



Andreas Henke

andreas.henke@mouse-klick.de

geboren am: 28. Februar 1974

hat erfolgreich den sechswöchigen Kurs bestanden:

Data Engineering und Data Science-Klarheit in den Schlagwort-Dschungel

von Prof. Dr. Felix Naumann

Der Kurs bestand aus den folgenden Einheiten

- Big Data und Data Science
- Skalierbares Datenmanagement
- Datenaufbereitung
- Informationsintegration
- Data Mining
- Text Mining
- Machine Learning

Gesamtergebnis

160,1 von 180,0 möglichen Punkten (88,9%)

Prof. Dr. Christoph Meinel
openHPI Direktor

Prof. Dr. Felix Naumann
Dozent



openHPI ist die Internet-Bildungsplattform des Hasso-Plattner-Instituts (HPI) für Digital Engineering gGmbH. Das HPI bildet gemeinsam mit der Universität Potsdam die Digital Engineering Fakultät der Universität Potsdam.

Potsdam, 23. Februar 2020

Online Verifizierung: <https://open.hpi.de/verify/xosel-bonyv-pedyd-bicid-ruzar>



openHPI Massive Open Online Course (7 Wochen)		
Lehrform	E-Learning-Angebot bestehend aus Video-Lektionen, Selbsttests, Lesematerial, Hausaufgaben, Abschlussprüfung	
Umfang des Kurses	2,5 SWS; 6 Kurswochen, 1 Prüfungswoche	
Leistungspunkte	2	
Kumulierung der ECTS Punkte mehrerer Kurse	Für 3 absolvierte Massive Open Online Kurse vergibt das Hasso-Plattner-Institut 6 ECTS Punkte	
Inhalte	Inhalte wie auf dem Zertifikat ausgewiesen; Ausführliche Inhaltsbeschreibung auf der Kursdetailseite des jeweiligen Kurses unter https://open.hpi.de/courses	
Prüfungen (Anzahl, Umfang)	7 (6 bewertete wöchentliche Hausaufgaben, 1 Abschlussklausur)	
Verifizierung der Prüfungsleistung	Überwachung der Prüfungsleistung per Webcam; mittels maschineller Gesichtserkennung wird überprüft, dass die Person, welche die Prüfungen ablegt, mit dem Kursanmelder übereinstimmt; das Foto der Person, die die Prüfung abgelegt hat, ist auf dem Zertifikat abgedruckt	
Teilnahmevoraussetzungen	Keine	
Prozentpunkte	Note als Zahl	Note im Zeugnis
100 bis 95	1,0	Sehr gut
94 bis 91	1,3	Sehr gut
90 bis 85	1,7	Gut
84 bis 80	2,0	Gut
79 bis 76	2,3	Gut
75 bis 70	2,7	Befriedigend
69 bis 65	3,0	Befriedigend
64 bis 61	3,3	Befriedigend
60 bis 55	3,7	Ausreichend
54 bis 50	4,0	Ausreichend
Unter 50	5,0	Nicht ausreichend

Data Engineering und Data Science - Klarheit in den Schlagwort - Dschungel

Von Prof. Dr. Felix Naumann

8 Januar 2020 - 26 Februar 2020

Kursbeschreibung:

Die Schlagwörter Künstliche Intelligenz, Data Science, Data Engineering, und Big Data dominieren seit einigen Jahren nicht nur die IT-Schlagzeilen. In unserem Kurs wollen wir diese Wörter mit grundlegendem Inhalt füllen und die typischen Arbeitsschritte eines Data Scientists nachvollziehen. Insbesondere schauen wir hinter die Kulissen und betrachten den oft mühsamen Weg der Daten bis sie endlich genutzt werden können um z.B. mittels maschinellem Lernen Modelle trainieren zu können. Dazu gehören die Datenbeschaffung, die Datenreinigung, und die Datenintegration. Anschließend lernen wir, wie man aus diesen Daten und auch aus Texten neue Erkenntnisse mittels Data Mining und maschinellem Lernen gewinnt. Der Abschluss bildet eine Diskussion über Ethik und Fairness bei der automatisierten Datenanalyse.

Aufbau des Kurses:

Woche 1: Big Data und Data Science

- Relationale DBMS, SQL, Data Warehouses, Verteilte DBMS, NoSQL & NewSQL Systeme, Big Data, Datenquellen

Woche 2: Data Science Anwendungen und Text Mining

- Daten-getriebene Anwendungen, Data Science, Recommender Systeme, Künstliche Intelligenz, Datenkompetenz, Ethik und Privatsphäre, Text Mining, Computerlinguistik

Woche 3: Skalierbares Datenmanagement

- OLTP und OLAP, Parallelisierung, Verteilte Dateisysteme, Batch- und Datenstromverarbeitung, Map/Reduce & Hadoop, Datenpartitionierung, Cloud

Woche 4: Datenaufbereitung

- Informationsqualität, Dimensionen und Auswirkungen schlechter Datenqualität, Datenvandalismus, Data Preparation, Datenreinigung, Duplikaterkennung, Datenfusion

Woche 5: Informationsintegration

- Autonomie und Heterogenität, Schema Matching & Mapping, Data Warehouses, Business Intelligence, Data Lakes, Datenherkunft, Virtuelle Integration, Deep Web

Woche 6: Statistik, Data Mining, Machine Learning

- Visualisierung, Risikokompetenz, Simpson-Paradoxon, deskriptive vs. prädiktive Analyse, Assoziationsregeln, Clustering, Klassifizierung, Entscheidungsbäume, Neuronale Netzwerke, Deep Learning, Fairness, Erklärbare KI