

Selbstkompetenz:

- Mit zielgerichtetem Zeit- und Selbstmanagement im Selbststudium arbeiten können.
- Eigenständig in Themengebiete einarbeiten und vertiefen können.
- Mit Teamfähigkeit Datenanalysestudien in Gruppen bearbeiten können.

5.2 Lerninhalte**1. Einführung**

- 1.1. Modulorganisation
 - 1.1.1.Lernziele und Gliederung
 - 1.1.2.Termine und Skript
 - 1.1.3.Prüfungsleistung und Literatur
- 1.2. Business Intelligence
 - 1.2.1.Motivation
 - 1.2.2.Prozess
 - 1.2.3.Instrumente
 - 1.2.4.Systeme

2. Data Mining – Grundlagen

- 2.1. Grundbegriffe
- 2.2. Untersuchungsfragestellungen
 - 2.2.1.Assoziation
 - 2.2.2.Klassifikation
 - 2.2.3.Prognose
 - 2.2.4.Projektion
- 2.3. Daten
 - 2.3.1.Merkmale und Fälle
 - 2.3.2.Skalenniveaus
 - 2.3.3.Standarddatenformate
- 2.4. Modellbildung
 - 2.4.1.Prinzip
 - 2.4.2.Hyperparameter
 - 2.4.3.Modellauszeichnung mit Predictive Markup Modeling Language (PMML)
- 2.5. Vorgehensmodell „Cross-Industry Standard Process for Data Mining“ (CRISP-DM)
 - 2.5.1.Business Understanding
 - 2.5.2.Data Understanding
 - 2.5.3.Data Preparation
 - 2.5.4.Modelling
 - 2.5.5.Evaluation
 - 2.5.6.Deployment
- 2.6. Werkzeuge
 - 2.6.1.Klassifikation
 - 2.6.2.Data Mining Suite

3. Data Mining – Data Preprocessing

Modulbeschreibung

- 3.1. Grundlagen
- 3.2. Data Reduction
 - 3.2.1. Aggregation
 - 3.2.2. Dimensionality Reduction
- 3.3. Data Transformation
 - 3.3.1. Discretization, Binarization, Re-Codification
 - 3.3.2. Standardisation
 - 3.3.3. Value Transformation
 - 3.3.4. Feature Creation
- 3.4. Data Cleansing
 - 3.4.1. Imputation of Missing Values
 - 3.4.2. Anomaly Detection
 - 3.4.3. Duplicates
- 3.5. Data Selection
 - 3.5.1. Sampling
 - 3.5.2. Balancing
 - 3.5.3. Feature Selection
- 3.6. Anwendungen
 - 3.6.1. Übersicht
 - 3.6.2. Recency Frequency Monetary Value (RFM)-Analyse
- 4. Data Mining – Assoziation mit Association Rule Learning**
 - 4.1. Grundlagen
 - 4.1.1. Funktionsprinzip und Anwendungen
 - 4.1.2. Verfahren und Vorgehen
 - 4.2. Modellbildung mit Apriori
 - 4.2.1. Übersicht zum Algorithmus
 - 4.2.2. Frequent Itemset Generation
 - 4.2.3. Association Rule Generation
 - 4.2.4. Eigenschaften des Algorithmus
 - 4.2.5. Hyperparameter
 - 4.2.6. Modell in PMML
 - 4.3. Modellbildung mit Continuous Association Rule Mining Algorithm (CARMA)
 - 4.3.1. Prinzip
 - 4.3.2. Hyperparameter
 - 4.3.3. Modellinterpretation
 - 4.4. Modellbewertung
 - 4.4.1. Evaluationsansatz und -kriterien
 - 4.4.2. Support, Confidence, Lift
- 5. Data Mining – Klassifikation mit Decision Trees**
 - 5.1. Grundlagen
 - 5.1.1. Funktionsprinzip und Anwendungen
 - 5.1.2. Verfahren und Vorgehen
 - 5.2. Modellbildung mit Classification & Regression Trees (C&RT)
 - 5.2.1. Übersicht des Algorithmus
 - 5.2.2. Counting
 - 5.2.3. Splitting
 - 5.2.4. Pruning
 - 5.2.5. Hyperparameter
 - 5.2.6. Modellinterpretation
 - 5.3. Modellbildung mit C5.0
 - 5.3.1. Funktionsprinzip
 - 5.3.2. Baumstruktur
 - 5.3.3. Splitting
 - 5.3.4. Hyperparameter

Modulbeschreibung

5.3.5. Modell in PMML

5.4. Modellbewertung

5.4.1. Confusion Matrix

5.4.2. Kennzahlen zur Klassifikationsgüte

5.4.3. Klassifikationskosten

5.4.4. Bedeutsamkeit

5.4.5. Konfidenz

6. Data Mining – Klassifikation mit Support Vector Machines

6.1. Grundlagen

6.1.1. Gegenstand und Eigenschaften

6.1.2. Vektoren und Hyperebenen

6.1.3. Funktionsprinzip

6.2. Modellbildung mit Linear Support Vector Machines

6.2.1. Klassifikationsproblem

6.2.2. Klassifikation

6.2.3. Learning

6.2.4. Hyperparameter

6.2.5. Modell in PMML

6.3. Modellbildung mit Non Linear Support Vector Machines

6.3.1. Nicht-lineare Trennbarkeit

6.3.2. Kernel Trick

6.3.3. Hyperparameter

6.3.4. Modell in PMML

6.4. Modellbewertung

6.4.1. Confusion Matrix

6.4.2. Kennzahlen zur Klassifikationsgüte

6.4.3. Klassifikationskosten

6.4.4. Bedeutsamkeit

6.4.5. Konfidenz

7. Data Mining – Evaluation der Klassifikation

7.1. Grundlagen

7.1.1. Elemente beim Supervised Learning

7.1.2. Loss Function

7.1.3. Objective Function

7.1.4. Regularization

7.1.5. Bias & Bias Variance Tradeoff

7.2. Overfitting

7.2.1. Under- & Overfitting

7.2.2. Validation Data

7.2.3. Cross-Validation

7.3. Evaluationsmetriken

7.3.1. Confusion Matrix Measures

7.3.2. Response, Lift, Gain, Profit, Return on Invest (ROI)

7.3.3. Receiver Operation Characteristic (ROC) & Area under the Curve (AUC)

8. Data Mining – Klassifikation mit Ensemble Learning

8.1. Grundlagen

8.1.1. Funktionsprinzip

8.1.2. Methodenübersicht

8.2. Ensemble Methods

8.2.1. Konzept

8.2.2. Bagging

Modulbeschreibung

8.2.3.Boosting

8.2.4.Stacking

8.3. Random Forrest

8.3.1.Funktionsprinzip

8.3.2.Algorithmus

8.3.3.Hyperparameter

8.3.4.Modellbewertung

8.4. Gradient Boosted Trees (GBT)

8.4.1.Funktionsprinzip

8.4.2.Algorithmus

8.4.3.Hyperparameter

8.4.4.Modellbewertung

9. Data Mining - Segmentierung mit Cluster Analysis

9.1. Grundlagen

9.1.1.Funktionsprinzip und Anwendungen

9.1.2.Verfahren und Vorgehen

9.2. Modellbildung mit k-Means

9.2.1.Distanzmaße

9.2.2.Algorithmus

9.2.3.Hyperparameter

9.2.4.Modell

9.3. Modellbildung mit Two-Step Clustering

9.3.1.Prinzip

9.3.2.Algorithmus

9.3.3.Hyperparameter

9.3.4.Modell

9.4. Modellinterpretation und -bewertung

9.4.1.Vorgehen und Artefakte

9.4.2.Häufigkeitsverteilungen und Parameter

9.4.3.Cluster-Interpretation

9.4.4.Kennzahlen zur Güte des Clustering

10. Big Data

10.1. Grundlagen

10.1.1. Definition und Aspekte („Volume, Variety, Velocity“)

10.1.2. Anwendungsfelder

10.2. Verteilte Systeme

10.2.1. Skalierbarkeit und CAP-Theorem

10.2.2. Architekturmuster (Shared Everything/Storage/Nothing)

10.2.3. Database Sharding

10.2.4. Database Clustering

10.3. Technologien

10.3.1. Dataspace versus Database

10.3.2. Distributed Storage

10.3.3. Distributed Data Processing

10.3.4. Distributed In-Memory Data Processing

10.3.5. Real-time Data Ingestion

10.3.6. Real-time Data In-Memory-Processing

10.3.7. Big Database Systems

10.3.8. Big Data Warehouse Infrastructure

Modulbeschreibung

10.4. Anwendungen

10.4.1. Data Lake

10.4.2. Near Real-time Data Warehouse

11. Data Warehousing (DWH)

11.1. Grundlagen

11.1.1. Motivation und Definition

11.1.2. Funktionsprinzip und Anwendungen

11.1.3. Architektur-Typen von DWH-Systemen

11.2. Basis-Architektur DWH-System

11.2.1. Übersicht

11.2.2. Data Warehouse-Manager

11.2.3. Monitor und Arbeitsbereich (Staging Area)

11.2.4. Extraction, Transformation, Loading (ETL)

11.2.5. Basisdatenbank (Operational Data Store, ODS)

11.2.6. Data Warehouse (DWH)

12. Online Analytical Processing (OLAP)

12.1. Grundlagen

12.1.1. Motivation und Definition

12.1.2. Funktionsprinzip und Anwendungen

12.1.3. Systeme und Vorgehen

12.2. Multidimensionale Datenmodellierung

12.2.1. Dimension (Dimension)

12.2.2. Kenngröße (Measure) und Fakt (Fact)

12.2.3. Würfel (Cube)

12.3. Multidimensionale Abfragen

12.3.1. Drill-down & Roll-up

12.3.2. Slice & Dice

12.3.3. Drill-across & Drill-through

12.4. Konzeptueller Entwurf

12.5. Logischer Entwurf

12.5.1. Star & Snowflake Schema

12.5.2. Galaxy & Fact Constellation Schema

12.6. Physischer Entwurf

12.6.1. Relational OLAP (ROLAP)

12.6.2. Multidimensional OLAP (MOLAP)

12.6.3. Hybrid OLAP (HOLAP)

12.7. Implementierung

12.7.1. Cube-Erstellung (manuell mit SQL)

12.7.2. Cube-Abfragen mit Multidimensional Expressions (MDX)

→ zu den Details: siehe Vorlesungsverzeichnis, Lehrveranstaltungsplan etc.

5 5.3 Modulkurzinformation (Dieser Absatz [max. 250 Zeichen] wird auf der FH-Webseite veröffentlicht, um Studieninteressierte bei der Wahl ihres Studiengangs zu unterstützen. Fokussieren Sie sich auf wesentliche Inhalte und Ziele, gern verbunden mit Aussagen zur Bedeutung des Moduls für das weitere Studium oder berufliche Tätigkeiten. Bitte formulieren Sie ganze Sätze, sprechen Sie die Adressaten direkt an und vermeiden Sie Fachtermini.)



Modulbeschreibung

6	<p>6.1 Teilnahmevoraussetzungen (<i>Formal</i>: Prüfung in Modul XY muss bestanden sein o. ä.; <i>Inhaltlich</i>: Modul XY sollte absolviert sein, folgende Kenntnisse sollten vorhanden sein, ...)</p>
	<p>6.2 Voraussetzungen für die Vergabe von Leistungspunkten (z. B. Bestehen der Prüfung, erfolgreicher Abschluss einer Studienleistung, regelmäßige und aktive Teilnahme)</p> <p>Bestehen der Prüfung</p>
	<p>6.3 Prüfungsformen und -umfang (z. B. Klausur, mündliche Prüfung, Hausarbeit, Präsentation, Portfolio, Dauer der Prüfung in Min.)</p> <p>Datenanalysestudie</p>
	<p>6.4 Voraussetzungen für die Zulassung zur Prüfung</p>
	<p>6.5 Gewichtung der Note bei Ermittlung der Endnote</p> <p>s. Prüfungsordnung/ -en für oben (Zeile 3) genannte Studiengänge*</p> <p><small>*Die Prüfungsordnungen der Studiengänge finden Sie in den Amtlichen Bekanntmachungen der FH Münster unter dem folgenden Link https://www.fh-muenster.de/hochschule/aktuelles/amtliche_bekanntmachungen/index.php?p=2,7.</small></p>
7	<p>7.1 Veranstaltungssprache/n <input checked="" type="checkbox"/> Deutsch <input type="checkbox"/> Englisch <input type="checkbox"/> Weitere, nämlich:</p>
	<p>7.2 Modulverantwortliche/r</p>
	<p>7.3 Hauptamtlich Lehrende (optional) Prof. Dr. Wolfgang Wicht</p>
	<p>7.4 Maximale Teilnehmerzahl (optional)</p>
	<p>7.5 Ergänzende Informationen (optional) (z. B. Literaturempfehlungen, weitere beteiligte Personen etc.)</p>