



Klausur im Fach

Big Data Anwendungen Sommersemester 2021

Angaben zur Klausur

Prüfer: Dr. Stephan Schosser

Datum: 19. Juli 2021

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 (Assoziation Rules)**(20 Punkte)**

Gegeben seien folgende Warenkörbe in einer Drogeriemarktkette:

- Mia: Geschirrspültabs, Zahnbürste, Thermometer, Rasierer
- Noah: Schnelltest, Mundschutz, Rasierer
- Emma: Zahnbürste, Thermometer, Mundschutz
- Paul: Rasierer, Geschirrspültabs, Schnelltest, Mundschutz
- Ella: Schnelltest, Rasierer, Thermometer
- Emil: Zahnbürste, Mundschutz, Rasierer, Thermometer

- (a) Bestimmen Sie für die obigen Warenkörbe die Frequent Itemsets mit Hilfe des Apriori Algorithmus. Alle Frequent Itemsets sollen einen Support von mindestens 55% besitzen.

(10**Punkte)**

- (b) Prüfen Sie für die zwei-elementigen Frequent Itemsets aus (a), ob sich von diesen Assoziation Rules mit einer Confidence von 75% ableiten lassen und geben Sie diese an. **(4 Punkte)**
- (c) Bestimmen Sie den Lift für die Association Rule „Rasierer -> Thermometer“ und erläutern Sie kurz, an diesem Beispiel die Vorteile des Lift gegenüber der Confidence. **(3 Punkte)**
- (d) Diskutieren Sie kurz welche Nachteile von Assoziation Rules mit Hilfe von Collaborative Filtering umgangen werden können. **(3 Punkte)**

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

Gegeben seien folgende Daten von Bäumen:

Wuchshöhe	Wurzelraum	Blattbreite	Blattform
20m	60m ³	11cm	Rundlich
30m	10m ³	6cm	Eiförmig
30m	210m ³	4cm	Gefächert
10m	110m ³	5cm	Länglich
20m	60m ³	1cm	Gefingert

- (a) Nennen Sie das Attribut, das selbst nach einer geeigneten Transformation keinen Beitrag zu einem Clustering bieten wird und begründen Sie dies kurz. **(2 Punkte)**
- (b) Erläutern Sie für die drei verbleibenden Attribute, warum diese ohne Transformation zu keinem sinnvollen Ergebnis führen werden. **(3 Punkte)**
- (c) Transformieren Sie die verbleibenden Attribute nach Streichen des Attributs aus (a) mit Hilfe der Min-Max-Skalierung. **(3 Punkte)**
- (d) Clustern Sie die Daten aus (c) mit Hilfe des hierarchisch agglomerativen Clusterings. Nutzen Sie dabei die Manhattan Distanz. **(10 Punkte)**
- (e) Leiten Sie von Ihrem Ergebnis ein Dendrogramm ab. **(2 Punkte)**

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

- (a) Nennen und erläutern sie zwei Maße, um beim Aufbau von Entscheidungsbäumen das nächste Splitattribut zu wählen. **(4 Punkte)**
- (b) Erläutern Sie warum die Kombination von Klassifikatoren sinnvoll ist und nennen sie die Ebenen auf denen kombinierte Klassifikatoren angepasst werden können. **(4 Punkte)**
- (c) Erläutern Sie worauf bei der Bewertung eines Klassifikators mit Hilfe einer Confusion-Matrix geachtet werden muss, wenn ein Klassifikator für Spam (d.h. bei dem eine Klasse in den Testdaten stark überrepräsentiert ist) bewertet werden soll. **(4 Punkte)**
- (d) Wie lassen sich Strukturen in sozialen Netzwerken nutzen, um Klassen für Teilnehmer abzuleiten, wenn für einen Teil der Teilnehmer die entsprechenden Klassen bekannt sind? **(4 Punkte)**
- (e) Erläutern Sie die Kennzahl Prestige in sozialen Netzwerken und deren ökonomische Bedeutung. Nennen Sie einen Vorteil von Prestige gegenüber der Zentralität. **(4 Punkte)**