

REduce PArk-Search-Time

Data-Analytics zur Reduktion von Parksuchzeiten

Wie können Sie Ihr Data-Analytics-Projekt strukturieren und dafür sorgen, den maximalen Nutzen für sich und Ihr Unternehmen zu erreichen? Anhand eines praktischen Beispiels zeigen wir Ihnen die Möglichkeiten, aber auch die Stolpersteine, die in der Auswertung von Daten liegen.



Gemeinsam mit der *Urban Software Institute GmbH* haben wir vom *FIR an der RWTH Aachen* im Rahmen des Forschungsprojekts „REPASt“ einen *Proof of Concept* entwickelt, in dem kommerziell und öffentlich zugängliche Daten zusammengeführt werden, um Kurzzeitparkern ein Werkzeug zur Reduktion von Suchzeiten an die Hand zu geben. Im Folgenden stellen wir Ihnen die genaue Vorgehensweise vor und zeigen auf, welche Besonderheiten Unternehmen bei der Umsetzung eines Data-Analytics-Projekts beachten sollten.

Data-Analytics-Lebenszyklus

Bei der Durchführung von Projekten, mit besonderem Fokus auf der Auswertung und Nutzung von Daten liegt, bietet sich laut *Shearer* ein Vorgehen entsprechend dem Modell „*Cross Industry Standard Process for Data Mining*“ (CRISP-DM) an. Dieses stellt eine strukturierte und effiziente Auswertung der Projektergebnisse sicher. Das Modell unterteilt den Prozess des Data-Minings in sechs Phasen (s. Bild 1). Deren Reihenfolge ist nicht statisch und ein Wechsel zwischen verschiedenen Phasen ist sogar erforderlich. Der äußere Kreis des Diagramms verdeutlicht den zyklischen Charakter. Ein Data-Mining-Prozess wird nach dem Erreichen einer Lösung fortgesetzt. Dabei gewonnene Erkenntnisse führen oft zu neuen, spezifischeren Fragen, von deren Lösung man wiederum in nachfolgenden Data-Mining-Prozessen profitiert.¹

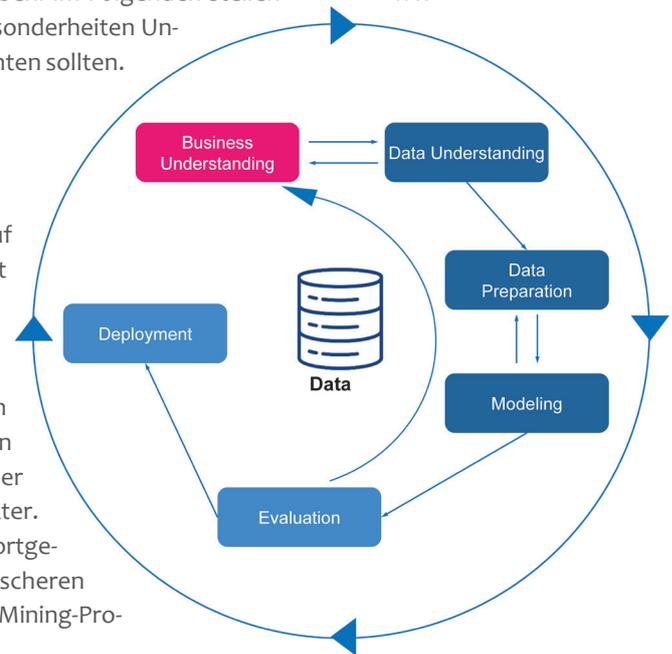


Bild 1:
CRISP-DM-Modell zur Durchführung von Data-Analytics-Projekten (eigene Darstellung)

Datengrundlage und Vorbereitung

Im Rahmen unseres Forschungsprojekts REPASt haben wir unterschiedliche Datenquellen ausgewertet. Diese beinhalten insbesondere Daten zu:

- Floating-Car-Data (kurz FCD)
- Parkscheinautomaten im Stadtgebiet Chemnitz²
- Meta-Informationen (Open Data) zu Parkplätzen im Stadtgebiet Chemnitz³

Da die Rohdaten nur eine Sammlung einzelner Datenpunkte darstellen, müssen wir diese in weiteren Verarbeitungsschritten aufbereiten. Einzelne Datenpunkte beinhalten Informationen zur Geoposition (Angaben zu Breiten- und Längengrad). Für die Fahrtenidentifizierung müssen wir diese Positionen jedoch konkreten Straßen oder Straßensegmenten zuordnen. Diese Zuordnung nehmen wir unter Zuhilfenahme des Barefoot-Map-Matchers vor, eines von *BMW Car IT* entwickelten Werkzeugs.

Für die Identifizierung des Parksuchverkehrs ist es wichtig, dass einzelne Fahrten extrahiert werden können. Hierfür bewerten wir die zeitlichen und örtlichen Dimensionen der Datenpunkte. So kann das Ende einer Fahrt vom Anfang der nächsten Fahrt unterschieden werden, wenn beispielsweise ein Intervall zwischen den Zeitstempeln der Datenpunkte vorliegt, das größer oder gleich vier Minuten ist.

¹ s. *Shearer* 2000, S. 14ff.

² Datenset Parkscheinautomaten der Stadtverwaltung Chemnitz, 2020: s. <http://portal-chemnitz.opendata.arcgis.com/datasets/parkscheinautomaten> (Link zuletzt geprüft: 10.08.2020)

³ Übersicht Parken der Stadtverwaltung Chemnitz, 2020: s. <https://www.chemnitz.de/chemnitz/de/unse-re-stadt/verkehr/auto-krad-ikw/parken/index.html> (Link zuletzt geprüft: 10.08.2020)

Darüber hinaus ist es wichtig, zwischen „Parken“, „Fahren in stockendem Verkehr“ und „Warten an einer Lichtsignalanlage“ zu unterscheiden. Hierfür verwenden wir die Informationen zur Fahrtrichtung. Ändern sich diese in einem 2-Minuten-Intervall um mehr als 100 Grad, können wir davon ausgehen, dass ein Parkvorgang stattgefunden hat. Diese Methodik funktioniert für viele Fälle in der Praxis sehr gut, kann jedoch noch weiter ausdifferenziert werden.

In einem weiteren Schritt verwenden wir Open-Street-Map (OSM) zur automatisierten Identifizierung von Parkflächen sowie zur Selektion wesentlicher Hauptstraßen, Autobahnen und Autobahnkreuze.

Architekturkonzept

Die Architektur des *Proofs of Concept* besteht aus vier Schichten. Die bereits beschriebenen unterschiedlichen Datenquellen bilden die Ausgangsbasis eines jeden Data-Analytics-Projekts. Die Anbindung der einzelnen Datenquellen in die Architektur ist nicht trivial und erfordert häufige Anpassungen. Solche Anpassungen betreffen insbesondere die technische Anbindungsmöglichkeit an eine Datenbank und die Definition passender Datenschemata.

In der Integrationsschicht (z. B. UrbanPulse) müssen die angebundenen Datenquellen schließlich vorgehalten und weiterverarbeitet werden (und beispielsweise von „Ausreißern“ bereinigt werden). Hierbei müssen wir für die entsprechenden Datenquellen passende Konnektoren entwickeln und einsetzen, die als Schnittstelle zwischen Datenquellen, dem Aufbau

von Datenbanken und Datenanalysen dienen.

In der Analytics-Schicht (z. B. *Jupyter Notebook*, *Python*, relevante Bibliotheken, Visualisierungswerkzeuge) analysieren wir die Daten und entwickeln für konkrete Problemstellungen die passenden Modelle, die wir dann testen und ausrollen.

Abschließend können wir aussagekräftige Visualisierungen bereitstellen, mit denen wir die Ergebnisse der Modellbildung und -nutzung veranschaulichen und eine Entscheidungsgrundlage schaffen. Hierfür bieten sich Open-Source-Tools wie z. B. *Kepler.gl* an.

Datenanalyse und Auswertung

Wir haben im Projekt exemplarisch Daten analysiert, die von der Stadt Chemnitz bereitgestellt wurden. Sie stammen aus dem Zeitraum vom 01.08.2018 bis zum 01.10.2019.

Die *Kepler.gl*-Demo selbst ist ein interaktiver *Proof of Concept*, in dem die Verteilung von Parksuchverkehr nachvollzogen werden kann. Es wird ein Kartenausschnitt von Chemnitz gezeigt. Hexagone darin visualisieren Bereiche, in denen FCD im Datenkorpus identifiziert werden und der Parksuchverkehr-Algorithmus angewendet wird. Dabei werden drei Klassen unterschieden:

- Gelb: In diesen Bereichen wurde kein signifikanter Parksuchverkehr identifiziert und klassifiziert.
- Rot: In diesen Bereichen wurde Parksuchverkehr in „mittlerem“ Ausmaß identifiziert und klassifiziert.
- Lila: In diesen Bereichen wurde Parksuchverkehr in „starkem“ Ausmaß identifiziert und klassifiziert.

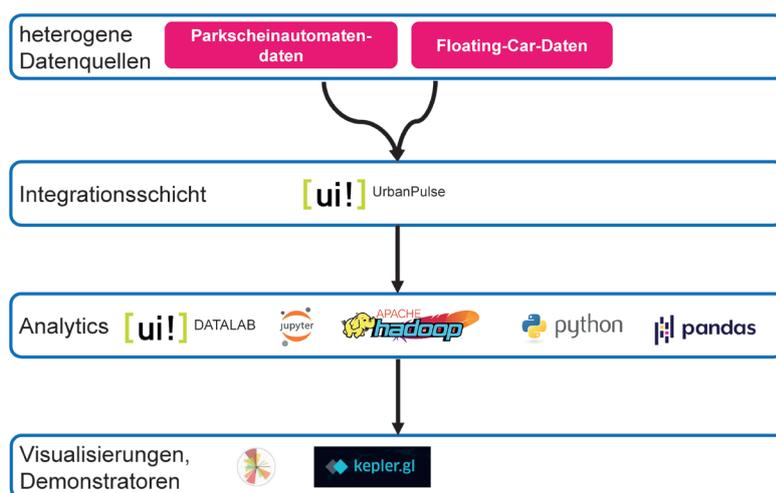


Bild 2:
Data-Analytics-Architektur
(eigene Darstellung i. A. a.
SHEARER 2000, S. 14)

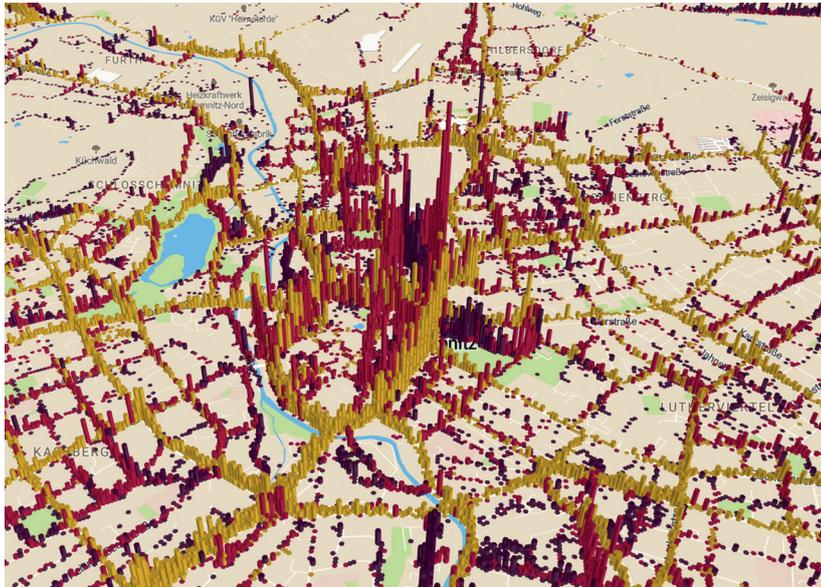


Bild 3:
Visualisierung des Parksuchverkehrs
in Chemnitz (eigene Darstellung)

Die Analyse zu Parksuchverkehr (s. Bild 3, S. 20) zeigt, dass insbesondere im Innenstadtbereich und hier vor allem rund um die Technische Hochschule Chemnitz wie auch in der näheren Umgebung des Hauptbahnhofs viel Parksuchverkehr herrscht („dunkelrote“ Hexagone). Hauptverkehrsstraßen hingegen zeigen keinen signifikanten Anteil am Parksuchverkehr. Dagegen erkennen wir außerhalb größerer Verkehrsstraßen, beispielsweise in Wohngebieten oder Einkaufszentren, ebenfalls starken Parksuchverkehr.

Bei der Strukturierung eines Data-Analytics-Projekts anhand des CRISP-DM-Modells geht man immer von einer konkreten Problemstellung aus, wie der des Parksuchverkehrs. Darauf aufbauend müssen die relevanten Datenquellen identifiziert und anschließend aufbereitet und verknüpft werden. Das vorgestellte Architekturkonzept kann hierfür als Grundlage dienen und an die unternehmensspezifischen Bedürfnisse angepasst werden, um den maximalen Erfolg für die eigenen Data-Analytics-Projekte zu erzielen.

Literatur:
SHEARER, C.: The CRISP-DM model: the new blueprint for data mining. In: Journal of data warehousing 5 (2000) 4, S. 13 – 22.

wi · go · heuser · bachleitner

Haben auch Sie Interesse daran, Ihr eigenes Data-Analytics-Projekt durchzuführen und bereits eine konkrete Problemstellung, die Sie lösen möchten? Zögern Sie nicht, uns anzusprechen. Unsere Experten beraten Sie gerne!

Mehr Informationen finden Sie unter: repast.fir.de



Rafael Götzten, M.Sc.
Projektmanager
Business-Transformation
FIR an der RWTH Aachen
Tel.: +49 241 47705-315
E-Mail: Rafael.Goetzen@fir.rwth-aachen.de



Tobias Heuser, M.Sc.
Urban Software
Institute GmbH
Data Scientist
Tel.: +49 (0) 176 34576073
E-Mail: tobias.heuser@ui.city



Markus Bachleitner
Urban Software
Institute GmbH
Director Smart Mobility Data
Tel.: +49 (0) 176 84958145
E-Mail: markus.bachleitner@ui.city