



Klausur im Fach

Big Data Anwendungen

Sommersemester 2022

Angaben zur Klausur

Prüfer: Dr. Stephan Schosser

Datum: 26. Juli 2022

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 programmierter Taschenrechner (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:

- A: Geschirrspültabs, Zahnbürste, Thermometer, Rasierer
- B: Schnelltest, Mundschutz, Rasierer
- C: Zahnbürste, Thermometer, Mundschutz
- D: Rasierer, Geschirrspültabs, Schnelltest, Mundschutz

- (a) Bestimmen Sie für die obigen Warenkörbe die Frequent Itemsets mit Hilfe des Apriori Algorithmus. Alle Frequent Itemsets sollen einen Support von 50% oder mehr 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) Stellen Sie sich vor, sie haben ein Frequent Itemset „Zahnbürste, Schnelltest, Rasierer“ gefunden. Nennen sie alle Assoziation Rules, die sie hiervon ableiten könnten. **(3 Punkte)**
- (d) Diskutieren Sie, ob Assoziation Rules oder Collaborative Filtering besser geeignet sind um Kunden auf Webshops andere Artikel zu empfehlen. **(3 Punkte)**

Aufgabe 2 (Klassifikation)**(20 Punkte)**

Gegeben seien folgende Daten zu Basketball Korbwürfen:

Position x	Position y	Wurfwinkel	Treffer
Rechts	3-Meter	Hoch	Ja
Rechts	Unter Korb	Gering	Ja
Rechts	Unter Korb	Hoch	Nein
Links	3 Meter	Mittel	Nein
Links	3 Meter	Gering	Ja

- (a) Führen Sie den ersten Split eines passenden Entscheidungsbaums durch. Sagen Sie dabei „Treffer“ voraus und nutzen Sie den Informationsgewinn. **(10 Punkte)**
- (b) Erläutern Sie, wie sie mit dem Naiv Bayes Algorithmus eine Klassifikation ermitteln könnten. **(5 Punkte)**
- (c) Wie lassen sich Attribute mit numerischen Werten bei der Klassifikation berücksichtigen? **(2 Punkte)**
- (d) Wie können Entscheidungsbäume für Regressionen genutzt werden? **(3 Punkte)**

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

- (a) Erläutern Sie, die Ebenen in welchen Dimensionen sie die Qualität von Entscheidungsbäumen verändern können. **(5 Punkte)**
- (b) Diskutieren Sie drei Verfahren, wie sie Entscheidungsbäume sinnvoll kombinieren können. **(3 Punkte)**
- (c) Erläutern Sie mit welcher Kennzahl Sie in sozialen Netzwerken Teilnehmer identifizieren können, die besonders für Werbung geeignet sind. **(2 Punkte)**
- (d) Erläutern Sie den k Nearest. Neighbor Algorithmus und welches Data Mining Problem dieser löst. **(5 Punkte)**
- (e) Erläutern Sie den Collaborative Filtering Algorithmus. **(5 Punkte)**