



IBM Performance 2011

Mehrwert und konkreter Nutzen durch
Data Mining

+ Advanced Analytics mit IBM SPSS

Fachhochschulstudiengänge Burgenland

Dr. Maximilian Kobler

Mag. Walter Strasser

Vorstellung & Kontaktdaten



Dr. Maximilian Kobler

- E-Mail: maximilian.kobler@fh-burgenland.at
- Telefon: + 43 590 10 60 136



Fachhochschul
Studiengänge



Burgenland

Bildung im
Herzen Europas.

Mag. Walter Strasser

- E-Mail: w.strasser@bmd.at



Inhaltsübersicht

*“The most exciting phrase to hear in science,
the one that heralds new discoveries, is not “Eureka!” (I found it)
but ‘that’s funny...’”*

Isaac Asimov

- Definition Data Mining
- Data Mining in Forschung & Lehre
- Anwendungsbeispiel Gebäudetechnik

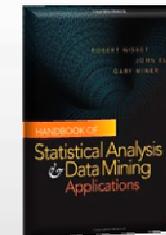
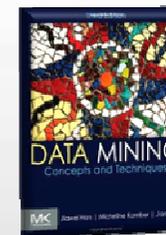
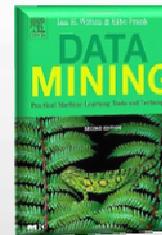
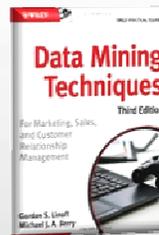
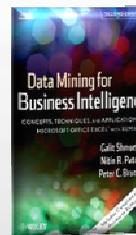
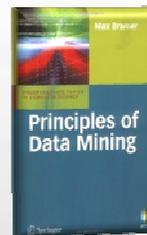
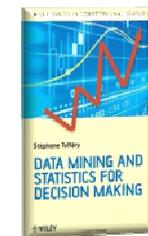
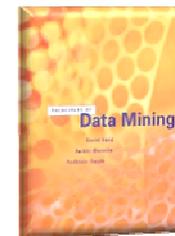
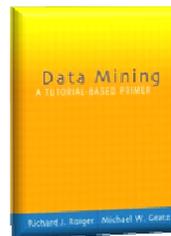
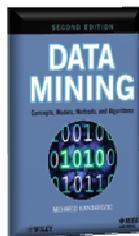
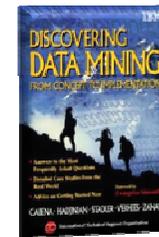
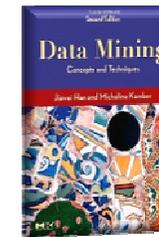


Data Mining



Versuch einer Definition

- “Das systematische Entdecken unbekannter wertvoller Informationen aus großen Datenmengen mit komplexen Methoden.”



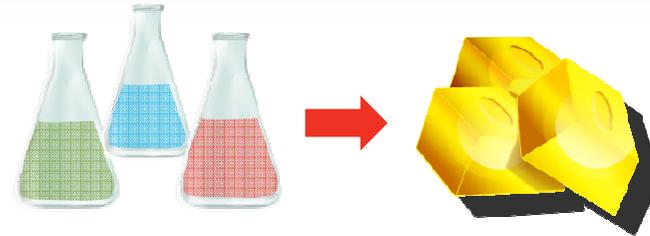
Data Mining Prozess



Alchemie

- Transmutation

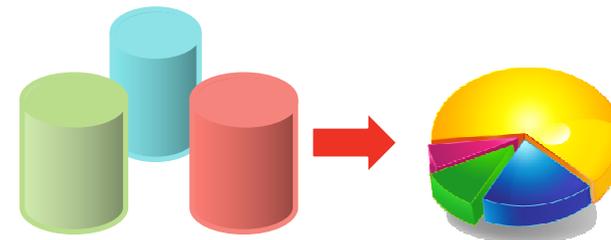
- Ziel: Stoffe auf chemischem Weg in Gold umzuwandeln
- Methode: Trial and Error



Data Mining

- Wissensexploration

- Ziel
 - Große Datenmengen automatisiert in Wissen umzuwandeln
 - Identifikation gültiger, vorher unbekannter, nützlicher Muster
 - Datenselektion, Datenzugriff, Datenaufbereitung, Transformation, Modellbildung, Modellanwendung, Evaluation
- Methode: Klassifikation, Schätzung, Vorhersage, Clusterbildung, Assoziation, Künstliche Intelligenz, (und Trial and Error)

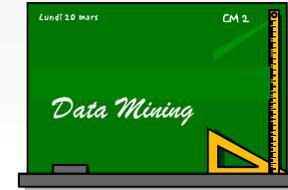


That's funny...



- Assoziation
 - Gleichzeitig auftretende Variablen
 - Warenkorbanalyse: Windeln und Bier werden zusammen gekauft
 - Bildanalyse: Panzer treten nur bei schönem Wetter auf
- Prädikation
 - Zeitverschobene Variablen (Time Warping)
 - Ereigniszusammenhänge: Nach Stromausfall steigt die Geburtenrate
- Klassifikation/Clustering
 - Kategorale Zielvariable
 - Marktsegmentierung: Kundengruppen
- Regression
 - Numerische Zielvariablen
 - Verkaufszahlen

Data Mining in Forschung & Lehre



Forschung

- Empirie, Explorative Datenanalyse, Hypothesengenerierung
- Bachelor, Diplomarbeiten
 - Vorhersage Modelle, Fragebogen Auswertung, Ursache-Wirkungsanalysen
- Forschungsprojekte
 - z.B. MAS Gebäudetechnik und -management

Lehre

- BAC Information, Medien & Kommunikation
 - Datenbanken, Visuelle Modellierung quantitativer Informationen, Marktanalysen, Datawarehouse & Data Mining, Information Retrieval
- MAS Business Process Management & Engineering
 - Datenmodellierung & Datenbanken, Technologiebasierte Managementkonzepte

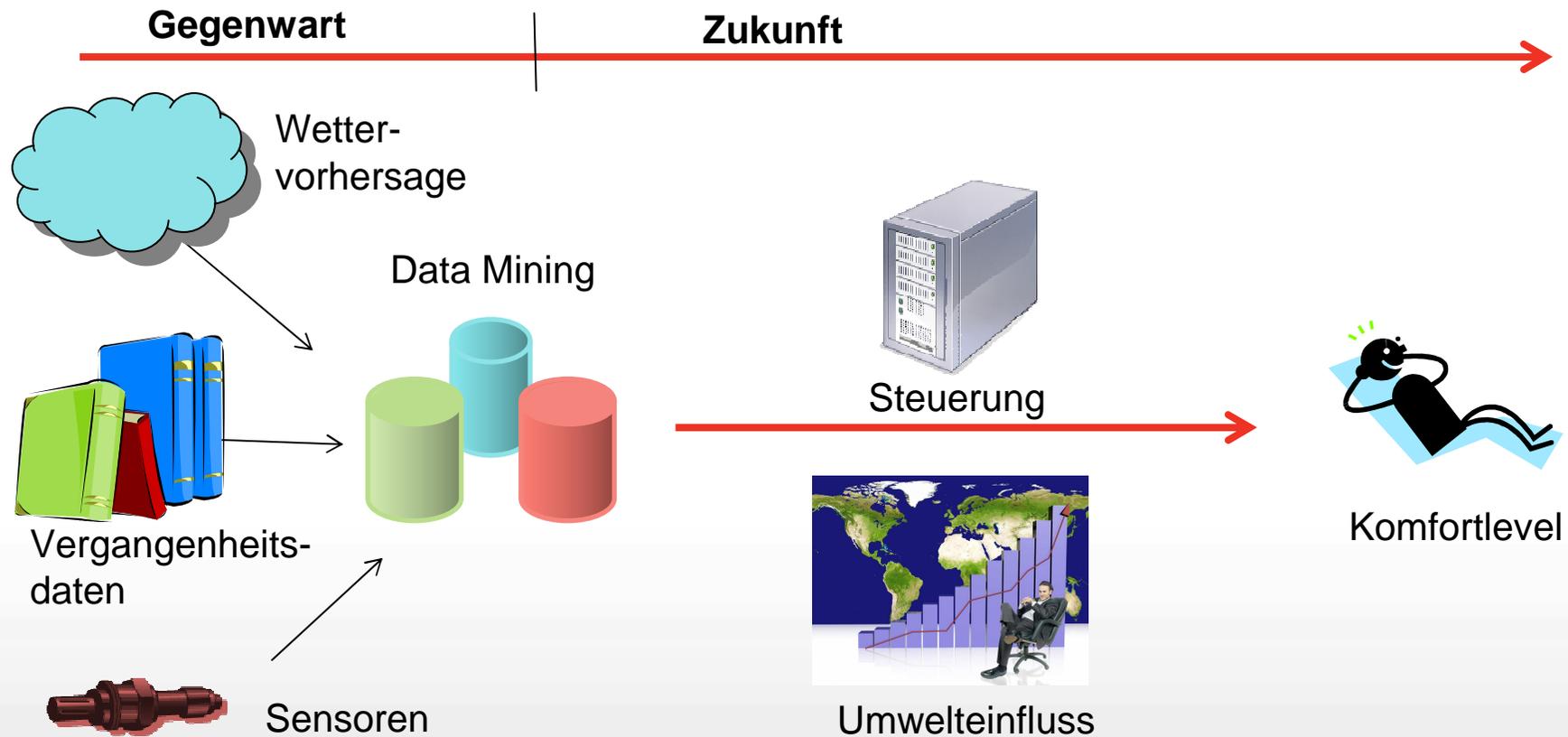


Anwendungsbeispiel Gebäudetechnik

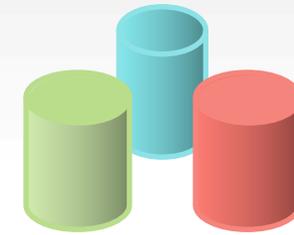


- Ziel: Komfortlevel abhängige prädikative Regelung
- Problem:
 - Zahlreiche Daten
 - Gebäudesensoren, Umweltdaten, zahlreiche Kennzahlen wie Durchlässigkeitsgrad, internen Wärmegewinn, Wärmekapazität, Daten
 - Unterschiedliche Meßniveaus und Zeitintervalle
 - Individuelle Präferenzen
 - Komfortlevel
 - Komplexe Steuerung (Heating, Ventilating, Air Conditioning)
 - Vielfältige (teils unbekannte bzw. schwer formalisierbare) Abhängigkeiten

Prädikative Regelung

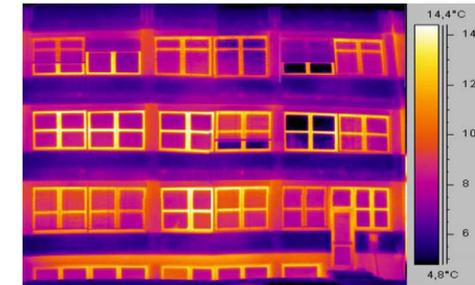


Knowledge Discovery (1)



■ Datenauswahl (Selection)

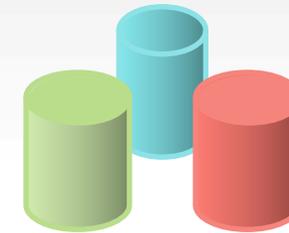
- Sensordaten (Temperatur, CO₂, Luftfeuchte, ...)
- Öffentliche Datenbanken (Feiertage, Klimadaten,...)
- Kommerzielle Datenbanken (Standortdaten, ..)



■ Daten

- Unterschiedliche Zeitauflösung (Sekunde, Stunde, Tag, Monat)
- Unterschiedliche Einheiten (z.B. metrisch)
- Unterschiedlichen Bezugspunkte (z.B. Meereshöhe)
- Ortsbezogene Informationen (z. B. Thermografische Daten)
- ...

Knowledge Discovery (2)



■ Datenbereinigung (Preprocessing)

Standortdaten

TRYWIEN DATEN ABGELEITET AUS DEM TRY ALPENVORLAND
 STATION: WIENHOHE WARTE STAT-NR.9501
 LAGE:48.10 GRAD N.<- B. 11.80 GRAD O.<-L. 200 METER ÜBER NN

Tag (fortlaufend)	Stunde	Lufttemperatur [°C]	rel. Feuchte [%]	Sonnenhöhe [°]	Azimuth [°]	Direkte Strahlung [W/m²]	Diffuse Himmelsstrahlung [W/m²]	Windgeschwindigkeit [m/s]	Windrichtung [°]	Bedeckung [%]
1	1	-0,5	80,0	0,00	0,00	0,0	0,0	0,0	0	63,0
1	2	-0,6	84,0	0,00	0,00	0,0	0,0	0,0	0	100,0
1	3	-0,5	89,0	0,00	0,00	0,0	0,0	0,0	0	100,0
1	4	-0,4	93,0	0,00	0,00	0,0	0,0	0,0	0	100,0

Klimadatenkatalog

Ortsbeschreibung					geogr. Breite		geogr. Länge		Globalstrahlung bei wirklicher Seehöhe												Diffu		
Planfeld Nr.	Seehöhe in m	Bundesland	Postleitzahl	Ortsname	°	dez.	°	dez.	Jan	Feb	Mrz	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Jan	Feb	
228	567	Sbg.	5342	Abersee	47	44	47,7333333	13 25	13,41666667	1032	1774	2813	3839	4635	5038	4994	4263	3165	2006	1150	805	719	1174
140	982	Tirol	9913	Abfaltersbach	46	45	46,75	12 32	12,53333333	1353	2172	3312	4269	4862	5222	5323	4782	3619	2337	1461	1125	886,6	1357
587	180	N.O.	3462	Absdorf	48	24	48,4	15 59	15,98333333	831	1499	2582	3826	4847	5355	5258	4496	3205	1845	915	611	597	1023
226	715	Sbg.	5441	Abtenau	47	34	47,5666667	13 21	13,35	1043	1790	2833	3857	4645	5045	4997	4276	3179	2020	1163	815,7	725,2	1182
162	380	O.O.	5122	Ach	48	9	48,15	12 50	12,83333333	929,4	1637	2719	3851	4721	5145	5061	4367	3185	1927	1044	720,4	658,4	1100

Sensordaten

Minute des Tages	Upv-Ist V	Upv-Soll V	Zac Ohm	E-Total kWh	h-Total h	Netz-Ein	Fehler	PV-Anlage Solarspannung V	PV-Anlage Sollwert-Upv V	PV-Anlage Netzstrom mA	PV-Anlage P-total W	PV-Anlage E-tota kWh
874,00	234	235	0,75	1451,5	5643,4	773	-----	234	235	886	202	1
875,00	234	235	0,75	1451,5	5643,4	773	-----	234	235	886	202	1
876,00	233	234	0,78	1451,5	5643,5	773	-----	233	234	886	202	1
877,00	233	234	0,77	1451,5	5643,5	773	-----	233	234	958	219	1
878,00	239	239	0,77	1451,5	5643,5	773	-----	239	239	1334	305	1
879,00	237	239	0,78	1451,5	5643,5	773	-----	237	239	1752	399	1
880,00	240	241	0,77	1451,5	5643,5	773	-----	240	241	2094	481	1
881,00	235	239	0,8	1451,5	5643,5	773	-----	235	239	1304	298	1
882,00	237	237	0,76	1451,5	5643,6	773	-----	237	237	1032	236	1
883,00	238	236	0,82	1451,5	5643,6	773	-----	238	236	2788	644	1
884,00	237	237	0,79	1451,6	5643,6	773	-----	237	237	2678	618	1
885,00	237	237	0,82	1451,6	5643,6	773	-----	237	237	2692	606	1

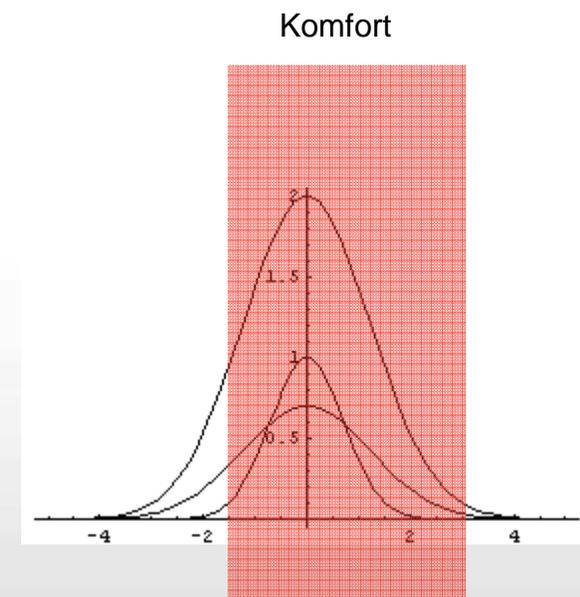
Knowledge Discovery (3)



- Datenumwandlung (Transformation)
- Operationalisieren, z.B. Komfortlevel



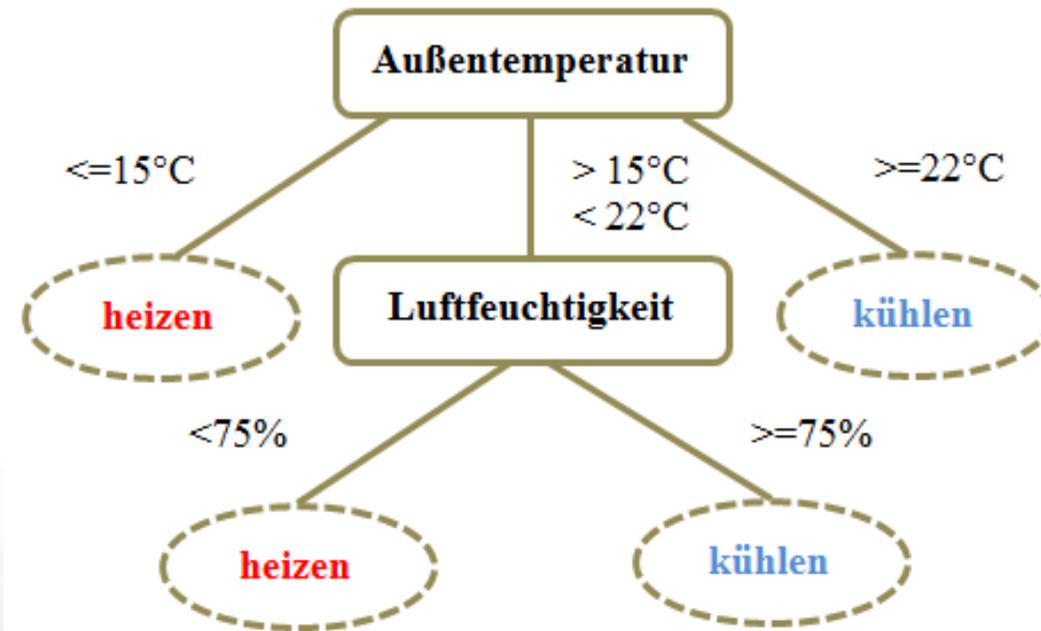
Parameter	Value Range
Air Velocity	0 - 0.2 m/s
Air Temperature	20 – 28°
Clothing	1 light, 2 heavy
Metabolic Rate	1 low, 2 high
Radiant Temperature	20 – 28°
Relative Humidity	35 - 55 %



Knowledge Discovery (4)



- Mustererkennung (Data Mining)
 - Clustering
 - Decision Tree
- Evaluation & Anwendung



Beispiel eines einfachen Entscheidungsbaumes

SPSS Beispiel



***** Analysis of Variance *****

28 cases accepted.
 0 cases rejected because of out-of-range factor values.
 0 cases rejected because of missing data.
 4 non-empty cells.
 1 design will be processed.

 ***** Analysis of Variance ----- design 1 *****
 Tests of Significance for TESTWERT using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	46.29	24	1.93	22.78	.000
GRUPPE	131.82	3	43.94	22.78	.000
(Model)	131.82	3	43.94		
(Total)	178.11	27	6.60		

R-Squared = .740
 Adjusted R-Squared = .708

 Estimates for TESTWERT
 --- Individual univariate .9500 confidence intervals

GRUPPE	Parameter	Coeff.	Std. Err.	t-Value	Sig.	t Lower -95%	CL- Upper
2	2.39285714	.45457	5.26402	.00002	1.45467	3.33104	
3	1.67857143	.45457	3.69267	.00114	.74039	2.61676	
4	-1.0357143	.45457	-2.27846	.03189	-1.97390	-.09753	



- A Ahmed, N E Korres, J Ploennigs, H Elhadi, K Menzel (2011): Mining building performance data for energy-efficient operation. *Advanced Engineering Informatics* 25 (2)
- A Ahmed, M Otreba, N E Korres, H Elhadi, K Menzel (2011): Assessing the performance of naturally day-lit buildings using data mining. *Advanced Engineering Informatics* 25 (2)
- H Doukas, C Nychtis, J Psarras (2008): Assessing energy-saving measures in buildings through an intelligent decision support model. *Building and Environment - Volume 44, Issue 2*
- M J Gacto, R Alcalá, F Herrera (2010): A multi-objective evolutionary algorithm for an effective tuning of fuzzy logic controllers in heating, ventilating and air conditioning systems. *Applied Intelligence*
- Yang Gao, E Tumwesigye, B Cahill, K Menzel (2010): Using Data Mining in Optimisation of Building Energy Consumption and Thermal Comfort Management. *2nd International Conference on Software Engineering and Data Mining SEDM*
- D Holcomb, W Li, S A Seshia (2009): Algorithms for Green Buildings: Learning-Based Techniques for Energy Prediction and Fault Diagnosis. *Technical Report - Electrical Engineering and Computer Sciences University of California at Berkeley*
- A Sureka, K V Indukuri (2008): Using Genetic Algorithms for Parameter Optimization in Building Predictive Data Mining Models. *Lecture Notes in Computer Science - Advanced Data Mining and Applications*



**Danke für Ihre
Aufmerksamkeit!**

- Simulationsergebnisse vs. Messdaten

