

BI Spektrum

EINE PUBLIKATION DES TDWI E.V.

Business Analytics: Passender Methodenmix bringt beste Resultate ab Seite 8

Best Practice:
**Analytics steuert
Fertigungsqualität**
Seite 36

Projektmanagement:
**Analytics-Prototypen
vom Lab in die
Produktion**
Seite 40

Interview:
**Kognitive Ansätze
spielen Hauptrolle
für Big Data**
Seite 32

Per Stricker,
ITGAIN



Sonderdruck zum Thema
Analytics/Digitale Transformation

bridging IT
Menschen Methoden Lösungen



Foto: Shutterstock

Predictive Analytics für intelligente Waschräume

Saubere Vorhersagen

Ein Beitrag von
Philip Rausch

Das Projekt „Predictive Cleaning“ zeigt an einem praktischen Einsatz die Relevanz der digitalen Transformation im Kontext Analytics auf und verbindet Data-Science-Methoden mit den von intelligenten Waschräumen generierten Sensordaten, um optimale Reinigungspläne je Waschraum aufzustellen.

Im Rahmen der digitalen Transformation gewinnt das Thema Analytics immer mehr an Relevanz und vielfach gelten Daten als wichtigster Rohstoff des 21. Jahrhunderts. Mit dem Einzug intelligenter Geräte und Sensoren in die Industrie, die Dienstleistungsbranche und den gesellschaftlichen Alltag werden permanent Daten erzeugt. Smarte Produkte können im Laufe ihres Lebenszyklus Informationen kommunizieren und generieren nicht nur Informationen zu Prozessschritten, sondern bestimmen auch Zukunftsschritte durch statistische Algorithmen (vgl. [Sch 15]).

Die intelligente Analyse, Aufbereitung, Nutzung und Visualisierung dieser verfügbaren Informationen wird bzw. ist für Unternehmen bereits heute ein entscheidender Wettbewerbsfaktor. So können Softwaresysteme und industrielle Produkte miteinander verknüpft werden, um intelligente Algorithmen für autonome Steuerungssysteme zu entwerfen (vgl. [LKY 14]).

Das Projekt „Predictive Cleaning“ beruht auf den von intelligenten Waschräumen erzeugten Sensordaten. Durch Data-Science-Methoden sollen die Häufigkeitsverteilungen (siehe Abbildung 1) auf Basis von Besuchermessdaten einen Lösungsansatz zur Erstellung optimaler Reinigungspläne je Waschraum liefern.

Im Zuge der digitalen Transformation werden analoge Prozesse nach und nach digitalisiert. Durch die Integration von industriellen Produkten wird es möglich, Predictive-Technologien auf Sensordaten und andere systembezogene Informationen anzuwenden (vgl. [LKY 14]). Dies geschieht nicht nur in der Industrie, sondern auch in der Dienstleistungsbranche, um bisher vollkommen analog abgewickelte Abläufe effizienter zu gestalten sowie verborgene Kostensenkungspotenziale zu heben.

Grundlage war die Idee, Waschräume bzw. die dort installierten Toiletten, Seifen- und Papierspender zu vernetzen und mit Sensoren auszustatten, um sekundengenau Daten über die Anzahl von Toilettenspülvorgängen, Füllständen von Seifen- und Papierspender sowie die Anzahl an Besuchern je Waschraum zu erfassen. Auf Basis dieser Daten soll neben der Qualität auch die Sauberkeit der Waschräume gesteigert, das heißt das Ambiente für Besucher angenehmer gestaltet werden. Für das zuständige Gebäudemanagement ist es somit möglich, Ressourcen nach Bedarf einzuteilen, also mit Hilfe der intelligenten Vernetzung aktuelle Nachfüllbedarfe zu ermitteln und Reinigungsarbeiten je nach Notwendigkeit durchzuführen.

Über parametrisierte und vordefinierte Berichte erhalten die Reinigungskräfte direkt Analysen und Informationen darüber, welche Waschräume Wartungsbedarf haben. So reagieren Gebäudemanager nicht mehr nur auf Beschwerden, sondern können ein proaktives und flexibles Reinigungskonzept verfolgen.

Um dies zu gewährleisten, werden die Sensordaten vom jeweiligen Gerät (Besucherzähler, Toilettenspülung, Spendereinheit) erfasst, über eine einheitliche Schnittstelle an ein zentrales Data Warehouse gesendet und dort als Grundlage zur Ausführung standardisierter Berichte gespeichert. Die auch auf mobilen Endgeräten (zum Beispiel Tablet) ausführbaren Berichte zeigen jedoch nur aggregiert die Auslastungssituation bzw. historische Daten je Waschraum und geben den Reinigungskräften keinen generierten Reinigungsplan vor. So ist es daher nicht möglich festzustellen, welches Kostensenkungspotenzial durch eine Änderung der Reinigungszeiten noch besteht. Diese Aufgabenstellung löst das im Folgenden dargestellte System.

Datenmanagement

Zur Erweiterung der bestehenden Lösung, insbesondere durch Einführung von (Predictive) Analytics, wurde ein Workshop durchgeführt, in dem die verfügbaren Rohdaten der Besucherzähler (Visitor Registration Unit, kurz VRU) bzw. der Toiletensensoren (Flush Registration Units, kurz FRU) analysiert wurden. Die Füllstände der Seifen- und Papierspender werden in dem hier beschriebenen Lösungsansatz vernachlässigt. Da die Daten der FRU-Einheiten sich nur in Bezug auf das Datenformat von den Daten der VRU-Einheiten unterscheiden, wird auf diese nicht separat eingegangen.

Tabelle 1 zeigt beispielhaft Zeilen aus einem Rohdatensatz einer VRU-Einheit, wobei zu beach-

PHILIP RAUSCH führt ein Team im Thema Microsoft Business Intelligence, leitet intern das Center of Excellence Analytics und verantwortet die zugehörige Sales Area Analytics der BridgingIT GmbH. Das Unternehmen beschäftigt über 500 Mitarbeiter und hat 2016 einen Umsatz von 69,9 Mio. € erwirtschaftet.

E-Mail: philip.rausch@bridging-it.de



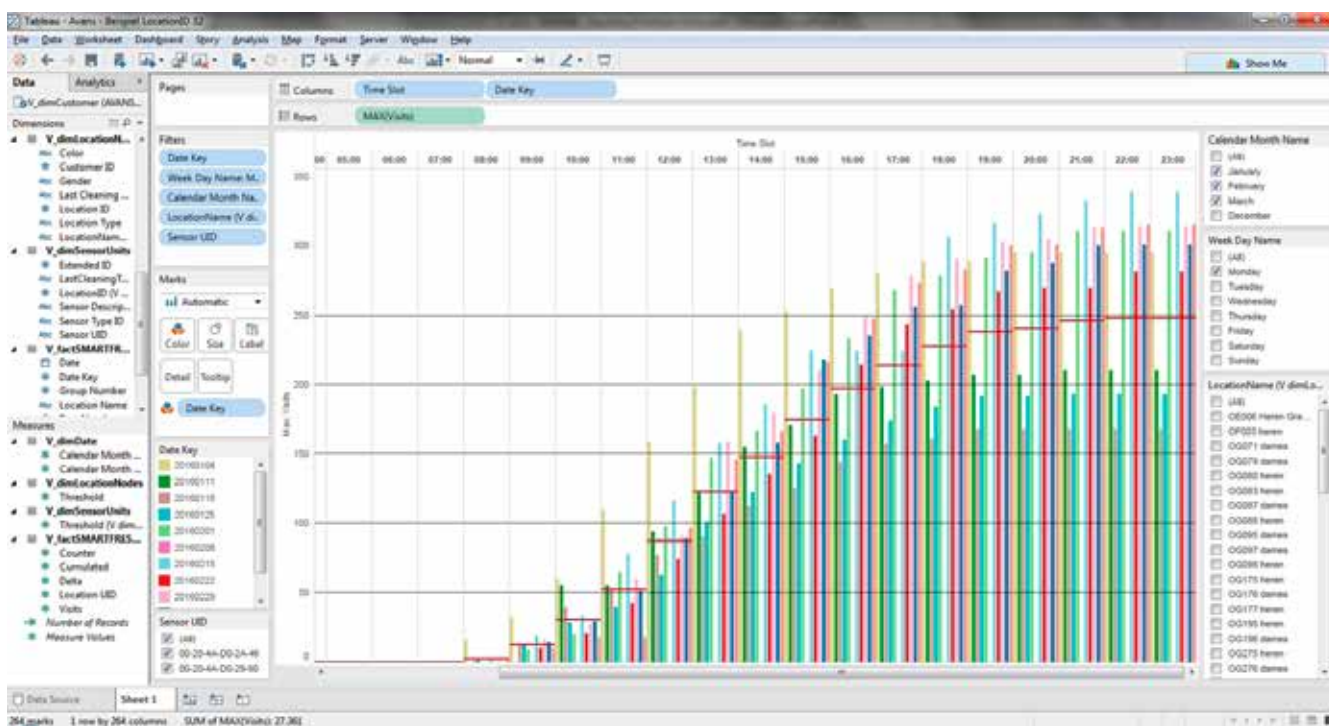
ten ist, dass der Besucherzähler eines Waschrums sowohl beim Eintritt als auch beim Verlassen einer Person den Wert in der Spalte „CounterIn“ um 1 erhöht. Der Zählerstand beginnt bei jeder VRU-Einheit mit dem Wert 0 und wird inkrementell, auch über Tagesgrenzen hinweg, stetig ganzzahlig fortgeschrieben.

Neben den in Tabelle 1 dargestellten Bewegungsdaten stehen auch Stammdaten zur Verfügung, wie beispielsweise die Zugehörigkeit von Sensoren zu Waschräumen. Diese Stamm- und Bewegungsdaten werden genutzt, um eine Datenbank mit Dimensions- und Faktentabellen im Star-Schema zu generieren, die als Grundlage für die im nächsten Abschnitt durchgeführte Datenanalyse dient. Zur Erstellung einer Zeitdimension wird ein selbstentwickeltes T-SQL-Skript verwendet. Die Rohdaten werden zunächst ohne Veränderung in diese SQL-Server-Datenbank (Sensormessdatenbank) geladen.

Datenanalyse

Nachdem die Rohdaten in die Sensormessdatenbank importiert sind, wird mit der Analyse der Daten und der Entwicklung des Lösungsansatzes begonnen: Wie in Abbildung 1 dargestellt, wird hierzu ein Ansatz basierend auf Häufigkeitsver-

Abb. 1: Besuchermessdaten: Häufigkeitsverteilung, gruppiert nach Stundenintervallen (Screenshot)



ManufactureYear	VRU_ID	TimeStamp	CounterIn	BatteryLevel	SignalQuality	SignalStrength	SoftwareVersion	DCU_ID	DCU_Version	Status
20	719	2015-09-28 21:52:27.000	30	3290	255	79	45	268	8	0
20	719	2015-09-28 22:04:08.000	35	3290	255	83	45	268	8	0
20	719	2015-09-28 22:54:38.000	37	3290	255	79	45	268	8	0
20	719	2015-09-28 23:45:09.000	38	3280	255	82	45	268	8	0
20	719	2015-09-28 23:49:02.000	39	3280	255	78	45	268	8	0
20	719	2015-09-29 00:00:42.000	40	3270	255	83	45	268	8	0
20	719	2015-09-29 00:04:36.000	41	3270	255	81	45	268	8	0
20	719	2015-09-29 01:06:46.000	42	3260	255	81	45	268	8	0
20	719	2015-09-29 01:14:33.000	56	3260	255	78	45	268	8	0
20	719	2015-09-29 01:18:27.000	62	3330	255	80	45	268	8	0

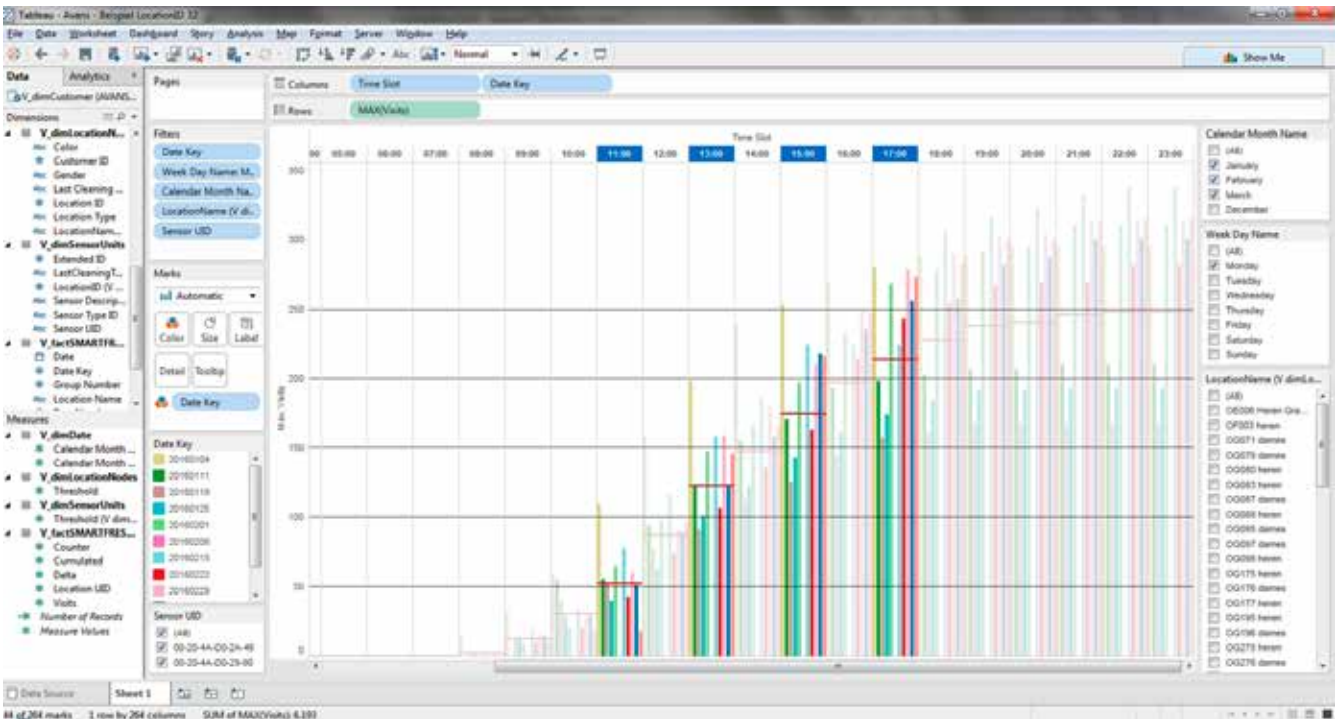
Tab. 1: Rohdatensatz der VRU-Einheit 20-719

lungen verfolgt, in dem pro Wochentag und Sensoreinheit die Besuchermessdaten kumuliert und als Verteilungsfunktion, gruppiert nach Stundenintervallen, unterteilt werden. Zur Datenanalyse und reinen Visualisierung des Lösungsansatzes wird Tableau verwendet. Die in dieses Tool integrierten Analytics-Funktionen werden hierbei nicht berücksichtigt bzw. angewandt.

Zur Bestimmung optimaler Reinigungsintervalle ist es notwendig, für jeden Waschraum sogenannte Schwellenwerte bzw. die Anzahl an Besuchern festzulegen, nach denen ein Raum gereinigt werden sollte. In Abbildung 1 ist dieser Schwellenwert durch horizontale graue Linien dargestellt. Er beträgt den Wert 50 bzw. ganzzahlige Vielfache dieses Wertes. Es wird definiert, dass bei Überschreitung der Durchschnittswerte (rote horizontale Linien) in einem Zeitintervall eine Reinigung zu Beginn bzw. während dieses Zeitintervalls erfolgen muss. Diesen Sachverhalt zeigt Abbildung 2, wobei die Reinigungszeiten blau hervorgehoben sind.

Um diese Analyse nicht wiederholt manuell für jeden Wochentag und jeden Waschraum separat mittels des in Abbildung 2 dargestellten Tableau-Dashboards durchzuführen, wird ein Algorithmus entwickelt, der basierend auf dem dargestellten Lösungsansatz auf Basis historischer und aktueller Messdaten automatisiert Reinigungspläne erstellt.

Abb. 2: Optimale Reinigungsintervalle für Montag gemäß vorgegebenem Schwellenwert (Screenshot)



Datenaufbereitung

Für die Anwendung des in Abbildung 2 visualisierten Algorithmus sind zahlreiche vorgelagerte Schritte notwendig: zum Beispiel Datenanreicherungen, Datentransformationen, Einfügen neuer und berechneter Felder. Der Besucherzähler muss zunächst partitioniert nach Tagen und Sensoreinheiten ermittelt werden. Danach muss er innerhalb der Partitionen kumuliert werden, damit neben der in den Rohdaten bereits vorhandenen fortlaufenden Besucherzählung über Tagesgrenzen hinweg auch ein aufsteigender Wert innerhalb eines Tages sowie das Delta zwischen zwei Messdatensätzen für Werte eines Tages zur Verfügung steht.

Da sowohl der Eintritt in den Raum als auch das Verlassen eines Besuchers von einer VRU-Einheit erfasst werden, muss der kumulierte Wert „Cumulated“ durch zwei geteilt werden, um einen „Visit“ zu erhalten. Die für die Ermittlung der optimalen Reinigungsintervalle aufbereiteten Sensordaten zeigt Tabelle 2. Diese Datensätze basieren auf den Werten von Tabelle 1, wobei von dieser die ersten vier Spalten (von links) zur Weiterverarbeitung verwendet werden. Die beiden Spalten „ManufacturerYear“ und „VRU_ID“ sind zu einer Spalte „Sensor UID“ zusammengefasst.

RowNumber	SensorUID	SensorDescription	DateKey	TimeStamp	Date	TimeSlot	Counter	Delta	Cumulated	Visits
1	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 18:49:15.000	2015-09-28	18:00	1	1	1	0
2	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 18:49:46.000	2015-09-28	18:00	1	0	1	0
3	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 18:53:39.000	2015-09-28	18:00	1	0	1	0
4	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 18:57:32.000	2015-09-28	18:00	2	1	2	1
5	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 19:01:25.000	2015-09-28	19:00	3	1	3	1
6	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 19:13:05.000	2015-09-28	19:00	5	2	5	2
7	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 19:16:58.000	2015-09-28	19:00	6	1	6	3
8	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 19:20:52.000	2015-09-28	19:00	7	1	7	3
9	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 19:48:04.000	2015-09-28	19:00	9	2	9	4
10	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 20:26:56.000	2015-09-28	20:00	11	2	11	5
11	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 20:30:49.000	2015-09-28	20:00	13	2	13	6
12	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 20:34:42.000	2015-09-28	20:00	14	1	14	7
13	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 20:38:35.000	2015-09-28	20:00	15	1	15	7
14	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 20:54:09.000	2015-09-28	20:00	16	1	16	8
15	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 20:58:02.000	2015-09-28	20:00	18	2	18	9
16	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 21:01:55.000	2015-09-28	21:00	19	1	19	9
17	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 21:09:42.000	2015-09-28	21:00	26	7	26	13
18	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 21:40:48.000	2015-09-28	21:00	28	2	28	14
19	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 21:52:27.000	2015-09-28	21:00	30	2	30	15
20	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 22:04:08.000	2015-09-28	22:00	35	5	35	17
21	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 22:54:38.000	2015-09-28	22:00	37	2	37	18
22	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 23:45:09.000	2015-09-28	23:00	38	1	38	19
23	20-719	1 ZOG Toilette Herren 02	20150928	2015-09-28 23:49:02.000	2015-09-28	23:00	39	1	39	19

Tab. 2: Anreicherung, Partitionierung und Kumulierung der Sensordaten

Datensäuberung

Neben der Analyse der Sensordaten sowie der Entwicklung des mehrstufigen Datentransformations- und Aggregationsprozesses ist speziell das Eliminieren redundanter Daten sowie die Korrektur falscher Werte durch den dieser Lösung zugrunde liegenden ETL-Prozess aufwendig. Im Rahmen eines eigenen Datensäuberungsmoduls, das als R-Skript realisiert wurde, werden die folgenden Fehlerfälle für VRU- und FRU-Daten abgefangen:

1. Ausreißer nach oben und nach unten, das heißt fallende und steigende Werte in der Zeitreihe
2. Sensordaten enthalten Lücken, das heißt, es gibt Tage bzw. Stundenintervalle ohne Werte
3. Mehrere aufeinanderfolgende Messdatensätze enthalten Ausreißer

Ein Beispiel für einen fehlerhaften Datensatz zeigt Tabelle 3. Dieser stammt von einem FRU-Datensatz und enthält die Fehler 1. und 2. Art.

Statistische Optimierung

Der hier visuell dargestellte Lösungsansatz ist innerhalb einer SQL-Server-Datenbank mittels gespeicherter Prozeduren, Tabellen, Views sowie T-SQL-Skripten realisiert. Er basiert auf Häufigkeitsverteilungen von Sensormessdaten, unterteilt in Stundenintervalle. Der Algorithmus ist modular bzw. schrittweise aufgebaut.

Nach dem Import der Rohdatensätze (vgl. Tabelle 1) werden diese in Dimensions- und Faktentabellen unterteilt. Die Spalte „TimeStamp“ wird in zwei Spalten „Date“ und „TimeSlot“ aufgeteilt und mit der Dimension „dimDate“ verknüpft. Die Spalte „SensorUID“ bildet die Verknüpfung zur Dimension Waschraum (vgl. Tabelle 2).

Im nächsten Schritt wird der fortlaufende Wert „CounterIn“ dazu verwendet, um innerhalb von Tagespartitionen je Sensoreinheit Verteilungsfunktionen auf Basis von Stundenintervallen zu erzeugen. Die Spalte „Delta“ enthält hierbei die Änderung zwischen zwei aufeinanderfolgenden Messwerten. Ausreißer oder Datenfehler (zum Beispiel negatives

Delta) werden mit Hilfe dieses berechneten Wertes korrigiert. Fehlende Daten bzw. Lücken, das heißt Stundenintervalle ohne Werte, werden mittels Interpolation geschlossen. Dubletten werden durch Gruppierung zu einem Messwert zusammengefasst.

Nach erfolgter Datensäuberung wird pro Wochentag und Waschraum bzw. Sensoreinheit für jedes Stundenintervall die Anzahl an kumulierten durchschnittlichen Besuchern errechnet (siehe Abbildung 2 – rote horizontale Linien). Auf Basis dieser Durchschnittswerte und der zuvor festgelegten individuellen Besuchergrenzwerte pro Raumeinheit ermittelt der Algorithmus durch Division beider Werte neben der Anzahl benötigter Reinigungsvorgänge auch das jeweilige Stundenintervall für den entsprechenden Reinigungsvorgang.

Der auf diese Weise generierte Soll-Reinigungsplan, aufgeteilt nach Stundenintervallen pro Wochentag, zeigt dem Gebäudemanagement die Einsparpotenziale pro Waschraum im Vergleich zu einem bestehenden Reinigungsplan je Waschraum auf. Der automatisierte Import von Sensordaten ist durch den Aufbau eines Massendatenverarbeitungsprozesses mittels SQL Server Integration Services realisiert und importiert die Rohdaten in die oben genannte Sensormessdatenbank.

Die in diese Sensormessdatenbank geladenen Datensätze werden mittels Tableau Dashboard visualisiert. In Tableau integrierte Analytics-Funktionen werden in diesem Ansatz nicht angewandt.

Der Datenimport kann je nach Bedarf mehrmals am Tag und vollautomatisiert durchgeführt werden. Hierbei lädt der mehrstufige Prozess zur historischen

BI-SPEKTRUM ist eine Fachpublikation des Verlags:
SIGS DATACOM GmbH | Lindlaustraße 2c | 53842 Troisdorf
Tel.: +49 (0) 22 41.2341-100 | Fax: +49 (0) 22 41.2341-199
E-mail: info@sigs-datacom.de
www.javaspektrum.de | www.objektspektrum.de
www.bi-spektrum.de

SIGS DATACOM
FACHINFORMATIONEN FÜR IT-PROFESSIONALS

Date	BloeSlaveID	Grp	GroupName	Hour	AbsValue	DifValue
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	15	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	15	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	15	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	15	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	15	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	15	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	17	4229	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	17	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	17	4493	0
2016/02/15	00-20-4A-B4-CA-56	69	OX272 dames	17	4493	0

Tab. 3: Ausreißer nach unten innerhalb der Zeitreihe sowie fehlender Wert bei „Hour“ 16

Datengrundlage aktuelle Rohdaten und ermittelt erneut unter Berücksichtigung der Schwellenwerte pro Sensoreinheit die optimalen Reinigungsintervalle je Waschraum. Die Pläne sind separiert nach Wochentag und Waschraum. Pro Waschraum kann darüber hinaus ein Wochenreinigungsplan mit gleichen Reinigungszeiten für jeden Wochentag generiert werden.

Der Algorithmus berücksichtigt dabei, dass für jeden Waschraum bzw. jede einzelne Sensoreinheit ein eigener Schwellenwert festgelegt werden kann. Falls kein Schwellenwert definiert wird, ist der Algorithmus ebenfalls in der Lage, einen optimalen Schwellenwert zu ermitteln.

Fazit und Ausblick

Der auf das Datenanalyseframework, bestehend aus ETL-Prozess, Datensäuberungsmodul und Sen-

sormessdatenbank, aufgesetzte Optimierungsalgorithmus generiert, basierend auf Häufigkeitsverteilungen, durch individuelle Vorgabe von Besuchergrenzwerten pro Raumeinheit einen täglichen Soll-Reinigungsplan in Stundenintervallen und zeigt die Einsparpotenziale pro Waschraum im Vergleich zum bestehenden Reinigungsplan auf.

Neben der Analyse der erhaltenen Sensordaten sind besonders die Aufbereitung und das Eliminieren von Ausreißern aufwendig. Die Schwellenwerte je Sensoreinheit werden in Abstimmung mit dem Gebäudemangement festgelegt und die damit generierten Soll-Reinigungspläne mit dem Ist-Plan verglichen. Bereits nach kurzer Zeit konnten Kosteneinsparungen erzielt und weitere Kunden für die Installation intelligenter Waschräume gewonnen werden.

Der Optimierungsalgorithmus soll im nächsten Schritt um geografische Informationen der Waschräume sowie um weitere Sensordaten ergänzt werden. So können zusätzlich auf Basis von Distanzen und Füllständen der in den Waschräumen installierten Seifenspender nicht nur die Reinigungspläne verfeinert, sondern auch Reinigungsrouten innerhalb einzelner Gebäude bzw. zwischen mehreren Gebäuden optimiert werden. Es kann auch ein Entscheidungsbaum-Algorithmus, wie in [Seg07] dargestellt, zur Prognostizierung regionsbasierter Bedarfe implementiert werden.

Literatur

- [Sch15] Schmidt, R. et al.: Industry 4.0 - Potentials for Creating Smart Products: Empirical Research Results. In: Business Information Systems, 18th International Conference, BIS 2015, Proceedings, Poznań/Polen, 24.-26. Juni 2015
- [Seg07] Segaran, T.: Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly Media, Beijing 2007, S. 142ff.
- [LKY14] Lee, J. / Kao, H. / Yang, S.: Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. Procedia CIRP, Vol. 16, 2014, S. 3-8