



Klausur im Fach

Big Data Anwendungen

Sommersemester 2019

Angaben zur Klausur

Prüfer: Dr. Stephan Schosser

Datum: 13. Juli 2019

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 (Klassifikation)**(20 Punkte)**

Gegeben seien folgende Trainingsdaten:

Trinken	Alkohol	Farbe	Deko
Ja	10%	Blau	Frucht
Ja	5%	Grün	Keine
Ja	10%	Grün	Frucht
Ja	0%	Grün	Keine
Nein	0%	Blau	Frucht
Nein	5%	Grün	Frucht

- (a) Leiten Sie das Attribut „Trinken“ mit Hilfe eines Entscheidungsbaums ab. Nutzen Sie hierfür als Splitkriterium den Gini-Index und entwickeln Sie solange neue Knoten bis entweder kein Splitattribut mehr verfügbar ist oder in einem Knoten alle Daten in der gleichen Klasse sind.

(10 Punkte)

- (b) Sagen Sie mit Hilfe des Entscheidungsbaums aus (a) die Empfehlung für folgende Beobachtung voraus (5%, Blau, Keine).

(3 Punkte)

- (c) Nennen Sie zwei andere Klassifikationsverfahren neben den Entscheidungsbäumen und diskutieren Sie kurz die Vorteile der Verfahren gegenüber Entscheidungsbäumen.

(4 Punkte)

- (d) Erläutern Sie wie Soziale Netzwerke zur Klassifikation genutzt werden können.

(3 Punkte)**Aufgabe 2 (Association Rules)****(20 Punkte)**

Gegeben seien folgende Daten:

Bier	Sekt	Wein	Gin	Nüsse	Chips
1	1	1	1	0	1
0	1	1	0	1	0
0	1	0	1	0	0
0	0	0	1	1	1
1	1	0	0	1	0
1	1	1	1	0	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) Geben Sie an, welche Vorteile Content Based Recommendations gegenüber Association Rules besitzen.

(3 Punkte)

- (d) Diskutieren Sie die Unterschiede zwischen Content Based Recommendations und Collaborative Filtering.

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

- (a) Diskutieren Sie welche Vorteile und Nachteile H-Trees gegenüber traditionellen Entscheidungsbäumen besitzen.

(5 Punkte)

- (b) Erläutern Sie den Datenkreislauf bei Big Data Lösungen.

(5 Punkte)

- (c) Erläutern Sie kurz, wie sie mit Hilfe eines sozialen Netzwerks entscheiden, welche Mitglieder für Werbung besonders relevant sind.

(4 Punkte)

- (d) Erläutern Sie, wie k Nearest Neighbor Klassifikation funktioniert. Diskutieren Sie dann Vor- und Nachteile des Verfahrens gegenüber anderen Klassifikationsalgorithmen.

(6 Punkte)