



**FACHHOCHSCHULE
WIENER NEUSTADT**
Austrian Network for Higher Education

Analysewerkzeuge

Customer Feedback analyzer

Challenge: Steigende Wichtigkeit von Bewertungen für die Entwicklung und Verbesserung von Produkten aller Art

Lösung: 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



**FACHHOCHSCHULE
WIENER NEUSTADT**
Austrian Network for Higher Education

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

Produktbewertungen div. In-House-Amazon-Produkte

- Quelle: [kaggle.com](https://www.kaggle.com)
- Enthält über 34.000 customer reviews in englischer Sprache
- Robuster Datensatz
- Ausschließlich Amazon-eigene Produkte, wie Kindle, Fire TV Stick, etc.
- Wichtigste Kategorien:
 - Produkt
 - Sternebewertung
 - Überschrift/Kurzbewertung
 - Textliche Bewertung

Data-Overview



**FACHHOCHSCHULE
WIENER NEUSTADT**
Austrian Network for Higher Education

1429_1.csv (46.72 MB)

19 of 21 columns Views

	A name	A asins	A brand	A categor...	A keys	A manufa...	reviews...	reviews...	A reviews...	reviews...	reviews...	reviews...	reviews...	reviews...	reviews...	reviews...	reviews...	reviews...		
1	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	B01AHB9CN2	Amazon	Electronics, iPad & Tablets, All Tablets, Fire Tablets, Tablets, Computers & Tablets	841667194676, amazon/53084484, amazon/b01a1b6cnc2, 8841667194676, allnewfirehddtabletdisplaywi16gbinc1udesspecialofferrsmagenta/5620486, al1newfirehddtabletdisplaywi16gbinc1udesspecialoffersmagenta...	Amazon	2017-01-13T00:00:00.000Z	2017-07-03T23:33:15Z	2017-06-07T09:04:00.000Z, 2017-04-07-04-30T00:45:00.000Z		TRUE			0	5	http://reviews.bestbuy.com/3545/5620486/reviews.htm?format=embedded&page=200,http://reviews.bestbuy.com/3545/5620486/reviews.htm?format=embedded&page=166	This product so far has not disappointed. My children love to use it and I like the ability to monitor control what content they see with ease.	Kindle		
2	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	B01AHB9CN2	Amazon	Electronics, iPad & Tablets, All Tablets, Fire Tablets, Tablets, Computers & Tablets	841667194676, amazon/53084484, amazon/b01a1b6cnc2, 8841667194676, allnewfirehddtabletdisplaywi16gbinc1udesspecialofferrsmagenta/5620486, al1newfirehddtabletdisplaywi16gbinc1udesspecialoffersmagenta...	Amazon	2017-01-13T00:00:00.000Z	2017-07-03T23:33:15Z	2017-06-07T09:04:00.000Z, 2017-04-07-04-30T00:45:00.000Z		TRUE			0	5	http://reviews.bestbuy.com/3545/5620486/reviews.htm?format=embedded&page=200,http://reviews.bestbuy.com/3545/5620486/reviews.htm?format=embedded&page=167	great for beginner or experience d person. Bought as a gift and she loves it	very fast		
3	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta	B01AHB9CN2	Amazon	Electronics, iPad & Tablets, All Tablets, Fire Tablets, Tablets, Computers & Tablets	841667194676, amazon/53084484, amazon/b01a1b6cnc2, 8841667194676, allnewfirehddtabletdisplaywi16gbinc1udesspecialofferrsmagenta/5620486, al1newfirehddtabletdisplaywi16gbinc1udesspecialoffersmagenta...	Amazon	2017-01-13T00:00:00.000Z	2017-07-03T23:33:15Z	2017-06-07T09:04:00.000Z, 2017-04-07-04-30T00:45:00.000Z		TRUE			0	5	http://reviews.bestbuy.com/3545/5620486/reviews.htm?format=embedded&page=200,http://reviews.bestbuy.com/3545/5620486/reviews.htm?format=embedded&page=168	Inexpensive tablet for him to use and learn on, step up from the NABI. He was thrilled with it, learn how to Skype	Beginner tablet for our 9 year old son.		

TextBlob

- Library zur Verarbeitung von Text-Daten
- Simple API Basis für 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

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)

Pakete

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 Machine Learning
- Relevante Module:
 - **StopWords:** Ziel ist die Ausfilterung von Füll-Wörtern (z.B. Pronomen und Bindewörter, also „I“, „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