziert – angefangen vom Import der Daten über deren Visualisierung und Analyse bis hin zur Modellierung und Prognose. Im Anschluss wird anhand eines kurzen Fallbeispiels im Kontext des Weltstahlpreises gezeigt, wie mit PwC's Predictive Analytics Suite Prognosen für eine bestimmte Plangröße generiert werden können. Abschließend werden die wichtigsten Punkte kurz zusammengefasst.

Funktionalitäten und Methodik der Anwendung

Die PwC's Predictive Analytics Suite ist vollständig in der auf statistische Anwendungen ausgerichteten Programmiersprache R entwickelt. Dies bietet unseren Mandanten die Möglichkeit, bewährte statistisch-ökonometrische Methoden sowie auch innovative Machine Learning-Algorithmen für unternehmensspezifische Prognosen zu nutzen. Die Predictive Analytics Suite ist eine cloudbasierte Anwendung. Sie verfügt über eine intuitive graphische Benutzeroberfläche, die auch Anwender ohne Programmierkenntnisse Schritt-für-Schritt durch das Tool führt und erlaubt, valide Prognosen zu erstellen. Grundsätzlich liegt der Analyse eine dreistufige Vorgehensweise zugrunde, bestehend aus Datenanalyse, Modellierung und Prognose (s. Abb. 1).

1 Datenanalyse

Die gemeinsam mit den Mandanten identifizierten, potentiellen Einflussfaktoren und die Plangröße werden zu Beginn der Datenanalyse in die Anwendung importiert. Neben einer direkten Schnittstelle für diverse Datenbanken besteht die Möglichkeit, die relevanten Zeitreihen (z.B.

Dr. Frauke Schleer-van Gellecom PricewaterhouseCoopers AG, Friedrich-Ebert-Anlage 35-37, 60327 Frankfurt am Main, E-Mail: frauke.schleer-van.gellecom@de.pwc.com

Alexander Häußer, PricewaterhouseCoopers AG,

Friedrich-Ebert-Anlage 35-37, 60327 Frankfurt am Main,

E-Mail: alexander.haeusser@de.pwc.com

Tobias Flath, PricewaterhouseCoopers AG, Bernhard-Wicki-Straße 8, 80636 München, E-Mail: tobias.flath@de.pwc.com

^{*}Kontaktperson: Dr. Gisela Kramer, PricewaterhouseCoopers AG, Friedrich-Ebert-Anlage 35-37, 60327 Frankfurt am Main, E-Mail: gisela.kramer@de.pwc.com







Abbildung 1: PwC's Predictive Analytics Suite (Quelle: PwC AG).

Monats- oder Quartalsdaten) von einer lokalen Datenquelle zu importieren. Die Daten werden zudem soweit sinnvoll für die weitere Analyse vorverarbeitet: Einzelne Beobachtungen (bspw. Ausreißer) werden manuell angepasst oder entfernt, und der gesamte Datensatz kann in zwei Teile, sog. Trainings- und Testdaten, zerlegt werden. Auf Basis der Trainingsdaten werden die Erklärungsmodelle gebildet, die übrigen Testdaten werden zur Evaluation der Prognosegüte verwendet.

Die Anwendung ermöglicht unterschiedlichst Arten der Datenvisualisierung. Neben der Darstellung einzelner oder mehrerer Zeitreihen in klassischen Liniendiagrammen können auch Verteilungen der Variablen in Form von Histogrammen oder Box-Whisker-Plots dargestellt werden. Zur Analyse der Beziehungen zwischen zwei oder mehr Variablen können diese gegeneinander in einem Scatterplot oder einer sog. Scatterplot-Matrix dargestellt werden. Letzteres bietet die Möglichkeit, die Interaktion mehrerer Variablen in einem einzigen Diagramm zu erfassen. Neben der rein graphischen Analyse werden die Daten einer umfangreichen Korrelationsanalyse unterzogen. Dazu werden die Wechselbeziehungen zwischen der Plangröße und sämtlichen Treibern ermittelt. Zum Beispiel misst die Kreuz-Korrelation den linearen Zusammenhang zwischen der Plangröße y(t) und einem potentiellen Treiber x(t-k), wobei k den zeitlichen Vorlauf des Treibers, das sogenannte "Time-Lag", bezeichnet (s. Abb. 2). Werden signifikante Wechselbeziehungen gefunden, deutet dies darauf hin, dass der Treiber einen zeitlichen Vorlauf vor der Plangröße besitzt. Solche Vorlaufbeziehungen werden in der Modellierung berücksichtigt und verbessern eine Prognose. Die so identifizierten Einflussgrößen dienen als Grundlage für das im nächsten Schritt zu entwickelnde Erklärungsmodell.

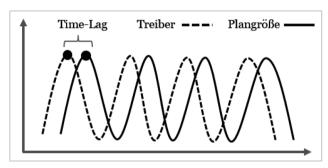


Abbildung 2: Vorlaufender Treiber und Plangröße (Quelle: PwC AG).

2 Modellierung

Nachdem die Daten in einem ersten Schritt visualisiert, analysiert und aus der Vielzahl der potentiellen Einflussgrößen die relevanten Treiber identifiziert wurden, können diese in einem nächsten Schritt in ein Erklärungsmodell überführt werden. Hierfür stehen verschiedene Modellierungsansätze zur Auswahl, die neben innovativen Machine Learning-Algorithmen auch statistisch-ökonometrische Methoden, beispielsweise die Berechnung einer multivariaten Regression, beinhalten. In multivarianten Regressionsmodellen werden die Beziehungen zwischen einer abhängigen Variablen (Plangröße) und mehreren unabhängigen Variablen (relevante Treiber) erfasst. Um den Einfluss der einzelnen Treiber auf die Plangröße zu quantifizieren, wird das Modell mittels Kleinste-Quadrate-Methode geschätzt. Es besteht die Möglichkeit, dass ein automatischer Optimierungsalgorithmus ein optimales Modell identifiziert. Dazu müssen lediglich die Variablen und eine Obergrenze der Time-Lags angegeben werden. Der Algorithmus prüft daraufhin sämtliche Kombinationen zwischen allen Variablen und Time-Lags und entscheidet selbstständig anhand eines statistischen Gütemaßes, welche Kombination ein geeignetes Modell liefert. Alternativ kann ein manuelles Regressionsmodell mit den gewünschten Treibern und den dazugehörigen Zeitverzögerungen (Time-Lags) erstellt werden, um kundenspezifische Besonderheiten zu berücksichtigen.

Danach erfolgt die Berechnung des identifizierten Erklärungsmodells auf Basis historischer Daten. Besonderheiten in den Zeitreihen wie etwa Saisonalität (z.B. Kalendereffekte) oder andere charakteristische Muster können ebenfalls berücksichtigt werden. Im Anschluss an die Modellierung wird das Erklärungsmodell einer umfangreichen Diagnose unterzogen. Zum einen erfolgt eine Überprüfung der statistischen Signifikanz und ökonomischen Plausibilität des Erklärungsmodells. Zum anderen wird anhand diverser statistischer Kennzahlen betrachtet, wie gut das Modell die Testdaten erklärt (Anpassungsgüte). Neben der rein quantitativen Beurteilung eines Erklärungsmodells werden die Ergebnisse auch graphisch mit den Trainingsdaten verglichen. Basierend auf einem identifizierten Modell erfolgt die Prognose der interessierenden Plangröße. Im folgenden Abschnitt wird die Erstellung einer solchen Prognose detailliert beschrieben.

3 Prognose und Evaluation

Im Falle der im vorherigen Schritt selektierten multivariaten Regression - ein mögliches Modell innerhalb der Predictive Analytics Suite - ist es notwendig, für jeden einzelnen Treiber eine individuelle Prognose durchzuführen. Diese werden anschließend im Regressionsmodell zusammengesetzt. Neben Punktprognosen der Plangröße werden die Ergebnisse zusätzlich in Bandbreiten, den sog. Prognoseintervallen (PI), angegeben. Diese bilden die Unsicherheit einer Prognose ab, indem sie festlegen, mit welcher Wahrscheinlichkeit ein zukünftiger Wert innerhalb des angegebenen Prognoseintervalls liegen wird. PI bilden somit einen objektiven Rahmen für planerische und strategische Entscheidungen, in dem Best- und Worst-Case-Szenarien abgeleitet werden können. Wenn der vorhandene Datensatz vorab in zwei Teile zerlegt wurde, kann dieser zur Bewertung der Prognosen verwendet werden. Anhand der Testdaten werden die Abweichungen zwischen den Ist- und den Prognosewerten, die sog. Prognosefehler, berechnet. Auf Basis dieser Prognosefehler werden diverse statistische Kennzahlen abgeleitet, die die Prognosegüte des Modells quantitativ beschreiben und somit eine Aussage über die Qualität des Prognosemodells erlauben.

In einem abschließenden Schritt besteht die Möglichkeit, anhand rollierender Prognose die Generalisierbarkeit des Modells, d.h. die Prognosequalität des Modells für verschiedene Zeiträume innerhalb des gesamten Betrachtungszeitraums, zu überprüfen. Die Ergebnisse der Datenanalyse und Prognose können in standardisierten Formaten, z.B. CSV und PDF, als Reports abgespeichert und weiterverarbeitet werden und so als Grundlage für fundierte planerische und strategische Entscheidungen dienen.

Anwendungsbeispiel

Anhand eines Anwendungsfalls mit der Plangröße "Weltstahlpreis" wird die Funktionsweise und Leistungsfähigkeit der Predictive Analytics Suite veranschaulicht. Am Anfang erfolgt die Zusammenstellung einer Longlist potenziell relevanter Einflussfaktoren auf die Plangröße, hier den Weltstahlpreis. Hierbei werden sowohl Makrofaktoren (z. B. Einkommensentwicklung, Demografie, Konsumklima) als auch Mesofaktoren (z. B. Branchenwachstum) berücksichtigt. Soweit vorhanden, finden auch Mikrofaktoren (z. B. Auftragseingang) Verwendung, die für das betrachtete Unternehmen spezifisch sind. Die Evaluation der Vielzahl der identifizierten potentiellen Treiber erfolgt im Anschluss auf Grundlage der visuellen Auswertung, sowie der Übersicht der Korrelationen zwischen der Plangröße und sämtlichen Treibern.

Im Fallbeispiel werden auf Grund der starken Korrelationen der Plangröße Weltstahlpreis z.B. mit den Treibern Metal Index, Fuel Energy Index, Iron Ore, Arca Steel Index und Arcelor Mittal (Aktienkurs) diese Variablen als potentielle Einflussgrößen identifiziert und bilden somit die Shortlist der Treiber. Diese selektierten Treiber werden anschließend detailliert in ihrer Beziehung zur Plangröße analysiert. (s. Abb. 3).

Die Wahl eines geeigneten Erklärungsmodells erfolgt nun auf Basis des automatischen Optimierungsalgorithmus. Dazu werden lediglich die vorausgewählten Treiber und eine Obergrenze für die Time-Lags gewählt. Der Optimierungsalgorithmus entfernt die Variable *Arcelor Mittal* und identifiziert automatisch ein optimales Erklärungsmodell.

Im dritten Schritt erfolgt basierend auf dem gewählten Erklärungsmodell die Berechnung der Prognosen (s. Abb. 4). Die graphische Auswertung zeigt, dass die tatsächlichen Werte innerhalb der Prognoseintervalle liegen, die jeweils die Grundlage für Best- und Worst-Case Szenarien darstellen und darüber hinaus sehr wenig von der Punktschätzung abweichen. Dies verdeutlicht den

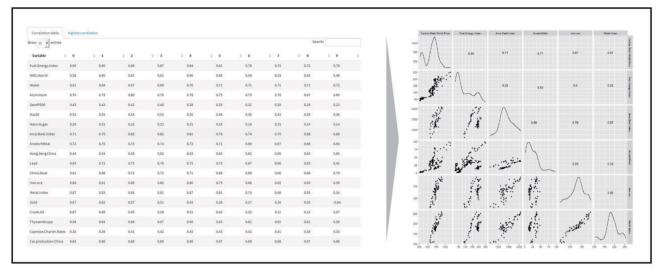


Abbildung 3: Der Selektionsprozess von der Long- zur Shortlist (Quelle: PwC AG).

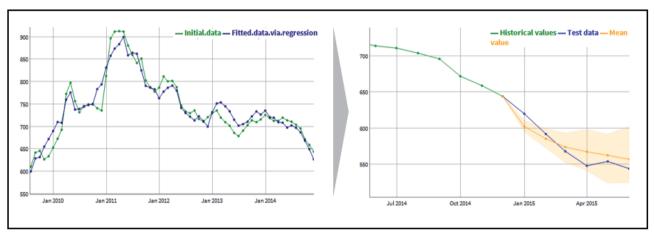


Abbildung 4: Vom Erklärungsmodell zur Prognose der Plangröße (Quelle: PwC AG).

Mehrwert, den PAS-gestützte Prognosen für eine effiziente Planung bieten und den Beitrag zur Optimierung der Unternehmenssteuerung, indem Kosteneinsparungen und eine proaktive Risikosteuerung ermöglicht werden.

Fazit

Die PwC's Predictive Analytics Suite ermöglicht es Mandanten, in Zusammenarbeit mit PwC die wesentlichen Treiber von Plangrößen zu identifizieren, neue Informationen zum Marktumfeld automatisch und schnell zu berücksichtigen und somit mittels valider Prognosen Entscheidungsgrundlagen zu optimieren. Die benutzerorientierte Anwendungskonzeption erlaubt es, die Vorteile bewährter und innovativer statistischer und ökonometrischer Modelle für eine Vielzahl von Fragestellung mit reduziertem Aufwand und hoher Mandantenorientierung zugänglich zu machen und die Herausforderung der Planung in einem interdependenten und sich schnell ändernden Wirtschaftsumfeld zu meistern.

Deskriptoren: Decision Support System, Datenanalyse, Entscheidungshilfe, Prognose, Innerbetriebliche Information

Literatur

Chatfield, C. (2003). The Analysis of Time Series: An Introduction. Chapman and Hall/CRC, 6th Edition.

Fox, J., Weisberg, H. S. (2011). An R Companion to Applied Regression. Sage Publications Inc., 2nd Edition.

Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press.

Hyndman, R.J, Athanasopoulos, G. (2013). Forecasting: Principles and Practice. OTexts.

Kapoor, S. G., Madhok, P., Wu, S. M. (1981). Modeling and Forecasting Sales Data by Time Series Analysis. Journal of Marketing Research, Vol. 18, No. 1, pp. 94–100.

Makridakis, S.G, Wheelwright, S.C., Hyndman, R.J. (1998). Forecasting: Methods and Applications. John Wiley & Sons, 3rd Edition.



Dr. Gisela Kramer
PricewaterhouseCoopers AG
Friedrich-Ebert-Anlage 35-37
60327 Frankfurt am Main
gisela.kramer@de.pwc.com

Gisela Kramer ist Partnerin im Geschäftsbereich Advisory im Bereich Finance & Regulation bei PricewaterhouseCoopers in Deutschland. Frau Kramer arbeitet seit mehr als 15 Jahren bei PwC und verantwortet in Deutschland die Bereiche Public Policy und PwC Economics.



Dr. Frauke Schleer-van Gellecom
PricewaterhouseCoopers AG
Friedrich-Ebert-Anlage 35-37
60327 Frankfurt am Main
frauke.schleer-van.gellecom@de.pwc.com

Frauke Schleer-van Gellecom ist Managerin im Geschäftsbereich Advisory im Bereich Finance & Regulation, Economics & Policy bei PricewaterhouseCoopers in Deutschland. Frau Schleer-van Gellecom leitet das deutsche PwC Economics Team. Vor ihrer Zeit bei PwC war sie am Zentrum für Europäische Wirtschaftsforschung tätig.



Alexander Häußer
PricewaterhouseCoopers AG
Friedrich-Ebert-Anlage 35-37
60327 Frankfurt am Main
alexander.haeusser@de.pwc.com

Alexander Häußer ist Consultant im Geschäftsbereich Advisory im Bereich Finance & Regulation, Economics & Policy bei PricewaterhouseCoopers in Deutschland. Herr Häußer wirkt bei der Weiterentwicklung der PwC's Predictive Analytics Suite mit.



Tobias Flath
PricewaterhouseCoopers AG
Bernhard-Wicki-Straße 8
80636 München
tobias.flath@de.pwc.com

Tobias Flath ist Senior Manager im Geschäftsbereich Advisory im Bereich Risk Consulting bei PricewaterhouseCoopers in Deutschland. Herr Flath arbeitet seit über zehn Jahren bei PricewaterhouseCoopers und hat sich auf strategisches und quantitatives Risikomanagement spezialisiert. Er verantwortet bei PwC den Bereich Risk Analytics.