

**FRAGE 1 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**



Nennen Sie die fünf wichtigsten Unterschiede zwischen KRIs und KPIs.

1. KRIs können finanziell oder nicht-finanziell sein, wohingegen KPIs immer nicht-finanziell sind.
2. KRIs werden entweder monatlich oder jährlich gemessen, wohingegen KPIs stündlich, täglich oder wöchentlich gemessen werden.
3. Verfügbarkeit von Informationen. Über die KRIs wird in Sitzungen berichtet, während die KPIs den Mitarbeitenden ständig zur Verfügung gestellt werden.
4. KRIs geben keine Auskunft über Probleme, während KPIs behebbare Probleme aufzeigen.
5. Das Spektrum der Schwerpunkte. Jeder KRI stellt eine Zusammenfassung vieler Aktivitäten dar, während sich jeder KPI auf eine bestimmte Aktion konzentriert.

# (Je 2 Punkte)



**FRAGE 2 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**



Nennen und beschreiben Sie kurz vier der wichtigsten Merkmale effektiver KPIs.

* Nicht-finanziell. Stattdessen ist eine finanzielle Kennzahl ein Indikator für Ergebnisindikatoren (RIs).
* Häufige Messung. Damit sich ein KPI positiv auswirkt, muss er häufig gemessen werden (stündlich, täglich oder wöchentlich). Das erleichtert den Spielraum für Verbesserungen.
* Einfach zu verstehen. Der gewählte KPI sollte für alle Mitarbeitenden des Unternehmens leicht verständlich sein.
* Ausschließlich mit einem Team verbunden. Es bewährt sich, einen KPI mit einem bestimmten Team in der Organisation bzw. ein beginnendes Problem direkt mit einem Team zu verknüpfen.
* Beitrag zur Erreichung des strategischen Ziels der Organisation. KPIs (Leistungsindikatoren) bilden die Grundlage der Strategie, die sicherstellen soll, dass die Ziele der Organisation erreicht werden.
* Nur kleinere negative Aspekte. Vor der Nutzung eines Leistungsindikators (KPI) muss er getestet werden, damit mögliche negative Folgen festgestellt werden können. Bedeutung und Schwere der möglichen Folgen müssen bewertet werden.

# (2 Punkte für jedes Merkmal; maximal 4 Merkmale)



**FRAGE 3 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



Nennen Sie drei Indikatoren, die zu der folgenden Kategorie gehören: Verständnis der internen Prozesse einer Organisation.

Auslastungsgrad (CUR), Projektplanabweichung (PSV), Anteil der vollständigen und rechtzeitigen Lieferungen (VRL/DIFOT), Prozessausfallzeit.

# (Je 2 Punkte)



**FRAGE 4 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



Nennen Sie die drei Stufen des von Parmenter (2020) beschriebenen KPI-Entwurfsprozesses.

Stufe 1: Das Unternehmen sollte sich voll und ganz für die Einführung von KPIs einsetzen und dafür ein **internes** KPI-Team einrichten. Das KPI-Team sollte sich aus hochqualifizierten Personen zusammensetzen, die mit den strategischen Zielen des Unternehmens bestens vertraut sind.

Stufe 2: Die **entscheidenden Erfolgsfaktoren** des Unternehmens müssen festgelegt werden.

Stufe 3: Die richtigen **Maßnahmen, die sich auf die Leistung der Organisation auswirken**, müssen gewählt werden.

# (2 Punkte pro Stichwort)



**FRAGE 5 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



Nennen Sie drei Indikatoren, die zu der folgenden Kategorie gehören: Verstehen der Kund:innen.

Promotorenüberhang (NPS), Kundenbindungsrate (CRR), Kundenreklamationen, Kundenfluktuationsrate

# (Je 2 Punkte, max. 3 Indikatoren)



**FRAGE 6 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**



Erklären Sie, wie der Promotorenüberhang (NPS = Net Promoter Score) gemessen werden kann.

Der NPS kann mit einer **Umfrage per E-Mail, online oder persönlich** gemessen werden. Zur Berechnung des Promotorenüberhangs (NPS) einer Organisation kann eine **Skala von 0 bis 10** verwendet werden.

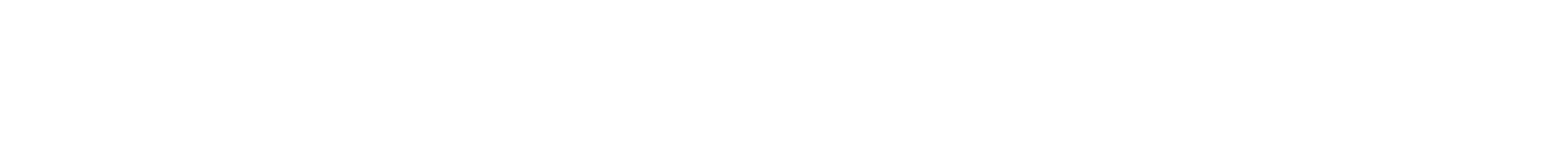
Die Kund:innen lassen sich in verschiedene Gruppen einteilen, darunter diejenigen, die **dem Unternehmen gegenüber loyal** sind und die Arbeit des Unternehmens fördern wollen, und diejenigen, die **unzufrieden** sind und sich daher darauf konzentrieren, das Ansehen der Marke zu schädigen. Der NPS kann berechnet werden, indem **der Prozentsatz der unzufriedenen Kund:innen vom Prozentsatz der zufriedenen Kund:innen abgezogen wird**.

# (2 Punkte pro Stichwort)



**FRAGE 7 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**



Definieren Sie die Kundenbindungsrate (CRR = Customer Retention Rate), insbesondere …

1. … wie die Rate gemessen werden kann,
2. … wo die Daten gesammelt werden können, und
3. … Angabe der Formel zur Berechnung der Kundenbindungsrate.

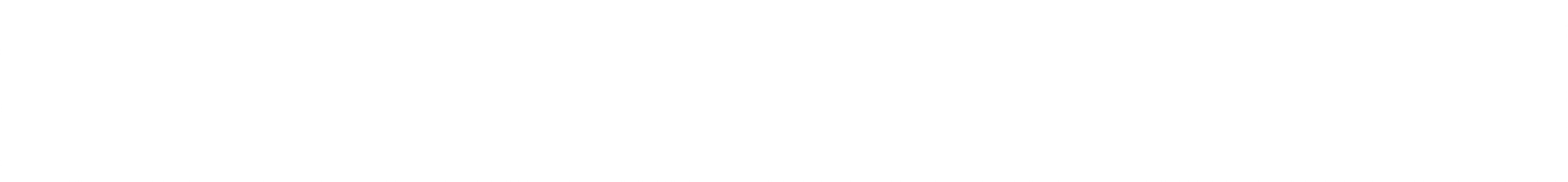
Die Kundenbindungsrate misst, wie loyal Kund:innen gegenüber einer Marke, einem Unternehmen oder einer Dienstleistung sind. **(3 Punkte)**

Die für die Kundenbindungsrate benötigten Daten können aus dem CRM-System einer Organisation **(3 Punkte)** oder über eine gründliche Untersuchung gesammelt werden.

**(1 Punkt)**

Die Formel zur Berechnung der Kundenbindungsrate CRR lautet wie folgt:

CRR = (Anzahl der Kund:innen zu Beginn) / (Anzahl der Kund:innen am Ende) **(3 Punkte)**



Ein Unternehmen hat 100 Kund:innen befragt, um seinen Promotorenüberhang (NPS) zu berechnen. Alle Kund:innen gaben an, wie wahrscheinlich sie das Produkt des Unternehmens an ihre Freunde weiterempfehlen würden. Die Antworten wurden in zwei Hauptgruppen unterteilt. 70 Kund:innen antworteten mit „sehr wahrscheinlich“ und 20 mit „wahrscheinlich nicht“.

Welchen Wert hat der Promotorenüberhang (NPS) in diesem Fall? Zeigen Sie Ihren Rechenweg.



**FRAGE 8 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**

NPS = (% der zufriedenen Kund:innen – % der unzufriedenen Kund:innen) **(3 Punkte)** NPS = [(70/100)\*100 – (30/100)\*100] **(3 Punkte für die Gleichung)** NPS = 40 % **(2 Punkte für die richtige Antwort)**



**FRAGE 9 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**



Nennen Sie vier Metriken, mit denen Sie Kundenreklamationen eindeutig messen können.

Die Gesamtzahl der von Kund:innen innerhalb eines Tages, einer Woche oder eines Monats eingereichten Reklamationen.

* Die durchschnittliche Zeit für die Auflösung einer Reklamation.

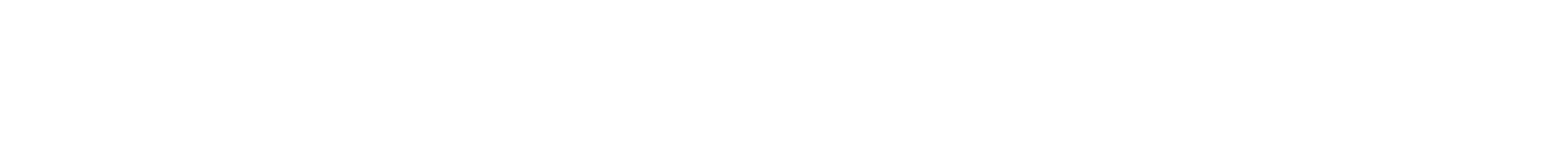
Die Häufigkeit der verschiedenen Arten an Reklamationen. Den Prozentsatz der Reklamationen, die gelöst werden konnten.

# (2 Punkte pro Metrik)



**FRAGE 10 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**



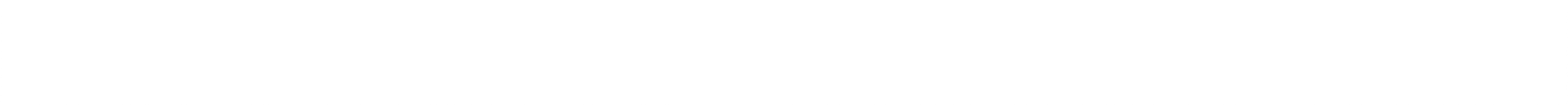
1. Welcher Faktor beeinflusst die Kosten für die Berechnung der Kundenbindungsrate (CRR) am stärksten?
2. Nennen Sie ein Beispiel für ein Unternehmen, für das die Berechnung einer CRR mit geringen Kosten verbunden ist, sowie ein Beispiel für ein Unternehmen, für das die Kosten dafür höher sind.
3. Die zur Bewertung der Kundenbindungsrate erforderlichen Kosten hängen von den Daten eines Unternehmens ab **(2 Punkte)**, insbesondere der Quantität und Qualität der Daten. **(2 Punkte)**
4. Eine Bank, die große Mengen an Informationen über ihre Kund:innen sammelt und speichert, kann beispielsweise die Kundenbindungsrate mit relativ geringen Kosten berechnen. **(3 Punkte)**

Für Unternehmen wie Restaurants oder Geschäfte, die nicht über große Mengen an Kundendaten verfügen, kann die Berechnung der Kundenbindungsrate dagegen eine Umfrage erfordern. Dies führt zu höheren Kosten und zusätzlichem Personaleinsatz. **(3 Punkte)**



**FRAGE 11 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



Nennen Sie zwei Faktoren, die sich auf die Häufigkeit der Erfassung von Kundenreklamationen auswirken. Erläutern Sie kurz.

* 1. Umfang der Organisation.

Je größer das Unternehmen ist, desto höher sind die Kosten für die Erfassung von Kundenreklamationen. **(3 Punkte)**

* 1. Art des Produkts oder der Dienstleistung.

Die Art des Produkts oder der Dienstleistung bestimmt die Häufigkeit, mit der Kundenreklamationen gesammelt werden. Ein Produkt, das erst kürzlich auf den Markt gekommen ist, erfordert zum Beispiel häufigeres Sammeln von Reklamationen. Auf diese Weise kann das Produkt entsprechend optimiert werden. **(3 Punkte)**



**FRAGE 12 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**



Erläutern Sie die möglichen Kosten, die mit der Berechnung der Kennzahl Promotorenüberhang (NPS = Net Promoter Score) zusammenhängen.

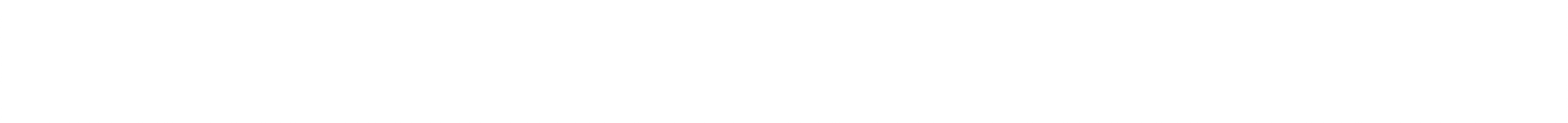
Die zur Berechnung des Promotorenüberhangs gesammelten Daten stammen aus **Umfragen** unter aktuellen Kund:innen. Die Kosten **hängen von der für die Sammlung dieser Daten gewählten Methode ab.** Die Kosten sind besonders hoch, wenn ein Unternehmen **Umfragen** in **Papierform oder persönlich** durchführen möchte.Zur Kostensenkung und Verringerung der an diesem Prozess beteiligten Mitarbeitenden sollte ein Unternehmen eine Möglichkeit finden, **die Sammlung der Daten zu automatisieren**.

# (2 Punkte pro Stichwort)



**FRAGE 13 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**



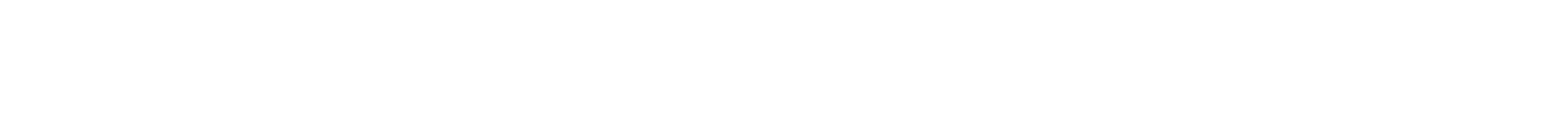
1. Nennen und erklären Sie einen Grund, warum es für Telekommunikationsunternehmen wichtig ist, ihre Kundenfluktuationsrate (CTR) zu minimieren.
2. Nennen Sie einen wichtigen Schwerpunktbereich bei der Minimierung der Kundenfluktuationsrate (CTR).
   * 1. Für Telekommunikations- und Finanzdienstleistungen ist es wichtig, die Kundenfluktuationsrate (Customer Turnover Rate, CTR) zu minimieren, weil die Kund:innen mit großer Wahrscheinlichkeit den **Anbieter wechseln**, wenn ihnen **günstigere Dienstleistungen** oder **bessere Abonnementverträge** angeboten werden. Zum Beispiel können Kund:innen von Telekommunikationsunternehmen problemlos ihren Anbieter wechseln, wenn sie ein besseres Angebot finden.
     2. Die Identifizierung von Kund:innen mit Wechselabsichten und ihre **Rückgewinnung** sollte ein Schwerpunkt für diese Unternehmen sein.

# (2 Punkte pro Stichwort)



**FRAGE 14 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**



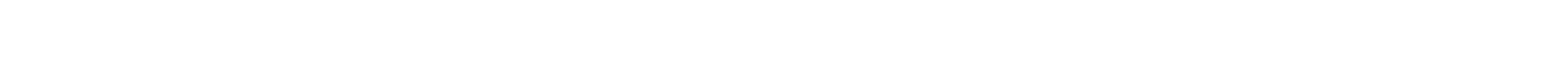
1. Erläutern Sie kurz den Begriff „Textmining“.
2. Erläutern Sie kurz den Begriff „natürliche Sprachverarbeitung“.
3. Definieren Sie kurz den Begriff „Korpus“.
4. Textanalyse oder „Textmining“ sind Begriffe, die den gesamten Prozess der **Identifizierung von Mustern in Textdaten** beschreiben.
5. Die Verarbeitung natürlicher Sprache (NLP = Natural Language Processing) ist eine der Hauptkomponenten des Textmining, mit denen Organisationen Erkenntnisse gewinnen möchten. Die Ermittlung der **Häufigkeit, mit der ein Wort in einem Text oder Dokument erscheint**, gibt einen Überblick über die wichtigsten Wörter in einem Textkorpus.
6. Ein Korpus ist eine **Sammlung geschriebener Texte oder Dokumente**, die mit verschiedenen Techniken zur Verarbeitung natürlicher Sprache bearbeitet werden.

# (2 Punkte pro Stichwort)



**FRAGE 15 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**



Erläutern Sie kurz den Begriff „semantischer Raum“.

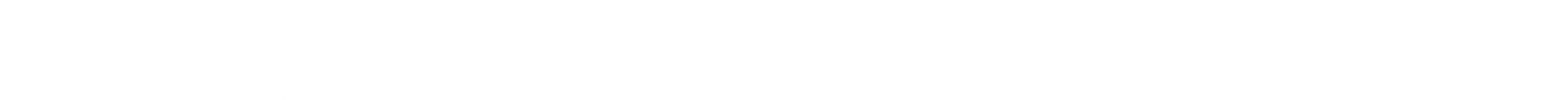
Die Verarbeitung natürlicher Sprache verändert die Art und Weise, wie wir eine Sprache sehen. Daher kann ein Dokument nicht nur als eine Folge von Wörtern betrachtet werden, sondern vielmehr als ein Punkt in einem **mehrdimensionalen Raum** (semantischer Raum). Je näher die **Punkte im Raum** beieinander liegen, desto höher ist die Verbindung zwischen den Dokumenten. Daher kann der semantische Raum als eine Karte betrachtet werden, welche **Ähnlichkeiten in der Bedeutung der Dokumente** in einem Korpus festlegt. Der Abstand zwischen den Punkten ist das Maß, das verwendet wird, um Entscheidungen in diesem semantischen Raum zu treffen.

# (2 Punkte pro Stichwort)



**FRAGE 16 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**



1. Wie kann die Kundenfluktuationsrate (CTR) gemessen werden?
2. Nennen Sie die bei diesem Vorgang verwendeten Formel.
   1. Für Unternehmen mit Kundenverträgen kann die CTR dadurch berechnet werden, dass ausgewertet wird, wie viele Kundenverträge verlängert wurden. **(3 Punkte)**

Werkzeuge für CRM und Geschäftsanalytik (Business Intelligence) können für Unternehmen mit einem größeren Kundenstamm sehr nützlich sein. **(2 Punkte)**

Mit ausgefeilten Data-Mining-Verfahren lassen sich Kund:innen aufspüren, die mit hoher Wahrscheinlichkeit das Unternehmen verlassen, und feststellen, welche Faktoren diese Abwanderung bedingen. Durch die Erkennung möglicher Fluktuations- und Kündigungsfaktoren kann das Unternehmen die notwendigen Maßnahmen festlegen und umsetzen, um so die Kund:innen zurückzugewinnen. **(2 Punkte)**

* 1. CTR = (Anzahl der verlorenen Kund:innen im Zeitraum t)/(Gesamtzahl der Kund:innen im Zeitraum t) **(3 Punkte)**



**FRAGE 17 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



Definieren Sie kurz den Auslastungsgrad (Capacity Utilization Rate, CUR) und beschreiben Sie, warum er ein wichtiger KPI für die Produktionsüberwachung ist.

Der Auslastungsgrad ist das maximale Produktionspotenzial, das unter Berücksichtigung der vollen Auslastung und der verfügbaren Ressourcen erreicht werden kann.

Mit dem Auslastungsgrad *(CUR = Capacity utilization rate)* lässt sich messen, wie effektiv ein Unternehmen seine Produktionskapazität nutzt. Diese Kennzahl misst die Differenz zwischen der tatsächlich erzeugten Menge an Produkten oder erbrachten Dienstleistungen und dem Produktionspotenzial, das bei voller Auslastung der Kapazitäten des Unternehmens erreicht werden könnte. **(3 Punkte)**

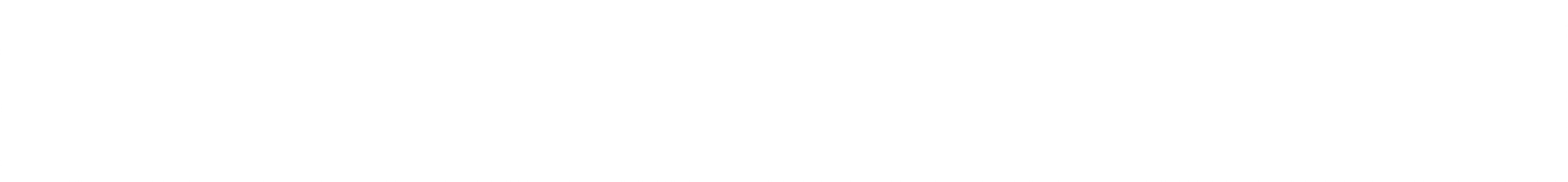
Dieser Indikator ist wichtig, weil er bestimmt, ob ein Anstieg der Produktion zu einem Anstieg der Produktionskosten führen würde. **(3 Punkte)**



**FRAGE 18 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**





Das Wort „Mode“ findet sich 15 Mal in Dokument A und 300 Mal in Dokument B. Dokument A ist ein Artikel mit 100 Wörtern, während Dokument B ein Buch mit 50.000 Wörtern ist.

1. Berechnen Sie die jeweilige Termhäufigkeit (TF) für dieses Wort in Dokument A und Dokument B.
2. In welchem Dokument erscheint das Wort „Mode“ seltener?

a)

TF(„Mode“, Dokument A) = 15/100 = 0,15 **(3 Punkte)** TF(„Mode“, Dokument B) = 300/50000 = 0,006 **(3 Punkte)**

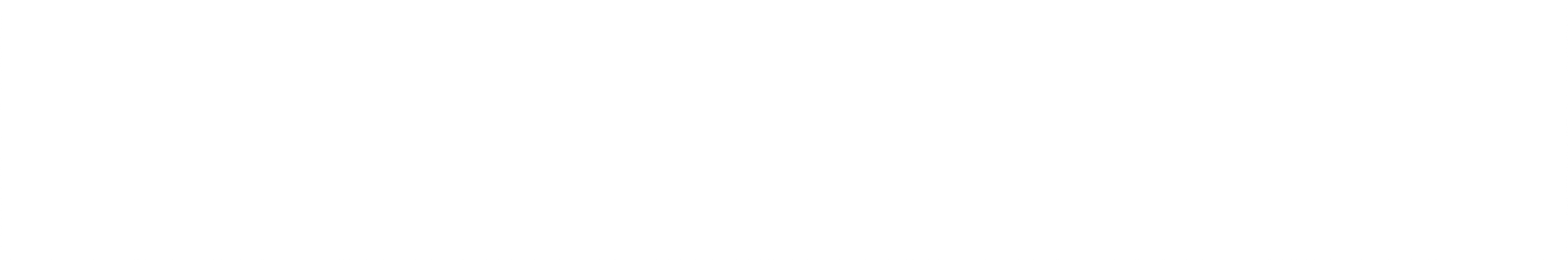
b) Je kleiner der TF-Wert, desto seltener ist ein Wort in einem Dokument. **(1 Punkt)**

Daher ist das Wort „Mode“ in Dokument B weniger häufig zu finden. **(3 Punkte)**



**FRAGE 19 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**



Das Wort „Blume“ kommt zehnmal in Dokument A und 200 Mal in Dokument B vor. Dokument A ist ein Artikel mit 500 Wörtern, und Dokument B ein Buch mit 10.000 Wörtern.

Der gesamte Korpus besteht aus 50.000 Wörtern, und das Wort „Blume“ kommt nur in Dokument A und Dokument B vor.

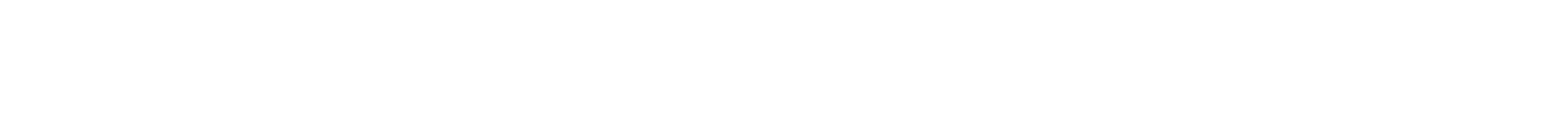
1. Berechnen Sie die Termhäufigkeit des Wortes „Blume“ in Dokument A.
2. Berechnen Sie die inverse Dokumentenhäufigkeit des Wortes „Blume“ in Dokument A.
3. Ermitteln Sie den endgültigen Wert von Tf-idf für das Wort „Blume“ in Dokument A.
4. TF(„Blume“, Dokument A) = 10/500 = 0,02 **(3 Punkte)**
5. IDF(„Blume“, Korpus) = log(50000/2)= 4,39 **(3 Punkte)**
6. TF\_IDF („Blume“, Dokument A, Korpus) = TF(„Blume“, Dokument A) \* IDF („Blume“, Korpus) **(1 Punkt)**

TF\_IDF(„Blume“, Dokument A, Korpus) = 0,02 \* 4,39 = 0,0878 **(3 Punkte)**



**FRAGE 20 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**



Am Standort eines Unternehmens können maximal 20.000 Einheiten pro Tag produziert werden. Die tatsächliche Produktion pro Tag liegt bei 16.000.

Berechnen Sie den Auslastungsgrad (CUR) für diesen Standort.

CUR = (Tatsächliche Kapazität zum Zeitpunkt t / Potenzielle Kapazität zum Zeitpunkt t) \* 100 **(3 Punkte)**

CUR = (16.000/20.000)\*100 **(3 Punkte)**

CUR = 80 % **(2 Punkte)**



Erläutern Sie kurz, was semantische Analyse bedeutet.



**FRAGE 21 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**

Die semantische Analyse beschreibt **den Prozess des Verstehens von Texten in natürlicher Sprache.** Im Gegensatz zur lexikalischen Analyse, bei der es um die Bedeutung eines bestimmten Wortes geht, konzentriert sich die semantische Analyse auf **größere Textstücke**.Sie kann die **Beziehungen zwischen Wörtern in Sätzen, Absätzen oder sogar ganzen Dokumenten** erkennen.Mit der semantischen Technologie wird die Struktur eines Satzes oder eines Dokuments analysiert, um so die wichtigsten Informationen zu identifizieren und das **Thema zu verstehen**, von dem der Text handelt. Durch den Einsatz von semantischer Analyse und natürlicher Sprachverarbeitung (NLP) lernen Maschinen, Texte besser zu verstehen und wertvolle Informationen für den Kundendienst und die Geschäftsanalytik (Business Intelligence) zu liefern.

# (2 Punkte pro Stichwort)



**FRAGE 22 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**



Nennen und beschreiben Sie kurz vier mögliche Kosten, die mit dem VRL-Indikator (vollständige und rechtzeitige Lieferung, DIFOT) verbunden sind.

Die Daten können mit Hilfe eines **Auftragsverfolgungssystems** gesammelt werden. Wenn ein solches bereits im Unternehmen vorhanden ist, **entstehen dadurch keine zusätzlichen Kosten**. Unternehmen ohne ein solches System können ihre Produkte durch **Dritte** liefern lassen, was jedoch **zusätzliche Kosten** verursacht. Alternativ kann die Organisation zur Berechnung der VRL (DIFOT) auch eine **Umfrage** durchführen, um den DIFOT zu berechnen, doch auch dies verursacht Kosten.

# (2 Punkte pro Stichwort)



**FRAGE 23 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



Erklären Sie die Kosten im Zusammenhang mit der Kundenfluktuationsrate (CTR).

Die Kosten für die Messung dieses Indikators hängen von der **Branche** und der **Gesamtzahl der Kund:innen** ab. Unternehmen, die Produkte oder Dienstleistungen anbieten, ohne eine nennenswerte Beziehung zu ihren Kund:innen aufzubauen, haben möglicherweise Schwierigkeiten, die Zahl der verlorenen Kund:innen im Auge zu behalten.

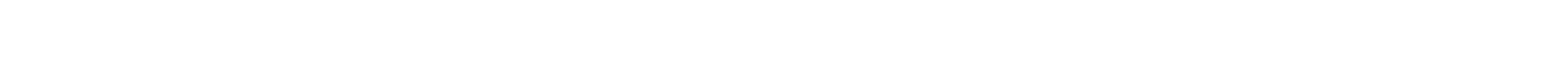
# Technologische Hilfsmittel und Systeme wie z. B. Werkzeuge für die Kundenpflege (CRM = Customer Relationship Management) sind von großem Nutzen, um die Kosten und den Personalaufwand für die Messung dieser Kennzahl zu begrenzen.

# (2 Punkte pro Stichwort)



**FRAGE 24 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**



Erläutern Sie kurz, was Themenmodellierung ist und wo sie eingesetzt werden kann.

Die Themenmodellierung ist eine Methode, die dabei hilft, große Mengen von Textdaten zu untersuchen, Ähnlichkeiten in Dokumenten zu finden und abstrakte Themen zu entdecken. **(2 Punkte)**

Sie wird meist bei Suchmaschinen verwendet, um eine Zeichenfolge mit den relevantesten Ergebnissen abzugleichen. **(2 Punkte)**

Die Themenmodellierung wird auch in Empfehlungssystemen eingesetzt, um benutzerspezifische Dienste anzubieten, um Lebensläufe zu überprüfen und Bewerber:innen nach den erforderlichen Fähigkeiten zu filtern **(2 Punkte)**, um den Kundendienst zu verbessern und Reklamationen zielgerichteter an die entsprechenden Abteilungen weiterzuleiten, sowie in anderen Systemen zur Informationsgewinnung **(2 Punkte).**



**FRAGE 25 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



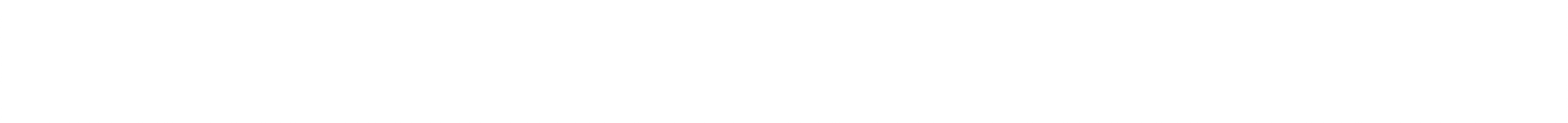
Nennen Sie fünf der wichtigsten Web-Metriken.

1. Anzahl der Besuche oder einmaligen Besuche **(2 Punkte)**
2. Auf der Website verbrachte Zeit **(1 Punkt)**
3. Absprungrate (Bounce Rate) **(1 Punkt)**
4. Ausstiegsrate (Exit Rate) **(1 Punkt)**
5. Konversionsrate **(1 Punkt)**



**FRAGE 26 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**

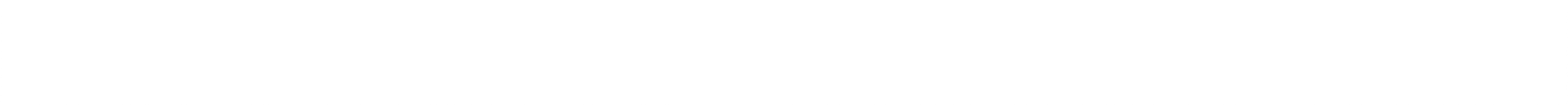


1. Mit welcher Häufigkeit wird der Auslastungsgrad (CUR) gemessen?
2. Wie häufig sollte der Auslastungsgrad (CUR) für eine einzelne Maschine berechnet werden?
3. Wie häufig sollte der Auslastungsgrad (CUR) für eine ganze Fabrik berechnet werden?
   1. Der Auslastungsgrad (CUR) kann **täglich** oder **wöchentlich** gemessen werden, was von der jeweiligen **Produktionslinie abhängt**.
   2. Wenn das Unternehmen den Auslastungsgrad (CUR) für eine einzelne Maschine berechnen muss, kann er **stündlich** gemessen werden.
   3. Für eine gesamte Fabrik sollte der Auslastungsgrad (CUR) **täglich, wöchentlich oder monatlich** gemessen werden. **(2 Punkte pro Stichwort)**



**FRAGE 27 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**



Nennen und definieren Sie kurz die beiden Metriken „messbar“ und „termingebunden“ im Zusammenhang mit dem Indikator für die Projektplanabweichung (PSV).

Bei Messung der Projektabweichung gegenüber der Zeitplanung ist ein allgemeiner Überblick möglich. Verzögerungen sind in der Regel mit zusätzlichen Kosten verbunden. **(3 Punkte)**

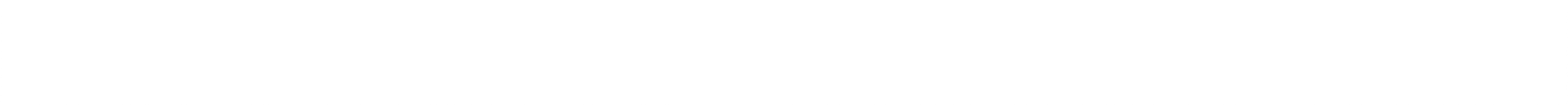
Je nach Dauer der Projekte kann eine Organisation entscheiden, ob sie die Projektplanabweichung (PSV = Project Schedule Variance) wöchentlich oder monatlich messen möchte. **(2 Punkte)**

Die Messung dieses Indikators ist relativ einfach, weil er lediglich einen Vergleich zwischen dem geplanten Fertigstellungstermin und dem tatsächlichen Fertigstellungstermin darstellt. In diesem Fall kann die Zeit in Tagen oder Wochen gemessen werden. **(3 Punkte)**



**FRAGE 28 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



Nennen Sie ein Beispiel dafür, wann eine hohe Absprungrate (Bounce Rate) nicht unbedingt auf mögliche Probleme mit einer Website hinweist.

Bei Blogs **(2 Punkte)** oder Websites mit nur einer Startseite **(2 Punkte)** ist eine hohe Absprungrate nicht als schlechtes Ergebnis zu werten, weil die Absprungrate in diesen Fällen nicht wichtig ist. **(2 Punkte)**



**FRAGE 29 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**



1. Erklären Sie, was eine Sitzung im Zusammenhang mit einer Website ist.
2. Erklären Sie, was ein Cookie im Zusammenhang mit einer Website ist.
3. Jede Sitzung gehört zu einem Besuch auf der Website. Sie beziehen sich in der Regel auf die **Erfahrung** eines **bestimmten Besuchenden** einer Website. Betrachten wir die Nutzung von JavaScript-Tags auf einer Website. In dem Moment, in dem Nutzende die Website betreten, erstellt das Analysewerkzeug eine eindeutige Sitzung für jeden Nutzenden. Alle Nutzenden sind mit einer **eindeutigen Sitzungs-ID** verknüpft, und **jede Aktion, die sie auf der Website durchführen, ist mit dieser eindeutigen Sitzungs-ID verbunden**.
4. Cookies hingegen werden zur Identifikation **einzelner Besucher:innen** einer Website genutzt. Wenn Nutzende ein Element auf einer Website anfordern, wird es mit einem eindeutigen Cookie verknüpft, das **auf dem Webbrowser des Nutzenden gespeichert wird**.Auch wenn die Nutzenden die Website wieder verlassen haben, **bleiben die eindeutigen Cookies im Browser gespeichert**.So werden Nutzende bei einem späteren Besuch derselben Website anhand des zuvor im Browser **gespeicherten eindeutigen Cookies erkannt**.

# (1 Punkt pro Stichwort)



**FRAGE 30 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



Nennen Sie die fünf Hauptschritte der Zeitleiste für die Informationsverarbeitung, die bei der traditionellen Datenanalyse verwendet wird.

* 1. Forschungsfrage **(1 Punkt)**
  2. Extraktion von Datenquellen in Ereignisprotokolle **(2 Punkte)**
  3. Vorverarbeitung der Daten **(1 Punkt)**
  4. Mining-Techniken zur Erkennung von Mustern **(1 Punkt)**
  5. Verbesserung durch Feedback und neue Ideen **(1 Punkt)**



**FRAGE 31 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**



Auf einer Website gab es insgesamt 100 Sitzungen. 50 Besucher:innen betraten und verließen die Website, ohne eine Aktion durchzuführen.

1. Geben Sie die Formel zur Berechnung der Absprungrate an.
2. Berechnen Sie den Wert der Absprungrate für die Website anhand der obigen Informationen.

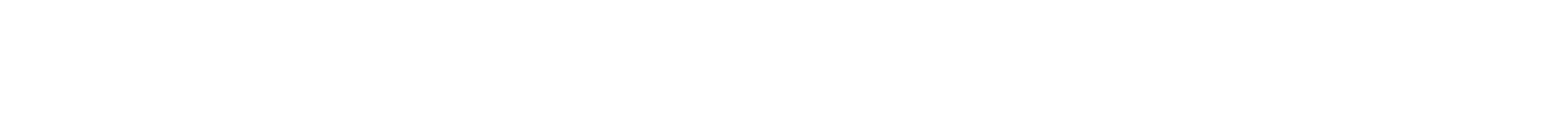
Absprungrate (%) = (Anzahl der Einzelseitensitzungen )/(Gesamtzahl der Sitzungen) **(3 Punkte für die Formel)**

Absprungrate = 50/100= 0,5 \*100 = 50 % **(2 Punkte für das Einsetzen der Werte) (3 Punkte für das Berechnen des richtigen Wertes)**



**FRAGE 32 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 01**



1. Nennen Sie drei Elemente, aus denen ein Petri-Netz besteht.
2. Nennen Sie drei weitere Modellierungssprachen (mit Ausnahme von Petri-Netzen), die für die Modellierung betrieblicher Geschäftsprozesse verwendet werden können.
   * 1. 1. P – Mengen an Plätzen (Stellen) **(1 Punkt)**
        1. T – Mengen von Transitionen (Übergängen) **(1 Punkt)**
        2. I, O – Eingangskanten und Ausgang **(1 Punkt)**
     2. 1. BPMN – Schreibweise für die Geschäftsprozessverwaltung
        1. UML – Unified Modelling Language (Modellierungssprache)
        2. YAWL – Yet another workflow language
        3. EPC – Event Driven Process Chain (Ereignisgesteuerte Prozesskette)

# (1 Punkt für jede Antwort aus diesen vier Optionen)



**FRAGE 33 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



Nennen und erläutern Sie fünf der wichtigsten Vorteile der internen Site-Suche.

Ähnlich den großen Suchmaschinen wie Google und Yahoo gibt es auf den meisten Websites einen internen Suchmechanismus. Dieser bietet die Möglichkeit, auf der gesamten Website nach einem bestimmten Stichwort zu suchen. Durch Überwachung der internen Website-Suchen ist es möglich, die **Absichten der Website-Besucher:innen einzuschätzen**.Mithilfe von Analysewerkzeugen wie Google Analytics kann ein Teil der Besucher:innen, die in einem bestimmten Zeitintervall eine Website-Suche durchgeführt haben, berechnet und zur **Darstellung von Suchtrends** genutzt werden.Solche Analysen liefern Berichte mit Metriken wie z. B. **eindeutige Suchanfragen, Suchabbrüche, Suchergebnisse usw.** So lassen sich Informationen über die **Interessen der Nutzenden** und die gesuchten Inhalte gewinnen. Die Genauigkeit der internen Website-Suche kann durch eine gekoppelte Überwachung der Suchanfragen und der **Absprungrate** bestimmt werden.Da die Absprungrate die Anzahl der Personen bezeichnet, welche die Website auf einer bestimmten Seite verlassen, liefert die Verknüpfung mit der internen Site-Suche Informationen darüber, welche Ergebnisse der internen Suche funktionieren.

# (2 Punkte pro Stichwort)



**FRAGE 34 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 01**



Nennen und definieren Sie kurz die vier Hauptaktivitäten des Process-Mining.

1. Sammeln der Daten: Diese Daten können mit jedem in Unternehmen genutzten Datensystem gesammelt werden, z. B. Excel-Tabellen, Datenbanken, Nachrichtenprotokolle, Transaktionsprotokolle und ERP-Systeme. Abhängig von den anfangs gestellten Fragen kann das Unternehmen unterschiedliche Ereignisprotokolle erstellen, die auf demselben Satz extrahierter Daten basieren.
2. Erkennen und Verstehen von Prozessen: Techniken zur Prozesserkennung erzeugen automatisch eine Darstellung des Prozessmodells, das den aktuellen Prozessen und zugehörigen Daten zugrunde liegt.
3. Konformitätsprüfung: Nach der Prozesserkennungsphase ist eine Konformitätsprüfung (Compliance) notwendig, um mögliche Abweichungen zu ermitteln. Ein Vergleich zwischen einem idealen Prozessmodell und dem aus der Prozesserkennung abgeleiteten Modell ist zwingend. Damit wird überprüft und festgehalten, ob das Informationssystem den Standards und Projektrichtlinien entspricht bzw. inwieweit es davon abweicht.
4. Verbesserung der aktuellen Geschäftsprozesse: Die Verbesserung von Geschäftsprozessen besteht darin, die Prozessleistung bezüglich der Zeit zu analysieren und mögliche Blockaden zu vermeiden.

# (1 Punkt pro Aktivität, 1 Punkt pro Definition)



**FRAGE 35 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 01**



Nennen und erläutern Sie kurz fünf Gründe, warum ein Unternehmen auf Process-Mining umsteigen sollte.

Abgleich der Prozesse. Process-Mining hebt Abweichungen zwischen dem ursprünglichen Prozessentwurf und den tatsächlichen Geschehnissen während der Ausführung der Prozesse hervor. Dadurch bietet sich ein besserer Einblick in die Auswirkungen, die eine bestimmte Aktivität auf die anderen Bereiche der Geschäftsprozesse des Unternehmens hat.

Perfekte und durchgehende Transparenz des Prozessmodells der Organisation. Aufgrund von Unsicherheiten und mangelnden Informationen können herkömmliche Methoden kein richtiges Verständnis des Prozessverhaltens vermitteln.

Eliminierung von Unterschieden in den Daten und Formaten. Daten können in einer Vielzahl von Formaten erzeugt oder gespeichert werden. Durch Process-Mining können sich Möglichkeiten zum Sammeln von Informationen aus verschiedenen Quellen zeigen.

Kürzere Bearbeitungszeit. Durch Process-Mining kann ein Prozessmodell automatisch in kürzerer Zeit und mit geringeren Kosten als mit herkömmlichen Prozessanalysemethoden erstellt werden.

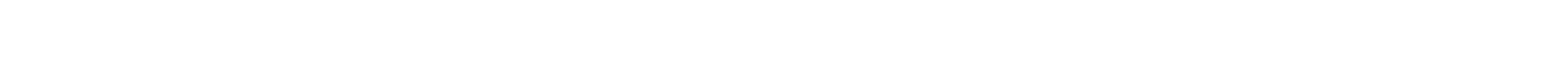
Datengesteuerte Entscheidungen. Die wesentlichen Informationen werden automatisch generiert und bieten die Möglichkeit, Änderungen in verschiedenen Projekten schneller vorzunehmen.

# (2 Punkte für jeden Vorteil)



**FRAGE 36 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



Nennen und erläutern Sie kurz die fünf verschiedenen Quellen des Datenverkehrs auf einer Website.

Eine Website kann Datenverkehr aus einer Vielzahl von Quellen erhalten. Zu den wichtigsten gehören die folgenden:

* Direkter Datenverkehr: Datenverkehr, der nicht von einer anderen verlinkten Website stammt.
* Datenverkehr aus bezahlten Suchen: Datenverkehr, der von Suchkampagnen stammt.
* Organischer Datenverkehr: Datenverkehr, der von Suchmaschinen kommt (nicht bezahlt).
* Sozialer Datenverkehr: Datenverkehr, der von sozialen Medienplattformen kommt (Facebook, Instagram, LinkedIn usw.).
* E-Mail-Datenverkehr: Datenverkehr, der von E-Mail-Kampagnen kommt.

# (2 Punkte pro Datenverkehrsquelle)



**FRAGE 37 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**



Erklären Sie die Analyse sozialer Medien (Social Media Analytics, SMA) und warum sie durchgeführt wird.

Die Analyse sozialer Medien (Social Media Analytics) konzentriert sich auf die Umwandlung aller auf Social-Media-Plattformen generierten Daten in aussagekräftige Geschäftseinblicke. Dieser analytische Prozess kann sowohl mit Wissenschaft als auch mit Kreativität in Verbindung gebracht werden, weil er die Extraktion und Analyse von Informationen mit ausgefeilten Techniken und die Interpretation dieser Analysen zur Erreichung der Ziele einer Organisation beinhaltet. **(3 Punkte)**

Die Analyse sozialer Medien (SMA = Social Media Analytics) nutzt kontinuierlich die auf Social-Media-Plattformen generierten Inhalte für die Echtzeitanalyse, wie Trends in den sozialen Medien entstehen und geteilt werden.

# SMA kann Unternehmen Informationen über Kundenmeinungen und die Stimmung in Bezug auf bestimmte Produkte liefern. (3 Punkte)



**FRAGE 38 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**



Nennen und erklären Sie kurz zwei Kodierungstechniken, die in einem semantischen Raum verwendet werden.

Zwei häufig verwendete Kodierungstechniken im semantischen Raum sind **Bag-of-Words** **(BoW)** und das **Produkt aus Termhäufigkeit und inverser Dokumentenhäufigkeit (Tf-idf, engl.: TF-IDF = term frequency times inverse document frequency)**.

Eine Bag-of-Words stellt in **vektorieller Form** die Anzahl und Häufigkeit der Wörter dar, während die Tf-idf-Vektoren die **Relevanz der Wörter** anhand der Wortbewertung (Score) darstellen. Diese beiden Techniken können Teil eines NLP-Prozesses sein oder separat in unterschiedlichen Kontexten eingesetzt werden. Zu den Anwendungen, in denen sie eingesetzt werden können, gehören Spam-Filterung und Stimmungserkennung (Sentimentanalyse).

# (2 Punkte pro Stichwort)



**FRAGE 39 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**



Nennen und erläutern Sie kurz drei der wichtigsten Gesichtspunkte zum Begriff „Big Data“.

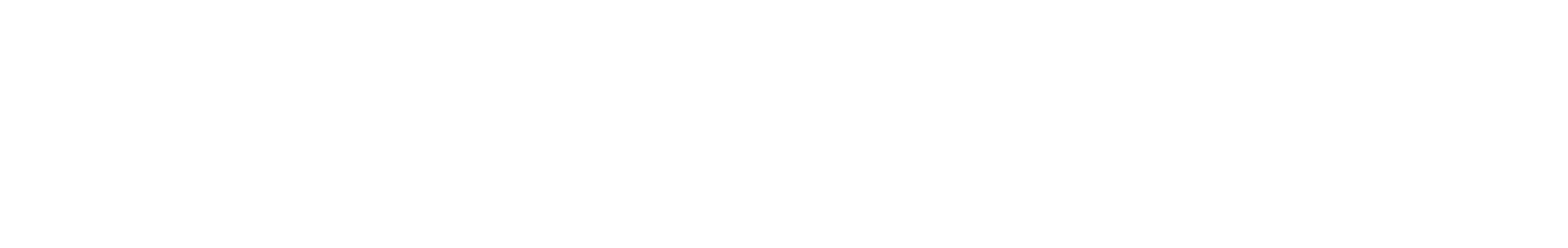
Die Daten, die auf den gängigen Social-Media-Plattformen generiert werden, lassen sich unter den folgenden drei Aspekten betrachten:

* Menge (Volume): Die große Anzahl von Nutzenden sozialer Medien und ihre Aktivitäten in diesen Netzwerken haben eine große Menge an Daten über soziale Plattformen erzeugt. **(2 Punkte)**
* Geschwindigkeit (Velocity): Dieses Konzept betrifft die Geschwindigkeit, mit der Daten generiert werden. Es wird beispielsweise geschätzt, dass Twitter-Nutzende 500 Millionen Tweets pro Tag verfassen. Mit Hilfe von Mining-Werkzeugen können Daten schon bei ihrer Erzeugung analysiert werden, was die Erstellung von Berichten in Echtzeit ermöglicht. **(2 Punkte)**
* Vielfältigkeit (Variety): Die auf sozialen Medienplattformen generierten Daten liegen in einem unstrukturierten Format vor. Dabei kann es sich um Text, Videodateien, Bilder usw. handeln. **(2 Punkte)**



**FRAGE 40 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**



Zwei Sätze sind vorgegeben:

Satz 1: „Das Buch war interessant, mit einer guten Struktur.“ Satz 2: „Fantastische Geschichte, gutes Buch.“

1. Legen Sie fest, aus welchen Wörtern unser Wortschatz zunächst bestehen soll.
2. Überlegen Sie, aus wie vielen Wörtern sich der Wortschatz zusammensetzen würde, wenn wir irrelevante Wörter wie z. B.: „das“, „war“, „mit“ und „einer“ entfernen würden.
3. Wie sehen die Häufigkeitsvektoren für den ersten und zweiten Satz nach der Entfernung irrelevanter Wörter aus?
   1. Der Wortschatz besteht zunächst aus den folgenden Wörtern: „das“, „Buch“, „war“, „interessant“, „mit“, „einer“, „guten“, „Struktur“, „fantastische“, „Geschichte“, „gutes“. **(1 Punkt)**
   2. Nach dem Entfernen irrelevanter Wörter würde der Wortschatz aus 7 Wörtern bestehen. **(3 Punkte)**
   3. Der Häufigkeitsvektor für den ersten Satz wäre Satz\_1 = [1 1 1 1 0 0 0]. **(3 Punkte)**

Der Häufigkeitsvektor für den zweiten Satz wäre Satz\_2 = [1 0 0 1 0 1 1] **(3 Punkte)**



**FRAGE 41 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**



Nennen und beschreiben Sie kurz vier Unterschiede zwischen der Analyse sozialer Medien (Social Media Analytics, SMA) und der Geschäftsanalyse.

1. SMA konzentriert sich auf die Interpretation von Daten, die von Social-Media-Plattformen gesammelt werden, um die Hauptziele und Strategien des Unternehmens zu erreichen. Geschäftsanalysen hingegen befassen sich mit der Interpretation der Unternehmensdaten, die regelmäßig in den Managementsystemen des Unternehmens gespeichert werden. **(3 Punkte)**
2. Bei der Geschäftsanalyse arbeiten wir in der Regel mit gut strukturierten Daten (die in Unternehmensdatenbanken gespeichert sind), oder mit Daten, die aus anderen Quellen in einem bekannten Format stammen. Auf der anderen Seite werden bei der Analyse sozialer Medien die Daten in verschiedenen Formen gesammelt (Text, Emoticons, Bilder, Videos, Hyperlinks usw.). **(2 Punkte)**
3. Geschäftsanalysen sind eng mit vertraulichen Daten verknüpft, die von einem Unternehmen gesteuert werden und in den Datenbanken des Unternehmens gespeichert sind. Dies führt zu mehr bürokratischen Verfahren und weniger Flexibilität bei der Verarbeitung dieser Daten. Daten aus sozialen Medien sind in der Regel für jeden zugänglich. **(2 Punkte)**
4. Während die traditionellen Medienformen auf dem Informationsfluss vom Herausgebenden zum Nutzenden (Verbrauchenden) beruhen, haben soziale Medienplattformen ein neues Modell geschaffen, bei dem der Nutzende sowohl Herausgebender als auch Verbrauchender ist. Bei SMA hängt der Wert der Daten davon ab, wie häufig diese geteilt bzw. verbreitet werden, wohingegen die Daten bei der Geschäftsanalyse nur für das Unternehmen von Bedeutung sind, das sie besitzt. **(3 Punkte)**



**FRAGE 42 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**



Nennen und erklären Sie kurz die vier Schritte, die für die Vorverarbeitung von Text erforderlich sind.

Entfernung von Stoppwörtern. Bei der Anwendung dieser Methode ist es wichtig, dass die Bedeutung des Satzes nicht verändert wird.

Stemming (Stammformreduktion) ist eine Methode, bei der die Endungen verschiedener Wörter entfernt werden (z. B. ergibt die Entfernung der Endung „es“ aus dem Wort „Wortes“ den Wortstamm „Wort“).

Lemmatisierung zielt darauf ab, Flexionsendungen zu entfernen, und liefert ein sogenanntes „Lemma“ (die Wortbasis bzw. den Wortstamm). Lemmatisierung ist genauer als Stemming, weil sie auf der Bedeutung des Wortes basiert. Diese Methode behandelt unregelmäßige Fälle effektiver. Im Falle des Wortes „bemerkt“ zum Beispiel identifiziert die Lemmatisierung „bemerken“ als die Grundform des Wortes. Durch das Stemming würde nur das „t“ wegfallen, wodurch dieses Wort in „bemerk“ umgewandelt würde, was selbst keine echte Bedeutung hat.

Rechtschreibfehler korrigieren.

# (2 Punkte pro Schritt)



**FRAGE 43 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**



Nennen und erklären Sie die fünf wichtigsten Schritte des SMA-Verfahrens.

1. Identifikation: Der erste Schritt besteht darin, die wichtigsten Informationen zu ermitteln. Bei der Ermittlung der relevantesten Informationsquellen ist es hilfreich, eine genaue Vorstellung des Ziels zu haben.
2. Datensammlung: Nach der Überlegung, welche Daten für ein gewünschtes Ergebnis gesammelt werden müssen, kann die Sammlung der Daten beginnen. Je nach Art der zu sammelnden Daten (Text, Bilder, Videos) können verschiedene Werkzeuge für die Extraktion verwendet werden. Für die Extraktion von Daten aus sozialen Medienplattformen kommen in den meisten Fällen API-Aufrufe (Application Programming Interface) zum Einsatz.
3. Datenbereinigung und ‑vorverarbeitung: Die gesammelten Daten sollten bereinigt werden, um unerwünschte Informationen zu entfernen. Bei Textdaten können durch Techniken zur Verarbeitung natürlicher Sprache unwichtige Informationen aus dem extrahierten Text entfernt werden.
4. Analyse der Daten: In dieser Phase können die vorverarbeiteten Daten analysiert werden, um daraus mögliche Erkenntnisse zu gewinnen.
5. Präsentation und Interpretation der Ergebnisse: Nach der Analyse der gesammelten Daten können diese in einer Form präsentiert werden, die für ein allgemeines Publikum leichter verständlich ist. Die gewonnenen Erkenntnisse lassen sich durch Visualisierungen wirkungsvoll an interessierte Parteien weitergeben und ihnen erklären.

# (2 Punkte pro Schritt)



**FRAGE 44 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**



Nennen und erklären Sie kurz drei Probleme, die bei der Anwendung der BoW-Methode auftreten können.

Seltenheit / dünne Besetzung. Vektoren mit vielen Nullen werden als dünnbesetzte Vektoren bezeichnet. Dünnbesetzte Vektoren können den Prozess erschweren, da sie erhebliche Rechenressourcen erfordern.

Vokabeln. Um einen weniger dünn besetzten Vektor zu erhalten, müssen die ins Vokabular aufzunehmenden Wörter sorgfältig ausgewählt werden.

Bedeutung. Bei Einsatz von BoW wird die Wortfolge oder die Struktur eines Satzes nicht berücksichtigt, was zu Problemen im allgemeinen Kontext des Dokuments führen kann.

# (2 Punkte pro Antwort)



**FRAGE 45 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**



Angenommen, die folgenden Werte für die Termhäufigkeit und die inverse Dokumentenhäufigkeit seien gegeben: TF(„Bleistift“, Dokument A) = 0,6

TF(„Bleistift“, Dokument B) = 0,7 IDF(„Bleistift“) = 4

1. Welchen Wert hat TF\_IDF(„Bleistift“, Dokument B, Korpus) auf Grundlage der oben genannten Informationen?
2. Geben Sie die Formel und den Endwert an.
   1. TF\_IDF(„Bleistift“, Dokument B, Korpus) = TF(„Bleistift“, Dokument B) \* IDF(„Bleistift“, Korpus) **(3 Punkte)**
   2. TF\_IDF = 0,7 \* 4 **(3 Punkte)** = 2,8 **(2 Punkte)**



**FRAGE 46 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



Erklären Sie, was der folgende Code bewirkt:

for line in file:

tweet = json.loads(line)

entities = tweet.get('entities', {})

h = entities.get('user\_mentions', [])

users\_in\_tweet = [tag['screen\_name'].lower() for tag in h]

Der Code zeigt eine Sammlung von Benutzer-Erwähnungen in der Zeitleiste eines bestimmten Nutzenden. **(3 Punkte)** Er lädt alle Zeilen einer json-Datei. **(2 Punkte)** Der obige Code greift auf die meistgenannten Nutzenden im Profil, von dem aus gesucht wird, zu. **(3 Punkte)**



**FRAGE 47 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



Erklären Sie, was der folgende Code bewirkt: import matplotlib.pyplot as plt

y = [count for t, c in tf.most\_common(20)] x = range(1, len(y)+1)

plt.bar(x, y) plt.ylabel(„Häufigkeit") plt.savefig('Verteilung.png')

Mit Hilfe der Bibliothek matplotlib wird **die Häufigkeit der Begriffe grafisch dargestellt**. Es wird ein **Diagramm mit 20 der häufigsten Begriffe und ihrer jeweiligen Häufigkeit** dargestellt. Die **Spaltenbezeichnung ist „Häufigkeit“.** Das Diagramm wird in einer **Datei namens „Verteilung.png“** gespeichert.

# (2 Punkte pro Stichwort)



**FRAGE 48 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**



Nennen Sie drei der wichtigsten Python-Pakete oder ‑Funktionen, die zur Implementierung von Tf-idf verwendet werden können.

1. Pandas-Paket
2. sklearn-Paket
3. TfidfVectorizer()-Funktion
4. fit\_transform()-Funktion
5. get\_feature\_names()-Funktion

# (2 Punkte für jede Antwort)



**FRAGE 49 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



Erklären Sie, was A/B-Testen bedeutet.

Die einfachste Art des Experimentierens sind so genannte A/B-Tests. A steht für die ursprüngliche Testvariable und B für die neue Version der ursprünglichen Variablen. **(1 Punkt)** A/B-Tests sind auch unter dem Begriff Split-Tests bekannt. Verschiedenen Personen werden verschiedene Versionen einer Website gezeigt, um zu ermitteln, mit welcher Variante am effektivsten neue Kund:innen gewonnen werden können. **(1 Punkt)** A/B-Tests können als Experiment betrachtet werden, dessen Ergebnisse die beste Option für Nutzende zur Erreichung der Hauptziele einer Website bestimmen. Die Messung der direkten Auswirkungen von Änderungen an der Website auf die Web-Metriken ermöglicht es, diejenigen Änderungen vorzunehmen, die sich positiv auswirken. **(2 Punkte)**

Durch Experimentieren mit Bildern, Schaltflächen oder anderen Gestaltungsmerkmalen lässt sich überprüfen, welche Version den besten Wert für die gewählte Metrik bietet. Vor dem Experimentieren mit den Komponenten einer Website sollte ein genauer Plan erstellt werden. **(2 Punkte)**



**FRAGE 50 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**



Die Funktionen defaultdict(), CountVectorizer() , fit\_transform() und get\_feature\_name() gehören zu den wichtigsten Python-Paketen, mit denen die Bag-of-Words-Technik implementiert werden kann.

1. Nennen Sie zwei weitere wichtige Pakete, die diese Funktion erfüllen können.
2. Nennen Sie für jedes Paket eine Funktion, mit der die Bag-of-Words-Technik implementiert werden kann, und erläutern Sie diese kurz.
3. Numpy-Paket:

Dies ermöglicht die Verarbeitung von Datenstrukturen in einer feldähnlichen Form (Array). Bei der BoW-Technik gibt die Funktion np.zeros beispielsweise ein neues, mit Nullen gefülltes Feld (Array) einer bestimmten Länge zurück.

1. NLTK-Paket:

Dies ist die wichtigste Python-Bibliothek, die zur Verarbeitung von Textinformationen genutzt wird. Bei der BoW-Technik wird zum Beispiel nltk.tokenize importiert, um den Text zu tokenisieren.

1. sklearn-Paket:

Eine Bibliothek für maschinelles Lernen in Python, die verschiedene Klassifizierungsalgorithmen unterstützt. Die wichtigste Funktion zur Erstellung des BoW-Modells in sklearn ist CountVectorizer(). Sie legt den n-Gramm-Bereich fest, in dem die Wörter tokenisiert werden sollen.

1. Pandas-Paket:

Pandas bietet eine bessere Enddarstellung der Vektoren. Mit Hilfe von Pandas-Datagrame() werden die Daten strukturiert und in einem gut strukturierten Format gedruckt.

# (1 Punkt für den Namen des Pakets, maximal 2 Punkte)

**(2 Punkte für die Erklärung der Funktion, maximal 4 Punkte)**



**FRAGE 51 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**



Erstellen Sie ein Python-Code-Snippet zur Berechnung des euklidischen Abstands zwischen zwei Vektoren. Verwenden Sie Python-Bibliotheken (numpy, scipy usw.).

Antwortmöglichkeit 1:

from scipy.spatial.distance import Euclidean **(3 Punkte für die Bestimmung des Pakets)**

review1 = [1 1 1 1 1 1 0 0 0 0 0] **(1 Punkt für die Bestimmung eines ersten Vektors)**

review2 = [0 0 1 0 1 0 1 1 2 0 0] **(1 Punkt für die Bestimmung eines zweiten Vektors)**

distance = euclidean(review1, review2) **(3 Punkte für die Bestimmung der Funktion)**

Antwortmöglichkeit 2:

import numpy as np **(3 Punkte zur Bestimmung des Pakets)**

point1 = np.array((1, 2, 3)) **(1 Punkt zur Bestimmung eines ersten Vektors)**

point2 = np.array((1, 1, 1)) **(1 Punkt zur Bestimmung eines zweiten Vektors)**

sum\_sq = np.sum(np.square(point1 - point2)) **(3 Punkte für die Bestimmung der Funktion)**

print(np.sqrt(sum\_sq))



**FRAGE 52 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**



Nennen und erklären Sie die vier Schritte, die bei der Durchführung eines A/B-Tests befolgt werden sollten.

1. Verstehen des Zwecks der Website.

Vor dem Beginn der Experimentierphase muss der Hauptzweck der Website klar sein. Welche Aspekte könnten auf der Website noch optimiert werden? Der Zweck der Website und die möglichen Optimierungsprozesse müssen in quantifizierbare Metriken umgesetzt werden. Wir müssen die Metriken definieren, die durch die Durchführung verschiedener Tests verbessert werden sollen. So können wir die Endergebnisse leichter überprüfen und die beste Variante unserer Website bestimmen.

1. Sammeln von Daten und Identifizieren möglicher Engpässe. Neben den wichtigsten Erfolgspunkten ist es auch wichtig, die Stellen der Website zu identifizieren, an denen häufig Kund:innen oder Nutzende die Website verlassen (mögliche Engpässe).
2. Definieren einer Hypothese.

Nachdem wir die Erfolgskennzahlen und mögliche Engpässe auf unserer Website ermittelt haben, sollten wir Hypothesen formulieren, die mögliche Wege zum Erreichen des eigentlichen Ziels der Website aufzeigen. Vor Beginn eines Experiments sollte eine Hypothese aufgestellt werden, um festzustellen, ob wir den richtigen Weg zum Erreichen der vordefinierten Erfolgsmetriken gefunden haben. Auch wenn ein Test fehlschlägt und negative Ergebnisse liefert, kann er zumindest zu anderen möglichen Hypothesen führen.

1. Starten der Tests.

Nachdem wir die möglichen Hypothesen festgelegt haben, müssen wir in der Reihenfolge der Prioritäten mit dem Experimentieren beginnen. Die Besucher:innen werden nach dem Zufallsprinzip ausgewählt und bekommen verschiedene Varianten der Website angeboten. Ihre Benutzererfahrung wird nachverfolgt, um diejenige Variante zu finden, die bessere Ergebnisse bei den definierten Erfolgsmetriken verspricht. Die Durchführung eines Tests kann zu notwendigen weiteren Experimenten führen, wodurch ein Kreislauf entsteht, der die Website zu ihrer bestmöglichen Version führt.

# (2 Punkte für jeden Schritt)



**FRAGE 53 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



Nennen und erläutern Sie kurz drei Vorteile und einen Nachteil von A/B-Tests.

A/B-Tests sind eine der **einfachsten (1 Punkt)** und **kostengünstigsten (1 Punkt)** Experimentiertechniken. Sie sind die beste Lösung für Unternehmen, die **gerade erst mit dem Testen beginnen**. **(2 Punkte)** A/B-Tests **erfordern in der Regel keine umfangreichen Ressourcen (2 Punkte)**, und die **erzielten Ergebnisse sind meist relativ einfach zu interpretieren (2 Punkte)**.

Das Testen von Seiten mit mehreren Varianten ist jedoch schwieriger zu interpretieren. Es ist leicht, die „Sieger-Testseite“ zu finden, aber es ist **unmöglich zu verstehen, welche genauen Elemente hinter diesem Erfolg stehen und warum. (2 Punkte)**



**FRAGE 54 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**



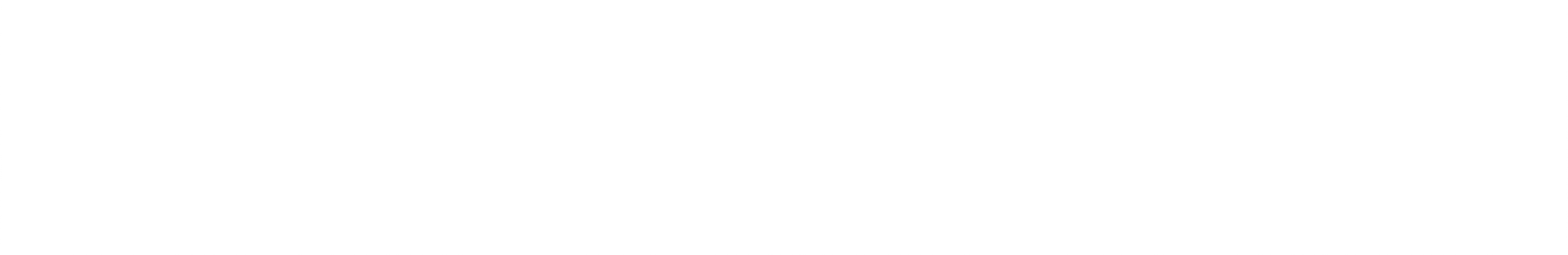
Nennen und erläutern Sie kurz zwei beliebte Techniken zur semantischen Analyse.

* 1. Die latente semantische Analyse (LSA) **(1 Punkt)** ist eine unüberwachte Technik in der natürlichen Sprachverarbeitung für die Informationssuche, welche die Beziehung zwischen einer Reihe von Wörtern oder Dokumenten analysiert. LSA ist ein Algorithmus, der mit Bag-of-Words-Vektoren oder TF-IDF-Vektoren funktioniert. **(3 Punkte)**
  2. Die LDA-Methode (Latent Dirichlet Allocation) **(1 Punkt)** nutzt die Dirichlet-Verteilung, eine Wahrscheinlichkeitsverteilung. Das LDA-Modell geht davon aus, dass jedes der Themen, aus denen ein Dokument in einem Korpus besteht, mit einer Wahrscheinlichkeitsverteilung über den Wörtern verteilt ist. Wenn dem LDA-Algorithmus ein aus verschiedenen Dokumenten bestehendes Korpus gegeben wird, versucht er herauszufinden, welche Themen diese Dokumente enthalten. **(3 Punkte)**



**FRAGE 55 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**



Gegeben seien die folgenden Werte:

Anzahl der Wörter im Dokument 1: 100

Anzahl der Wörter in Dokument 1 zum Thema 1: 20 α = 0,1

Anzahl der Themen: 3

1. Geben Sie die Formel zur Berechnung des Zusammenhangs zwischen Thema 1 und Dokument 1, also p(Thema 1 | Dokument 1).
2. Berechnen Sie das Ergebnis für die angegebenen Werte.

p(Thema t|Dokument d) = (Anzahl\_dt + α)/(Anzahl\_d – 1 + Anzahl\_t \* α) **(3 Punkte für die Ermittlung der Formel)**

p(Thema t|Dokument d) = (20+0,1) / (100 – 1 + 3 \* 0,1) = 0,2 **(2 Punkte für die Berechnung des Zählers) (2 Punkte für die Berechnung des Nenners)**

# (3 Punkte für die Berechnung des Quotienten)



**FRAGE 56 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**



Nennen und beschreiben Sie kurz vier Komponenten einer Webseite, auf der ein A/B-Test durchgeführt werden kann.

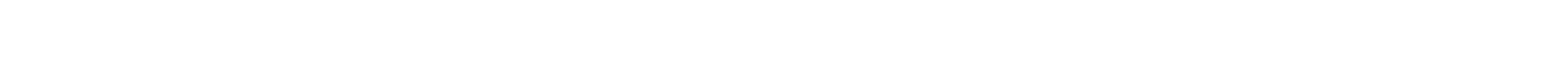
1. Überschriften: Da Überschriften zu den ersten Dingen gehören, welche die Nutzenden einer Website sehen, sollte die beste Kombination von Wörtern (nicht zu kurz und nicht zu lang) ermittelt und mit verschiedenen Textgrößen und Schriftarten getestet werden.
2. Inhalt: Besucher:innen lassen sich in zwei Hauptgruppen einteilen: diejenigen, die gerne lange Texte lesen, und diejenigen, die kurze Zusammenfassungen bevorzugen. Mit A/B-Tests können wir experimentieren und eine Variante mit langem Inhalt und eine andere mit kürzerem Inhalt erstellen. Die Ergebnisse der A/B-Tests zeigen, welche der Versionen von den Nutzenden besser angenommen wird. So ergibt sich eine Version, die zu höheren Konversionsraten, niedrigeren Absprungraten usw. führt.
3. Design und Layout: Design und Layout einer Seite sollten die Besucher:innen nicht verwirren und ihnen klare Informationen bieten. Der Inhalt sollte einfach und übersichtlich sein. Da die Startseite die erste besuchte Seite ist, muss sie perfekt optimiert werden. Deshalb sollte sie die erste Seite sein, die A/B-Tests unterzogen wird. Es können verschiedene Bilder, Videos und Textstile getestet werden, um die Variante mit der höchsten Konversionsrate zu finden.
4. Navigationsleisten. Die Navigation sollte strukturiert und einfach zu verstehen sein, so dass die Nutzenden die gesuchten Informationen leicht finden können.
5. Rezensionen. Die Anzeige von Bewertungen (Rezensionen), Empfehlungen und Referenzen auf einer Website kann für den Betreiber der Website von direktem Nutzen sein. Mit A/B-Tests lässt sich feststellen, ob es für ein Unternehmen eine gute Idee ist, Rezensionen auf seine Website zu stellen.

# (2 Punkte pro Komponente, max. 4 Komponenten)



**FRAGE 57 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



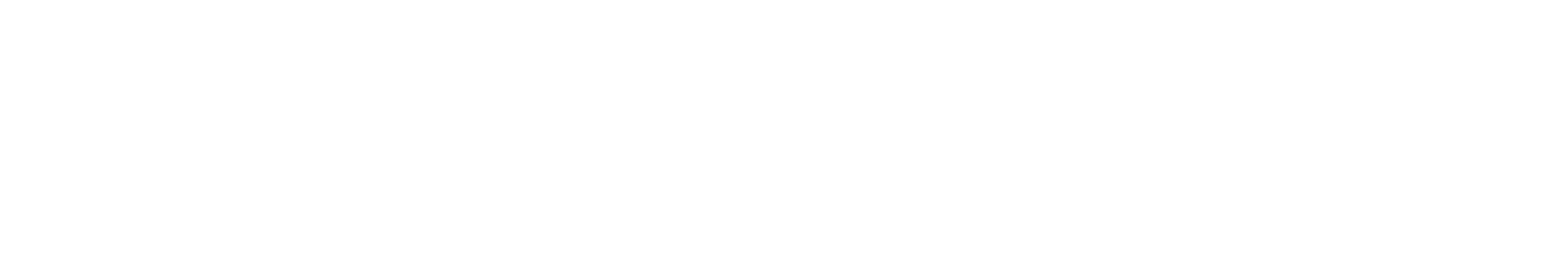
Erklären Sie, warum multivariate Tests (MVT) ein hohes Risiko des Scheiterns bergen.

A/B-Tests können zum Erfolg oder Misserfolg führen, aber in beiden Fällen haben wir die Möglichkeit, etwas Neues über die Komponenten einer Website zu lernen. **(2 Punkte)** Im Falle von MVT ist es überhaupt nicht rentabel, mehrere Kombinationen zu testen und kein Ergebnis zu erzielen. **(2 Punkte)** Der Zeit- und Ressourcenaufwand für die Durchführung der Tests ist im Vergleich zu A/B-Tests zu hoch. **(2 Punkte)**



**FRAGE 58 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**



Gegeben seien die folgenden Werte:

Anzahl der Erwähnungen des Wortes „Auto“ in Thema 1: 40 Gesamtzahl der Wörter in Thema 1: 70

β = 0,1

Anzahl der Wörter im Vokabular: 10

1. Geben Sie die Formel zur Berechnung der Signifikanz des Wortes „Auto“ in Thema 1 an: p(Wort „Auto“

| Thema 1)

1. Berechnen Sie das Ergebnis für die angegebenen Werte.

p(Wort w| Thema t) = (Anz\_wt + β)/(ƩAnz\_wt – 1 + Anz\_v \* β) **(3 Punkte für die Bestimmung der Formel)**

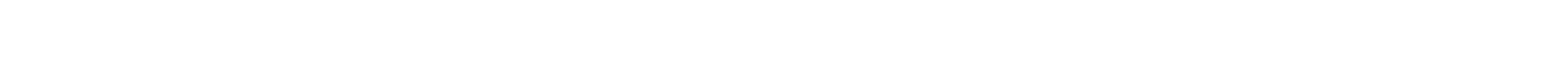
p(„Auto“| Thema 1) = (40 + 0,1) / (70 – 1 + 10\*0,1) = 0,57 **(2 Punkte für die Berechnung des Zählers) (2 Punkte für die Berechnung des Nenners)**

# (3 Punkte für die Berechnung des richtigen Wertes)



**FRAGE 59 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 02**



Definieren und erklären Sie die Formel, mit der bei der LDA-Methode einem Wort ein neues Thema zugeordnet wird.

p(w| t,d) = p(Thema t |Dokument d) \* p(Wort w | Thema t) **(3 Punkte)**

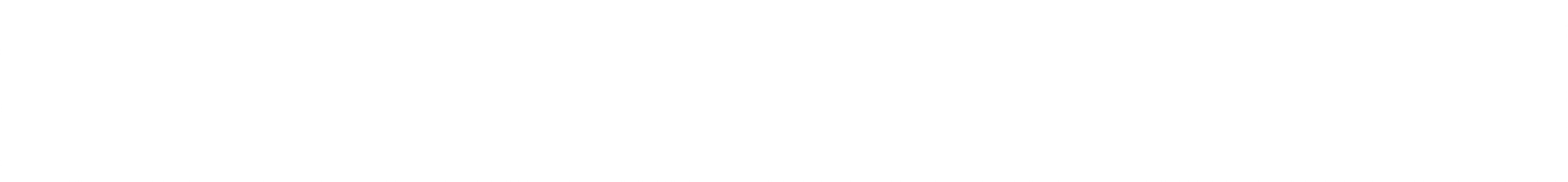
Die Wahrscheinlichkeit, dass ein Thema zu einem Dokument gehört, wird multipliziert mit der Wahrscheinlichkeit, dass ein Wort zu einem bestimmten Thema gehört. **(3 Punkte)**

Das Wort wird dem neuen Thema zugeordnet, bei dem die höchste Wahrscheinlichkeit errechnet wurde. **(2 Punkte)**



**FRAGE 60 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



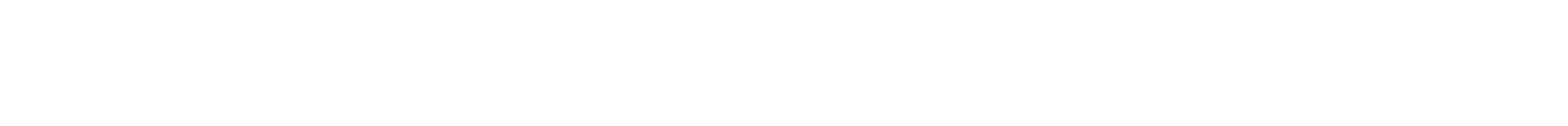
Angenommen wir haben eine Website für ein Online-Unternehmen. Die Startseite der Website hat zwei wichtige Elemente: Bilder und Schaltflächen. Wir entscheiden uns für zwei Versionen jedes Elements, um zu prüfen, welche Kombination besser abschneidet.

Bestimmen Sie die Formel, mit der die Anzahl der möglichen Kombinationen berechnet wird, und ermitteln Sie damit den Wert für diesen Fall.

Anzahl der Variationen = Anzahl der Versionen des ersten Elements \* Anzahl der Versionen des zweiten Elements \* Anzahl der Versionen des dritten Elements **(3 Punkte)**

Anzahl der Variationen = 2 \* 3 = 4 Kombinationen **(3 Punkte)**

Insgesamt erhalten wir 4 verschiedene Kombinationen für die Website.**(3 Punkte)** Jede dieser Kombinationen wird getestet, um diejenige mit der potenziell höchsten Konversionsrate zu finden. **(1 Punkt)**



1. Erklären Sie, wofür die Funktion TfidfVectorizer() in der Python-Implementierung des LDA-Algorithmus verwendet wird.
2. Nennen Sie zwei Parameter, die an diese Funktion übergeben werden können.



**FRAGE 61 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**

* 1. Die Funktion TfidfVectorizer wandelt eine Sammlung von Rohdokumenten in eine Matrix von Tf-idf-Merkmalen um.

# (3 Punkte)

* 1. Zu den Parametern, die der Funktion übergeben werden können, gehören: use\_idf, max\_features, smooth\_idf, stopwords, lowercase, ngram\_range, max\_df, min\_df usw.) **(1,5 Punkte pro Parameter, maximal 3 Punkte)**



**FRAGE 62 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**



1. Erklären Sie, wofür die Funktion TruncatedSVD() in der Python-Implementierung des LSA-Algorithmus verwendet wird.
2. Nennen Sie zwei Parameter, die an diese Funktion übergeben werden können.

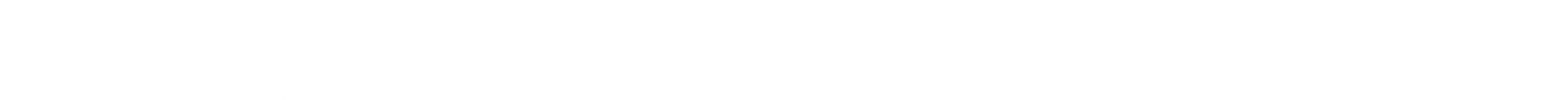
Die Funktion TruncatedSVD wird für die Dimensionalitätsverringerung mittels SVD (Singulärwertzerlegung) verwendet. Sie ist Teil des sklearn-Pakets und wird zur Implementierung von LSA genutzt. **(3 Punkte)**

Einige der Parameter, die an diese Funktion übergeben werden können, sind: n\_components, algorithm, n\_iter, random\_state, tol) **(1,5 Punkte pro Parameter, maximal 3 Punkte)**



**FRAGE 63 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**



1. Erklären Sie, wie SVD die Dimensionen verringert.
2. Benennen und bestimmen Sie die Dimensionen jeder der erstellten Matrizen.
3. Die Singulärwertzerlegung (SVD) ist eine Methode der Matrixfaktorisierung. Die Singulärwert-Zerlegung einer Matrix B mit m x n wäre: B = P \* L \* K^T **(3 Punkte)**, mit:

B → Matrix mit m x n

P → m x k linke singuläre Matrix

L → k x k diagonale Matrix

K → n x k rechte singuläre Matrix K^T → k x n, die Transponierte von K **(3 Punkte)**

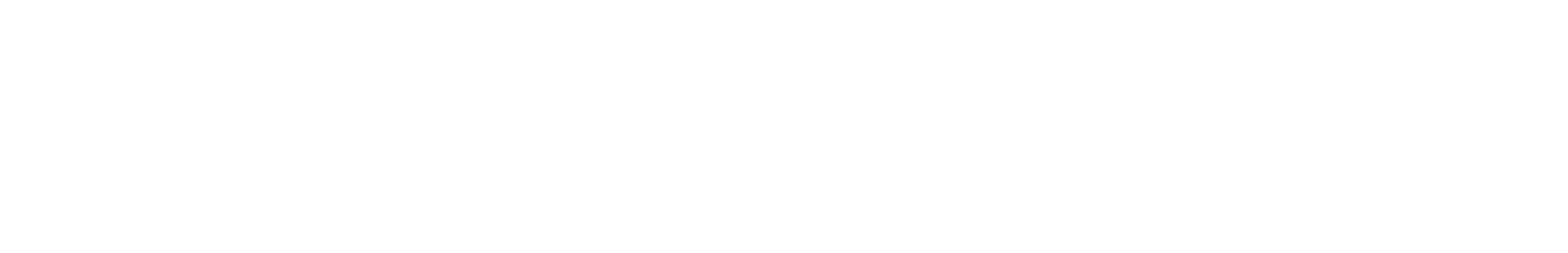
1. In diesem Fall hat SVD die Dimensionen durch Auswahl eines k-Wertes verringert. **(1 Punkt)**

k ist ein Parameter, der die Anzahl der Themen angibt, die wir finden möchten. Dieser Parameter kann mit Hilfe eines Silhouettenkoeffizienten oder der Themenkohärenz bestimmt werden. **(3 Punkte)**



**FRAGE 64 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**



Gegeben seien die folgenden Werte:

p(Thema 1 | Dokument 1) = 0,04

p(Thema 2 | Dokument 1) = 0,2

p(„Zuhause“|Thema 1) = 0,78

p(„Zuhause“|Thema 2) = 0,55

1. Berechnen Sie die neue Wahrscheinlichkeit, mit der das Wort „Zuhause“ mit jedem Thema in Verbindung gebracht werden kann.
2. Welchem Thema wird das Wort „Zuhause“ schließlich zugeordnet?

p(„Zuhause“, Thema 1) = p(Thema 1|Dokument 1) \* p(„Zuhause“| Thema 1) = 0,04 \* 0,78 = 0,0312 **(2 Punkte für die Ermittlung der Formel) (2 Punkte für die Berechnung des richtigen Wertes)**

p(„Zuhause“, Thema 2)= p(Thema 2|Dokument 1) \* p(„Zuhause“| Thema 2) = 0,2 \* 0,55= 0,11

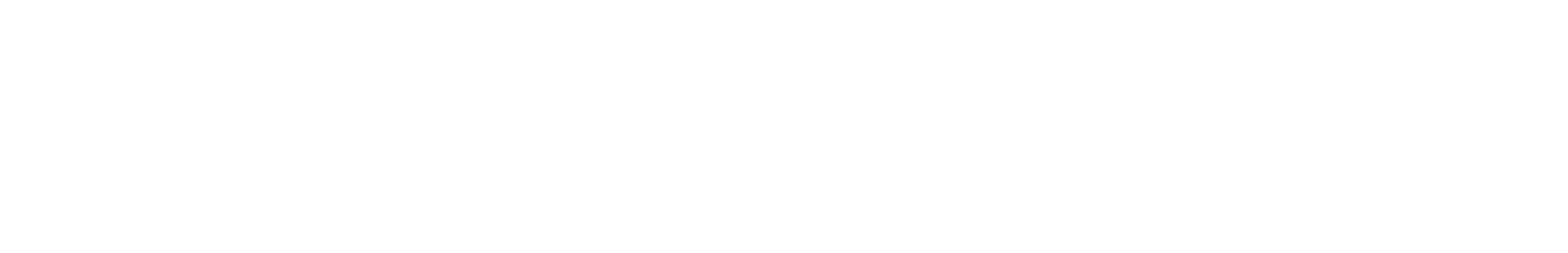
# (2 Punkte für die Ermittlung der Formel) (2 Punkte für die Berechnung des richtigen Wertes)

Das Wort „Zuhause“ wird dem zweiten Thema zugeordnet, da hierfür die Wahrscheinlichkeit höher ist. **(2 Punkte)**



**FRAGE 65 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 02**



Gegeben seien die folgenden Werte: p(Thema 1 | Dokument 1) = 0,32

p(Thema 2 | Dokument 1) = 0,21

p(„Telefon“|Thema 1) = 0,24

p(„Telefon“|Thema 2) = 0,56

1. Berechnen Sie die neue Wahrscheinlichkeit, mit der das Wort „Telefon“ mit jedem Thema in Verbindung gebracht werden kann.
2. Welchem Thema wird das Wort „Telefon“ schließlich zugeordnet?

p(„Telefon“, Thema 1) = p(Thema 1|Dokument 1) \* p(„Telefon“|Thema 1) = 0,32\*0,24 = 0,0768

# (2 Punkte für die Ermittlung der Formel) (2 Punkte für die Berechnung des richtigen Wertes)

p(„Telefon“, Thema 2) = p(Thema 2|Dokument 1) \* p(„Telefon“|Thema 2) = 0,21\*0,56 = 0,1176

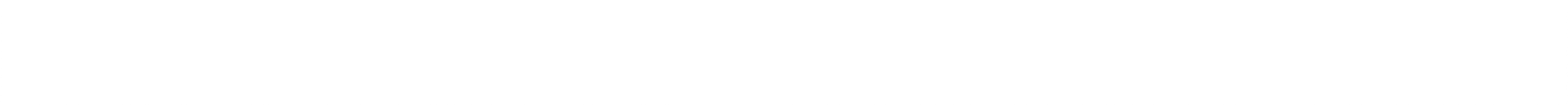
# (2 Punkte für die Ermittlung der Formel) (2 Punkte für die Berechnung des richtigen Wertes)

Das Wort „Telefon“ wird dem zweiten Thema zugeordnet, da hierfür die Wahrscheinlichkeit höher ist. **(2 Punkte)**



**FRAGE 66 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 02**

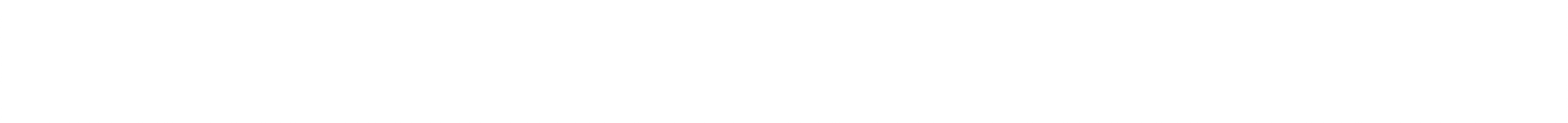


1. Definieren Sie kurz die Funktion fit\_transform() .
2. Welche Aufgabe erfüllt fit\_transform() ?
3. Die Methode fit\_transform() aus der Sklearn-API erstellt eine Matrix der Dokumente im Korpus und berechnet das Tf-idf-Maß für jedes Wort. **(3 Punkte)**
4. Sie berechnet den Mittelwert und die Varianz jedes der in unseren Daten vorhandenen Merkmale. **(3 Punkte)**



**FRAGE 67 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



1. Erklären Sie die Metrik „Ausstiegsrate (Exit Rate)“.
2. Wie können wir unterscheiden zwischen den Nutzenden, welche die Website betreten und keine Aktion durchgeführt haben, und den Nutzenden, die gefunden haben, was sie suchten, und dann die Website wieder verlassen haben?
3. Mit der Ausstiegsrate wird die Anzahl der Personen berechnet, die eine Website von einer bestimmten Seite aus verlassen haben. Sie kann den Zeitpunkt oder die Seite der Website anzeigen, an dem ein Nutzender das Interesse verloren und die Seite verlassen hat. Obwohl diese Kennzahl als Leck einer Website angesehen werden kann, zeigt sie nicht immer ein schlüssiges Ergebnis. **(3 Punkte)**
4. Die Ausstiegsrate sollte zusammen mit anderen wichtigen Metriken wie der Absprungrate und der Konversionsrate berechnet und ausgewertet werden. Der Hauptunterschied zwischen Ausstiegsrate (Exit Rate) und Absprungrate (Bounce Rate) besteht darin, dass die Absprungrate auf den Sitzungen basiert, die auf einer bestimmten Seite beginnen und enden. Andererseits ist die Ausstiegsrate der Prozentsatz der zuletzt in einer Sitzung aufgerufenen Seiten. **(3 Punkte)**



**FRAGE 68 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



Erklären Sie anhand eines Beispiels, was eine Konversionsrate ist und warum ihre Messung wichtig ist.

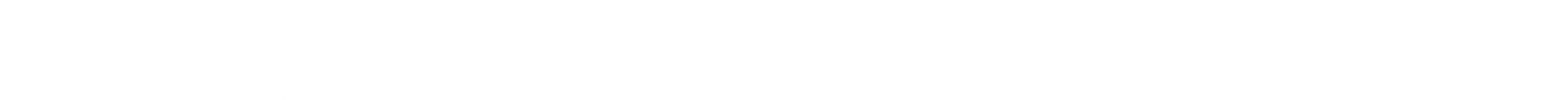
Um besser zu verstehen, was Konversionsraten sind, können wir ihre Nutzung auf einer E-Commerce-Website (elektronischer Handel) beschreiben. Die Konversionsrate beschreibt den Erfolg (also die getätigten Bestellungen im elektronischen Handel) geteilt durch die Anzahl der Besuche (visits) bzw. Einzelbesucher:innen (unique visitors). Je nach Zielsetzung können wir uns für Besucher:innen oder Einzelbesucher:innen entscheiden. **(3 Punkte)**

Die Konversionsrate ist hilfreich, um zu verstehen, ob die Besucher:innen der Website ein Produkt kaufen oder eine Bestellung aufgeben möchten. Wir können Besuche wählen, wenn wir annehmen, dass ein Nutzender bei jedem Besuch eine Bestellung aufgibt, und Einzelbesucher:innen, wenn wir glauben, dass Nutzende die Website mehrmals besuchen, bevor sie eine Bestellung aufgeben. **(3 Punkte)**



**FRAGE 69 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**

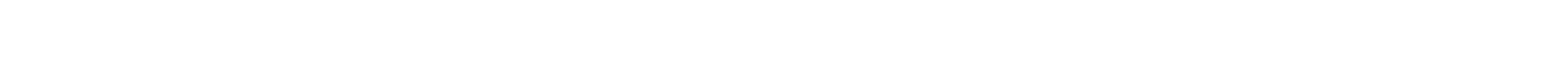


1. Erklären und veranschaulichen Sie die qualitative Dimension der Metrik zur Kundenbindung.
2. Bestimmen Sie zwei Methoden, die zur qualitativen Messung der Kundenbindung verwendet werden können.
3. Nutzende können eine positive oder negative Erfahrung mit einer Website machen. **(2 Punkte)** Wenn ein Nutzender beispielsweise nach einem bestimmten Artikel sucht und ihn aufgrund der unübersichtlichen Anordnung der Informationen auf der Website nicht findet, scheint er zwar gebunden zu sein, ist aber überhaupt nicht zufrieden. Die Bindung als emotionales Engagement der Nutzenden zu sehen, ist ziemlich schwierig, weil keine Analysewerkzeuge die Gedanken und Emotionen der Nutzenden beim Navigieren auf einer Website messen können. **(3 Punkte)**
4. Mögliche Methoden zur qualitativen Messung der Bindung sind z. B. die Durchführung von Online-Umfragen bei Kund:innen, die Nutzung von Daten aus der Marktforschung oder die Berechnung der Kundenbindung über einen längeren Zeitraum. **(1,5 Punkte pro richtige Antwort, maximal 3 Punkte)**



**FRAGE 70 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



Erklären Sie die quantitative Dimension der Metriken zur Kundenbindung.

Die Bindung auf der Website kann von einer geringen Bindung bis hin zu einer hohen Bindung reichen. Es ist einfach, der Bindung von Nutzenden durch eine quantitative Metrik einen Wert zuzuordnen. **(3 Punkte)**

Nutzende besuchen entweder die Website und verbringen viel Zeit mit der Suche und dem Lesen der Inhalte, oder sie verbringen nicht viel Zeit damit und beschäftigen sich nicht mit der Website. Dies lässt sich mit Web-Analysewerkzeugen leicht messen. Es ist jedoch wichtig zu verstehen, wie der gemessene Wert für das Unternehmen ein positives Ergebnis bringen kann. **(3 Punkte)**



**FRAGE 71 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**



Nennen und erklären Sie kurz die vier Hauptschritte der Clickstream-Analyse.

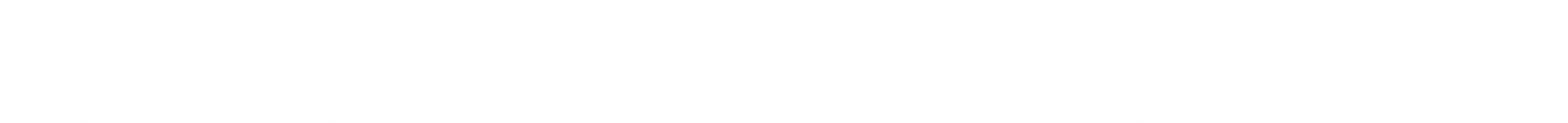
1. Verstehen: Es gibt mehrere Gründe, warum Clickstream-Analysen durchgeführt werden können. Zunächst einmal muss ein Unternehmen seine Bedürfnisse und das Ergebnis verstehen, das es sich von der Analyse von Clickstream-Daten erhofft, anstatt die Datenverkehrskanäle zu überprüfen oder den Inhalt zu verbessern.
2. Sammeln: Die Kenntnis des Ziels erleichtert den zweiten Schritt, da nur die für die Unternehmensziele relevanten Daten gesammelt werden müssen. Eines der besten Analysewerkzeuge für Clickstream-Daten ist Google Analytics. Dieses Werkzeug bietet eine Vielzahl kostenloser Optionen wie die Anzahl der Besucher:innen, ihre Interaktion mit der Website usw. Für jede Sitzung werden die relevanten Web-Metriken angezeigt und können leicht verglichen werden.
3. Analyse: Im Analyseschritt müssen die im zweiten Schritt gesammelten Informationen verglichen und gegenübergestellt werden. Dabei ist es wichtig, Muster oder Ausnahmen in der Art und Weise zu finden, wie verschiedene Nutzende mit einer Website interagieren. Die große Menge an Informationen, die bei einer Web-Analyse gesammelt werden kann, sollte ebenfalls genutzt werden.
4. Verbessern: Nach der Analyse der gesammelten Daten und der festgestellten Muster können Schlussfolgerungen gezogen werden, um die Benutzererfahrung auf der Website weiter zu verbessern und die Bindung der Nutzenden zu verstärken.

# (2 Punkte pro Schritt)



**FRAGE 72 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



1. Erklären Sie den Begriff „Segmentierung“.
2. Erklären Sie, warum „Segmentierung“ eingesetzt wird.
3. Nennen Sie zwei verschiedene Arten der Segmentierung, die auf einer Website durchgeführt werden können.
   1. Verschiedene Arten von Nutzenden können aus unterschiedlichen Gründen auf eine Website zugreifen. Dies erfordert eine gute Segmentierungsstrategie, um jedem Benutzertyp eine wirkungsvolle und nützliche Erfahrung zu bieten. Die Segmentierung hilft dabei, sich auf bestimmte Segmente zu konzentrieren und Ideen zu liefern, die praktisch umgesetzt werden können. **(3 Punkte)**
   2. Berichte und grafische Darstellungen der Umsätze aus diesen verschiedenen Quellen stellen die beste Möglichkeit dar, um dem Management die Auswirkungen der Web-Analyse und die zu betrachtenden Schritte für die Zukunft aufzuzeigen. **(3 Punkte)**
   3. Zu den Segmentarten gehörten unter anderem:
      1. Untergruppen von Nutzenden
      2. Untergruppen von Sitzungen
      3. Untergruppen von Treffern

# (2 Punkte pro Antwort, maximal 4 Punkte)



**FRAGE 73 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



Geben Sie drei mögliche Verwendungszwecke für Clickstream-Daten an.

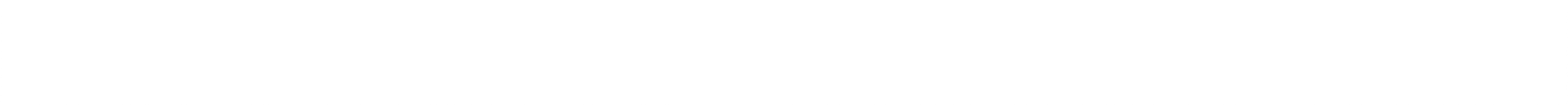
1. Kundenerfahrung verstehen
2. Benutzende segmentieren
3. Künftige Verbesserungen vorhersagen
4. Unterscheiden der Nutzenden, je nachdem über welche Kanäle sie die Website erreicht haben
5. Identifizieren der Kundenbindung der Nutzenden

# (2 Punkte für jede richtige Option, maximal 6 Punkte)



**FRAGE 74 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**



Nennen und erläutern Sie kurz zwei Methoden, mit denen die Wirksamkeit einer SEO-Kampagne bewertet werden kann.

1. WebMaster Tools: Die gängigen Suchmaschinen (Google, Yahoo usw.) bieten den Dienst Webmaster Tools an, der Statistiken über den Fortschritt der SEO-Bemühungen liefert. **(2 Punkte)** Zur Nutzung der Webmaster Tools müssen Sie sich bei der Suchmaschine anmelden und sich authentifizieren, indem Sie eine HTML-Datei in den Stammordnern der eigenen Website ablegen. Da jede Suchmaschine die Seiten auf unterschiedliche Weise kategorisiert, liefert die Verwendung der Webmaster Tools wichtige Informationen über den Zustand einer Website **(3 Punkte)**
2. Direkter Datenverkehr: Zur Überprüfung, ob eine SEO-Kampagne funktioniert oder nicht, ist zu untersuchen, ob die Anzahl der Seiten, auf die der direkte Datenverkehr zugreift, seit Beginn der SEO-Bemühungen gestiegen ist. **(3 Punkte)**



**FRAGE 75 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



Erklären und veranschaulichen Sie, wie eine nachverfolgbare URL erstellt werden kann.

Nachverfolgbare URLs zeigen, welche E-Mails Kund:innen auf eine Website bringen und welche E-Mails nicht effizient sind. Dazu müssen am Ende einer URL sogenannte UTM-Parameter (Urchin Tracking Module) stehen. UTM-Parameter sind Tags, die Informationen über die Person speichern, die auf den Link geklickt und eine Website aufgerufen hat. **(3 Punkte)**

Laut Delane (o. D.) gehören die folgenden UTM-Parameter zu den wichtigsten zur Erstellung einer nachverfolgbaren URL für weitere Analysen in Google Analytics:

UTM\_source: Dies beschreibt die Quelle oder den Ursprung des Verweises (Referrer, z. B. einen Newsletter). **(2 Punkte)**

UTM\_medium: Dies definiert die Marketingaktivität (täglich, wöchentlich, monatlich) **(2 Punkte)**

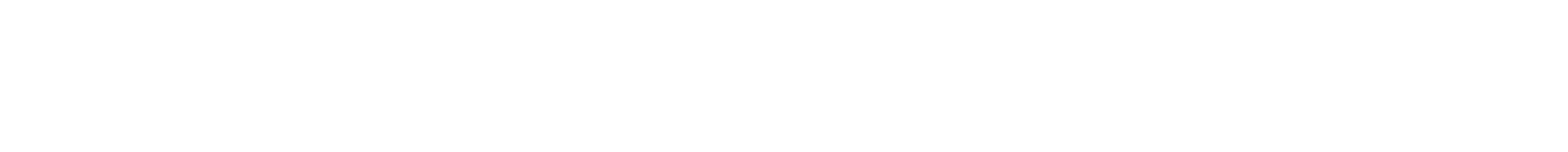
UTM\_name: Details oder Name der Kampagne. **(2 Punkte)**

Ein Beispiel für eine generierte URL kann folgendes sein: https://www.example.com?utm\_source=newsletter[&utm\_medium=weekly&utm\_campaign=autumn\_sales](http://www.example.com/?utm_source=newsletter&utm_medium=weekly&utm_campaign=autumn_sales) **(1 Punkt)**



**FRAGE 76 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**



Ein Unternehmen hat insgesamt 1.000 E-Mails mit einer nachverfolgbaren URL versandt. 100 dieser E-Mails waren nicht zustellbar und wurden zurückgeschickt.

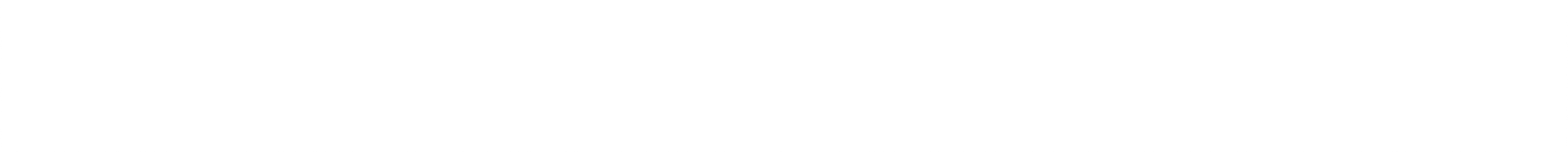
1. Erklären Sie kurz, was die Erfolgsquote (delivery rate) einer E-Mail-Kampagne ist.
2. Berechnen Sie die Erfolgsquote der E-Mail-Kampagne dieser Organisation.

Die Erfolgsquote (delivery rate) gibt Aufschluss darüber, ob eine E-Mail erfolgreich Nutzende auf die Website gelockt hat. **(3 Punkte)** Delivery\_rate = [(Anzahl der gesendeten E-Mails – Anzahl der Rückläufer)/(Anzahl der gesendeten E-Mails)] \* 100 % **(3 Punkte)** Delivery\_rate = [(1000-100)/100] \* 100 % = 90 % **(2 Punkte)**



**FRAGE 77 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



Ein Unternehmen hat insgesamt 2.000 E-Mails mit einer nachverfolgbaren URL versandt. 500 dieser E-Mails wurden zurückgeschickt, und nur 1.000 E-Mails wurden geöffnet.

1. Erklären Sie kurz, was die Öffnungsquote (open rate) einer E-Mail-Kampagne ist.
2. Berechnen Sie die Öffnungsquote der E-Mail-Kampagne dieser Organisation.
   1. Die Öffnungsquote ist die Gesamtzahl der geöffneten E-Mails **(2 Punkte)**, obwohl Bilder und Skripte in der Regel aus Sorge vor Viren blockiert werden **(2 Punkte)**.
   2. Open\_rate = [(Anzahl der geöffneten E-Mails)/(Anzahl der gesendeten E-Mails)] \* 100 % **(3 Punkte)**

Open\_rate = [1000/2000] \* 100 % = 50 % **(3 Punkte)**



**FRAGE 78 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



Ein Unternehmen hat insgesamt 3.000 E-Mails mit einer nachverfolgbaren URL versandt. 600 der E-Mails wurden zurückgeschickt und nur 1.000 E-Mails wurden geöffnet. Von den 700 Nutzenden, die auf die URL geklickt haben, waren nur 400 auf der Website aktiv, während die anderen die Website ohne weitere Aktivität wieder verließen.

1. Erklären Sie kurz, was die Klickrate einer E-Mail-Kampagne ist.
2. Berechnen Sie die Klickrate der E-Mail-Kampagne dieser Organisation.
3. Die Klickrate ist unentbehrlich, um die Effizienz einer an verschiedene Nutzende gesendeten E-Mail zu verstehen **(2 Punkte)**. Sie misst, wie viele Nutzende auf den URL-Link geklickt haben **(2 Punkte)**.
4. Click\_rate = [(Anzahl der Klicks)/(Anzahl der geöffneten E-Mails)] \* 100 % **(3 Punkte)**

Click\_rate= [700/1000] \* 100 % = 70 % **(3 Punkte)**



**FRAGE 79 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



Ein Unternehmen hat insgesamt 3000 E-Mails mit einer nachverfolgbaren URL versandt. 600 der E-Mails wurden zurückgeschickt und nur 1000 E-Mails wurden geöffnet. Von den 700 Nutzenden, die auf die URL geklickt haben, waren nur 400 auf der Website aktiv, während die anderen die Website ohne weitere Aktivität wieder verließen.

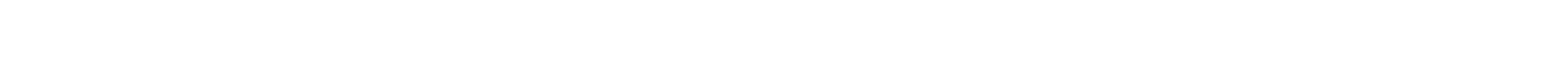
1. Erklären Sie kurz, was die Absprungrate (bounce rate) einer E-Mail-Kampagne ist.
2. Berechnen Sie die Absprungrate der E-Mail-Kampagne dieser Organisation.
3. Die Absprungrate zeigt, welche Seiten verbessert werden müssen. Sie zeigt, wie viele Besuche auf der Website zu keiner Nutzerbindung geführt haben. **(2 Punkte**) Eine E-Mail-Kampagne mit einer Reaktionsquote von 100 % aber einer Absprungrate von 99 % ist überhaupt nicht effektiv, weil sich die Nutzenden in diesem Fall nicht richtig mit der Website beschäftigen. **(2 Punkte)**
4. Bounce\_rate = [(Anzahl der Besuche mit einem einzigen Seitenaufruf)/(Anzahl der Besuche)] \* 100 % **(3 Punkte)**

Bounce\_rate = [300/700] \* 100 % = 42,8 % **(3 Punkte)**



**FRAGE 80 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



Nennen und erklären Sie kurz die drei grundlegenden, in Empfehlungssystemen verwendeten Beziehungen bzw. Interaktionen.

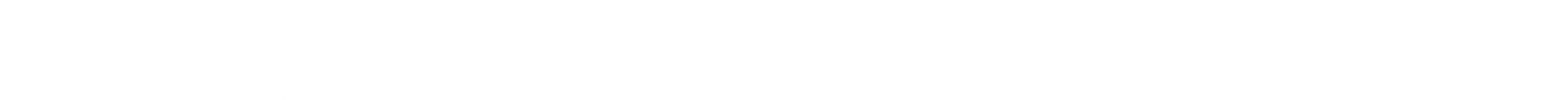
1. Nutzender-Nutzender: Diese Systeme geben Empfehlungen, die auf gemeinsamen Freunden, ähnlichem Alter usw. basieren.
2. Nutzender-Produkt: Diese Systeme geben Empfehlungen, die auf den Interessen oder dem Beruf der Nutzenden basieren.
3. Produkt-Produkt: Diese Systeme geben Empfehlungen, die auf ähnlicher Musik, Büchern, Videos usw. basieren.

# (jeweils 2 Punkte)



**FRAGE 81 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F1/Lektion 03**



1. Erklären Sie den inhaltsbasierten Ansatz für Empfehlungssysteme.
2. Nennen Sie drei Vorteile und drei Nachteile des inhaltsbasierten Ansatzes.
   1. Ein inhaltsbasiertes Empfehlungssystem findet eine Übereinstimmung zwischen einem Artikel und dem Profil eines Nutzenden, um so die beste Empfehlung zu ermitteln. **(2 Punkte)** Mit inhaltsbasierten Methoden wird versucht, auf der Grundlage von Attributen eines Artikels oder Merkmalen eines Nutzenden ein Modell zu erstellen, das die Nutzender-Produkt-Integration erklärt. Wir können zum Beispiel versuchen, das Verhalten von Jugendlichen zu modellieren, die bestimmte Musikgenres bevorzugen, und von Erwachsenen, die normalerweise andere bevorzugen. **(2 Punkte)**
   2. Zu den wichtigsten Vorteilen inhaltsbasierter Empfehlungssysteme gehören die folgenden:
      1. Die Empfehlungen basieren auf den bisherigen Aktivitäten der Nutzenden.
      2. Es ist nicht notwendig, auf die Daten anderer Nutzender zugreifen zu können.
      3. Das Hinzufügen neuer Nutzender und Produkte ist ganz einfach.

# (1 Punkt pro Vorteil, maximal 3 Punkte)

Zu den wichtigsten Nachteilen inhaltsbasierter Empfehlungssysteme gehören die folgenden:

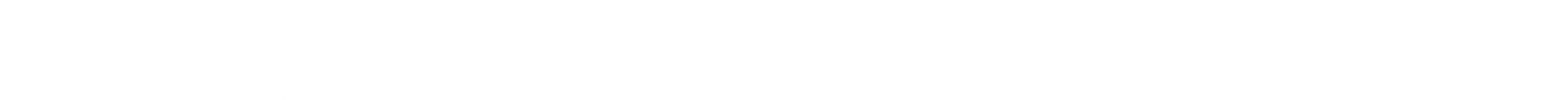
1. Nicht alle Inhalte können durch Stichwörter beschrieben werden (Videos, Bilder).
2. Es besteht die Gefahr einer Überspezialisierung.
3. Auf der Grundlage der Bewertungszahl (Score) für Artikel und Interessen gilt: Je niedriger die Bewertungszahl, desto schlechter die Empfehlung.
4. Es gibt keine bekannte Historie für neue Nutzende im System.

# (1 Punkt pro Nachteil, maximal 3 Punkte)



**FRAGE 82 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**



1. Erklären Sie den Ansatz der kollaborativen Filterung für Empfehlungssysteme.
2. Was bedeutet der Begriff „Kaltstart“ in diesem Zusammenhang?
3. Empfehlungssysteme mit kollaborativem Filter erstellen Empfehlungen, indem sie Ähnlichkeiten zwischen den Präferenzen der Nutzenden einer Plattform prüfen. **(2 Punkte)**

Bei diesem Ansatz basieren die empfohlenen Artikel auf den Meinungen anderer Nutzender mit ähnlichen Interessen. Die Interaktionen zwischen Nutzenden und Produkten werden in einer Nutzender-Produkt-Interaktionsmatrix gespeichert. Die Algorithmen im Zusammenhang mit kollaborativem Filtern werden in zwei Gruppen unterteilt: Speicherbasiert (funktionieren mit Vergangenheitsdaten ohne Modell) und modellbasiert (erklären die Interaktion zwischen Nutzenden und Produkt anhand eines Modells). **(3 Punkte)**

1. Ein „Kaltstart“ ist ein bekanntes Problem bei diesem Ansatz und bezieht sich auf die Interaktion zwischen neuen Nutzenden oder neuen Produkten bzw. das Fehlen ausreichender Daten für gute Empfehlungen. Da die Anzahl der Interaktionen zu Beginn gering ist, ist es schwierig, zu diesem Zeitpunkt passende Vorschläge zu machen **(3 Punkte)**.



**FRAGE 83 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F1/Lektion 03**



Nennen und erklären Sie kurz drei in Empfehlungssystemen genutzte Ansätze.

* 1. Wissensbasierte Ansätze: Diese Art von Empfehlungssystemen basiert auf explizitem Wissen, das durch die Kombination von Produkten, den Präferenzen eines Nutzenden und Empfehlungskriterien gesammelt wird.
  2. Hybride Ansätze: Solche Systeme basieren auf einer Kombination aus verschiedenen Ansätzen, nämlich kollaborative Filterung und inhaltsbasierte Filterung.
  3. Graph-basierte Ansätze: Empfehlungssysteme, bei denen Daten in Form eines Graphen dargestellt werden, wobei die Knoten Nutzende oder Artikel sein können und die Kanten die Interaktionen zwischen Nutzenden oder Artikeln darstellen.
  4. Kontextabhängige Empfehlungssysteme: Diese Systeme passen sich an verschiedene Kontexte an, in denen sich Nutzende befinden (Standort, Personen in der Nähe usw.).
  5. Inhaltsbasierte Ansätze: Ein inhaltsbasiertes Empfehlungssystem findet eine Übereinstimmung zwischen einem Artikel und dem Profil eines Nutzenden, um die beste Empfehlung zu ermitteln.
  6. Ansätze für das kollaborative Filtern: Empfehlungssysteme mit kollaborativem Filter erstellen Empfehlungen, indem sie Ähnlichkeiten zwischen den Präferenzen der Nutzenden einer Plattform prüfen.

# (2 Punkte für jede Option ähnlich denen der obigen Liste, maximal 6 Punkte)



**FRAGE 84 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F1/Lektion 03**



Betrachten Sie die Daten in der nachstehenden Tabelle, wobei 1 = gefällt, 0 = gefällt nicht bedeutet, und ein Leerzeichen, dass der Nutzende das Lied noch nicht angehört hat:

1. Welches Lied sollte Ardit gemäß dem kollaborativen Filteransatz (Artikel-Artikel) als nächstes empfohlen werden?

Erläutern Sie Ihre Überlegungen.

1. Welches Lied sollte Anne gemäß dem kollaborativen Filteransatz (Nutzender-Nutzender) als nächstes empfohlen werden? Erläutern Sie Ihre Überlegungen.
   * 1. **Artikel-Artikel:** Anne und James mögen die Lieder „Everything“ und „When you are Gone“. Da diese beiden Nutzenden diese beiden Lieder mögen, können wir daraus schließen, dass Ardit, der „Everything“ mag, auch „When you are Gone“ mögen könnte. Dieses Lied wird ihm also empfohlen. **(2 Punkte für die Begründung, 2 Punkte für die richtige Antwort)**
     2. **Nutzender-Nutzender** Anne und James mögen beide „Everything“ und „When you are Gone“, mögen aber „Million Reasons“ nicht. Da diese beiden Nutzenden ähnliche Interessen zu haben scheinen, können wir zu dem Schluss kommen, dass Anne das Lied „When we were Young“ mögen würde, weil James es auch mag **(2 Punkte für die Begründung, 2 Punkte für die richtige Antwort)**.



**FRAGE 85 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



Nennen und erklären Sie kurz die wichtigste Python-Bibliothek, die in jedem der vier SMA-Schritte verwendet wird.

Die wichtigsten Python-Bibliotheken, die in den einzelnen Schritten von SMA verwendet werden, sind:

NumPy (Numerical Python): Dies ermöglicht die Verarbeitung von Datenstrukturen in einer Array-ähnlichen Form.

SciKit-learn: Dies ist eine Bibliothek für maschinelles Lernen in Python, die verschiedene Klassifizierungsalgorithmen unterstützt (Support Vector Machines, k-means, usw.).

Pandas: Pandas ist eine Python-Bibliothek für die Analyse und Manipulation von Daten.

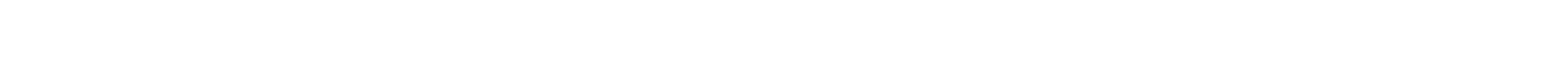
NLTK: Dies ist die wichtigste Python-Bibliothek, die zur Verarbeitung von Textinformationen genutzt wird.

# (1 Punkt pro Name und 1 Punkt pro Erklärung, maximal 8 Punkte)



**FRAGE 86 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**



Nennen Sie drei Methoden zur Darstellung der Endergebnisse in SMA.

1. Grafiken
2. Wortwolken
3. Diagramme
4. Plots
5. Infografiken
6. Textinhalt

# (2 Punkte für jede richtige Antwort, maximal 6 Punkte)



**FRAGE 87 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



Nennen und erläutern Sie kurz vier der wichtigsten Attribute, die zu einem Tweet gehören.

1. created\_at: Die UTC-Zeit der Tweet-Erstellung.
2. id: Eindeutige Kennung eines Tweets.
3. text UTF-Text des Beitrags.
4. in\_reply\_to\_status\_id: Wenn der Tweet eine Antwort ist, enthält dieses Attribut die ID des ursprünglichen Tweets.
5. in\_reply\_to\_user\_id\_str: Wenn der Tweet eine Antwort ist, enthält dieses Attribut die Benutzer-ID.
6. user: Dieses Feld enthält Informationen zu den Attributen des Nutzenden (ID, Standort, Anzahl der Freunde, Beschreibung usw.).
7. Coordinates: Geografischer Ort, an dem der Tweet gepostet wurde.
8. Retweet\_count Die Anzahl der Retweets.

# (1 Punkt pro Name und 1 Punkt pro Erklärung, maximal 8 Punkte)



**FRAGE 88 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**



Nennen und erklären Sie die beiden wichtigsten APIs für den Zugriff auf die Daten von Twitter.

1. REST-API: Dies steht für Representational State Transfer, ein Begriff, der im Jahr 2000 von Roy Fielding definiert wurde. Die REST-API basiert auf dem REST-Architekturstil und ermöglicht die Interaktion mit RESTful Web Services. **(3 Punkte)**
2. Streaming-API: Diese API bietet eine völlig andere Möglichkeit, auf Twitter-Daten zuzugreifen. Beim Herstellen einer Verbindung wird ein Datenstrom geöffnet. Wenn neue Tweets generiert werden, gelangen sie in die Verbindung und werden von der Anwendung gelesen. Diese Art von API hat eine begrenzte Kapazität, abhängig von der Menge an Informationen, die extrahiert werden sollen. **(3 Punkte)**



**FRAGE 89 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**

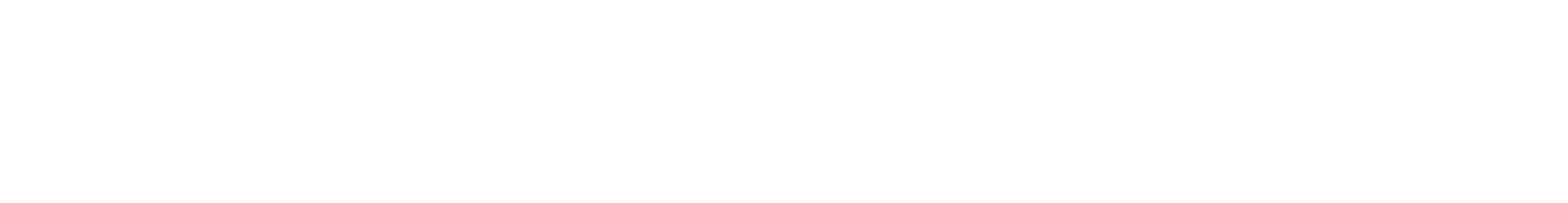


1. Nennen Sie die beiden Marken (Token), die bei Twitter zu einem Nutzenden bzw. zu einer Anwendung gehören.
2. Nennen Sie die Python-Bibliothek, welche die Interaktion mit der Twitter-API ermöglicht.
3. Nennen Sie zwei Funktionen, die zum Erstellen einer authentifizierten Verbindung dienen können, und die jeweiligen Parameter, die an diese Funktionen übergeben werden sollten.
   1. Die Token für die Anwendung sind API\_KEY und API\_SECRET, und die Token für das Konto des Entwicklers sind ACCESS\_TOKEN und ACCESS\_TOKEN\_SECRET. **(2 Punkte)**
   2. Die Python-Bibliothek, die für die Interaktion mit der Twitter-API importiert werden sollte, ist tweepy. **(2 Punkte)**
   3. Die folgenden beiden Funktionen können zum Erstellen einer authentifizierten Verbindung genutzt werden:
      1. OAuthHandler. **(2 Punkte)** Die Parameter, die an diese Funktion übergeben werden sollten, sind: API\_KEY **(0,5 Punkte)** und API\_SECRET **(0,5 Punkte)**.
      2. set\_access\_token. **(2 Punkte)** Die Parameter, die an diese Funktion übergeben werden sollten, sind: ACCESS\_TOKEN **(0,5 Punkte)** und ACCESS\_TOKEN\_SECRET **(0,5 Punkte)**.



**FRAGE 90 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**



a. Erklären Sie, was der folgende Code bewirkt:

with open('tweets.csv' % topic, 'w') as f:

writer.writerows(data)

b. Erläutern Sie kurz jede der vier im obigen Code verwendeten Funktionen.

writer.writerow(["Id","Created\_at","Retweet\_count","Text"])

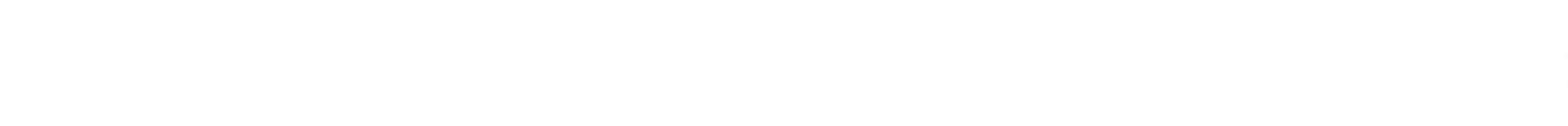
writer = csv.writer(f)

1. Der Code definiert einen Codeblock in Python, der die Attribute (id, created\_at, text) in eine tweets.csv-Datei schreibt. **(2 Punkte)**
2. 1. open('tweets.csv' % topic, 'w') erstellt und öffnet eine csv-Datei namens tweets.csv im Schreibmodus. **(2 Punkte)**
   1. csv.writer(f) gibt ein Writer-Objekt zurück, das die Benutzerdaten in abgegrenzte Zeichenketten auf dem Objekt f konvertiert. **(2 Punkte)**
   2. writerow(["Id", "Created\_at", "Retweet\_count", "Text"]) schreibt die Parameter id, created\_at, retweet\_count und text in das Dateiobjekt des Writers. **(2 Punkte)**
   3. writerows(data) schreibt jede Zeile des Datenrahmens (data) in die csv-Datei. **(2 Punkte)**



**FRAGE 91 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**



Erklären Sie die folgenden wesentlichen Twitter-Begriffe:

1. Nennen Sie die wichtigsten Elemente, die ein Tweet enthalten kann.
2. Nennen Sie die drei wichtigsten Merkmale, die zu jedem Tweet gehören.
3. Ein Tweet hat ein Limit von 280 Zeichen und kann Hashtags, Benutzer-Erwähnungen, URLs usw. enthalten. Auch wenn ein Tweet eigentlich nur eine kurze Zeichenkette ist, gehören auch zwei Metadatenelemente dazu: Entitäten und Orte. **(3 Punkte)**

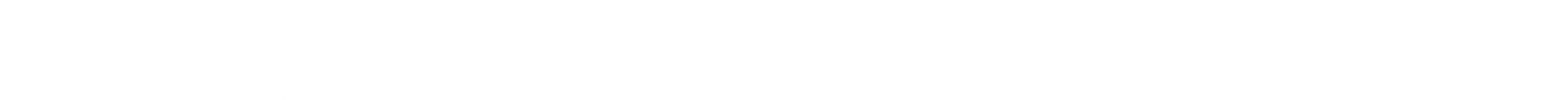
Zu den Entitäten gehören Hashtags, URLs und andere Medien, die mit einem bestimmten Tweet verknüpft sein können. Der Ort zeigt die geografische Lokalisierung an, wo ein Tweet erzeugt wurde.

1. Zu den Hauptmerkmalen, die in jedem Tweet zu finden sind, gehören die folgenden: der **Autor**, die **Textnachricht**, eine **ID**; **benutzerbezogene Daten**, die **Anzahl der Retweets** und das **Erstellungsdatum**. Neben diesen Hauptmerkmalen **kann ein Tweet über 150 Attribute enthalten**, darunter Koordinaten, Favoriten, Ort und retweet\_count. **(1 Punkt pro Stichwort, maximal 3 Punkte)**



**FRAGE 92 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**



1. Erläutern Sie kurz, was das Oauth-Protokoll ist.
2. Nennen und erklären Sie kurz einen Vorteil von Oauth.
3. Alle Verbindungen erfordern ein Authentifizierungsverfahren, und im Fall von Twitter ist dies Open Authorization (OAuth). OAuth ist ein offener Autorisierungsrahmen für REST-APIs. Er bietet Anwendungen die Möglichkeit, begrenzten Zugriff auf die Daten eines Nutzenden zu erhalten, ohne das Passwort des Nutzenden eingeben zu müssen. **(3 Punkte)**
4. Bei dieser Art der Authentifizierung wird das Passwort des Nutzenden niemals mit Dritten ausgetauscht, wodurch das Vertrauen der Nutzenden garantiert wird. **(3 Punkte)**



**FRAGE 93 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**



Nennen und erklären Sie kurz die drei Arten von Streaming-APIs.

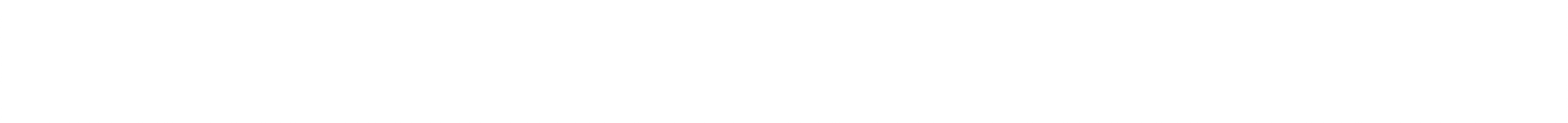
* 1. Öffentliche Datenströme, also Datenströme (Streams), die öffentliche Tweets enthalten.
  2. Benutzer-Datenströme, also Datenströme zu den Tweets eines bestimmten Nutzenden.
  3. Site-Datenströme, also Datenströme, die sich auf mehr als einen Nutzenden beziehen.

# (1 Punkt pro API und 1 Punkt pro Erklärung)



**FRAGE 94 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**



1. Erklären Sie die Ratenbegrenzung der Twitter-API und ihren Zweck genau.
2. Erläutern Sie, wie die Überschreitung der Ratenbegrenzung vermieden werden kann.
3. Erklären Sie, wie eine Verzögerung von 20 Sekunden zwischen zwei verschiedenen API-Anfragen erzeugt werden kann.
   * 1. Twitter hat eine Ratenbegrenzung für die Anzahl der Anfragen, die eine bestimmte Anwendung innerhalb eines bestimmten Zeitrahmens an die API stellen kann. Diese Grenzen wurden festgelegt, um eine zuverlässige API bereitzustellen, da Tausende von Entwickler:innen Anfragen an die Twitter-API stellen. **(3 Punkte)**

Die Ratenbegrenzungen hängen von der Art der verwendeten Authentifizierungsmethode ab und werden auf der Grundlage der angegebenen Zugriffsmarke (Access Token) festgelegt. Wenn die API-Ratenbegrenzungen erreicht werden, wird eine Fehlermeldung angezeigt, und der reguläre Zugang kann bei mehrfachen Anfragen gesperrt werden. Diese Funktion soll potenziellen Missbrauch verhindern. **(3 Punkte)**

* + 1. Ein Python-Modul namens time() kann verwendet werden, um eine Verzögerung zwischen zwei Anfragen festzulegen und so ein mögliches Überschreiten der zulässigen Ratenbegrenzung zu vermeiden. **(2 Punkte)**
    2. Mit der time\_sleep()-Funktion können wir eine Verzögerung von 20 Sekunden zwischen zwei API-Anfragen erzeugen. Die Funktion wird wie folgt aufgerufen: time\_sleep(20). **(2 Punkte)**



**FRAGE 95 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**



Beantworten Sie die nachstehenden Fragen zum folgenden Code:

import tweepy

data=[[tweet.id\_str,tweet.created\_at,tweet.retweet\_count, tweet.text.encode("utf-8")] for tweet in tweets]

Der Befehl tweepy.Cursor(API.search, q = topic, lang='en').items(20) besteht aus drei Elementen.

1. Beschreiben Sie den Zweck jeder dieser Komponenten.
2. Benennen Sie die Attribute, die im Datenrahmen gespeichert werden.

tweets = tweepy.Cursor(API.search, q = topic, lang='en').items(20)

topic="#covid"

API = tweepy.API(a)

a.set\_access\_token(ACCESS\_TOKEN, ACCESS\_TOKEN\_SECRET)

a = tweepy.OAuthHandler(API\_KEY,API\_SECRET)

ACCESS\_TOKEN\_SECRET = ''

ACCESS\_TOKEN = ''

API\_SECRET = ''

API\_KEY = ''

1. Der Befehl tweepy.Cursor(API.search, q = topic, lang='en').items(20) besteht aus den folgenden Elementen.
   1. API.search und q=topic → Suche nach Tweets zum Thema #covid **(3 Punkte)**
   2. lang='en' → Nur Tweets in englischer Sprache sammeln **(2 Punkte)**
   3. items(20) → Nur die letzten 20 Tweets zu diesem Thema sammeln **(2 Punkte)**
2. Die im Datenrahmen gespeicherten Attribute des Tweets sind **id\_str, created\_at** und **retweet\_count (1 Punkt pro Stichwort)**



**FRAGE 96 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**



Betrachten Sie den folgenden Code:

with open('user.jsonl', 'w') as file:

file.write(json.dumps(status.\_json)+"\n")

1. Erklären Sie jeden der vier Parameter der Funktion Cursor.
2. Wofür wird der „for“-Zyklus in diesem Codeblock verwendet?
3. Wie viele Tweets werden insgesamt mit diesem Codeblock gesammelt?

for status in page:

count=100).pages(2):

screen\_name='user1',

for page in Cursor(API.user\_timeline,

1. 1. API.user\_timeline liefert die letzten Statusmeldungen eines Nutzenden **(2 Punkte)**.
   1. screen\_name='user1' definiert den Namen des Nutzenden, dessen Status gesammelt wird. **(1 Punkt)**
   2. count=100 bestimmt die Anzahl der zu sammelnden Statusmeldungen. **(2 Punkte)**
   3. pages(2) gibt die Seite der zu sammelnden Ergebnisse an. **(2 Punkte)**
2. Der for-Zyklus wird verwendet, um die gesammelten Tweets in einer JSON-Zeilendatei zu speichern. **(2 Punkte)**
3. Die Gesamtzahl der mit diesem Codeblock gesammelten Tweets beträgt 200. **(1 Punkt)**



**FRAGE 97 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 04**



Gegeben sei der folgende Code:

with open('user\_timeline.jsonl', 'r') as file:

]

hashtags.update(hashtags\_tweet)

1. Wie wird die Counter-Funktion verwendet?
2. Erklären Sie die Bedeutung des gegebenen Codeblocks und das Endergebnis, das er liefert.

hashtags\_tweet = get\_hashtags(tweet)[tag['text'].lower()for tag in h

h = entities.get('hashtags', [])

entities = tweet.get('entities',{})

tweet = json.loads(i)

for i in file:

hashtags = Counter()

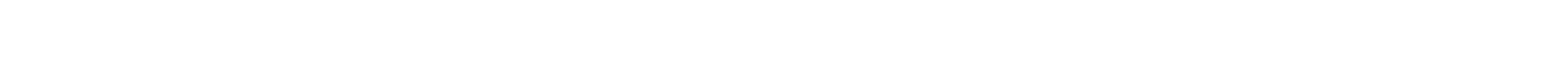
1. Die Funktion collections.Counter berechnet die Anzahl der Zeichenketten (Strings), wobei Strings (in diesem Fall Hashtags) die Schlüssel in einem Wörterbuch sind und der mit ihnen verknüpfte Wert ihre jeweilige Häufigkeit ist. **(2 Punkte)**
2. In dem vorgegebenen Code lesen wir die .json-Datei (user\_timeline.jsonl) Zeile für Zeile **(1 Punkt)** Der Tweet selbst wird in einem Wörterbuch gespeichert **(1 Punkt)**, und alle Entitäten dieses Tweets werden in einer Variablen namens „entities“ gespeichert. **(1 Punkt)** Da wir besonders an den erwähnten Hashtags interessiert sind, können wir von diesen Entitäten aus mit der Funktion get() auf die Hashtags der Entität zugreifen. **(2 Punkte)** Die Hashtags werden alle mit der Funktion lower() in Kleinbuchstaben umgewandelt. **(1 Punkt)**

Die Funktion hashtags.update() addiert die Anzahl der Elemente in hashtags\_tweet zur Anzahl der Elemente im Objekt Counter (hashtag). **(2 Punkte)**



**FRAGE 98 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



Erklären Sie, was die Facebook Graph API ist und wie sie funktioniert.

Die Facebook Graph API ist eine HTTP-basierte API zur Bearbeitung von Daten, zum Posten von Fotos oder Stories usw. Der Name dieser API basiert darauf, dass Facebook weitgehend als sozialer Graph bekannt ist.

Facebook selbst hat eine Graphdatenstruktur, die aus Knoten (Personen oder Nutzenden) und den Verbindungen zwischen diesen Knoten besteht. **(2 Punkte)**

Die Zugriffsmarken bieten einen sicheren Weg, um ohne Passwort auf die Daten der Nutzenden zuzugreifen. **(1 Punkt)** Knoten in der Graph-API sind eindeutige Objekte mit einer eindeutigen ID. **(1 Punkt)**

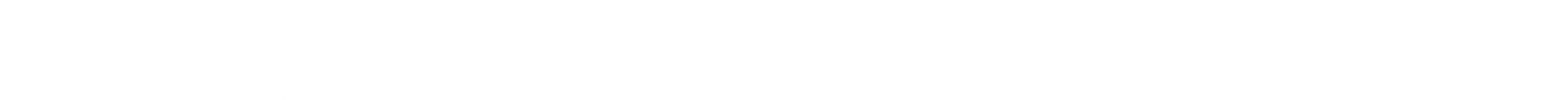
Einige der Knoten im Facebook-Graph sind: Benutzerknoten (jede einzelne Person, die im Netzwerk angemeldet ist), Gruppen und Beiträge. Jeder Knoten ist mit Feldern wie ID, Name und Beschreibung verknüpft. **(2 Punkte)**

Ein besonderer Knoten ist der Knoten /me, der die aktuelle Person bestimmt, die die API-Aufrufe tätigt. **(2 Punkte)**



**FRAGE 99 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 04**

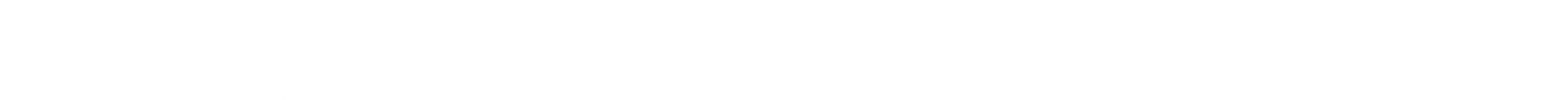


1. Was war die Facebook Graph Search?
2. Welche Methode hat sie ersetzt?
3. Die Facebook Graph Search nutzte natürliche Sprachverarbeitung für Suchanfragen **(2 Punkte)** Wenn zum Beispiel der folgende Satz „Freunde, die in Italien leben und Motorrad fahren“ in die Suchleiste eingegeben wird, gibt es relevante Informationen über die Interessen der eigenen Kontakte. **(2 Punkte)**
4. Dieses Werkzeug wurde durch die Stichwortsuche ersetzt. **(2 Punkte)**



**FRAGE 100 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



1. Erklären Sie das von Facebook entwickelte Open-Graph-Protokoll.
2. Geben Sie ein Beispiel dafür, wie es funktioniert.
3. Das Open-Graph-Protokoll (OGP) ist ein Protokoll, das im April 2010 von Facebook erstellt wurde, um Webseiten in den Facebook Social Graph zu integrieren. Wie der Name schon sagt, bietet dieses Protokoll eine Möglichkeit, den sozialen Graphen durch Integration vieler neuer Objekte (Webseiten) zu öffnen. **(3 Punkte)**
4. Wir stellen uns eine Webseite vor, auf der die Möglichkeit „Gefällt mir auf Facebook“ (engl.: Like on Facebook) angeboten wird. Facebook integriert RDFa-Metadaten auf einer solchen Seite, so dass sie in den Aktivitätsdatenstrom von Facebook aufgenommen wird. **(3 Punkte)** RDFa (Resource Description Framework in Attributes) ist eine HTML5-Erweiterung mit der Möglichkeit, HTML/XML-Dokumenten attributbezogene Erweiterungen hinzuzufügen. Dies erleichtert die Suche und erlaubt eine bessere Anbindung. **(2 Punkte)**



**FRAGE 101 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



Nennen und erläutern Sie kurz die vier Hauptkomponenten des Facebook-Werkzeugs Graph API Explorer.

* 1. Facebook App Dropdown: Ein Dropdown-Menü zur Auswahl der zu testenden Anwendung.
  2. Nutzender oder Seite: Ein Menü zur Auswahl zwischen Anwendungs-, Benutzer- oder Seiten-Tokens für die gewünschte Anwendung.
  3. Berechtigung hinzufügen: Ein Menü, mit dem der aktuellen Anwendung Berechtigungen hinzugefügt werden können. Die Standardberechtigung ist public\_profile. Es gibt jedoch drei weitere Gruppen von Berechtigungen: Berechtigungen für Benutzerdaten (z. B. email und photos\_), Seiten für Ereignisgruppen (z. B. page\_read\_engagement, page\_events) und andere (z. B. Instagram\_basic und publish\_video).
  4. Zugriffsmarke (Access Token): Nach Auswahl der richtigen Optionen und Berechtigungen kann die Zugriffsmarke (Access Token) erstellt und in diesem Feld angezeigt werden.
  5. Das Feld Abfrage: Hier können die auszuführenden Abfragen definiert werden.
  6. Code/Debug-Informationen abrufen: So kann der Code automatisch nach einer Abfrage im Abfragefeld erstellt werden.

# (1 Punkt pro Komponente und 1 Punkt pro Erklärung, maximal 8 Punkte)



**FRAGE 102 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 04**



Erklären Sie, was der folgende Code bewirkt.

import json

FACEBOOK\_TEMP\_TOKEN=" "

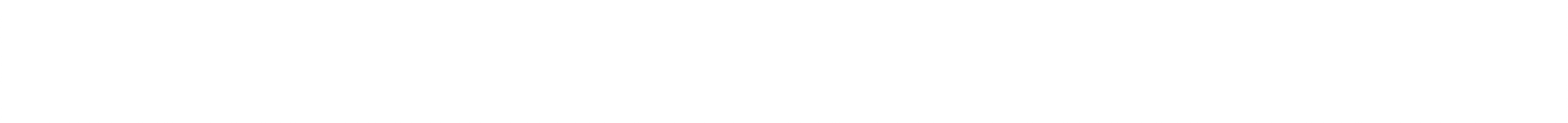
graph=facebook.GraphAPI(FACEBOOK\_TEMP\_TOKEN) profile=graph.get\_object('me',fields='id,name,email') print(json.dumps(profile,indent=4))

Der Code erforscht und druckt die Entwicklerinformationen **(2 Punkte)** im JSON-Format **(3 Punkte)**, insbesondere die Felder id **(1 Punkt)**, name **(1 Punkt)** und email **(1 Punkt)**.



**FRAGE 103 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**



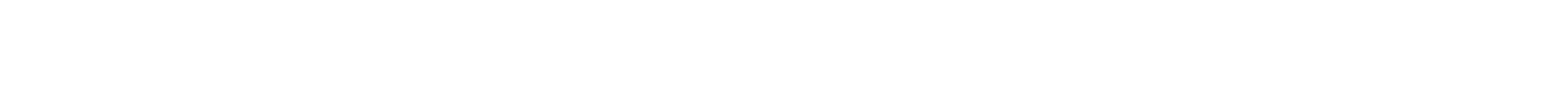
1. Beschreiben Sie ein Beispiel aus der Praxis, bei dem A/B-Tests eingesetzt wurden.
2. Erklären Sie, warum in dieser Situation A/B-Tests eingesetzt wurden und was das Ergebnis der A/B-Tests war.

Die Präsidentschaftskampagne von Obama ist ein gutes Beispiel dafür, denn diese Kampagne stützte sich stark auf Datenanalysen und Experimente. Im November 2007 bestand einer der größten Engpässe auf der Kampagnen-Website darin, die Besucher:innen der Website davon zu überzeugen, sich für E-Mails anzumelden. Der Schritt zur E-Mail-Anmeldung wurde optimiert, um mehr E-Mail-Abonnements zu erzielen. Die optimierte Seite generierte im Jahr 2008 zusätzliche 2,8 Millionen Anmeldungen mit E-Mail-Adressen. Weitere Experimente bei der Obama-Kampagne haben gezeigt, dass Testmethoden dabei helfen, objektive Entscheidungen zu treffen. Obwohl die Mehrheit des Teams der Meinung war, dass ein Video auf der Landeseite der Kampagne am besten funktioniere, zeigten Tests, dass das Foto der Familie Obama die meisten Klicks brachte.

# (2 Punkte für die Erwähnung eines realen Beispiels für A/B-Tests)

# (3 Punkte für die Erwähnung dessen, warum A/B-Tests verwendet wurden)

**(3 Punkte für die Darstellung des Endergebnisses des A/B-Tests im Beispiel)**



1. Erklären Sie, wie A/B-Tests das Design und Layout einer Seite beeinflussen.
2. Welche Seite einer Website sollte zuerst einem A/B-Test unterzogen werden? Erklären Sie, warum das so ist.



**FRAGE 104 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**

* + 1. Das Design und das Layout einer Seite sollten die Besucher:innen einer Website nicht verwirren, sondern ihnen vielmehr klare Informationen liefern. **(2 Punkte)** Der Inhalt sollte einfach und übersichtlich sein, und die beste Variante der Website kann durch A/B-Tests ermittelt werden. **(3 Punkte**)
    2. Da die Startseite die erste besuchte Seite ist, muss sie perfekt optimiert werden und sollte die erste Seite sein, die A/B-Tests unterzogen wird. **(3 Punkte)**



**FRAGE 105 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**



Erklären Sie, warum MVT komplexer ist als A/B-Tests.

MVT erfordert **umfangreiche Anstrengungen** der gesamten Organisation. Es handelt sich um ein **komplexes** Verfahren, das von **erfahrenen Personen** durchgeführt werden muss. A/B-Tests sind auch für Laien einfach zu handhaben. MVT erfordert jedoch **Expert:innen**, wenn gute Ergebnisse erzielt werden sollen.

# (2 Punkte pro Stichwort)



**FRAGE 106 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



Erklären Sie, was MVT ist und wie es funktioniert.

MVT steht für engl. Multivariate Testing, also Multivariates Testen. Es handelt sich dabei um eine Form von A/B-Tests, bei denen mehrere Elemente gleichzeitig verändert werden und dann die Leistung jeder Kombination der veränderten Komponenten bewertet wird. **(3 Punkte)**

Zu jeder der möglichen Kombinationen werden Daten gesammelt, um festzustellen, welche Kombination von Elementen besser funktioniert. Diese Testmethode wird vor allem dann eingesetzt, wenn man radikale Änderungen an einer Website mit mehr als einer geänderten Komponente vornehmen möchte. **(3 Punkte)**



**FRAGE 107 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**



Nennen und erklären Sie zwei der gängigsten multivariaten Testmethoden.

Zwei der am häufigsten verwendeten multivariaten Tests sind vollständige Versuchspläne (full factorial testing) und Teilfaktorpläne (fractional factorial testing). **(2 Punkte)**

Beim vollständigen Versuchsplan wird der eingehende Website-Datenverkehr gleichmäßig auf die verschiedenen Testkombinationen verteilt. **(3 Punkte)**

Beim Teilfaktorplan wird mit einer kleineren Anzahl von Kombinationen experimentiert und dann versucht zu extrapolieren, was mit den anderen Kombinationen passieren wird. **(3 Punkte)**



**FRAGE 108 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



Vergleichen Sie die beiden Testmethoden: vollständige Versuchspläne (full factorial testing) und Teilfaktorpläne (fractional factorial testing).

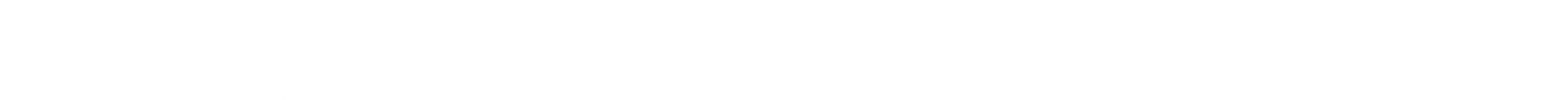
Beide Methoden haben Vor- und Nachteile. Vollständige Versuchspläne (full factorial testing) liefern **detaillierte Informationen über jede mögliche Kombination**, allerdings **auf Kosten des Zeitaufwands**. Der Zeitaufwand für das Testen einer großen Anzahl von Kombinationen kann beträchtlich sein. Andererseits gehen bei Teilfaktorplänen (fractional factorial testing) **manche Daten** über die Kombinationen **verloren**, wobei die **Ergebnisse** aber auch **schneller** vorliegen. Außerdem erfordern sie auch **weniger Datenverkehr**, und sind daher eine gute Option für das Testen von Websites mit geringem Datenverkehr.

# (2 Punkte pro Stichwort)



**FRAGE 109 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



1. Definieren Sie kurz multivariate Tests.
2. Nennen und erläutern Sie kurz drei Hauptvorteile von multivariaten Tests.
3. Multivariate Tests sind eine einfache Methode, die lediglich einige JavaScript-Tags auf einer Website erfordern. **(1 Punkt)** Die durch MVT gesammelten Daten liefern sofort ein besseres Verständnis für die Vorlieben der Mehrheit der Nutzenden, was bei der Bestimmung der optimalen Version einer Website hilft. **(3 Punkte**)
4. 1. Die Möglichkeit, mehrere Komponenten einer Seite zu testen.

Angenommen wir beabsichtigen, Änderungen an zwei Komponenten einer Website vorzunehmen und möchten mit dem Testen beginnen. Bei A/B-Tests sind zwei aufeinanderfolgende Tests erforderlich. Zunächst wird die erste Komponente einem A/B-Test unterzogen, um die beste Version zu finden. Dann wird die zweite Komponente einem A/B-Test unterzogen, um die beste Variante zu finden.

* 1. Die Verwendung von multivariaten Tests anstelle von A/B-Tests ermöglicht es, jede mögliche Kombination zwischen den beiden Elementen zu testen und festzustellen, welche Kombination am besten funktioniert. Daher kann mit MVT die beste Kombination verschiedener Elemente auf einer Seite gefunden werden. So kann Zeit gespart werden, weil nicht mehrere A/B-Tests durchgeführt werden müssen.
  2. Erkennen der Elemente, die keinen positiven Beitrag auf einer Seite leisten.

MVT findet zwar die beste Kombination von Elementen auf einer Seite, erkennt aber auch die Elemente, die nicht zur Konversionsrate der Website beitragen. So können z. B. ungeeignete Textinhalte, Bilder, Videos und ähnliches auf der Seite verschoben oder ganz entfernt werden.

* 1. Bestimmen der Struktur einer Seite.

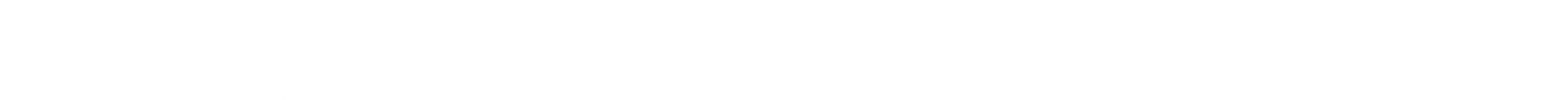
Durch das Testen verschiedener Arten von Elementen an unterschiedlichen Positionen kann MVT bei der Strukturierung der Seite helfen. Die ersten Elemente, die Besucher:innen sehen, bestimmen in der Regel ihre Bereitschaft, die Website weiter zu erkunden. Daher sollten die wichtigsten Komponenten zuerst angezeigt werden, um Aufmerksamkeit zu erregen. MVT-Tests stellen die richtige Methode dar, wenn Unsicherheiten über die Struktur der Start- oder Landeseite bestehen.

# (2 Punkte pro Vorteil, maximal 6 Punkte)



**FRAGE 110 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



1. Definieren Sie kurz multivariate Tests.
2. Nennen und erklären Sie drei Hauptnachteile von multivariaten Tests.
3. Multivariate Tests sind eine einfache Methode, die lediglich einige JavaScript-Tags auf einer Website erfordern. **(1 Punkt)** Die durch MVT gesammelten Daten liefern sofort ein besseres Verständnis für die Vorlieben der Mehrheit der Nutzenden, was bei der Bestimmung der optimalen Version einer Website hilft. **(3 Punkte**)
4. 1. Sie benötigen mehr Datenverkehr.

MVT sollte vor allem bei Websites mit hohem Datenverkehrsaufkommen eingesetzt werden. Nur dann ist es möglich, aussagekräftige Ergebnisse zu erzielen. Wenn weniger frequentierte Websites mit MVT getestet werden sollen, lässt sich als Kompromiss eine geringere Anzahl von Kombinationen testen.

* 1. MVT ist komplexer als A/B-Tests.

MVT erfordert umfangreiche Anstrengungen der gesamten Organisation. Es handelt sich um ein komplexes Verfahren, das von erfahrenen Personen durchgeführt werden muss. A/B-Tests sind auch für Laien einfach zu handhaben. MVT erfordert jedoch professionelles Testen durch Experten.

* 1. Zeitaufwand.

A/B-Tests lassen sich rasch einrichten und liefern schnelle Antworten. Da MVT komplexer ist, dauern die Einrichtung und der Betrieb hingegen länger. Jedes Unternehmen muss selbst entscheiden, ob der Zeitaufwand für die Durchführung multivariater Tests akzeptabel ist und ob sie lohnende Ergebnisse liefern werden.

* 1. Hohes Risiko des Scheiterns.

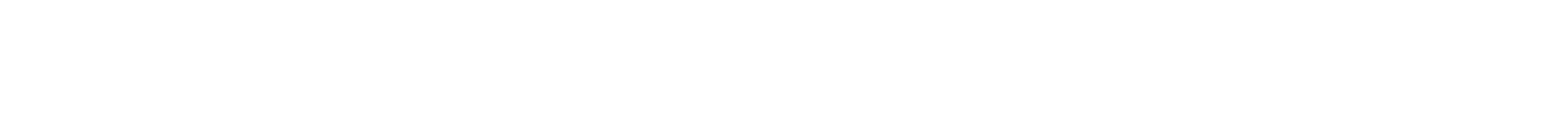
A/B-Tests können zum Erfolg oder Misserfolg führen, aber in jedem Fall können wir etwas Neues über Website-Komponenten lernen. Bei MVT ist es überhaupt nicht rentabel, mehrere Kombinationen ergebnislos zu testen. Der Zeit- und Ressourcenaufwand für die Durchführung der Tests ist im Vergleich zu A/B-Tests zu hoch.

(2 Punkte pro Nachteil)



**FRAGE 111 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**



Nennen Sie ein Beispiel, bei dem zur Prüfung von Elementen zwei aufeinanderfolgende A/B-Tests durchgeführt werden müssen, während beim multivariaten Testen alle möglichen Komponenten gleichzeitig geprüft werden können.

Angenommen wir beabsichtigen, Änderungen an einer Website vorzunehmen und möchten mit dem Testen beginnen. Ziel ist es, die Anzeige eines Bildes oder eines Videos auf der Startseite zu vergleichen und zu testen, welche der beiden Schaltflächen „Jetzt ausprobieren“ oder „Jetzt kaufen“ besser funktioniert. Bei A/B-Tests sind zwei aufeinanderfolgende Tests erforderlich. Zuerst wird das Bild gegen das Video getestet, und das Bild gewinnt. Danach werden die Schaltflächen getestet, und die Schaltfläche „Jetzt kaufen“ gewinnt.

Die Nutzung multivariater Tests anstelle von A/B-Tests ermöglicht es, jede mögliche Kombination zu testen und festzustellen, welche Kombination am besten funktioniert. Daher kann mit MVT die beste Kombination verschiedener Elemente auf einer Seite gefunden werden. So kann Zeit gespart werden, weil nicht mehrere A/B-Tests durchgeführt werden müssen.

**(3 Punkte für die Erklärung, was bei einem A/B-Test passieren würde)**

# (3 Punkte für die Erklärung, was bei einem multivariaten Test passieren würde)

**(2 Punkte für die Begründung, warum multivariate Tests in diesem Beispiel die bessere Auswahl darstellen)**



**FRAGE 112 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



1. Welche Schwäche hat der Epsilon-Greedy-Algorithmus?
2. Nennen Sie zwei andere Methoden, die effizienter sind als der Epsilon-Greedy-Algorithmus.
3. In Fällen, in denen schon sicher ist, dass eine Komponente besser ist als eine andere, bedeutet die Verwendung des Epsilon-Greedy-Algorithmus, dass immer noch eine gewisse Zeit mit der Untersuchung der schlechteren Komponente verbracht wird. Dies wird als **Übererkundung (Over-exploring)** bezeichnet. **(2 Punkte)**
4. Andere Multi-armed-Bandit-Algorithmen, wie z. B. der **Softmax-Algorithmus (oder Boltzmann-Exploration)** und der **UCB-Algorithmus (Upper Confidence Bound Algorithm)** finden die beste Variante einer Seite effizienter als der Epsilon-Greedy-Algorithmus.

# (2 Punkte pro Stichwort, maximal 4 Punkte)



**FRAGE 113 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



Erklären Sie Multi-armed-Bandit-Tests.

Der Multi-armed-Bandit-Test ist eine neue Form des Experimentierens, die es ermöglicht, den **Datenverkehr dynamisch den Versionen mit besserer Leistung zuzuweisen** und nicht den Versionen mit langsamerem Datenverkehr.

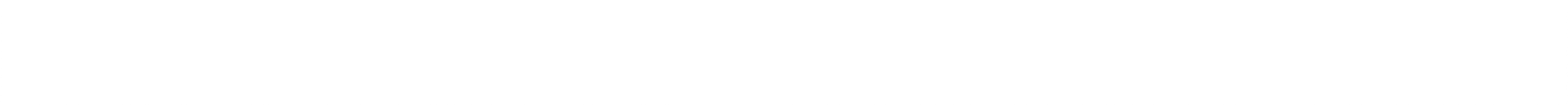
Der Multi-armed-Bandit-Test funktioniert als ein **Kompromiss zwischen Erkunden und Ausbeuten**. Ein Multi-armed-Bandit-Algorithmus erforscht und weist den Großteil des Datenverkehrs dynamisch der Version zu, die eine höhere Gewinnwahrscheinlichkeit zeigt (Ausbeutung).

# (2 Punkte pro Stichwort)



**FRAGE 114 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



Erklären Sie die beiden Begriffe „Ausbeutung“ und „Erkundung“ und ihre Verbindung zu Multi-armed-Bandit-Algorithmen.

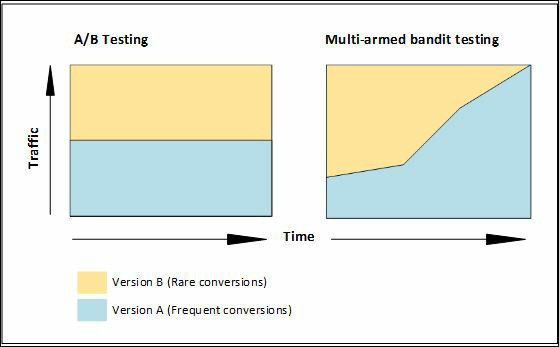
Die Begriffe Exploration (Erkundung) und Ausbeutung (Exploitation) sind die beiden wesentlichen Konzepte von Multi-armed-Bandit-Algorithmen. Um diese Konzepte besser zu verstehen, denken wir an einen Wissenschaftler, der ständig versucht zu experimentieren (Exploration), und an eine Unternehmerin, die ihre Gewinne maximieren möchte (Exploitation bzw. Ausbeutung). **(3 Punkte)**

Bei klassischen A/B-Tests wird erkundet (Exploration), um statistisch signifikante Ergebnisse zu ermitteln. Bei A/B-Tests wird die Konversionsrate für jede Variante berechnet, und die Version mit der besten Leistung wird zum Sieger erklärt. Der Multi-armed-Bandit-Test hingegen funktioniert als ein Kompromiss zwischen Erkunden und Ausbeuten. **(3 Punkte)**



**FRAGE 115 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



1. Erklären Sie, was in den Fällen von A/B-Tests bzw. Multi-armed-Bandit-Tests passiert.
2. Was sind die beiden Hauptunterschiede zwischen diesen beiden Testmethoden?
3. Die Abbildung zeigt den Unterschied zwischen A/B-Tests und Multi-armed-Bandit-Algorithmen, insbesondere was im Laufe der Zeit passiert und wie die beste Version ausgewählt wird. **(3 Punkte)**
4. Wir sehen, dass bei A/B-Tests der Datenverkehr gleichmäßig auf die beiden verschiedenen Versionen aufgeteilt wird **(3 Punkte)**, während bei Multi-armed-Bandit-Tests die Höhe des Datenverkehrs sich verändert, je nachdem, welche Version besser abschneidet. **(3 Punkte)**

Bei A/B-Tests wird die Konversionsrate für jede Variante berechnet, und die Version mit der höchsten Leistung wird zum Sieger erklärt. Der Multi-armed-Bandit-Test hingegen funktioniert als ein Kompromiss zwischen Erkunden und Ausbeuten. **(1 Punkt)**



**FRAGE 116 VON 315**

**DLBDSEDA01\_Offen\_mittel\_F2/Lektion 05**

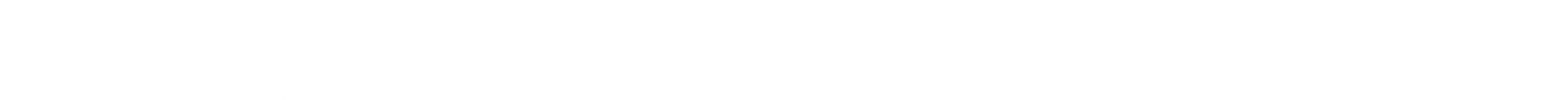


1. In welcher Art von Szenario wären A/B-Tests besser als Multi-armed-Bandit-Tests?
2. In welcher Art von Szenario wäre ein Multi-armed-Bandit-Test besser als ein A/B-Test?
3. A/B-Tests sind die perfekte Wahl, wenn das Ziel eines Unternehmens darin besteht, die statistischen Ergebnisse der einzelnen Varianten einer Website zu erfahren, was bei der Entwicklung anderer Produkte nützlich ist. **(3 Punkte)**
4. Multi-armed-Bandit-Algorithmen eignen sich am besten, wenn ein Unternehmen keine Zeit hat, statistische Erkenntnisse zu sammeln **(3 Punkte)**, sondern vielmehr eine Kennzahl in kurzer Zeit optimieren möchte. **(2 Punkte**.



**FRAGE 117 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



1. Was ist ein „gieriger“ (greedy) Algorithmus?
2. Was ist ein Epsilon-Greedy-Algorithmus?
3. Ein gieriger Algorithmus ist ein Algorithmus, der in einem bestimmten Moment die beste Aktion auswählt, ohne an die Konsequenzen zu denken. **(3 Punkte)**
4. Der Epsilon-Greedy-Algorithmus ist mit einem Greedy-Algorithmus verwandt. Es wechselt zufällig zwischen den beiden Möglichkeiten Erkunden und Ausbeuten. Die Logik dahinter ist ähnlich wie das Werfen einer Münze vor jeder Entscheidung zwischen zwei bestehenden Möglichkeiten. **(3 Punkte)**



**FRAGE 118 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



Erklären Sie die drei Schritte, die beim Epsilon-Greedy-Algorithmus befolgt werden.

Der Epsilon-Greedy-Algorithmus durchläuft folgende Schritte:

Ein neuer Nutzender besucht die Website. Der Epsilon-Greedy-Algorithmus wählt nach dem Zufallsprinzip zwischen den beiden Optionen Erkundung und Ausbeutung aus. Der Algorithmus trifft die Entscheidung im Wesentlichen auf der Grundlage eines Münzwurfs. Angenommen die Münze fällt mit einer Wahrscheinlichkeit von Epsilon auf Zahl. **(3 Punkte)**

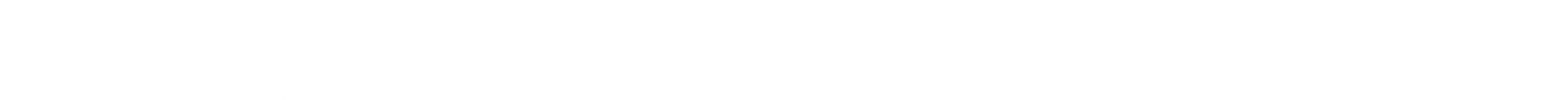
Wenn die Münze auf Kopf fällt, wird jedoch die Ausbeutung gewählt. Der Algorithmus prüft die zuvor aufgezeichneten Informationen, um herauszufinden, welches Bild bis zu diesem Zeitpunkt die höchste Konversionsrate erzielt hat. **(3 Punkte)** Dieses Bild zeigt er dann dem aktuellen Besuchenden der Website. **(1 Punkt)**

Wenn die Münze jedoch Zahl zeigt, wird erkundet. Der Algorithmus experimentiert dann mit den Bildern und muss daher zufällig zwischen den beiden möglichen Bildern wählen. In diesem Fall beträgt die Wahrscheinlichkeit, dass eines der beiden Bilder angezeigt wird, 50 %. Wenn eine Münze Zahl zeigt, wird den Besucher:innen auf der Homepage Bild A angezeigt, und bei Kopf Bild B. **(3 Punkte)**



**FRAGE 119 VON 315**

**DLBDSEDA01\_Offen\_leicht\_F2/Lektion 05**



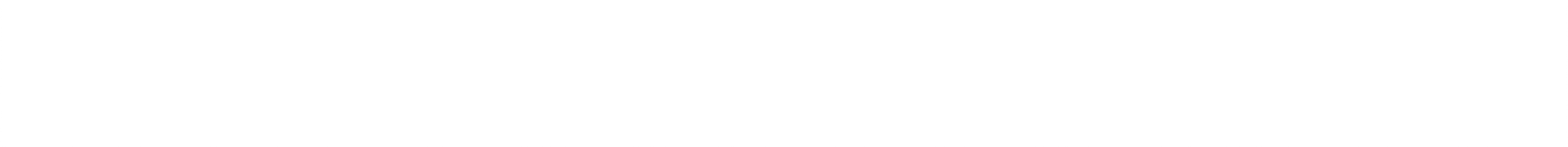
1. Erklären Sie den Begriff „einarmiger Bandit“.
2. Erklären Sie den Begriff „Belohnung“.
3. Der Name „Multi-armed-Bandit“ („mehrarmiger Bandit“) basiert auf den Spielautomaten in einem Casino, die als **einarmige Banditen** bezeichnet werden, weil Spielende mit dem Ziehen eines einzigen Hebels Geld gewinnen oder verlieren können. Die Methode des mehrarmigen Banditen (Multi-armed-Bandit) ist vergleichbar mit der Wahl eines Spielautomaten, der eine höhere Gewinnchance bietet, und von dieser Metapher leitet der Algorithmus seinen Namen ab. Ein Arm ist der Hebel, der bei einer Aktion gezogen wird.
4. In Multi-armed-Bandit-Algorithmen bezeichnet der Begriff „Belohnung“ das erfolgreiche Erreichen eines Ziels, z. B. ein Klick, eine Neukundenanmeldung usw. Eine Belohnung muss etwas sein, das mathematisch berechnet werden kann. Dadurch wird sichergestellt, dass die Belohnung aus verschiedenen Versionen berechnet und verglichen werden kann, um die am besten passende Version zu finden.

# (3 Punkte für jede Definition)



**FRAGE 120 VON 315**

**DLBDSEDA01\_Offen\_schwer\_F2/Lektion 05**



1. Nennen und erläutern Sie drei Werte, die in die Definition eines Algorithmus der Epsilon-Greedy-Klasse aufgenommen werden können.
2. Erklären Sie, wozu die Methoden define\_arm und update\_information in einer Implementierung des Epsilon-Greedy-Algorithmus dienen.
3. Epsilon: Dieser Begriff bezieht sich auf die Häufigkeit, mit der eine der Optionen gewählt wird (Erkunden/Ausbeuten). **(2 Punkte)**

Count (Zähler): In diesem Feld (Array) wird gespeichert, wie oft jede Version (jeder Arm) gewählt wurde. In einem Beispiel mit zwei Armen, bei dem der erste Arm nicht und der zweite Arm zweimal gewählt wurde, würde das Zählfeld als [0,2] erscheinen. **(2 Punkte)**

Values (Werte): Ein Feld (Array), in dem die Anzahl der mit einem bestimmten Arm verbundenen Belohnungen gespeichert wird. **(2 Punkte)**

1. 1. define\_arm. Jedes Mal, wenn wir den Erkundungsmodus aufrufen, wird die Methode define\_arm (ohne Argumente) aufgerufen, um den Verweis (Index) auf den gewählten Arm zurückzugeben. **(2 Punkte)**

2. update\_information. Der gewählte Arm gibt dem System eine Belohnung. Die Aktualisierungsfunktion (Update) benötigt drei Argumente: ein Objekt, den numerischen Index des zuvor gewählten Arms und die Belohnung für diesen Arm. Diese Funktion aktualisiert die vorherigen Informationen und nimmt die notwendigen Änderungen für die nächsten Aktionen vor. **(2 Punkte)**





**FRAGE 121 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Die SMART-Kriterien werden als Methode zur Bewertung der wesentlichen KPIs verwendet.

Wofür steht das „T“?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Time-bound (Termingebunden).* |
| Temporary (Vorübergehend) |
| Turnover (Umsatz) |
| Technological (Technologisch) |



**FRAGE 122 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Was sollte ein Unternehmen verstehen, bevor es eine KPI-Strategie entwickelt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Die historischen Leistungen der Organisation. |
| Die Marketingstrategie des Unternehmens |
| *Die wichtigsten Ziele, die das Unternehmen zu erreichen versucht.* |
| Die technischen Systeme, die das Unternehmen verwendet |







**FRAGE 123 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Die SMART-Kriterien werden als Methode zur Bewertung der wesentlichen KPIs verwendet.

Wofür steht das „S“?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Sales (Umsatz) |
| Standard |
| Savings (Ersparnisse) |
| *Specific (Spezifisch)* |



**FRAGE 124 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Welcher der folgenden ist ein wichtiger Leistungsindikator (Key Performance Indicator)?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Die gestrigen Umsätze |
| Nettogewinn vor Steuern |
| Kapitalrendite (ROI) |
| *Kundenreklamationen (Customer Complaints)* |







**FRAGE 125 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Die SMART-Kriterien werden als Methode zur Bewertung der wesentlichen KPIs verwendet. Wofür steht das „M“?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Mature (Ausgereift) |
| Monetary (Monetär) |
| *Measurable (Messbar)* |
| Marketable (Marktfähig) |



**FRAGE 126 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Welche der folgenden Perspektiven ist im Balanced Scoreboard einer Organisation enthalten?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Externe Prozesse |
| Arbeitgeber |
| *Lernen* |
| Unterrichten |







**FRAGE 127 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Wie häufig werden die als Ergebnisindikatoren eingestuften Maßnahmen in der Regel durchgeführt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Täglich |
| Stündlich |
| Wöchentlich |
| *Monatlich* |



**FRAGE 128 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Die SMART-Kriterien werden als Methode zur Bewertung der wesentlichen KPIs verwendet.

Wofür steht das „A“?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Attainable (Erreichbar)* |
| Advanced (Fortgeschritten) |
| Adjustable (Einstellbar) |
| Advertisable (Werbefähig) |







**FRAGE 129 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Die SMART-Kriterien werden als Methode zur Bewertung der wesentlichen KPIs verwendet.

Wofür steht das „R“?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Repairable (Reparierbar) |
| *Relevant* |
| Receivable (Forderung) |
| Reparation (Wiedergutmachung) |



**FRAGE 130 VON 315**

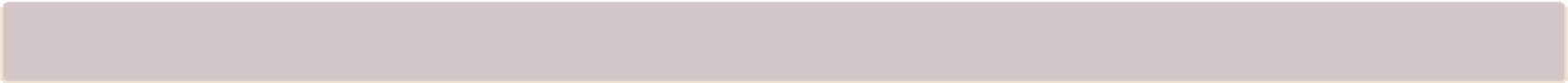
**DLBDSEDA01\_MC\_mittel/Lektion 01**



Welche der folgenden Eigenschaften beschreibt die Position des Chief Measurement Officer?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Ein neuer Mitarbeitender, der über KPIs informiert ist. |
| Eine Person, deren einzige Aufgabe darin besteht, die endgültigen KPI-Werte zu erklären. |
| *Eine Person, die den größten Teil ihrer Arbeitszeit der KPI-Implementierung widmet.* |
| Eine Person, die entscheidet, welche KPIs von der Organisation berücksichtigt werden müssen. |







**FRAGE 131 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Welcher der folgenden Indikatoren ist ein wichtiger Ergebnisindikator?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Kundenbindungsrate(Customer Retention Rate) |
| Promotorenüberhang (NPS = Net Promoter Score) |
| *Reingewinn (Net Profit)* |
| Kundenreklamationen (Customer Complaints) |



**FRAGE 132 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Was verbirgt sich hinter dem Begriff „KPI“?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Key planning indicators |
| Key predicting index |
| *Key performance indicator* |
| Key production indicators |







Welche der folgenden Eigenschaften gehört zu KPIs?

**Wählen Sie eine Antwort aus:**



Wofür steht der Begriff „NPS“?

**Wählen Sie eine Antwort aus:**



**FRAGE 133 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**

|  |
| --- |
| Finanzen |
| *Häufige Messung* |
| Schwierig zu verstehen |
| Gemessen auf jährlicher Basis |



**FRAGE 134 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**

|  |
| --- |
| Net performance score |
| *Net promoter score* |
| Net production score |
| Net performance standard |







**FRAGE 135 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Ein NPS kann folgendermaßen gemessen werden:

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Befragung der Mitarbeitenden |
| Analyse historischer Daten |
| Durchführung von Marketing-Kampagnen |
| *Durchführung von Umfragen* |



**FRAGE 136 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**



Welche der folgenden Aussagen über die Häufigkeit der Messung des CRR-Indikators für die Kundenbindung ist richtig?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Sie hängt nicht von der Art des Produkts oder der Dienstleistung ab. |
| Die Laufzeit des Vertrags hat keinen Einfluss auf die Häufigkeit der Messung der Kundenbindungsrate CRR. |
| Je länger der Einkaufszyklus ist, desto einfacher ist es, den Überblick über die CRR zu behalten. |
| *Sie hängt von der Art des Produkts oder der Dienstleistung ab.* |







**FRAGE 137 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**



Wie wird der Indikator für die Kundenfluktuationsrate gemessen?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Anzahl der wiederkehrenden Kund:innen/Gesamtzahl der Kund:innen |
| *Anzahl der verlorenen Kund:innen/Gesamtzahl der Kund:innen* |
| Anzahl der verlorenen Kund:innen/Anzahl der neuen Kund:innen |
| Anzahl der Neukund:innen/Gesamtzahl der Kund:innen |



**FRAGE 138 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Was bedeutet der Begriff „CRR“ im Zusammenhang mit KPIs?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Company‘s registered rate |
| Current return rate |
| *Customer retention rate* |
| Customer received reactions |







**FRAGE 139 VON 315**

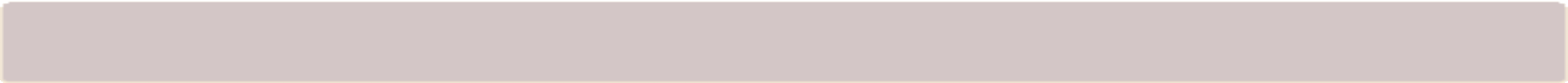
**DLBDSEDA01\_MC\_schwer/Lektion 01**



Was misst der Indikator PSV?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Den Unterschied zwischen der geschätzten Zeit für die Fertigstellung eines Projekts und der tatsächlich benötigten Zeit.* |
| Die Aufteilung zwischen der Zeit, die für den Start des Projekts benötigt wird, und der Zeit, die für seine Umsetzung benötigt wird. |
| Die für die Planung eines zukünftigen Projekts benötigte Zeit. |
| Die für die Umsetzung eines Projekts benötigte Zeit. |



**FRAGE 140 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Wofür steht der Begriff „CUR“ im Zusammenhang mit KPIs?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Capacity unload rate |
| Capacity utility received |
| *Capacity utilization rate* |
| Customer underlying reactions |







**FRAGE 141 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Wofür steht der Begriff „PSV“ im Zusammenhang mit KPIs?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Project schedule variance* |
| Partial service variance |
| Production service variance |
| Performance standard variance |



**FRAGE 142 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



Jedes Dokument in einem Korpus wird dargestellt als:

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Ein Vektor mit einer Länge, die der Gesamtzahl der Wörter im Dokument entspricht. |
| Ein Vektor mit einer Länge, die der Anzahl der Einzelwörter im Dokument entspricht. |
| *Ein Vektor mit einer Länge, die dem Vokabular des gesamten Korpus entspricht.* |
| Ein Vektor mit einer Länge, die der Anzahl der Doppelwörter im Dokument entspricht. |







**FRAGE 143 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



Wie viele Bi-Gramme können aus folgendem Satz erzeugt werden: „Sonntag war ein sonniger Tag.“

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| 5 |
| *4* |
| 32 |
| 3 |



**FRAGE 144 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Wofür steht der Begriff „DIFOT“ im Zusammenhang mit KPIs?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Deposit in full, on time |
| Develop in full, on time |
| Demand in full, on time |
| *Delivery in full, on time* |







Was ist der niedrigste zulässige DIFOT-Wert?

**Wählen Sie eine Antwort aus:**



**FRAGE 145 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Welche der folgenden Wortformen stellt das Wort „doubled“ nach dem Stemming dar?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| doubly |
| doubled |
| *doubl* |
| double |



**FRAGE 146 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**

|  |
| --- |
| 0,5 |
| *0,95* |
| 0,8 |
| 0,3 |







Was ist in der Indikatormessung „Prozessausfallzeit“ enthalten?

**Wählen Sie eine Antwort aus:**



**FRAGE 147 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



Was ist ein dünnbesetzter Vektor?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Ein Vektor, bei dem alle Werte gleich Null sind. |
| Ein Vektor, bei dem alle Werte ungleich Null sind. |
| *Ein Vektor mit vielen Nullen.* |
| Ein Vektor mit duplizierten Werten. |



**FRAGE 148 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**

|  |
| --- |
| Testzeit |
| *Wartung* |
| Mitarbeitende, die das Unternehmen verlassen |
| Zeit für die Projektplanung |







**FRAGE 149 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**



Wie wird der Grad der Maschinenausfallzeit berechnet?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| (Zeit zur Behebung von Fehlern/Produktionszeit) \* 100 |
| (Zeitaufwand für Wartung + Zeitaufwand für Fehlerbehebung/Produktivzeit) \* 100 |
| (Wartungszeit/Produktionszeit) \* 100 |
| *(Tatsächliche Produktionszeit/geplante Produktionszeit) \* 100* |



**FRAGE 150 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Wofür wird die Funktion w.isalpha() in Python verwendet?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Zur Entfernung der Stoppwörter aus dem Text |
| Zur Umwandlung des Texts in Kleinbuchstaben |
| *Zur Prüfung, ob eine Zeichenkette Text darstellt* |
| Zum Auffinden des Wortlemmas im Text |







Wofür steht das Kürzel Tf-idf?

**Wählen Sie eine Antwort aus:**



**FRAGE 151 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**



Die Höhe der Prozessausfallzeit sollte folgendermaßen gemeldet werden:

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Wenn ein vordefinierter Schwellenwert überschritten wird.* |
| Wöchentlich. |
| Stündlich. |
| Wenn eine Maschine oder ein Prozess länger als einen Tag nicht funktioniert. |



**FRAGE 152 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**

|  |
| --- |
| *Term frequency – inverse document frequency* |
| Topic finder – inverse document finder |
| Term function – inverse document function |
| Text filtering – inverse data filtering |







Wenn ein Wort zwölfmal in einem Dokument A mit insgesamt 200 Wörtern vorkommt, wie hoch ist dann der TF-Wert für dieses Wort in Dokument A?

**Wählen Sie eine Antwort aus:**



**FRAGE 153 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Welche der folgenden Methoden wird zur Modellierung von Geschäftsprozessen verwendet?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| ERP |
| *UML* |
| CMS |
| CRM |



**FRAGE 154 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**

|  |
| --- |
| *0,06* |
| 0,12 |
| 0,94 |
| 0,44 |







**FRAGE 155 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Was bedeutet der Buchstabe P in einem Petri-Netz?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Point |
| Parent |
| *Places* |
| Positions |



**FRAGE 156 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**



Das Wort „Globus“ kommt in fünf von 300 Dokumenten in einem bestimmten Korpus vor.

Wie hoch ist der IDF-Wert für dieses Wort?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| 60 |
| –1,77 |
| *1,77* |
| 0,01 |







**FRAGE 157 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Was bedeutet der Buchstabe T in einem Petri-Netz?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Time |
| *Transitions* |
| Terminal |
| Temporary |



**FRAGE 158 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**



Wie groß ist der euklidische Abstand der beiden Frequenzvektoren v1 = [1 0 1] und v2 = [0 0 1]?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| 2 |
| 0 |
| *1* |
| 0,5 |







**FRAGE 159 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Was bedeutet der Begriff I/O in einem Petri-Netz?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Input arc/output arc* |
| Input place/output place |
| Input time/output time |
| Input value/output value |



**FRAGE 160 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Wie kann die dünne Besetzung einer Vektor-Darstellung im NLP verringert, also dichter gemacht werden?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Durch die Berücksichtigung vonVokabularen mit n-Token-Sequenzen. |
| Durch das Ordnen der Wörter im Wortschatz, abhängig von der Struktur des Ausgangssatzes. |
| *Durch sorgfältige Auswahl der Wörter, die in den Wortschatz aufgenommen werden sollen.* |
| Durch Schreiben jedes Worts im Vokabular in Kleinbuchstaben. |







**FRAGE 161 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**



Welche der folgenden Web-Metriken dient als Grundlage für die Berechnung anderer Web-Metriken?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Ausstiegsrate (Exit Rate) |
| Auf der Website verbrachte Zeit |
| Absprungrate (Bounce Rate) |
| *Anzahl der Besuche* |



**FRAGE 162 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Was bedeutet der Buchstabe M in einem Petri-Netz?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Matrix |
| Mutual |
| Margin |
| *Marking* |







Welches der folgenden Werkzeuge gibt einen Überblick darüber, was auf einer bestimmten Website passiert?

**Wählen Sie eine Antwort aus:**



**FRAGE 163 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Welche Rolle spielt eine Kante in einem Petri-Netz?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Sie verbindet zwei Transitionen. |
| Sie verbindet zwei Orte. |
| *Sie verbindet Plätze (Stellen) mit Transitionen oder andersherum.* |
| Sie verbindet eine Eingabestelle mit einer Ausgabestelle |



**FRAGE 164 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**

|  |
| --- |
| ProM |
| Minit |
| Celonis |
| *Google Analytics* |







**FRAGE 165 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**



Was stellt eine Marke (Token) in einem Petri-Netz dar?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Den Zustand eines Netzes.* |
| Die Anzahl der Eingabestellen. |
| Die Anzahl der Ausgabestellen. |
| Die Anzahl der Schaltsequenzen. |



**FRAGE 166 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Die Metrik „Auf der Website verbrachte Zeit“ (Verweildauer) berechnet die Zeit, die ein Nutzender auf einer bestimmten Seite verbringt.

Für welche Art von Seite ist die Berechnung der Zeit, die Nutzende mit dem Navigieren verbracht haben, schwierig?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Für die Landeseite. |
| Für die zweite aufgerufene Seite. |
| *Für die zuletzt aufgerufene Seite.* |
| Für die erste aufgerufene Seite. |







Welche der folgenden Aussagen ist richtig?

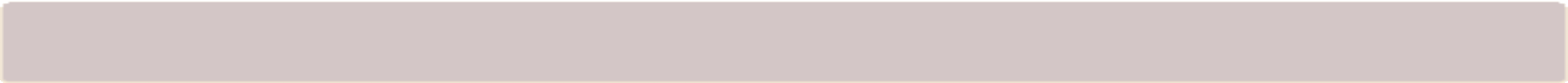
**Wählen Sie eine Antwort aus:**



**FRAGE 167 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**

|  |
| --- |
| *In der Konformitätsprüfungsphase wird das abgeleitete Prozessmodell mit einem idealen Prozessmodell verglichen*. |
| UML und BPMN sind typische Formate zum Speichern von Ereignisprotokollen. |
| In der Prozesserkennungsphase werden Modelle repariert, die nicht der Realität entsprechen. |
| Herkömmliche Data-Mining-Methoden können nicht zur Verbesserung von Geschäftsprozessen eingesetzt werden. |



**FRAGE 168 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welcher der folgenden Werte gilt in der Regel als eine ungefähre durchschnittliche Sitzungsdauer?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Drei Minuten.* |
| Dreißig Minuten. |
| Eine Stunde. |
| Eine Minute. |







**FRAGE 169 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Mit welcher Methode kann ein guter Durchschnittswert für die Sitzungsdauer ermittelt werden?

Antwortmöglichkeit 1: Mit einer benutzerfreundlichen Oberfläche. Antwortmöglichkeit 2: Mit ansprechenderen Bildern.

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Nur Antwortmöglichkeit 1 ist richtig. |
| *Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig.* |
| Weder die Antwortmöglichkeit 1 noch die Antwortmöglichkeit 2 ist richtig. |
| Nur Antwortmöglichkeit 2 ist richtig. |



**FRAGE 170 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**



Welcher der folgenden ist ein wichtiger Grund für die Umstellung auf Process-Mining?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Durch Process-Mining gibt es weniger Datenschutzprobleme. |
| Process-Mining benötigt weniger Hardware-Ressourcen. |
| Process-Mining benötigt weniger technisches Personal. |
| *Durch Process-Mining werden mögliche Blockaden aufgedeckt*. |







Welches der folgenden Werkzeuge wird von der akademischen Gemeinschaft häufig für Process-Mining genutzt?

**Wählen Sie eine Antwort aus:**



**FRAGE 171 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 01**

|  |
| --- |
| Xero |
| Slack |
| Google Analytics |
| *ProM* |



**FRAGE 172 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**



Was ist die Absprungrate?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Der Prozentsatz der Nutzenden, die eine Aktion auf der Website ausführen. |
| Der Prozentsatz der Nutzenden, welche die Website mehr als drei Minuten lang besuchen. |
| *Der Prozentsatz der Nutzenden, die eine Website wieder verlassen, ohne eine Aktion durchzuführen.* |
| Der Prozentsatz der Nutzenden, welche die Website weniger als drei Minuten lang besuchen. |







Welche der folgenden Aussagen ist richtig?

**Wählen Sie eine Antwort aus:**



**FRAGE 173 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Wie kann die Absprungrate berechnet werden?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Anzahl der neuen Besucher:innen/Gesamtzahl der Besucher:innen |
| Gesamtzahl der Besucher:innen/Gesamtzahl der Sitzungen |
| Anzahl der Einzelseitensitzungen/Gesamtzahl der Sitzungen |
| Gesamtzahl der Sitzungen/Gesamtzahl der Besucher:innen |



**FRAGE 174 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**

|  |
| --- |
| Nach dem Process-Mining gibt es in der Regel keine weiteren Verbesserungen in einer Organisation. |
| Process-Mining erhöht den Zeitaufwand für die Erstellung eines guten Prozessmodells. |
| *Process-Mining ermöglicht ein besseres Verständnis von Geschäftsprozessen.* |
| Herkömmliche Prozessanalyse-Werkzeuge schneiden besser ab als Process-Mining-Werkzeuge. |







**FRAGE 175 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 01**



Der Fachbereich des Process-Mining baut auf Folgendem auf:

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Ergebnis- und Leistungsindikatoren. |
| CRMs und Geschäftsverwaltungssysteme. |
| Statistik und Innovation. |
| *Prozessmodellierung und Data-Mining.* |



**FRAGE 176 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Die Metrik für die Kundenbindung lautet:

Antwortmöglichkeit 1: Die Kennzahl für die Kundenbindung ist eine quantitative Kennzahl. Antwortmöglichkeit 2: Die Kennzahl für die Kundenbindung ist eine qualitative Kennzahl.

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Nur Antwortmöglichkeit 1 ist richtig. |
| Nur Antwortmöglichkeit 2 ist richtig. |
| *Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig.* |
| Weder Antwortmöglichkeit 1 noch Antwortmöglichkeit 2 ist richtig. |







**FRAGE 177 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 01**



Welches der folgenden Formate ist ein typisches Format für die Speicherung von Ereignisprotokollen?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Excel-Rechenblätter |
| CSV |
| *XES* |
| UML |



**FRAGE 178 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**



Was zeigt an, dass eine SEO-Kampagne richtig funktioniert hat?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Ein Anstieg der Anzahl der Seiten, die durch direkten Datenverkehr aufgerufen werden.* |
| Ein Anstieg der Anzahl der Seiten, die durch organischen Datenverkehr aufgerufen werden. |
| Ein Anstieg der Anzahl der Seiten, die durch sozialen Datenverkehr aufgerufen werden. |
| Ein Anstieg der Anzahl der Seiten, die durch E-Mail-Datenverkehr aufgerufen werden. |







Wie wird die Erfolgsquote (delivery rate) berechnet, wenn eine Website über eine erstellte URL aufgerufen wird?

**Wählen Sie eine Antwort aus:**



**FRAGE 179 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**



Was passiert bei der Merkmalsextraktion in der natürlichen Sprachverarbeitung (NLP)?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Die Länge der einzelnen Wörter wird berechnet. |
| Die Länge der einzelnen Sätze wird berechnet. |
| *Die Wörter werden in eine Vektordarstellung umgewandelt.* |
| Die Wörter werden zu ähnlichen Themen gruppiert. |



**FRAGE 180 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**

|  |
| --- |
| Delivery\_rate = (Anzahl der eindeutigen Besucher:innen)/(Anzahl versendeter E-Mails) |
| *Delivery\_rate = (Anzahl der gesendeten E-Mails – Anzahl der nicht zustellbaren E-Mails)/(Anzahl der gesendeten E-Mails)* |
| Delivery\_rate = (Anzahl der gesendeten E-Mails)/(Anzahl der nicht zustellbaren E-Mails) |
| Delivery\_rate = (Anzahl der nicht zustellbaren E-Mails)/(Anzahl der gesendeten E-Mails) |







**FRAGE 181 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Wie wird die Öffnungsquote (open rate) berechnet, wenn eine Website über eine erstellte URL aufgerufen wird?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Open\_rate = (Anzahl der Besucher:innen)/(Anzahl der geöffneten E-Mails) |
| Open\_rate = (Anzahl der Besucher:innen)/(Anzahl der gesendeten E-Mails) |
| *Open\_rate = (Anzahl der geöffneten E-Mails)/(Anzahl der gesendeten E-Mails)* |
| Open\_rate = (Anzahl der gesendeten E-Mails)/(Anzahl der geöffneten E-Mails) |



**FRAGE 182 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



BoW und TF-IDF können verwendet werden bei:

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Themendefinition. |
| *Spam-Filterung.* |
| Marketing-Strategie. |
| Seitenoptimierung. |







Welche der folgenden Aussagen ist richtig?

**Wählen Sie eine Antwort aus:**



**FRAGE 183 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Was ist eine API?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Application programming interface* |
| Abstract programming interface |
| Authentication programming interface |
| Application page index |



**FRAGE 184 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**

|  |
| --- |
| Zu den Stoppwörtern gehört eine Liste aller Wörter mit weniger als vier Buchstaben. |
| *Zu den Stoppwörtern gehören Wörter, die in der Vorverarbeitungsphase herausgefiltert werden.* |
| Zu den Stoppwörtern gehört eine Liste von Interpunktionszeichen, die entfernt werden. |
| Zu den Stoppwörtern gehören alle Wörter, die nicht zu einem bestimmten Thema gehören. |







Welche der folgenden Aussagen ist richtig?

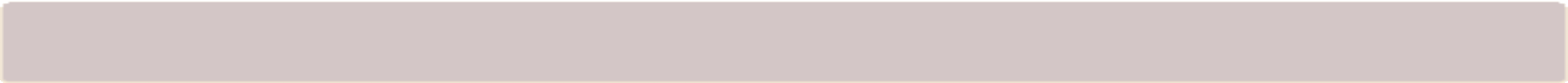
**Wählen Sie eine Antwort aus:**



**FRAGE 185 VON 315**

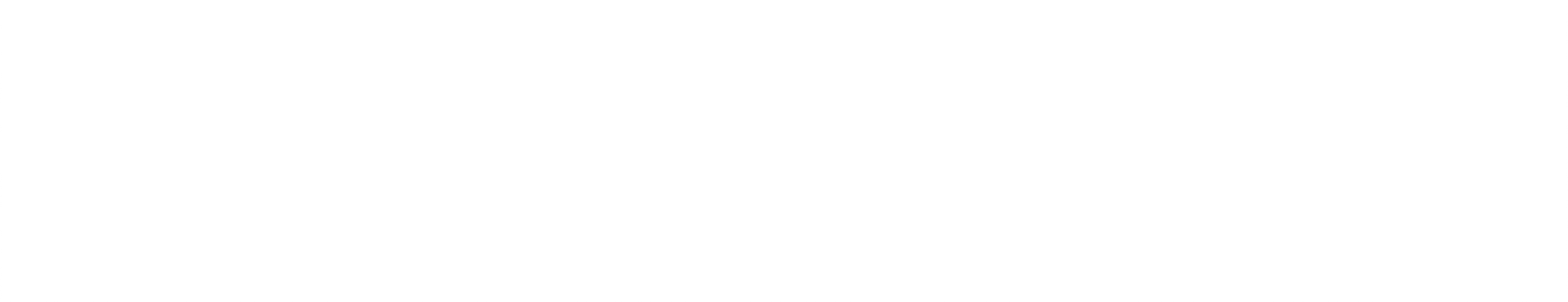
**DLBDSEDA01\_MC\_schwer/Lektion 02**

|  |
| --- |
| Die BoW-Technik berücksichtigt die Reihenfolge, in der die Wörter in einem Dokument gefunden werden. |
| Stemming ist eine Technik, bei der Flexionsendungen entfernt werden. |
| *Je näher die Punkte im semantischen Raum beieinander liegen, desto höher ist die Verbindung zwischen den Dokumenten.* |
| Im semantischen Raum wird ein Dokument als eine Sammlung von Wörtern betrachtet. |



**FRAGE 186 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Welche Arten von Daten umfasst der Begriff „Big Data“?

Antwortmöglichkeit 1: Strukturierte Daten. Antwortmöglichkeit 2: Unstrukturierte Daten.

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig.* |
| Weder Antwortmöglichkeit 1 noch Antwortmöglichkeit 2 ist richtig. |
| Nur Antwortmöglichkeit 2 ist richtig. |
| Nur Antwortmöglichkeit 1 ist richtig. |







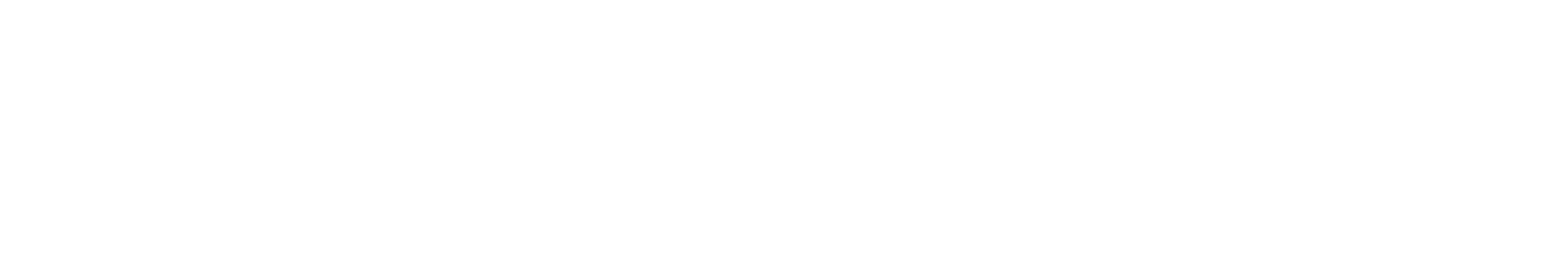
Wo gilt die DSGVO?

**Wählen Sie eine Antwort aus:**



**FRAGE 187 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Welche Python-Funktion wird für die Erstellung eines BoW-Modells genutzt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| fit\_transform() |
| defaultdict() |
| get\_feature\_name() |
| *CountVectorizer()* |



**FRAGE 188 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**

|  |
| --- |
| Nur in Kanada. |
| Nur in den USA. |
| *Nur in EU-Ländern.* |
| In allen Ländern. |







**FRAGE 189 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



Welche der folgenden Aussagen über die Analyse sozialer Medien ist richtig?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Die Daten sind meist für jedermann zugänglich.* |
| Die Daten sind normalerweise vertraulich. |
| Die Daten werden aus den Verwaltungssystemen der Organisation gesammelt. |
| Die Daten liegen meist in einem strukturierten Format vor. |



**FRAGE 190 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



Die Tf-idf-Funktion gibt die Bedeutung … wieder.

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *eines bestimmten Wortes in einem Korpus* |
| eines bestimmten Wortes in einem Dokument des Korpus |
| eines Wortes in einem bestimmten Text |
| eines Wortes in einem Satz |







Welche der folgenden Aussagen über Geschäftsanalysen ist richtig?

**Wählen Sie eine Antwort aus:**



**FRAGE 191 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



Welche der folgenden Beschreibungen bezeichnet Wörter mit höherer Bedeutung?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Wörter mit einem kleineren IDF-Wert. |
| *Wörter mit einem größeren IDF-Wert.* |
| Wort mit einem IDF-Wert von Null. |
| Wörter mit einem Tf-idf-Maß von Null. |



**FRAGE 192 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**

|  |
| --- |
| Die Daten sind in der Regel für die Öffentlichkeit zugänglich. |
| Der Wert der Daten hängt vom Grad der Teilung bzw. Veröffentlichung ab. |
| Die Daten liegen in der Regel sowohl in strukturierten als auch in unstrukturierten Formaten vor. |
| *In der Regel handelt es sich um historische Daten.* |







**FRAGE 193 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Welche der folgenden Techniken gehört zur Textbereinigung bei der natürlichen Sprachverarbeitung?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Tokenisierung |
| Wortzählung |
| *Lemmatisierung* |
| Wort-Vektorisierung |



**FRAGE 194 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



SciKit-learn ist eine Python-Bibliothek, die Folgendes bietet:

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Verarbeitung von Textinformationen. |
| Verarbeitung von Datenstrukturen in feldähnlicher Form. |
| *Unterstützung unterschiedlicher Klassifizierungsalgorithmen.* |
| Datenanalysen und Datenmanipulation. |







**FRAGE 195 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Welcher der folgenden Algorithmen ist ein Klassifizierungsalgorithmus?

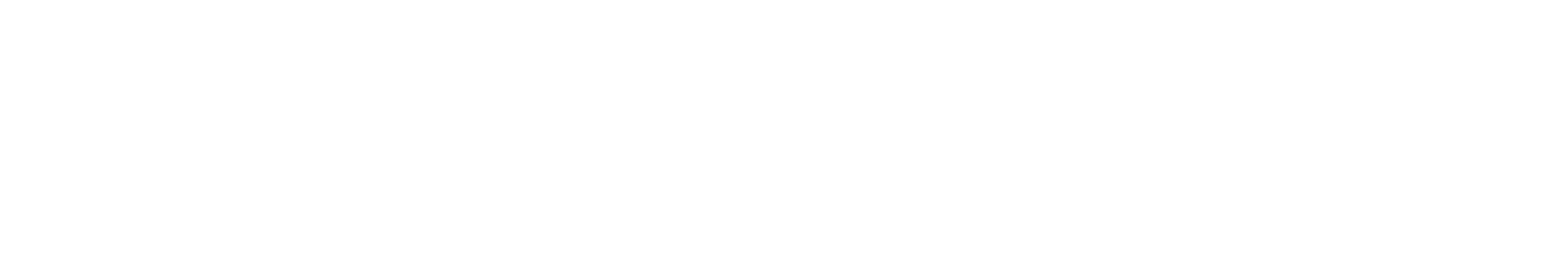
**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Support Vector Machines* |
| SVD |
| IDF |
| TF-IDF |



**FRAGE 196 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



Mit welcher der folgenden Python-Funktionen kann eine Zeichenkette in n-Gramme unterteilt werden?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| w.isalpha() |
| w.lower() |
| get\_feature\_name() |
| *word\_tokenize()* |







Welcher Begriff gehört zum Konzept von „Big Data“?

**Wählen Sie eine Antwort aus:**



**FRAGE 197 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Welche der folgenden Aussagen trifft auf die semantische Analyse zu?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Sie beschreibt den Prozess des Verstehens von Texten in natürlicher Sprache.* |
| Sie konzentriert sich auf die Bedeutung eines bestimmten Wortes. |
| Sie wird zur Analyse von Satzstrukturen genutzt. |
| Sie prüft nur die Häufigkeit eines Wortes, aber nicht seine Bedeutung. |



**FRAGE 198 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**

|  |
| --- |
| *Vielfalt* |
| Virtualisierung |
| Visuell |
| Vektor |







Welche der folgenden Techniken gehört zur semantischen Analyse?

**Wählen Sie eine Antwort aus:**



**FRAGE 199 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**

|  |
| --- |
| *Latent Dirichlet allocation* |
| OAuth |
| Multi-armed-Bandit |
| Analyse der Entitäten |



**FRAGE 200 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**



Pandas ist eine Datenbibliothek, die Folgendes bietet:

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Unterstützung unterschiedlicher Klassifizierungsbibliotheken. |
| *Datenanalysen und ‑manipulation.* |
| Verarbeitung von Datenstrukturen in feldähnlicher Form. |
| Verarbeitung von Textinformationen. |







Welche Plattform hat die meisten Einschränkungen und Überprüfungsprozesse, damit Anwendungen die Daten sicher nutzen können?

**Wählen Sie eine Antwort aus:**



**FRAGE 201 VON 315**

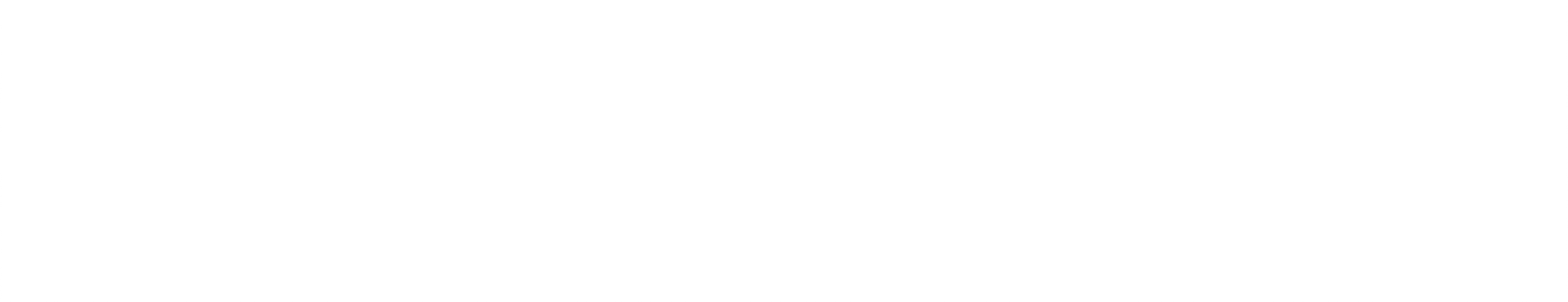
**DLBDSEDA01\_MC\_mittel/Lektion 04**

|  |
| --- |
| YouTube |
| Pinterest |
| *Facebook* |
| Twitter |



**FRAGE 202 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



In welcher(n) Situation(en) wird die latente semantische Analyse verwendet?

Antwortmöglichkeit 1: Textklassifizierung Antwortmöglichkeit 2: Reduzierung der Dimensionen

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig.* |
| Weder Antwortmöglichkeit 1 noch Antwortmöglichkeit 2 ist richtig. |
| Nur Antwortmöglichkeit 2 ist richtig. |
| Nur Antwortmöglichkeit 1 ist richtig. |







**FRAGE 203 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Welche LSA-Methode wird für die Dimensionsreduzierung einer Dokumenten-Begriffsmatrix genutzt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| TF-IDF |
| BoW |
| *SVD* |
| MVT |



**FRAGE 204 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Welches der folgenden Werkzeuge wird für das Experimentieren mit Webseiten eingesetzt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Google Analytics |
| Graph API |
| *VWO* |
| ProM |







**FRAGE 205 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Was ist die einfachste Form des Experimentierens?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Der Epsilon-Greedy-Algorithmus |
| Multi-armed-Bandit |
| *A/B-Tests* |
| Multivariate Tests |



**FRAGE 206 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**



Nachdem die Klassifizierung mit Tf-idf abgeschlossen ist, wird eine Matrix A (m x n) erstellt.

Wofür stehen die Zeilen in dieser Matrix?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Für die Anzahl der Absätze. |
| Für die Anzahl der Einzelwörter. |
| *Für die Anzahl der Dokumente.* |
| Für die Anzahl der Themen. |







**FRAGE 207 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**



Nachdem die Klassifizierung mit Tf-idf abgeschlossen ist, wird eine Matrix A (m x n) erstellt.

Wofür stehen die Spalten in dieser Matrix?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Für die Anzahl der Einzelwörter.* |
| Für die Anzahl der Themen. |
| Für die Anzahl der Dokumente. |
| Für die Anzahl der Absätze. |



**FRAGE 208 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Was wird in einer Teststrategie als Konversion bezeichnet?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Der Punkt, an dem Nutzende die gewünschte Aktion auf der Website ausführen.* |
| Der Punkt, an dem die beste Version der Seite bestimmt wird |
| Der Punkt, an dem Nutzende die Website verlassen. |
| Der Punkt, an dem der Datenverkehr zwischen den Testvarianten aufgeteilt wird. |







**FRAGE 209 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Warum ist es bei LSA notwendig, die Dimensionen der während der Tf-idf-Klassifizierung erstellten Matrix zu reduzieren?

Antwortmöglichkeit 1: Die Matrix ist dünnbesetzt. Antwortmöglichkeit 2: Die Matrix ist verrauscht.

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Nur Antwortmöglichkeit 1 ist richtig. |
| Weder Antwortmöglichkeit 1 noch Antwortmöglichkeit 2 ist richtig. |
| *Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig.* |
| Nur Antwortmöglichkeit 2 ist richtig. |



**FRAGE 210 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**



Wie werden die Besucher:innen bei A/B-Tests aufgeteilt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Abhängig von ihrer IP-Adresse. |
| *Nach dem Zufallsprinzip, unabhängig von anderen Faktoren.* |
| Abhängig von der Zeit des Zugriffs auf die Website. |
| Je nach ihren Vorlieben. |







Welches ist die gängigste Testmethode für Unternehmen, die gerade erst mit dem Testen begonnen haben?

**Wählen Sie eine Antwort aus:**



Wofür steht der Begriff „SVD“?

**Wählen Sie eine Antwort aus:**



**FRAGE 211 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**

|  |
| --- |
| *A/B-Tests* |
| Multi-armed-Bandit-Tests |
| Multivariate Tests |
| Sicherheitstests |



**FRAGE 212 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**

|  |
| --- |
| Simple vector decomposition |
| Secure vector data |
| Sparse vector document |
| *Singular value decomposition* |







Welche der folgenden Methoden kann verwendet werden, um die k-Anzahl der bei SVD gewählten Themen zu bestimmen?

**Wählen Sie eine Antwort aus:**



Welche der folgenden Aussagen ist richtig?

**Wählen Sie eine Antwort aus:**



**FRAGE 213 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**

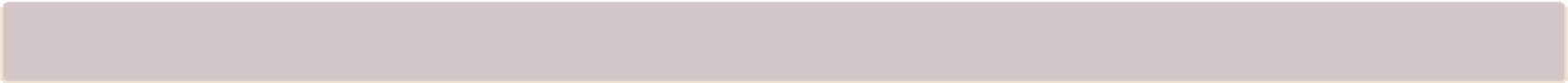
|  |
| --- |
| Silhouettenkoeffizient |
| LDA |
| IDF |
| Themenmodellierung |



**FRAGE 214 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**

|  |
| --- |
| A/B-Tests sind eine der kompliziertesten und teuersten Experimentiertechniken. |
| A/B-Tests erfordern in der Regel umfangreiche Ressourcen für ihre Durchführung. |
| *Es ist schwierig zu verstehen, welche Elemente bei A/B-Tests die Gewinnerseite beeinflussen.* |
| Es ist schwierig zu verstehen, welche Elemente die Gewinnerseite bei multivariaten Tests beeinflussen. |







Welche der folgenden Aussagen ist richtig?

**Wählen Sie eine Antwort aus:**



Gegeben sei die folgende Funktion:

TruncatedSVD(n\_components=2, algorithm='randomized', n\_iter=10).

Was bezeichnet der Parameter „n\_components“?

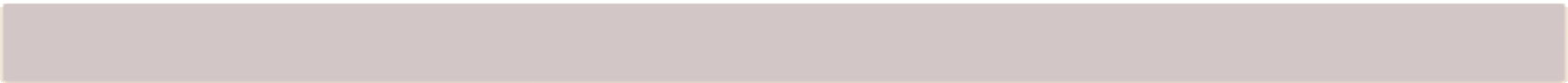
**Wählen Sie eine Antwort aus:**



**FRAGE 215 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**

|  |
| --- |
| *Vor dem Beginn der Experimentierphase muss der Hauptzweck der Website klar sein.* |
| Der letzte Schritt der Teststrategie besteht darin, zu definieren, was für die Website als Konversion gilt und was nicht. |
| Ein Test, der fehlschlägt oder negative Ergebnisse liefert, darf nicht zu weiteren Hypothesen führen. |
| Der beste Zeitpunkt für A/B-Tests ist, wenn alle Web-Metriken auf dem gewünschten Niveau sind. |



**FRAGE 216 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**

|  |
| --- |
| Die Anzahl der Zeilen. |
| *Die Anzahl der Themen.* |
| Die Anzahl der Spalten. |
| Die Anzahl der Wörter. |







Gegeben sei die folgende Funktion:

TfidfVectorizer(use\_idf=True, max\_features=5, smooth\_idf=True),

Was bezeichnet der Parameter „max\_features“?

**Wählen Sie eine Antwort aus:**



**FRAGE 217 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Was beeinflusst die Anzahl der Kombinationen in einem multivariaten Test?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Die verfügbare Zeit. |
| Das technische Personal. |
| Die Ressourcen des Unternehmens. |
| *Die Anzahl der geänderten Komponenten.* |



**FRAGE 218 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**

|  |
| --- |
| Begriffe mit einer Dokumentenhäufigkeit von weniger als 5 werden nicht berücksichtigt. |
| *Es werden nur die besten Merkmale berücksichtigt.* |
| Begriffe mit einer Dokumentenhäufigkeit von mehr als 5 werden nicht berücksichtigt. |
| Es werden fünf oder mehr Merkmale berücksichtigt. |







Welche der folgenden Aussagen ist richtig?

**Wählen Sie eine Antwort aus:**



Multivariate Tests sind …

**Wählen Sie eine Antwort aus:**



**FRAGE 219 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**

|  |
| --- |
| *LDA nutzt die Dirichlet-Verteilung.* |
| LDA steht für Latent Distributed Algorithm. |
| LDA kann die in diesen Sätzen enthaltenen Themen nicht automatisch erkennen. |
| LDA ist ein überwachter Algorithmus. |



**FRAGE 220 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**

|  |
| --- |
| eine Art von Test, bei dem nur ein Element getestet wird. |
| eine Art von Test, die nicht mit anderen Testmethoden verbunden ist. |
| *eine verallgemeinerte Art von A/B-Tests.* |
| eine Art von Multi-armed-Bandit-Tests. |







Ein Modul zur Durchführung des LDA-Algorithmus gibt es im …

**Wählen Sie eine Antwort aus:**



**FRAGE 221 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Wenn wir auf einer Website die beiden Elemente Überschrift und Navigationsleiste (von denen es jeweils drei Varianten gibt) ändern möchten, wie hoch ist dann die Gesamtzahl der Kombinationen?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Neun* |
| Vier |
| Sechs |
| Acht |



**FRAGE 222 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**

|  |
| --- |
| Numpy-Paket. |
| pandas-Paket. |
| SciPy-Paket. |
| *sklearn-Paket.* |







**FRAGE 223 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Was ist der erste Schritt im LDA-Algorithmus?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Auswahl einer Anzahl m von zu berücksichtigenden Wörtern. |
| Überprüfung jedes Dokuments, um jedem Wort ein Thema zuzuordnen. |
| *Auswahl einer Zahl k von Themen.* |
| Aktualisieren des Themas für jedes Wort. |



**FRAGE 224 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**



Der Parameter α (Alpha) im LDA-Algorithmus basiert auf …

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *der Mischung der Themen, die extrahiert werden sollen.* |
| der Anzahl der berücksichtigten Dokumente. |
| der Anzahl der Wörter im Wortschatz. |
| der Anzahl der Wörter in jedem Dokument. |







Der Parameter β (Beta) im LDA-Algorithmus definiert …

**Wählen Sie eine Antwort aus:**



**FRAGE 225 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**

|  |
| --- |
| *die Kontrolle über die Verteilung der Wörter für jedes Thema.* |
| die Kontrolle der Verteilung des Wortes über den gesamten Korpus. |
| die Kontrolle der Verteilung des Themas über den gesamten Korpus. |
| die Kontrolle über die Verteilung des Themas für jedes Dokument. |



**FRAGE 226 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**



Mit welcher Gleichung (Formel) wird bei LDA die Beziehung zwischen einem Thema und einem Dokument berechnet?

Die Abkürzungen haben dabei die folgenden Bedeutungen:

no\_dt = Anzahl der Wörter in einem Dokument, die zum Thema gehören no\_d = Anzahl der Wörter in einem Dokument

no\_t = Anzahl der Themen

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| p(topic t|document d) = (no\_dt)/(no\_d) |
| *p(topic t|document d) = (no\_dt+α)/(no\_d-1+no\_t\*α)* |
| p(topic t|document d) = (no\_dt+α)/(no\_d-1) |
| p(topic t|document d) = (no\_dt+α)/(no\_t\*α) |







Welchen typischen Wert erhalten die Parameter α und β?

**Wählen Sie eine Antwort aus:**



**FRAGE 227 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 02**



Mit welcher Gleichung wird die Bedeutung eines bestimmten Wortes innerhalb eines Themas berechnet?

Die Abkürzungen haben dabei die folgenden Bedeutungen: no\_wt = Anzahl der Wörter, die zum Thema gehören no\_v = Gesamtzahl der Wörter im Wortschatz

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *p(word w| topic t)= (n\_wt+β)/(Ʃno\_wt-1+no\_v\*β)* |
| p(word w| topic t)= (n\_wt+β)/(Ʃno\_wt+no\_v) |
| p(word w| topic t)= (n\_wt+β)/(Ʃno\_wt-1) |
| p(word w| topic t)= (n\_wt)/(Ʃno\_wt-1+no\_v\*β) |



**FRAGE 228 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 02**

|  |
| --- |
| *Weniger als 1* |
| Mehr als 1 |
| Größer als 10 |
| Weniger als 0 |







**FRAGE 229 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 02**



Welche der folgenden Aussagen über den SVD-Algorithmus ist richtig?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Er zerlegt eine Matrix in genau zwei andere Matrizen. |
| Er reduziert die Anzahl der Dimensionen um genau 1. |
| Er erhöht die Anzahl der Raumdimensionen des Datensatzes. |
| *Er ist eine Methode der Matrixfaktorisierung.* |



**FRAGE 230 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**



Welcher Begriff bezeichnet einen bestimmten Besuch auf einer Website?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Tags |
| Cookies |
| *Sitzungen* |
| Benutzer-ID |







Welcher Begriff beschreibt eine Technologie, die für die Identifikation eindeutiger bzw. unterscheidbarer Besucher:innen einer Website verwendet wird?

**Wählen Sie eine Antwort aus:**



**FRAGE 231 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**

|  |
| --- |
| Tags |
| Sitzungen |
| *Cookies* |
| Benutzer-ID |



**FRAGE 232 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Wie wird die durchschnittliche Sitzungsdauer in Google Analytics gemessen?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Gesamtzeit aller Sitzungen / die Anzahl der besuchten Seiten |
| Die auf einer Seite verbrachte Zeit / die Gesamtzahl der Sitzungen |
| Die auf einer Seite verbrachte Zeit / die Gesamtzahl der besuchten Seiten |
| *Gesamtdauer der Sitzung / Anzahl der Sitzungen* |







**FRAGE 233 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**



Eine hohe Absprungrate ist in der Regel **kein** Problem bei …

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Nachrichten-Websites. |
| Websites für soziale Medien. |
| e-Commerce-Websites. |
| *Websites mit einer einzigen Seite.* |



**FRAGE 234 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Wie beschreibt die Ausstiegsquote?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Die Anzahl der Personen, die eine Website von einer bestimmten Seite aus verlassen haben.* |
| Die Anzahl der Personen, die ohne etwas zu kaufen die Website verlassen haben. |
| Die Anzahl der Personen, die eine Website schon nach weniger als einer Minute wieder verlassen haben. |
| Die Anzahl der Personen, die ohne eine Aktion die Website wieder verlassen haben. |







**FRAGE 235 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Wie wird die Konversionsrate auf einer E-Commerce-Website berechnet?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Anzahl der Sitzungen / Anzahl der Einzelbesucher:innen |
| Anzahl der Besucher:innen / Anzahl der Sitzungen |
| *Die Anzahl der Bestellungen / die Anzahl der (Einzel-)Besucher:innen oder die Anzahl der Besuche* |
| Anzahl der Einzelbesucher:innen / Gesamtzahl der Besucher:innen oder Besuche |



**FRAGE 236 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welche der folgenden Möglichkeiten wird in Google-Analytics-Berichten **nicht** analysiert?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Die Art des Browsers* |
| Seitenansichten |
| Durchschnittliche Besuchsdauer |
| Durchschnittliche Seitenladezeit |







**FRAGE 237 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**



Bei den Standardeinstellungen von Google Analytics gilt eine Sitzung als beendet, …

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| wenn der Nutzende 30 Minuten lang inaktiv war. |
| wenn der Nutzende einen anderen Browser öffnet. |
| wenn der Nutzende eine andere Browser-Registerkarte (Tab) öffnet. |
| nach 30 Minuten, unabhängig davon, ob der Nutzende aktiv oder inaktiv ist. |



**FRAGE 238 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welcher der folgenden Schritte gehört **nicht** zu den Clickstream-Analysen?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Sammeln von notwendigen Informationen |
| *Vorverarbeitung der gesammelten Daten* |
| Verbesserung der Benutzererfahrung |
| Verstehen der Bedürfnisse der Organisation |







Wofür steht SEO?

**Wählen Sie eine Antwort aus:**



**FRAGE 239 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**



Mit welcher Web-Metrik kann die Absicht der Besucher:innen einer Website eingeschätzt werden?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Absprungrate |
| *Interne Website-Suche* |
| Ausstiegsrate |
| E-Mail-Kampagnen |



**FRAGE 240 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**

|  |
| --- |
| System encryption optimization |
| Search engine output |
| Session engine optimization |
| *Search engine optimization* |







**FRAGE 241 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Was ist direkter Datenverkehr?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Datenverkehr, der von sozialen Medienplattformen kommt. |
| Datenverkehr, der von E-Mail-Kampagnen kommt. |
| Datenverkehr, der von Suchkampagnen stammt. |
| *Datenverkehr, der nicht von einer anderen verlinkten Website stammt.* |



**FRAGE 242 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Was ist organischer Datenverkehr?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Datenverkehr, der von Suchkampagnen stammt. |
| Datenverkehr, der von sozialen Medienplattformen kommt. |
| Datenverkehr, der nicht von einer anderen verlinkten Website stammt. |
| *Datenverkehr, der von Suchmaschinen kommt (nicht bezahlt).* |







Was sollte am Ende einer URL-Adresse eingefügt werden, um Informationen über die Person zu speichern, die auf einen Link geklickt hat?

**Wählen Sie eine Antwort aus:**



**FRAGE 243 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**

|  |
| --- |
| Seitenumbruch |
| Suchanfrage |
| Filterabfragen |
| *UTM-Parameter* |



**FRAGE 244 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welche der folgenden URL-Parameter sollten Teil einer Tracking-Kampagne (Nachverfolgung) sein?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| UTM\_string |
| *UTM\_medium* |
| UTM\_input |
| UTM\_search |







**FRAGE 245 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welchen Wert hat die Quelle bzw. der Ursprung des Verweises in der folgenden URL?

https://[www.option.com?](http://www.option.com/) utm\_source=newsletter&utm\_medium=monthly&utm\_campaign=Example

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Newsletter* |
| Example |
| Monthly |
| Option |



**FRAGE 246 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Wie wird die Klickrate berechnet, wenn eine Website über eine erstellte URL aufgerufen wird?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Click\_rate = (Anzahl der Klicks)/(Anzahl der nicht zustellbaren E-Mails) |
| Click\_rate = (Anzahl der Klicks)/(Anzahl der gesendeten E-Mails ) |
| Click\_rate = (Anzahl der Klicks)/(Anzahl der Einzelbesucher:innen) |
| *Click\_rate = (Anzahl der Klicks)/(Anzahl der geöffneten E-Mails)* |

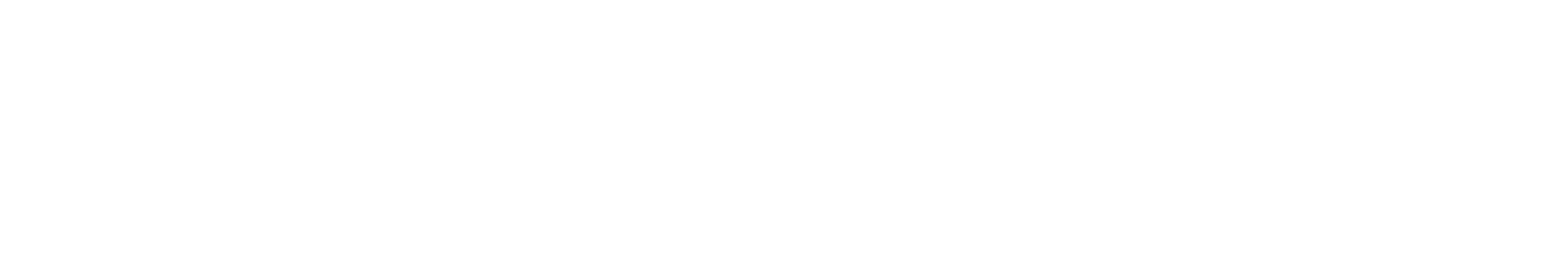






**FRAGE 247 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Wie wird die Absprungrate berechnet, wenn der Zugriff auf eine Website über eine erstellte URL erfolgt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Bounce\_rate = (Anzahl der Besucher:innen, die eine Aktion ausführen)/(Anzahl der Besuche) |
| Bounce\_rate = (Anzahl der Klicks)/(Anzahl der Besuche ) |
| *Bounce\_rate = (Anzahl der Besuche mit einer einzigen Seitenansicht)/(Anzahl der Besuche)* |
| Bounce\_rate = (Anzahl der Einzelbesucher:innen)/(Anzahl der Besuche) |



**FRAGE 248 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Ein Empfehlungssystem ist …

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| ein manuelles Verfahren. |
| *eine Reihe von Methoden zum Filtern von Informationen.* |
| nicht auf früheren Informationen basierend. |
| eine E-Commerce-Website. |







Welche der folgenden Aussagen über den inhaltsbasierten Ansatz ist richtig?

**Wählen Sie eine Antwort aus:**



**FRAGE 249 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welcher der folgenden Ansätze wird in Empfehlungssystemen genutzt?

**Wählen Sie eine Antwort aus:**

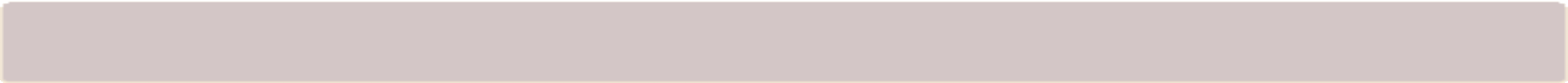
|  |
| --- |
| Multi-armed-Bandit |
| *Inhaltsbasierter Ansatz* |
| Analyse der Merkmale (Entitäten) |
| Tf-idf |



**FRAGE 250 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**

|  |
| --- |
| *Damit wird ein Modell erstellt, das die Interaktionen zwischen Nutzenden und Produkt auf der Grundlage der Eigenschaften der Nutzenden erklärt.* |
| Er basiert auf den Meinungen anderer Nutzender mit ähnlichen Interessen. |
| Die Algorithmen werden in zwei Gruppen unterteilt: speicherbasiert und modellbasiert. |
| Er prüft die Ähnlichkeit zwischen den Vorlieben der Nutzenden einer Plattform. |







Welche der folgenden Aussagen bezieht sich auf kollaborative Filteransätze?

**Wählen Sie eine Antwort aus:**



**FRAGE 251 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welcher der folgenden Punkte ist ein Nachteil von inhaltsbasierten Empfehlungssystemen?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Es besteht keine Gefahr einer Überspezialisierung. |
| Sie basieren auf den vergangenen Aktivitäten des Nutzenden. |
| *Es gibt keine bekannte Historie für neue Nutzende im System.* |
| Inhalte können immer in Stichwörtern beschrieben werden. |



**FRAGE 252 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**

|  |
| --- |
| Damit wird ein Modell erstellt, das die Interaktionen zwischen Nutzenden und Produkt auf der Grundlage der Eigenschaften der Nutzenden erklärt. |
| *Die Algorithmen werden in zwei Gruppen unterteilt: speicherbasiert und modellbasiert.* |
| Je weniger Nutzende, desto effektiver ist das Empfehlungssystem. |
| In der Anfangsphase ist es einfacher, Empfehlungen abzugeben. |







Welche Kombination nutzt der hybride Ansatz in Empfehlungssystemen?

**Wählen Sie eine Antwort aus:**



**FRAGE 253 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welcher Ansatz kann für Empfehlungssysteme genutzt werden?

Antwortmöglichkeit 1: Wissensbasierte Ansätze Antwortmöglichkeit 2: Inhaltsbasierter Ansatz

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Nur Antwortmöglichkeit 2 ist richtig. |
| *Nur Antwortmöglichkeit 1 ist richtig.* |
| Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig. |
| Weder Antwortmöglichkeit 1 noch Antwortmöglichkeit 2 ist richtig. |



**FRAGE 254 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**

|  |
| --- |
| Kontextabhängige und wissensbasierte Ansätze |
| *Kollaboratives Filtern und inhaltsbasierter Ansatz* |
| Wissensbasierter Ansatz und Graph-basierter Ansatz |
| Kontextabhängige und Graph-basierte Ansätze |







**FRAGE 255 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 03**



Welches der folgenden Ähnlichkeitsmaße wird in Empfehlungssystemen genutzt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Euklidischer Abstand* |
| k-nächste-Nachbarn-Methode |
| Analyse der wichtigsten Komponente |
| Singulärwertzerlegung |



**FRAGE 256 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 03**



Zwei Werte für die Kosinus-Ähnlichkeit sind wie folgt angegeben:

cos(user1,place1) = 0,24, cos(user2,place1) = 0,0 Welche der folgenden Aussagen ist richtig?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Der Cosinus-Ähnlichkeitswert zeigt, dass der erste Platz für keinen Nutzenden relevant ist. |
| Da diese Werte für die Kosinus-Ähnlichkeit nicht zulässig sind, sind die Berechnungen ungenau. |
| *Die Kosinus-Ähnlichkeit zeigt, dass der erste Ort für den ersten Nutzenden relevanter ist.* |
| Die Kosinus-Ähnlichkeit zeigt, dass der erste Ort für den zweiten Nutzenden relevanter ist. |







Welche der folgenden Aussagen beschreibt die Nutzung des KNN-Algorithmus beim kollaborativen Filtern richtig?

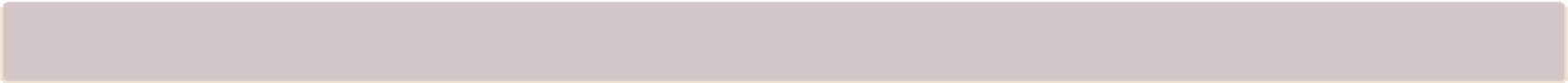
**Wählen Sie eine Antwort aus:**



**FRAGE 257 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 03**

|  |
| --- |
| Der KNN-Algorithmus verknüpft einen neuen Nutzenden X direkt mit der am wenigsten häufigen Kategorie. |
| *Der KNN-Algorithmus zählt die Anzahl der Vorkommen jeder Kategorie für die k Nutzenden bzw. Produkte.* |
| Die KNN-Methode vergleicht verschiedene Nutzende auf der Grundlage der ihnen zugeordneten Merkmale. |
| Der KNN-Algorithmus ist eine einfache Methode und ein unüberwachter Algorithmus. |



**FRAGE 258 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Welcher der folgenden Dienste ist ein Microblogging-Dienst?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Snapchat |
| Facebook |
| *Twitter* |
| Instagram |







Wofür steht der Begriff „SMA“?

**Wählen Sie eine Antwort aus:**



**FRAGE 259 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Welche der folgenden Social-Media-Plattformen ist der Marktführer?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| TikTok |
| Pinterest |
| *Facebook* |
| Twitter |



**FRAGE 260 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**

|  |
| --- |
| Social media algorithms |
| Social media application |
| Social media accessibility |
| *Social media analytics* |







**FRAGE 261 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Was macht eine Daten-API?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Eine Daten-API sorgt für den Authentifizierungsprozess. |
| Eine Daten-API sorgt für die Initialisierung der Verbindung. |
| Eine Daten-API sorgt für die Verarbeitung der Daten. |
| *Eine Daten-API bietet einen strukturierten Zugriff auf die Daten.* |



**FRAGE 262 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Was verbirgt sich hinter dem Begriff „GDPR“?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *General Data Protection Regulation* |
| Global Data Privacy Rules |
| General Data Protocol Regulations |
| Global Data Processing Regulation |







**FRAGE 263 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Was ist der Hauptzweck der Nutzung eines Jupyter-Notebooks?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Gewährleistung eines höheren Maßes an Privatsphäre für die Nutzenden. |
| Berechnung der Zeit für die Ausführung. |
| Automatische Kategorisierung der Daten. |
| *Live-Programmierung und Erstellung von Visualisierungen.* |



**FRAGE 264 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



Was sind RDFa-Metadaten?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Ein Javascript-Tag |
| *Eine HTML5-Erweiterung* |
| Eine HTTP-Verbindung |
| Eine CSS-Datei |







Welches der folgenden Facebook-Projekte ist veraltet?

**Wählen Sie eine Antwort aus:**



**FRAGE 265 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**

|  |
| --- |
| *Facebook Graph Search* |
| Facebook Local |
| Facebook for Developers |
| OGP |



**FRAGE 266 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**



Wie viele Zeichen darf ein Tweet derzeit enthalten?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| 100 |
| 1000 |
| 200 |
| *280* |







**FRAGE 267 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**



Wie sieht die Netzwerktopologie von Twitter aus?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Einseitig gerichtet* |
| Richtungsunabhängig |
| Zweiseitig gerichtet |
| Multidirektional |



**FRAGE 268 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



Welche der folgenden Elemente gehören zu den Metadaten eines Tweets?

Antwortmöglichkeit 1: Entitäten Antwortmöglichkeit 2: Orte

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Nur Antwortmöglichkeit 1 ist richtig. |
| Nur Antwortmöglichkeit 2 ist richtig. |
| Weder die Antwortmöglichkeit 1 noch die Antwortmöglichkeit 2 ist richtig. |
| *Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig.* |







**FRAGE 269 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**



Zu den Twitter-Entitäten gehören …

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| der in einem Tweet enthaltene Text. |
| *Hashtags.* |
| der Standort. |
| Benutzer-IDs. |



**FRAGE 270 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



Wie wird ein Tweet in JSON ausgedrückt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Datum-Text-Paar. |
| *Schlüssel-Werte-Paare.* |
| Datum-ID-Paar. |
| ID-Text-Paar. |







Welche der folgenden Aussagen über die Twitter-API ist richtig?

**Wählen Sie eine Antwort aus:**



**FRAGE 271 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



Wie viele Attribute kann ein Tweet enthalten?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Mehr als 1000 |
| *Mehr als 150* |
| Weniger als 150 |
| Weniger als 10 |



**FRAGE 272 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**

|  |
| --- |
| Die Streaming-API ist nützlich, wenn wir auf bereits veröffentlichte Tweets zugreifen möchten. |
| *Die REST-API ist nützlich, wenn wir auf bereits veröffentlichte Tweets zugreifen möchten.* |
| Oauth ermöglicht die Interaktion mit der Streaming-API. |
| Die REST-API öffnet einen Datenstrom zum Sammeln der kommenden Tweets. |







**FRAGE 273 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



Bei der Verwendung von Oauth …

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *wird das Passwort der Nutzenden niemals an Dritte weitergegeben.* |
| ist ein unbegrenzter Zugriff auf die Daten eines Nutzenden möglich. |
| ist die Privatsphäre des Nutzenden gefährdet. |
| wird das Passwort des Nutzenden nur einmal ausgetauscht. |



**FRAGE 274 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 04**



Welche Art von Tweets wird von Site-Datenströmen gesammelt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Tweets mit Bezug auf mehr als einen Nutzenden.* |
| Tweets mit Bezug auf einen bestimmten Nutzenden. |
| Tweets mit Bezug auf eine bestimmte Anfrage. |
| öffentliche Tweets. |







request\_1()

Der obige Code …

**Wählen Sie eine Antwort aus:**

request\_2() ….

time.sleep(20)



Welche der folgenden Marken (Token) wird bei der Erstellung einer Anwendung auf der Twitter Developer Platform erzeugt?

**Wählen Sie eine Antwort aus:**



**FRAGE 275 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**

|  |
| --- |
| führt die zweite Anfrage 20 Sekunden lang aus. |
| führt die erste Anfrage 20 Sekunden lang aus. |
| *erzeugt eine Verzögerung von 20 Sekunden zwischen zwei API-Anfragen.* |
| führt die zweite Anfrage nur aus, wenn die erste Anfrage weniger als 20 Sekunden gedauert hat. |



**FRAGE 276 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**

|  |
| --- |
| Ausgabe-Token |
| *Zugriffsmarke (Access Token)* |
| Geheime ID |
| Eingabe-Token |







Welche auf der Registerkarte „Berechtigungen“ in der Twitter Developer Plattform auswählbare Option ist die sicherste?

**Wählen Sie eine Antwort aus:**



**FRAGE 277 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**

|  |
| --- |
| Lesen und Schreiben |
| Nachrichten schreiben und weiterleiten |
| *Nur Lesen* |
| Schreiben |



**FRAGE 278 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**



Die Zugriffsmarken (Access Tokens) spiegeln … wider.

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *das Entwicklerkonto* |
| die Verbindungsanfrage |
| den API-Aufruf |
| eine bestimmte Anwendung |

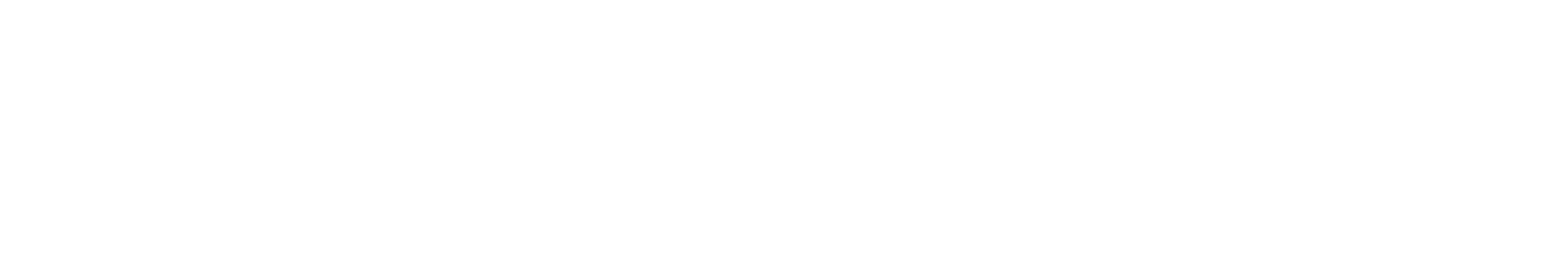






**FRAGE 279 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Wie lautet die Python-Funktion zum Sammeln aller Tweets der Konten, denen gefolgt wird?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| get\_timeline() |
| load\_timeline() |
| user\_timeline() |
| *home\_timeline()* |



**FRAGE 280 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**



Wie viele Tweets werden mit dem folgenden Befehl gesammelt?

Cursor(API.user\_timeline, screen\_name='nytimes', count=200).pages(2):

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| 100 |
| *400* |
| 200 |
| 2 |







**FRAGE 281 VON 315**

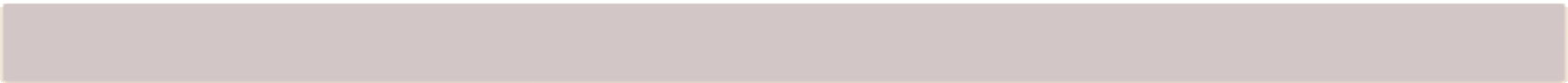
**DLBDSEDA01\_MC\_schwer/Lektion 04**



Was führt der Befehl hashtags.most\_common(10) aus?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Er listet die zehn am häufigsten genannten Hashtags in zufälliger Reihenfolge auf. |
| *Er ordnet die Schlüssel vom größten zum kleinsten und listet die zehn am häufigsten genannten Hashtags auf.* |
| Er ordnet die Schlüssel vom kleinsten zum größten und listet die zehn am häufigsten genannten Hashtags auf. |
| Es listet Hashtags auf, aber und nur wenn mehr als zehn Hashtags gefunden werden. |



**FRAGE 282 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 04**



Wofür kann die matplotlib-Bibliothek in Python verwendet werden?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Damit kann eine Feldstruktur (Array) zum Speichern der Termhäufigkeit erstellt werden. |
| Damit können die Wörter tokenisiert und die Termhäufigkeit gezählt werden. |
| Zur Berechnung der Termhäufigkeit. |
| *Zur Visualisierung der Termhäufigkeit in einem grafischen Format.* |







Was besagt das Pareto-Prinzip?

**Wählen Sie eine Antwort aus:**



**FRAGE 283 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**

|  |
| --- |
| 20 % der Begriffsvorkommen stammen aus 80 % der Einzelbegriffe. |
| 50 % der Begriffsvorkommen stammen aus 50 % der Einzelbegriffe. |
| *80 % der Begriffsvorkommen stammen aus 20 % der Einzelbegriffe.* |
| 60 % der Begriffsvorkommen stammen aus 40 % der Einzelbegriffe. |



**FRAGE 284 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Was besagt das Zipfsche Gesetz?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Das häufigste Wort wird dreimal so oft gefunden wie das zweithäufigste, und so weiter. |
| *Das häufigste Wort wird doppelt so oft gefunden wie das zweithäufigste, und so weiter.* |
| Das häufigste Wort wird viermal so oft gefunden wie das zweithäufigste, und so weiter. |
| Das häufigste Wort wird fünfmal so oft gefunden wie das zweithäufigste, und so weiter. |







Die Facebook Graph API ist …

**Wählen Sie eine Antwort aus:**



**FRAGE 285 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**



Welche der folgenden Antwortmöglichkeiten stellt sowohl einen sozialen Graphen als auch einen Interessengraphen dar?

Antwortmöglichkeit 1: Twitter Antwortmöglichkeit 2: YouTube

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Nur Antwortmöglichkeit 2 ist richtig. |
| Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig. |
| *Nur Antwortmöglichkeit 1 ist richtig.* |
| Weder Antwortmöglichkeit 1 noch Antwortmöglichkeit 2 ist richtig. |



**FRAGE 286 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 04**

|  |
| --- |
| eine FTP-basierte API. |
| eine SMTP-basierte API. |
| *eine HTTP-basierte API.* |
| eine IP-basierte API. |







Worauf bezieht sich der Begriff „Abbruch“?

**Wählen Sie eine Antwort aus:**



**FRAGE 287 VON 315**

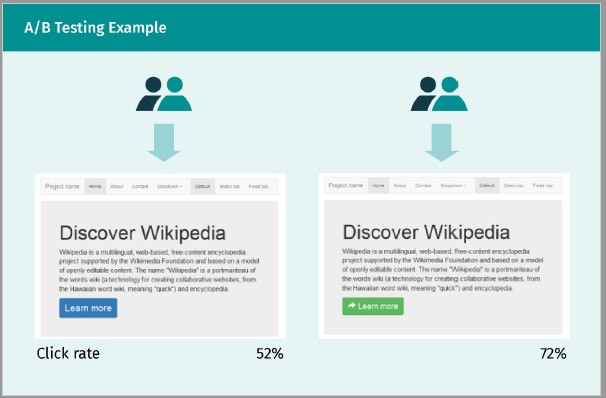
**DLBDSEDA01\_MC\_mittel/Lektion 05**

|  |
| --- |
| Besucher:innen, die von der Suchseite aus auf mehr als eine Website klicken. |
| Besucher:innen, die die Suchseite nach einer Sekunde verlassen. |
| *Besucher:innen, die ohne etwas anzuklicken die Suchseite wieder verlassen.* |
| Besucher:innen, die sich beschweren, dass die Suchseite nicht funktioniert. |



**FRAGE 288 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Analysieren Sie das Bild. Welche Version der Website hat am besten abgeschnitten?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Es gibt nicht genügend Daten, um die leistungsstärkste Website zu ermitteln. |
| Die beiden Versionen bringen die gleiche Leistung. |
| Die linke Version hat besser abgeschnitten. |
| *Die rechte Version hat besser abgeschnitten.* |







**FRAGE 289 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Unter welchem anderen Namen sind A/B-Tests auch bekannt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Website-Tests |
| Versionstests |
| Vergleichstests |
| *Split-Tests* |



**FRAGE 290 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Welche der folgenden Komponenten können mit A/B-Tests überprüft werden?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Navigationsleisten* |
| HTML-Tags |
| Javascript-Tags |
| XML-Tags |







Welcher der folgenden Werte für die Konversionsrate würde die beste Variante in einem multivariaten Test ergeben?

**Wählen Sie eine Antwort aus:**



**FRAGE 291 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**

|  |
| --- |
| *9,5 %* |
| 6,7 % |
| 2,6 % |
| 4,3 % |



**FRAGE 292 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**



Wann werden multivariate Tests normalerweise eingesetzt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Bei geringem Datenverkehr. |
| Wenn wir nur eine Komponente ändern möchten. |
| Wenn wir nicht viel Zeit zum Experimentieren haben. |
| *Wenn wir radikale Veränderungen vornehmen wollen.* |







Welche der folgenden Aussagen ist richtig?

**Wählen Sie eine Antwort aus:**



**FRAGE 293 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Was ist eine häufig verwendete Technik bei multivariaten Tests?

**Wählen Sie eine Antwort aus:**

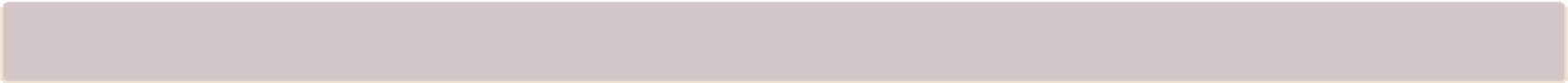
|  |
| --- |
| Der Epsilon-Greedy-Algorithmus |
| Der Softmax-Algorithmus |
| Multi-armed-Bandit |
| *Vollständige Versuchspläne (full factorial testing)* |



**FRAGE 294 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**

|  |
| --- |
| Teilfaktorpläne und vollständige Versuchspläne benötigen die gleiche Zeit für ihre Durchführung. |
| Die Durchführungszeit für vollständige Versuchspläne hängt von der Art der Daten ab, während dies bei Teilfaktorplänen nicht der Fall ist. |
| Teilfaktorpläne benötigen für die Durchführung mehr Zeit als vollständige Versuchspläne. |
| *Vollständige Versuchspläne benötigen für die Durchführung mehr Zeit als Teilfaktorpläne.* |







Welche der folgenden Aussagen ist richtig?

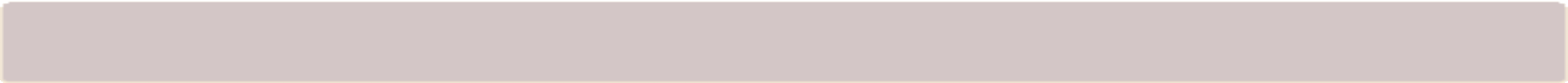
**Wählen Sie eine Antwort aus:**



**FRAGE 295 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**

|  |
| --- |
| Für die Durchführung vollständiger Versuchspläne ist ein höherer Datenverkehr erforderlich als für Teilfaktorpläne. |
| Weder vollständige Versuchspläne noch Teilfaktorpläne erfordern für ihre Durchführung ein bestimmtes Datenverkehrsaufkommen. |
| Vollständige Versuchspläne und Teilfaktorpläne benötigen für ihre Durchführung das gleiche Datenverkehrsaufkommen. |
| *Für die Durchführung von Teilfaktorplänen ist ein höherer Datenverkehr erforderlich als für vollständige Versuchspläne.* |



**FRAGE 296 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Welcher der folgenden Punkte ist ein Vorteil von multivariaten Tests?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Der Zeitaufwand für die Einrichtung und den Betrieb ist gering. |
| *Alle Kombinationen der verschiedenen Komponenten können getestet werden.* |
| Sie können auf Websites mit hohem Datenverkehrsaufkommen eingesetzt werden. |
| Sie haben kein hohes Risiko des Scheiterns. |







Was ist ein Nachteil von multivariaten Tests im Vergleich zu A/B-Tests?

**Wählen Sie eine Antwort aus:**



**FRAGE 297 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**

|  |
| --- |
| *Die Einrichtung und der Betrieb dauern länger.* |
| Mehrere Komponenten können nicht damit getestet werden. |
| Sie können keine Elemente erkennen, die nichts beitragen. |
| Sie helfen nicht bei der Strukturierung der Seiten. |



**FRAGE 298 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Welche Testmethode ist für technisch nicht versierte Personen am einfachsten zu nutzen?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Multivariate Tests |
| Teilfaktorpläne |
| Multi-armed-Bandit-Tests |
| *A/B-Tests* |







Woher hat der Multi-armed-Bandit-Algorithmus seinen Namen?

**Wählen Sie eine Antwort aus:**



**FRAGE 299 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**

|  |
| --- |
| Von einem stochastischen Prozess. |
| *Von Spielautomaten in Kasinos.* |
| Von einer Art der Wahrscheinlichkeitsverteilung. |
| Von Verkaufsautomaten für Lebensmittel. |



**FRAGE 300 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Welche Schwäche hat der Epsilon-Greedy-Algorithmus?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *Eine gewisse Zeit wird mit der Erkundung der Option verbracht, die nicht gewinnt.* |
| Die Hälfte der Zeit wird mit der Erkundung der Option verbracht, die nicht gewinnt. |
| 90 Prozent der Zeit werden mit dem Ausbeuten der Option verbracht, die nicht gewinnt. |
| 10 Prozent der Zeit werden mit dem Ausbeuten der Option verbracht, die nicht gewinnt. |







**FRAGE 301 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Was sollte in die Definition einer Epsilon-Greedy-Klasse aufgenommen werden?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Die Anzahl der durchzuführenden Erkundungen. |
| Die Anzahl der Arme. |
| *Die Anzahl der mit einem bestimmten Arm verbundenen Belohnungen.* |
| Die Anzahl der durchzuführenden Ausbeutungen. |



**FRAGE 302 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Was ist **kein** Multi-armed-Bandit-Algorithmus?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Der Softmax-Algorithmus |
| *Vollständige Versuchspläne* |
| Der UCB-Algorithmus |
| Der Epsilon-Greedy-Algorithmus |







Welcher der folgenden Begriffe gehört zum Multi-armed-Bandit-Test?

**Wählen Sie eine Antwort aus:**



**FRAGE 303 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**



Wie wird beim Multi-Armed-Bandit-Test der Datenverkehr aufgeteilt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| 90 % an Version A und 10 % an Version B |
| *Dynamisch* |
| Gleichmäßig |
| Manuell |



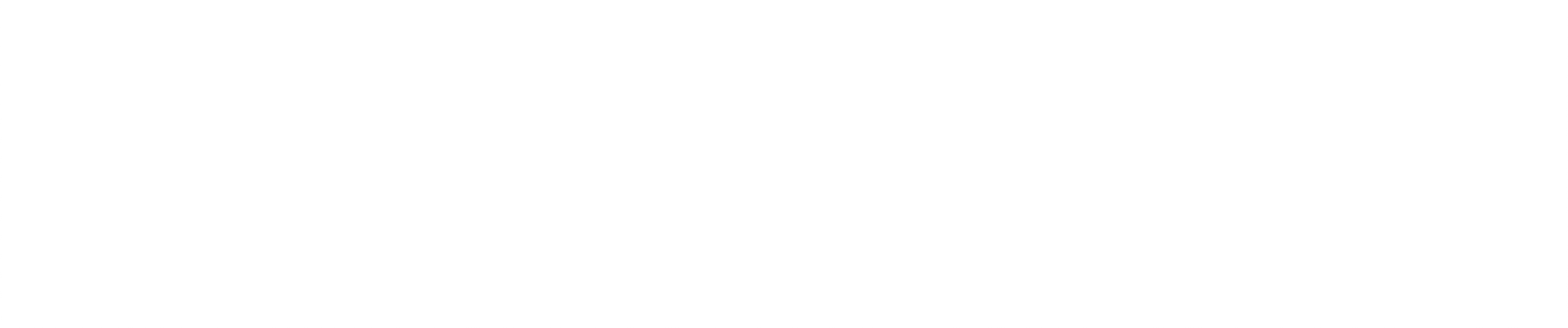
**FRAGE 304 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**

|  |
| --- |
| Kodierung |
| Erweiterbar |
| *Ausbeutung* |
| Verschlüsselung |







Welche Testmethode(n) sollte(n) verwendet werden, wenn eine Organisation statistische Ergebnisse für jede Variante ihrer Website wissen möchte?

Antwortmöglichkeit 1: A/B-Tests

Antwortmöglichkeit 2: Multi-armed-Bandit-Tests

**Wählen Sie eine Antwort aus:**



**FRAGE 305 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Der Multi-armed-Bandit-Algorithmus …

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| *weist den Datenverkehr dynamisch dem Arm mit der höchsten Erfolgsquote zu.* |
| weist 80 Prozent des Datenverkehrs dem Arm mit der höchsten Erfolgsquote und 20 Prozent dem Arm mit der niedrigsten Erfolgsquote zu. |
| weist den Datenverkehr dynamisch dem Arm mit der niedrigsten Erfolgsquote zu. |
| verteilt den Datenverkehr gleichmäßig auf den Arm mit der niedrigsten und den mit der höchsten Erfolgsquote. |



**FRAGE 306 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**

|  |
| --- |
| *Nur Antwortmöglichkeit 1 ist richtig.* |
| Nur Antwortmöglichkeit 2 ist richtig. |
| Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig. |
| Weder Antwortmöglichkeit 1 noch Antwortmöglichkeit 2 ist richtig. |







**FRAGE 307 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Welche Testmethode(n) sollte(n) verwendet werden, wenn eine Organisation keine Zeit zum Sammeln statistischer Befunde hat?

Antwortmöglichkeit 1: A/B-Tests

Antwortmöglichkeit 2: Multi-armed-Bandit-Tests

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Weder die Antwortmöglichkeit 1 noch die Antwortmöglichkeit 2 ist richtig. |
| Antwortmöglichkeit 1 und Antwortmöglichkeit 2 sind beide richtig. |
| *Nur Antwortmöglichkeit 2 ist richtig.* |
| Nur Antwortmöglichkeit 1 ist richtig. |



**FRAGE 308 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**



Welche der folgenden Aussagen über einen guten Multi-armed-Bandit-Algorithmus ist richtig?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Er konvergiert nie. |
| *Er konvergiert irgendwann.* |
| Er teilt die Ressourcen gleichmäßig auf beide getesteten Versionen auf. |
| Er erhöht langsam die Erkundungszeit. |







Welche der folgenden Aussagen ist richtig?

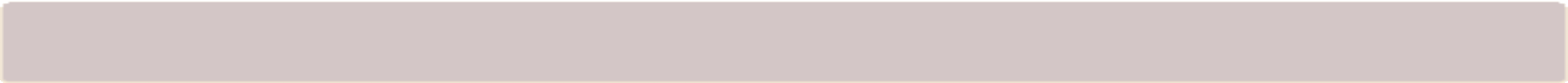
**Wählen Sie eine Antwort aus:**



**FRAGE 309 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**

|  |
| --- |
| *Ein gieriger Algorithmus wählt in einem bestimmten Moment die beste Aktion aus, ohne an die Konsequenzen zu denken.* |
| Ein gieriger Algorithmus wählt die beste zukünftige Aktion aus, ohne an die Konsequenzen zu denken. |
| Ein gieriger Algorithmus berücksichtigt keine Informationen, die in vorherigen Schritten gesammelt wurden. |
| Ein gieriger Algorithmus wählt die beste Aktion zu einem bestimmten Zeitpunkt und berücksichtigt dabei auch zukünftige Konsequenzen. |



**FRAGE 310 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**



Was macht ein Epsilon-Greedy-Algorithmus?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Er erkundet stets. |
| Er beutet stets aus. |
| Er erkundet nie und beutet nie aus. |
| *Er wechselt zufällig zwischen Erkunden und Ausbeutung.* |







Was kann hypothetisch für die Beeinflussung des Entscheidungsprozesses in einem Epsilon-Greedy-Algorithmus genutzt werden?

**Wählen Sie eine Antwort aus:**



**FRAGE 311 VON 315**

**DLBDSEDA01\_MC\_leicht/Lektion 05**

|  |
| --- |
| Die Wahrscheinlichkeit p(n) = 0,1. |
| Ein Wurf mit einem Würfel. |
| *Ein Wurf mit einer Münze.* |
| Die Wahrscheinlichkeit p(n) = 0,3. |



**FRAGE 312 VON 315**

**DLBDSEDA01\_MC\_schwer/Lektion 05**



Was passiert in der Ausbeutungsphase des Epsilon-Greedy-Algorithmus?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Der Algorithmus wählt mit der Wahrscheinlichkeit von 1 – Epsilon eine der Varianten aus. |
| Der Algorithmus wählt zufällig zwischen den Varianten aus. |
| *Der Algorithmus prüft die höchste Konversionsrate bis zu diesem Punkt.* |
| Der Algorithmus wählt mit der Wahrscheinlichkeit Epsilon eine der Varianten aus. |







Angenommen wir werfen eine Münze. Sie fällt mit einer Wahrscheinlichkeit von Epsilon auf Zahl. Bei Zahl werden beide Varianten der Website erkundet. Fällt sie auf Kopf, wird ausgebeutet. Wie hoch ist die Wahrscheinlichkeit, dass der Algorithmus ausbeutet?

**Wählen Sie eine Antwort aus:**



**FRAGE 313 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**

|  |
| --- |
| Epsilon/6 |
| Epsilon/3 |
| Epsilon/2 |
| *1 – Epsilon* |



**FRAGE 314 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**



Angenommen wir werfen eine Münze. Sie fällt mit einer Wahrscheinlichkeit von Epsilon auf Zahl. Bei Zahl werden beide Varianten der Website erkundet. Fällt sie auf Kopf, wird ausgebeutet. Wie hoch ist die Wahrscheinlichkeit, dass der Algorithmus die beste Option auswählt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| Epsilon/6 |
| 1 – Epsilon |
| *Epsilon/2* |
| Epsilon/4 |







**FRAGE 315 VON 315**

**DLBDSEDA01\_MC\_mittel/Lektion 05**



Angenommen wir werfen eine Münze. Sie fällt mit einer Wahrscheinlichkeit von Epsilon auf Zahl. Bei Zahl werden beide Varianten der Website erkundet. Fällt sie auf Kopf, wird ausgebeutet. Wie hoch ist die Wahrscheinlichkeit, dass der Algorithmus die schlechteste Option auswählt?

**Wählen Sie eine Antwort aus:**

|  |
| --- |
| 1 – Epsilon |
| Epsilon/4 |
| *Epsilon/2* |
| Epsilon/6 |

