



Prof. Dr. Dr. Lars Schmidt-Thieme Institut für Informatik
 Prof. Dr. Christof Wecker Institut für Erziehungswissenschaft
 Prof. Dr. Jürgen Sander Vizepräsident für Lehre & Institut für Mathematik

1. Qualität und Originalität des Lehrkonzepts

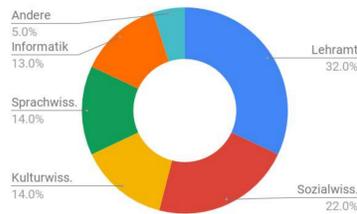
- mikromodular aufgebaute Einführung in die Data Science
 - Einheiten von max. 10 Minuten
- methodische Mikromodule
 - für alle Studierenden gleich

& Beispiel-Mikromodule

- fächerspezifisch
- Theorie — Vorlesung & Übung
- & Praxis — Stationen-Praktikum
- & Innovation — interdisziplinäres Proseminar

2. Institutionelle Breitenwirkung

- 1. Stufe: Studierende aller Studiengänge mit Statistik-Modul
 - 41% aller Studierenden
- 2. Stufe: alle Studierenden



3. Bedarfsorientierung

- Fachdisziplinen:
 - interdisziplinäre Arbeitsgruppe *Data Literacy*
 - interdisziplinäre Projekte und Seminare
- Arbeitswelt & Gesellschaft:
 - 35 Partnerunternehmen sowie Experten aus der Bildungspolitik, der Schulverwaltung und der Bildungsforschung
 - studentische Projekte mit Unternehmen, Schulen und anderen Organisationen
 - Alumni-Befragung
 - Data Science-Projekttag

4. Strukturen und Prozesse zur nachhaltigen Etablierung

- *Research and Innovation Center Data Science (RICDS)*
 - 3 methodische Data Science-Arbeitsgruppen
 - Datenintensiv arbeitende Gruppen aus der Computerlinguistik, Psychologie, Erziehungswissenschaft, Politikwissenschaft etc.
- curriculare Einbindung in Wahlbereiche aller Studiengänge
- interdisziplinäre Arbeitsgruppe *Data Literacy*
- Kooperation mit
 - Methodenbüro
 - Zentrum für Lehrerbildung (CeLeB)
 - Fachstelle für Forschungsdatenmanagement

5. Expertise und Zusammensetzung des Konsortiums



Prof. Dr. Dr. Lars Schmidt-Thieme	Prof. Dr. Ulrich Heid	Prof. Dr. Wolf Schünemann	Prof. Dr. Renate Soellner	Prof. Dr. Christof Wecker	Prof. Dr. Martin Sauerwein
Maschinelles Lernen	Computerlinguistik	Politikwissenschaft	Psychologie	Erziehungswissenschaft	Geographie

Aufbau der beiden Zertifikate

Data Literacy-Zertifikat (16 ECTS)

<i>Einführung in die Data Science</i>	Vorlesung & Übung	6 ECTS
<i>Data Science Lab</i>	Stationen-Praktikum	6 ECTS
<i>Data Science Proseminar</i>	Proseminar	4 ECTS

Data Science-Zertifikat (28 ECTS)

<i>Einführung in die Data Science</i>	Vorlesung & Übung	6 ECTS
<i>Data Science Lab</i>	Stationen-Praktikum	6 ECTS
<i>Data Science Proseminar</i>	Proseminar	4 ECTS
Data Science-Wahlmodul	verschiedene	6 ECTS
Data Science-Wahlmodul	verschiedene	6 ECTS

Kompetenzen

1. Daten kodieren und visualisieren.
2. Modelle für Abhängigkeiten von Zielmerkmalen von Prädiktormerkmalen entwerfen.
3. Modelle für Abhängigkeiten von Zielmerkmalen von Prädiktormerkmalen aus Daten lernen.
4. Modelle für Struktur in Daten entwerfen.
5. Modelle für Struktur in Daten lernen.
6. Modelle analysieren und vergleichen.
7. Datenbasierte Schlussfolgerungen über Korrelationen aus Daten ziehen und von Kausalitäten abgrenzen.
8. Modelle für Daten verschiedener Modalitäten (Texte, Bilder, Relationen etc.) entwerfen.
9. Modelle für Daten verschiedener Modalitäten (Texte, Bilder, Relationen etc.) lernen.
10. **Aufsätze der eigenen Disziplin mit Data Science-Methodik verstehen.**
11. Datensammlung und Verarbeitung mit Berücksichtigung des Datenschutzes entwerfen.
12. Experimente reproduzierbar dokumentieren.

Ein Beispiel

Beyond Linear Regression: Learning Embeddings

For **nominal predictors**, linear regression learns only **level-specific bias terms**:

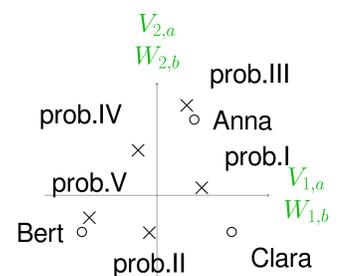
$$\hat{y}(a, b) := u + v^T \mathbf{1}_a + w^T \mathbf{1}_b = u + v_a + w_b \quad (1)$$

More expressive models embed the **nominals** in a latent space by associating their levels with **latent features**:

$$\hat{y}(a, b) := u + (V \mathbf{1}_a)^T W \mathbf{1}_b = u + V_{:,a}^T W_{:,b} \quad (2)$$

Education: Predicting Student Performance

student <i>a</i>	problem <i>b</i>	perf. <i>y</i>
Anna	problem I	100
Anna	problem III	90
Anna	problem V	20
Bert	problem I	10
Bert	problem II	90
Bert	problem V	80
Clara	problem I	80
Clara	problem II	90
Clara	problem IV	10



Linguistics: Predicting Relations between Words

word <i>a</i>	word <i>b</i>	is_a <i>y</i>
king	man	true
king	human	true
king	tool	false
queen	woman	true
queen	human	true
queen	plant	false

