

Analyse von Twitter-Posts zu Aktiengesellschaften mit Hilfe von Methoden der Textanalyse

Oliver Dankwart,
Forschungsgruppe Data Science, Institut für Digitalen Wandel (IDW), Master Digital Business

Einführung

- Soziale Netzwerke als essenzielle Treiber des modernen Internets und der Kommunikation.
- Einfluss von Twitter-Beiträgen auf Unternehmen in Form von möglichen Kapitalverlusten und Branding-Einbußen.
- Monitoring der sozialen Netzwerke durch automatische Analysen unter Anwendung von Methoden der Textanalyse wichtig für Unternehmen, um frühzeitig negative Beiträge zu moderieren/beiträge zu moderierentive Beiträge zu moderieren.

Methodik

Datenaufbereitung

- Zusammenführen der Datenquellen (CSV-Dateien), Filterung auf Teildatensatz (Tweets aus 2019, nur Google & Co.) und Bereinigung der übrigen Datenzeilen.
- Vorverarbeitung der Textdaten durch die Entfernung von Sonderzeichen (URLs, #, @-Nutzernamen, etc.), Tokenisierung, Part-of-Speech-Tagging, Lemmatisierung und Erzeugung des Bag of Words-Models.

Topic Detection

- Anwendung des Latent Dirichlet Allocation-Algorithmus (LDA) zur Erkennung der Themen.
- Optimierung des Topic-Modells anhand der Topic Anzahl K und Hyperparametern α und η unter Auswertung quantitativer Kennzahlen (Perplexity, Coherence-Metrik Cv).

Sentiment Analyse

- Lexikon-basierte Sentiment-Analyse mit Valence Aware Dictionary and Sentiment Reasoner (VADER-Algorithmus).
- Automatisches Einordnen der Twitter-Beiträge in Polaritäten positiv, negativ und neutral.

Evaluierung der Ergebnisse

- Grafische Darstellung durch Zusammenführung der gefundenen Themen und den Häufigkeiten der Stimmungsrichtungen pro Thema.
- Qualitative Auswertung durch manuelle Analyse und Nutzung des Analysetools pyLDAvis.

Zielsetzung

- Zweiteilige Analyse eines Twitter-Datensatzes zu Aktiengesellschaften aus der Technologie-Branche (GAFAM).
- Initiale Erkennung häufig genannter Themen durch eine Topic Detection mit folgender Sentiment-Analyse anhand der gefundenen Themen.
- Grafische Aufbereitung mit anschließender Analyse der kombinierten Ergebnisse.

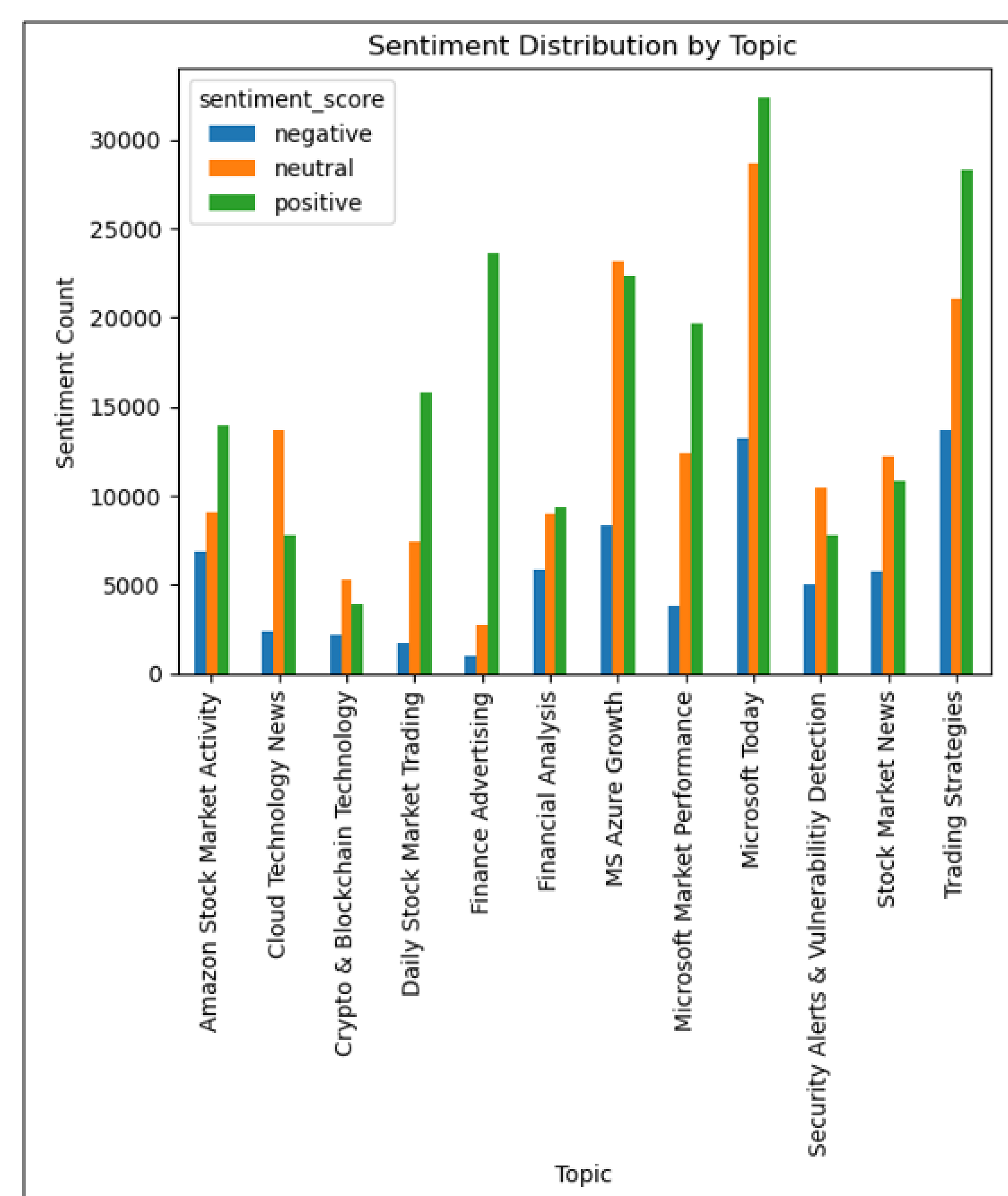
Ergebnis

Ergebnisse der Topic Model-Optimierung

Topic K	α	η	Coherence	Perplexity
6	auto	symmetric	0,42997	-8,12379
6	auto	auto	0,42997	-8,11135
12	symmetric	symmetric	0,41917	-9,32112
12	symmetric	auto	0,41917	-9,31267
12	asymmetric	symmetric	0,41677	-9,34484
12	asymmetric	auto	0,41677	-9,33637
12	auto	symmetric	0,41418	-9,30352
12	auto	auto	0,41418	-9,29516
6	asymmetric	symmetric	0,40660	-8,12314
6	asymmetric	auto	0,40660	-8,11032

- Topic Anzahl 12 mit symmetrischem α und η führt zu bester Coherence/Perplexity-Kombination

Endergebnis Topic Detection & Sentiment Analyse



- Ergebnisse im Diagramm dienen der weiteren Praxisanwendung, bspw. Entscheidungsfindung in Unternehmen oder bei Finanzmarktakteuren

Zusammenfassung

- Optimierte Topic Detection erkennt relevante Themen in Social Media-Datensätzen.
- Sentiment-Analyse verleiht Analyse eine weitere Dimension (Stimmung des Themas).
- Vielfältige Anwendungsmöglichkeiten der Ergebnisse bei Aktiengesellschaften, in der Finanzbranche selbst oder für Kleinanleger.

Ergebnisse Topic-Modell K=12

Topicname	Top-Wörter pro Topic
Microsoft Azure Growth	earnings, report, growth, year, company, azure, business, q, revenue, ai
Security Alerts & Vulnerability Detection	alert, cap, security, windows, pt, insider, infosec, dowjones, push, netsec
Stock Market News	stock, stockmarket, news, nasdaq, holding, stocks, invest, market, money, daytrading
Crypto & Blockchain Technology	move, volume, profit, bitcoin, return, average, area, blockchain, odds, partner
Amazon Stock Market Activity	amazon, gain, short, corporation, contract, trade, market, deal, post, stocks
Daily Stock Market Trading	trading, option, day, trade, friday, free, try, breakout, trades, sign
Cloud Technology News	cloud, tech, new, range, rating, min, huge, monday, calls, data
Financial Analysis	price, target, sale, interest, pay, oversold, holder, need, change, partnership
Microsoft Market Performance	microsoft, stock, position, management, stake, value, million, market, declined, share
Finance Advertising	get, make, use, yesterday, join, com, robinhood, sony, link, please
Trading Strategies	buy, call, go, sell, 's, time, year, say, think, take
Microsoft Today	today, week, day, chart, msft, see, look, hold, put, break

Kontakt

Prof. Dr. Wolfram Höpken
Leiter IDW / Forschungsgruppe Data Science
+49 751 501 9764
wolfram.hoepken@rwu.de

Literatur

1. D. M. Blei, "Probabilistic Topic Models," Commun. ACM, Jg. 55, Nr. 4, S. 77-84, Apr. 2012, issn: 0001-0782. doi: 10.1145/2133806.2133826. Adresse: <https://doi.org/10.1145/2133806.2133826>.
2. R. Churchill und L. Singh, "The Evolution of Topic Modeling," S. 1-35, Nov. 2022. doi: 10.1145/3507900. Adresse: <https://doi.org/10.1145/3507900>
3. A. Giachanou und F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," Bd. 49, Association for Computing Machinery, 2016. doi:10.1145/2938640. Adresse: <https://doi.org/10.1145/2938640>.