

Analysewerkzeuge

Customer Feedback analyzer

Aufgabe



<u>Challenge</u>: Steigende Wichtigkeit von Bewertungen für die Entwicklung und Verbesserung von Produkten aller Art

<u>Lösung</u>: Entwicklung eines Modells, welches die Kundenmeinung quantifiziert und eine entsprechende Beantwortung der vorgenommenen Bewertung vornehmen kann.

Erfolgskriterien:

- Modellentwicklung um Kundenbewertungen zu erhalten
- Quantifizierung der besagten Kundenmeinung und entsprechende Antwort modellieren
- Nutzbarkeit in sozialen Medien (Antworten auf Kundenmeinungen)

Geplante Herangehensweise



- Data Scraping, um möglichst viele Einzelbewertungen zu erhalten
- Text Mining/linguistische Methoden zur Bewertung Sentiment-Analyse
- Datenaufbereitung Insights
- Möglichkeiten zur automatisierten Antwort prüfen

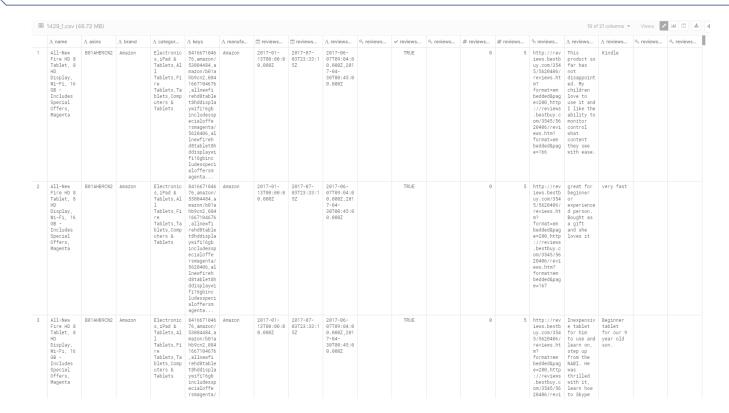
Data-Overview



Produktbewertungen div. In-House-Amazon-Produkte

- Quelle: <u>kaggle.com</u>
- Enthält über 34.000 customer reviews in englischer Sprache
- Robuster Datensatz
- Ausschließlich Amazon-eigene Produkte, wie Kindle, Fire TV Stick, etc.
- Wichtigste Katgeorien:
 - Produkt
 - Sternebewertung
 - Überschrift/Kurzbewertung
 - Textliche Bewertung

Data-Overview







TextBlob

- Library zur Verarbeitung von Text-Daten
- Simple API Basis f
 ür natural language processing (NLP) mit Python
- für Python 2 als auch 3 verfügbar
- Ansätze, wie Erkennen von Wortarten, Extraktion von Substantiven, Stimmungsanalyse und auch Klassifizierungen möglich
- Auswertung von zwei Metriken: polarity & subjectivity

FACHHOCHSCHULE WIENER NEUSTADT Austrian Network for Higher Education

TextBlob

Wesentliche TextBlob Objekte:

- blob.sentences: Anzeige der einzelnen Sätze aus einem Text
- blob.tokens: Anzeige der einzelnen Wörter
- blob.tags: Anzeige der Wortarten
- blob.noun_phrases: Zusammenfassung von zusammenhängenden Wörtern
- blob.sentiment: ermittelt Polarity und Subjectivity

Lemmatization: Stammwort von Wörtern herunterbrechen (bsp. driven – auf drive)



TextBlob

Polarity Score (dt. Gegensätzlichkeit – wie positiv oder negativ ein Wort/Text ist)

- innerhalb der Werte (– 1.0) und (1.0)
- Auf Wort-für-Wort-basis, jedes Wort hat einen eigenen Wert
- Wert X (Bewertung aller Worte) geteilt durch die Anzahl der Worte = Polarity mean, Wert
- Wort "not" oder "no" vor einem positiv konnotierten Wort wird mit x -0.3 bzw. x -0.5 multipliziert/gewichtet

Subjectivity Score (dt. Subjektivität)

- innerhalb der Werte 0.0 und 1.0
- 0.0 ist sehr objektiv, 1.0 ist sehr subjektiv



Natural Language Toolkit (NLTK)

- NLP-Library mit umfangreichen Packages zu Mashine Learning
- Relevante Module:
 - StopWords: Ziel ist die Ausfilterung von Füll-Wörtern (z.B. Pronomen und Bindewörter, also "l",
 "is")
 - POS-Tag: Tag bestimmt, was das Wort ist (Verb, Adjektiv, Substantiv) gebraucht um zu lemmatizen; Ziel: Topics bilden
 - PorterStemmer: Alternative Art zu Lemmatizen (Nennform); Prefixes und Suffixes werden eliminiert (Wortstamm)
 - **Tokenize**: Review in einzelne Sätze, und Sätze in einzelne Wörter oder zerlegen



Diverse Darstellungs-Pakete

- **Seaborn**: Daten-Visualisierungs Library basierend auf matplotlib, pandas-Integration
- **Ipywidgets/IPython.display**: Für die GUI Button
- Wordcloud



Sonstiges

Sklearn: CountVectorizer Documentmatrix – pro Review oder latent distribution – topic finding (LDA – welche Topics im Text vorkommen können, anhand der Verteilung der Wörter) → Auf Reviews jedoch schwer applizierbar, da zu geringe Sample Size