



KLAUSUR ZUM BACHELORMODUL  
„PROBEKLAUSUR VORLESUNG SYMBOLISCHE PROGRAMMIERSPRACHE“  
PROBEKLAUSUR,  
DR. BENJAMIN ROTH  
KLAUSUR AM

|                 |   |
|-----------------|---|
| VORNAME:        | <input type="text"/>  |
| NACHNAME:       | <input type="text"/>  |
| MATRIKELNUMMER: | <input type="text"/>  |
| STUDIENGANG:    | <input type="checkbox"/> B.Sc. Computerlinguistik, <input type="checkbox"/> B.Sc. Informatik, <input type="checkbox"/> Magister |
|                 | <input type="checkbox"/> anderer:   |

Die Klausur besteht aus **7 Aufgaben**. Die Punktzahl ist bei jeder Aufgabe angegeben. Die Bearbeitungsdauer beträgt **45 Minuten**. Bitte überprüfen Sie, ob Sie ein vollständiges Exemplar erhalten haben. Tragen Sie die Lösungen in den dafür vorgesehenen Raum im Anschluss an jede Aufgabe ein. Falls der Platz für Ihre Lösung nicht ausreicht, benutzen Sie bitte **nur** die ausgeteilten Zusatzblätter! Verwenden Sie einen dokumentenechten Kugelschreiber oder Füller, **keine** Bleistifte. Es sind **keine Hilfsmittel** zugelassen. Geben Sie Programmcode immer in **Python** an. **Sie können Fragen auf Englisch bearbeiten**. Bitte tragen Sie **zuerst**, d.h., bevor Sie die Aufgaben lösen, auf **allen** Seiten Ihren Namen ein und füllen Sie die Titelseite aus.

| Aufgabe                            | mögliche Punkte | erreichte Punkte |
|------------------------------------|-----------------|------------------|
| 1. Evaluierung von Klassifikatoren | 4               |                  |
| 2. Naive Bayes Klassifikator       | 6               |                  |
| 3. Objektorientierung              | 5               |                  |
| 4. Klassifikation und Clustering   | 3               |                  |
| 5. NLTK and Lexical Information    | 6               |                  |
| 6. WordNet                         | 3               |                  |
| 7. POS Tagging                     | 3               |                  |
| Summe                              | 30              |                  |
| Note                               |                 |                  |

**Einwilligungserklärung (optional)**

Hiermit stimme ich einer Veröffentlichung meines Klausurergebnisses in der Veranstaltung „PROBEKLAUSUR Vorlesung Symbolische Programmiersprache“ vom unter Verwendung meiner Matrikelnummer im Internet zu.

Datum: \_\_\_\_\_ Unterschrift: \_\_\_\_\_

NAME:

---

## Aufgabe 1 **Evaluierung von Klassifikatoren**

Gegeben ein binärer Klassifikator für die Klassen True und False.

- (a) Geben Sie Formel zur Berechnung von Precision, Recall und F1-Measure an (für Klasse True). Erklären Sie alle verwendeten Variablen.

- (b) Geben Sie Formel zur Berechnung der Accuracy an. Erklären Sie alle verwendeten Variablen.

4 PUNKTE

NAME:

---

## Aufgabe 2 Naive Bayes Klassifikator

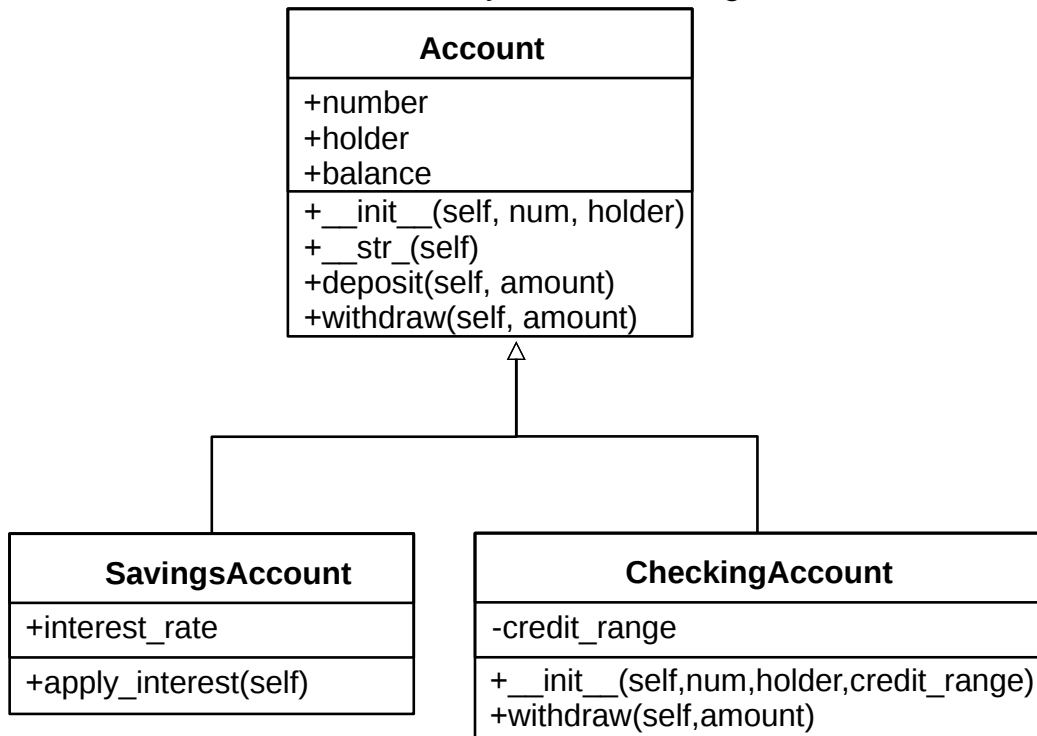
- (a) Wie lautet bei einem Binären Textklassifikator (Naive Bayes) das Entscheidungskriterium?
- (b) Was sind die sogenannten Log-Odds, und wie ergeben sie sich aus dem Entscheidungskriterium?
- (c) Erklären Sie das Konzept der **bedingten Unabhängigkeitsannahme** am Beispiel der Berechnung der Wahrscheinlichkeit für  $P(\text{Text}|\text{Label})$ .

6 PUNKTE

NAME: \_\_\_\_\_

### Aufgabe 3 Objektorientierung

Gegeben die UML-Klassenhierarchie unten. `annesAccount` ist ein Instanzobjekt von `SavingsAccount`, und `stefansAccount` ist ein Instanzobjekt von `CheckingAccount`.



Geben Sie für die folgenden Aufrufe jeweils an, ob sie definiert sind, und wenn ja, in welcher Klasse die aufgerufene Methode definiert wurde:

- `SavingsAccount(2, "Anne")`
- `CheckingAccount(5, "Stefan", 300)`
- `str(annesAccount)`
- `str(stefansAccount)`
- `annesAccount.deposit(200)`
- `stefansAccount.deposit(500)`
- `annesAccount.withdraw(300)`
- `stefansAccount.withdraw(300)`
- `annesAccount.apply_interest()`
- `stefansAccount.apply_interest()`

5 PUNKTE

NAME:

---

## Aufgabe 4 Klassifikation und Clustering

- (a) Was ist der Unterschied zwischen überwachtem Lernen (supervised) und unüberwachtem Lernen (unsupervised) (Antwort in einem Satz)?

(1 Punkt)

(2 Punkte)

- (b) Welche der folgenden Aussagen sind wahr?

- (a) K-means ist ein unüberwachter Algorithmus.
- (b) K-nearest neighbors ist ein überwachter Algorithmus
- (c) Naive Bayes ist ein unüberwachter Algorithmus.
- (d) Lesk ist ein überwachter Algorithmus

3 PUNKTE

NAME:

## Aufgabe 5 NLTK and Lexical Information

(a) Definieren Sie folgende Begriffe:

(2 Punkte)

(a) Token

(b) Type

(c) Collocation

(d) Bigram

(b) Nennen Sie zwei Beispiele der Lexika, die es im NLTK gibt. Für welche NLP Aufgaben können Sie diese Lexika verwenden?

(2 Punkte)

(c) Gegeben folgender Programmcode:

(2 Punkte)

```
1 import nltk
2
3 text = nltk.corpus.genesis.words("english-kjv.txt")
4 bigrams = nltk.bigrams(text)
5 cfd = nltk.ConditionalFreqDist(bigrams)
6
7 print(list(cfd["living"]))
8 >>> ['creature', 'thing', 'soul', '.', 'substance', ',']
9
10 print(list(cfd["living"].values()))
11 >>> [7, 4, 1, 1, 2, 1]
12
13 result = cfd["living"].max()
```

Was wird in der Zeile 13 berechnet?

Was ist der Inhalt der Variable result?

6 PUNKTE

NAME:

---

## Aufgabe 6 WordNet

(a) Erklären Sie kurz die Idee des Lesk-Algorithmus.

(1 Punkt)

(b) Die unten angegebene Tabelle zeigt 2 Bedeutungen von dem Wort "bank". In welcher Bedeutung wird dieses Wort laut Lesk-Algorithmus im Satz "Where do you bank in this town?" benutzt? Begründen Sie Ihre Antwort.

(2 Punkte)

| Sense                  | Definition   |
|------------------------|--|
| Synset("bank.v.03")    | do business with a bank or keep an account at a bank |
| Synset('deposit.v.02') | put into a bank account                              |

3 PUNKTE

NAME:

---

## Aufgabe 7 POS Tagging

Gegeben die Hypothese: Ein Satz endet niemals mit einer Präposition (preposition).  
Beschreiben Sie, wie Sie diese Hypothese mit Hilfe von NLTK verifizieren können.

3 PUNKTE