

Data-Mining

Datenbankschürfung, Datenschürfung,

KDD – Knowledge Discovery in
Databases

Wissensentdeckung in Datenbanken

Inhalt

- Allgemein
 - Typische Aufgaben
 - Wissenentdeckungsprozess
 - Beispiel Klassifikationsproblem
 - Nächste Nachbarn Verfahren
 - Entscheidungsbaumverfahren
 - Visuelles Data-Mining
-

Allgemein

- Es geht darum in großen Datenmengen allgemeine Gesetzmäßigkeiten zu extrahieren, von dessen Existenz man nichts weiß und welcher Art diese sein könnten.
-

Typische Aufgaben 1

Klassifikation:

Ist der Kunde kreditwürdig oder nicht?

Prognose:

Wie wird sich der Dollarkurs entwickeln?

Konzeptbeschreibung:

Welche Eigenschaften haben defekte Fahrzeuge?

Typische Aufgaben 2

Segmentierung:

Wie lassen sich meine Kunden in aussagekräftiger Weise in Untergruppen einteilen?

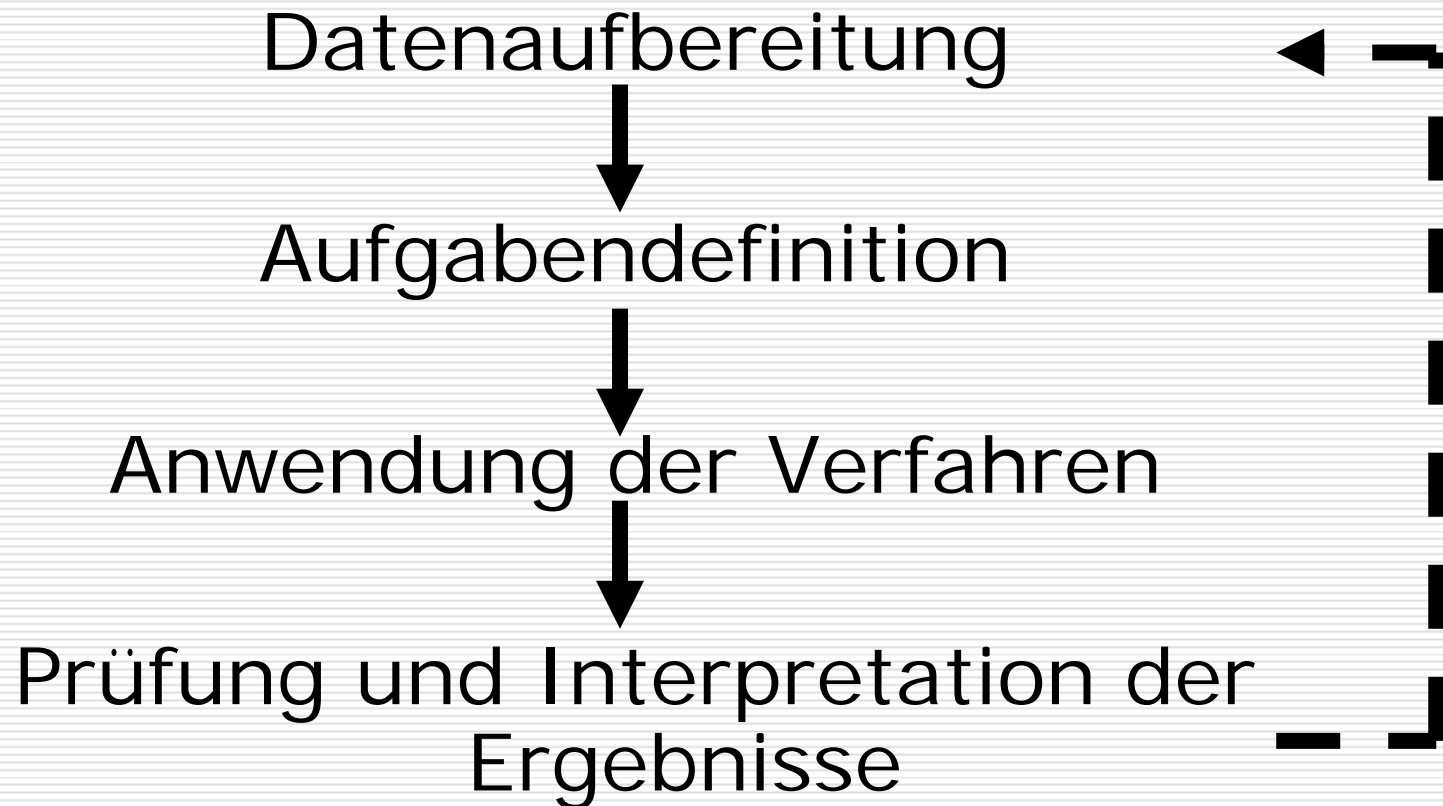
Abweichungsanalyse:

Gibt es Umsatzschwankungen?

Abhängigkeitsanalyse:

Welche Produkte werden häufig zusammen gekauft?

Wissensentdeckungsprozess



Datenaufbereitung

- Umfang ca. 60 – 90% des Gesamtaufwandes
 - Auswahl aussichtsreicher Datensätze
 - Vereinheitlichung des Datenformates
 - Fehler und Ausreißer entfernen
 - Datenmenge reduzieren (opt.)
-

Beispiel Klassifikationsproblem 1

□ Beispiel:

- Betrugsbekämpfung eines Kreditkartenunternehmens
- Datensätze von Zahlungsvorgängen und die Ergebnisse der Überprüfung

□ Aufgabe:

Entscheidung ob eine Detailprüfung nötig ist oder nicht.

Beispiel Klassifikationsproblem 3

Nr.	Zahlungsart	Kartentyp	Betrag	Klasse
1	offline	standard	700	Betrug
2	online	premium	800	OK
3	offline	premium	700	Betrug
4	online	temporär	500	OK
5	offline	premium	800	OK
6	offline	premium	600	Betrug
7	offline	premium	1100	OK
8	offline	premium	900	OK
9	offline	temporär	900	Betrug
10	offline	temporär	800	Betrug
11	online	standard	700	OK
12	online	premium	1000	Betrug
13	offline	standard	600	Betrug
14	online	temporär	700	OK
15	offline	standard	300	OK
16	offline	standard	400	Betrug

Fall:

online

premium

1020

Beispiel Klassifikationsproblem 2

Nr.	Zahlungsart	Kartentyp	Betrag	Klasse
1	offline	standard	700	Betrug
2	online	premium	800	OK
3	offline	premium	700	Betrug
4	online	temporär	500	OK
5	offline	premium	800	OK
6	offline	premium	600	Betrug
7	offline	premium	1100	OK
8	offline	premium	900	OK
9	offline	temporär	900	Betrug
10	offline	temporär	800	Betrug
11	online	standard	700	OK
12	online	premium	1000	Betrug
13	offline	standard	600	Betrug
14	online	temporär	700	OK
15	offline	standard	300	OK
16	offline	standard	400	Betrug

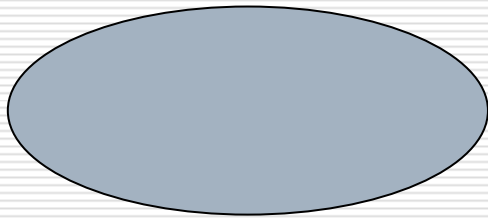
Fall:
online
premium
1020

Nächste Nachbarn Verfahren

Nr.	Zahlungsart	Kartentyp	Betrag	Klasse
1	offline	standard	700	Betrug
2	online	premium	800	OK
3	offline	premium	700	Betrug
4	online	temporär	500	OK

Fall: offline, premium, 500

Entscheidungsbaumverfahren



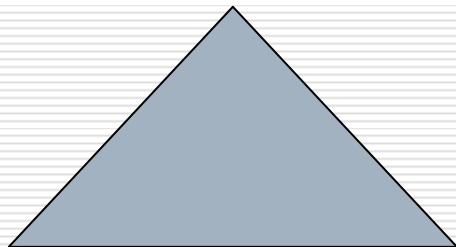
Frage

- Astgabel



Mögliche Antworten

- Zweige



Entscheidung

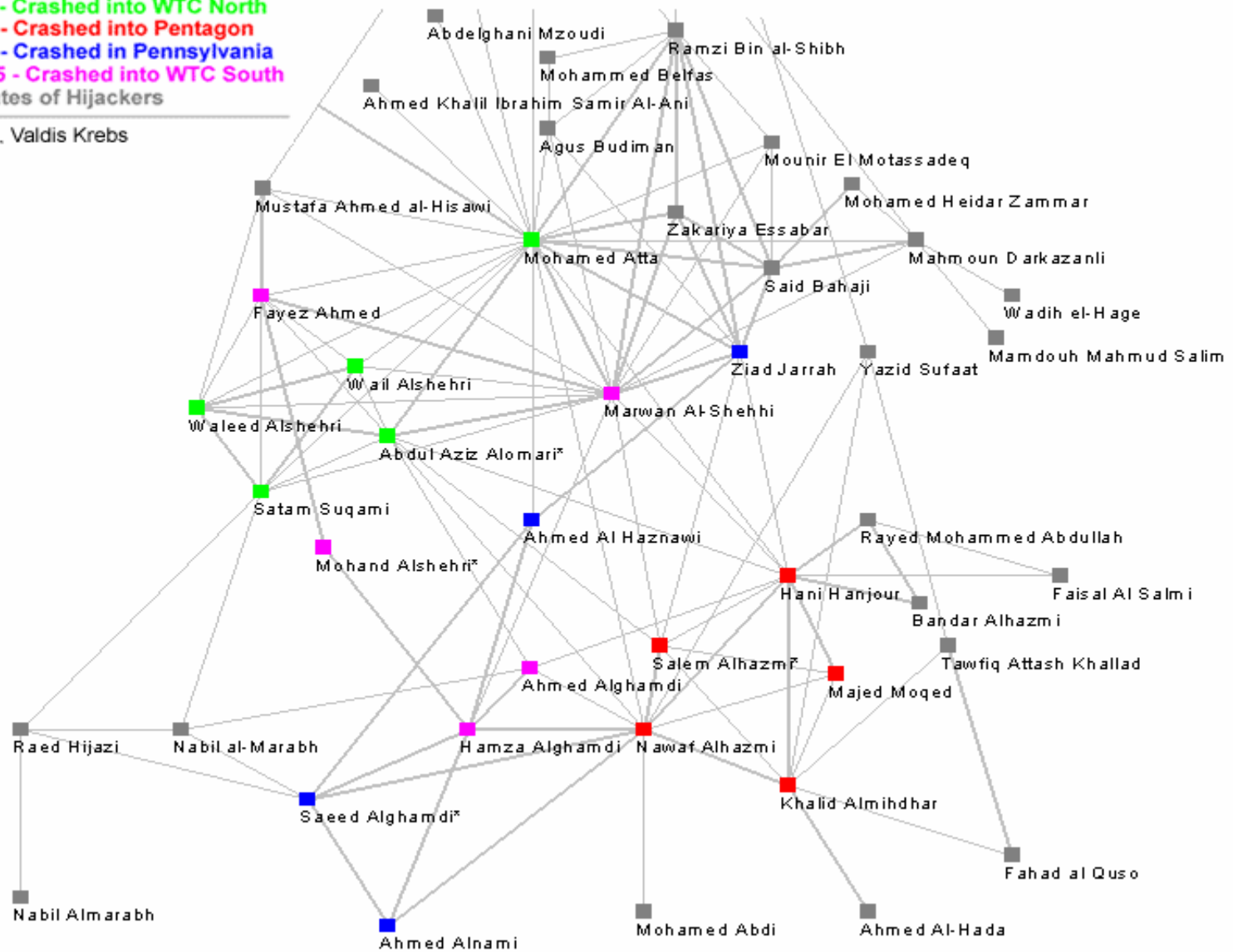
- Blätter

Entscheidungsbaumverfahren 2

Nr.	Zahlungsart	Kartentyp	Betrag	Klasse
1	offline	standard	700	Betrug
2	online	premium	800	OK
3	offline	premium	700	Betrug
4	online	temperorär	500	OK
5	offline	premium	800	OK
6	offline	premium	600	Betrug
7	offline	premium	1100	OK
8	offline	premium	900	OK
9	offline	temperorär	900	Betrug
10	offline	temperorär	800	Betrug
11	online	standard	700	OK
12	online	premium	1000	Betrug
13	offline	standard	600	Betrug
14	online	temperorär	700	OK
15	offline	standard	300	OK
16	offline	standard	400	Betrug

- Flight AA #11 - Crashed into WTC North
- Flight AA #77 - Crashed into Pentagon
- Flight UA #93 - Crashed in Pennsylvania
- Flight UA #175 - Crashed into WTC South
- Other Associates of Hijackers

Copyright © 2002, Valdis Krebs



MAP of the MARKET

Map Your Portfolio Mutual Fund Map

SmartMoney.com

Controls

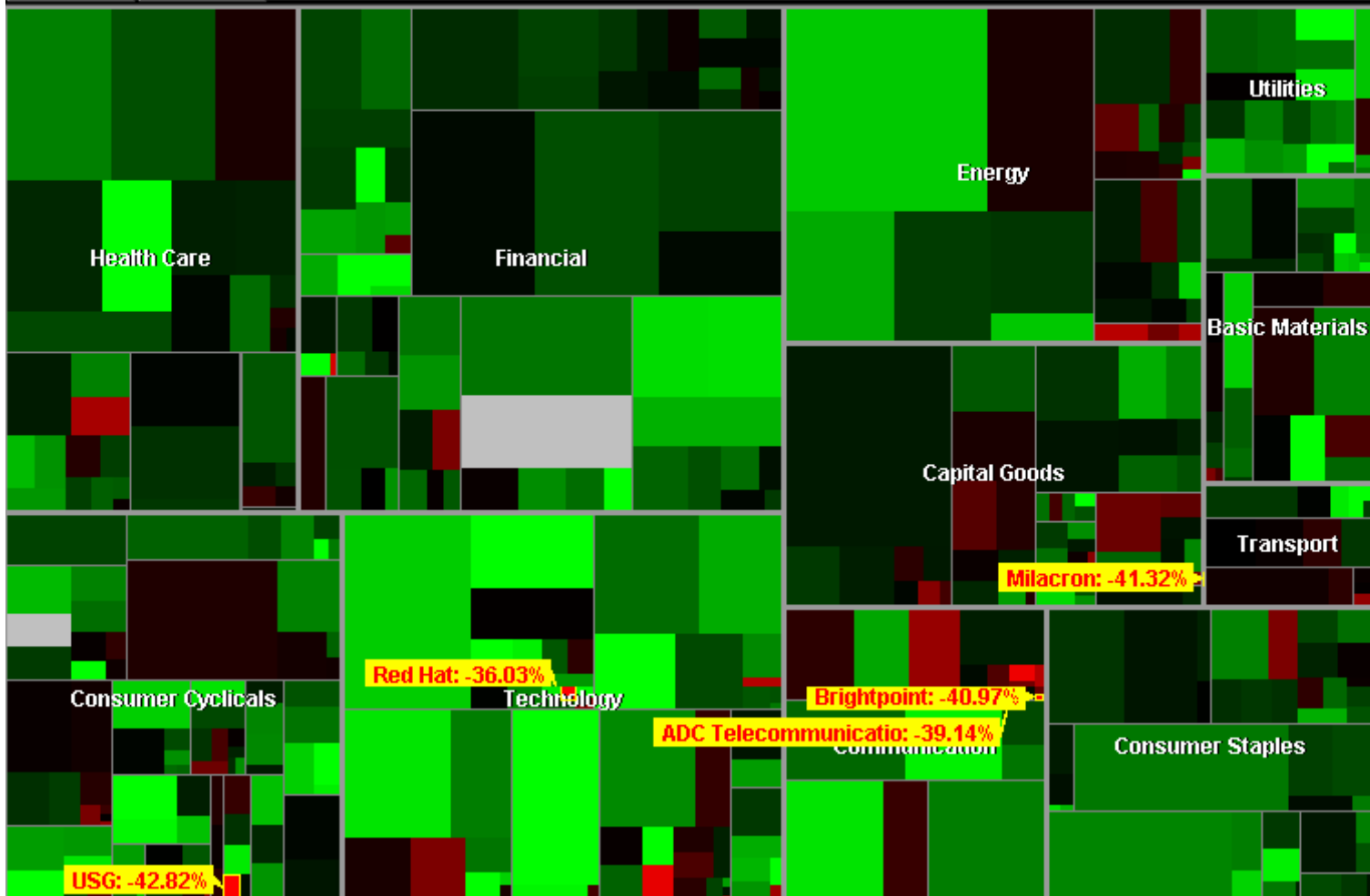
Instructions

No Headlines

DJIA 12193.01 +914.40 +8.11%

Nasdaq 2424.07 +213.70 +9.67%

10:42 am Nov. 30



Legend

Map Control Panel

Color key (% change)

-40% -20% +0% +20% +40%

News

No Headlines

Show change since

Close 26 Weeks

52 Weeks YTD

Highlight Top 5

Gainers Losers

No highlights

Find (name or ticker)

Color scheme

red/green

blue/yellow

Java Applet Window

Adblock

Quellen

- ❑ <http://de.wikipedia.org/wiki/Data-Mining>
 - ❑ Spektrum der Wissenschaft November 2002
 - ❑ <http://www.orgnet.com/hijackers.html>
 - ❑ <http://www.smartmoney.com/marketmap/>
 - ❑ <http://www.spektrum.de/artikel/804202>
-