



Klausur im Fach

Big Data Anwendungen

Sommersemester 2018

Angaben zur Klausur

Prüfer: Dr. Stephan Schosser

Datum: 25. Juli 2018

Prüfungsnummer: 21807

Persönliche Angaben (in Druckbuchstaben ausfüllen)

Nachname: _____ Vorname: _____

Matrikelnummer: _____ Fakultät: _____

Bewertung (wird vom Prüfer ausgefüllt)

Aufgabe	1	2	3	Gesamtpunkte	Note
Punkte					

Zugelassene Hilfsmittel

- Nicht-programmierbarer Taschenrechner
ohne Kommunikations- oder Datenverarbeitungsfunktion (lt. Aushang des Prüfungsamtes)

Hinweise zur Klausur

- Die Bearbeitungszeit für diese Klausur beträgt 60 Minuten.
- Die Klausur besteht aus 3 Aufgaben, von denen 3 Aufgaben zu bearbeiten sind.
- Die Klausur umfasst 2 Seiten.
- Die Heftung dieser Unterlagen darf nicht gelöst werden.

Hinweise zur Bearbeitung

- Bitte tragen Sie oben auf diesem Deckblatt zuerst Ihre persönlichen Daten ein.
- Bitte prüfen Sie die Vollständigkeit der Klausur.
- Sie sind dafür verantwortlich, dass das Aufsichtspersonal Ihre Klausur erhält.
- Viel Erfolg beim Lösen der Klausuraufgaben!

Aufgabe 1 (Association Rules)**(20 Punkte)**

Gegeben seien folgende Daten:

Bettdecke	Matratze	Bettgestell	Kissen	Bezug	Lampe
1	1	1	1	0	1
0	1	0	1	1	0
0	0	1	1	0	0
0	0	1	1	1	1
1	0	0	1	0	0
1	1	1	1	1	0

- (a) Der Wert 1 steht für „gekauft“, der Wert 0 für „nicht gekauft“. Leiten Sie mit Hilfe des Apriori Algorithmus die Frequent Itemsets ab, die einen Support von mindestens 1/2 besitzen. **(10 Punkte)**
- (b) Identifizieren Sie auf der Basis der Frequent Itemsets aus (a) alle Association Rules mit minimaler Confidence von 3/4. **(5 Punkte)**
- (c) Diskutieren Sie die Nachteile des Apriori Algorithmus gegenüber FP-Growth. **(2 Punkte)**
- (d) Geben Sie an, welchen Vorteil Recommender Systeme gegenüber Association Rules besitzen. **(3 Punkte)**

Aufgabe 2 (Clustering)**(20 Punkte)**

Gegeben seien folgende Daten:

Alter	Einkommen	Kontostand	Kreditwürdig
18	10.000	-1.000	Ja
17	2.000	500	Nein
25	4.000	1.000	Ja
40	3.000	2.000	Ja

- (a) Nennen Sie alle Attribute für die mit Hilfe des Jaccard-Index Abstände ermittelt werden können und begründen Sie kurz. **(2 Punkte)**
- (b) Nennen Sie zwei Verfahren zur Transformation von Daten auf ähnliche Wertebereiche. Nutzen Sie eines davon und transformieren Sie die Daten so, dass Sie mit Hilfe der Manhattan-Distanz sinnvoll geclustert werden können. **(5 Punkte)**
- (c) Nutzen Sie hierarchisch agglomeratives Clustering zum Clustern der Daten aus (b).
[Sollten Sie keine Lösung für (b) haben, nutzen Sie die Ausgangsdaten] **(10 Punkte)**
- (d) Zeichnen Sie ein Dendrogramm für die Daten aus (c).
[Sollten Sie keine Lösung für (c) haben, zeichnen Sie ein fiktives Dendrogramm und nennen Sie für alle Elemente, wie Sie diese konstruieren würden] **(3 Punkte)**

Aufgabe 3 (Sonstiges)**(20 Punkte)**

- (a) Grenzen Sie Big Data von traditioneller Datenanalyse ab. **(5 Punkte)**
- (b) Diskutieren Sie welche Vorteile CDH-Trees gegenüber traditionellen Entscheidungsbäumen besitzen. **(5 Punkte)**
- (c) Erläutern Sie, wie Association Rules für die Lösung von Klassifikationsproblemen genutzt werden können. **(5 Punkte)**
- (d) Beschreiben Sie, wie Support Vector Machines funktionieren. **(5 Punkte)**