

Übungsblatt 5

Abgabe: Bis **Montag, den 08.07.2024, bis 23:59 Uhr** über Moodle. Die Übungsblätter sind in Gruppen von drei (in Ausnahmefällen zwei) Studierenden zu bearbeiten. Die Lösungen sind, sofern nicht anders angegeben, auf nach Aufgaben getrennten **PDFs** über Moodle abzugeben. Alle Teilaufgaben einer Aufgabe sind in einem PDF hochzuladen. Pro Gruppe genügt es, wenn eine Person die Lösung der Gruppe abgibt. Zur Bewertung wird der zuletzt hochgeladene Stand herangezogen. Vermerken Sie auf allen Abgaben Ihre **Namen**, Ihre **CMS-Benutzernamen** und Ihre **Abgabegruppe (z.B. Gruppe 123)** aus Moodle. Benennen Sie die hochgeladenen PDF-Dateien nach dem Schema: A<Aufgabe>-<Person1>-<Person2>-<Person3>.pdf, bspw. A3-Musterfrau-Mustermann-Beispiel.pdf für Aufgabe 3 von Lisa Musterfrau, Peter Mustermann und Karla Beispiel. Die Auflistung der Namen kann in beliebiger Reihenfolge erfolgen. Beachten Sie die Informationen im Moodle-Kurs <https://hu.berlin/sds24>.

Aufgabe 1 (Hypothesentests)

2 + 4 + 4 = 10 Punkte

Ein Softwareunternehmen plant, eine neue Version ihrer KI-basierten App zu veröffentlichen. Bevor es die überarbeitete Version in vollem Umfang für alle Nutzer freischaltet, soll ein A/B-Test durchgeführt werden, um zu prüfen, ob die durchschnittliche Reaktionszeit der App sich in der neuen Version verschlechtert hat, z.B. aufgrund höheren Rechenaufwands.

Zu diesem Zweck werden die Nutzer zufällig in zwei Gruppen eingeteilt, die unterschiedliche Versionen der App nutzen: Gruppe A nutzt die neue Version (Stichprobe A), während Gruppe B die bisherige Version verwendet (Stichprobe B). Die nachfolgenden durchschnittlichen Reaktionszeiten der App (in Sekunden) wurden jeweils innerhalb eines Tages gemessen. *Hinweis:* Man geht von einer Normalverteilung der Reaktionszeiten aus.

Stichprobe A: 0.24, 0.22, 0.20, 0.25

Stichprobe B: 0.2, 0.19, 0.22, 0.18

Das Unternehmen möchte jetzt feststellen, ob die neue App-Version tatsächlich zu einer signifikant höheren Reaktionszeit führt. Hierzu sollen folgende Aufgaben gelöst werden:

- Formulieren Sie die Nullhypothese H_0 und die Alternativhypothese H_A für das beschriebene Szenario. *Hinweis:* Gehen Sie von einem einseitigen Test aus.
- Implementieren Sie einen einseitigen Zweistichproben-T-Test, der ermittelt, ob es mit der neuen App-Version (Stichprobe A) tatsächlich zu einer signifikant höheren Reaktionszeit kommt als mit der vorherigen App-Version (Stichprobe B). Führen Sie den T-Test für die beiden Stichproben mit einem Signifikanzniveau von 5% durch. Kann die Nullhypothese zugunsten der Alternativhypothese verworfen werden? Interpretieren Sie kurz und stichhaltig den errechneten p-Wert in Bezug auf das vorliegende Szenario. Nutzen Sie ausschließlich die Open-Source-Bibliotheken NumPy und die Funktion `stats.t.cdf` aus SciPy. *Hinweis:* Nutzen Sie zur Berechnung der Stichprobenvarianz die Bessel-Korrektur.

- (c) Führen Sie einen Mann-Whitney-U-Test durch, um ohne Annahmen über die Verteilungen beider Stichproben auf eine erhöhte Reaktionszeit zu prüfen. Führen Sie den Test für die beiden Stichproben mit einem Signifikanzniveau von 5% durch. Nutzen Sie die Open-Source-Bibliothek NumPy und die Klasse `mannwhitneyu` aus SciPy.

Kann die Nullhypothese zugunsten der Alternativhypothese verworfen werden? Argumentieren Sie kurz und stichhaltig hinsichtlich etwaiger statistischer Annahmen, weshalb bei widersprüchlichen Resultaten entweder (1) dem Ergebnis des T-Tests, oder (2) dem Ergebniss des Mann-Whitney-U-Test mehr Glauben zu schenken ist.

Nutzen Sie zur Beantwortung der Fragen aus (a), (b) und (c) das in Moodle bereitgestellte Python-Template "aufgabe3.py". Im Template gibt es jeweils eine Funktion je Teilaufgabe, in welcher die Aufgabe programmatisch gelöst werden soll. Die erwarteten Datentypen der Rückgabe entnehmen Sie der Beschreibung zu Beginn jeder Funktion.

Verändern Sie **nicht** die Signatur der definierten Funktionen. Darüber hinaus können Sie beliebige Hilfsfunktionen definieren. Sie dürfen lediglich die Open-Source-Bibliotheken NumPy und `stats.t.cdf` sowie `mannwhitneyu` aus SciPy nutzen.

Aufgabe 2 (Entscheidungsbäume)**8 + 2 = 10 Punkte**

Ein Finanzdienstleistungsunternehmen will einen schnellen Entscheidungsbaum zur Prüfung der Kreditwürdigkeit ihrer Kunden und Kundinnen entwickeln. Hierfür sollen drei Attribute (Einkommen, Alter sowie Schulden) benutzt werden, um die Kreditwürdigkeit als binäre Klassifikation (ja oder nein) zu entscheiden. Es wurde der folgende Datensatz nach aufwendiger Prüfung erhoben:

Einkommen (in Tsd. Euro)	Alter	Schulden	Kreditwürdigkeit
30	jung	nein	ja
50	mittel	nein	ja
30	alt	ja	nein
50	mittel	nein	ja
30	jung	ja	nein
50	alt	nein	ja
50	mittel	ja	nein
50	jung	nein	ja
30	alt	nein	nein
70	mittel	nein	ja

- (a) Erstellen Sie einen binären Entscheidungsbaum mithilfe des ID3 Algorithmus (VL Folie 25) sowie den Splitting-Strategien (VL Folie 27). Geben Sie Ihre Rechenwege, Zwischenergebnisse und stichhaltige Beschreibungen an, wie Sie Attribute auswählen und splitten. Skizzieren Sie den finalen Entscheidungsbaum. *Hinweis:* Wenn die Entropie unendlich groß wird, ersetzen sie diese mit 0 in Ihren Berechnungen.
- (b) Leiten Sie aus dem Entscheidungsbaum aus Aufgabe 2 (a) textuelle Regeln für die Vertriebsabteilung des Finanzdienstleistungsunternehmens ab, die beschreiben, wann ein Kunde oder eine Kundin kreditwürdig ist. Halten Sie die Regeln so kurz wie möglich.

Aufgabe 3 (Evaluationsmaße)**3 + 5 + 2 = 10 Punkte**

Das Finanzdienstleistungsunternehmen aus Aufgabe 2 entwickelt ebenfalls einen Klassifikator zur Identifizierung von betrügerischen Kreditkartenkäufen auf Grundlage von Transaktionsdaten und Kundenverhalten. Im Rahmen einer ersten Studie hat das Team zehn Transaktionen analysiert und mit dem Klassifikator Vorhersagen zum Kaufverhalten getroffen. Die tatsächlichen Ergebnisse wurden anschließend durch eine aufwendige Untersuchung ausgewertet und sind in der folgenden Tabelle dargestellt:

Transaktion	Wahres Ergebnis	Vorhergesagtes Ergebnis
1	Echt	Echt
2	Echt	Echt
3	Echt	Missbrauch
4	Echt	Missbrauch
5	Missbrauch	Missbrauch
6	Echt	Echt
7	Echt	Echt
8	Echt	Missbrauch
9	Echt	Echt
10	Missbrauch	Missbrauch

- (a) Erstellen sie jeweils eine Wahrheitsmatrix (auch Konfusionsmatrix genannt) für die Klasse “Echt“ bzw. “Missbrauch“. Berechnen Sie damit die Accuracy der Klassifikation. Geben Sie Ihren Rechenweg an.
- (b) Berechnen Sie für jede Klasse die Evaluationsmaße Precision und Recall einzeln sowie den Macro-F1-Score gesamt für die Klassifikationsergebnisse. Geben Sie Ihren Rechenweg an.
- (c) Angenommen, das Team entscheidet, es ist besonders wichtig, dass wenn eine Transaktion als echt eingestuft wird, dies auch stimmt; selbst wenn das bedeutet, dass einige echte Transaktionen als missbräuchlich eingestuft werden. Welches Evaluationsmaß aus Teilaufgabe 3 (b) wäre in diesem Kontext am aussagekräftigsten? Begründen Sie Ihre Antwort.