

Data Mining – Möglichkeiten und Anwendungsfelder

15. Symposium on Privacy and Security, Zürich, 31. 8. 2010



Dr. Stefan Rüping
Fraunhofer IAIS

© Fraunhofer Institut für intelligente
Analyse- und Informationssysteme IAIS

 **Fraunhofer**
IAIS

Data Mining – Möglichkeiten und Anwendungsfelder

- Vorstellung
- Was ist Data Mining?
- Aktuelle Entwicklungen im Data Mining
- Privacy-preserving Data Mining

© Fraunhofer Institut für intelligente Analyse-
und Informationssysteme IAIS

 **Fraunhofer**
IAIS

Fraunhofer-Gesellschaft



- Joseph von Fraunhofer, Physiker und Unternehmer
- Fraunhofer-Mission:
 - Innovation durch Spitzenforschung und Umsetzung im Markt für Kunden
 - Finanzierung: 33% Forschungsprojekte, 33% Wirtschaftserträge, 33% Grundfinanzierung
- 57 Institute, 40 Standorte, 15.000 Mitarbeiter, 1,2 Mrd. € Gesamthaushalt
- Highlights: z.B. MP3



© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS

 **Fraunhofer**
IAIS

Fraunhofer IAIS: Intelligente Analyse- und Informationssysteme



- 240 Mitarbeiter: Wissenschaftler, Projektengineure, Technisches Personal und Verwaltung
- Standort: Fraunhofer Campus Schloss Birlinghoven/Bonn
- Bereiche
 - Adaptive Reflective Teams (ART)
 - Knowledge Discovery (KD)
 - Medienproduktion (MP)
 - NetMedia (NM)
 - Virtual Environments (VE)
- Gemeinsame Forschungsgruppen und Kooperationen mit

 b-it

 universität**bonn**
RHEINISCHE-FREIWOHLFELT-UNIVERSITÄT

 Fachhochschule
Bonn-Rhein-Sieg

© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS

 **Fraunhofer**
IAIS

Ausgewählte Kunden & Partner



© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS



Was ist Data Mining?

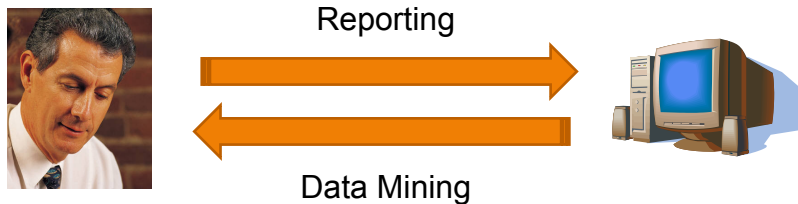


© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS



Was ist Data Mining?

- *“Knowledge Discovery is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.”* Fayyad et al. (1996)
- Beispiel: *Kunden über 55 Jahre aus ländlichen Gebieten haben eine um 35% niedrigere Kündigungswahrscheinlichkeit*

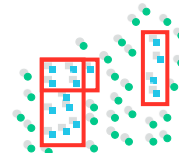
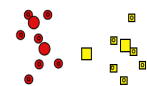
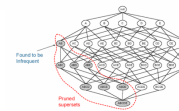
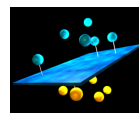
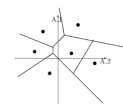
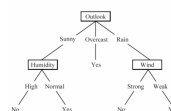


© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS

Fraunhofer IAIS

Data Mining Verfahren

- Entscheidungsbäume
- Nächste Nachbarn (kNN)
- Support-Vector-Machines
- Assoziationsregeln
- Subgruppensuche
- Clustering
- Graph Mining
- ...alle frei verfügbar in Data Mining Toolkits



sas

SPSS AN IBM COMPANY

WEKA The University of Waikato

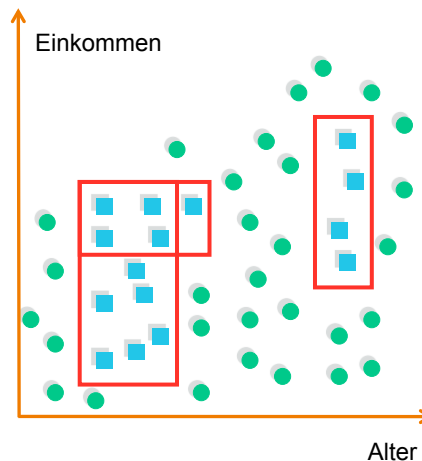
rapid-i REPORT THE FUTURE

© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS

Fraunhofer IAIS

Verfahren: Klassifikation

- Aufgabe
 - Gegeben Beispiele welche Objekte in welche Klassen fallen
 - Sage die Klassen für neue Objekte vorher
- Einsatzgebiete zum Beispiel
 - Neukundengewinnung
 - Kündigunganalyse
 - Betrugserkennung



Verfahren: Clustering

- Aufgabe
 - Gegeben eine Menge von Objekten
 - Finde Untergruppen in den Daten
- Beispiele
 - Kunden-Profilung
 - Werbeplatzierung



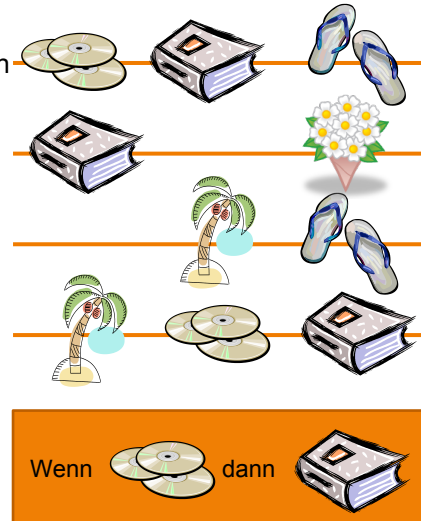
Typische Data-Mining-Verfahren: Assoziationsregeln

Aufgabe

- Gegeben eine Menge von Transaktionen von Objekten
- Finde zusammengehörige Objekte

Beispiele

- Cross-Selling
- Empfehlungssysteme



Was ist Data Mining?

Data Mining ist:



Data Mining ist nicht:



Auf einem Schiff befinden sich 26 Schafe und 10 Ziegen. Wie alt ist der Kapitän? Baruk (1989)

Was ist Data Mining?

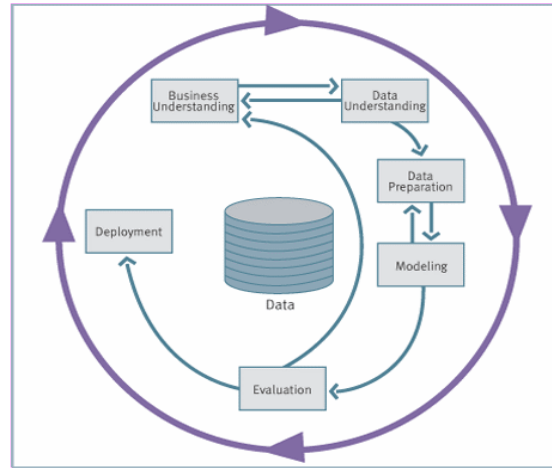
- „Auf einem Schiff befinden sich 26 Schafe und 10 Ziegen. Wie alt ist der Kapitän?“
- „Auf einem Schiff befinden sich 26 Schafe und 10 Ziegen. Der Kapitän heißt Wilhelm. Wie alt ist der Kapitän?“
 - → **Wahrscheinlich schon im Rentenalter**
- „Auf einem Schiff befinden sich 26 Schafe und 10 Ziegen. Der Kapitän heißt Kevin. Wie alt ist der Kapitän?“
 - → **Wahrscheinlich um die 20**

Typische Muster

- Sozio-demographische Daten
 - „gleich und gleich gesellt sich gern“
- Collaborative Filtering
 - „Kunden die dieses Buch gekauft haben, haben auch die folgenden Bücher gekauft...“
- ...

CRISP-DM Prozess

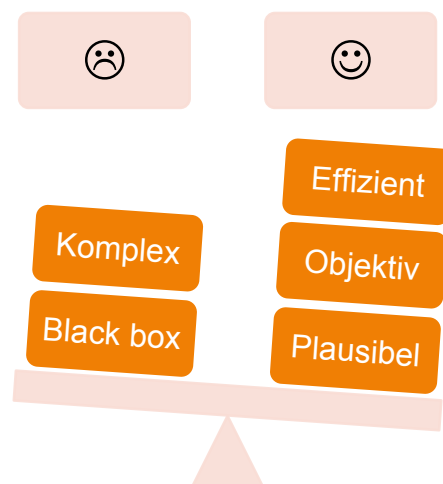
- Cross-Industry Standard Process for Data Mining
- Phasen
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment



© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS

Fraunhofer IAIS

Privacy-relevante Aspekte von Data Mining



© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS

Fraunhofer IAIS

Aktuelle Entwicklungen im Data Mining

Unstrukturierte Daten verstehen

- Texte
- Blogs
- Videos

YouTube Broadcast Yourself™ Startseite Videos Kanäle Suchen

How To Unlock a Kryptonite Lock With a Bic Pen

0.05 / 2.28 360p 38.094 Aufrufe

Erfahrungsbericht

“ Das Mieseste, was ich an Service erlebt habe ”

★☆☆☆☆ 24. Mar 2005

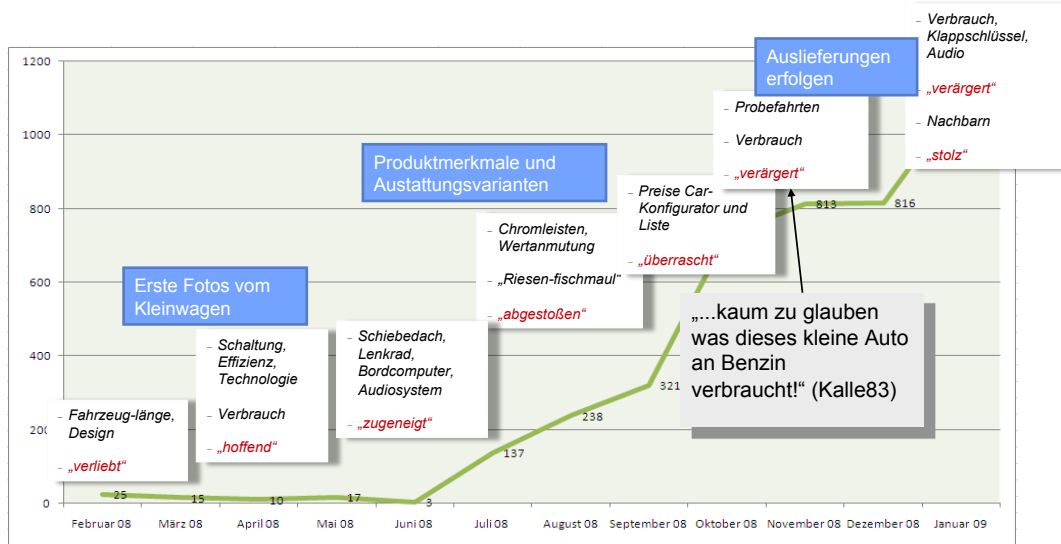
<p>Pro: Das Pro suche ich noch immer</p> <p>Kontra: So ziemlich alles</p> <p>Empfehlenswert: Nein</p>	<p>Details: Kundenservice: Erreichbarkeit: Kompetenz: mehr</p>
--	--

Dieser Erfahrungsbericht wurde von 28 Ciao Mitgliedern durchschnittlich als **sehr hilfreich** bewertet

Folgender Sachverhalt:
Mein Fernsehgerät der Firma Neckermann hat's innerhalb der Garantiezeit geschafft, den Geist an

Beispiel Marktforschung – PKW-Produkteinführung

Themen und Emotionen mittels Smart Semantics aus Online-Foren extrahiert



* anonymisiert

© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS



Data.gov v / data.gov v.uk

DISCOVER. PARTICIPATE. ENGAGE.

FEATURED DATASET: GEOGRAPHIC DISTRIBUTION OF VA EXPENDITURES (GDG)

The Geographic Distribution of VA Expenditures (GDG) is an annual report that shows estimated VA expenditures for major programmatic areas by geographic area (state, county, and congressional district). The major programmatic areas are: Compensation and Pensions, Readjustment, Education and Vocational Rehabilitation, Insurance, Construction, and Medical and Administrative.

Open Government Directive Mandates Open Data

The Open Government Directive, released on December 8, 2009, instructs all federal agencies to provide high-value information to the public online in open, accessible, machine-readable formats. Read More...

Join the Dialogue

When Data.gov launched last May, we promised to evolve this initiative with the public. We are taking that step.

10/26/2009 5 **Val-O-Mo Farm, Inc., Elmwood, WI 54740**

10/26/2009 Worker was operating a skid-steer cleaning out a dairy cattle barn near an outdoor manure slurry pit. The skid-steer and the worker fell off the end of the push-off platform into the manure slurry pit, trapping the worker in the vehicle. Worker died of suffocation due to inhalation of manure.

work zone and was struck by a motor vehicle.

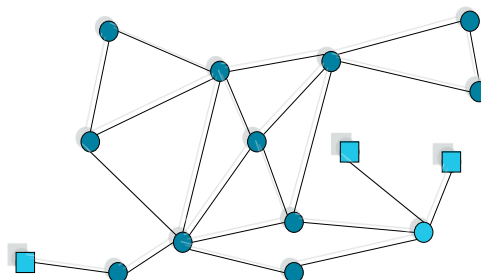
009 Worker was dismantling a 40-foot high scaffold and fell.

© Fraunhofer Institut für intelligente Analyse- und Informationssysteme IAIS

IAIS

Techn I gien - Graphmining

- Analyse komplexer Netzwerke
 - Welche Netzwerke sind typisch?
 - Wie hängen A und B zusammen?
 - Netzwerkmuster als Betrugsindikatoren
 - Wer ist innen, wer ist Außenseiter?
- Anwendungen
 - Webseiten / User-Tracking
 - Soziale Netzwerke
 - Unternehmens-Abhängigkeiten



Privacy-preserving Data Mining

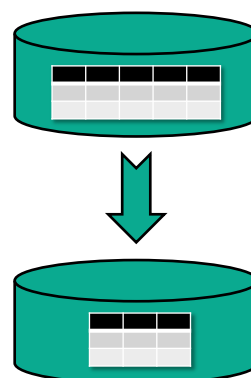
Privacy-preserving Data Mining

- Vereinbart Data Mining und Datenschutz
 - Rechtlichen Fragen
 - Öffentliche Meinung
- Privacy by Design
 - Entwicklung von Datenschutz-kompatiblen Analyseansätzen
 - Soviel Anonymität wie nötig, soviel Wissen wie möglich



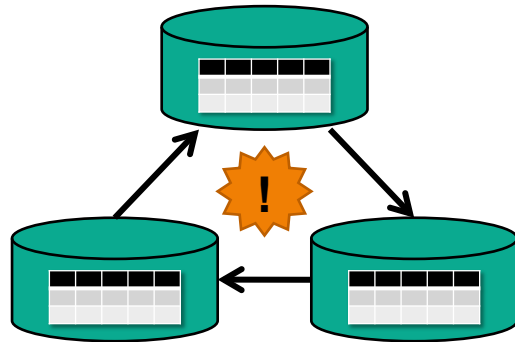
Privacy-preservati n durch Vergrößerung v n Daten

- Wie kann man Daten herausgeben ohne personenbezogene Informationen zu veröffentlichen?
 - Geschlecht, Geburtstag und Postleitzahl identifizieren 87% der US-Bevölkerung eindeutige
- K-Anonymität: jedes Muster passt auf mindestens k Personen
 - Auf Daten
 - Im Algorithmus

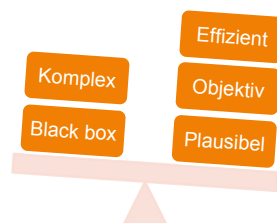


Privacy-preservativ durch kryptographische Verfahren

- Integration kryptographischer Ansätze in Data-Mining-Verfahren
 - Gesamtergebnis wird berechnet, ohne dass beteiligte Parteien ihre Daten aufdecken müssen
- Vorteil
 - Garantierter Datenschutz
 - Kein Informationsverlust



Fazit



Kontakt

Fraunhofer Institut IAIS

Schloss Birlinghoven

53754 St. Augustin

www.iais.fraunhofer.de

Dr. Stefan Rüping

Tel: +49 (0) 2241 14 2072

Mail: stefan.rueping@iais.fraunhofer.de

