

---

**Modulbezeichnung:** Data Warehousing und Knowledge Discovery in Databases (DWKDD) 5 ECTS

(Data Warehousing and Knowledge Discovery in Databases)

Modulverantwortliche/r: Richard Lenz

Lehrende: Thomas Ruf, Richard Lenz

---

Startsemester: SS 2021

Dauer: 1 Semester

Turnus: jährlich (SS)

Präsenzzeit: 60 Std.

Eigenstudium: 90 Std.

Sprache: Deutsch und Englisch

---

**Lehrveranstaltungen:**

Ausschlussbedingung: Dieses Modul darf nur abgelegt werden, wenn keine der im Modul enthaltenen Lehrveranstaltungen auch noch in einem anderen Modul enthalten ist, das bereits abgelegt wurde.

Knowledge Discovery in Databases (SS 2021, Vorlesung, 2 SWS, Richard Lenz et al.)

Data Warehousing (SS 2021, Vorlesung, 2 SWS, Thomas Ruf)

---

**Empfohlene Voraussetzungen:**

siehe Lehrveranstaltungen

---

**Inhalt:**

**KDD**

1. Introduction
2. Know Your Data
3. Data Preprocessing
4. Mining Frequent Patterns, Associations and Correlations: Basic Concepts and Methods
5. Advanced Frequent Pattern Mining
6. Classification: Basic Concepts
7. Classification: Advanced Methods
8. Cluster Analysis: Basic Concepts and Methods
9. Cluster Analysis: Advanced Methods
10. Outlier Detection
11. Trends and Research Frontiers in Data Mining

**DW**

Das "Data Warehouse" stellt sowohl im theoretischen Bereich der Datenbankforschung als auch in der praktischen Anwendung in der Wirtschaft ein "Hot Topic" dar. Die breite Fächerung des Themengebietes macht es notwendig, sich dem Begriff "Data Warehouse" von verschiedenen Blickwinkeln zu nähern. Die Hauptpunkte der Vorlesung liegen dabei in der Diskussion der unterschiedlichen Architekturansätze, der zugehörigen Datenmodelle und den verarbeitungstechnischen Grundlagen. Weiterhin wird auf aktuelle Diskussionspunkte wie die Realisierungstechniken ROLAP und MOLAP oder die Aggregatbildung, -verwendung und -haltung eingegangen. Die Vorlesung ist geprägt von einer Mischung aus theoretischen Grundlagen und praktischer Anwendung von neuen Forschungsergebnissen. Um einen stärkeren Praxisbezug zu erreichen, werden mehrere Vorträge von Firmenvertretern bestritten, die zum einen von ihren Erfahrungen bei dem Aufbau eines "Data Warehouses" berichten und zum anderen den Stand der Technik in einschlägigen Data Warehouse- Systemen darstellen.

**Lernziele und Kompetenzen:**

**DW**

Die Studierenden

- kennen verschiedene Anwendungsgebiete des Data Warehousing und charakterisieren diese;
- modellieren multidimensionale Datenbanken;
- entwickeln ein Grundverständnis für die Abbildung multidimensionaler Datenstrukturen in Datenbanken und bewerten die Vor- und Nachteile der verschiedenen Ansätze in konkreten Anwendungssituationen;
- setzen grundlegende Verfahren der Datenauswertung in Data Warehouse-Systemen für konkrete Anwendungssituationen ein;

- erläutern die Architektur und den Betrieb technischer Data Warehouse-Systeme und erläutern die sich hieraus ergebenden Implikationen im Betrieb solcher Systeme;
- erklären die von Praxisvertretern vorgestellten Einsatzszenarien und Systemansätzen für Data Warehouse-Systeme;
- schätzen aktuelle Entwicklungen im Themengebiet ein.

#### *Fachkompetenz*

##### *Wissen*

KDD: Die Studierenden

- kennen den typischen KDD-Prozess;
- kennen Verfahren zur Vorbereitung von Daten für das Data Mining;
- kennen die Definition von Distanz- bzw. Ähnlichkeitsfunktionen für die verschiedenen Typen von Attributen;
- sind vertraut mit dem Prinzip des Apriori-Algorithmus zur Bestimmung von Mengen häufiger Elemente (frequent itemsets);
- kennen den FP-Growth-Algorithmus zum schnellen Auffinden von Mengen häufiger Elemente;
- können die Definitionen von Support und Confidence für Assoziationsregeln wiedergeben;
- können die Ermittlung von Assoziationsregeln auf der Basis von Mengen häufiger Elemente beschreiben;
- sind in der Lage, die Vorgehensweise bei Klassifikationsaufgaben wiederzugeben;
- können darlegen, wie ein Entscheidungsbaum auf einem Trainingsdatensatz erzeugt wird;
- können das Prinzip der Bayes'schen Klassifikation darstellen;
- können verschiedene Clustering-Verfahren aufzählen;
- können den Ablauf von k-Means-Clustering beschreiben;
- kennen die verschiedenen Arten von Ausreißern.

##### *Verstehen*

KDD: Die Teilnehmer können

- Distanz- oder Ähnlichkeits-Funktionen auf einem speziellen Datenbestand definieren;
- Attribute eines Datensatzes auf ihre Bedeutung für die Analyse hin überprüfen und ggf. Attributwerte geeignet transformieren.

#### *Lern- bzw. Methodenkompetenz*

KDD:

Die Studierenden

- müssen ein hohes Maß an Eigeninitiative zur Aneignung des Stoffes mitbringen und werden darin bestärkt;
- müssen die englische Sprache so weit beherrschen, dass sie der Vorlesung folgen können;
- verwenden gezielt ein Lehrbuch zur Vorlesung;
- bekommen zahlreiche Hinweise auf zugrundeliegende Literatur, die sie bei Bedarf herausuchen und durcharbeiten müssen.

#### *Selbstkompetenz*

KDD:

Die Teilnehmer

- müssen ggf. fehlendes Vorwissen selbständig nacharbeiten (falls sie diese Lehrveranstaltung trotzdem wählen);
- müssen sich selbst die Zeit für die Vor- und Nachbereitung des Stoffs einteilen.

#### *Sozialkompetenz*

KDD:

Die Teilnehmer

- eignen sich den Stoff in einer Gruppe mit extrem verschiedenen Hintergründen an;
- brauchen das Gespräch mit anderen Teilnehmern zur Aneignung des Stoffs;
- müssen sich in der Diskussion auch selbst der englischen Sprache bedienen.

#### **Literatur:**

siehe Lehrveranstaltungen

---

#### **Verwendbarkeit des Moduls / Einpassung in den Musterstudienplan:**

Das Modul ist im Kontext der folgenden Studienfächer/Vertiefungsrichtungen verwendbar:

**[1] Informatik (Master of Science)**

(Po-Vers. 2010 | TechFak | Informatik (Master of Science) | Gesamtkonto | Wahlpflichtbereich | Säule der softwareorientierten Vertiefungsrichtungen | Vertiefungsrichtung Datenbanksysteme | Data Warehousing und Knowledge Discovery in Databases)

**[2] International Information Systems (IIS) (Master of Science)**

(Po-Vers. 2014w | ReWiFak | International Information Systems (IIS) (Master of Science) | Informatics | Informatics Electives | Data Management II | Data Warehousing und Knowledge Discovery in Databases)

**[3] International Information Systems (IIS) (Master of Science)**

(Po-Vers. 2017w | ReWiFak | International Information Systems (IIS) (Master of Science) | Gesamtkonto | Informatics | Informatics Electives | Data Management II | Data Warehousing und Knowledge Discovery in Databases)

**[4] International Information Systems (IIS) (Master of Science)**

(Po-Vers. 2018w | ReWiFak | International Information Systems (IIS) (Master of Science) | Informatics | Core courses (Informatics) | Data and knowledge (Informatics - Core) | Data Warehousing und Knowledge Discovery in Databases)

---

**Studien-/Prüfungsleistungen:**

Data Warehousing und Knowledge Discovery in Databases (Prüfungsnummer: 228988)

(englische Bezeichnung: Data Warehousing and Knowledge Discovery in Databases)

Prüfungsleistung, mehrteilige Prüfung

Anteil an der Berechnung der Modulnote: 100%

weitere Erläuterungen:

Die Bewertung der Prüfungsleistung setzt sich zusammen aus 1/2 Note der Klausur in Data Warehousing (von etwa 60 Minuten Dauer) und 1/2 Note der mündlichen Prüfung in KDD (von etwa 30 Minuten Dauer). Zum Bestehen der Modulprüfungen müssen beide Teilprüfungen bestanden werden.

Prüfungssprache: Deutsch oder Englisch

Erstablingung: SS 2021, 1. Wdh.: WS 2021/2022

1. Prüfer: Richard Lenz

Data Warehousing und Knowledge Discovery in Databases (Prüfungsnummer: 824544)

(englische Bezeichnung: Data Warehousing and Knowledge Discovery in Databases)

Prüfungsleistung, mehrteilige Prüfung

Anteil an der Berechnung der Modulnote: 100%

weitere Erläuterungen:

Die Bewertung der Prüfungsleistung setzt sich zusammen aus 1/2 Note der Klausur in Data Warehousing (von etwa 60 Minuten Dauer) und 1/2 Note der mündlichen Prüfung in KDD (von etwa 30 Minuten Dauer). Zum Bestehen der Modulprüfungen müssen beide Teilprüfungen bestanden werden.

Prüfungssprache: Deutsch oder Englisch

Erstablingung: SS 2021, 1. Wdh.: WS 2021/2022

1. Prüfer: Richard Lenz

Data Warehousing und Knowledge Discovery in Databases (Prüfungsnummer: 743223)

(englische Bezeichnung: Data Warehousing und Knowledge Discovery in Databases)

Prüfungsleistung, mehrteilige Prüfung

Anteil an der Berechnung der Modulnote: 100%

weitere Erläuterungen:

Die Bewertung der Prüfungsleistung setzt sich zusammen aus 1/2 Note der Klausur in Data Warehousing (von etwa 60 Minuten Dauer) und 1/2 Note der mündlichen Prüfung in KDD (von etwa 30 Minuten Dauer). Zum Bestehen der Modulprüfungen müssen beide Teilprüfungen bestanden werden.

Prüfungssprache: Deutsch oder Englisch

Erstablingung: SS 2021, 1. Wdh.: WS 2021/2022

1. Prüfer: Lenz/Ruf (T10145)

