



# INTELLIGENTE DATENANALYSE IN MATLAB

Einführungsveranstaltung

# Überblick

- Organisation.
- Literatur.
- Inhalt und Ziele der Vorlesung.
- Beispiele aus der Praxis.

# Organisation

- Vorlesung/Übung + Projektarbeit.
- 4 Semesterwochenstunden.
- Vorlesung: Mo 10:00-11:30, PC-Pool 04.1.03.
- Übung: Mo 12:00-13:30, PC-Pool 04.1.03.
- Babylon-Account für PC-Pool bei Herrn Glöde (Raum 04.1.09) beantragen.
- Webseite mit Terminen, Folien und Downloads:  
<http://www.cs.uni-potsdam.de/ml>

# Organisation

- Vorlesung:
  - Grundlagen Maschinelles Lernen und MATLAB.
  - Konzepte/Algorithmen der intelligenten Datenanalyse.
  
- Übung:
  - Implementierung der Algorithmen in MATLAB.
  - Anwenden der Verfahren auf Praxis-Problemstellungen.

# Organisation

## □ Projektarbeit:

- Selbständiges Bearbeiten eines Problems aus der Praxis:
  - Problemstellung und geeignetes Verfahren identifizieren.
  - Verfahren in MATLAB implementieren und auf gegebene Daten anwenden.

## □ Prüfung:

- Vorstellung der Projektarbeit.
- Mündliche Prüfung zu Themen der Vorlesung/Übung.
- Prüfungsnote: 50% Prüfung + 50% Projektarbeit.

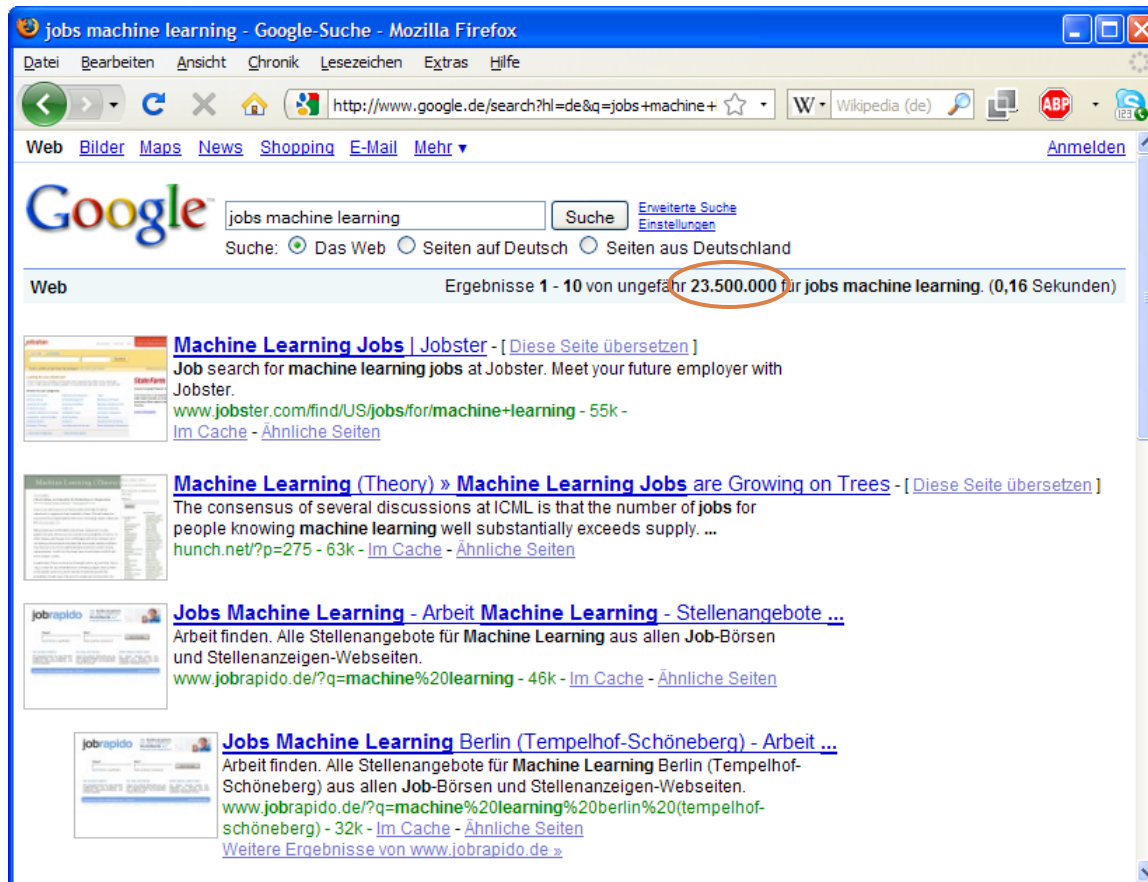
# Literatur

- Chris Bishop: Pattern Recognition and Machine Learning.
- Richard O. Duda: Pattern Classification.
- David G. Stork: Computer Manual in MATLAB to Accompany Pattern Classification.
- MATLAB-Demos & -Tutorials im Internet, z.B.  
<http://www.mathworks.de/products/matlab/demos.html>  
[http://www.mathworks.de/academia/student\\_center/tutorials/launchpad.html](http://www.mathworks.de/academia/student_center/tutorials/launchpad.html)

# Inhalte und Ziele der Vorlesung

- Grundlagen im Arbeiten mit MATLAB.
- Mathematische Grundlagen der Optimierung, Statistik und Numerik.
- Methoden des Maschinellen Lernens:
  - Klassifikation & Regression.
  - Clustering.
  - Ranking, Recommendation und Reputation.
  - Dimensionsreduktion und Visualisierung, usw.
- Selbständiges Durchführen von Datenanalysen, d.h. bearbeiten von Problemen des Maschinellen Lernens.

# Motivation



jobs machine learning - Google-Suche - Mozilla Firefox

http://www.google.de/search?hl=de&q=jobs+machine+

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

Ergebnisse 1 - 10 von ungefähr 23.500.000 für jobs machine learning. (0,16 Sekunden)

**Machine Learning Jobs | Jobster** - [ Diese Seite übersetzen ]  
Job search for machine learning jobs at Jobster. Meet your future employer with Jobster.  
[www.jobster.com/find/US/jobs/for/machine+learning](http://www.jobster.com/find/US/jobs/for/machine+learning) - 55k -  
[Im Cache](#) - [Ähnliche Seiten](#)

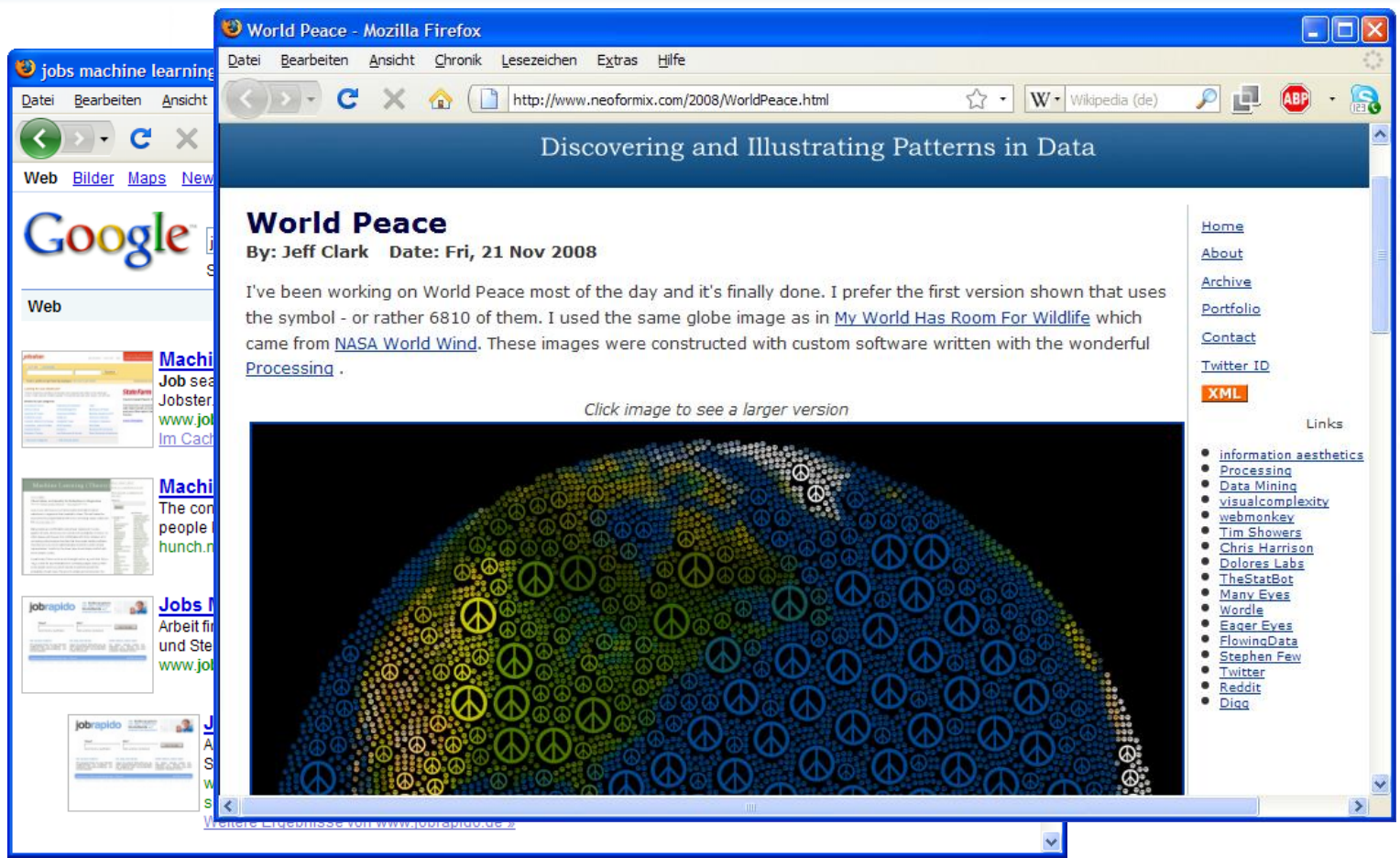
**Machine Learning (Theory) » Machine Learning Jobs are Growing on Trees** - [ Diese Seite übersetzen ]  
The consensus of several discussions at ICML is that the number of jobs for people knowing machine learning well substantially exceeds supply. ...  
[hunch.net/?p=275](http://hunch.net/?p=275) - 63k - [Im Cache](#) - [Ähnliche Seiten](#)

**Jobs Machine Learning - Arbeit Machine Learning - Stellenangebote ...**  
Arbeit finden. Alle Stellenangebote für Machine Learning aus allen Job-Börsen und Stellenanzeigen-Webseiten.  
[www.jobrapido.de/?q=machine%20learning](http://www.jobrapido.de/?q=machine%20learning) - 46k - [Im Cache](#) - [Ähnliche Seiten](#)

**Jobs Machine Learning Berlin (Tempelhof-Schöneberg) - Arbeit ...**  
Arbeit finden. Alle Stellenangebote für Machine Learning Berlin (Tempelhof-Schöneberg) aus allen Job-Börsen und Stellenanzeigen-Webseiten.  
[www.jobrapido.de/?q=machine%20learning%20berlin%20\(tempelhof-schoeneberg\)](http://www.jobrapido.de/?q=machine%20learning%20berlin%20(tempelhof-schoeneberg)) - 32k - [Im Cache](#) - [Ähnliche Seiten](#)  
[Weitere Ergebnisse von www.jobrapido.de »](#)



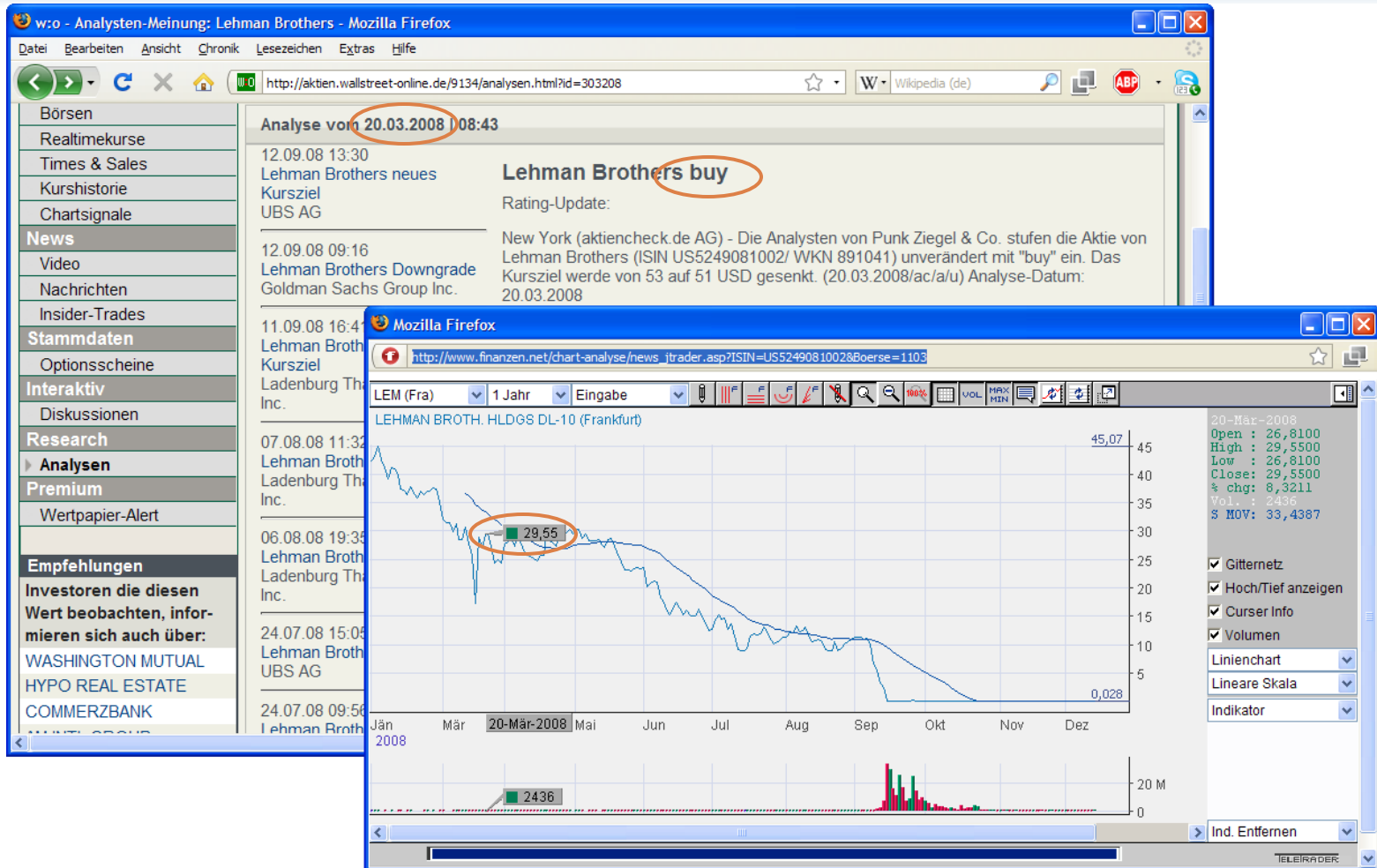
# Motivation



# Was ist Intelligente Datenanalyse?

- Gegeben: Beobachtungen/Daten.
- Gesucht: Zusammenhänge in den Daten.
- Ansatz: Modell finden welches die Beobachtungen bzw. bestimmte Teile davon gut erklärt.
  - Datenvorverarbeitung und geeignete Repräsentation der Daten wählen.
  - Finden eines Modells durch Maschinelles Lernen:
    - Modellierung des Lernproblems: Bestimmen von gegebenen und gesuchten Größen, Performanzmaß/Zielkriterium, Modellraum usw.
    - Lösen des Lernproblems: Algorithmus zum Finden eines geeigneten Modells.
  - Anwenden des Modells auf neue Daten.

# Beispiel: Modellierung von Risiken



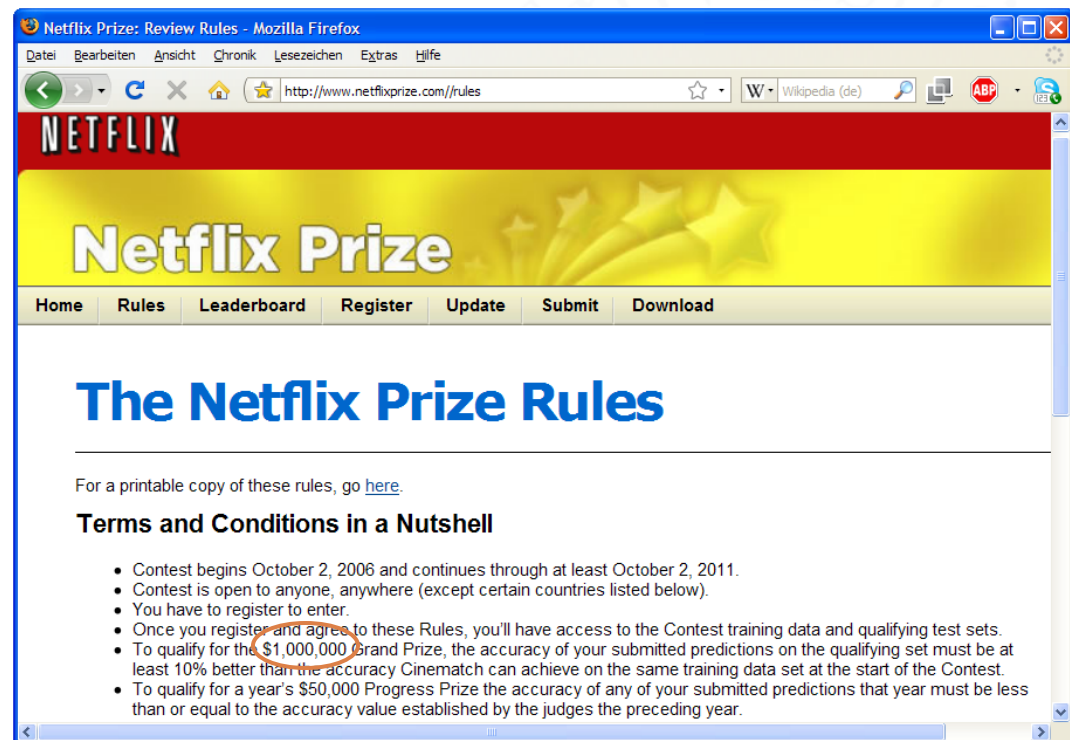
# Beispiel: Cross-/Up-Selling

- Entdecken von Mustern in Datenbanken:
  - ▣ Welche Produkte anbieten?
  - ▣ Wie und wo Produkte platzieren & bewerben?
  - ▣ Kaufvorschläge & Bundles generieren.



# Beispiel: Empfehlungen

- Basierend auf Nutzerverhalten/-bewertungen:
  - Vorhersage von neuen Bewertungen.
  - Produkt-empfehlungen für andere Nutzer.



# Beispiel: Spam Filtern

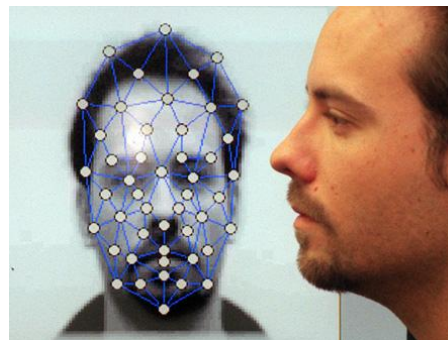
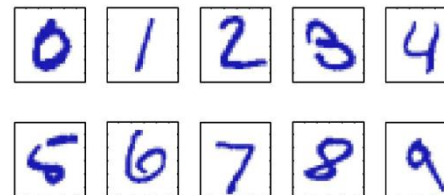
- Klassifikation aufgrund von Text, Bildern, URLs ...
- Große Datenmenge beim Training (>1 Mio Emails).
- Klassifikationsproblem mit Gegenspieler:
  - Gegenspieler verändert Verhalten in Reaktion auf gelerntes Modell.
- Hohe Genauigkeit bei legitimen Emails.



# Beispiel: Mustererkennung

## □ Sprach-, Text- und Bildererkennung:

- Handschrifterkennung
- Gesichtserkennung
- Sprachsteuerung
- uvm.







# Wir stellen ein

- Wir suchen
  - Studentische Mitarbeiter, gern im Zusammenhang mit Studien- und Diplomarbeit.
  - Wissenschaftliche Mitarbeiter.
  - Promotionsstipendiaten.
  
- Wir bieten
  - Erstklassiges Forschungs- und Arbeitsumfeld.
  - Interessante Themen und Problemstellungen, z.T. Drittmittelprojekte mit Industriepartnern.