**Lernen eines Naive-Bayes-Klassifikators für Restaurant-Bewertungen in Tripadvisor**

|  |  |
| --- | --- |
| **Name** | Jannik Bikowski, Till Emme, Dominik Ramsauer |
| **E-Mail** | [Jannik.Bikowski@stud.uni-regensburg.de](mailto:Jannik.Bikowski@stud.uni-regensburg.de), [Till.Emme@stud.uni-regensburg.de](mailto:Till.Emme@stud.uni-regensburg.de),  [Dominik.Ramsauer@stud.uni-regensburg.de](mailto:Dominik.Ramsauer@stud.uni-regensburg.de) |
| **Fachsemester** | 5, 5, 5 |
| **Matrikelnummer** | 1768542, 1750770, 1737690 |
| **Thema:** | „Lernen eines Naive-Bayes-Klassifikators für Restaurant-Bewertungen in Tripadvisor“ |
| **Kurs** | Informationssysteme (LV-Nr. 36670b) |
| **Semester** | SS 2017 |
| **Kursleiter** | Prof. Dr. Bernd Ludwig |

Inhaltsverzeichnis

[1. Ziele der Arbeit 1](#_Toc491163585)

[2. Theoretischer Hintergrund 1](#_Toc491163586)

[3. Theoretischer Hintergrund 1](#_Toc491163587)

[Literaturverzeichnis 2](#_Toc491163588)

Hisejfsiejfosikefjalsejkfalsßeifjalskfjlkajsdddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddd

# Ziele der Arbeit

Das Ziel der Arbeit ist die Implementierung eines Naive-Bayes-Klassifikators, der anhand vorher gesammelter Daten von Restaurant-Bewertungen die Gesamtbewertung eines Restaurants vorhersagt. Hierfür wurden Bewertungen von 100 Regensburger Restaurants der Touristikwebsite „TripAdvisor“ gesammelt. Es wurden dafür für jedes Restaurant die Teilbewertungen für Service, Küche, Qualität und Einrichtung, sowie die Gesamtbewertung als Datengrundlage benutzt.

In dieser Arbeit soll zunächst der theoretische Hintergrund beleuchtet werden. Im Anschluss wird näher auf die praktische Implementierung des Klassifikators eingegangen. Der implementierte Algorithmus soll dann mithilfe von Optimierungsversuchen verbessert und die Accuracy mit dem Machine-Learning-Tool „Weka“ verglichen werden. Die Ergebnisse werden abschließend im Fazit präsentiert.

# Theoretischer Hintergrund

In der modernen Gesellschaft spielen Informationssysteme eine immer größere Rolle. Besonders in der angewandten Informatik kommen diese zum Einsatz, zum Beispiel in der automatischen Sprachverarbeitung aber auch in der Bioinformatik, der Krebsforschung oder Psychologie. Ein wesentlicher Bestandteil solcher Informationssysteme ist die Lösung von Klassifizierungsproblemen. Um Informationen zu verarbeiten müssen Daten in festgelegte Kategorien eingeordnet werden. Zur Lösung der genannten Probleme können zahlreiche Ansätze verfolgt werden, z.B. Maximum-Entropie -Modell oder Naives-Bayes-Klassifikation. Letzterer wird im Folgenden genauer erläutert.

Der Naive-Bayes-Klassifikator versucht jedes einzelne Objekt derjenigen Klasse zuzuordnen, zu der es am wahrscheinlichsten gehört. Für ein gegebenes Objekt *o* soll für jede Klasse *c* die bedingte Wahrscheinlichkeit der Klasse *P(c|o)* berechnet werden. Diejenige Klasse mit der höchsten Wahrscheinlichkeit wird die vom Klassifikator vorhergesagte Klasse.

Daraus ergibt sich folgende Formel:

C:\Users\Dominik-PC\AppData\Local\Microsoft\Windows\INetCache\Content.Word\1.png

Mit dem Bayes’schen Theorem folgt:

C:\Users\Dominik-PC\AppData\Local\Microsoft\Windows\INetCache\Content.Word\2.png

*P(o)* kann weggelassen werden, weil es für alle Klassen gleich ist.

C:\Users\Dominik-PC\AppData\Local\Microsoft\Windows\INetCache\Content.Word\3.png

1. Modellannahme: Die gewählten Merkmale sind angemessen, um das Objekt zu repräsentieren. *x1, …, xn* bilden die Merkmale des Objekts.

C:\Users\Dominik-PC\AppData\Local\Microsoft\Windows\INetCache\Content.Word\4.png

2. Modellannahme: Die gewählten Merkmale sind voneinander unabhängig.

C:\Users\Dominik-PC\AppData\Local\Microsoft\Windows\INetCache\Content.Word\5.png

# Implementierung

# Optimierung und Vergleich mit Weka

# Fazit

Literaturverzeichnis

Gräf u. a. 2011  
D. Gräf – S. Großmann – P. Klimczak – H. Krah – M. Wagner, Filmsemiotik. Eine Einführung in die Analyse audiovisueller Formate, Schriften zur Kultur- und Mediensemiotik 3 (Marburg 2011)

Hickethier 2012  
K. Hickethier, Film- und Fernsehanalyse 5(Stuttgart 2012)

Kuchenbuch 2005  
T. Kuchenbuch, Filmanalyse. Theorien, Methoden, Kritik, UTB 2648 2(Wien 2005)

Mikos 2015  
L. Mikos, Film- und Fernsehanalyse, UTB M 2([s.l.] 2015)

Wright 2016  
R. Wright, House of Cards. Chapter 42 (2016)