

Predictive Analytics

Was ist Predictive Analytics?

Bei der prädiktiven Analytik geht darum eine Statistik z.B. über das Verhalten einer Person zu erstellen. Die einzelnen Informationen werden dann aus den Datenmengen extrahiert und verglichen, um Zusammenhänge in diesen Datenmengen zu entdecken und eine Vorhersage des Verhaltensmusters der ausgewählten Person zu treffen. Meist sind es wiederkehrende Entscheidungen, die epochal sind bei der Prognose der Handlungsweise eines Nutzers.

Prädiktive Analytik ist eine statistische und analytische Technik um ein Modell zu entwickeln, dass die zukünftigen Ereignisse und Handlungsweisen darstellt. Das Modell unterscheidet sich stark nach welchen Vorhersagen man sich richtet (vgl. URL: <https://www.the-digital-insurer.com/wp-content/uploads/2013/12/78-Predictive-Modeling-White-Paper.pdf>, Charles Nyce: *Predictive Analytics White Paper*, S.1, Stand 01.11.2017).

Das wichtigste bei dieser und wie bei vielen anderen analytischen Techniken ist es zuverlässige Daten mit hoher Qualität zu haben, um eine relative saubere Analyse durchzuführen. „Predictive Analytics“ ist in vielen Branchen gegenwärtig sehr ausgeprägt vorhanden. Von dem Vorhersehen von Flugpreisen, Kundenkaufverhaltens oder das Aufstellen von Prognosen hinsichtlich der Behandlungserfolge an Patienten gibt es ein breites Spektrum der Einsatzmöglichkeit dieses Verfahrens. Jedoch gibt es einige Bedenken bei der Ausführungsgewalt der „Predictive-Analytics-Software“. Falls die Software so entworfen ist, dass der Endbenutzer falsche Parameter eingeben kann, so wird es zur einer falschen Vorhersage kommen und einen gravierenden Schaden beim Unternehmen verursachen. Um dem entgegenzuwirken ist es unabdinglich die Software so zu gestalten, dass es keine falschen Parameter geben kann (vgl. URL: <https://news.sap.com/germany/was-ist-eigentlich-predictive-analytics/>, Stand 01.11.2017).

„Predictive Analytics Process“

Bei dem Ablauf der prädiktiven Analytik gibt es eine feste Reihenfolge, die jedes Mal abgearbeitet werden muss um eine vollwertige Analyse zu bekommen.

1. **Projekt festlegen**
 - ➔ Definieren der Projekt Ergebnisse.
 - ➔ Identifizieren der Datensätze mit dem man arbeiten wird.
2. **Datensammlung**
 - ➔ „Data Mining“ – Daten von mehreren Quellen werden für die Analyse vorbereitet. Dieser Schritt bietet einen großen Einblick auf die Benutzerinteraktion.
3. **Datenanalyse**
 - ➔ Die Analyse von Daten ist ein Prozess der Untersuchung, Säuberung, Veränderung und Modellierung von Daten für das Ziel der Entdeckung von nützlichen Informationen.
4. **Statistik**
 - ➔ Die statistische Analyse ermöglicht es, die Annahmen und Hypothesen zu validieren und anhand von statistischen Standardmodellen zu testen.
5. **Modellieren**
 - ➔ „Predictive Modeling“ bietet die Möglichkeit, automatisch präzise Vorhersagungsmodelle für die Zukunft zu erstellen.

6. Entwicklung

- ➔ „Predictive Model Deployment“ bietet die Möglichkeit, die Analyseergebnisse in den täglichen Entscheidungsfindungsprozess einzubringen, um Ergebnisse, Berichte und Ausgaben zu erhalten, indem die auf der Modellierung basierenden Entscheidungen automatisiert werden.

7. Modellüberwachung

- ➔ Die Modelle werden verwaltet und überwacht, um die Leistung zu überprüfen und sicherzustellen, dass sie das erwartete Ergebnis liefert.

Quelle: (vgl. URL: <https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/>, Stand 01.11.17)

API (Application Programming Interface)

Google selbst bietet die Google Cloud Prediction API an, die vor allem Funktionen der „Predictive Analytics“ ermöglicht. Diese RESTful-API ermöglicht es die Erstellung eines Modells für maschinelles Lernen und lässt damit zu Daten zu analysieren und verschiedene Funktionen zu Ihren Anwendungen hinzuzufügen. Leider ist die API veraltet und wird ab dem 30.04.2018 außer Dienst gestellt.

Welche Aufgaben hat die Prediction API?

Die Prediction API führt zwei Aufgaben durch:

- ➔ Bei einem neuen Element wird ein numerischer Wert vorhergesagt, der für dieses Element gilt. Dieser Wert basiert auf ähnlich bewerteten Beispielen in den Trainingsdaten.
- ➔ Bei einem neuen Element wird eine Kategorie gewählt, die es am besten beschreibt. Voraussetzung dafür ist ein Satz von ähnlich kategorisierten Elementen in den Trainingsdaten.

Hier findet man auch den ganzen Leitfaden zur Einbindung der API:

Quelle: (vgl. URL: https://cloud.google.com/prediction/docs/developer-guide#what_is_the_prediction_api, Stand 01.11.17)

Der Nachfolge von Google Cloud Prediction API ist „Cloud Machine Learning Engine“. Wie Cloud Prediction erfüllt es alle gewünschten Funktionen und verfügt über das leistungsstarke TensorFlow-Framework, das hinter vielen Produkten von Google steht. TensorFlow selbst, ist eine Open Source Software-Bibliothek für numerische Berechnungen mit Hilfe von Datenflussgraphen.

Besonderheiten von „Cloud Machine Learning Engine“:

- ➔ Einfache Modelle für maschinelles Lernen anfertigen, die mit allen Datentypen und -größen arbeiten können.
- ➔ TensorFlow-Framework (Deep Learning Funktionen)
- ➔ Skalierbare prognostische Analysen

Quelle: (vgl. URL: <https://cloud.google.com/ml-engine/>, Stand 01.11.17)