

# Textklassifikation

Sharon Dickson Igwe & Jessica Kenner

# Gliederung

- Was ist Textklassifikation?
- Wofür braucht man sie?
- Wie funktioniert sie?
- Welche Schwierigkeiten gibt es?
- Beispiel

# Was ist Textklassifikation?

- Eine maschinelle Lernmethode ➡ große Textmengen vordefinierten Kategorien zuordnen
- Kann große Textmengen analysieren, strukturieren und organisieren/kategorisieren

# Wofür braucht man sie?

- Ein großer Teil aller Daten/Informationen sind unstrukturiert
- Digitale Information werden immer mehr
  - Dadurch entsteht der Bedarf für Algorithmen zum
  - ➔ Sortieren, Filtern und Kategorisieren dieser Information
- Besonders nützlich in Bereichen wie Echtzeit-Analysen, Vereinheitlichung von Kriterien und Skalierbarkeit

# Wie funktioniert sie?

- Kann auf unterschiedliche Weise umgesetzt werden
  - ➔ Einerseits **automatisiert** und andererseits **manuell**
- Um riesige Textdateien schneller zu analysieren, ist es sinnvoll, auf maschinelle Lernalgorithmen zurückzugreifen
  - ➔ Arbeiten nicht nur schneller, sie sind in der Regel auch viel genauer und günstiger in der Handhabung
- Es gibt drei Arten von automatisierter Textklassifikationen
  - ➔ Ruled-based systems, Machine Learning based systems, Hybrid Systems

# Welche Schwierigkeiten gibt es?

## Ruled-based systems

- Durch handgefertigte linguistische Regeln werden Texte in organisierte Gruppen klassifiziert
- Verständlich
- Können leicht modifiziert werden
  
- Sehr zeitaufwendig
- Erfordern Fachwissen
- Teuer

# Welche Schwierigkeiten gibt es?

## Machine Learning based systems

- Keine vordefinierten Regeln
- Textklassifikation lernt von früheren Klassifikationen
- Genauer und flexibler als RBS
  
- Brauchen viele Trainingsdaten
- Kann viel Rechenleistung benötigen

# Welche Schwierigkeiten gibt es?

## Hybrid systems

- Mischform
- Machine Learning wird mit regelbasiertem System kombiniert
- Verwendet, um Ergebnisse zu verbessern

# Beispiel

## Machine Learning based system:

1. Dataset
2. Text classification tool
3. Testing/Improving
4. Deploy

# Beispiel

## 1. Dataset

- Daten sammeln (viele Daten -> verlässlichere Vorhersagen)
- Manuell oder durch Text-Mining

# Beispiel

## 2. Text classification tool

- Gesammelte Daten an Algorithmus füttern
- Textklassifikator erstellen

# Beispiel

## 3. Testing and Improving

- Datensätze werden (nach und nach) ausgewertet, bis ausgelernt
- Ziel: Jeder Datensatz erhöht Genauigkeit

# Beispiel

## 4. Deploy

- System kann eingesetzt werden, wenn es ausgelernt hat und zufriedenstellende Ergebnisse liefert

# Quellen

- [Textklassifikation - Bedeutung und Funktionen - datasolut GmbH](#) (Zuletzt aufgerufen: 25.04.22)
- [Textklassifikationen – Anwendungsbereiche im Internet \(clickworker.de\)](#) (Zuletzt aufgerufen: 25.04.22)
- [Text Classification: What it is And Why it Matters \(monkeylearn.com\)](#) (Zuletzt aufgerufen: 25.04.22)

**Vielen Dank für Ihre  
Aufmerksamkeit**