

Modulhandbuch für das weiterbildende Zertifikatsstudium Data Professional Basics

**Abschluss:
Data Professional (CBS)**

**Stand: März 2025
SPO Version 1 vom 14.02.2026**

Abkürzungen:

- ECTS European Credit Transfer and Accumulation System
- CP Credit Points, ECTS-Punkte
- h Stunden
- SWS Semesterwochenstunden
- SoSe Sommersemester
- WiSe Wintersemester
- SPO Studien- und Prüfungsordnung

Erklärungen:

Modul: Zusammenschluss mehrerer Lehrveranstaltungen zu einer thematisch **zusammenhängenden Einheit** mit gemeinsamem Lernziel.

Workload: Angabe des **Arbeitsaufwands** der Studierenden, der mit dem beschriebenen Modul bzw. der beschriebenen Lehrveranstaltung verbunden ist. Umfasst sind nicht nur Präsenzzeiten, sondern auch Vor- und Nachbereitung der Lehrveranstaltung sowie Zeiten für die Prüfungsvorbereitung. Gemessen wird der Workload in Stunden (h), die sich aus dem Modulumfang in Form von Leistungspunkten, sogenannten Credit Points, ergeben (s.u.).

Credit Points (CP): Credit Points geben den Umfang des Lernens auf Basis von Kompetenzen und den damit verbundenen Arbeitsaufwand (Workload) an. **Ein Credit Point** entspricht an der HKA einem Workload von **30 Arbeitsstunden**.

Inhaltsverzeichnis

Modul 1 Praxisorientierte Datenanalyse für KMU: Grundlagen und Übungen	4
Modul 2 Datenvorverarbeitung, -analyse und -visualisierung	7
Modul 3 Text Data Analytics	9

Modulname: <i>Praxisorientierte Datenanalyse für KMU: Grundlagen und Übungen</i>	
Modulübersicht	
EDV-Bezeichnung:	DPBZ110
Modulverantwortliche(r):	Prof. Dr. Dennis Janka, Prof. Dr. Peer Küppers, Prof. Dr. Jannik Strötgen
Modulumfang (ECTS):	3 CP
	18 h Vorlesung, 72 h Selbststudium, inkl. Seminar- und Prüfungsvorbereitung und Prüfung
Einordnung (Semester):	1
Inhaltliche Voraussetzungen:	keine
Voraussetzungen nach SPO:	keine
Fachkompetenzen:	Die Teilnehmenden können nach Abschluss des Moduls den Nutzen von Daten im unternehmerischen Kontext einordnen und verschiedene Arten von strukturierten und unstrukturierten Daten unterscheiden. Sie sind in der Lage, einfache Visualisierungen und Kennzahlen dem Datentyp entsprechend auszuwählen, mit geeigneter Software zu berechnen und fundiert zu interpretieren. Sie verstehen grundlegende Datenstrukturen, Datenintegrität und relationale Modellierung. Darüber hinaus wenden sie Datenanalyse- und Visualisierungstechniken in Tabellenkalkulations- und BI-Tools sicher an. Die Teilnehmenden verfügen über Grundkenntnisse in der Analyse von Textdaten und sind sich der sprachlichen Mehrdeutigkeiten und deren Einfluss auf die Textanalyse bewusst. Sie können Unternehmensdaten strukturiert aufbereiten und interpretieren, um fundierte Entscheidungen zu unterstützen.
Methodenkompetenzen:	Die Teilnehmenden sind nach Abschluss des Moduls in der Lage, Daten systematisch zu analysieren und zu verarbeiten. Sie setzen Strukturierungs- und Aggregationstechniken gezielt ein, um aus Daten wertvolle Informationen zu gewinnen. Sie nutzen geeignete Werkzeuge zur Visualisierung und Präsentation der Analyseergebnisse. Zudem erkennen sie Schwierigkeiten bei der Verarbeitung von Textdaten und sind für die Herausforderungen der maschinellen Verarbeitung natürlicher Sprache sensibilisiert. Die Teilnehmenden können Konzepte der Datenanalyse auf betriebliche Fragestellungen übertragen.
Sozial- und Selbstkompetenzen:	Die Teilnehmenden sind nach Abschluss des Moduls in der Lage, Datenanalysen verständlich zu präsentieren und die Ergebnisse fundiert zu interpretieren. Sie entwickeln eine analytische Denkweise, die eine datengetriebene Entscheidungsfindung unterstützt. Sie arbeiten selbstständig mit Daten und Analysetools, um betriebliche Fragestellungen zu lösen. Darüber hinaus bewerten sie kritisch die Qualität von Daten und erkennen deren Einfluss auf unternehmerische Entscheidungen. Die Teilnehmenden reflektieren Textanalysen kritisch, insbesondere im Hinblick auf mögliche Verzerrungen und Interpretationsspielräume.
Prüfungsleistungen:	Die theoretischen Kenntnisse der Studierenden werden in einer schriftlichen Klausur (Dauer 60 min, ggf. elektronisch in ILIAS), einer Projektarbeit oder in einer mündlichen Prüfung (Dauer 20 min) bewertet. Die Prüfungsart wird rechtzeitig zu Semesterbeginn bekannt gegeben.
Verwendbarkeit:	Dieses Modul ermöglicht Teilnehmenden ohne tiefergehende Vorkenntnisse in der Datenanalyse die Grundlagen für eine Teilnahme an Modul 2 und Modul 3 zu schaffen.

Lehrveranstaltung: Einführung in die Datenanalyse
EDV-Bezeichnung: DPBZ111
Dozent/in: Prof. Dr. Dennis Janka
Umfang (SWS): 6 h
Turnus: jährlich
Art und Modus: Vorlesung, Arbeit in Kleingruppen, Übung, Diskussion
Lehrsprache: deutsch
Inhalte: • Datengetriebene Anwendungsfälle im Unternehmenskontext • Strukturierte und unstrukturierte Daten, Taxonomie • Visualisierungen: z.B. Histogramme, Streuplots sowie deren Interpretation und Erstellung mit geeigneter Software • Statistische Kennzahlen (Lage- und Streumaße, Korrelation, Regressionsparameter), deren Interpretation und Berechnung mit geeigneter Software
Empfohlene Literatur: M. Cetinkaya-Rundel, J. Hardin: Introduction to Modern Statistics. OpenIntro, 2024. https://openintro-ims.netlify.app/
Anmerkungen: -

Lehrveranstaltung: Datenmanagement und -analyse – Von der Modellierung bis zur Visualisierung
EDV-Bezeichnung: DPBZ112
Dozent/in: Prof. Dr. Peer Küppers
Umfang (SWS): 6 h
Turnus: jährlich
Art und Modus: Vorlesung, Arbeit in Kleingruppen, Übung, Diskussion
Lehrsprache: deutsch
Inhalte: • Datenkonsistenz und -integrität • Konzeptionelle Datenmodellierung und Überführung in relationale Strukturen • Datenauswertungen mit Tabellenkalkulationssoftware, bspw. Excel (Umsetzung von Filter-, Gruppierungs-, Aggregations- und Verweisfunktionalitäten sowie Einführung in Pivot-Tabellen) • Datenauswertungen mit BI-Tools, bspw. PowerBI (Grundkonzepte)
Empfohlene Literatur: Kemper, A., Eickler, A. (2015). Datenbanksysteme: Eine Einführung. 10. Auflage. Berlin: De Gruyter. Schels, I. (2023). Business Intelligence mit Excel: Datenanalyse und Reporting mit Power Query, Power Pivot und Power BI Desktop. 3. Auflage. München: Carl Hanser Verlag.
Anmerkungen: -

Lehrveranstaltung: Text Data Literacy und Grundlagen Text Data Analytics
EDV-Bezeichnung: DPBZ113
Dozent/in: Prof. Dr. Jannik Strötgen
Umfang (SWS): 6 h
Turnus: jährlich
Art und Modus: Vorlesung, Arbeit in Kleingruppen, Übung, Diskussion
Lehrsprache: deutsch
Inhalte: • Was unterscheidet Textdaten von anderen Daten? • Herausforderungen bei der Verarbeitung von Textdaten (z.B. Unstrukturiertheit, Mehrdeutigkeit, Kontextabhängigkeit, Syntax und Semantik) • Texte als Wissensschatz (z.B. Bag-of-Words Modell, Metadaten und Kontext, Informationsextraktion) • Anwendungsbeispiele aus der Praxis (z.B. News Monitoring, Review Analysis, Sentiment Analysis)
• Sensibilisierung für Ambiguitäten in Sprache (z.B. Homonyme, Synonyme, Polysemie sowie Ironie, Sarkasmus und die Bedeutung von Domänenwissen) • Einfache Text Analytics Anwendung mit Hilfe von Häufigkeitsanalysen
Empfohlene Literatur: Einzelne Kapitel aus Dan Jurafsky and James H. Martin: Speech and Language Processing, 3rd edition (draft) https://web.stanford.edu/~jurafsky/slp3/
Anmerkungen: -

Modulname: Datenvorverarbeitung, -analyse und -visualisierung**Modulübersicht**

EDV-Bezeichnung: DPBZ120

Modulverantwortliche(r): Prof. Dr. Dennis Janka, Prof. Dr. Peer Küppers

Modulumfang (ECTS): 3 CP

18 h Vorlesung, 72 h Selbststudium, inkl. Seminar-, Prüfungsvorbereitung und Prüfung

Einordnung (Semester): 1

Inhaltliche Voraussetzungen:

Modul 1 oder äquivalente Kenntnisse

Voraussetzungen nach SPO:

keine

Fachkompetenzen:

Die Teilnehmenden erwerben ein vertieftes Verständnis für Unsicherheiten statistischer Kennzahlen. Sie besitzen nach Abschluss des Moduls ein konzeptionelles Verständnis für Techniken des maschinellen Lernens und sind in der Lage, statistische Vorhersagemodelle anzuwenden. Sie vertiefen ihre Kenntnisse in der relationalen Datenmodellierung und können Daten effizient verknüpfen. Darüber hinaus erwerben sie Grundkenntnisse in SQL und setzen entsprechende Datenanalysen praktisch um. Die Teilnehmenden arbeiten sicher mit Analysewerkzeugen (z.B. RapidMiner) für explorative Analysen und die Datenvorverarbeitung. Sie können komplexe, interaktive BI-Dashboards mit Power BI erstellen und dabei Hierarchien, Drill-Downs und DAX gezielt einsetzen.

Methodenkompetenzen:

Die Teilnehmenden sind nach Abschluss des Moduls in der Lage, die Qualität von Daten fachgerecht einzuschätzen. Sie verstehen die Grenzen und Unsicherheiten statistischer Aussagen. Sie wenden Techniken der Datenbereinigung und -transformation systematisch an, um Daten für Analysen optimal aufzubereiten. Zudem beherrschen sie die Anwendung von Workflows zur Automatisierung der Datenanalyse. Die Teilnehmenden können große Datenmengen verknüpfen und aggregieren, umstrukturierte und flexible Analysen durchzuführen. Sie setzen Datenvisualisierungstechniken gezielt ein, um Informationen klar und effektiv zu kommunizieren.

Sozial- und Selbstkompetenzen:

Die Teilnehmenden sind nach Abschluss des Moduls fähig, komplexe Datenanalysen klar zu präsentieren und fundiert zu interpretieren. Sie reflektieren kritisch die Qualität von Daten und erkennen deren Einfluss auf Entscheidungsprozesse. Sie arbeiten selbstständig mit Analyse- und BI-Tools und wenden diese gezielt auf betriebliche Anwendungsfälle an. Darüber hinaus entwickeln sie eine datengetriebene Denkweise zur Optimierung von Geschäftsprozessen.

Prüfungsleistungen:

Die theoretischen Kenntnisse der Studierenden werden in einer schriftlichen Klausur (Dauer 60 min, ggf. elektronisch in ILIAS), einer Projektarbeit oder in einer mündlichen Prüfung (Dauer 20 min) bewertet. Die Prüfungsart wird rechtzeitig zu Semesterbeginn bekannt gegeben.

Verwendbarkeit:

Das Modul Datenvorverarbeitung, -analyse und -visualisierung ist Grundlage für weiterführende Module im Bereich Datenvorverarbeitung, fortgeschrittene Datenanalyse und Datenvisualisierung.

Lehrveranstaltung: Fortgeschrittene Techniken der Datenanalyse
EDV-Bezeichnung: DPBZ121
Dozent/in: Prof. Dr. Dennis Janka
Umfang (SWS): 9 h

Turnus: jährlich
Art und Modus: Vorlesung, Arbeit in Kleingruppen, Übung, Diskussion
Lehrsprache: deutsch
Inhalte: • Stichprobenerhebung: Beobachtungsstudien und Experimente, Kausalität • Unsicherheiten statistischer Kennzahlen • Bootstrapping zur Berechnung von Unsicherheiten, Umsetzung in geeigneter Software • Konzept maschinelles Lernen, Beispiele, Überblick über verschiedene Modelle • Anwendung einfacher statistischer Modelle anhand ausgewählter Beispiele
Empfohlene Literatur: M. Cetinkaya-Rundel, J. Hardin: Introduction to Modern Statistics. OpenIntro, 2024. https://openintro-ims.netlify.app/
Anmerkungen: -

Lehrveranstaltung: Datenbanken, Datenvorverarbeitung und Visualisierung
EDV-Bezeichnung: DPBZ122
Dozent/in: Prof. Dr. Peer Küppers
Umfang (SWS): 9 h
Turnus: jährlich
Art und Modus: Vorlesung, Arbeit in Kleingruppen, Übung, Diskussion
Lehrsprache: deutsch
Inhalte: • Erweiterte Konzepte der Datenmodellierung und relationalen Datenhaltung (Normalisierung, Verknüpfung von Datenquellen, Joins, Abbildung komplexer Datenstrukturen in relationalen Systemen) • Einführung in SQL (Filterung, Gruppierung, Aggregation, Joins) • Anwendungsfall SQL: Analyse eines realistischen Datensatzes unter Verwendung eines Datenbankmanagementsystems (bspw. MySQL) • Datenanalyse mit einem Analytics-Tool (bspw. RapidMiner): Grundlagen Workflow-basierter Analysesysteme, explorative Analysen, Anwendung statistischer Methoden • Strategien zur Bereinigung und Transformation von Daten (Datenqualität, fehlende Werte, Ausreißererkennung, Handling von „imbalanced data“) • Datenvorverarbeitung und Umgang mit Zeitreihendaten • Erweiterte Visualisierungskonzepte mit einem BI-Tool, bspw. PowerBI • Ausblick Data Warehousing
Empfohlene Literatur: Kemper, A., Eickler, A. (2015). Datenbanksysteme: Eine Einführung. 10. Auflage. Berlin: De Gruyter. Schels, I. (2023). Business Intelligence mit Excel: Datenanalyse und Reporting mit Power Query, Power Pivot und Power BI Desktop. 3. Auflage. München: Carl Hanser Verlag. Shmueli, G., Bruce, P., Gedeck, P. und Patel, N. (2023). Machine Learning for Business Analytics: Concepts, Techniques, and Applications in RapidMiner. Hoboken, NJ: Wiley.
Anmerkungen: -

Modulname: Text Data Analytics**Modulübersicht**

EDV-Bezeichnung: DPBZ130

Modulverantwortliche(r): Prof. Dr. Jannik Strötgen

Modulumfang (ECTS): 3 CP

18 h Vorlesung, 72 h Selbststudium, inkl. Seminar-, Prüfungsvorbereitung und Prüfung

Einordnung (Semester): 1

Inhaltliche Voraussetzungen:
keineVoraussetzungen nach SPO:
keine**Fachkompetenzen:**

Die Teilnehmenden verstehen nach Abschluss des Moduls verschiedene Textdatenquellen und deren Eigenschaften. Sie können grundlegende Methoden der Textverarbeitung und -vorverarbeitung anwenden, wie Tokenisierung, Lemmatisierung und Stopword Removal. Sie beherrschen grundlegende Techniken der Textanalyse, wie Wortfrequenzanalysen, tf-idf und Ähnlichkeitsmaße. Zudem verfügen sie über Kenntnisse in Natural Language Processing (NLP)-Techniken, insbesondere in Named Entity Recognition und Relation Extraction. Die Teilnehmenden sind mit grundlegenden Methoden der Textklassifikation und des Textclustering vertraut. Sie verstehen zudem die wichtigsten rechtlichen und ethischen Aspekte der Textdatenverarbeitung, insbesondere im Hinblick auf Datenschutz und Bias.

Methodenkompetenzen:

Die Teilnehmenden sind nach Abschluss des Moduls in der Lage, Textdaten aus verschiedenen Quellen zu sammeln und für die Analyse vorzubereiten. Sie können Textdaten analysieren und visualisieren, z.B. durch Wortfrequenzanalysen, Word Clouds und Ähnlichkeitsmaße. Sie wenden grundlegende Techniken der Textvorverarbeitung, wie Tokenisierung, Stemming und Lemmatisierung, sicher in der Praxis an. Zudem besitzen sie die Fähigkeit, NLP-Techniken zur Extraktion strukturierter Informationen anzuwenden, darunter Named Entity Recognition und Temporal Tagging. Die Teilnehmenden erstellen und nutzen einfache Textklassifikatoren und Textclustering-Anwendungen (z.B. Sentiment Analysis, Topic Modeling). Sie setzen geeignete Tools und Software, wie RapidMiner oder Python, gezielt zur Textanalyse und -verarbeitung ein.

Sozial- und Selbstkompetenzen:

Die Teilnehmenden können nach Abschluss des Moduls Ergebnisse aus der Textanalyse verständlich und zielgruppengerecht präsentieren. Sie reflektieren kritisch die Prozesse der Textanalyse und können mögliche Fehlerquellen oder Bias in den Daten erkennen und adressieren. Sie wenden Methoden und Tools zur Analyse von Textdaten eigenständig an, um betriebliche oder wissenschaftliche Fragestellungen zu lösen. Darüber hinaus gehen sie verantwortungsvoll mit Daten um und berücksichtigen ethische sowie datenschutzrechtliche Aspekte in der Praxis.

Prüfungsleistungen:

Die theoretischen Kenntnisse der Studierenden werden in einer schriftlichen Klausur (Dauer 60 min, ggf. elektronisch in ILIAS), einer Projektarbeit oder in einer mündlichen Prüfung (Dauer 20 min) bewertet. Die Prüfungsart wird rechtzeitig zu Semesterbeginn bekannt gegeben.

Verwendbarkeit:

Das Modul Text Data Analytics ist Grundlage für weiterführende Module im Bereich Advanced Text Data Analytics oder Natural Language Processing.

Lehrveranstaltung: Text Data Analytics
EDV-Bezeichnung: DPBZ131
Dozent/in: Prof. Dr. Jannik Strötgen
Umfang (SWS): 18 h
Turnus: jährlich
Art und Modus: Vorlesung, Arbeit in Kleingruppen, Übung, Diskussion
Lehrsprache: deutsch
Inhalte: • Einführung in Text Data Analytics und Überblick über Einsatzbereiche und Anwendungsfälle (z.B. Kundenservice, Fake-News-Erkennung) • Arten von Textdaten (z.B. Nachrichtentexte, Soziale Medien, Emails) • Datenbeschaffung (manuelles Sammeln vs. automatisierte Scraping-Techniken) und erstes Verarbeiten von Textdaten mit geeigneten Tools und Datenformaten (z.B. txt, xml, json) • Grundlagen Textvorverarbeitung (z.B. Satzsegmentierung, Tokenization, Stemming, Lemmatization, Stop Word Removal) und Umsetzung mit geeigneten Tools (z.B. RapidMiner oder Python-Bibliotheken) • Einfache Techniken zur Textanalyse wie Wortfrequenz-Analyse (z.B. Word Clouds, Term-Häufigkeiten), Relevanzbewertung von Wörtern (z.B. tf-idf) • Texte als mathematische Vektoren (z.B. Vector Space Model) und Ähnlichkeitsmaße für Textdaten (z.B. Cosine Similarity) • Grundlagen des Natural Language Processing (z.B. Named Entity Recognition, Relation Extraction, Temporal Tagging) • Text Analytics Methoden für Textklassifikation (z.B. Sentiment Analyse) und Textclustering (z.B. Topic Modelling) • Ethik und Datenschutz (z.B. rechtliche Aspekte bei der Verarbeitung von Textdaten sowie ethische Überlegungen und Bias in Textdaten)
Empfohlene Literatur: Einzelne Kapitel aus Dan Jurafsky and James H. Martin: Speech and Language Processing, 3rd edition (draft) https://web.stanford.edu/~jurafsky/slp3/
Anmerkungen: -