

Entwicklung und Implementierung eines Prognose-Moduls zur Hochrechnung und Auswertung statistischer Daten

Prognose der GdW Jahresstatistik

# Datenbasis

## GdW

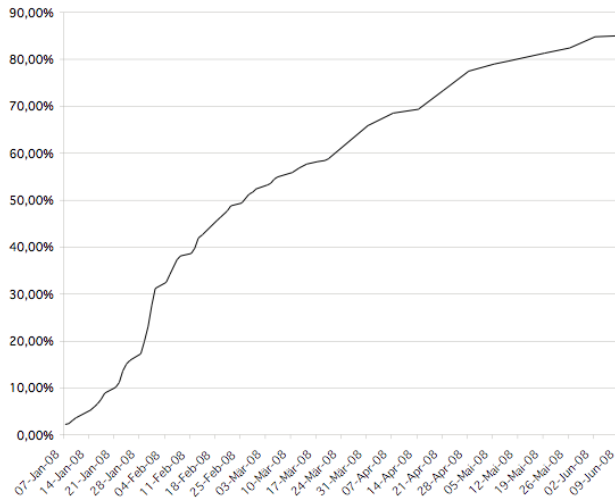
- Bundesverband deutscher Wohnungs- und Immobilienunternehmen
- Spitzenverband mit 14 Regionalverbänden
- 3000 Wohnungsunternehmen mit ca. 6,2 Mio Wohnungen



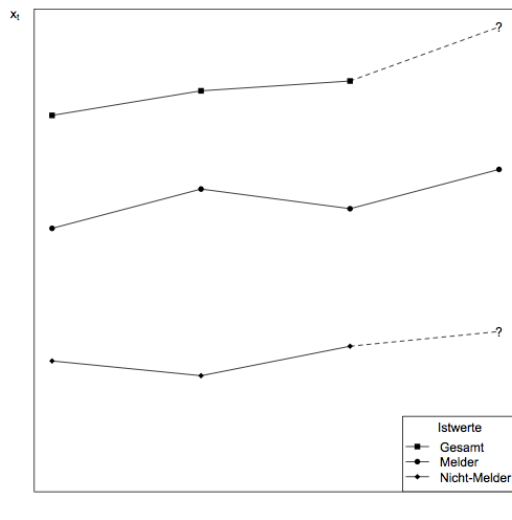
## GdW Jahresstatistik

- Die Wohnungsunternehmen als Grundgesamtheit
- ca. 240 Kennziffern, z.B.
  - Anzahl der Wohnungseinheiten
  - Nettokaltmiete
  - Bilanzsumme
  - ...

# Ziel der Diplomarbeit



- Rückläufe erstrecken sich fast über ein halbes Jahr
- Prognose aller Kennziffern zu einem früheren Zeitpunkt erwünscht



- Gesamtsumme einer Kennziffer Melder + progn. Nicht-Melder
  - Nicht nur Gesamtsumme interessant, sondern auch Prognoseergebnisse von Teilmengen und einzelnen Unternehmen der Grundgesamtheit
- **Bottom-Up-Prognose** für konsistente und transparente Ergebnisse

# Heterogenität der Datenbasis

## - Kennziffern

- Anz. der erhobenen Jahre
- unterschiedliche Interkorrelationen und Varianzen
- Typ (Boolean, Skala, Ganzzahlenwert)

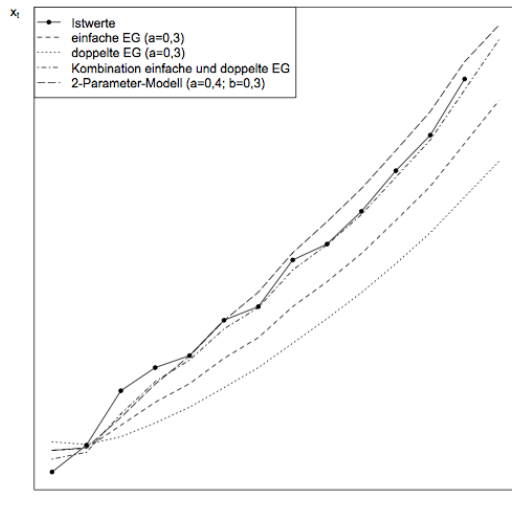
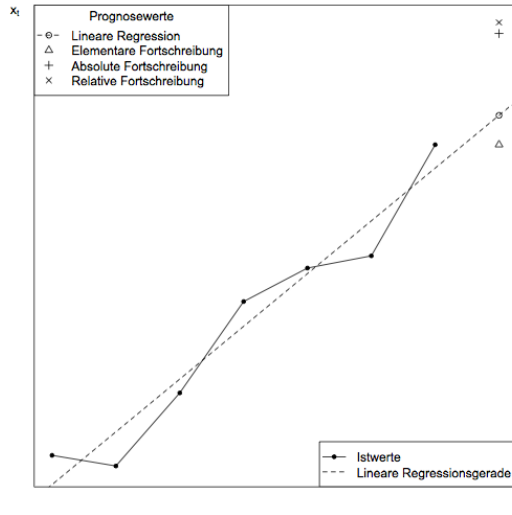
➤ kein Prognoseverfahren universell anwendbar

## - Mitgliedsunternehmen

- regionale Zuordnung (Verbandsbereiche, Ost-West, ...)
- Größe (Anz. der Wohnungseinheiten, Umsatzerlöse, ...)
- Rechtsform (Gesellschaft, eG, ...)

➤ Repräsentativität der Melder eingeschränkt

# Prognoseverfahren



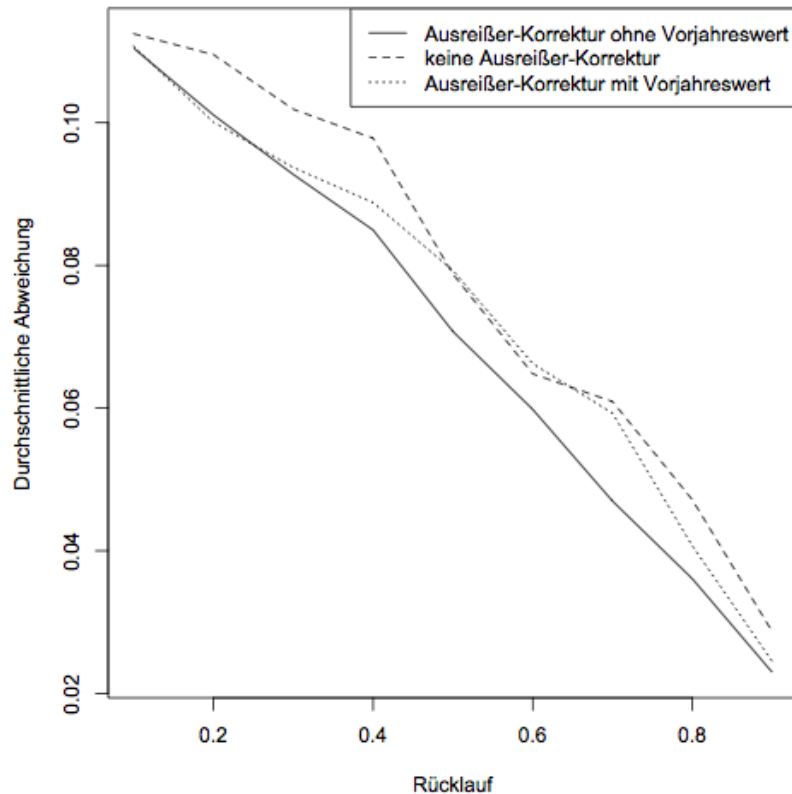
## Längsschnittanalyse

- Fortschreibung
- Lineare Regression
- Exponentielle Glättung
  - 2-Parameter-Modell
  - doppelte exp. Glättung
- ...

➤ vorteilhaft bei wenigen Meldern und mehreren Vergangenheitswerten (max. 5)

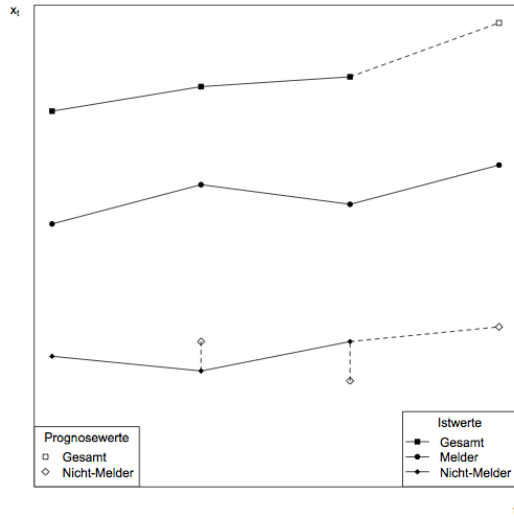
# Ausreißer

Möglichkeiten zur Erkennung:



- Standardabweichung
- Interquartilsabstand ist robuster gegen Ausreißer
- Wegen weniger Vergangenheitswerte (max. 5) nur dezent einsetzbar (Ausreißer wird in Richtung Median korrigiert)

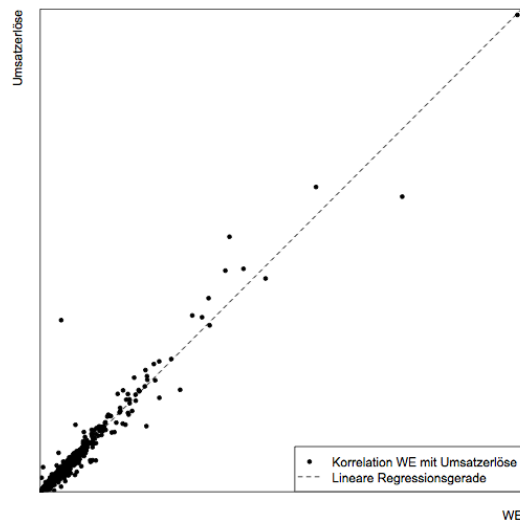
# Prognoseverfahren



## Querschnittsanalyse

- Melder als Indikator mit Trendfaktor
- Melder Durchschnittswert
- Weitere korrelierende Kennziffer als Indikator (bsp. Wohnungseinheiten)
- ...

➤ vorteilhaft bei vielen Meldern und guter Repräsentativität



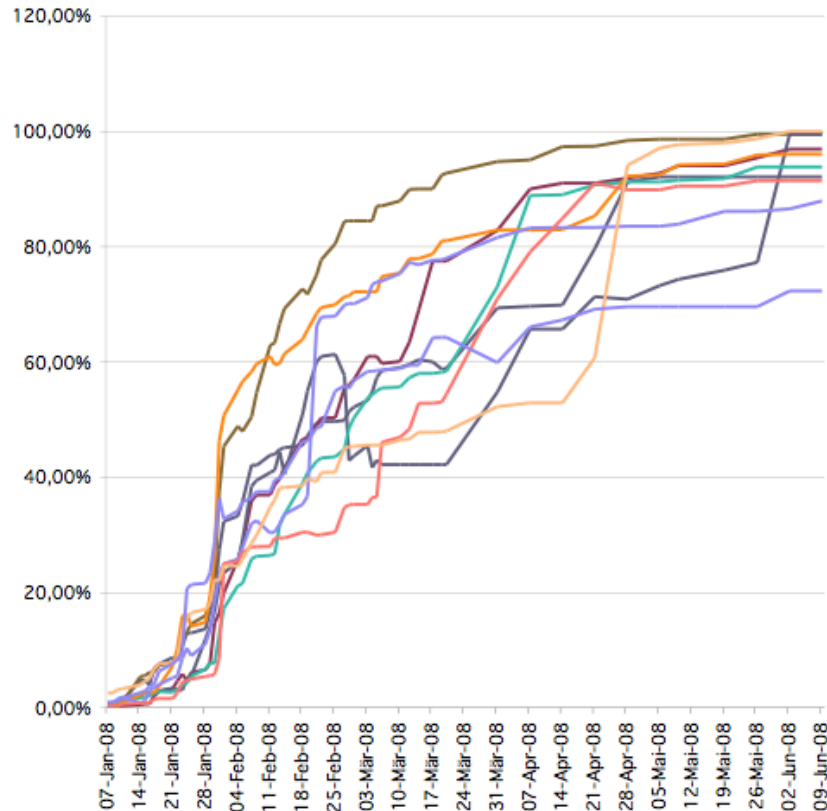
# Heterogenität der Datenbasis

- Kennziffern
  - Anz. der erhobenen Jahre
  - unterschiedliche Interkorrelationen und Varianzen
  - Typ (Boolean, Skala, Ganzzahlenwert)
  
- kein Prognoseverfahren universell anwendbar
  
  
- Mitgliedsunternehmen
  - regionale Zuordnung (Verbandsbereiche, Ost-West, ...)
  - Größe (Anz. der Wohnungseinheiten, Umsatzerlöse, ...)
  - Rechtsform (Gesellschaft, eG, ...)
  
- Repräsentativität der Melder eingeschränkt



# Repräsentativität

## Rückläufe der Verbände im Vergleich:



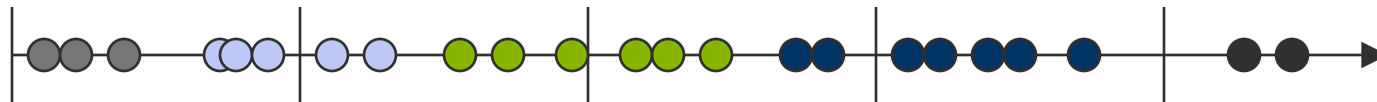
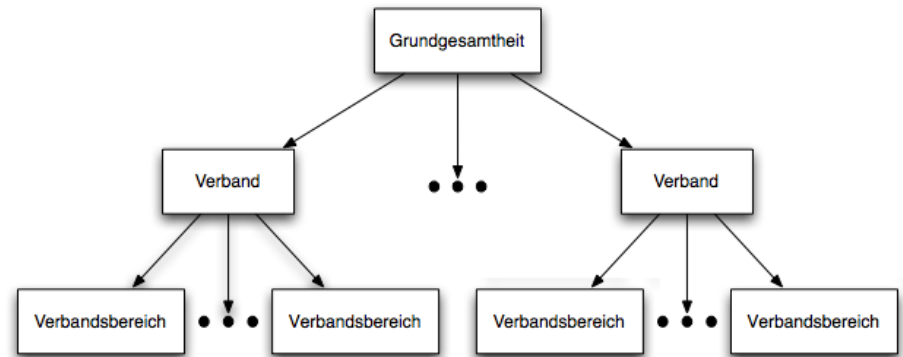
- Streuung schränkt die Repräsentativität der Melder ein
  - Bei weiteren Unternehmensmerkmalen ähnliche Streuung (Rechtsform, Größe)
- **Clusterung** der Grundgesamtheit

# Clusterung

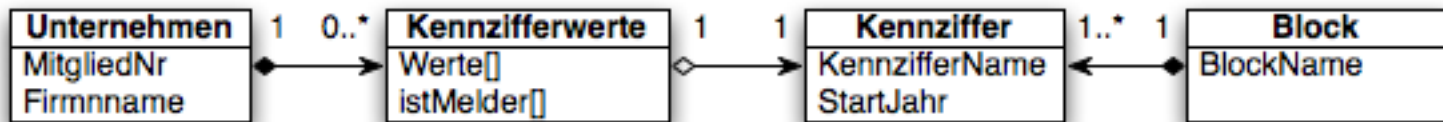
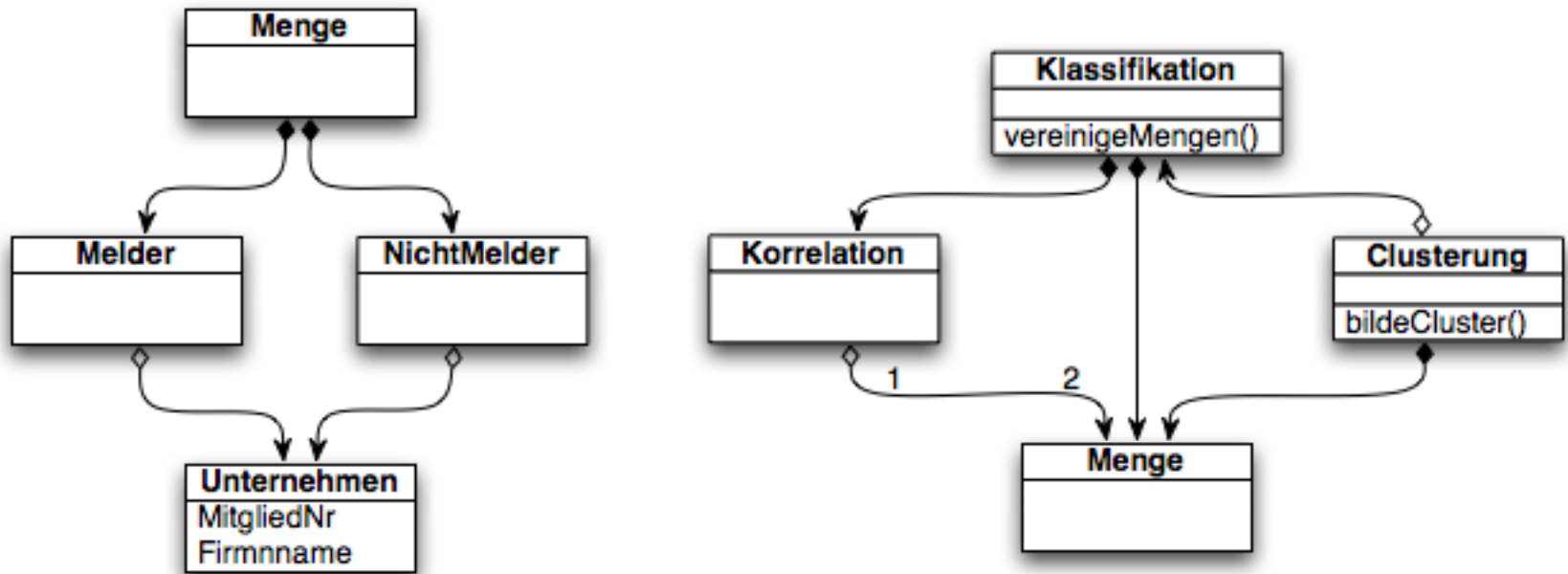
## Homogenisierung durch Bildung von Clustern

### Clusterung der Unternehmen

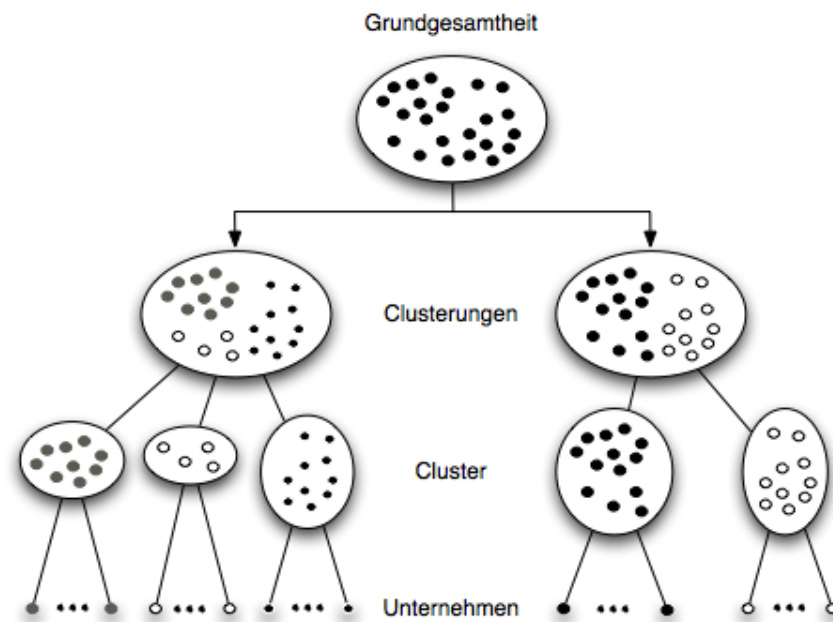
- Hierarchische Clusterung
  - nach Verband/Verbandsbereich
  - nach Rechtsform
  - nach Rechtsform und Verband
  
- Clusteranalyse
  - nach Größe des Unternehmens (Anzahl der Wohnungseinheiten)
  - nach gemeldeter Kennziffergröße des letzten Jahres



# Klassendefinitionen



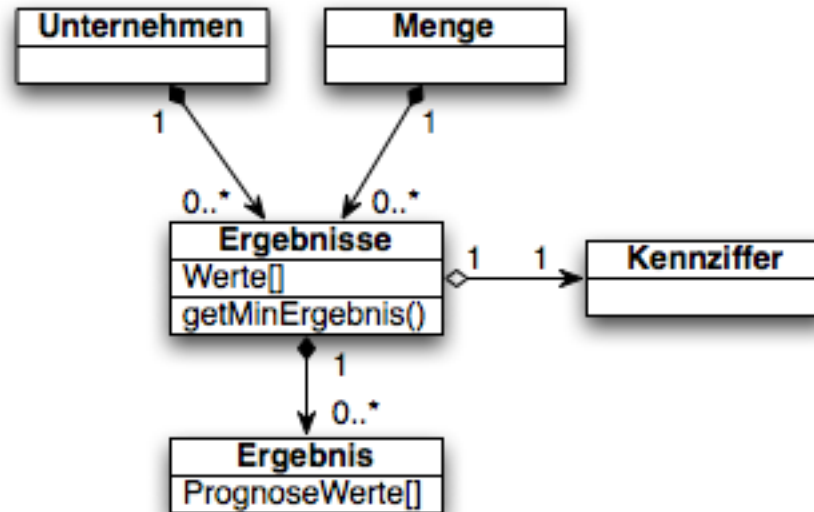
# Prognose-Algorithmus



- Berechnung verschiedener Clusterungen der Grundgesamtheit
- Berechnung der clusterspezifischen Faktoren und Mittelwerte anhand der Melder-Werte
- Hochrechnung der Nicht-Melder durch verschiedene Prognoseverfahren
- Auswahl des „besten“ Unternehmensergebnisses
- Auswahl des „besten“ Clusterergebnisses
- Auswahl der „besten“ Clusterung
- Speicherung der Prognosewerte mit dem besten globalen Verfahren



# Klassendefinition

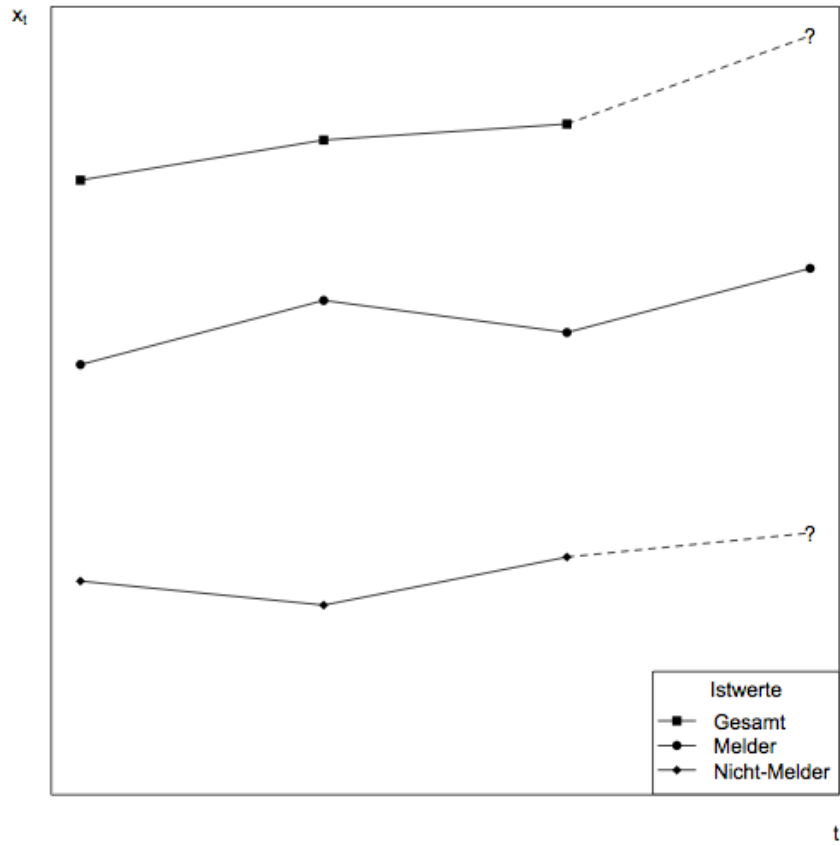


Verwaltung der Prognoseergebnisse

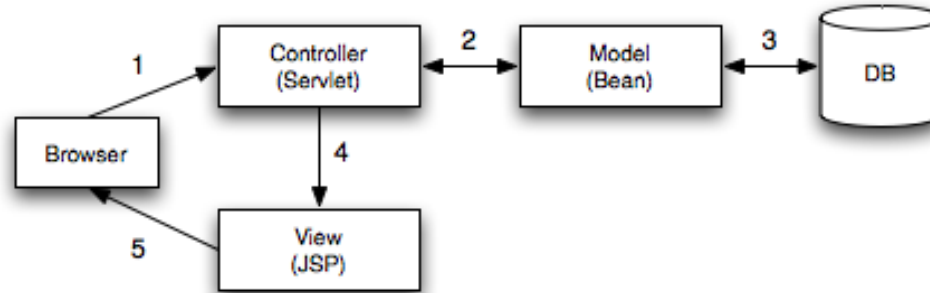
# Ausnahmefälle

Weniger als 4-malig erhobene Kennziffern:

- Analyse der Fälle durch randomisierte Auswahl von Unternehmen mit Kennzifferwerten des Vorjahres
- Definition von Hybridverfahren
- Auswahl eines Prognoseverfahrens und einer Clustering regel-basiert



# Prognose-Tool (Zugang)



## Model-View-Controller von Java für Webentwicklung

- ermöglicht den Web-Zugriff per Browser für die Verbände





# Prognose-Tool (Suche)


Wechsel zu Ad-hoc Modus

Startseite
**Suche**
Logout

Suchbegriff:  Suchen 12 von insgesamt 240 Ergebnissen gefunden

Nr.	Block	Ziffer	Datenfeldname	Beschreibung	Feldtyp
1	E	4.2	c622000	Gesamtwohnfläche der Mietwohnungen im eigenen Wohnungsbestand am 31.12.2008	Zahlenwert
2	F	9	c642000a	Gesamtwohnfläche der leer stehenden eigenen Wohnungen	Zahlenwert
3	I	1.1.5	c812000b	Gesamtwohnfläche dieser Wohnungen	Zahlenwert
4	I	1.2.2	c818000c	Gesamtwohnfläche dieser Wohnungen (warme Betriebskostenvorauszahlung im Dezember 2008)	Zahlenwert
5	I	1.3.2	c812000a	Gesamtwohnfläche dieser Wohnungen (Wohnungen, für die keine Betriebskostenvorauszahlungen erhoben werden)	Zahlenwert
6	L	3.9	c642000b	Gesamtwohnfläche der leer stehenden für Dritte verwalteten Wohnungen	Zahlenwert
7	P	5.1	c898513	Fernwärme   beheizte Fläche	Zahlenwert
8	P	5.2	c898523	Gaskessel   beheizte Fläche	Zahlenwert
9	P	5.3	c898533	Ölkessel   beheizte Fläche	Zahlenwert
10	P	5.4.1	c898543	weitere Energieträger   beheizte Fläche	Zahlenwert
11	P	5.4.2	c898553	weitere Energieträger   beheizte Fläche	Zahlenwert
12	P	5.5	c898563	Hausbedarfsstrom   beheizte Fläche	Zahlenwert

## Suche nach einer Kennziffer



# Prognose-Tool (Prognoseergebnisse)

Wechsel zu Ad-hoc Modus



<a href="#">Startseite</a>	<a href="#">Suche</a>	<a href="#">Logout</a>
----------------------------	-----------------------	------------------------

Klassifikation: +Rechtsform +Wohnungseinheit +OstWest

Block	Ziffer	Datenfeldname	Beschreibung	Feldtyp
E	4	c614000	eigener Wohnungsbestand am 31.12.2008	Zahlenwert

Verbandsbereiche	Melder/Gesamt WU	Melder WU (%)	Melder/Gesamt WE	Melder WE (%)	Schwankung (%)	Schwankung Melder (%)	Trend (%)	Ergebnis
Saarland	6/13	46,2%	5094/11959	42,6%	4,7%	0,1%	-0%	2684+3717= <u>6401</u>
Sachsen - VSWG & VSWU	324/336	96,4%	590129/590423	100%	0,2%	2,1%	-2,1%	547562+294= <u>547856</u>
Thüringen - ptw & vtw	169/172	98,3%	288410/288410	100%	0,1%	2%	-2%	270352+0= <u>270352</u>
Sachsen-Anhalt - VdWg & VdW	178/201	88,6%	374018/380210	98,4%	0,6%	2,8%	-2,8%	350412+4615= <u>355027</u>
Brandenburg	162/205	79%	334345/362482	92,2%	0,7%	1,9%	-1,8%	310812+24024= <u>334836</u>
Nordrhein-Westfalen	383/422	90,8%	895179/948603	94,4%	0,3%	0,3%	-0,3%	827839+45363= <u>873202</u>
Rheinland-Pfalz - vdw Rheinland Westfalen	23/25	92%	19672/19832	99,2%	0,2%	0%	-0%	18213+160= <u>18373</u>
Niedersachsen	122/159	76,7%	242143/293586	82,5%	0,2%	0,5%	+0,4%	222227+46793= <u>269020</u>
Bremen/Bremerhaven	7/16	43,8%	72200/82493	87,5%	0,3%	0,6%	-0,5%	56996+9859= <u>66855</u>
Hamburg	50/67	74,6%	271317/281544	96,4%	0,1%	0,1%	-0,1%	267366+10227= <u>277593</u>
Schleswig-Holstein	51/71	71,8%	113401/132756	85,4%	5,1%	0%	+0%	96074+11109= <u>107183</u>
Mecklenburg-Vorpommern	139/150	92,7%	286323/290661	98,5%	0,3%	0,3%	+0,3%	253846+4136= <u>257982</u>
Hessen	102/135	75,6%	244561/263178	92,9%	2,5%	3,7%	+3,5%	217820+17962= <u>235782</u>
Rheinland-Pfalz - PdW südwest	29/39	74,4%	64316/66138	97,2%	1,9%	0,7%	-0,7%	57632+1793= <u>59425</u>
Bayern	312/457	68,3%	446895/547757	81,6%	0,5%	6,7%	+6%	389537+72185= <u>461722</u>
Württemberg	118/130	90,8%	232256/284953	81,5%	2,6%	2,4%	+2,1%	159027+29319= <u>188346</u>
Baden	121/131	92,4%	171154/172572	99,2%	2,5%	0,2%	-0,2%	140513+1154= <u>141667</u>
Berlin-West	59/62	95,2%	363280/366693	99,1%	1,1%	5,1%	+5,3%	309932+774= <u>310706</u>
Westdeutschland	1/2	50%	94/94	100%	28,6%	0%	0%	94+0= <u>94</u>
Berlin-Ost	47/48	97,9%	255206/255251	100%	0,3%	8%	+8,6%	235808+45= <u>235853</u>
<b>Gesamt c614000</b>	<b>2403/2841</b>	<b>84,6%</b>	<b>5269993/5639595</b>	<b>93,4%</b>	<b>0,2%</b>	<b>0,8%</b>	<b>+0,7%</b>	<b>4734746+283529=<u>5018275</u></b>

Melder-Anteil

Varianz

# Prognose-Tool (Prognoseergebnisse)

Wechsel zu Ad-hoc Modus

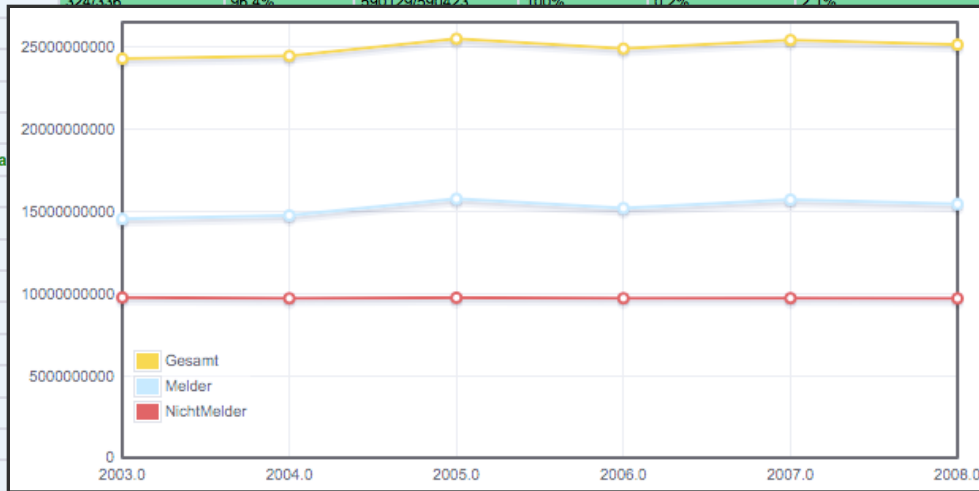


<a href="#">Startseite</a>	<a href="#">Suche</a>	<a href="#">Logout</a>
----------------------------	-----------------------	------------------------

Klassifikation: +Rechtsform +Wohnungseinheit +OstWest

Block	Ziffer	Datenfeldname	Beschreibung	Feldtyp
E	4	c614000	eigener Wohnungsbestand am 31.12.2008	Zahlenwert

Verbandsbereiche	Melder/Gesamt WU	Melder WU (%)	Melder/Gesamt WE	Melder WE (%)	Schwankung (%)	Schwankung Melder (%)	Trend (%)	Ergebnis
Saarland	6/13	46,2%	5094/11959	42,6%	4,7%	0,1%	-0%	2684+3717= <u>6401</u>
Sachsen - VSWG & VSWU	324/336	96,4%	590129/590423	100%	0,2%	2,1%	-2,1%	547562+294= <u>547856</u>
Thüringen - ptw & vtw							-2%	270352+0= <u>270352</u>
Sachsen-Anhalt - VdWg & VdW							-2,8%	350412+4615= <u>354827</u>
Brandenburg							-1,8%	310812+24024= <u>334836</u>
Nordrhein-Westfalen							-0,3%	827839+45363= <u>873202</u>
Rheinland-Pfalz - vdw Rheinland Westfa							-0%	18213+160= <u>18373</u>
Niedersachsen							+0,4%	222227+46793= <u>269020</u>
Bremen/Bremerhaven							-0,5%	56996+9859= <u>66855</u>
Hamburg							-0,1%	267366+10227= <u>277593</u>
Schleswig-Holstein							+0%	96074+11109= <u>107183</u>
Mecklenburg-Vorpommern							+0,3%	253846+4136= <u>257982</u>
Hessen							+3,5%	217820+17962= <u>235782</u>
Rheinland-Pfalz - PdW südwest							-0,7%	57632+1793= <u>59425</u>
Bayern							+6%	389537+72185= <u>461722</u>
Württemberg							+2,1%	159027+29319= <u>188346</u>
Baden							-0,2%	140513+1154= <u>141667</u>
Berlin-West	59/62	95,2%	363280/366693	99,1%	1,1%	5,1%	+5,3%	309932+774= <u>310706</u>
Westdeutschland	1/2	50%	94/94	100%	28,6%	0%	0%	94+0= <u>94</u>
Berlin-Ost	47/48	97,9%	255206/255251	100%	0,3%	8%	+8,6%	235808+45= <u>235853</u>
<b>Gesamt c614000</b>	<b>2403/2841</b>	<b>84,6%</b>	<b>5269993/5639595</b>	<b>93,4%</b>	<b>0,2%</b>	<b>0,8%</b>	<b>+0,7%</b>	<b>4734746+283529=<u>5018275</u></b>



Melder-Anteil

Varianz

# Zusammenfassung

- verschiedene Prognoseverfahren sind wegen der Heterogenität der Kennziffern erforderlich
- Clusterung erhöht die Repräsentativität der Melder
- Konkurrenz der Prognoseverfahren und Clusterungen erhöht die Flexibilität des Algorithmus und verringert die regel-basierte Hochrechnung
- Hochrechnung von Ausnahmefällen verläuft regel-basiert
- Bewertung der Prognoseergebnisse durch die Varianz, den Melder-Anteil und die Anzahl der erhobenen Jahre möglich

Ende des Vortrags

Vielen Dank für Ihre Aufmerksamkeit