Big Data im Maschinen- und Anlagenbau
- Einsatzmöglichkeiten am Beispiel des Projekts smartTCS -

Andreas Varwig

Lehrstuhl für Informationsmanagement und Wirtschaftsinformatik (IMWI)
Universität Osnabrück

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Agenda

- Einführung
- Maschinendiagnostik
- Advanced Analytics und Predictive Maintenance
- Unterstützte Prozessführung in Wartungsprozessen

Agenda

- Einführung
- Maschinendiagnostik
- Advanced Analytics und Predictive Maintenance
- Unterstützte Prozessführung in Wartungsprozessen

Einführung: Andreas Varwig

Andreas Varwig

E-Mail: andreas.varwig@uos.de
Tel.: +49 (0) 541 969 4813

Universität Osnabrück
Fachgebiet Informationsmanagement und Wirtschaftsinformatik
Katharinenstraße 3
49074 Osnabrück

Lebenslauf

seit 2016:
Wissenschaftlicher Mitarbeiter am Fachgebiet für Informations-management und Wirtschaftsinformatik, Universität Osnabrück

2013-2016:
Senior Consultant, Mieschke Hofmann und Partner, Gesellschaft für Management- und IT-Beratung mbH, Ludwigsburg, Experte für Predictive Analytics und Big Data

2012-2013:
Consultant, ifb AG, Köln, Experte für Predictive Analytics

2010-2012:
Dozent, FOM Hochschule für Ökonomie & Management, Dozent für „Finanzmathematik“ und „Investition und Finanzierung“

2009-2012:
Wissenschaftlicher Mitarbeiter am Lehrstuhl für Finanzwirtschaft, Universität Bremen

Expertise & Projekthistorie (Auszug)

Seit 2016: smartTCS (öffentlich, durch das BMBF gefördertes Forschungsprojekt)
Entwicklung von Use-Cases, Services und technischen Modulen zur Unterstützung technischer Kundendienstleistungen durch eine Kooperationsplattform.
Rolle: Experte Data Mining & Predictive, Entwickler

2014-2016: Internationaler Süßwarenhersteller
Entwicklung von Analysen zur Mustererkennung und Käuferverhaltensvorhersage (SAP Predictive Analysis, R, SAP HANA) auf Basis von Transaktionsdaten.
Rolle: (Teil-) Projektleitung, Experte Data Mining & Predictive, Entwickler, Coach

2015-2016: Landmaschinenhersteller
Clusteranalyse von Sensor- und Geo-Daten zur automatisierten Bestimmung von Tätigkeits- und Bewegungsprofilen und Optimierung von Fahrzeugkonfigurationen (Hadoop und R)
Rolle: Projektleitung, Experte Data Mining & Predictive, Entwickler

2014-2016: Internationaler Automobilzulieferer
Implementierung von Reportings (SAP BO auf SAP HANA) und Evaluierung von Software-lösungen zum Data & Text Mining zum Predictive Quality Management
Rolle: Experte Data Mining & Predictive, Entwickler, Coach

Veröffentlichungen (Auszug)

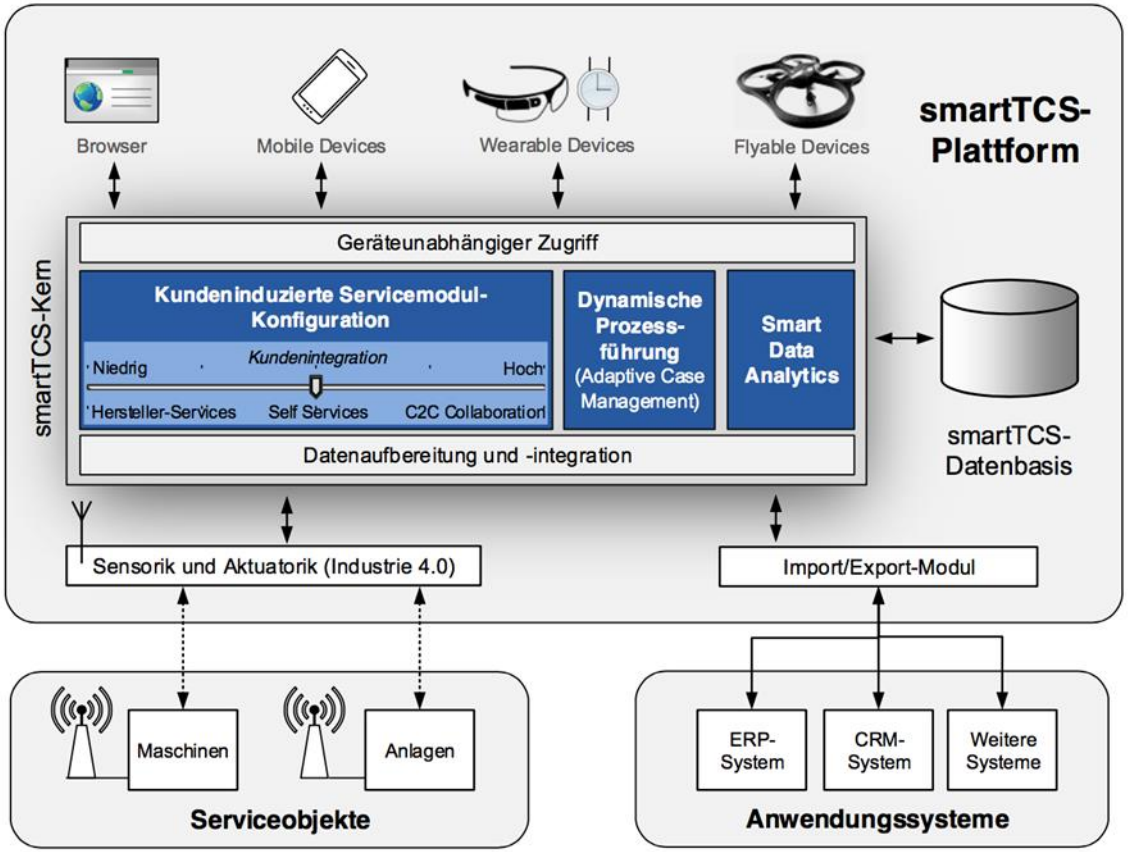
Varwig, A., Kammler, F., Thomas, O. (2017), Responding to the Forecast: Towards the Integration of Machine State Prediction and Required Maintenance Services, Informatik 2017, Lecture Notes in Informatics (LNI) (Eibl M., Gaedke M., (Hrsg.)), im Druck.

Thomas, O., Varwig, A., Kammler, F., Zobel, B., Fuchs, A. (2017), DevOps: IT-Entwicklung im Industrie 4.0-Zeitalter - Flexibles Reagieren in einem dynamischen Umfeld, in HMD Praxis der Wirtschaftsinformatik, Nr. 54 (2), S. 178-188.

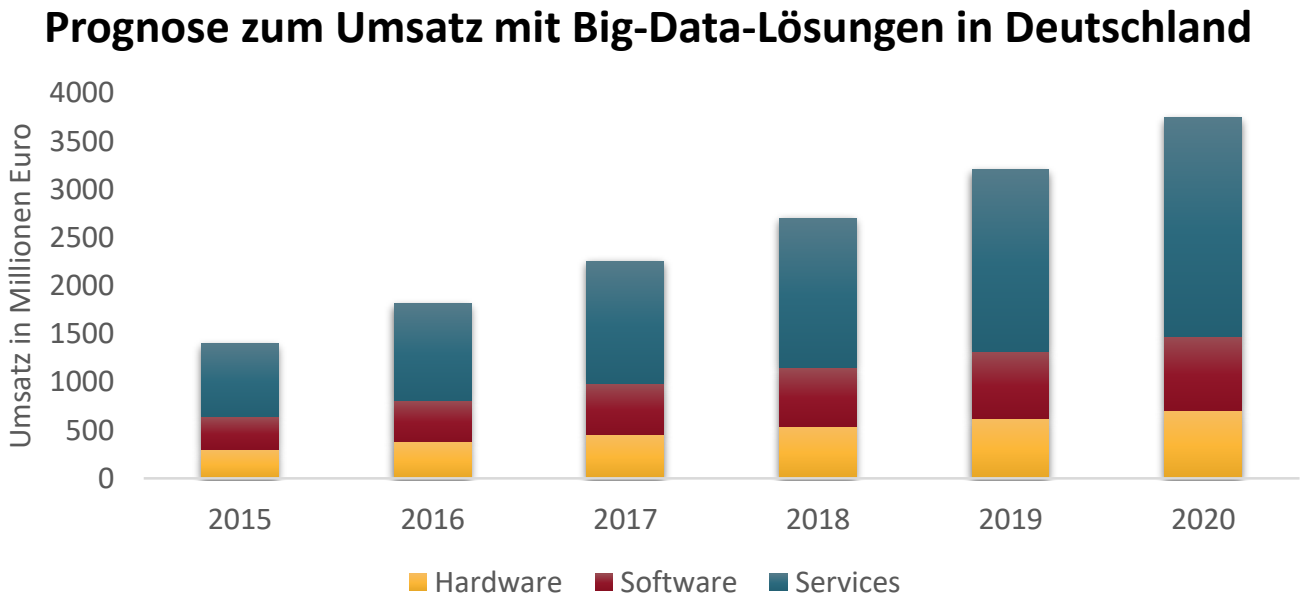
Poddig, T., Varmaz A., Varwig, A. (2013), Centralized resource planning and Yardstick competition, in Omega, Nr. 44 (1), S. 112-118.

Fieberg, C., Varmaz A., Varwig, A. (2013), RMatlab-app2web: Web Deployment of R/MATLAB Applications, in Journal of Statistical Software, Nr. 54 (5), S. 1-11.

Einführung: smartTCS-Projektziel



Im Rahmen von smartTCS soll eine Plattformlösung konzipiert und implementiert werden, welche die Kooperation verschiedener Unternehmen bei der Durchführung verschiedenster technischer Kundendienstleistungen ermöglicht und verbessert.

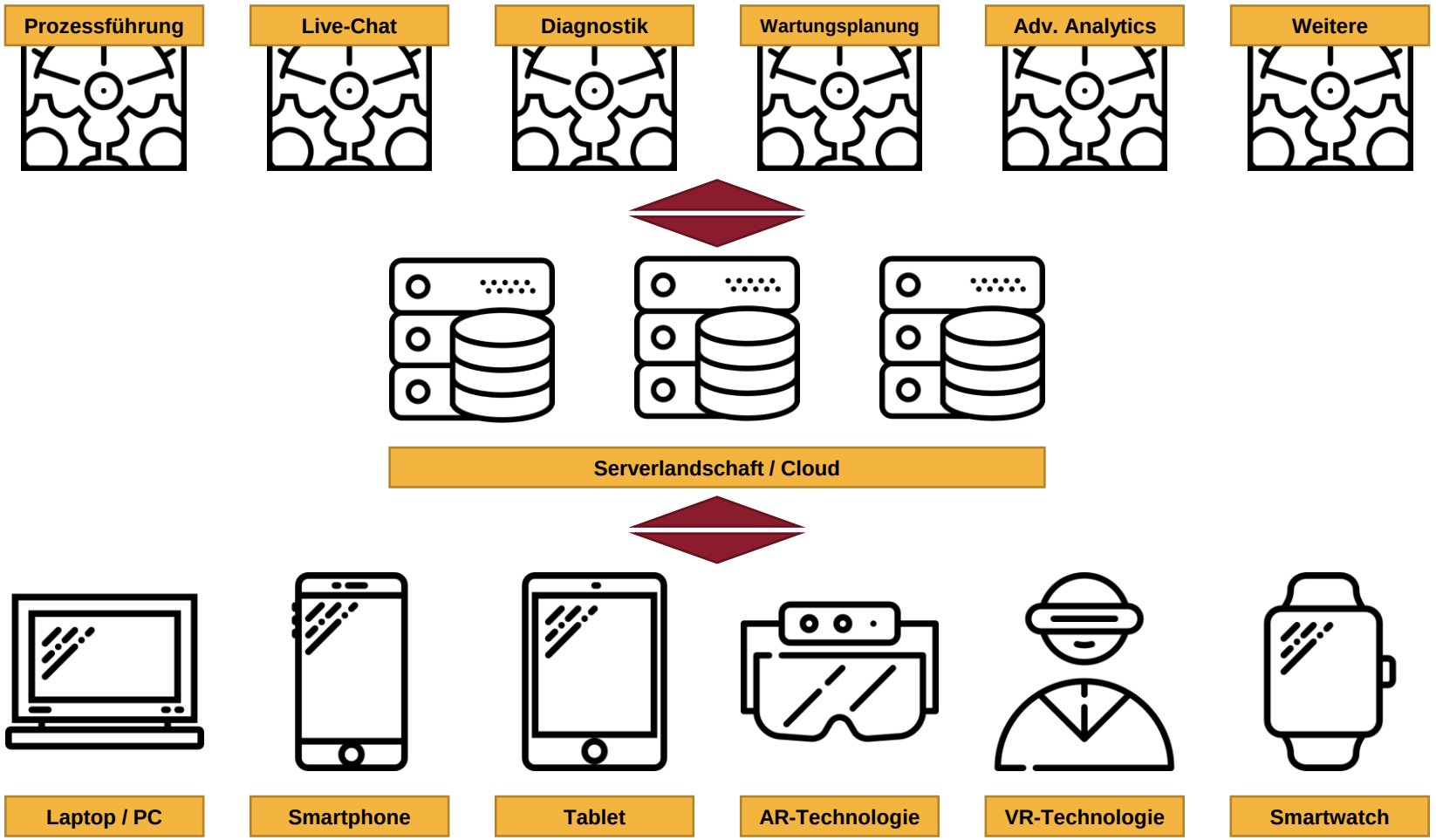


Quelle: Statista (ID 603793) - Experton 2015

Der Begriff „Big Data“ wird häufig dazu verwendet, um die sich verändernde Beschaffenheit von zu verarbeitenden Datenmengen zu beschreiben. Dabei wird oftmals auf die charakterisierenden Merkmale, die 4 Vs, verwiesen: Volumen, Velocity, Variety und Veracity.

Big Data ist jedoch mehr als eine Beschreibung großer Datenmengen. Big Data kann die Grundlage verschiedener neuer Geschäftsmodelle darstellen und bietet ein immenses Ertragspotential!

Einführung: smartTCS-Architekturskizze



Modulare Plattformarchitektur

Bisher wurden bereits zahlreiche potentielle technische Module konzipiert und in ersten Prototypen umgesetzt. Die Module können separat und integriert betrieben und auf verschiedenen Endgeräten genutzt werden.

Bei der Entwicklung der Kooperationsplattform setzt die Universität Osnabrück auf eine Kombination verschiedener Software-Tools und Technologien. Den Kern bildet das IMWI-Hadoop-Cluster.

Agenda

- Einführung
- Maschinendiagnostik
- Advanced Analytics und Predictive Maintenance
- Unterstützte Prozessführung in Wartungsprozessen

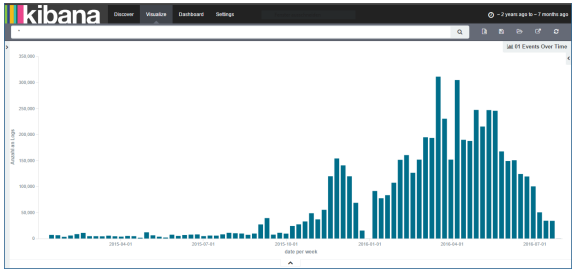
Maschinendiagnostik: Datenbasen

Maschinensensoren

Datenbeschaffenheit:
Instabile Informationszusammensetzung, gemischt strukturierte und unstrukturierte Daten, Vermischung multipler Signaltypen, sehr hohe Datenfrequenz

Vorteil:
Hohe Aktualität (Echtzeitnähe), hohe Prognosegüte, unverfälschte Daten, …

Herausforderungen:
Automatisierte Datenvorverarbeitung (insb. dynamische Informationsverdichtung), teils ist die Übertragung großer Datenmengen notwendig, Mitunter wird Rechenkapazität auf dem Endgerät benötigt, …

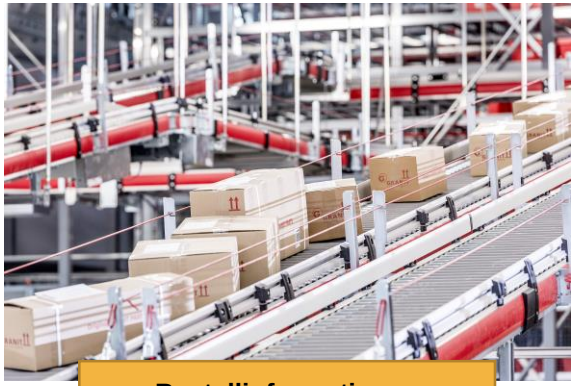


Fehlerlogs

Datenbeschaffenheit:
Weitestgehend standardisierte Daten mit einer stabilen Datenstruktur, häufig Zeilenweise Logs, Fehlereventbasierte Datensatzerzeugung

Vorteil:
Mittlere Aktualität, hohe Informationsabdeckung

Herausforderungen:
Geringere Schätzgenauigkeit, Approximation des Maschinenzustands notwendig, Häufig ist auch die Integration in ein Informationssystem notwendig (i.S.v. Management Dashboards) notwendig.



Bestellinformationen

Datenbeschaffenheit:
Hochstandardisierte Daten aus bestehenden Informationssystemen (Ersatzteilbestellungen)

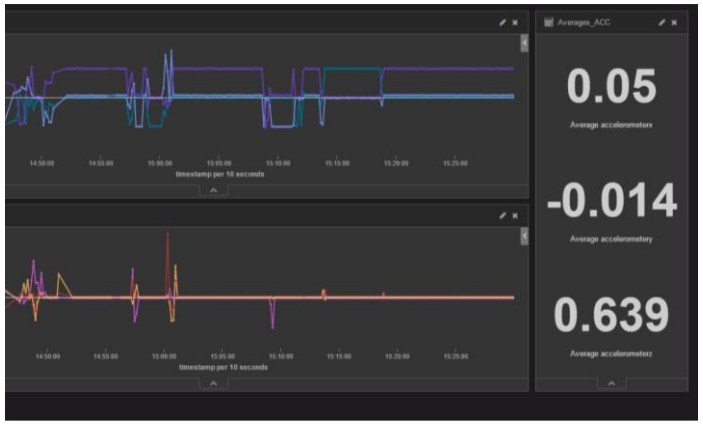
Vorteil:
Hohe Informationsverfügbarkeit, Nachfrage-Forecasting ist ein breit erschlossenes Forschungsgebiet

Herausforderungen:
Kein zwingender Zusammenhang zwischen Order und Defekt, sehr geringe Schätzgenauigkeit

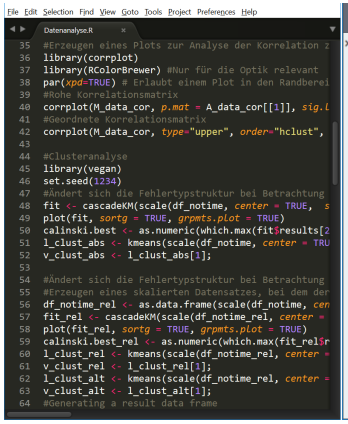
Maschinendiagnostik: Entwicklung technischer Module

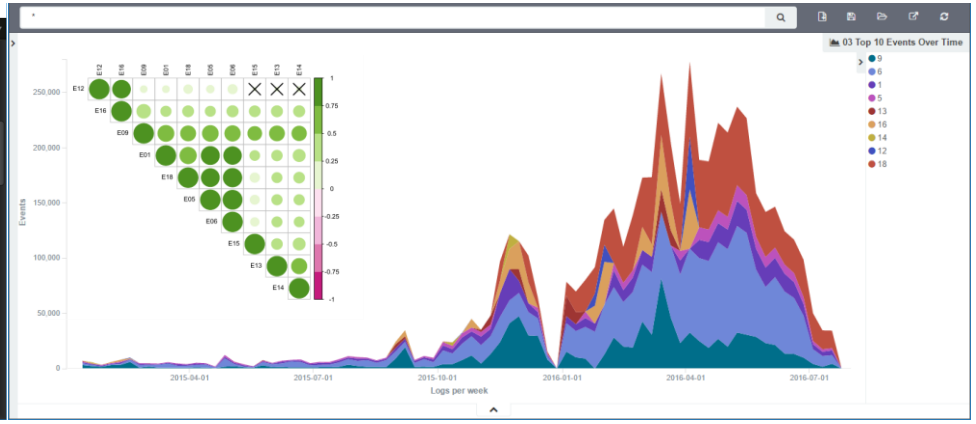
Sensor-streaming

Sensordatenstreaming



(Echtzeit-) Analyse





Ausgangssituation und Vorgehensweise

- Es wurden Sensorinformationen (insg. mehr als 10.000.000 Zeilen gemischter numerischer und Freitextinformationen) verschiedener Anlagen (bewusst ohne zusätzliche erklärende Informationen) zur Analyse zur Verfügung gestellt.
- In einem mehrstufigen Verfahren wurden die Dateien harmonisiert, in eine Fakten-tabelle überführt und mittels der Skript-sprache R weiterverarbeitet.
- In der automatisierten Aufbereitung wurden zusammenfassende Informationen (172 Datenspalten) zu 6183 Maschinen-durchläufen erkannt.

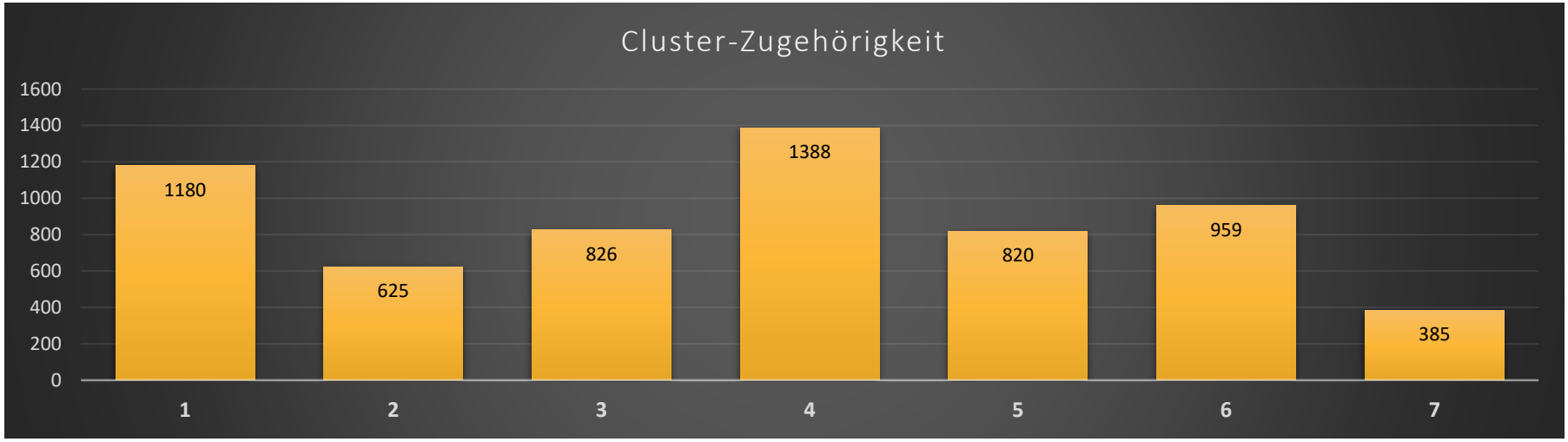
Cluster.r

```r
library(RMySQL)
library(psych)
library(pastecs)
library(FactoMineR)

con <- dbConnect(MySQL(),
                 user="imwi", password="passwort",
                 dbname="AV_Kunde", host="127.0.0.1")
on.exit(dbDisconnect(con))
rs <- dbSendQuery(con, "select * from FACT_DATA;")
data <- fetch(rs, n=-1)
dbSendQuery(con, 'drop table if exists R_DESC, R_COR,
    R_VAR')
dbDisconnect(con)

#Simple analysis
#Data normaliziation using data.Normalization from
clusterSim
#PCA
data.pca <- prcomp(data, center = TRUE, scale. = TRUE)
print(data.pca) #way too big to vizualize
plot(data.pca, type = "l")
summary(data.pca)

result <- PCA(data)

source('/home/andreas/R/Kunde/noClust.R')
n <- noclust(data) #Took about 6 hours to detect
calinski -> n = 7

fit <- kmeans(data, centers = 7)
ds <- cbind(data, cluster = as.factor(fit$cluster))

```

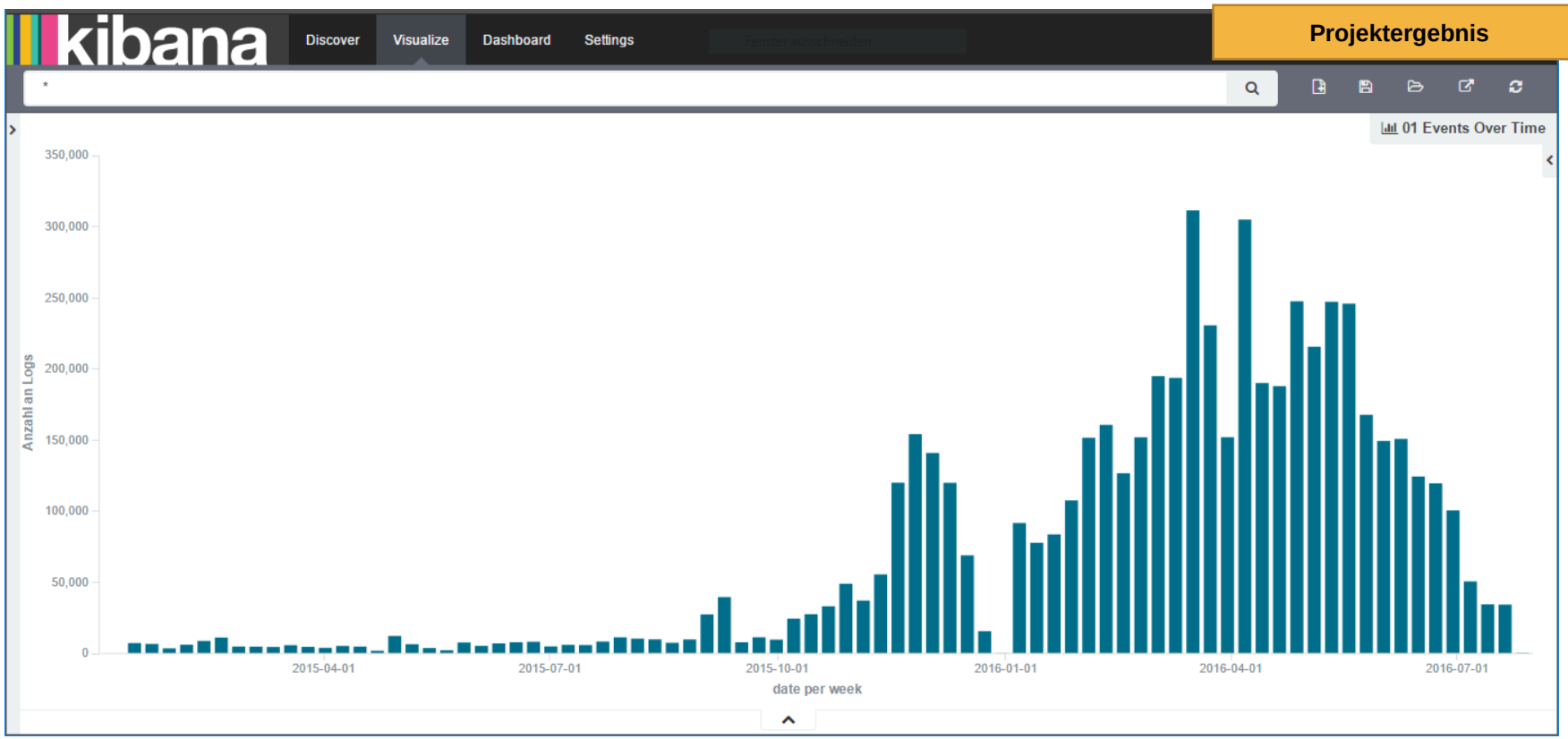
Zwischenergebnisse und nächste Schritte

Projektergebnis

- Die Anzahl erklärender Spalten konnten konnte um mehr als 90% reduziert werden
- In der Analyse wurden 7 unterschiedliche Maschinenzustände identifiziert
- Der nächste Schritt der Untersuchung ist eine Plausibilitätsprüfung in Zusammenarbeit mit dem Maschinenbetreiber



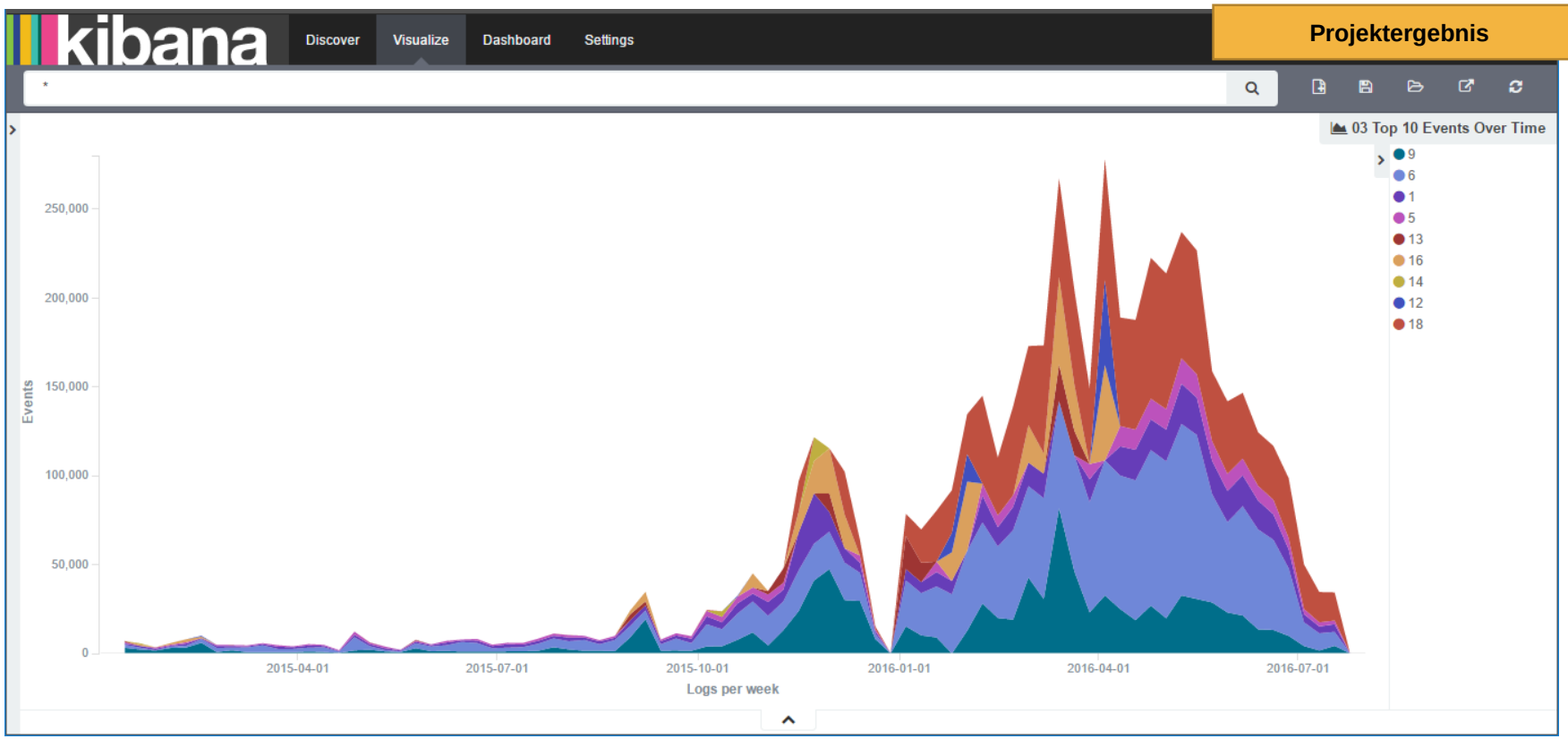
Maschinendiagnostik: Logdaten-Analyse (1)



Dargestellt: Häufigkeit aller Fehlerevents, wochenweise aggregiert

In den bereitgestellten Daten ist ein klarer Trend für eine steigende Häufigkeit der Fehlermeldungen etwa bis Ende Q1 2016 erkennbar. Dies lässt Rückschlüsse auf a) eine Veränderung der Anzahl der Log-Quellen (Anzahl der betrachteten Anlagen), b) eine Veränderung des Maschineneinsatzes oder c) eine zunehmende Fehleranfälligkeit (etwa auf Grund von Maschinenverschleiß) schließen. Weil in den bereitgestellten Logs keine Maschinen-IDs und keine (offensichtlichen) Tätigkeitsprofile hinterlegt sind, kann keine der Schlussfolgerungen ausgeschlossen werden.

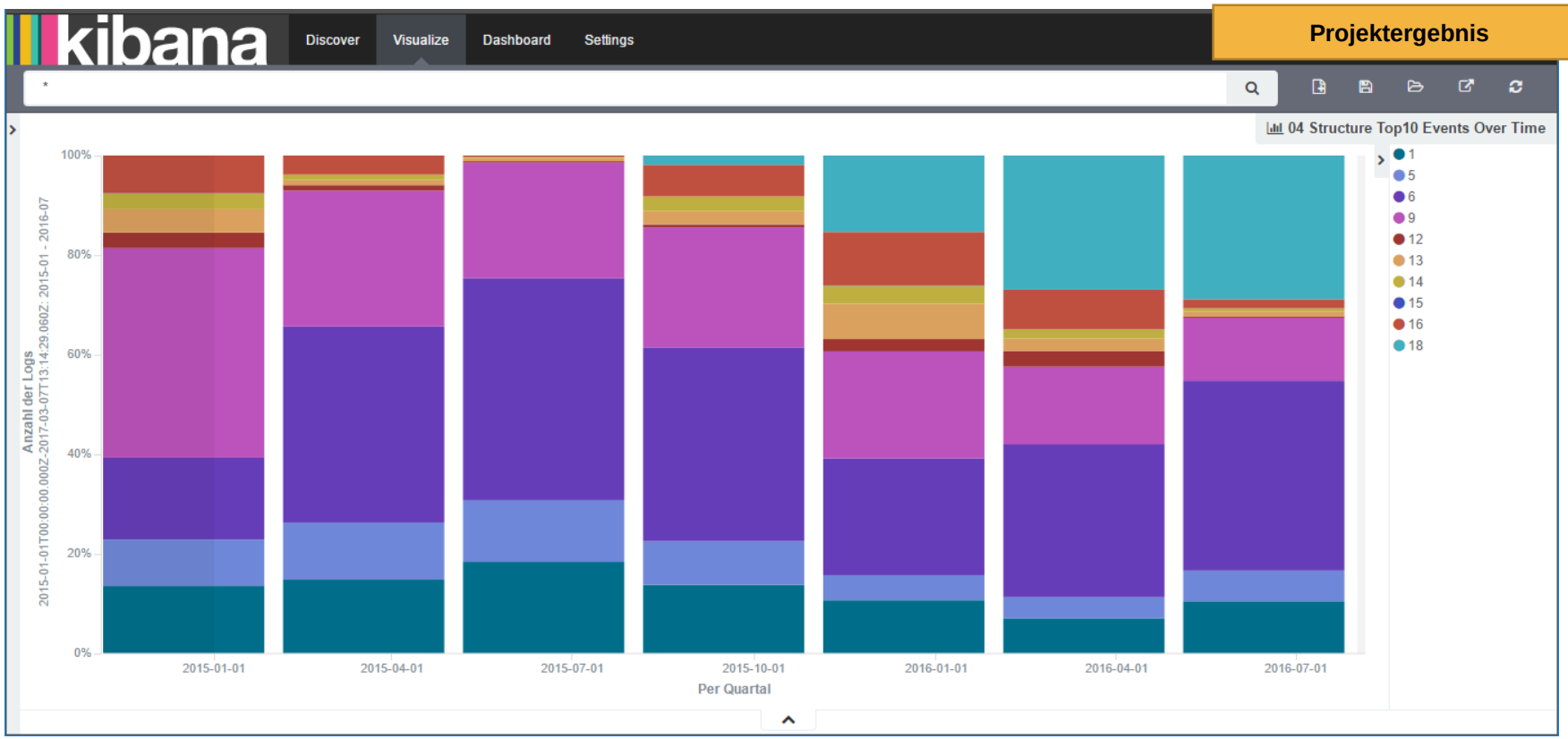
Maschinendiagnostik: Logdaten-Analyse (2)



Dargestellt: Häufigkeit aller event_ID-Einträge, wochenweise aggregiert (gestapelt)

Für einzelne Zeitstempel (genau auf die Millisekunde) treten Mehrfachnennungen der gleichen Event_ID auf. Diese werden als Detailinformationen eines Fehlers interpretiert und der entsprechende Fehler damit nur einfach gezählt. Diese Darstellung lässt die Identifikation von Top-Events (Events die über die Zeit stabil am häufigsten gemeldet werden) zu.

Maschinendiagnostik: Logdaten-Analyse (3)



Dargestellt: Relative Häufigkeit aller event_ID-Einträge, quartalsweise aggregiert

Die Events „06“ und „09“ treten im gesamten Zeitablauf in allen betrachteten Aggregationsstufen durchgängig am häufigsten auf. Die Struktur der Fehlerhäufigkeiten scheint sich jedoch im Zeitablauf zu verändern. Während bspw. Event-Meldungen des Typs „18“ in den ersten 3 Quartalen nahezu nicht auftreten, steigt die relative Bedeutung des Fehlertyps in den folgenden Quartalen stetig an. Dies ist ein Indiz für eine sich verändernde Fehleranfälligkeit (im Sinne eines veränderten Maschineneinsatzes oder im Sinne von zunehmendem Verschleiß).

Maschinendiagnostik: Rechtliche Herausforderungen

Datenschutz

- Welche Daten dürfen überhaupt verknüpft werden?
- Dürfen „personenbezogene Daten“ überhaupt regelmäßig ausgewertet werden?
- Inwiefern spielt ein unternehmens-übergreifender Datenaustausch hier eine Rolle?

Haftung

- Wie lässt sich die „Haftungsregelung“ beim Ausfall (oder bei Fehlfunktionen) einzelner Softwaredienste auf einer „Multi-Software-Anbieter-Plattform“ sinnvoll ausgestallten?
- Wer haftet, in welchem Maße, für „unrechtmäßig erschlichene“ Informationen durch Dritte?

Agenda

- Einführung
- Maschinendiagnostik
- Advanced Analytics und Predictive Maintenance
- Unterstützte Prozessführung in Wartungsprozessen

Predictive Maintenance: Erweiterte Datenanalyse

Um anhand der vorhandenen Daten Vorhersagen über Ausfallwahrscheinlichkeiten zu treffen und entsprechende Frühwarnsysteme zu konstruieren, bieten sich verschiedene Analyseansätze an. Diese umfassen u.a.:

(1) Ermittlung von Interdependenzen (i.S.v. Korrelationen, Hauptkomponenten und Granger-Kausalitäten) zwischen den Auftrittshäufigkeiten einzelner Fehlermeldungstypen zur Informations-reduktion und „Entkomplizierung“ eines entsprechenden Frühwarnsystems. Falls bspw. Rückschlüsse über 5 andere Fehler durch die Beobachtung einer einzelnen Häufigkeit gezogen werden können, ist es nicht notwendig, die Meldungen über diese 5 Fehler permanent zu überwachen.
Entweder muss ein über die Zeit konstanter Zusammenhang angenommen oder eine entsprechende Untersuchung regelmäßig wiederholt werden.

(2) Bestimmung von „Meldungszuständen“ bzw. Ermittlung der Veränderungen der Zusammensetzung unterschiedlicher Fehlermeldungen zu einzelnen Beobachtungspunkten. Wenn bspw. die Fehlerhäufigkeitsstruktur an einem bestimmten Tagesabschnitt, an einem Tag oder in einer Woche stark von den sonstigen Beobachtungen abweicht, kann dies als Indikator für ein Frühwarnsystem dienen.

(3) Modellierung von Event-Häufigkeiten als autoregressive (AR) Prozesse, um erwartbare Fehleranzahlen zu ermitteln und „Konfidenzintervalle“ abzuleiten, die visualisiert werden können. Beim Überschreiten eines erwartbaren Grenzwertes könnte bspw. eine Warnung ausgelöst werden.
Hierfür muss angenommen werden, dass die Anzahl der Log-Quellen über den Betrachtungshorizont konstant bleibt.

Predictive Maintenance: Methodenkanon zur Ausfallanalyse

Cluster-Analyse

Einsatzzweck
Erstanalyse zur Fehler- und Zustandstypisierung

Input & Output
I: Matrix mit unbegrenzter Anzahl an Informationen (Spalten) und zeilenweisen Beobachtungen
O: Identifikation (Spalte) der Anzahl und Art der unterschiedlichen Maschinenzustände im Datensatz

Methoden
- K-Means-Algorithmus
- Fuzzy C-Means

Klassifikation

Einsatzzweck
Zuordnung von neuen Informationen zu bekannten Zuständen (Frühwarnsys.)

Input & Output
I: Trainiertes Klassifikationsmodell & Informationsmatrix (mit stabiler Datenstruktur)
O: Zuordnung neuer Beobachtungen zu einem der im trainierten Analysemodell enthaltenen Maschinenzustände

Methoden
- Bayes-Klassifikatoren
- Entscheidungsbäume (C4.5)
- (Fuzzy-)ANNs
- SVMs

Korrelationsanalyse

Einsatzzweck
Informationsverdichtung und Bestimmung von Abhängigkeiten

Input & Output
I: Matrix mit begrenzter Anzahl an Informationen (Spalten) und zeilenweisen Beobachtungen
O: Sets mit „Korrelationskoeffizienten“ für die einzelnen Spalten / Hauptkomponenten

Methoden
- Lineare Korrelation (Pearson)
- Hauptkomponentenanalyse
- Uni- und multivariate Regressionen
- VAR-Modelle

Zeitreihenanalyse

Einsatzzweck
Vorhersage von Maschinenzuständen und -verschleiß

Input & Output
I: Zeitreihen einzelner Beobachtungen (1 Spalte) mit zeitlich äquidistanten Beobachtungen je Zeile.
O: Forecast über die Entwicklung der jeweiligen Zeitreihe.

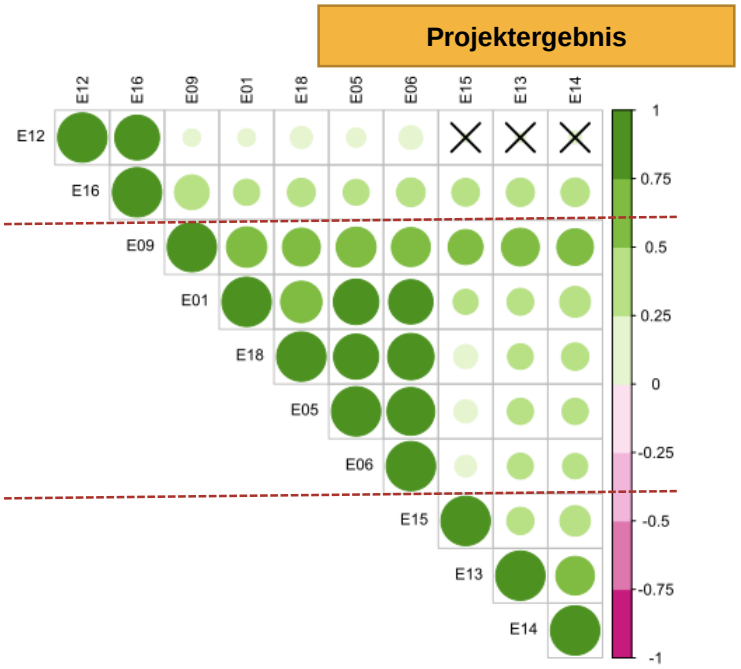
Methoden
- AR, ARMA, ARIMA
- Holt-Winters
- Monte-Carlo-Simulation

- Alle beschriebenen Methoden sind in Wissenschaft und Praxis anerkannte Verfahren.
- Die Algorithmen wurden in der freien Skriptsprache R (teils selbst) implementiert.
- Alle Skripte können individuell angepasst und erweitert werden.

Predictive Maintenance: Zusammenhänge zwischen Fehlertypen sind nachweisbar

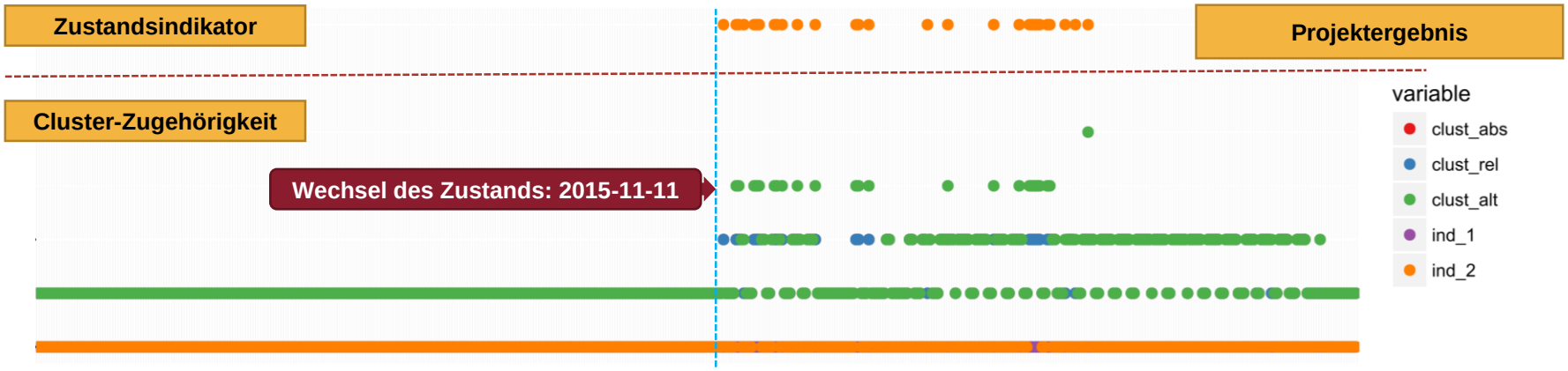
Durch eine Analyse der Korrelation der täglich auftretenden Event-IDs konnten 3 unterschiedliche Event-Typ-Gruppen ermittelt werden.

- Die Events „12“ und „16“ scheinen abhängig von einander, jedoch weitestgehend unabhängig von den anderen Fehlern aufzutreten. Insbesondere für das Event „12“ können nur sehr geringe lineare Abhängigkeiten zu anderen Meldungen nachgewiesen werden.
- Die Events „09“, „01“, „18“, „05“ und „06“ scheinen sehr stark positiv voneinander abzuhängen. Ein häufigeres Auftreten eines dieser Events lässt dementsprechend ein ebenfalls häufigeres Auftreten der anderen vermuten. Hier böte sich das gezielte Monitoring des Eventtypen an, welcher bei Auftreten die meisten Kosten verursacht.
- Die Events „15“, „13“ und „14“ scheinen in keinem starken Zusammenhang zu anderen Events zu stehen. Ein Auftreten dieser Events lässt dementsprechend keine Rückschlüsse auf andere zu.



Dargestellt ist eine Korrelationsmatrix, die über ein hierarchisches Clustering-Verfahren sortiert wurde. Während die Farbe Grün auf einen positiven Zusammenhang hindeutet (wenn x steigt, steigt auch y), zeigt Rot eine negative Korrelation an. Es wurde keine negative Korrelation erkannt. Die Größe der dargestellten Kreise deutet auf die Stärke des Zusammenhangs. Ein Kreuz zeigt an, dass kein statistisch signifikanter Zusammenhang besteht.

Predictive Maintenance: Struktur der Fehlerhäufigkeiten verändert sich

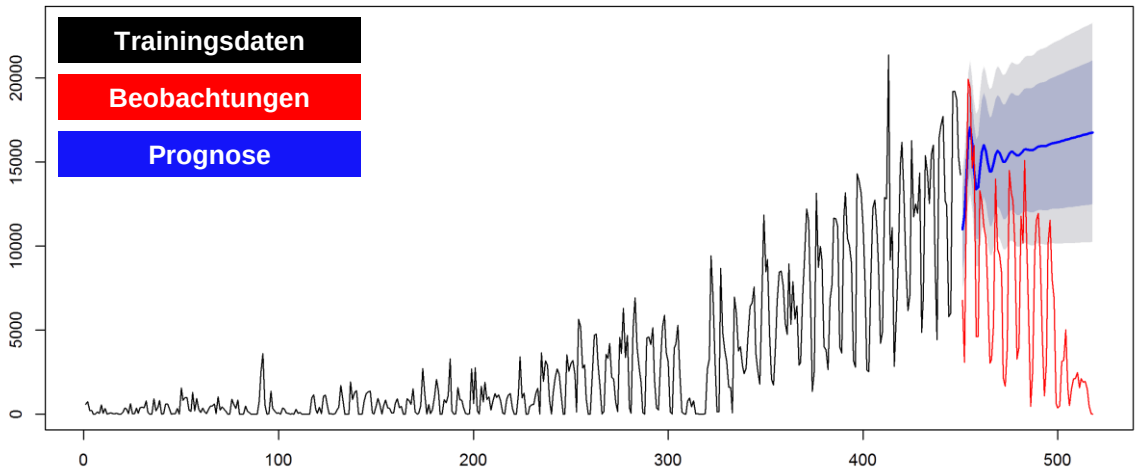


Dargestellt ist die Zuordnung der Tage im Beobachtungszeitraum zu einem „Messzustands-Cluster“ (Höhe). Die Farben Rot, Blau und Grün zeigen dabei die Zuordnung unter Verwendung verschiedener Verfahren und unter der Berücksichtigung von 2 (Blau und Rot) bzw. 4 (Grün) möglichen Clustern an. Die Farben Violett und Orange, oberhalb der gestrichelten Linie, deuten auf stark ungewöhnliche Maschinenzustände hin (Violett tritt nicht auf).

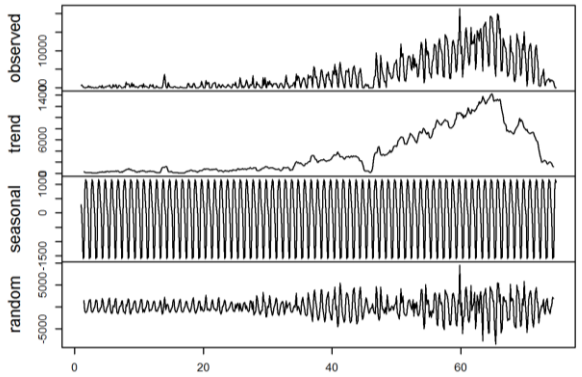
Verschiedene Klassifizierungsverfahren deuten darauf hin, dass es mindestens 2 (Calinksi), bzw. mindestens 4 (Bayes-IC), unterschiedliche Zustände (im Sinne von Tages-Typen) gibt. In der Darstellung erscheinen blaue und rote Markierungen (unterschiedliche Verfahren für die Zuweisung zweier Zustände) ausschließlich in den gleichen Clustern (Rot wird komplett von Blau überlagert). Das deutet darauf hin, dass die Zuweisung der 2 Zustände eindeutig ist. Ab dem 11.11.2015 treten wiederholt Fehlerhäufigkeitszusammensetzungen auf, die sich stark von den anfänglichen Beobachtungen unterscheiden.
Betrachtet man außerdem eine Zuordnung zu 4 Clustern (Grün), kann man „besonders auffällige“ Tage identifizieren. Dies eignet sich bspw. als Indikator für ein Frühwarnsystem.

Predictive Maintenance: Verfügbare Datenhistorie erlaubt nur kurze Prognosen

Forecast für Tageshäufigkeiten von event-ID-06 (ARIMA(3-1-2) with drift)



Projektergebnis



	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-2.27169	1834.028	1160.992	NaN	Inf	0.8272194	-0.03070048	NA
Test set	-8997.01132	10516.155	9215.374	-Inf	Inf	6.5660523	0.73155075	61.86201

Dargestellt ist ein Forecasting der zu erwartenden täglichen Fehlerhäufigkeiten für das Event „06“ unter Verwendung eines ARIMA(3-1-2)-Modells inkl. der entsprechenden Teststatistiken.

Es erscheint grundsätzlich möglich, die zu erwartenden Tageshäufigkeiten für das betrachtete Event kurzfristig mit einer hohen Genauigkeit zu prognostizieren. In der Darstellung ist dies daran ersichtlich, dass die blaue Linie (Forecast) zu Anfang der Prognose sehr nah an den tatsächlichen Beobachtungen (rote Linie) liegt. Mit zunehmender Länge des Forecasts wird die Prognose jedoch ungenau. Etwa nach 10 Tagen ist im konkreten Fall eine zu starke Abweichung zwischen Beobachtungen und Forecast feststellbar. Hierfür kann es jedoch verschiedene Gründe geben. Forecasts für andere Events sowie unter Betrachtung unterschiedlicher Aggregationsstufen bleiben zu untersuchen.

Predictive Maintenance: Rechtliche Herausforderungen

Eigentum

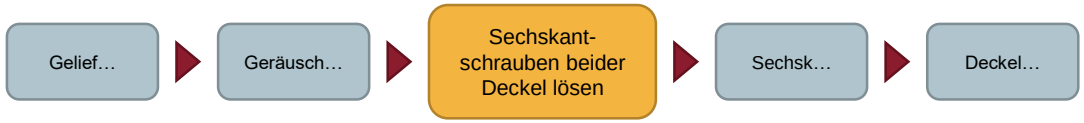
Maschinendaten und darauf angewandte Analysemethoden stellen ein erhebliches Einkommenspotential dar. Nicht nur können dadurch Wartungskosten, die im Gewährleistungszeitraum anfallen, verringert werden, es können auch neue Dienstleistungen angeboten und Kundenzufriedenheit, -bindung und -gewinnung verbessert werden.

- Wem aber gehören die Daten, die eine Maschine während des Betriebs erzeugt?
- (Wie) Können die Eigentumsverhältnisse an den erzeugten Daten im Fall von Maschinen- und Anlagenleasing resp. Miete geregelt werden?
- Kann man einen internationalen Anspruch auf Datenfreigabe überhaupt durchsetzen?

Agenda

- Einführung
- Maschinendiagnostik
- Advanced Analytics und Predictive Maintenance
- Unterstützte Prozessführung in Wartungsprozessen

Prozessführung: Prozessunterstützung mit AR-Brillen



Eines der zentralen Dienstleistungsmodule, die im Rahmen von smartTCS konzipiert und prototypisch bereits auf Basis verschiedener Datenbrillen implementiert wurde, ist die interaktive Prozessführung. Mit diesem Modul ist es möglich, die Durchführung einzelner Dienstleistungen gezielt aus der Ferne zu unterstützen und zu überwachen. In bestimmten Fällen ist es durch das Unterstützungssystem sogar möglich, Dienstleistungen komplett durch Kunden zu erbringen.

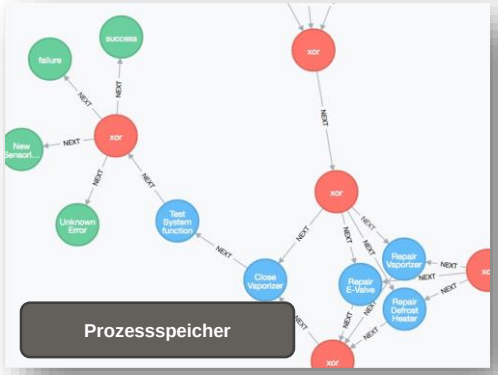
Die Prozessführung beinhaltet neben einer reinen Übersicht vorangegangener, aktueller und nächster Tätigkeiten auch die Bereitstellung von kontextspezifischen Detailinformationen und den Zugriff auf Live-Kommunikationssysteme zur Echtzeit-Remote-Unterstützung im Bedarfsfall.

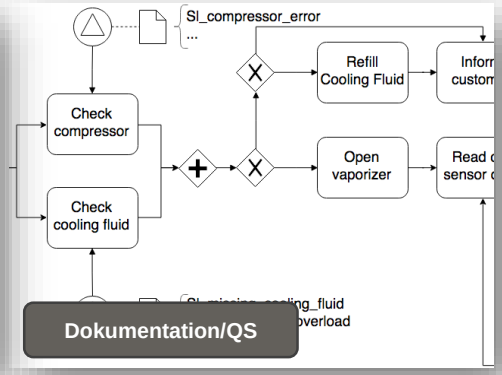
Prozessführung: Entwicklung technischer Module (1)

Prozess-führung

Prozessaufnahme

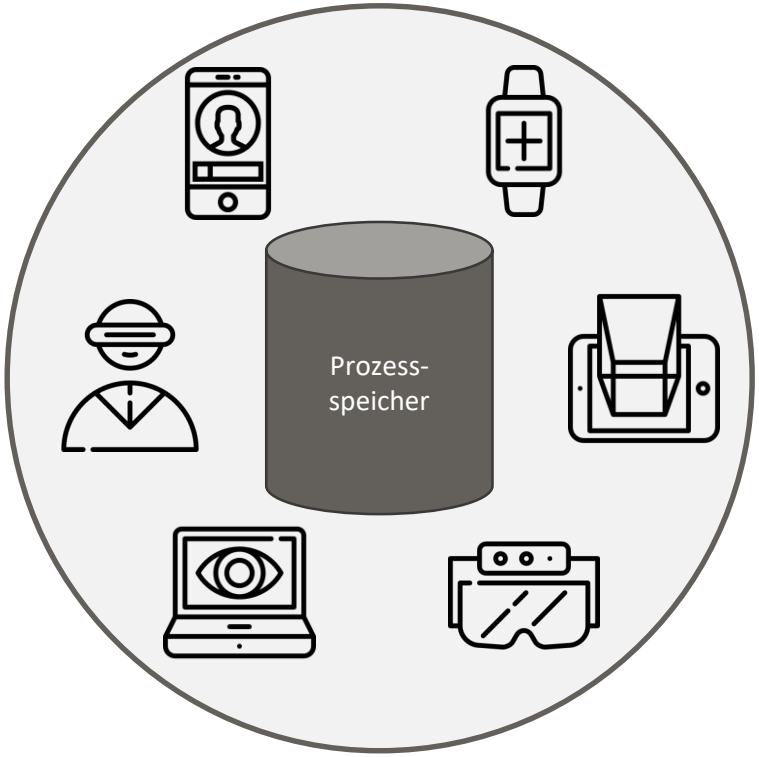
Repräsentation





Prozessführung: Entwicklung technischer Module (2)

Prozess-führung



Prozessführung: Rechtliche Herausforderungen

Haftung

- Wer haftet bei fehlerhaft durchgeführten, angeleiteten Wartungsarbeiten?

- Und wie ist ggf. die Beweispflicht im Hinblick auf die Fehlerursache?

„Globalisierung“

- Gibt es international andere Anforderungen, Berechtigungen und Vorgaben Informationen zu erheben und bereitzustellen?

Vielen Dank für Ihre Aufmerksamkeit!
Haben Sie Fragen?

- smartTCS-Projektteam der Universität Osnabrück -

Prof. Oliver Thomas
oliver.thomas@uos.de

Friedemann Kammler
friedemann.kammler@uos.de

Andreas Varwig
andreas.varwig@uos.de

Backup: Big Data im Maschinen- und Anlagenbau
- Einsatzmöglichkeiten am Beispiel des Projekts smartTCS -

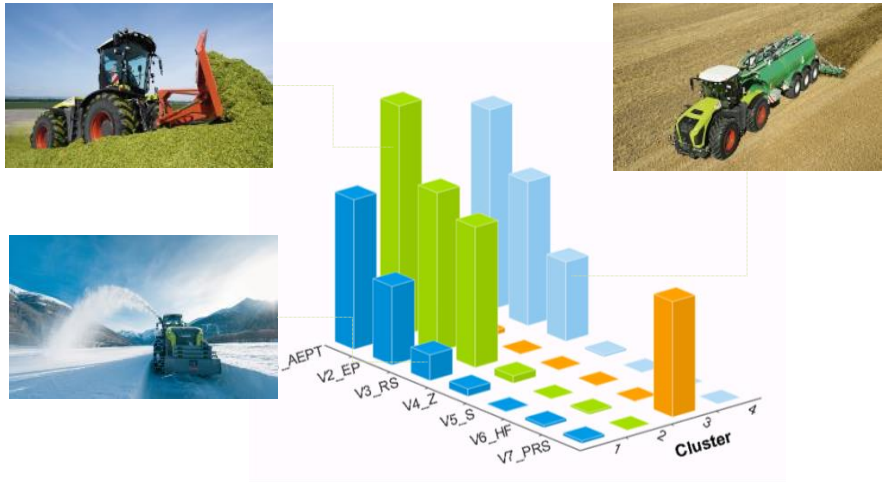
Andreas Varwig

Lehrstuhl für Informationsmanagement und Wirtschaftsinformatik (IMWI)
Universität Osnabrück

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Projekthistorie: Andreas Varwig (1)

Predictive Maintenance durch Aktivitäten-und Belastungstracking
Streaming und Echtzeitanalyse von Sensordaten



Ergebnisse:

- Anhand der Sensordaten konnten diverse Applikationen (Tätigkeitscluster) identifiziert und bislang unerwartete Einsatzgebiete ermittelt werden.
- Diese Tätigkeitscluster bilden nun die Grundlage für Belastungs- und Lebensdauer-prognosen.

Motivation:

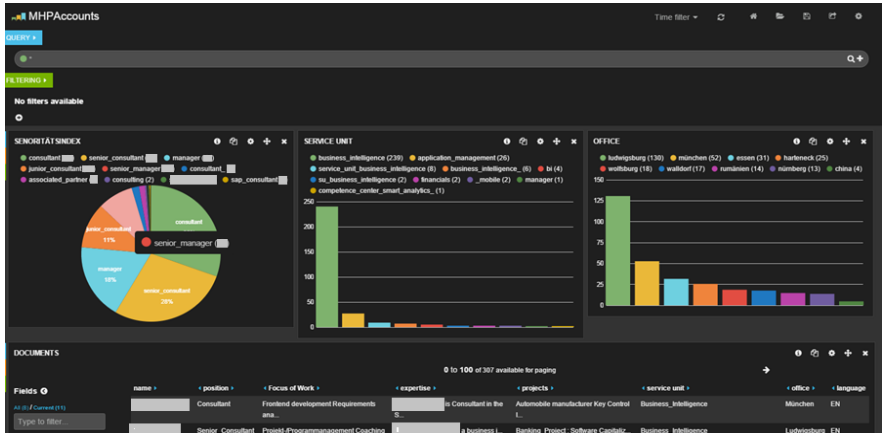
- Der Projektpartner möchte die tatsächlichen Einsatzgebiete und Anwendungsfelder der verkauften Nutzfahrzeuge ermitteln. Hierzu sollen simultan die Daten von hunderten Sensoren ausgewertet werden.
- Dadurch sollen auch die Material- und Werkzeugbelastungen und deren erwartete Lebensdauer bestimmt werden.

Eingesetzte Technologien und Methoden:

- Streaming von Sensordaten mittels Apache Storm in ein Hortonworks Hadoop-Cluster
- Cluster- und Anomalieanalyse durch Fuzzy-K-Means-Clustering und Neuronale Netze (PNNs) in R
- Informationssysteme und Dashboards mit HTML5, Kibana und SAP BO Design Studio

Projekthistorie: Andreas Varwig (2)

Strategische Personalplanung mit Webcrawling und Text-Mining Bedarfsanalyse und Predictive Knowledge Management



Ergebnisse:

- Es wurde eine easy-to-use (mobile/ Freitextsuche) Oberfläche zur Auswertung vorhandener Kompetenzen bereitgestellt.
- Eine live-Schlagwortanalyse in Stellenanzeigen (Stepstone, Indeed, Monster, Karriereportale von Konkurrenten) verschafft nun auch on-demand Einblicke in die strategische Personalplanung von Konkurrenten.

Motivation:

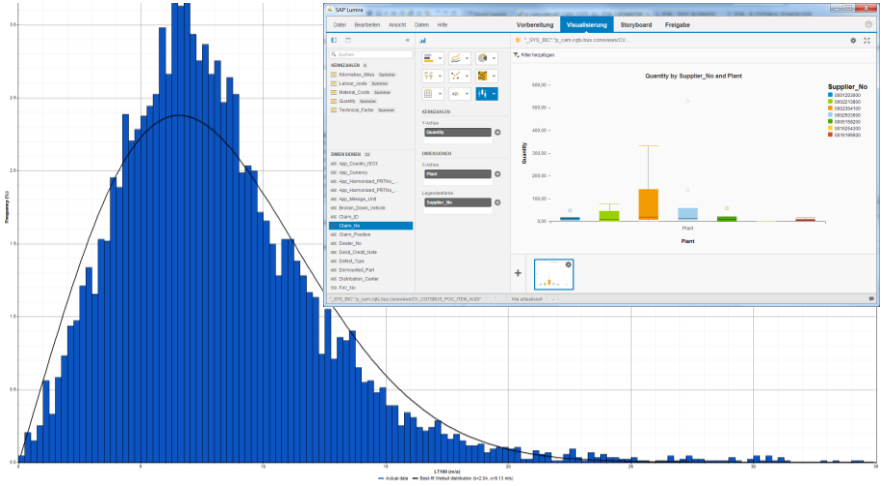
- Beim Projektpartner besteht, trotz tausender vorhandener Mitarbeiterprofile, Ungewissheit über die im Unternehmen vorhandenen Kenntnisse, Fähigkeiten und Software-Erfahrungen.
- Außerdem sollen strategische Kompetenz-bedarfe für anstehende Einstellungen ermittelt werden.

Eingesetzte Technologien und Methoden:

- Webcrawling mit Apache Nutch
- Text Mining (Word Stemming und Häufigkeits-analysen) mit R und Rapidminer
- Kibana Dashboards

Projekthistorie: Andreas Varwig (3)

Predictive Quality Management durch Materialausfall-Forecasting
Berechnung von Ausfallwahrscheinlichkeiten und Frühwarnsysteme



Ergebnisse:

- Bislang unbekannte Ursachen für die Beanstandung bestimmter Sensorgruppen konnten identifiziert werden.
- Konsolidierte Erfassung von Mängeln ermöglicht ein taggenaues Tracking und die Identifikation von Unregelmäßigkeiten
- „Erhebliche Vereinfachung der Qualitätssteuerung“

Motivation:

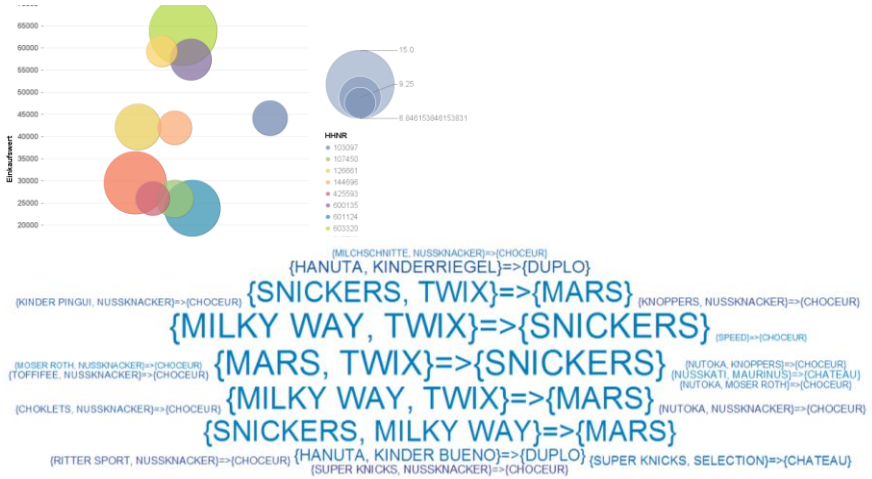
- Eine neue Strategie zur nachhaltigen Optimierung der Lebenszyklen und Minimierung der Reklamationen von Sensoren und Baugruppen soll umgesetzt werden.
- Bislang fehlten umfassende Frühwarnsysteme, die eine ungewöhnliche Häufung von Mängeln auf Einzelproduktebene aufzeigen und Ursachenforschung ermöglichen.

Eingesetzte Technologien und Methoden:

- Konsolidierung von Reklamationsfällen und Materialmängeln in SAP HANA
- Automatisiertes Forecasting von Konfidenzintervallen für Defektzahlen und Reklamationen aller vorhandenen Sensortypen mit Weibull-Regressionen und Pfadsimulationen in R

Projekthistorie: Andreas Varwig (4)

Kundenclustering und Kampagnensteuerung mit Bondaten
Cross-Selling-Potentiale und Real Time Marketing



Ergebnisse:

- Auf Basis der Bondaten konnten (je nach gewählter Trennschärfe der Analyseverfahren) 7-28 Einkäufergruppenprofile ermittelt werden, die eine gezielte Steuerung von Preis- und Cross-Selling-Kampagnen ermöglichen.
- Außerdem konnten erwartete „Up-Lift-Quoten“ zur monetären Bewertung von Kampagnen-erfolgen berechnet werden.

Motivation:

- Der Projektpartner ist Mitglied im Arbeitskreis „Real Time Marketing“ und verfügt über die GfK über Bondaten und Kundeninformationen von ~1 Mio. Einkäufen.
- Auf Basis der Bondaten sollen Algorithmen zur Bestimmung von Kundengruppen (Clustern), Verhaltensprofile und Absatzwahrscheinlich-keiten abgeleitet werden.

Eingesetzte Technologien und Methoden:

- Ermittlung von Kundenclustern mit SAP Predictive Analytics
- Bestimmung von Cross-Selling-Potentialen durch Assoziations- und Musteranalysen (Apriori, FPGrowth) mit R

Verwendung von R zur erweiterte Datenanalyse

Zur Durchführung der beschrieben Analysen wird die freie Skriptsprache R eingesetzt. Hierbei kann weitestgehend auf bereits bestehende, kostenlos und frei verfügbare Implementierungen von Analysealgorithmen zurückgegriffen werden. Lediglich für einzelne, datenspezifische und individuelle Bearbeitungsschritte musste eigener R-Code erstellt werden. Insgesamt sind so ca. 200 Zeilen Code (inkl. Kommentar) entstanden.

Alle nachfolgend beschriebenen Ergebnisse wurden unter Betrachtung von, nach Event-IDs und Tagen, aggregierten Fehlerhäufigkeiten ermittelt. Es bleibt zu untersuchen, ob sich die Ergebnisse bei anders aggregierten Daten bestätigen lassen.