

Einleitung und Zielsetzung

In der heutigen digitalen Ära spielt die Verarbeitung großer Datenmengen eine zentrale Rolle in der Unternehmensanalyse und Entscheidungsfindung. Insbesondere im Finanzsektor können tiefere Einblicke in Unternehmensberichte wertvolle Informationen über die zukünftige Entwicklung eines Unternehmens liefern. Diese Arbeit befasst sich mit der Sentimentanalyse von 10-K Berichten der Nasdaq100-Unternehmen, um die emotionalen Tendenzen in den Berichten zu bewerten und deren potenziellen Einfluss auf die Aktienkurse zu untersuchen. Diese Analysen sollen aufzeigen, ob es signifikante Unterschiede im Sentiment zwischen den verschiedenen Items und/oder den verschiedenen Branchen gibt, und ob diese Veränderungen einen Einfluss auf die Aktienkursentwicklung haben.



Abb. 1: Die Abbildung zeigt die Verteilung der Sentimentscores pro Sektor für die Items 1, 1A, 7, 7A und den Gesamtsentimentscore (N=93; Manufacturing: N=42, Services: N=30, Transportation & Others: N=10, Retail Trade: N=10, Mining: N=1).

Quellen:

- Araci, D. (2019, August 27). *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models*. <http://arxiv.org/pdf/1908.10063>
- Huang, A., Wang, H., & Yang, Y. (2020). FinBERT—A Deep Learning Approach to Extracting Textual Information. *SSRN Electronic Journal*. Advance online publication. <https://doi.org/10.2139/ssrn.3910214>
- Loughran, T., & McDonald, B. (2020). Textual Analysis in Finance. *Annual Review of Financial Economics*, 12(1), 357–375. <https://doi.org/10.1146/annurev-financial-012820-032249>
- SEC. *API Documentation: SEC Filings API*. <https://sec-api.io/docs>

Methodik

Datenbeschaffung:

Die Daten wurden mittels der SEC API beschafft. Teil der Analyse waren die 10-K Berichte aktueller NASDAQ100 Unternehmen, die im Zeitraum von 2000 bis 2023 veröffentlicht wurden. Der Fokus lag dabei auf den Items 1, 1A, 7 und 7A dieser Berichte.

Sentimentanalyse:

Zur Sentimentanalyse wurde das Transformer-Modell FinBERT (Modell: yiyanghkust/finbert-tone) verwendet (Python 3.10.8). Die Sentimentscores wurden satzweise berechnet, wobei für jeden Bericht und jedes Item positive, negative und neutrale Scores ermittelt wurden. Zusätzlich wurde ein Gesamtscore berechnet, der die Differenz zwischen positiven und negativen Scores darstellt.

Die Unternehmen wurden basierend auf ihren Standard Industrial Classification (SIC) Codes in verschiedene Sektoren unterteilt. Um die Unterschiede im Sentiment zwischen den Sektoren und den Items zu analysieren, wurde eine Varianzanalyse (ANOVA) durchgeführt.

Korrelation mit Aktienkursen:

Die prozentuale Veränderung des Aktienkurses wurde vom Berichtstag bis ein Jahr danach berechnet. Um eine sinnvolle Analyse der Daten zu gewährleisten, wurde die Veränderung des Sentimentscores eines Jahres mit der Aktienkursveränderung des darauffolgenden Jahres korreliert. Zur Untersuchung potenzieller Zusammenhänge zwischen den Veränderungen der Sentimentscores und den Aktienkursentwicklungen wurde eine Pearson-Korrelationsanalyse durchgeführt.

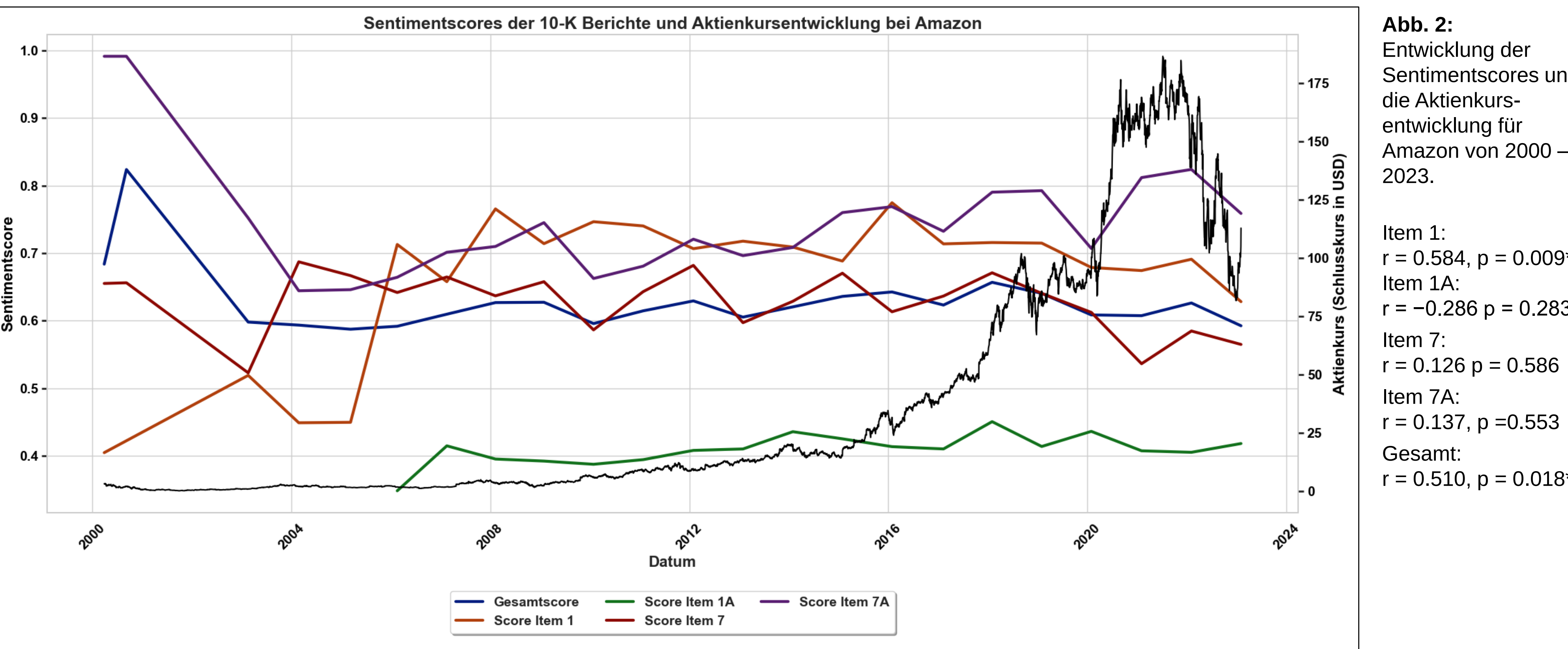


Abb. 2: Entwicklung der Sentimentscores und die Aktienkursentwicklung für Amazon von 2000 – 2023.

Item 1:
 $r = 0.584$, $p = 0.009^{**}$
Item 1A:
 $r = -0.286$, $p = 0.283$
Item 7:
 $r = 0.126$, $p = 0.586$
Item 7A:
 $r = 0.137$, $p = 0.553$
Gesamt:
 $r = 0.510$, $p = 0.018^{*}$

Hypothesen: Ergebnisse und Interpretation

H1: Es gibt signifikante Unterschiede im Sentiment zwischen den verschiedenen Abschnitten (Items) der 10-K Berichte.

Die Ergebnisse der ANOVA zeigten signifikante Unterschiede zwischen den Sentimentscores der verschiedenen Items, $F(3,6128) = 2544.455$, $p < 0.0001$ und damit dass bestimmte Abschnitte tendenziell positiver oder negativer formuliert sind als andere, was möglicherweise auf unterschiedliche Schwerpunkte oder Risiken hinweist, die in diesen Abschnitten behandelt werden.

H2: Das Sentiment in den 10-K Berichten variiert signifikant zwischen den unterschiedlichen SIC Sektoren.

Die ANOVA-Analyse untersuchte die Unterschiede im Sentiment zwischen verschiedenen SIC-Sektoren für die Items 1, 1A, 7, 7A sowie den Gesamtsentimentscore. Die wichtigsten Ergebnisse sind:

Item	F-Wert	p-Wert
Item 1	$F(3,88) = 9.64$	$p < 0.0001$
Item 1A	$F(3,88) = 4.27$	$p < 0.01$
Item 7	$F(3,88) = 0.92$	$p > 0.43$
Item 7A	$F(3,88) = 3.23$	$p < 0.03$
Gesamt	$F(3,88) = 9.31$	$p < 0.0001$

Diese Ergebnisse zeigen, dass bestimmte Branchen konsistent positivere oder negativere Berichte verfassen, möglicherweise aufgrund von branchenspezifischen Herausforderungen, Chancen oder regulatorischen Bedingungen.

H3: Veränderungen des Sentimentscores eines Unternehmens von einem Jahr auf das nächste korrelieren mit den entsprechenden Veränderungen des Aktienkurses.

Die Korrelationen zwischen den Veränderungen der Sentimentscores und den Veränderungen der Aktienkurse wurden analysiert. Die Ergebnisse zeigten:

Item	Korrelationskoeffizient r	P-Wert
Veränderung Sentiment Item 1	0.0006	0.9830
Veränderung Sentiment Item 1A	0.0066	0.8224
Veränderung Sentiment Item 7	-0.0275	0.2907
Veränderung Sentiment Item 7A	-0.0388	0.1373
Veränderung Gesamtsentimentscore	0.0092	0.7211

Die Korrelationen zwischen den Veränderungen der Sentimentscores und den Veränderungen der Aktienkurse sind sehr schwach und statistisch nicht signifikant ($p > 0.05$). Dies deutet darauf hin, dass auf aggregierter Ebene keine starken Zusammenhänge zwischen den Sentimentveränderungen in den 10-K Berichten und den nachfolgenden Aktienkursveränderungen bestehen. Es wurde jedoch festgestellt, dass bei einzelnen Unternehmen wie Amazon signifikante Korrelationen existieren, was darauf hinweist, dass eine unternehmensspezifische Analyse von Vorteil sein könnte.