

ÜBUNGSKLAUSUR

**Lehrveranstaltung:
Big Data**

Kurs:

Datum:

Dozent/in: Dr. Holger Stephan

Erlaubte Hilfsmittel: keine

Dauer: 30 min

max. zu erreichende Punktzahl: 30

Hinweise:

- Überprüfen Sie bitte zunächst, ob Sie alle 3 Seiten (inkl. dieses Deckblattes) vorliegen haben.
- Schreiben Sie auf jedes Lösungsblatt Ihre Matrikelnummer. Bitte schreiben Sie **nicht Ihren Namen**, sondern nur Ihre Matrikelnummer auf die Klausur!
- Schreiben Sie nur mit einem dokumentenechten Stift (kein Bleistift, kein Rotstift!).
- Lassen Sie bitte auf der rechten Seite 3 cm Rand.

Matrikelnummer:

Aufgabe	1	2	3	4	5	6	Σ
erreichbar	3	3	8	10	3	3	30
Punkte							

Aufgabe 1: Typen von Data Analytics
3 Pkte

Was ist das Besondere an Big Data? Nennen Sie die 5 V.

Aufgabe 2: Typen von Data Analytics
3 Pkte

Je nach Fragestellung unterscheidet man zwischen Descriptive (1), Diagnostic (2), Predictive (3) und Prescriptive (4) Data Analysis. Ordnen Sie zu!

	(1)	(2)	(3)	(4)
Wie viel Personal brauchen wir nächstes Wochenende?				
Wieviel von Produkt x haben wir 2021 verkauft?				
Warum sind die Rohstoffpreise 2019 gestiegen?				
Was war die Niederschlagsmenge letztes Jahr?				
Welches neue Produkt verspricht den höchsten Gewinn?				
Wann muss die Pumpe das nächste Mal gewartet werden?				

Aufgabe 3: BD-Storage und -Processing
8 Pkte

Big Data wird oft in so genannten NoSQL-Datenbanken abgespeichert.

- a) Nennen Sie vier Untertypen von NoSQL-Datenbanken.
(nur den Namen, keine Handelsprodukte) und beschreiben Sie sie kurz.
- b) Sie müssen eine große Menge an Sensordaten bestehen aus Temperaturwerten und Bildern abspeichern. Welchen Typ von NoSQL Datenbank wählen Sie? Warum?

Aufgabe 4: Machine Learning
10 Pkte

- a) Was versteht man unter „Unsupervised Learning“?
Beschreiben Sie die Methode, was sind die Voraussetzungen (besonders im Vergleich zu anderen ML Typen), geben Sie ein praktisches Beispiel. (mind. 4 ganze Sätze)
- b) Unter Deep Learning versteht man Maschinelles Lernen mit Hilfe von Neuronalen Netzen. Warum heißt es „Neuronal“? Wie lernt ein Neuronales Netz?
- c) Welche Gefahren sehen Sie, wenn unvollständige Datensätze zum Training einer KI verwendet werden?

Aufgabe 5: Visualisierung

3 Pkte

Was sind die Gründe, warum man Daten (gut) visualisieren sollte?

Aufgabe 6: Governance

3 Pkte

Warum ist es wichtig, einen Data Owner zu haben?