

Analyse der prädiktiven Fähigkeit von Twitter-Sentiments auf die Entwicklung des Börsenkurses von Technologieunternehmen

Flurina Fiona Baumann, Nadine Belinda Brunner
und Kim Oliver Tokarski

Zusammenfassung

Die Datenmengen vervielfachen sich in der heutigen Zeit konstant, was zum Begriff Big Data geführt hat. Durch diese Datenmengen entsteht ein neues Potenzial, Fragen zu beantworten. Eine dieser Fragestellungen, welche mithilfe von Big Data untersucht werden kann, ist, inwiefern die Social-Media-Daten die Veränderung von Börsenkursen voraussagen können. Diese Studie untersucht die prädiktive Fähigkeit von Twitter-Nachrichten im Zusammenhang mit einem Technologieunternehmen und dessen Börsenkurs anhand von zwei Anwendungsfällen. Konkret wird anhand der Twitter-Nachrichten mithilfe einer Sentimentanalyse die Stimmung der Twitter-Nutzer mit den Veränderungen des Börsenkurses verglichen. Diese Analyse wird anhand der Technologieunternehmen Facebook und Amazon vorgenommen. In einem ersten Schritt wird untersucht, ob eine Beziehung zwischen den Twitter-Sentiments und dem Börsenkurs besteht. In einem zweiten Schritt, ob die Twitter-Sentiments eine Voraussagekraft für die Veränderung des Börsenkurses haben. Die Auswertung zeigt bei beiden Unternehmen eine positive Korrelation der Twitter-Sentiments und des Börsenkurses auf. Weiter konnte mithilfe der Granger-Analyse eine signifikante Voraussagekraft der Twitter-Sentiments für die Börsenkurse beider Unternehmen ermittelt werden. Die Twitter-Sentiments können die Börsenkurse 13 h voraussagen.

F. F. Baumann (✉)
Bern, Schweiz

N. B. Brunner
MuttENZ, Schweiz

K. O. Tokarski
Departement Wirtschaft, Berner Fachhochschule, Bern, Schweiz
E-Mail: kim.tokarski@bfh.ch

9.1 Ausgangslage und Problemstellung

„Knowledge is Power“ diese Aussage ist in der heutigen Zeit sehr treffend (Mayer-Schönberger und Cukier 2013). Der Gesellschaft und somit auch den Unternehmen stehen immer mehr Daten zur Verfügung (Wrobel, Voss et al. 2015). Das durchschnittliche Unternehmen hatte bereits 2014 circa 427-mal so viele Daten wie jemals in der US Kongressbibliothek gespeichert wurde (Davenport und Paulus 2014). Davon haben 25 % der Daten einen potenziellen Wert für die Gesellschaft und die Unternehmen (Davenport und Paulus 2014).

Seit 2011 hat das Interesse der Gesellschaft an Big Data zugenommen (Ward und Barker 2013). Aufgrund der Allgegenwärtigkeit des Begriffs und somit der Herkunft aus den Gebieten Wissenschaft, Industrie und Medien gibt es laut Ward und Barker (2013) keine klare Definition. Die gängigsten Definitionen beinhalten jedoch die sogenannten vier V's: **Volume, Velocity, Variety und Veracity**. Dabei sind die am häufigsten in der Literatur genannten Eigenschaften von Big Data Volume, Variety und Velocity (Bendler et al. 2018; Dijcks 2013; Dorschel 2015; Gluchowski und Chamoni 2016; IBM 2017; Inmon und Linstedt 2015; Mayer-Schönberger und Cukier 2013; Wrobel et al. 2015). Die Eigenschaft „Veracity“ wurde aufgrund der stetig wachsenden Social-Media-Daten erst später hinzugefügt. Diese Social-Media-Daten sind benutzergeneriert und verlangen deshalb eine weitere Eigenschaft der Begrifflichkeit „Big Data“ (Bendler et al. 2018). Die Eigenschaft „Volume“ sollte dabei nicht als notwendige Eigenschaft angesehen werden. Wird aufgrund einer Spezialisierung eines Themas nur ein geringes Datenvolumen erreicht, handelt es sich dann trotzdem noch um Big Data. Jedoch gilt die Verwaltung von einem hohen Datenvolumen mit strukturierten Daten in einer relationalen Datenbank nicht als Big-Data-Anwendung (Dorschel 2015).

Im Bereich Big Data Analytics gibt es bereits viele Anwendungsmöglichkeiten. Im Speziellen wird Big Data Analytics bereits oft für unternehmerische Fragestellungen verwendet. Beispielsweise kann das Cross-Selling und Up-Selling mithilfe von Reputationsystemen erhöht oder logistische Fragestellungen können anhand von Verkehrs- und Fahrzeugdaten besser beantwortet werden (Davenport und Paulus 2014; EMC und Greenplum 2012; Marr 2016; Dorschel 2015). Es gibt aber auch im gesellschaftlichen Bereich Anwendungsmöglichkeiten. Neben der Möglichkeit, mit Big Data Analytics beispielsweise Ausbrüche von Epidemien besser vorauszusagen, um dementsprechend zu reagieren, kann diese auch in der Planung und Bewirtschaftung der öffentlichen Infrastruktur helfen (King und Hajnal 2014; Manyika et al. 2018). Wie das Beispiel zeigt, wird Big Data Analytics oft zur Voraussage von Ereignissen genutzt.

Gerade im Bereich von prädiktiven Fragestellungen gibt es einen Anwendungsfall, welcher sowohl für Unternehmen wie auch für Privatpersonen genutzt werden kann. Dieser Anwendungsfall steht in Verbindung mit der Theorie des Homo Oeconomicus. Gemäß dieser Theorie strebt jeder Marktteilnehmer u. a. eine Nutzenmaximierung an (Mankiw 2004). Dies betrifft insbesondere auch die Akteure am Aktienmarkt. Bei einer Investition im Aktienmarkt stellt sich seit jeher die Frage, wie sich die entsprechenden Aktien entwickeln und ob sich eine Investition lohnt oder nicht (Scherbaum 2017).

Bisher mussten sich die Investoren mehrheitlich auf die Unternehmensinformationen und Ratingagenturen verlassen, wenn sie die Investitionsentscheidungen trafen (Scherbaum 2017). Der Aktienkurs wird jedoch aus dem Zusammenspiel von Angebot und Nachfrage eines Aktientitels gebildet. Diese Kauf- und Verkaufsentscheidung der Anleger basieren wiederum auf deren Erwartungen, welche unter anderem anhand der gesellschaftlichen Stimmung gebildet werden (Nguyen et al. 2015; Scherbaum 2017). Aus diesem Grund ist es bei der Fragestellung nach der Entwicklung eines Aktienkurses besonders wichtig die Erwartungen der Anleger zu verstehen.

Bereits mehrere Studien haben sich mit der Fragestellung befasst, welche Daten diese Erwartungen der Anleger beziehungsweise die allgemeine Stimmung im Zusammenhang mit dem Unternehmen voraussagen. Dabei haben sich Social-Media-Daten als erfolgreiche Barometer herausgestellt. In der Mehrheit der betreffenden Studien konnte einen Zusammenhang oder sogar eine prädiktive Aussagekraft von Social-Media-Daten für die Aktienkurse festgestellt werden. Die bisher untersuchten Studien beschäftigen sich jedoch mehrheitlich mit spezifischen Fragestellungen in diesem Bereich. Aus diesem Grund ist für das Verständnis der Voraussagekraft von Social-Media-Daten für Börsendaten noch immer ein großes Potenzial vorhanden.

Die vorliegende Studie soll einen weiteren Einblick in die Möglichkeiten von Big Data Analytics im Zusammenhang mit der Voraussage von Aktienkursen geben und dabei aufzeigen, dass eine Analyse von Big Data ohne ein großes Budget möglich ist.

9.2 Stand der Forschung

Bereits seit den späten 1990er-Jahren haben sich Forschende mit dem Zusammenhang beziehungsweise der prädiktiven Fähigkeit von Social-Media-Kanälen und der Veränderung von Börsenkursen beschäftigt. Im folgenden Abschnitt wird der Stand der Forschung anhand von drei Kriterien verglichen:

- Form der Analyse der Social Media Daten
- Unterschied im gewählten Kommunikationskanal
- Art der untersuchten Börsentitel

Dabei wird jeweils untersucht, ob die Studien einen signifikanten Zusammenhang der beiden Datenquellen nachweisen können.

9.2.1 Analyse Social Media

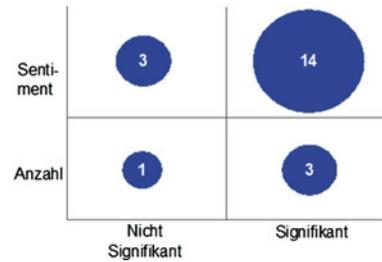
Der Zusammenhang zwischen den Nachrichten auf Social Media oder anderen ähnlichen Kanälen und den Veränderungen an der Börse wurde bereits in verschiedenen Formen erforscht. Als einer der ersten Forscher hat sich Wysocki (1998) mit diesem Thema

auseinandergesetzt. Dabei hat er sich im Gegensatz zu späteren Forschungsarbeiten vor allem auf die Anzahl der Nachrichten beschränkt, ohne dabei den Informationsgehalt der Nachrichten zu berücksichtigen. Neben Wysocki (1998) haben auch noch Reed (2016) und Mao et al. (2012) sich mit der Anzahl der Nachrichten als Variable befasst. Reed (2016) hat herausgefunden, dass die Intensität der Nachrichten mit wirtschaftlichem Inhalt einen signifikant negativen Effekt auf die Börsenkurse hat. Mao, Wei et al. (2012) haben bei der Analyse der Nachrichtenvolumen nur die Nachrichten evaluiert, welche inhaltlich einen Bezug zur Börse hatten. Viele der neueren Forschungsarbeiten haben sich jedoch bei der Analyse der Nachrichten nicht auf die Anzahl der Nachrichten beschränkt, sondern auch auf den emotionalen Inhalt der Nachrichten. Dabei gab es nur wenige Forschungsarbeiten, welche keinen signifikanten Zusammenhang zwischen den öffentlichen Nachrichten und der Veränderung an der Börse aufzeigten. Vu et al. (2012) konnten beispielsweise durch die Reaktion beziehungsweise Stimmung auf Twitter-Nachrichten mit einer Wahrscheinlichkeit von 75 % bis 83 % die Börsenkurse von vier ausgewählten Technologieaktien voraussagen. Dabei haben die Forscher Nachrichten mit bestimmten Schlüsselwörtern zur Marke oder deren Produkten erfasst und als positiv (bullish), negativ (bearish) oder neutral eingestuft. Auch alle anderen Forschungsarbeiten mit einer Sentimentanalyse haben die Klassifizierung der Nachrichten in diese drei Gruppen genutzt. Gesamthaft haben 17 Forschungsarbeiten für die Analyse der Social-Media-Nachrichten die Methode der Sentimentanalyse verwendet. Davon konnten 14 Studien einen signifikanten Zusammenhang zwischen den Sentiments der Social-Media-Nachrichten und der Veränderung der Börsenkurse feststellen (Bollen et al. 2011; Antweiler und Frank 2004; Sprenger und Welp 2010; Oh und Sheng 2011; Liew und Wang 2016; Zhang et al. 2011; Forbergskog und Blom 2014; Vu et al. 2012; Rao und Srivastava 2012; Ranco et al. 2015; Zheludev et al. 2014; Mao et al. 2011; Souza et al. 2015; Nguyen et al. 2015). Drei weitere Forschungsarbeiten konnten keinen signifikanten Zusammenhang ermitteln (Brown 2012; Das und Chen 2007; Das et al. 2005). Vier der bisherigen Forschungsarbeiten haben für die Analyse der Social-Media-Nachrichten keine Sentimentanalyse verwendet, sondern lediglich die Anzahl der Nachrichten gemessen. Davon konnten drei Studien wiederum einen signifikanten Zusammenhang zwischen der Anzahl von Social-Media-Nachrichten und der Veränderung des Börsenkurses feststellen (Wysocki 1998; Reed 2016; Mao et al. 2012, während eine Studie kein signifikantes Ergebnis ausweisen konnte (Tumarkin und Whitelaw 2001). Die Abb. 9.1 zeigt die Analyse der Social-Media-Nachrichten im Zusammenhang mit dem Forschungsergebnis. Wie der Übersicht zu entnehmen ist, dominieren die Forschungsarbeiten, welche einen signifikanten Zusammenhang nachweisen konnten, stark.

9.2.2 Kommunikationskanal

Bei den bisherigen Studien wurde meist Twitter als Bezugsquelle für die Voraussage der Veränderung an der Börse verwendet. Bei den 13 Arbeiten, welche Twitter-Daten

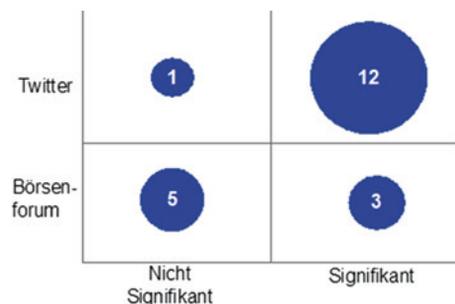
Abb. 9.1 Übersicht
Forschungsergebnisse Analyse
Social Media. (Eigene
Darstellung)



verwendet haben, konnte der Zusammenhang auch in fast allen Fällen nachgewiesen werden (Bollen et al. 2011; Reed 2016; Liew und Wang 2016; Zhang et al. 2011; Forbergskog und Blom 2014; Vu et al. 2012; Rao und Srivastava 2012; Ranco et al. 2015; Zheludev et al. 2014; Mao et al. 2012; Mao et al. 2011; Souza et al. 2015). Nur die Ergebnisse von Brown (2012) waren nicht aussagekräftig. Wobei hier auch zu erwähnen ist, dass es sich nur um ein Testdataset handelte. Als Börsenforum wurden verschiedene Kanäle verwendet. Beispielsweise haben Tumarkin und Whitelaw (2001) und Antweiler und Frank (2004) Nachrichten des Börsenforums RangingBull.com analysiert. Auch MessageBoard (Nguyen et al. 2015) und Stocktwits.com (Oh und Sheng 2011; Sprenger und Welp 2010) wurden als Nachrichtenforum mit Börsenbezug genannt. Gesamthaft haben acht Studien die Social Media Daten von Börsenforen bezogen. Dabei konnten fünf der Studien einen signifikanten Zusammenhang zwischen den Social-Media-Nachrichten und den Veränderungen der Börsenkurse aufweisen (Wysocki 1998; Antweiler und Frank 2004; Oh und Sheng 2011; Nguyen et al. 2015; Sprenger und Welp 2010). Drei weitere Studien konnten keine Signifikanz feststellen (Tumarkin und Whitelaw 2001; Das und Chen 2007; Das et al. 2005).

Neben den börsenbezogenen Nachrichtenforen und Twitter wurde kaum ein anderer Social-Media-Kanal für den Bezug der Stimmungsdaten genannt. Twitter ist, wie auch der Auflistung von erfolgreichen Forschungsarbeiten in der Abb. 9.2 entnommen werden kann, ein gutes Instrument für die Voraussage der Veränderung an der Börse. Microblogging ist aufgrund des hohen Volumens, der Echtzeitanalyse und der Kürze der Nachrichten besonders für eine starke Voraussagekraft der Veränderungen an der Börse

Abb. 9.2 Übersicht
Forschungsergebnisse
Kommunikationskanal.
(Eigene Darstellung)



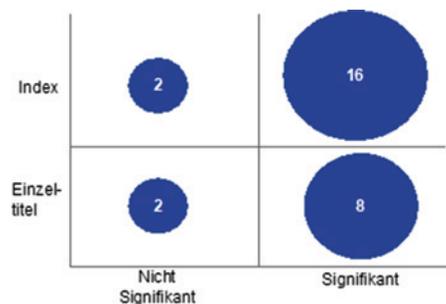
geeignet (Oh und Sheng 2011). Insbesondere ist Twitter ein Informationsinstrument, welches nicht nur die großen Geschichten beinhaltet, sondern auch sehr regionale Informationen abbildet. Diese Informationen sind auf Twitter meist bereits vor der Veröffentlichung auf den offiziellen Nachrichtenportalen vorhanden (Souza, Kolchyna et al. 2015). Twitter hat auch gemäß Souza Kolchyna et al. (2015) nachweislich eine höhere Aussagekraft bezüglich der Veränderungen an der Börse als herkömmliche Nachrichten.

9.2.3 Börsentitel

Für die beobachteten Werte an der Börse wurden bisher sowohl Indizes wie auch einzelne Wertschriftentitel verwendet. Als Indizes wurde oft der Dow Jones Industrial Average (DJIA) untersucht (u. a. Bollen et al. 2011; Reed 2016; Rao und Srivastava 2012). Daneben wurde auch der Index Standard & Poors 500 (S&P 500) oft als Messgröße eingesetzt (u. a. Sprenger und Welppe 2010; Zhang et al. 2011; Mao 2012). Wenn in den Forschungsarbeiten einzelne Aktientitel beobachtet wurden, handelt es sich meistens um eine eher größere Anzahl von Aktien. Dabei wurden oft Aktien von größeren und bekannteren Unternehmen miteinbezogen. Beispielsweise wurde bei Souza et al. (2015) unter anderem die Veränderung der Aktien von Abercrombie & Fitch Co. und Nike Inc. miteinbezogen, aber nie als Fokus der Untersuchung betrachtet. Besonders oft wurden auch Technologieunternehmen in die Beobachtung miteinbezogen (Vu et al. 2012; Das et al. 2005; Tumarkin und Whitelaw 2001; Das und Chen 2007; Liew und Wang 2016; Rao und Srivastava 2012). Der Abb. 9.3 kann entnommen werden, dass eine sich eine größere Anzahl von Untersuchungen mit Indizes als Dateninput für die Veränderung der Börsenkurse beschäftigt hat.

Von insgesamt 18 Studien, welche einen Indizes untersucht haben, konnten auch 16 Studien einen signifikanten Zusammenhang aufweisen (Bollen et al. 2011; Sprenger und Welppe 2010; Oh und Sheng 2011; Reed 2016; Zhang et al. 2011; Forbergskog und Blom 2014; Rao und Srivastava 2012; Ranco et al. 2015; Zheludev et al. 2014; Mao et al. 2012; Mao et al. 2011). Nur zwei weitere Studien konnten keinen signifikanten Zusammenhang ermitteln (Brown 2012; Das und Chen 2007). Zehn weitere Forschungsarbeiten haben

Abb. 9.3 Übersicht Forschungsergebnisse Index. (Eigene Darstellung)



einzelne Wertschriftentitel als Vergleichswert herangezogen. Davon konnten auch acht Studien einen signifikanten Zusammenhang aufweisen (Wysocki 1998; Antweiler und Frank 2004; Liew und Wang 2016; Vu et al. 2012; Rao und Srivastava 2012; Mao et al. 2012; Souza et al. 2015; Nguyen et al. 2015). Zwei weitere Studien konnten wiederum keinen signifikanten Zusammenhang ermitteln (Tumarkin und Whitelaw 2001; Das et al. 2005).

9.2.4 Erkenntnisse aus der Theorie und Forschungslücke

Die bisherigen Forschungsarbeiten zeigen auf, dass es mehrheitlich einen Zusammenhang zwischen den Social-Media-Daten und den Veränderungen der Börsenkurse gibt. Die Sentimentanalyse als Stimmungsbarometer spiegelt sich in den Ergebnissen als besonders aussagekräftig wider. Aufgrund der neuen Möglichkeiten durch die Sentimentanalyse wurden in den letzten Jahren kaum mehr nur die Anzahl der Nachrichten analysiert. Die Sentimentanalyse ist somit ein empfehlenswertes Instrument zur Voraussage der Börsenkurse anhand von Social-Media-Nachrichten. Weiter wurde der Social-Media-Kanal Twitter mehrfach erfolgreich für ähnliche Studien verwendet und eignet sich aufgrund der Eigenschaft des Microbloggings besonders für die Analyse. Bisher haben sich die meisten Forschungsarbeiten mit der Voraussage der Veränderung von ganzen Indizes oder Branchen befasst. Die Forschungsarbeiten, welche sich einzelnen, ausgewählten Aktientitel widmen, haben meist große international tätige Unternehmen gewählt. Dabei wurden nur selten individuelle Aktien von Technologieunternehmen untersucht. Da die Aktien von Technologieunternehmen aufgrund von hohen Renditepotenzialen und andererseits auch hohen Risiken starken Schwankungen unterworfen sind, eignen sich die Aktientitel jedoch besonders für die Analyse (Ciolli 2017; Investopedia 2018).

9.2.5 Hypothesen und theoretisches Modell

Bisher gab es keine Studie, welche sich mit der Frage beschäftigt, ob die Stimmung gegenüber spezifischen Technologieunternehmen auf Social Media einen Zusammenhang mit den Veränderungen der Börsenkurse dieser spezifischen Technologieunternehmen aufweist bzw. eine Voraussagekraft für die Entwicklung der Börsenkurse hat. Die Resultate der Studien zu Technologie-Indizes und einzelner Börsentitel anderer Branchen weisen stark darauf hin, dass die Stimmung auf Social Media einen Zusammenhang und sogar eine Voraussagekraft für die einzelnen Börsentitel von Technologieunternehmen haben. Das Twitter als Social-Media-Plattform für Börsenthemen bereits mehrmals erfolgreich in ähnlichen Studien genutzt wurde, zeigt, dass Twitter eine angemessene Plattform für die vorliegende Studie ist. Aufgrund der oben

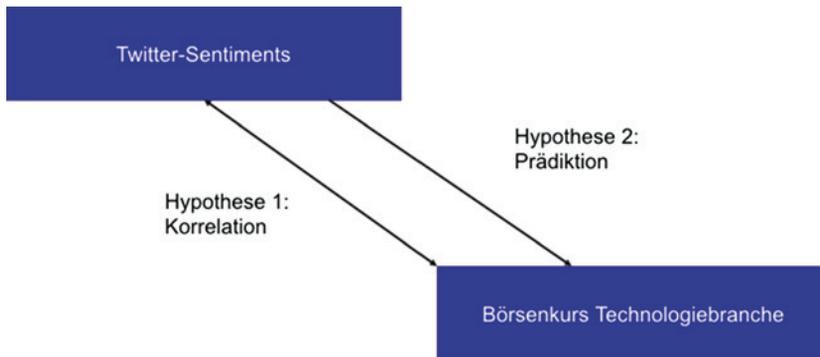


Abb. 9.4 Theoretisches Modell. (Eigene Darstellung)

erwähnten Begründungen, basierend auf bisherigen Forschungsergebnissen, werden folgende Hypothesen aufgestellt:

- **Hypothese 1**
Die Twitter-Sentiments im Zusammenhang mit einzelnen Technologieunternehmen weisen eine Korrelation mit der Entwicklung des Börsenkurses dieser Unternehmen auf.
- **Hypothese 2**
Die Twitter-Sentiments im Zusammenhang mit einzelnen Technologieunternehmen haben eine Voraussagekraft für die Entwicklung des Börsenkurses dieser Unternehmen.

Aus den oben aufgeführten Hypothesen ergibt sich für den Zusammenhang der Datensätze Twitter-Sentiments zu den einzelnen Technologieunternehmen und der Entwicklung der Börsenkurse derselben Unternehmen das theoretische Modell gemäß Abb. 9.4.

9.3 Methodisches Vorgehen

9.3.1 Rahmenbedingungen und Plattform

Die Definition der Rahmenbedingungen bezüglich der Plattform und der Unternehmen dient als Leitplanke für die Umsetzung.

In der vorliegenden Arbeit wurden die Sentiments der Tweets von der Twitter-Plattform verwendet. Dies aufgrund der größten Akzeptanz dieser Plattform in der Finanzbranche und der bereits erfolgreichen Nutzung in vielen Forschungsbereichen (Sprenger und Welppe 2010).

Nach (Nguyen et al. 2015) gibt es zwei verschiedene Arten von Tweets auf Twitter. Eine Art Tweet beginnt ein neues Thema die zweite Art Tweets, ein sogenannter Retweet, antwortet auf ein Thema (Nguyen et al. 2015). Die verschiedenen Arten von Tweets führen zu einem komplizierten Kommunikationsnetzwerk. In der vorliegenden Arbeit wurde die Art der Tweets vernachlässigt und unabhängig verwendet.

9.3.2 Unternehmen

Für diese Studie wurden drei Technologieunternehmen gewählt, da diese Branche an der Börse besonders interessant im Sinne des Wachstums sowie Entwicklungspotenzials ist (Ciolli 2017; Investopedia 2018). Aus den Technologieunternehmen wurden Tesla, Facebook und Amazon als Studienobjekte ausgewählt. Dies weil diese drei Unternehmen gemäß (Forbes Technology Council 2018) in den Top fünf der interessantesten Technologieaktien vom Jahr 2018 vertreten sind und aktuell in den Medien besonders präsent (Kawa 2018). Ein weiterer Punkt für die Auswahl ist der Name der Unternehmen, respektive deren Produkt oder Service. Das Produkt beziehungsweise der Service ist direkt mit dem Unternehmensnamen verbunden. Die Tweets können deshalb direkt mit dem Unternehmensnamen Tesla, Amazon oder Facebook gefiltert werden (Nguyen et al. 2015). Da die Unternehmenssprache Englisch ist, werden für die Studie nur englischsprachige Tweets berücksichtigt (Kroker 2013).

Neben der oben aufgeführten Rahmenbedingung soll die Umsetzung kostenlos sein. Dies wurde in der Umsetzung respektive in der Auswahl der Tools berücksichtigt.

9.3.3 Datenpipeline

Um die beschriebenen Hypothesen zu testen, wird dem Ablauf gemäß Abb. 9.5 gefolgt. In einem ersten Schritt wird das Datenformat definiert, welches aus strukturierten oder unstrukturierten Daten bestehen kann. In einem zweiten Schritt müssen Entscheidungen zu den Bezugsplattformen und den Möglichkeiten zum Bezug der Daten getroffen

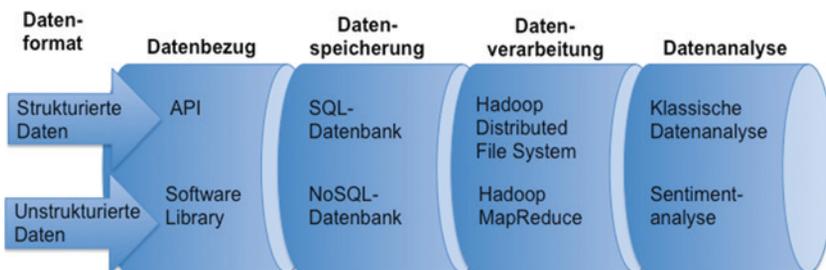


Abb. 9.5 Datenpipeline. (Eigene Darstellung)

werden. Weiter müssen die bezogenen Daten in einer Datenbank gespeichert werden, welche den gewählten Inputdaten entspricht. Bei einer großen Datenmenge kann außerdem ein Datenverarbeitungssystem miteinbezogen werden, um die Datenmenge verarbeitbar zu machen. Als letzter Schritt müssen die analysiert und interpretiert werden, um die Fragestellung zu beantworten.

Nachstehend wird vertieft auf die Schritte der Datenpipeline eingegangen. Die einzelnen Schritte beinhalten nur die für diese Forschungsarbeit notwendigen Überlegungen und sind daher nicht abschließend zu betrachten.

9.3.4 Datenformat

Daten können in zwei verschiedene Typen eingeteilt werden (Inmon und Linstedt 2015). Einerseits gibt es die strukturierten Daten, welche ein klares und gleichbleibendes Format aufweisen. Auf der anderen Seite gibt es die unstrukturierten Daten, welche keiner standardisierten Struktur folgen und dementsprechend auch mit Maschinen nicht einfach zu verarbeiten sind (Inmon und Linstedt 2015). In der vorliegenden Arbeit wurden beide Datenformate verwendet. Die Twitter-Daten sind aufgrund des freien Textformats als unstrukturierte Daten zu betrachten. Die Börsendaten sind in der Länge und Form klar definiert und folgen daher einer klaren Struktur.

9.3.5 Datenbezug

Für den Datenbezug der Twitter- und Börsendaten wurden verschiedene Anwendungen evaluiert. Die Abb. 9.6 zeigt als Übersicht auf, welche Anwendungen für den Bezug der zwei Twitter-Daten und der Börsendaten gewählt werden.

Damit Tweets von Twitter und der Börse gesammelt werden können, wird eine API (Mao et al. 2012; Ranco et al. 2015) und eine Library (Baumann 2018) benötigt. Twitter bietet die Standard-API, die Premium-API und die Enterprise-API an. Die Enterprise-API sowie die Premium-API sind kostenpflichtig, während die Standard-API kostenlos ist. Aus diesem Grund wird in der vorliegenden Arbeit die Standard-API verwendet (Twitter 2018). Innerhalb der Standard-API kann die Streaming-API oder die Searching-API verwendet werden (Twitter 2018). Die Streaming-API erlaubt die Tweets in Echtzeit herunterzuladen, jedoch nicht in der Vergangenheit zu suchen. Für eine Voraussage des Aktienkurses wird normalerweise die Streaming-API von Twitter verwendet (Vu et al. 2012). Allerdings ist für eine Streaming-API eine permanent laufende Infrastruktur mit Servern notwendig. In der vorliegenden Arbeit wird aufgrund der fehlenden kostenintensiven Infrastruktur die Searching-API verwendet. Diese API kann für Suchanfragen verwendet werden (Rao und Srivastava 2012; Ranco et al. 2015). Die Tweets können bis zu sieben Tage in die Vergangenheit bezogen werden. Die Library, als zweite Komponente für den Datenbezug, übernimmt die Kommunikation mit

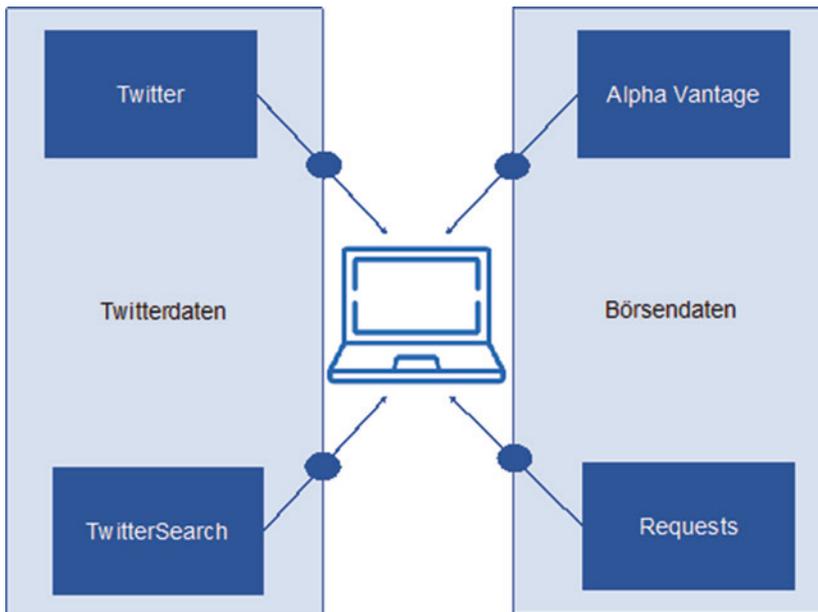


Abb. 9.6 Datenbezug. (Eigene Darstellung)

der Twitter-API anhand öffentlich verfügbarer Funktionen (educalingo 2018; Technopedia 2018). Grundsätzlich wurden Libraries für die Programmiersprache Python mit einer Internetrecherche evaluiert und auf die Eigenschaften kostenlos, installierbar, funktionsfähig respektive mit einer Lösung der Paginierung für Suchresultate, getestet. Die TwitterSearch Library, von der Technischen Universität München (Koepp 2018) beinhaltet nach einer Fehlerkorrektur der Paginierung alle Eigenschaften und wird deshalb als geeignet betrachtet und in dieser Studie verwendet (Baumann 2018).

Die Börsenkurse werden mittels Python-Skript von der Plattform Alpha Vantage bezogen (Alpha Vantage 2018). Diese Plattform bietet die Börsendaten bereits in Stundenintervallen an und die Daten sind in die Vergangenheit beziehbar. Die Anforderungen an diese Library sind geringer als bei der Twitter-API, daher fand hier keine Evaluation der Library statt und es wurde die populäre Python Library „requests“ verwendet (Reitz 2018).

9.3.6 Datenspeicherung

Damit die Daten, welche von Twitter bezogen wurden, verarbeitet werden können, müssen die Daten in einer Datenbank gespeichert werden (Baumann 2018). Dafür gibt es unter anderem SQL- und NoSQL-Datenbanken. Die SQL eignet sich für strukturierte Daten während NoSQL vor allem für die Speicherung von unstrukturierten Daten

Abb. 9.7 Datenbanken.
(Eigene Darstellung)



auszeichnet (Moniruzzaman und Hossain 2013; Litzel 2017). Die Datenbanken können wie in Abb. 9.7 aufgezeigt verwendet werden.

Da das Datenformat der beiden Datenströme sowohl strukturiert wie auch unstrukturiert ist, wird in der vorliegenden Studie eine NoSQL-Datenbank benötigt. Da die Tweets von der Twitter-API im Format JSON geliefert werden, wird eine Datenbank gewählt, welche das JSON-Format unterstützt. Dies soll einen möglichst robusten und einfachen Import ermöglichen. Weiter ist bei der Verwendung einer Datenbank für Big Data besonders wichtig, dass genügend Platz vorhanden ist und Abfragen gemacht werden können. Dafür eignen sich die NoSQL-Datenbanken CouchDB oder MongoDB. Die CouchDB vom Entwicklungsteam des Apache-Projekts erfüllt diese Bedingungen. Da sie außerdem eine Library für die Programmiersprache Python zur Verfügung stellt, wird die CouchDB für diese Studie verwendet (Apache CouchDB 2018).

9.3.7 Datenverarbeitung

Die herkömmliche Datenverarbeitung kann aufgrund des großen Datenvolumens bei Big Data nicht effizient und schnell genug arbeiten. Aus diesem Grund gibt es das Datenverarbeitungssystem Hadoop (Marz und Warren 2015). Hadoop ist eine Open-Source-Plattform, welche aus mehreren Komponenten besteht. Unter anderem beinhaltet Hadoop ein Software-Framework für Speicheroptimierung und eine Verarbeitungskomponente für Big Data (Gang-Hoon et al. 2014). Da die Datenmenge in dieser Studie noch schnell genug verarbeitet werden kann, wird auf ein Datenverarbeitungssystem verzichtet.

9.3.8 Datenanalyse

Damit die Daten verarbeitet und analysiert werden können, müssen unstrukturierte Twitter-Daten zu strukturierten Daten, wie in Abb. 9.8 aufgezeigt, umgewandelt werden. Dies erfolgt mithilfe einer Sentimentanalyse. Die Sentimentanalyse untersucht die Stimmung in einem Tweet. Diese Stimmung wird als Zahl zwischen 1 und -1 repräsentiert. Eine Sentimentanalyse wird oft mittels „Machine Learning“ vortrainiert (Baumann 2018; Graf 2018). In dieser Studie wird das bereits bestehende Natural Language Toolkit (NLTK) verwendet (Bird et al. 2015). Zum NLTK gehört das Text-Analyse-Tool „Valence Aware Dictionary and sentiment Reasoner“ (Vader), welches vortrainiert ist, um die Stimmung von Tweets in englischer Sprache zu ermitteln (NLTK 2017). Das Text-Analyse-Tool „Vader“ ist speziell für die Analyse von Stimmungen in sozialen

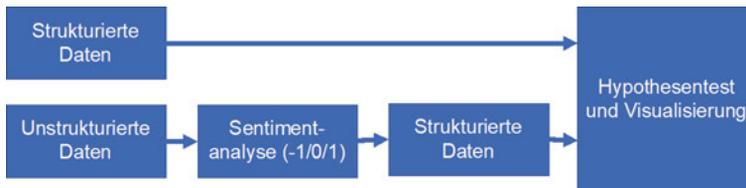


Abb. 9.8 Datenanalyse. (Eigene Darstellung)

Medien programmiert (Hutto 2018). Auch (Brown 2012) verwendet für die Sentimentanalyse erfolgreich das Natural Language Processing Toolkit in Python. Die Sentimentanalyse wurde mittels eines Skripts für alle gesammelten Tweets angewendet. Dabei wurden alle Tweets aus der CouchDB geladen, analysiert und mit dem eruierten Sentiment-Wert wieder in der CouchDB gespeichert. In einem nächsten Schritt werden für die Analysen der Daten, die Sentiments der einzelnen Tweets mittels arithmetischem Mittel pro Stunde gruppiert.

Die Börsenkurse werden in einem regelmäßigen Intervall von einer Stunde gespeichert, um eine zeitnahe Veränderung beobachten zu können. Da die Börse nicht durchgehend geöffnet ist, gibt es Lücken in den Datensätzen. Diese Lücken werden mit der Python-Library „pandas“ überbrückt (pandas 2018; Baumann 2018). Die Lücken wurden jeweils mit dem letzten bekannten Börsenkurs gefüllt.

Um die Hypothesen zu testen, wurden im letzten Schritt der Abb. 9.8 zwei verschiedene Analysen durchgeführt. Für die **Hypothese 1** wurde untersucht, ob einen Zusammenhang zwischen den Twitterdaten und den Börsendaten desselben Unternehmens besteht. Aus diesem Grund wird eine **Korrelationsanalyse** durchgeführt. Die **Hypothese 2** wird anhand der **Granger-Analyse** getestet. Die Granger-Analyse untersucht, ob die Twitter-Daten die Börsenkurse voraussagen können und welches Timelag (Zeitdifferenz) dazwischen liegt.

9.4 Ergebnisse

9.4.1 Beschreibung des Datensatzes

Daten der Tweets und der Börsenkurse wurden in einem Zeitraum vom 25. März 2018 bis zum 20. Mai 2018 bezogen. Für den Datenbezug wurde zur Optimierung der Tweets nicht nur die Unternehmensbezeichnung verwendet. Beim Unternehmen wurde beispielsweise Facebook zusätzlich zum Begriff Facebook auch der Hashtag #facebook eingeschlossen. Als erster Schritt wurde der Datenbezug der Twitter-Daten zum Unternehmen Tesla angestoßen. Nach einer Fehlerbehebung folgten dann am 30. April 2018 die Einstellungen für die Unternehmen Facebook und Amazon. Aufgrund eines Funktionsfehlers ist der Zeitraum vom 16. April bis zum 29. April nicht brauchbar.

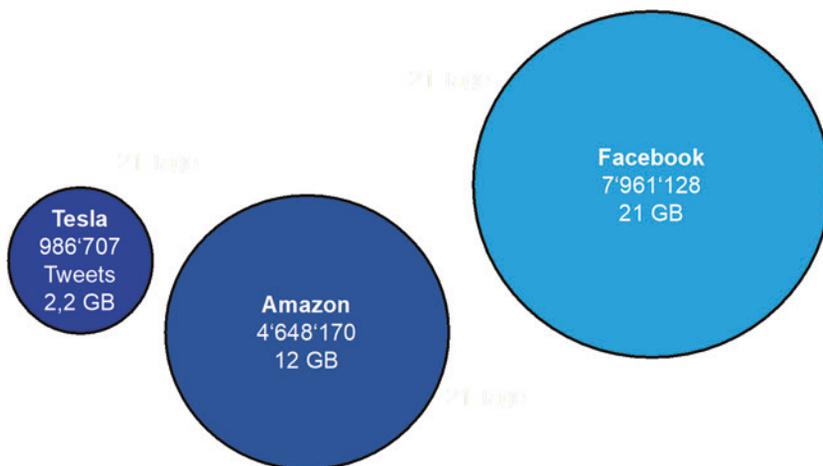
Tab. 9.1 Zeiträume. (Eigene Darstellung)

Unternehmen	Zeitraum		Zeitraum mit Funktionsfehler		Anzahl verwendete Tage
	Von	Bis	Von	Bis	
Tesla	25.03.2018	20.05.2018	16.04.2018	29.04.2018	21
Facebook	30.04.2018	20.05.2018			21
Amazon	30.04.2018	20.05.2018			21

Dies betrifft jedoch nur das Unternehmen Tesla, welches nun in der Mitte der Daten einen Zeitraum von 14 Tagen aufweist, welcher nicht genutzt werden kann. Eine Übersicht der Zeiträume ist der Tab. 9.1 zu entnehmen.

Innerhalb der Zeiträume, welche in der Tab. 9.1 aufgeführt sind, konnte je nach Unternehmen eine unterschiedliche Anzahl Tweets bezogen werden. Die Abb. 9.9 zeigt die bezogenen Tweets pro Unternehmen.

Wie der Abb. 9.9 zu entnehmen ist, konnten von Tesla über den gesamten Zeitraum, im Verhältnis zu Facebook und Amazon, nur wenige Daten gespeichert werden. Dabei ist noch zu bedenken, dass sich diese Anzahl Tweets von Tesla auf zwei verschiedene Zeiträume, mit einer Lücke von 14 Tagen dazwischen, beziehen. Daher ist für den durchgehenden Zeitraum vom 30. April 2018 bis zum 20. Mai 2018 nur noch etwa die Hälfte der in der Abb. 9.9 verwendeten Tweets nutzbar. Aufgrund der geringen Anzahl an brauchbaren Daten von Tesla und des großen Unterschieds der Anzahl Daten zu den zwei anderen Unternehmen, wurde Tesla als Untersuchungsobjekt ausgeschlossen. Zukünftige Forschungsarbeiten könnten das Unternehmen als Untersuchungsobjekt integrieren. Im vorliegenden Fall wurden nun aber lediglich Amazon und Facebook im Rahmen der Studie analysiert.

**Abb. 9.9** Bezogene Tweets. (Eigene Darstellung)

Spitzenwerte

Bereits bei der ersten Betrachtung des Verlaufs der Anzahl Tweets, der Twitter-Sentiments und der Veränderung der Börsendaten kann bei beiden Unternehmen ein leichtes Muster erkannt werden. Die Abb. 9.10 und die Abb. 9.11 zeigen jeweils den Verlauf dieser drei Werte für das Unternehmen Facebook und Amazon auf.

Im Folgenden sind die Amazon-Daten dargestellt.

Wie Abb. 9.10 und Abb. 9.11 zu entnehmen ist, gibt es zum Teil starke Bewegungen und trotz der kurzen Zeitdauer der Beobachtung einige Spitzenwerte. Wie in der Tab. 9.2 entnommen werden kann, sind diese Spitzenwerte auch tatsächlich einem Ereignis mit Zusammenhang zum Unternehmen zuzuordnen.

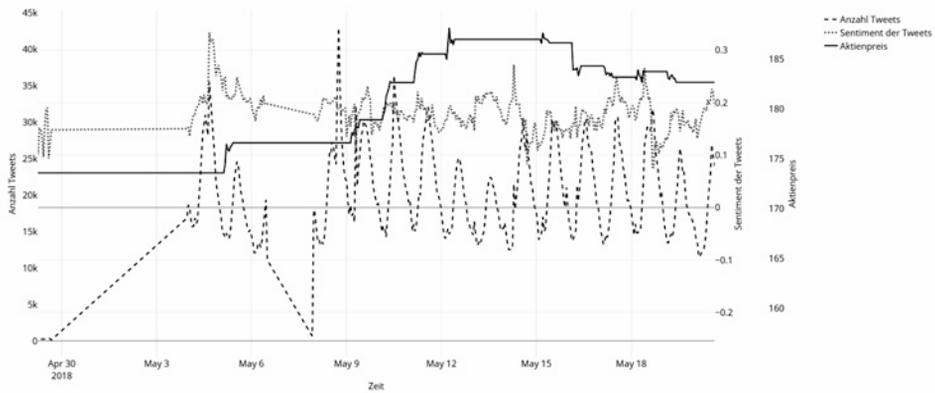


Abb. 9.10 Facebook Daten. (Eigene Darstellung)

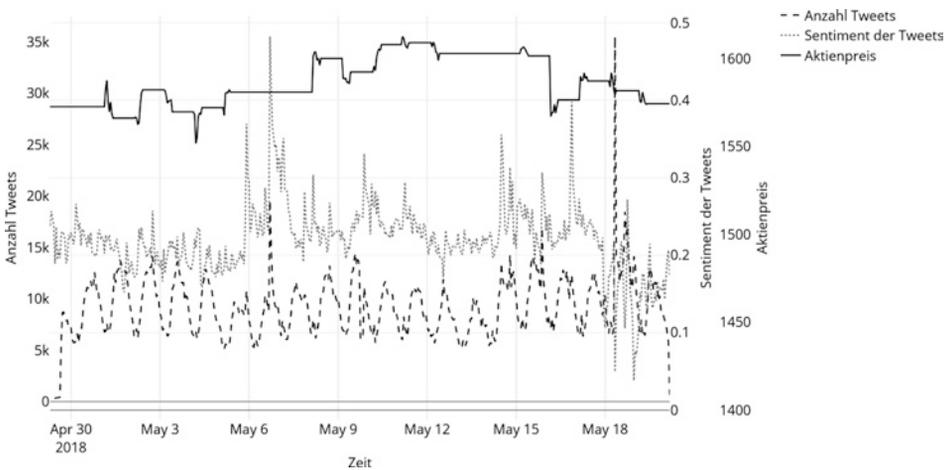


Abb. 9.11 Amazon Daten. (Eigene Darstellung)

Tab. 9.2 Spitzenwerte. (Eigene Darstellung)

Facebook	04.05.2018	0,333	Entlassung Mitarbeiter wegen Prahlerei mit Nutzerdaten (Redaktion 2018)
	14.05.2018	0,272	Eigene Kryptowährung für Facebook (Bluewin 2018b)
	18.05.2018	0,266	Partnerschaft mit Thinktank (Möller 2018) Facebook startet Werbeeinblendungen in Stories (Kroll 2018)
	09.05.2018	0,106	Führungsumbau (Neue Zürcher Zeitung 2018)
	14.05.2018	0,082	Facebook äußert Verdacht über Missbrauch von Nutzerdaten durch diverse Apps (finanzen.ch 2018a)
	18.05.2018	0,076	Partnerschaft mit Thinktank (Möller 2018) Facebook startet Werbeeinblendungen in Stories (Kroll 2018)
Amazon	06.05.2018	0,482	Amazon prüft Einstieg ins Bankgeschäft (Prior 2018)
	16.05.2018	0,398	Streik bei Amazon (finanzen.ch 2018b) Neugestaltung der Lebensmittelläden (Wingfield 2018)
	05.05.2018	0,369	Gewinnmaschine Amazon (Schürmann 2018) Amazon-Chef hebt ab (Beglinger und Beglinger 2018)
	18.05.2018	0,185	Wachstum der Amazon-Aktie (Sasse 2018) Amazon Prime Erhöhung der Preise (Rauffmann 2018) Mögliche Einmischung der Politik bei Amazon führt zu Verlierern auf beiden Seiten (Mullainathan 2018)

9.4.2 Bezug zu Big Data

Die Definition von Big Data erfolgt, wie in der Einleitung erwähnt, anhand der vier Eigenschaften Datenvolumen, Datenvielfalt, Datengeschwindigkeit und Richtigkeit. Der verwendete Datensatz weist diese Eigenschaften gemäß der Tab. 9.3 erwähnten Form auf.

Das totale verwendete Datenvolumen überschreitet bereits die Kapazitäten von bspw. dem Programm Excel (Microsoft 2018). Auch das Kriterium der Datenvielfalt wurde in dieser Studie erfüllt, da es sich sowohl um strukturierte wie auch unstrukturierte Daten handelt. Durch die sich ständig verändernden Twitter- und Börsendaten wurde die Datengeschwindigkeit ebenfalls erreicht. Das Kriterium Richtigkeit wurde aufgrund der direkt vom User generierten Daten ebenfalls erfüllt. Ein möglicher Unsicherheitsfaktor dabei ist jedoch die Sentimentanalyse. Da der Algorithmus jedoch bereits mehrmals erfolgreich getestet wurde, kann von einer „korrekten“ Analyse ausgegangen werden. Veränderungen und Weiterentwicklungen sind im Rahmen des technologischen Fortschritts möglich. Auch andere Algorithmen im genannten Kontext könnten das Untersuchungsergebnis verändern. Für die Zeitpunkte der Untersuchung gelten die zuvor dargestellten Aussagen.

Im Folgenden wird nun auf den Test der Hypothesen eingegangen.

Tab. 9.3 Big-Data-Eigenschaften. (Eigene Darstellung)

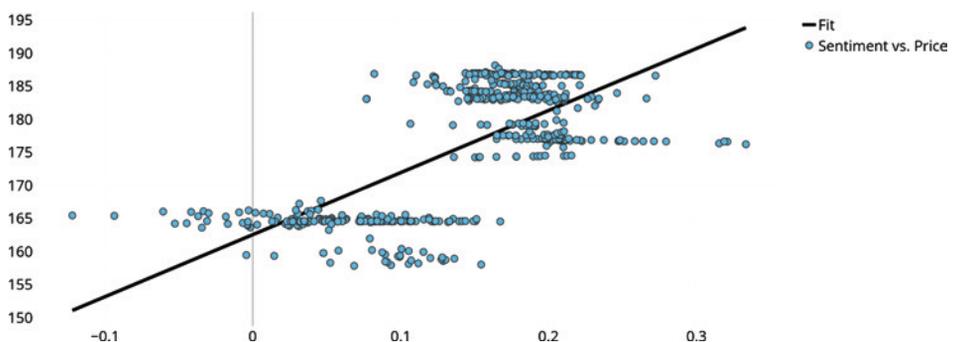
Eigenschaften	Inhalt	Verwendeter Datensatz
Datenvolumen/Volume	Anzahl und Größe von Daten	• 32,6 Gigabyte
Datenvielfalt/Variety	Quellen und Arten von Daten	• Quellen: Twitter und Alpha Vantage • Arten: strukturiert und unstrukturiert
Datengeschwindigkeit/Velocity	Datengenerierung in hoher Geschwindigkeit	• 32,6 Gigabyte innerhalb 21 Tagen • Stetige Veränderung
Richtigkeit/Veracity	Vollständigkeit und Verlässlichkeit der Dateninhalte	• User-generierte Textdaten vorhanden

9.4.3 Hypothesentest

9.4.3.1 Korrelationsanalyse

9.4.3.1.1 Facebook

Das Streudiagramm der Abb. 9.12 zeigt die Beziehungspunkte zwischen den Twitter-Sentiments im Zusammenhang mit Facebook (x-Achse) und dem Börsenkurs von Facebook (y-Achse). Die meisten Datenpunkte befinden sich auch hier eher bei der Orientierungslinie. Fast alle Punkte befinden sich außerdem im positiven Bereich des Streudiagramms. Der Spearman's Korrelationskoeffizient zeigt ebenfalls eine eher starke positive Beziehung mit dem Wert $r=0,557$. Die Signifikanz ist mit dem Wert von $p=0,00$ sehr hoch. Diese Analyse zeigt, dass die Twitter-Sentiments zu Facebook und der Börsenkurs von Facebook sich zyklisch bewegen und einen starken Zusammenhang aufweisen.

**Abb. 9.12** Streudiagramm – Facebook. (Eigene Darstellung)

9.4.3.1.2 Amazon

Das Streudiagramm in Abb. 9.13 zeigt die Beziehung der Twitter-Sentiments zu Amazon und der Börsenkurse von Amazon. Wie das Streudiagramm zeigt, liegen die Punkte weit auseinander und weniger um die Orientierungslinie herum. Auch der Spearman's Korrelationskoeffizient zeigt mit einem Wert von $r=0,319$ nur einen moderaten Zusammenhang auf. Die Signifikanz für diesen Wert ist mit $p=0,00$ ebenfalls gegeben und weist auf, dass die gemessenen Daten aussagekräftig sind. Daher kann davon ausgegangen werden, dass zwischen den Twitter-Sentiments zu Amazon und den Börsenkursen von Amazon eine moderate positive Beziehung besteht.

9.4.3.1.3 Schlussfolgerung

Die **Hypothese 1** lautet:

- „Die Twitter-Sentiments im Zusammenhang mit einzelnen Technologieunternehmen weisen eine Korrelation mit der Entwicklung des Börsenkurses dieser Unternehmen auf.“

Wie die Analysen der Spearman's Korrelationskoeffizienten zeigen, ist die Korrelation zwischen den Twitter-Sentiments eines spezifischen Technologieunternehmens und dem Börsenkurs desselben Technologieunternehmens bei diesen zwei Beispielen gegeben. **Es weisen beide Analysen einen Zusammenhang auf und somit ist die Hypothese 1 angenommen.**

Im Folgenden werden nun die Ergebnisse der Analysen im Kontext der Prädikation dargestellt.

9.4.3.2 Granger-Analyse

9.4.3.2.1 Facebook

In der Abb. 9.14 ist die Wahrscheinlichkeit einer Voraussage der Börsenkurse von Facebook anhand der Twitter-Sentiments zu Facebook dargestellt. Es wurde hier wieder ein

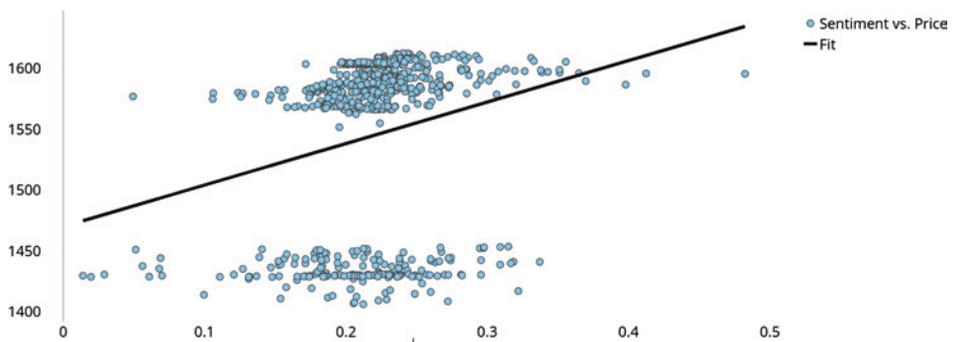


Abb. 9.13 Streudiagramm – Amazon. (Eigene Darstellung)

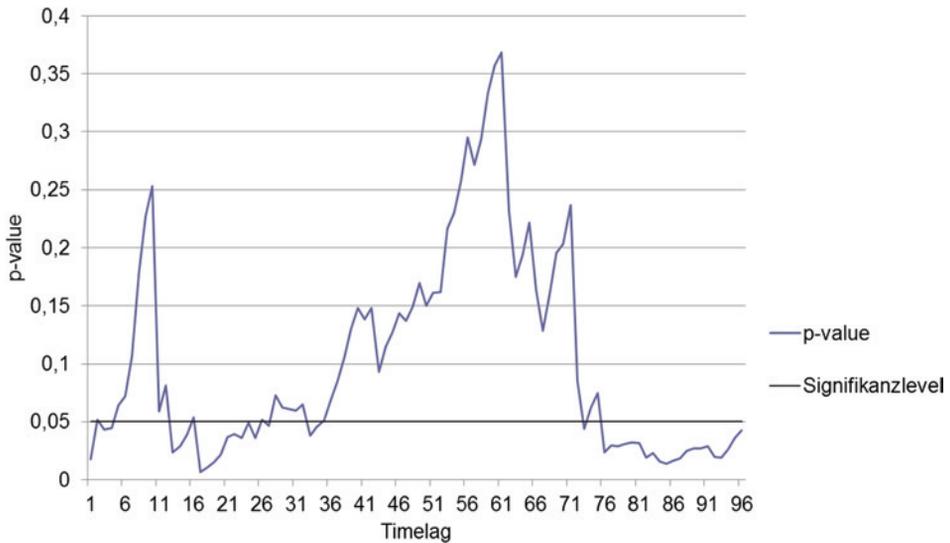


Abb. 9.14 Granger-Analyse – Facebook. (Eigene Darstellung)

Timelag von 1 bis 96 h analysiert. Der p-Wert in der Abb. 9.14 fällt mehrfach unter den Grenzwert von $p=0,05$. Das erste Mal ist der Grenzwert bei Stunde 13 unterschritten. Bei der Stunde 28 steigt der p-Wert wieder über den Grenzwert und ab Stunde 76 liegt der p-Wert wieder unter dem geforderten Grenzwert. Somit ist anzunehmen, dass die Twitter-Sentiments zu Facebook 13 im Voraus Börsenkurs voraussagen können.

9.4.3.2.2 Amazon

Die Abb. 9.15 zeigt den Wert der Wahrscheinlichkeit, dass die Twitter-Sentiments zu Amazon die Börsenkurse von Amazon voraussagen können. Der p-Wert liegt auch bei Amazon bei der Stunde 13 erstmals unter der minimalen Grenze von $p=0,05$ bei $p=0,006$. Dies bedeutet, dass eine Wahrscheinlichkeit von mehr als 99 % besteht, dass die Twitter-Sentiments zu Amazon die Börsenkurse von Amazon voraussagen. Der Börsenkurs kann folglich auch hier durch die Twitter-Daten 13 h im Voraus eruiert werden.

9.4.3.2.3 Schlussfolgerung

Die **Hypothese 2** lautete:

- „Die Twitter-Sentiments im Zusammenhang mit einzelnen Technologieunternehmen haben eine Voraussagekraft für die Entwicklung des Börsenkurses dieser Unternehmen.“

Die Hypothese 2 kann anhand der getesteten zwei Unternehmen angenommen werden. Die Granger-Analyse bei Facebook und Amazon hat einen **signifikanten Wert** und ist somit erfolgreich für die **Voraussage der Börsenkurse anhand von Twitter-Sentiments**.

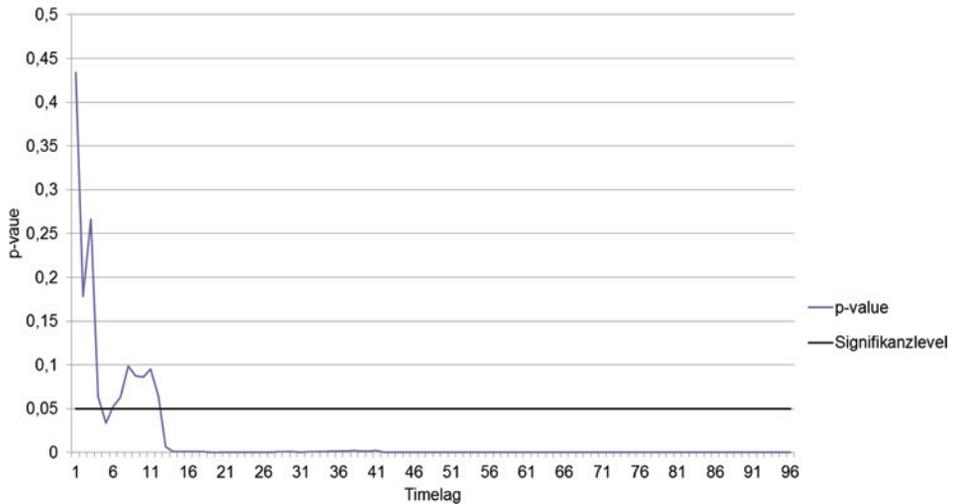


Abb. 9.15 Granger-Analyse – Amazon. (Eigene Darstellung)

Die Hypothese 2 ist somit angenommen. Bei 13 h Timelag weisen beide Unternehmen einen signifikanten p-Wert auf. Somit können die Daten von Twitter anhand der Sentiments die Bewegung der Börsenkurse 13 h im Voraus voraussagen.

9.5 Fazit und kritische Diskussion

Die verwendeten Daten in der vorliegenden Studie entsprechen grundsätzlich der Definition von Big Data. Somit kann sichergestellt werden, dass die Analysemethoden aus dem Bereich Big Data Analytics in dieser Studie angewendet werden können. Die Spearman's Korrelationsanalyse der Twitter-Sentiments und Börsenkurse zeigt bei beiden Unternehmen eine positive Korrelation. Auch wenn die Korrelation bei Amazon nur moderat ist. Somit ist die Hypothese 1 angenommen. Die Granger-Analyse weist ebenfalls einen signifikanten Wert aus und zeigt somit, dass die Börsenkurse durch mit Sentiment analysierte Twitter-Daten vorausgesagt werden können. Dabei ist herauszuheben, dass das Timelag bei beiden Unternehmen einen Wert von 13 h anzeigt. **Somit zeigen die verarbeiteten Twitter-Daten ab 13 h im Voraus die Veränderung der Börsenkurse an.**

9.5.1 Methodisches Vorgehen

Das methodische Vorgehen beinhaltete mehrere Entscheidungen, welche das Resultat möglicherweise beeinflusst haben. Ein umfassender Einflussfaktor war die Zeitdauer der Studie. Die Forschungsarbeit wurde in rund drei Monaten realisiert. Da keine

historischen Tweets bezogen werden konnten, war der Zeitraum des Datenbezugs eingeschränkt. Ein längerer bzw. veränderter Zeitraum könnte zu Veränderungen des Ergebnisses führen. Hier sei auf eine Replikation mit einem veränderten Zeitraum für zukünftige Studien hingewiesen.

Beim Bezug der Twitter-Daten gab es ebenfalls einige Entscheidungen, welche Einfluss auf das Ergebnis genommen haben könnten. Beispielsweise wurde die Sprache der Tweets auf Englisch eingeschränkt. Diese Entscheidung war darauf zurückzuführen, dass die Sentimentanalyse anhand von bestehenden Bausteinen durchgeführt werden musste. Diese Bausteine waren sprachlich oft eingeschränkt, was dazu führte, dass für jede Sprache ein eigener Baustein hätte eingebaut werden müssen. Daher wurde die Sprache auf die gängigste Sprache von Tweets beziehungsweise auf die Unternehmenssprache der Unternehmen reduziert. Zu bedenken ist aber, dass an der Börse nicht nur Anleger der englischen Sprache teilnehmen. Somit wurde im Datensatz der Tweets und der Anleger ein unterschiedlicher Teil der Population untersucht. Diese Tatsache könnte durchaus einen Einfluss auf die Resultate haben. Ein weiterer Punkt, welcher teilweise auch mit den verschiedenen Sprachregionen zusammenhängt, ist das Problem der Zeitzonen. Da in der Granger-Analyse der Unterschied der Stunden analysiert wird, könnte eine Einbindung von anderen Zeitzonen das Resultat verändern. Neben der Einbindung von anderen Sprachen und dementsprechend auch Regionen, wäre auch ein Optimierungspotenzial beim Befehl des Datenbezugs möglich. Bei dieser Studie wurden nur Tweets berücksichtigt, welche den Unternehmensnamen enthielten. Möglicherweise sind durch diese eingeschränkten Suchparameter Tweets nicht erfasst worden, welche im Kontext zum Unternehmen stehen, aber dieses nicht explizit erwähnen oder ausschreiben. Ebenfalls ist die Auswahl der Unternehmen kritisch zu hinterfragen. Obwohl diese zwei Unternehmen zu den bekanntesten Technologieunternehmen gehören, würde doch die Analyse von weiteren Unternehmen ein deutlicheres Bild der Thematik darstellen.

Neben den oben erwähnten Punkten gibt es noch weitere Überlegungen, welche in dieser Ausarbeitung unklar sind und weiter untersucht werden müssten. Beispielsweise wäre es möglich, dass neben den Tweets noch weitere Einflussfaktoren miteinbezogen werden müssten, um die Voraussage zu verbessern. Weiter ist mit dem gewählten Time-lag von 1 bis 96 bei der Granger-Analyse nicht ganz klar, was nach 96 h passiert.

9.5.2 Handlungsempfehlungen

Für weitere Forschungsarbeiten in diesem Gebiet könnten gemäß den Erläuterungen in den vorherigen Kapiteln folgende Punkte berücksichtigt werden:

- Zeitdauer der Studie
- Sprache der Tweets und Zeitzonen
- Begriffssuche der Tweets
- Auswahl und Anzahl der analysierten Technologieunternehmen

- Einbindung weiterer Einflussfaktoren
- Erweiterung der Timelags
- Machine Learning für Sentimentanalyse

Abschließend wird von den Autoren empfohlen, diese Studie zur Validierung zu replizieren und anhand der Handlungsempfehlungen zu optimieren.

Literatur

- Alpha Vantage. (2018). Alpha Vantage – Free APIs for Realtime and Historical Financial Data, Technical Analysis, Charting, and More! Alpha Vantage. <https://www.alphavantage.co/>. Zugegriffen: 27. Mai 2018.
- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3), 1259–1294. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2004.00662.x>. Zugegriffen: 16. Apr. 2018.
- Apache CouchDB. (2018). <http://couchdb.apache.org/>. Zugegriffen: 27. Mai 2018.
- Baumann, J. (2018). Technische Erarbeitung Umsetzung. Expertengespräch geführt im Rahmen der Studie.
- Beglinger, M. (2018). Amazon-Chef Jeff Bezos hebt ab | NZZ Neue Zürcher Zeitung. <https://www.nzz.ch/gesellschaft/amazon-chef-jeff-bezos-hebt-ab-ld.1390402>. Zugegriffen: 3. Juni 2018.
- Bendler, J., Wagner, S., Brandt, T., & Neumann, D. (2018). Informationsunschärfe in Big Data. [springerprofessional.de. https://www.springerprofessional.de/informationsunschaerfe-in-big-data/3424280?fulltextview=true](https://www.springerprofessional.de/informationsunschaerfe-in-big-data/3424280?fulltextview=true). Zugegriffen: 22. Apr. 2018.
- Bird, S., Klein, E., & Loper, E. (2015). NLTK Book. <http://www.nltk.org/book/>. Zugegriffen: 12. Mai 2018.
- Bluewin. (2018). Arbeitet Facebook an einer eigenen Kryptowährung? <https://www.bluewin.ch/de/digital/arbeitet-facebook-an-eigener-kryptowaehrung-100140.html>. Zugegriffen: 3. Juni 2018.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
- Brown, E. D. (2012). Will twitter make you a better investor? A look at sentiment, user reputation and their effect on the stock market. *Proc. of SAIS*, 7.
- Ciulli, J. (2017). Stock market volatility is back – And tech stocks are taking a beating business insider. <http://uk.businessinsider.com/stock-market-news-volatility-is-back-tech-stocks-taking-beating-2017-12>. Zugegriffen: 16. Mai 2018.
- Das, S., Martínez-Jerez, A., & Tufano, P. (2005). eInformation: A clinical study of investor discussion and sentiment. *Financial Management*, 34(3), 103–137.
- Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375–1388.
- Davenport, T. H., & Paulus, P. (2014). *Big data @ work: Chancen erkennen, Risiken verstehen*. München: Vahlen.
- Dijcks, J.-P. (2013) Oracle: Big data for the enterprise. <http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf>. Zugegriffen: 31. Dez. 2018.
- Dorschel, J. (2015). *Praxishandbuch Big Data*. Wiesbaden: Springer Fachmedien.
- educalingo. (2018). Programmbibliothek. <https://educalingo.com/de/dic-de/programmbibliothek>. Zugegriffen: 20. Mai 2018.

- EMC und Greenplum. (2012). Ten Practical Big Data Benefits | Stories | Data Science Series Data Science Series. <http://datascienceseries.com/stories/ten-practical-big-data-benefits>. Zugegriffen: 16. Mai 2018.
- finanzen.ch. (2018a). Facebook klemmt rund 200 Apps nach Datenskandal ab | 14.05.18 finanzen.ch. <https://www.finanzen.ch/nachrichten/aktien/Facebook-klemmt-rund-200-Apps-nach-Daten-skandal-ab-1024442857>. Zugegriffen: 3. Juni 2018.
- finanzen.ch. (2018b). Wieder Streik bei Amazon angekündigt | 16.05.18 finanzen.ch. <https://www.finanzen.ch/nachrichten/aktien/Wieder-Streik-bei-Amazon-angekueendigt-1024660684>. Zugegriffen: 3. Juni 2018.
- Forbergskog, J.-O., & Blom, C. R. (2014). Twitter and stock returns. <https://brage.bibsys.no/xmlui/handle/11250/94935>. Zugegriffen: 21. Apr. 2018.
- Forbes Technology Council. (2018). 12 Tech companies to watch in 2018 forbes. <https://www.forbes.com/sites/forbestechcouncil/2018/02/12/12-critical-tech-stocks-to-watch-in-2018/#fc354cc7d783>. Zugegriffen: 16. Mai 2018.
- Gang-Hoon, K., Trimi, S., & Ji-Hyong, C. (2014). Big-data applications in the government sector. *Communications Of The ACM*, 57(3), 78–85.
- Gluchowski, P., & Chamoni, P. (Hrsg.). (2016). *Analytische Informationssysteme: Business Intelligence-Technologien und -Anwendungen* (5., vollständig überarbeitete Aufl.). Berlin: Springer Gabler.
- Graf, L. (2018). Technische Erarbeitung und Umsetzung. Expertengespräch geführt im Rahmen der Studie.
- Hutto, C. J. (2018). vaderSentiment: VADER Sentiment Analysis. <https://github.com/cjhutto/vader-Sentiment>. Zugegriffen: 1. Juni 2018.
- IBM. (2017). Big Data Analytics | IBM Analytics IBM Analytics. <https://www.ibm.com/analytics/hadoop/big-data-analytics>. Zugegriffen: 31. Dez. 2017.
- Inmon, W. H., & Linstedt, D. (2015). *Data Architecture: A primer for the data scientist – Big data, data warehouse and data vault*. Waltham: Elsevier Inc.
- Investopedia. (2018). Technology Sector Investopedia. https://www.investopedia.com/terms/t/technology_sector.asp. Zugegriffen: 16. Mai 2018.
- Kawa, L. (2018). Tech stocks fear premium just jumped to a 13-year high. In: Bloomberg.com, 28.03.2018. <https://www.bloomberg.com/news/articles/2018-03-28/tech-fear-premium-jumps-to-13-year-high-on-regulatory-anxiety>. Zugegriffen: 15. Mai 2018.
- King, S., & Hajnal, I. (2014). *Big Data: Potential und Barrieren der Nutzung im Unternehmenskontext*. Wiesbaden: Springer VS.
- Koepp, Christian (2018) TwitterSearch: A Python library to easily iterate tweets found by the Twitter Search API. <https://github.com/ckoepp/TwitterSearch>. Zugegriffen: 27. Mai 2018.
- Kroker, M. (2013). Häufigste Sprachen auf Twitter: Englisch vor Japanisch & Spanisch – Deutsch nicht in Top-10. Kroker's Look@IT. <http://blog.wiwo.de/look-at-it/2013/12/16/haufigste-sprachen-auf-twitter-englisch-vor-japanisch-spanisch-deutsch-nicht-in-top-10/>. Zugegriffen: 2. Juni 2018.
- Kroll, S. (2018). Facebook startet Werbeeinblendungen in Stories INTERNET WORLD Business. <https://www.internetworld.de/online-marketing/facebook-marketing/facebook-startet-werbeeinblendungen-in-stories-1539373.html>. Zugegriffen: 3. Juni 2018.
- Liew, J. K.-S., & Wang, G. Z. (2016). Twitter sentiment and IPO performance: A cross-sectional examination. *Journal of Portfolio Management*, 42(4), 129–135.
- Litzel, N. (2017). Was ist NoSQL? <https://www.bigdata-insider.de/was-ist-nosql-a-615718/>.
- Mankiw, N. G. (2004). *Grundzüge der Volkswirtschaftslehre* (3., überarb. Aufl.). Stuttgart: Schäffer-Poeschel.

- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2018). Big data: The next frontier for innovation, competition, and productivity | McKinsey & Company. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>. Zugegriffen: 1. Apr. 2018.
- Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, twitter and search engine data. [arXiv:1112.1051](https://arxiv.org/abs/1112.1051) [physics, q-fin].
- Mao, Y., Wei, W., Wang, B., & Liu, B. (2012). Correlating S&P500 Stocks with Twitter Data. In: Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research. New York, NY, USA: ACM. S. 69–72.
- Marr, B. (2016). *Big data in practice: How 45 successful companies used big data analytics to deliver extraordinary results*. Hoboken: Wiley.
- Marz, N., & Warren, J. (2015). *Big Data – Principles and best practice of scalable real-time data systems*. New York: Manning Publications Co.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. London: John Murray.
- Microsoft. (2018). Spezifikationen und Beschränkungen in Excel – Excel. <https://support.office.com/de-de/article/Spezifikationen-und-Beschr%C3%A4nkungen-in-Excel-1672b34d-7043-467e-8e27-269d656771c3>. Zugegriffen: 2. Juni 2018.
- Möller, M. (2018). Facebook geht Partnerschaft mit Thinktank ein. In: Tages-Anzeiger, 18.05.2018. <https://www.tagesanzeiger.ch/digital/social-media/facebook-geht-partnerschaft-mit-thinktank-ein/story/14515281>. Zugegriffen: 3. Juni 2018.
- Moniruzzaman, A. B. M., & Hossain, S. A. (2013). NoSQL database: New era of databases for big data analytics – Classification, characteristics and comparison. *International Journal of Database Theory and Application*, 6(4), 14.
- Mullainathan, S. (20. Mai 2018). When the President takes On Amazon, nobody wins. *The New York Times*. <https://www.nytimes.com/2018/05/18/business/trump-amazon-economy.html> (03.06.2018).
- Neue Zürcher Zeitung. (9. Mai 2018). Facebook ordnet die Führung wichtiger Produkte neu | NZZ. *Neue Zürcher Zeitung*. <https://www.nzz.ch/wirtschaft/facebook-ordnet-die-fuehrung-wichtiger-produkte-neu-ld.1384417>. Zugegriffen: 3. Juni 2018.
- Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, 42(24), 9603–9611.
- NLTK. (2017). nltk.sentiment.vader — NLTK 3.3 documentation. https://www.nltk.org/_modules/nltk/sentiment/vader.html. Zugegriffen: 2. Juni 2018.
- Oh, C., & Sheng, O. (2011). Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. *Iciss. Citeseer*. S. 1–19.
- pandas. (2018). PyPI. <https://pypi.org/project/pandas/>. Zugegriffen: 1. Juni 2018.
- Prior, G. (2018). Amazon – Droht jetzt auch Gefahr für Banken? *finanzen.net*. <http://www.finanzen.net/nachricht/aktien/euro-am-sonntag-meinung-amazon-droht-jetzt-auch-gefahr-fuer-banken-6139808>. Zugegriffen: 3. Juni 2018.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetic, I. (2015). The effects of twitter sentiment on stock price returns. *PLoS one*, 10, e0138441.
- Rao, T., & Srivastava, S. (2012). Analyzing Stock Market Movements Using Twitter Sentiment Analysis. In: *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. Washington, DC, USA: IEEE Computer Society. S. 119–123.
- Rauffmann, T. (2018). Amazon hebt Preise für Amazon Prime an *finanzen.net*. <http://www.finanzen.net/nachricht/aktien/bald-auch-in-deutschland-amazon-hebt-preise-fuer-amazon-prime-an-6142378>. Zugegriffen: 3. Juni 2018.

- Redaktion. (2018). 04.05.2018 – Prahlerei mit Nutzerdaten: Facebook feuert Mitarbeiter | Home24 fährt weiter Verluste ein | Instagram integriert Bezahlfunktion <https://onlinehaendler-news.de/>. <https://www.onlinehaendler-news.de/e-commerce-tipp/31589-prahlerei-nutzerdaten-facebook-feuert-mitarbeiter-home24-verluste-instagram-bezahlfunktion.html>. Zugegriffen: 3. Juni 2018.
- Reed, M. (2016). A study of social network effects on the stock market. *Journal of Behavioral Finance*, 17(4), 342–351.
- Reitz, K. (2018). Requests: HTTP for Humans – Requests 2.18.4 documentation Requests: HTTP for Humans. <http://docs.python-requests.org/en/master/>. Zugegriffen: 27. Mai 2018.
- Sasse, R. (2018). Amazon Aktie Kursziel Finanztrends. <https://www.finanztrends.info/amazon-aktie-kursziel/>. Zugegriffen: 3. Juni 2018.
- Scherbaum, C. A. (2017). *So funktioniert die Börse* (3. Aufl.). Freiburg: Haufe Lexware.
- Schürmann, L. (2018). Amazon Web Services: Der Aufstieg von Jeff Bezos' Gewinnmaschine manager magazin. <http://www.manager-magazin.de/unternehmen/handel/amazon-web-services-der-aufstieg-von-jeff-bezos-gewinnmaschine-a-1206196.html>. Zugegriffen: 3. Juni 2018.
- Souza, T. T. P., Kolchyna, O., Treleaven, P. C., & Aste, T. (2015). Twitter sentiment analysis applied to finance: A case study in the retail industry. <https://arxiv.org/pdf/1507.00784.pdf>.
- Sprenger, T. O., & Welpel, I. M. (2010). Tweets and trades: The information content of stock microblogs. Rochester, NY: Social Science Research Network, SSRN Scholarly Paper No. ID 1702854.
- Technopedia. (2018). Software Library Techopedia.com. <https://www.techopedia.com/definition/3828/software-library>. Zugegriffen: 20. Mai 2018.
- Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57(3), 41–51.
- Twitter. (2018). Twitter developer platform developer. <https://developer.twitter.com/content/developer-twitter/en.html>. Zugegriffen: 27. Mai 2018.
- Vu, T.-T., Chang, S., Ha, Q. T., & Collier, N. (2012). An experiment in integrating sentiment features for tech stock prediction in Twitter. In: Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data. Gehalten auf der Workshop on Information Extraction and Entity Analytics on Social Media Data, Mumbai, India: The COLING 2012 Organizing Committee. S. 23–38.
- Ward, J. S., & Barker, A. (2013). Undefined by data: A survey of big data definitions. University of St Andrews, UK, pp. 1–2. <https://arxiv.org/pdf/1309.5821.pdf>.
- Wingfield, N. (17. Mai 2018). Next up at Amazon-run whole foods: Half-priced Halibut steaks. *The New York Times*. <https://www.nytimes.com/2018/05/16/technology/whole-foods-discount-amazon-prime.html>. Zugegriffen: 3. Juni 2018.
- Wrobel, S., Voss, H., Köhler, J., Beyer, U., & Auer, S. (2015). Big data, big opportunities. *Informatik-Spektrum*, 38(5), 370–378.
- Wysocki, P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. Rochester, NY: Social Science Research Network, SSRN Scholarly Paper. <https://papers.ssrn.com/abstract=160170>. Zugegriffen: 16. Apr. 2018.
- Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter “I hope it is not as bad as I fear”. *Procedia – Social and Behavioral Sciences*, 26, 55–62.
- Zheludev, I., Smith, R., & Aste, T. (2014). When Can Social Media Lead Financial Markets? In: *Scientific Report no. 4*, pp. 1–12. <https://doi.org/10.1038/srep04213>.

Baumann, Flurina (M.Sc. BA/flurina.baumann@gmail.com) Berufsbegleitendes konsekutives Masterstudium in Business Administration mit Vertiefung Business Development an der Berner Fachhochschule. Mehrjährige Tätigkeit im Bankensektor, Grosshandel und Industrie u. a. als Marketing Manager und ICT Business Analyst. Aktuell Führungsfunktion in der IT eines internationalen Industrieunternehmens.

Brunner, Nadine Belinda (M.Sc. BA/nadinebelinda.brunner@gmail.com) Berufsbegleitendes konsekutives Masterstudium in Business Administration mit Vertiefung Business Development an der Berner Fachhochschule. Mehrjährige Tätigkeit im Immobilien und Non-Profit Sektor, u. a. als Betriebswirtschaftliche Leitung. Aktuell Führung eines Standorts eines der grössten und führenden medizinischen Unternehmungen in der Schweiz.

Tokarski, Kim Oliver (Prof. Dr./kim.tokarski@bfh.ch) Professor für Entrepreneurship und Unternehmensführung und Leiter der Abteilung Weiterbildung am Departement Wirtschaft der Berner Fachhochschule. Er ist Gastprofessor an der Wirtschaftsuniversität Bukarest (ASE Bukarest), Rumänien. Seine Forschungs- und Lehrtätigkeiten liegen in den Bereichen Entrepreneurship, Social Entrepreneurship, Innovation, Digitalisierung, Strategie sowie Leadership. Er ist Autor unterschiedlicher Publikationen in den genannten Themenkontexten.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäss nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

