



Leitfaden zur Einführung einer Predictive Analytics Lösung

Inhalt

Darauf sollten Unternehmen bei der Implementierung einer Predictive Analytics-Lösung achten.....	3
Voraussetzung: Leistungsstarke Datenbank-Infrastruktur	3
Auswahl der Datensätze	3
Methodik	4
Folgende Vorgehensweise ist zu empfehlen.....	5
Die sechs Schritte der prädiktiven Analytik	5
Die gewonnenen Erkenntnisse erfolgreich in die Geschäftspraxis übertragen.....	6
Gute Prognose, schlechte Prognose	6
Fazit.....	6

Darauf sollten Unternehmen bei der Implementierung einer Predictive Analytics-Lösung achten

Voraussetzung: Leistungsstarke Datenbank-Infrastruktur

Grundlegende Voraussetzung für die effiziente Datenanalyse ist ein leistungsfähiges Data Warehouse, beziehungsweise eine geeignete Big Data-Infrastruktur, die auch große Datenmengen schnell auswertet. Im Rahmen von Predictive Analytics ist es notwendig, Daten aus unterschiedlichsten Quellen zu aggregieren und zu analysieren. Sämtliche Datenquellen (Maschinen, Prozesse, Produkte, Wetterdaten etc.) und Systeme (ERP, CRM, BI etc.) müssen miteinander integriert werden.

Eine weitere Herausforderung ist es, die große Vielfalt von Datentypen und Formaten sowie die Qualität der einzelnen Daten umfassend zu erfassen und daraus entsprechende Erkenntnisse zu gewinnen. So sollten die Daten alle Kriterien für hohe Qualität erfüllen wie Korrektheit, Konsistenz, Vollständigkeit, Aktualität oder Einheitlichkeit. Insbesondere für Predictive Analytics müssen genügend historische Daten vorhanden sein, damit sich diese mit geeigneten Prognosetechniken aussagekräftig in die Zukunft fortschreiben lassen.

Auswahl der Datensätze

Nach der Analyse der Geschäftsprozesse und dem Festlegen der Ziele folgt die Auswahl der Daten. Anstatt die komplette Datenbasis zu verwenden ist es meist wirtschaftlich sinnvoller, die Daten zuzuschneiden, die später über Algorithmen und Vorhersagemodelle ausgewertet werden.

Zunächst sollten die Daten durch das Einbringen von Domänenwissen reduziert werden. Anschließend sollten Daten die für den Use Case eindeutig irrelevant sind, entfernt werden. Erst dann können Daten, die statistisch nicht relevant sind oder mit anderen Daten stark korrelieren (einfaches Beispiel: Temperatur in Celsius und Temperatur in Fahrenheit) aus den Modellen heraus genommen werden.

Die Auswahl der Datensätze hängt immer von der Fragestellung ab, lässt sich aber Schritt für Schritt erweitern. Interne Daten werden meist um externe Informationen ergänzt, um aussagekräftige Prognosen zu erhalten.

Methodik

Es gibt viele unterschiedliche Methoden und Vorhersagemodelle, um aus Daten Prognosen abzuleiten. Sie reichen von klassischen Data-Mining-Methoden wie Clustering oder Regressionsanalyse über Elemente der Spieltheorie bis hin zum maschinellen Lernen. Bei letzterem werden die Algorithmen so trainiert, dass sie aus den vorliegenden Daten lernen, selbstständig ein Datenmodell erzeugen und dieses für Prognosen oder Entscheidungen einsetzen.

Prognosen sind immer nur so gut wie die zugrundeliegenden Daten. Entsprechend wichtig ist eine qualitätsgesicherte, solide Datenbasis. Vor allem beim Supervised Learning, bei dem die Stammdaten die Grundlage für das Lernen des Algorithmus bilden, können mit einer durch Stammdatenmanagement qualitativ hochwertig aufbereiteten Datengrundlage deutlich bessere Vorhersagen erzielt werden als beim Machine Learning mit einem unbehandelten Daten Set.

Folgende Vorgehensweise ist zu empfehlen

Die sechs Schritte der prädiktiven Analytik

1. Daten identifizieren. Potenziell wertvolle Daten liegen oft an mehreren schwer zugänglichen Stellen vor, sowohl intern (Datensilos in Unternehmensanwendungen) als auch extern (Social Media, Regierungsdaten und andere öffentliche oder lizenzierte Datenquellen). Datenvisualisierungstools können dabei helfen, die Daten aus verschiedenen Quellen dahingehend zu untersuchen, welche für ein Predictive Analytics-Projekt relevant sein könnten.
2. Daten aufbereiten. Dieser Punkt ist für die prädiktive Analytik eine zentrale Herausforderung. Viele Anwender von Predictive Analytics verbringen mehr als drei Viertel ihrer Zeit mit der Aufbereitung der Daten: Berechnung von aggregierten Feldern, Entfernen von Fremdzeichen, Ausfüllen fehlender Daten oder Zusammenführen mehrerer Datenquellen.
3. Vorhersagemodell erstellen. Prädiktive Analysewerkzeuge beinhalten Dutzende verschiedener statistischer und maschineller Lernalgorithmen, die Data Scientists wählen können, um das beste prädiktive Modell auszuführen. Welche Algorithmen am besten geeignet sind, hängt von der Art und Vollständigkeit der Daten und der Art der gewünschten Vorhersage ab. Analysten führen die Analyse zunächst auf einer Teilmenge der Daten durch, die als Trainingsdaten bezeichnet werden, und legen Testdaten beiseite, die sie zur Bewertung des Modells verwenden wollen.
4. Modell bewerten. Bei der prädiktiven Analytik geht es nicht um das Absolute, sondern um Wahrscheinlichkeiten. Modelle werden anhand von Metriken bewertet, die stark vom zugrundeliegenden Use Case abhängen. Um die Vorhersagekraft zu bewerten, verwenden Datenanalytiker das Modell zur Vorhersage des Testdatensatzes. Wenn das prädiktive Modell den Testdatensatz vorhersagen kann, ist es ein Kandidat für den Einsatz.
5. Umsetzbare Rezepte liefern. Eine Vorhersage ist wenig wertvoll, wenn sie es nicht ermöglicht, eine prädiktive Gelegenheit zu ergreifen oder ein negatives Ereignis zu vermeiden. Stakeholder müssen lernen, sich auf die Vorhersagen von Modellen zu verlassen, und diejenigen, die die Modelle erstellen, müssen von ihren Partnern im Unternehmen lernen, was die umsetzungsfähigsten Erkenntnisse sind.
6. Effektivität des Modells verbessern. Vorhersagemodelle sind nur so genau wie die Daten, die in sie eingespeist werden. Im Laufe der Zeit können sie ihre Effektivität verschlechtern oder verbessern. Um Modelle auf ihre Effektivität und Werthaltigkeit zu überprüfen, müssen neu gesammelte Daten durch die bestehenden Algorithmen wiederholt werden. Wenn das Modell ungenauer wird, muss der Data Scientist das Modell anpassen (z.B. durch Veränderung von Parametern in den Algorithmen) und/oder zusätzliche Daten suchen und einspeisen.

Die gewonnenen Erkenntnisse erfolgreich in die Geschäftspraxis übertragen

Angesichts der Komplexität der Predictive Analytics-Projekte mit vielen Variablen und Einflussfaktoren raten wir schrittweise vorzugehen und mit einem kleinen Pilotprojekt zu starten, das in einem laufenden Prozess ständig weiterentwickelt wird. Es besteht immer das Risiko, dass die Vorhersagemodelle nicht das erwartete Ergebnis bringen. Das kann unterschiedlichste Gründe haben, z.B. kann es an mangelnder Datenqualität liegen. Daher ist es wichtig, das Projekt schrittweise zu erarbeiten und die Datenquellen sukzessive zu erweitern.

Gute Prognose, schlechte Prognose

Ob sich der ganze Aufwand für das entsprechende Vorhersagemodell gelohnt hat, zeigt ein Vergleich der Prognose mit der Realität. Eine weitere Kenngröße für die Qualität einer Prognose wäre der Vergleich mit bislang im Unternehmen eingesetzten Methoden (zum Beispiel Regressionsanalyse in Verbindung mit dem Bauchgefühl des Managers) oder Konkurrenz-Algorithmen. Die statistische Auswertung der Prognosen und die korrekte Maßzahl für die Prognosegüte ist ein wichtiger Bestandteil eines Projektes.

Erweist sich die Prognose als stichhaltig, gilt es, die Ergebnisse in die Geschäftspraxis zu übertragen, sodass sie einen Mehrwert ergeben.

Fazit

Es gibt nicht die eine richtige Predictive-Analytics-Methode für alle Unternehmen. Welche Methode und welcher Algorithmus am besten funktionieren, hängt vom individuellen Ziel und der Fragestellung ab.

Predictive Analytics ist kein einmaliger Vorgang; die Unternehmen müssen ihre Analyse auf neue Daten wiederholen, um sicherzustellen, dass die Modelle immer noch wirksam sind und um auf Veränderungen der Kundenwünsche und der Wettbewerber reagieren zu können. Viele Unternehmen werten die Daten wöchentlich oder sogar kontinuierlich aus.

Warum Sie mit it-novum sprechen sollten...

Wir setzen diese Business Intelligence- und Big Data-Vorteile gewinnbringend für Unternehmen um:

- 360-Grad Blick auf Ihre Kunden
- Fachabteilungen werten dank Self-Service Analytics Big Data-Daten selbst aus
- Identifikation neuer Umsatzquellen durch intelligente Nutzung von Unternehmensdaten
- Kosteneinsparung durch Einsatz eines Data Warehouse
- Vermeidung des aufwändigen und fehleranfälligen Excel-Chaos

Wenn Sie diese Vorteile auch in Ihrem Unternehmen nutzen wollen, sollten wir uns kennenlernen!

it-novum bietet umfangreiche Dienstleistungen für Ihr Big Data Analytics Projekt:

- Data Engineering
- Implementierung von Data Warehouses und Data Lakes
- Pentaho/SAP Connector für die Verarbeitung von SAP-Daten
- Pentaho/HVA Connector für die Analyse von Video Streams
- Predictive Analytics und Machine Learning
- Dashboards und Data Visualization
- Embedded Analytics

Führend in Business Open Source-Lösungen und -Beratung

it-novum ist das führende IT-Beratungsunternehmen im deutschsprachigen Raum für Geschäftslösungen auf Basis von Open Source. Von unserem Hauptsitz in Fulda und den Niederlassungen Wien und Thalwil aus betreuen wir Mittelstandskunden und Großunternehmen sowie den öffentlichen Sektor.

Mit unserer Expertise in Beratung, Training und Support helfen wir Unternehmen, aus ihren Daten Erkenntnisse zu gewinnen und ihre datengetriebenen Projekte erfolgreich zu gestalten.



Ihr Ansprechpartner

Christopher Keller
Director Big Data Analytics & IoT

✉ +49 661 103 724

☎ christopher.keller@it-novum.com



it-novum GmbH Deutschland

Hauptsitz Fulda: Edelzeller Straße 44 • 36043 Fulda • Telefon +49 661 103 434

it-novum Zweigniederlassung Österreich

Ausstellungsstraße 50 / Zugang C • 1020 Wien • Telefon +43 1 205 774 1041

it-novum Vertrieb Schweiz

Seestraße 97 • 8800 Thalwil • Telefon +41 44 722 75 55