Big Data Masterclass

Big Data & Datenbanken · Anbieter: IU Akademie IU Internationale Hochschule

FÖRDERUNG & KOSTEN

Kurspreis: auf Anfrage

Förderfähig über: Bildungsgutschein

AZAV-zertifizierter Bildungsträger — Förderung durch die Agentur für Arbeit möglich

Kursdaten auf einen Blick

Durchführung	Online
Lernformat	Self-Paced
Dauer	8 Monate in Vollzeit oder 16 Monate in Teilzeit
Unterrichtssprache	Englisch
Kursstart	Startet alle 14 Tage
Vorkenntnisse	Keine Vorerfahrung
Abschluss	Zeugnis auf Hochschulniveau
Inklusive	Berufscoaching, Notebook / Tablet (geliehen)

Bildungsträger & Kontakt

IU Akademie IU Internationale Hochschule
Juri-Gagarin-Ring 152, 99084 Erfurt
Telefon: 030 22027283
E-Mail: weiterbildung@iu-akademie.de
Web: https://www.iu-akademie.de

Kurs online ansehen

01 Kursbeschreibung

Unsere digitalisierte Welt erzeugt eine enorme Menge an Daten, die von Unternehmen und Organisationen besser genutzt werden müssen. Denn sie geben wertvolle Einblicke und neue Erkenntnisse, die wichtig sind, um den Wettbewerbsvorteil zu sichern. Mit Deiner Online-Weiterbildung Big Data Masterclass hast Du die Daten im Griff – und im Blick. Deine Weiterbildung macht Dich fit für die grundlegenden Konzepte und Techniken der Datenanalyse. Du lernst, wie man Daten sammelt, bereinigt und strukturiert, um sie für die weiterführende Analyse vorzubereiten. Dazu erkundest Du verschiedene Datentypen und -formate und eignest Dir Tools und Technologien an, die zur Verarbeitung großer Datenmengen verwendet werden. Kenntnisse über Hadoop, Apache Spark, NoSQL-Datenbanken und andere Big Data-Frameworks sind ebenfalls mit dabei. Und Du wirst Daten mithilfe von Visualisierungstechniken wie Diagramme, Grafiken und Dashboards effektiv präsentieren können. Im Bereich Machine Learning erfährst Du, wie Du Algorithmen trainierst, um Vorhersagen und Mustererkennung durchzuführen. Mit dem Data Engineering wirst Du Dich ebenfalls intensiv beschäftigen. Damit Du später in der Lage bist, Datenpipelines zu entwerfen und zu entwickeln, komplexe Datenprobleme zu lösen und die Datenqualität zu gewährleisten.

02 Lehrplan & Module

1. Introduction to Data Science

Data science emerged as a multi-disciplinary field aimed at creating value from data. This course starts with an overview of data science and related fields and then defines data types and sources. Special focus is put on the assessment of data quality and electronic data processing.
Use of data-driven methods has become vital for businesses, and this course outlines how data-driven approaches can be integrated within a business context and how operational decisions can be made using data-driven methods.
Finally, this course highlights the importance of statistics and machine learning in the field of data science and gives an overview of relevant methods and approaches.
Kursinhalte
Introduction to Data Science Definition of the term „data science“
Data science and related fields
Data science activities
Data Data types and data sources
The 5Vs of data
Data curation and data quality
Data engineering
Data Science in Business Identification of use cases
Performance evaluation
Data-driven operational decisions
Cognitive biases
Statistics Importance of statistics for data science
Important statistical concepts
Machine Learning Role of machine learning in data science
Overview of machine learning approaches
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Workbook

2. Data Quality and Data Wrangling

The goal of data science can be summarized as the extraction of insights (hence, value) from data. It is self-evident that this objective cannot be successfully achieved based on unreliable and untrustworthy data. This course aims at establishing the notion of data quality and the pertinent methods for data quality management. Furthermore, techniques for acquiring data as well as formatting and tidying data in order to make it suitable for subsequent analytical treatment are covered.
Kursinhalte
Data Quality Introduction to data quality
Data quality dimensions and issue types
Data Quality Management Data governance and stewardship
Activities and processes
Data Acquisition Web scraping
Data APIs
Working with Common Data Formats Text-based formats (CSV, XML, JSON)
Binary formats (HDF 5, Parquet, Arrow)
Tidy Data Structuring
Cleansing
Enrichment
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Hausarbeit

3. Data Science Software Engineering

A core part of data science is creating value from data. This means not only the creation of sophisticated predictive models but also the development of these models according to modern software development principles.
This course gives a detailed overview of the relevant methods and paradigms which data scientists need to know in order to develop enterprise-grade models.
This course discusses traditional and agile project management techniques, highlighting both the Kanban and Scrum approaches. It explores relevant software development paradigms such as test-driven development, pair programming, mob programming, and extreme programming.
Special focus is given to the topic of testing and the consideration of how to bring a model into a production environment.
Kursinhalte
Traditional Project Management Requirements engineering
Waterfall model
Rational unified process
Agile Project Management Criticism of the waterfall model
Introduction to SCRUM
Introduction to Kanban
Testing Why testing?
Unit trests
Integration tests
Performance monitoring
Software Development Paradigms Test-driven development (TDD)
Pair programming
Mob programming
Extreme programming
From Model to Production Continuous delivery
Continuous integration
Building a scalable environment
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Examen, 90 Minuten

4. Project: From Model to Production Environment

This project course will give students hands-on experience in the challenging task of bringing a predictive model into a production environment. Students will need to consider practical aspects such as data storage and processing, as well as constraints such as service availability and the maximum amount of time a model is allowed to run due to external project requirements.
Through this course, students will obtain holistic overview of the integration of predictive models into enterprise-grade applications or services.
Kursinhalte
This project course focuses on practical aspects of ensuring that a predictive model can run in a production environment. The students start with a chosen use case and model and then evaluate the requirements which need to be fulfilled so that the model can be used as part of an enterprise application or app. Students need to evaluate requirements in terms of data storage, processing and throughput, and availability of the service, as well as the persistency, serving, and versioning of the model itself. Monitoring the execution of model predictions and raising alerts in cases of operational issues is a core part of building a reliable model pipeline. All relevant artifacts and considerations are documented by the students in a project report.
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Projektbericht

5. Big Data Technologies

Data are often considered the “new oil”, the raw material from which value is created. To harness the power of data, the data need to be stored and processed on a technical level. This course introduces the four “Vs” of data, as well as typical data sources and types.
The course discusses the most common data storage formats encountered in modern systems, focusing both on text-based as well as binary data formats.
Handling large amounts of data poses significant challenges for the underlying infrastructure. The course discusses the most important distributed and streaming data handling frameworks which are used in leading edge applications.
Kursinhalte
Data Types and Data Sources The 4Vs of data: volume, velocity, variety, veracity
Data sources
Data types
Working with Common Data Formats Text-Based Formats (CSV, XML, JSON)
Binary Formats (HDF5, Parquet, Arrow)
NoSQL data stores Introduction and motivation
Approaches and technical concepts
Distributed Systems Hadoop & MapReduce
Hadoop file system (HDFS)
Spark
DASK
Streaming Frameworks Spark streaming
Kafka
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Examen, 90 Minuten

6. Cloud Computing

Many of the recent advances in data science, particularly machine learning and artificial intelligence, rely on comprehensive data storage and computing power. Cloud computing is one way of providing that power in a scalable way, without considerable upfront investment in hardware and software resources. This course introduces the area of cloud computing together with its enabling technologies. Moreover, the most cutting-edge advances like serverless computing and storage are illustrated. Finally, a thorough overview on popular cloud offerings, especially in regard to analytics capabilities, is given.
Kursinhalte
Introduction to Cloud Computing Fundamentals of Cloud computing
Cloud Service Models
Benefits and Risks
Enabling Technology Virtualization and Containerization
Storage Technology
Networks and RESTful Services
Serverless Computing Introduction to Serverless Computing
Benefits
Limitations
Established Cloud Platforms General Overview
Google Cloud Platform
Amazon Web Services
Microsoft Azure
Platform Comparison
Data Science in the Cloud Provider-independent services and tools
Google Data Science and Machine Learning Services
Amazon Web Services Data Science and Machine Learning Services
Microsoft Azure Data Science and Machine Learning Services
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Examen, 90 Minuten

7. Data Engineering

This course explores concepts of data engineering. Data engineering is concerned with the infrastructure aspects of data science such as data storage and provision, as well as the provisioning of suitable operational environments.
After laying out foundational notions and concepts of the discipline, this course addresses important developments in storage technology; aspects of systems architecture for processing data at scale; containerization as a modern take on virtualization; and the logic of data pipelines and associated operational aspects. Important issues pertaining to data security and protection are also given appropriate attention.
Kursinhalte
Foundations of Data Engineering Reliability
Scalability
Maintainability
NoSQL In Depth Fundamentals of NoSQL
Established NoSQL solutions
Architectures for Data Processing at Scale Batch processing architectures
Architectures for stream and complex event processing
Lambda architecture
Containerization In Depth Docker containers
Container management
Governance & Security Data protection
Data security
Data governance
Operational Aspects Defining principles of DataOps
Building and maintaining data pipelines
Metrics and monitoring
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Examen, 90 Minuten

8. Projekt: Data Engineering

The focus of this course is the implementation of a real-world data engineering use case in the form of a student portfolio.
To this end, students choose a project subject from the various sub-domains of data engineering. Examples include setting up a Docker container environment or dockerized service; implementing a data pipeline according to DataOps principles; and setting up an NoSQL data store.
The goal is for students to demonstrate they can transfer theoretical knowledge to an implementation scenario that closely mimics practical work in a professional data engineering setting.
Kursinhalte
This course covers the practical implementation of approaches and techniques covered in the preceding methodological course in a project-oriented setting. Each participant must produce a portfolio detailing and documenting the work. Porfolio themes are chosen from a list, or suggested by the students in accord with the tutor.
Fakten zum Modul: Niveau: Bachelor · Unterrichtssprache: English · 5 ECTS-Punkte · Modulprüfung: Portfolio

03 Enthaltene Zertifikate

Trägerzertifikat: Zeugnis auf Hochschulniveau

04 Voraussetzungen

Vorkenntnisse

Niveau: Keine Vorerfahrung

Zulassungsvoraussetzung: gute Deutschkenntnisse auf B2-Niveau.

Zugangsempfehlungen: Englisch auf B2 Niveau; Python and SQL Programming oder gleichwertige Vorkenntnisse; Python Programming Basics oder gleichwertige Vorkenntnisse

Die Weiterbildung ist auf dem inhaltlichen Niveau eines Bachelor Studiengangs. Eine Weiterbildung auf Bachelor-Niveau vermittelt grundlegende Kenntnisse und Fähigkeiten in einem bestimmten Fachbereich.

Technische Voraussetzungen

Laptop mit Internetzugang.

05 Zielgruppen & Berufsperspektiven

- Arbeitssuchende
- Fachkräfte

Mögliche Berufsbilder: Data Scientist, Data Engineer, Data Analyst

NÄCHSTER SCHRITT

Kursdetails online ansehen und Anbieter kontaktieren

IU Akademie IU Internationale Hochschule · Juri-Gagarin-Ring 152, 99084 Erfurt
030 22027283 · weiterbildung@iu-akademie.de · https://www.iu-akademie.de

Online ansehen