

Aufgabenheft zu

Klausur: Modul 32711 – Business Intelligence

Termin: 15.09.2017, 11:30–13:30 Uhr

Prüfer: Univ.-Prof. Dr. U. Baumöl



Bitte lesen Sie die folgenden Hinweise sorgfältig, bevor Sie mit der Bearbeitung der Klausur beginnen!

Aufbau und Umfang der Klausur

Die Klausur umfasst vier Aufgaben mit der folgenden Punkteverteilung:

Aufgabe	1	2	3	4	Summe
Maximal erreichbare Punktzahl	20	35	35	10	100

Für die Bearbeitung dieser Klausur stehen Ihnen 120 Minuten zur Verfügung. Die Klausur besteht aus einem *Aufgabenheft* und einem *Lösungsbogen*. Die Aufgaben umfassen 6 Seiten. Bitte prüfen Sie Ihre Klausur auf Vollständigkeit.

Hinweise zur Verwendung des Lösungsbogens

1. Tragen Sie auf dem Deckblatt des Lösungsbogens Ihre Matrikelnummer sowie Ihren Namen und Vornamen ein. Versehen Sie zusätzlich jedes Lösungsblatt mit Ihrer Matrikelnummer.
2. Die Lösungen müssen in den vorgesehenen Raum auf dem Lösungsbogen eingetragen werden. Lösungen außerhalb des vorgesehenen Raumes werden nicht in die Bewertung einbezogen.
3. Markieren Sie bei Multiple-Choice Aufgaben die von Ihnen gewählten Alternativen durch ein Kreuz (X) an der dafür vorgesehenen Stelle im Lösungsbogen.
4. Sorgen Sie für eindeutige Eintragungen im Lösungsbogen.
5. Notizen können auf den Rückseiten der Aufgabenblätter gemacht werden. Diese Anmerkungen werden nicht in die Bewertung einbezogen.
6. Sie dürfen das Aufgabenheft vom Lösungsbogen trennen. Das Aufgabenheft muss nicht abgegeben werden. Es müssen jedoch am Ende der Klausur *alle* Blätter des Lösungsbogens *zusammengeheftet* abgegeben werden. Trennen Sie keine einzelnen Blätter ab.

Hinweise zur Bewertung der Aufgaben

Jede vollständig richtig gelöste Aufgabe oder Teilaufgabe wird mit der an Ort und Stelle angegebenen Punktzahl bewertet.

Für die Aufgabe 1 (Einfach-Wahlaufgaben 1 aus n) gilt: Es darf nur *ein* Kreuz pro Teilaufgabe gesetzt werden. Richtig gelöste Teilaufgaben werden mit der angegebenen Punktzahl bewertet. Nicht oder falsch beantwortete Teilaufgaben werden mit Null Punkten bewertet.

Für die Aufgabe 4 (Richtig – Falsch) gilt: Richtig gelöste Teilaufgaben werden mit der anteiligen Punktzahl bewertet. Nicht oder falsch beantwortete Teilaufgaben werden mit Null Punkten bewertet.

Für die Aufgaben 2 und 3 gilt: Teilweise richtig gelöste Aufgaben oder Teilaufgaben können mit einer entsprechend verminderten Punktzahl bewertet werden.

Hinweise zur Verwendung eines Taschenrechners

Die Verwendung eines Taschenrechners ist dann und nur dann erlaubt, wenn dieser einer der folgenden Modellreihen angehört:

- Casio fx86 oder Casio fx87
- Texas Instruments TI 30 X II
- Sharp EL 531

Die Verwendung anderer Taschenrechnermodelle wird als Täuschungsversuch gewertet und mit der Note „nicht ausreichend“ (5,0) sanktioniert.

Ob ein Taschenrechner einer der Modellreihen angehört, können Studierende selbst überprüfen, indem sie die vom Hersteller auf dem Rechner angebrachte Modellbezeichnung mit den oben angegebenen Bezeichnungen vergleichen: Bei *vollständiger* Übereinstimmung ist das Modell erlaubt. Ist die auf dem Rechner angebrachte Modellbezeichnung umfangreicher, enthält aber eine der oben angegebenen Bezeichnungen *vollständig*, ist das Modell ebenfalls erlaubt. In allen anderen Fällen ist das Modell nicht erlaubt. *Eventuelle Vorgänger- oder Nachfolgemodelle, die nicht in der oben aufgeführten Liste enthalten sind, sind ebenfalls nicht erlaubt.*

Aufgabe 1**(20 P)**

Markieren Sie im Lösungsbogen die zutreffende Aussage. Es ist nur eine Aussage pro Teilaufgabe korrekt, d. h. Sie müssen sich für *eine* Aussage entscheiden und dürfen nur *ein* Kreuz setzen.

- 1.1 „Getting data in and getting data out“ steht sinnbildlich für eine einseitige Fokussierung der Business Intelligence auf das technische Datenmanagement. Welche der folgenden Fragen stellt sich bezüglich „getting data in“? (5 P)
- a) Wie kann der Datenzugriff möglichst nutzerfreundlich gestaltet werden?
 - b) Welche Zugriffskonzepte machen den Zugriff schneller?
 - c) Wie sollen die aufbereiteten Daten zur Verfügung gestellt werden?
 - d) Welche Datenquellen müssen wie angeschlossen werden?
- 1.2 Der subjektive Informationsbedarf der Beteiligten einer Entscheidung entspricht nicht immer dem objektiven Informationsbedarf, den das zugrundeliegende Entscheidungsproblem bedingt. Was ist nach dem Modell von *Picot et al.* (2003) *nicht*, auch nicht unter ansonsten idealisierten Bedingungen, möglich, wenn eine solche Abweichung von subjektivem und objektivem Informationsbedarf gegeben ist. (5 P)
- a) Die Informationsnachfrage entspricht dem subjektiven Informationsbedarf.
 - b) Der Informationsstand entspricht dem objektiven Informationsbedarf.
 - c) Das Informationsangebot wird vollständig ausgeschöpft.
 - d) Das Informationsangebot entspricht dem objektiven Informationsbedarf.

1.3 Welches ist *keine* Dimension des Comprehensive Decision Model? (5 P)

- a) Akteur
- b) Organisation
- c) Applikationen
- d) Geschäftsmodell

1.4 Welches ist *kein* Element der Dimension „Kunde“ in der Business Model Ontology? (5 P)

- a) Vertriebskanäle
- b) Kundenprozess
- c) Wettbewerbsvorteile
- d) Kundenbeziehung

Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!

Aufgabe 2**(35 P)**

Clusteranalyse wird, z. B. im Marketing, zur Bildung von Kundensegmenten verwendet. Aus dem Lehrbrief wissen Sie, dass das Ziel einer Clusteranalyse die Gruppierung von Instanzen gemäß objektiv nachvollziehbarer Kriterien ist. Eine wichtige Rolle bei einigen Verfahren dazu spielen die Distanzfunktion $dist$ und die Distanzmatrix \mathbf{D} . Für zwei Instanzen x_i, x_j und $j < i$ wird $\mathbf{D}_{i,j} = dist(x_i, x_j)$ gesetzt. Tabelle 1 zeigt ein Beispiel für eine Distanzmatrix.

Tabelle 1: Beispielhafte Distanzmatrix

	x_1	x_2	x_3	x_4	x_5
x_1	0	0	0	0	0
x_2	14	0	0	0	0
x_3	6	9	0	0	0
x_4	13	8	15	0	0
x_5	12	4	11	1	0

Hierarchisch-agglomerative Clusteringalgorithmen basieren darauf, aus einer Distanzmatrix von *Instanzen* sukzessive Distanzmatrizen von *Clustern* zu berechnen. Dazu haben Sie im Lehrbrief ein generisches Verfahren kennengelernt, das sich verschiedener sogenannter linkage-Methoden bedient. Tabelle 2 zeigt ein Beispiel für Zwischenergebnisse nach jedem Schritt eines hierarchisch-agglomerativen Clusterings auf einer hier nicht gezeigten Sequenz von Distanzmatrizen. *Hinweis: es handelt sich nicht um das Clustering zur Distanzmatrix in Tabelle 1!*

Tabelle 2: Teilergebnisse einer Durchführung von hierarchisch-agglomerativem Clustering

Schritt	kleinste Distanz	Cluster
0	1	$\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}$
1	2	$\{x_1\}, \{x_2, x_4\}, \{x_3\}, \{x_5\}$
2	4	$\{x_1\}, \{x_2, x_4, x_5\}, \{x_3\}$
3	7	$\{x_1, x_3\}, \{x_2, x_4, x_5\}$
4	-	$\{x_1, x_2, x_3, x_4, x_5\}$

Bearbeiten Sie auf Grundlage obiger Ausführungen die folgenden Aufgaben:

- 2.1 Nennen Sie umgangssprachlich die fünf Schritte, die zum hierarchisch-agglomerativen Clustering durchgeführt werden müssen! (10 P)
- 2.2 Führen Sie hierarchisch-agglomeratives Clustering nach dem single-linkage Verfahren für die Distanzmatrix in Tabelle 1 durch! Geben Sie in jedem Schritt die bis dahin erzeugten Cluster, die kleinste Distanz und die neue Distanzmatrix an! Weitere Angaben sind nicht erforderlich. (15 P)
- 2.3 Zeichnen Sie das Dendrogramm, welches sich aus dem beispielhaften Clustering in Tabelle 2 ergibt! (10 P)

Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!

Aufgabe 3**(35 P)**

Um aus Daten Informationen und letztlich Wissen generieren zu können, sind mehrere Schritte notwendig. Das Laden der Daten in das Data Warehouse (DWH) bildet dabei einen elementaren Schritt zwischen der Sammlung der Daten und der Analyse.

- 3.1 Ein DWH besteht aus verschiedenen Komponenten. Ein Beispiel für eine Komponente ist die Basisdatenbank. Diese ist der Kern des DWH und erfüllt u. a. die Distributionsfunktion des DWH. Nennen Sie fünf weitere Komponenten eines DWH! Nennen Sie zu jeder Komponente eine Funktion, und erläutern Sie diese in ein bis zwei Sätzen! (10 P)
- 3.2 DWHs sind dispositive Datenhaltungssysteme und werden von operativen Datenhaltungssystemen separiert betrachtet (Kemper, Baars und Mehanna, 2010). Verdeutlichen Sie diese Abgrenzung, indem Sie die Eigenschaften der Daten in einer Abbildung oder Tabelle einander gegenüber stellen! Betrachten Sie hierzu die Eigenschaften der Daten strukturiert nach Dimensionen! Erläutern Sie die Abbildung, indem Sie Bezug zu den folgenden Eigenschaften des DWH nach *Inmon* (2000) herstellen: Themenorientierung, Integration, Zeitraumbezug und Nicht-Volatilität. (15 P)
- 3.3 Nehmen Sie kritisch zu folgender These Stellung. Nennen Sie dazu Pro- und Contra-Argumente, und formulieren Sie abschließend Ihre eigene Meinung zu der These: „Der Einsatz von Data Marts ist ineffizient im Vergleich zum Einsatz eines zentralen DWH“! (10 P)

Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!

Aufgabe 4**(10 P)**

Überprüfen Sie die folgenden Aussagen auf ihre Richtigkeit. Kennzeichnen Sie im Lösungsbogen *uneingeschränkt* zutreffende Aussagen mit einem Kreuz bei „Richtig“ und bei allen anderen Aussagen mit einem Kreuz bei „Falsch“.

- 4.1 Die Datenlatenz ist die Zeit, die nach Eintreten eines Ereignisses benötigt wird, um die das Ereignis betreffenden Daten für die weitere Analyse bereitzustellen.
- 4.2 Die Analyselatenz beinhaltet die Zeit für die Analyse und Bereitstellung der Daten sowie für die daraus folgende Entscheidung und Handlung.
- 4.3 Es wird angenommen, dass eine Information anfänglich umso wertvoller ist, je länger ihre Halbwertszeit ist.
- 4.4 Vorausschauende Analysemöglichkeiten in BI-Systemen versprechen prinzipiell negative Latenzen.
- 4.5 Mit der zunehmenden Verkürzung von Latenzzeiten sinken auch die mit der Verkürzung verbundenen Kosten.
- 4.6 Advanced Analytics beinhalten Verfahren zur Integration von Daten und Inhalten sowie Predictive Analytics.
- 4.7 Die strukturierte Sortierung von Wörtern im Text Mining wird Bag-of-Words genannt.
- 4.8 Die Sammlung von Dokumenten, auf die sich eine Analyse im Text Mining bezieht, wird Korpus genannt.
- 4.9 Im Text Mining wird eine unstrukturierte Repräsentation eines Dokuments in eine strukturierte Repräsentation als Vektor überführt.
- 4.10 In-Memory Analytics ist ein Ansatz, der durch die stark fallenden Hardwarepreise möglich wurde.

Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!

Matrikelnr.

--	--	--	--	--	--	--

Name _____

Vorname _____

Lösungsbogen

Klausur: Modul 32711 - Business Intelligence

Termin: 15.09.2017 11:30 – 13:30 Uhr

Prüfer: Univ.-Prof. Dr. U. Baumöl

Aufbau und Bewertung der Klausur

Aufgabe	1	2	3	4	Summe
Maximal erreichbare Punktzahl	20	35	35	10	100
Erreichte Punktzahl					

Datum:

Note:

Unterschrift des Prüfers

--	--	--	--	--	--	--

Aufgabe 1 (ankreuzen)

(20 P)

	a)	b)	c)	d)
1.1				
1.2				
1.3				
1.4				

Aufgabe 2

(35 P)

2.1

--	--	--	--	--	--	--

--	--	--	--	--	--	--

2.2

--	--	--	--	--	--	--

--	--	--	--	--	--	--

--	--	--	--	--	--	--

2.3

Lösungsbogen

Ihre Matrikelnummer:

--	--	--	--	--	--	--

Klausur
Business Intelligence

Unterschrift _____
SS 2017

8

Aufgabe 3

(35 P)

3.1

--	--	--	--	--	--	--

--	--	--	--	--	--	--

3.2

--	--	--	--	--	--	--

--	--	--	--	--	--	--

--	--	--	--	--	--	--

3.3

--	--	--	--	--	--	--

Aufgabe 4 (Ein Kreuz bei Richtig oder Falsch eintragen)

(10 P)

	Richtig	Falsch
4.1		
4.2		
4.3		
4.4		
4.5		
4.6		
4.7		
4.8		
4.9		
4.10		