



WESSLING

Quality of Life





Forum Labor- und Qualitätsmanagement
Mainz, 04.09.2019

Big Data – Ein Hype, der vorübergeht? (Zusammenfassung)

Dr. Michael Winterstein
WESSLING GmbH
48341 Altenberge
michael.winterstein@wessling.de



Entscheidungsbäume

Künstliche
Intelligenz

Data Mining

Industrie 4.0

Deep
Learning

Neuronale
Netze

Predictive Maintenance

PLS

Smart Data

Text Mining

Machine
Learning

Digitalisierung

Weckt dies Neugier
oder Angst?

Was soll ich damit?

Muss ich das verstehen?

Hilft mir das in meinem Job,
welchen Nutzen habe ich
davon?

„Es kommt von überall her. **Es weiß alles.** Sein Name ist Big Data.“ (Dilbert/Scott Adams)

Charakteristik von Big Data: Die 5 V

- **V**olumen (Datenvolumen)
- **V**eracity (Wahrhaftigkeit, Fehleranteile...)
- **V**elocity (Daten in Echtzeit verarbeiten)
- **V**ariety (Datenvielfalt – Texte, Bilder, Videos, Sensordaten)
- **V**ariability (Daten verändern sich schnell)

„Ziel von Big Data ist nicht die kommerzielle Welt, sondern die Wissenschaft selbst, die nicht mehr benötigt würde.“

Hypothesen, Modelle, Kausalitäten sind überflüssig, weil einem Big Data zeige, wie die komplizierte Wirklichkeit wirklich sei.

Verständnis ist ohnehin in Anbetracht der Komplexität hoffnungslos.

Big Data würde uns von diesem umständlichen und untauglichen Hilfsmittel „Wissenschaft“ befreien.

(Chris Anderson, 2008)

WARUM DATA MINING? – Wir messen doch schon alles!



Tycho Brahe (1546-1601)

Umfassende Datenreihen über die Stellung des Mars

Aber: keine Vorhersage möglich

MESSEN



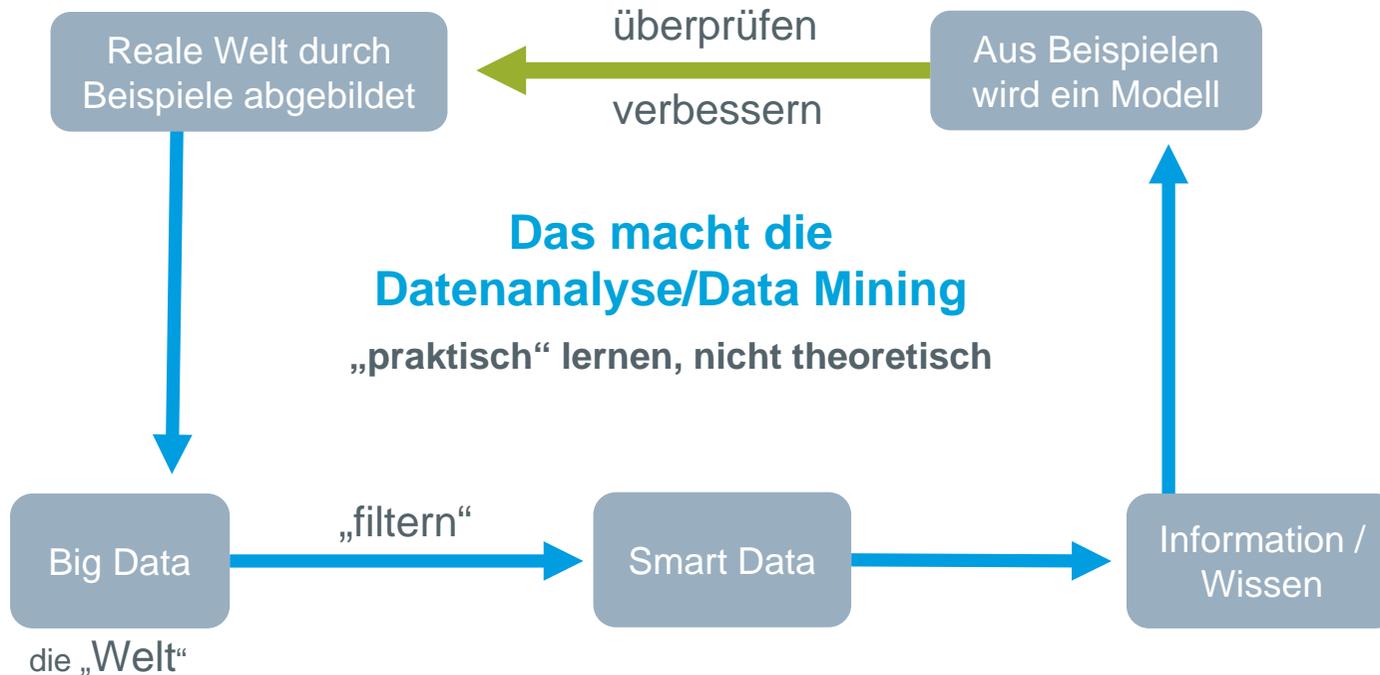
Johannes Kepler (1571-1630)

Leitet aus diesen Datenreihen die drei Keplerschen Gesetze ab.

Nun: Vorhersagen möglich

DATA MINING

Aus Daten zu neuem Wissen! Neues Wissen zu finden ist aber **sooo schwer....**
(Data Mining als Ausweg)

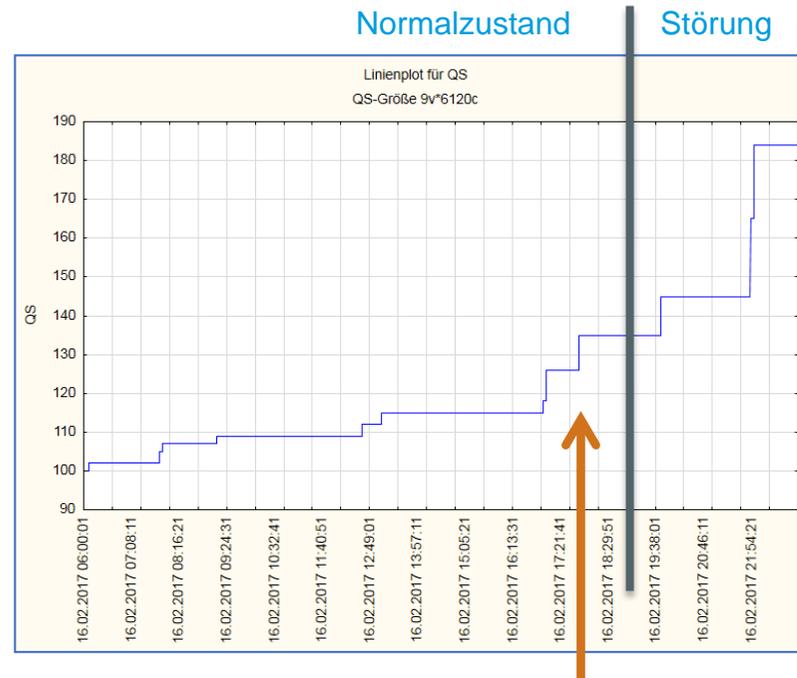
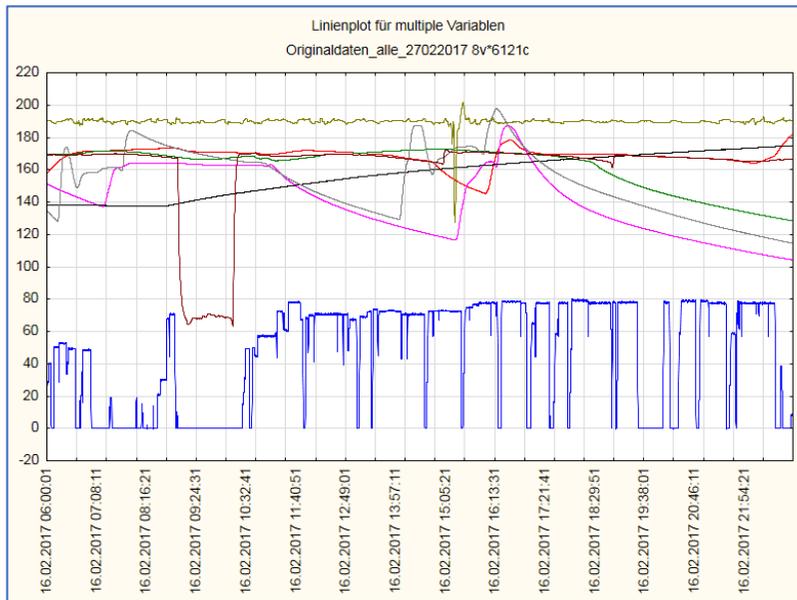


**Keine Angst vor Statistik/Datenanalyse –
alles Job der Data Scientists**
eigene Kapazitäten oder Partner suchen.

Wäre es nicht toll, wenn man Folgendes lösen könnte?

- Non-Target-Analyse (untersuche nur Veränderungen!)
- Bilddateien auswerten (Chromatogramme)
- Automatische Fehlererkennung im Labor (Ziel: Nullfehlerproduktion)
- Angebotserstellung (automatische Preisregulierungen)
- Rechnungskontrollen/Zahlungseingänge
- Vorbeugende Wartung teurer Messgeräte (kein Ausfall mehr!)
- Auswertung Fachliteratur/Webseiten
- „Rettung“ und ständige Verfügbarkeit von Fachwissen
- Auffinden aller Unterlagen zur Probe x ohne Suchen
- Automatische Aktualisierung aller Akkreditierungsunterlagen
- Papierloses Labor

Wenn Sie 1400 Merkmale messen, erkennen Sie dann noch die wesentlichen Zusammenhänge?
Bei mehr als 3 Dimensionen hat der Mensch eben ein Problem, die Mathematik jedoch nicht...



Erkennt „einfaches Muster“ –
wie, das muss man nicht verstehen...(kann man aber!)

Hier eingreifen!
Ziel: „Big Data“

- **KI** – Wissenschaft, wie Computer vom Menschen lernen können (Automatisierung intelligenten Verhaltens)
- **Data Mining/Text Mining/Web Mining** (Wissenszuwachs aus Daten/Texten/Web)
- **Machine Learning** – wie Maschinen/Computer lernen, Wissen aus Erfahrung (Daten) zu generieren – gibt sehr viele, tolle Verfahren!
- **Neuronale Netze** – Technik des Lernens. Vom Gehirn abgeschaut, Parallelverarbeitung unzähliger Verbindungen und ihren Gewichten
- **Deep Learning (tiefergehendes Lernen)** – mit vielschichtigen Neuronalen Netzen in gewaltigen Datenmengen Muster finden (Bilder, Sprachen erkennen)
- **Support Vector Machines** – Transformation in einen höherdimensionalen Hyperraum, in dem Entscheidungen einfacher werden
- **Entscheidungsbaumverfahren** – Suche nach besttrennendsten Merkmalen, abgeleiteten Regeln, oft sehr verständlich

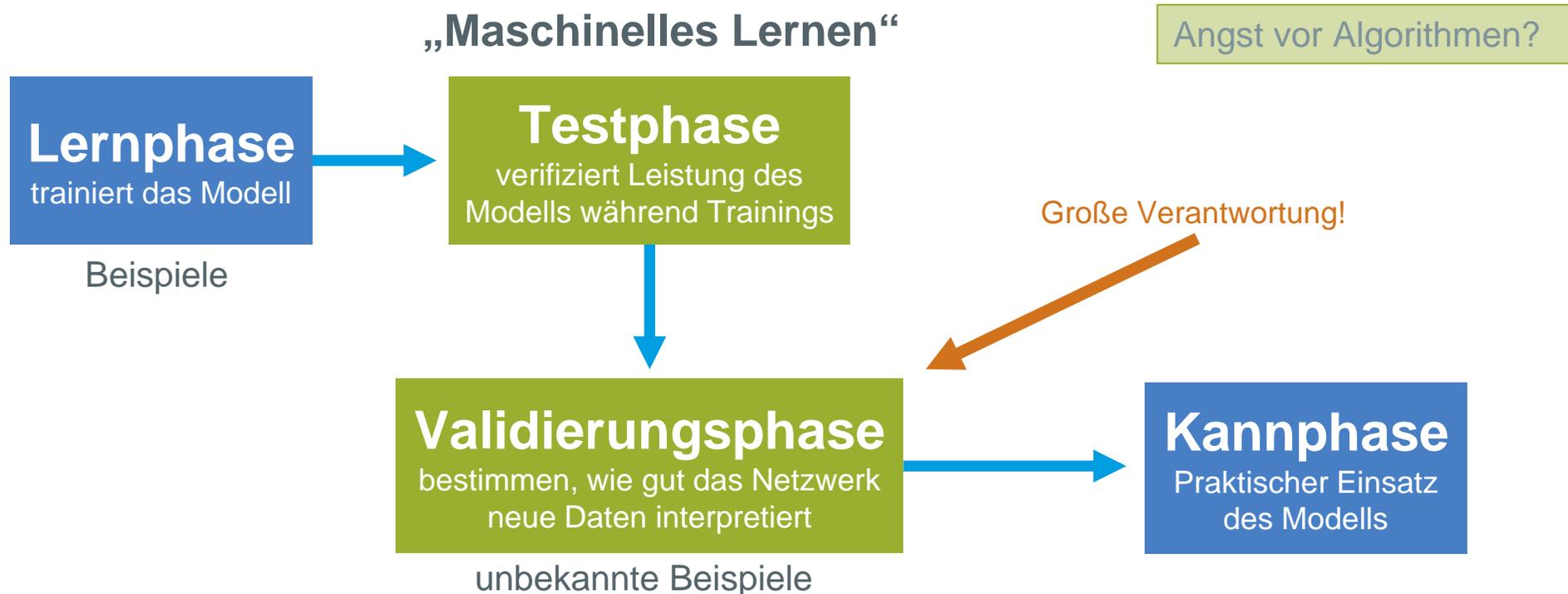
Methoden-Mix → **Data Scientist** (heute, aber später...?)

PHASEN BEIM LERNEN AUS BEISPIELEN

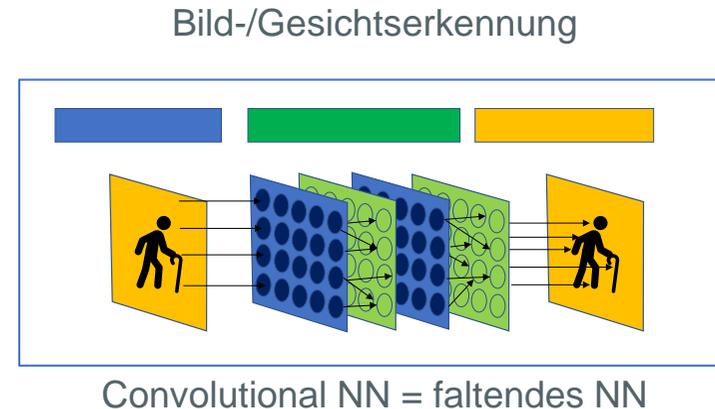
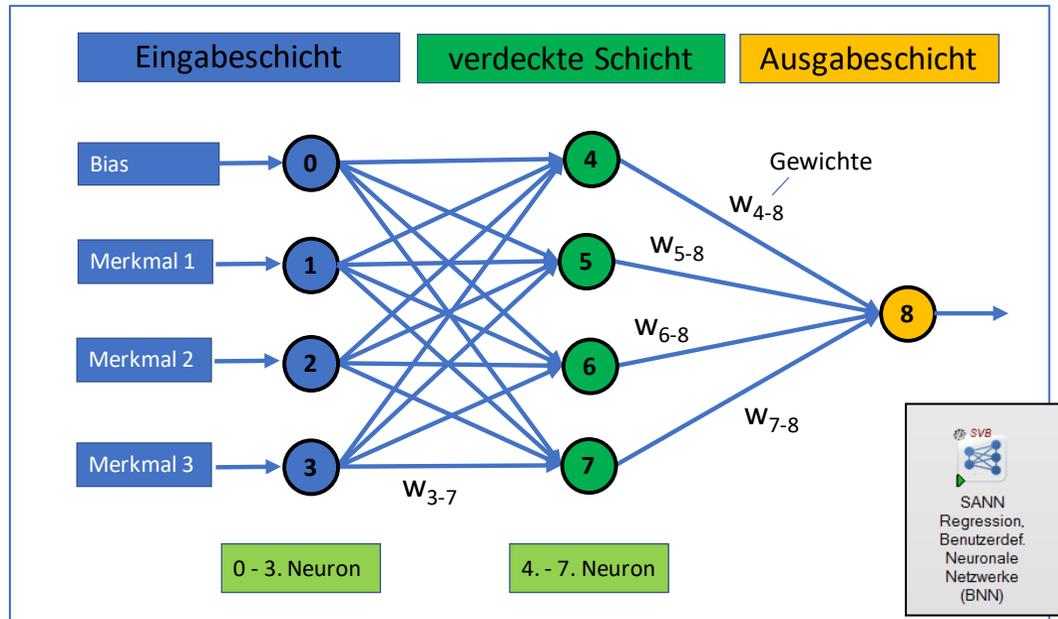
Eine Möglichkeit des Lernens = „Lernen aus Beispielen“ = Historie/**Erfahrungen**

Historisches Wissen = Beispiele in Datenbanken (ideal für Computer!).

Wenn diese Beispiele die Welt repräsentieren und der Computer genau dieses Wissen hat, muss man da Angst haben?
Doch nur, wenn Beispiele falsch/nicht repräsentativ sind!



Mehrschichtiges Backpropagation-Netzwerk

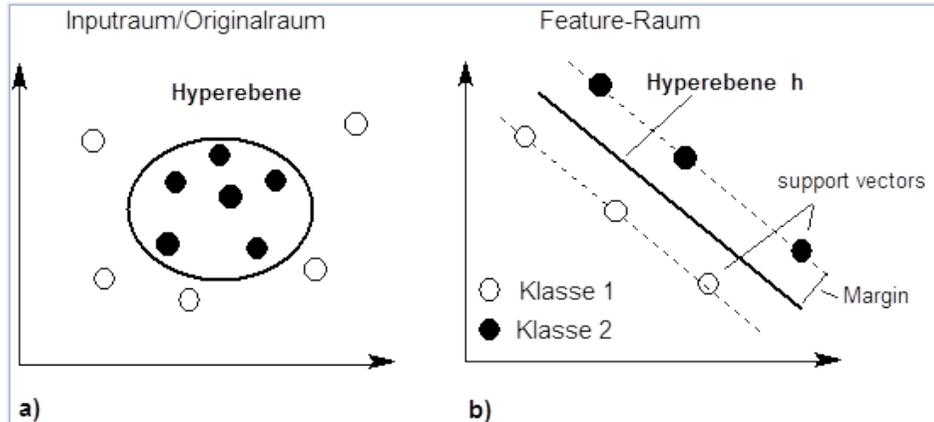


Überwachtes Lernen: Lernt aus Beispielen (mit Lehrer).

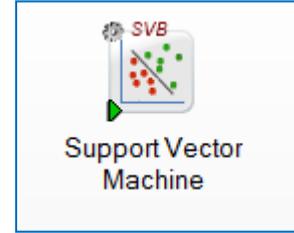
Nichtüberwachtes Lernen: Findet Zusammenhänge/Ähnlichkeiten selbst.

Kann z. B. ein schlecht messbares Merkmal (8) durch andere abbilden (1-3)
MSPC – Multivariate statistische Prozesskontrolle

Einfache Separation nicht möglich!

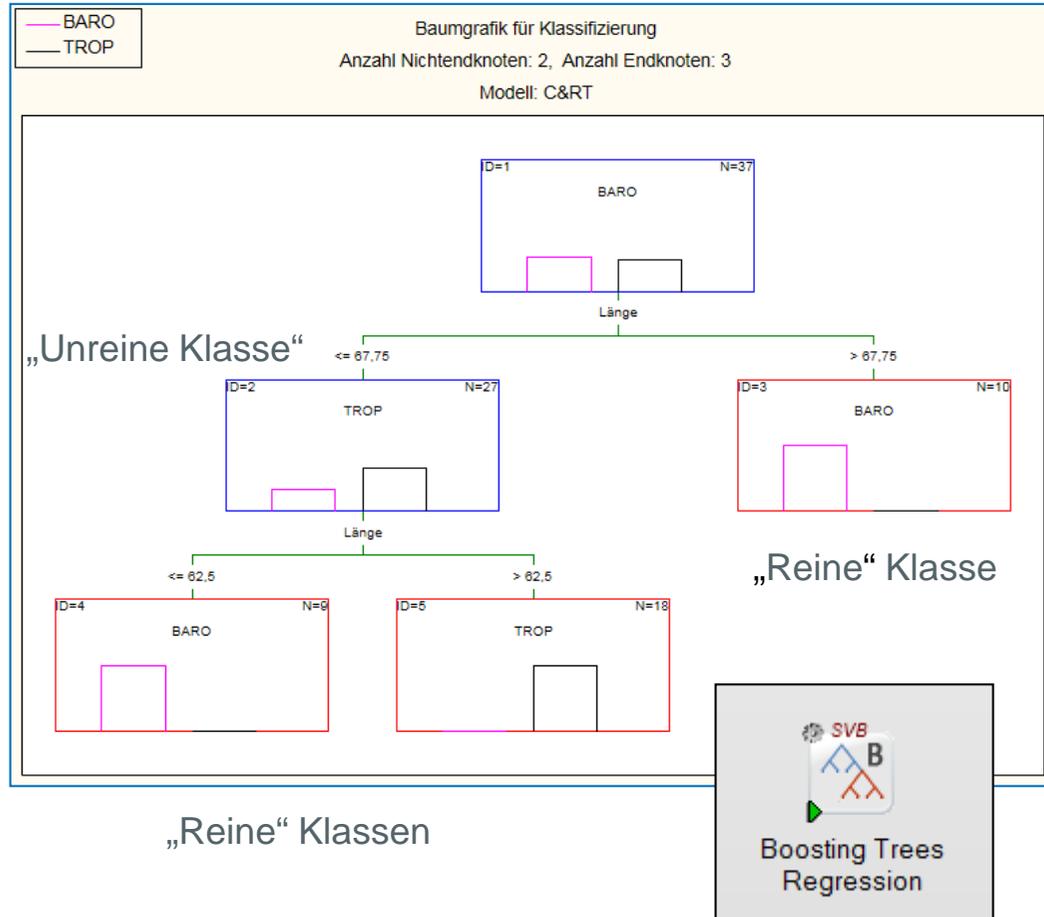


Transformation in höherdimensionalen Feature Raum,
dort ganz einfach zu trennen



Muss nicht auf alle Punkte
schauen, nur auf die
entscheidenden an der
Trennebene.

Findet die Merkmale, die am besten trennen und erstellt ein Modell. Regeln sind **verständlich**.



Bspl. Spamfilter (Text Mining)

Spam: Variable Importance



57 WEBSEITEN – HÄUFIGSTE WÖRTER SUCHEN



57 Webseiten aktuell auswertbar, häufigste Wörter dargestellt.

Allgemeine

- Andere lösen Ihre Probleme nicht – also machen Sie es!
- Beschäftigen Sie sich ein wenig mit Möglichkeiten/ Algorithmen
- Daten sind wertvoll – nutzen Sie die Möglichkeit, mit Werkzeugen dort Wissen herauszuholen!
- Zeigen Sie, dass Sie Chef und KI (nur) Werkzeug ist.

Strukturelle

- Nicht Frage, ob man sich damit beschäftigt, sondern wie!
- Eigene Kapazitäten, oder Partner binden (Prozessverständnis entscheidend)

Konkrete

- Start mit überschaubarem Projekt (mit Geräteherstellern?)
- Erstellen Sie eine Road-Map
- Partner/Verbündete in Firma finden
- Den Nutzen managergerecht darstellen

Problem?

Daten vorhanden?

Starten!



WESSLING

Quality of Life