

# Predictive Analytics

---

VON PAUL BAUMGARTEN



# Der Fall „Target“

---

Werbemail an minderjährige mit Babyartikeln regt Vater auf.

Vater beschwert sich beim Manager – der entschuldigt sich.

Manager ruft nach ein paar Tagen nochmal an – Vater entschuldigt sich.

## **Was war passiert?**



Target sammelt Daten aller Käufer (Kreditkartenzahlung o.Ä.)

Erstellt „pregnancy score“ aus Kaufverhalten

- Calcium, Zink und Magnesium -> erste 20 Wochen
- Wattebäusche und Seifen
- Geruchsfreie Seife und ganz viel Watte -> bald kommt's

# Der Fall „Target“

---

# Wie kommt sowas an?

---

“If we send someone a catalog and say, ‘Congratulations on your first child!’ and they’ve never told us they’re pregnant, that’s going to make some people uncomfortable,..”

“We are very conservative about compliance with all privacy laws. **But even if you’re following the law, you can do things where people get queasy.**”

Andrew Pole, Statistiker bei Target

# Gliederung

---

Anwendungsgebiet

Erste Schritte

Regression

Maschinelles Lernen

Aktuelle Forschung

Konferenzen

# Anwendung

---

Verhaltensanalyse -> Betrüger erkennen

Lukrativste Kunden indentifizieren, Cross-Selling Chances, Kundenverhalten

Ressourcenmanagement (z.B.: Überbuchungsplanung)

Risikobewertung z.B.: Kreditwürdigkeit, Versicherungen

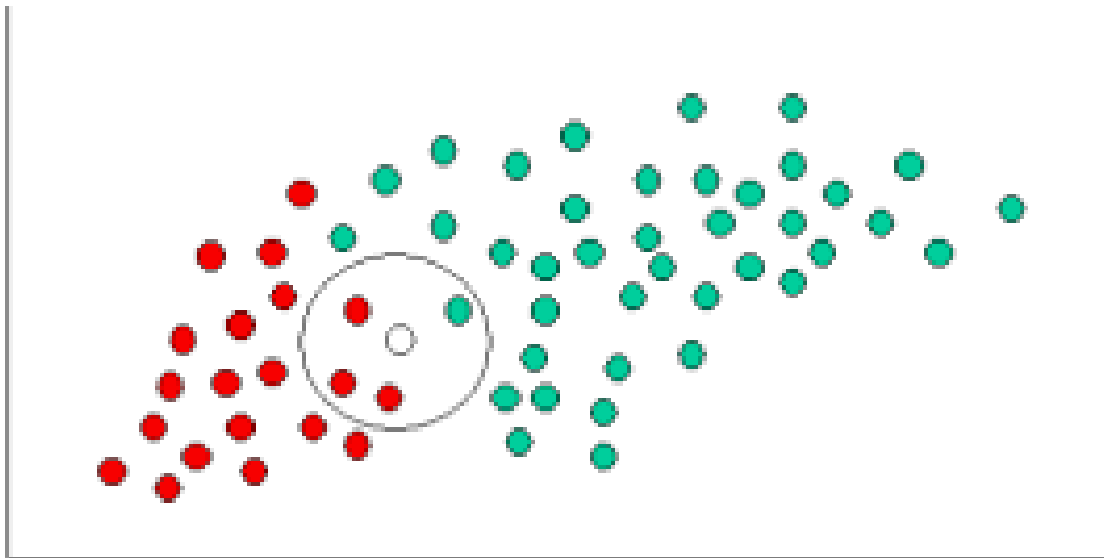
DETECTING FRAUD

OPTIMIZING  
MARKETING  
CAMPAIGNS

IMPROVING  
OPERATIONS

REDUCING RISK

# Bayes-Klassifikation



<http://www.statsoft.com/textbook/naive-bayes-classifier>

40 Grüne, 20 Rote

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

$$\text{Prior}(\text{red}) = 20 / 60$$

$$\text{Prior}(\text{green}) = 40 / 60$$

$$\text{Likelihood}(\text{red}) = 3 / 20$$

$$\text{Likelihood}(\text{green}) = 1 / 40$$

$$\text{posterior}(\text{red}) = (20 / 60) * (3 / 20) = 1 / 20$$

$$\text{posterior}(\text{green}) = (40 / 60) * (1 / 40) = 1 / 60$$

Neuer Punkt wird Grün,  
da  $1/20 > 1/60$

# Problematik

---

Nicht immer ideales Cluster vorhanden

Abhängigkeiten zwischen den Features

Beispiel Ansteckung einer Krankheit

1. Schwaches Immunsystem
2. Keimexposition
3. Fehlernährung

$$P(1) = 1/10$$

$$P(2) = 4/5$$

$$P(3) = \frac{1}{2}$$

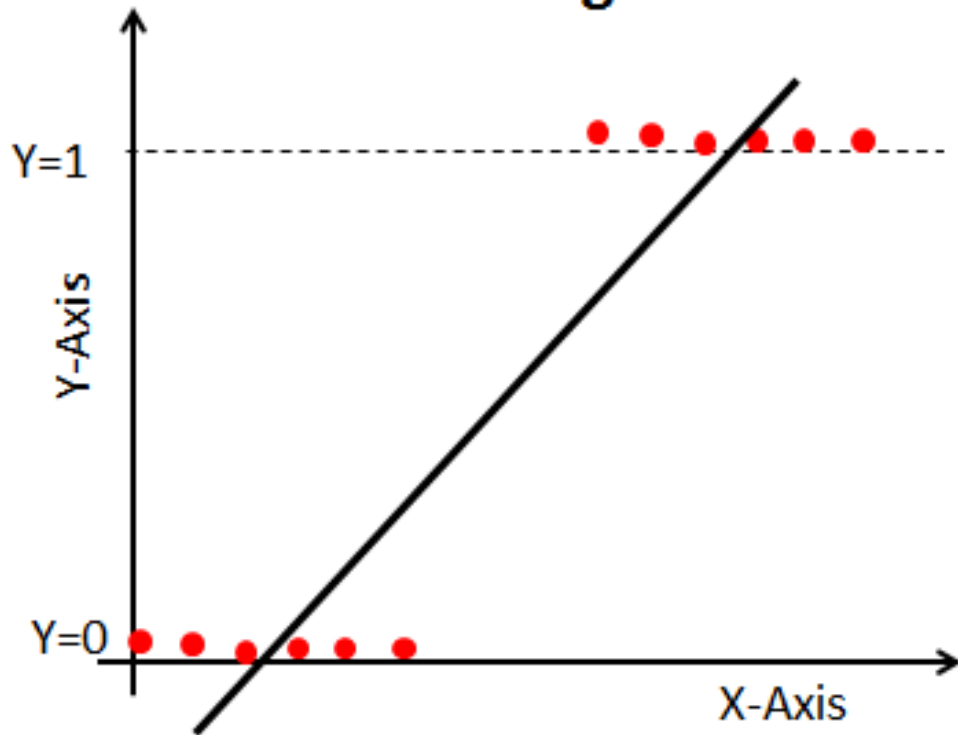
Aber wenn 3  $\rightarrow$   $P(1)=7/10$   $\rightarrow$  P(3) hat mehr Einfluss (gewichtet)  $\rightarrow$  Widerspruch zur Unabhängigkeitsannahme



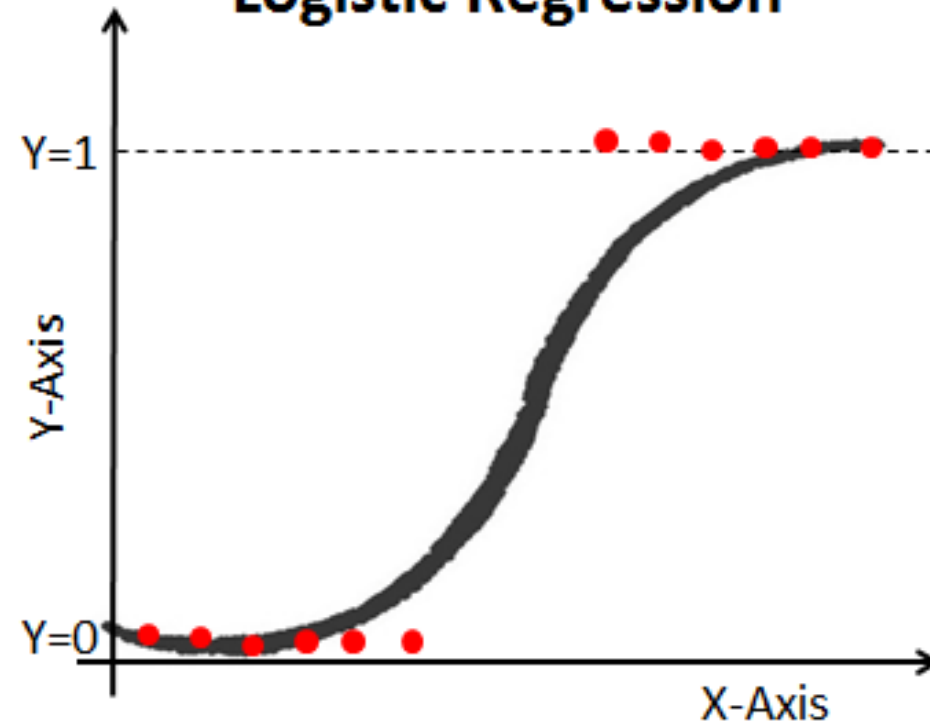
# Regression

---

## Linear Regression



## Logistic Regression



<https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python>

# Maschinelles Lernen

---

Neuronale Netze

Support Vector Maschinen

Naive Bayes

K-nearest Neighbours

# Aktuelle Forschung – Random Forest

---

Konzept nicht neu (1990er)

- Entscheidungsbäume + Demokratie

Parallelsierbar -> schnell

Erklärbar

Gute Ergebnisse

# Arbeiten

---

G. Bejarano, M. Jain, A. Ramesh, A. Seetharam and A. Mishra, "Predictive Analytics for Smart Water Management in Developing Regions," *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, Taormina, Sicily, Italy, 2018, pp. 464-469.

P. Anantharaman, M. Qiao and D. Jadav, "Large Scale Predictive Analytics for Hard Disk Remaining Useful Life Estimation," *2018 IEEE International Congress on Big Data (BigData Congress)(BIGDATA CONGRESS)*, San Francisco, CA, USA, 2018, pp. 251-254.

# Konferenzen

---

## Directed:

Predictive Analytics Konferenz, Wien 2018

Predictive Analytics World, Berlin 2018

## Related:

*2018 IEEE International Congress on Big Data (BigData Congress)(BIGDATA CONGRESS), San Francisco, CA, USA, 2018, pp. 251-254.*

*2018 IEEE International Conference on Smart Computing (SMARTCOMP), Taormina, Sicily, Italy, 2018*

2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)

International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)

# Quellen

---

Breiman L., Random forests. In Machine Learning, 2001

*Hand, D. J.; Yu, K. (2001). "Idiot's Bayes — not so stupid after all?". International Statistical Review. 69 (3): 385–399.*

<http://www.statsoft.com/textbook/naive-bayes-classifier>

<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#34cd9fb46668>

<https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1& r=1&hp>

<https://intl.target.com/>

[https://www.sas.com/de\\_at/insights/analytics/predictive-analytics.html](https://www.sas.com/de_at/insights/analytics/predictive-analytics.html)