

Korrelation von Social Sentiments und Kursentwicklungen an den Börsen

Bachelorarbeit

eingereicht von: **Daniel Huk**
Matrikelnummer: 01425743

im Fachhochschul-Bachelorstudiengang Wirtschaftsinformatik (0470)
der Ferdinand Porsche FernFH

zur Erlangung des akademischen Grades eines
Bachelor of Arts in Business

Betreuung und Beurteilung: MMag. Vera Glinz

Wiener Neustadt, 02.10.2022

Ehrenwörtliche Erklärung

Ich versichere hiermit,

1. dass ich die vorliegende Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Inhalte, die direkt oder indirekt aus fremden Quellen entnommen sind, sind durch entsprechende Quellenangaben gekennzeichnet.
2. dass ich diese Bachelorarbeit bisher weder im Inland noch im Ausland in irgendeiner Form als Prüfungsarbeit zur Beurteilung vorgelegt oder veröffentlicht habe.

Wien, 02.10.2022

Signiert von: Daniel Michael Huk
Datum: 03.10.2022 00:02:58
<small>Dieses mit einer qualifizierten elektronischen Signatur versehene Dokument hat gemäß Art. 25 Abs. 2 der Verordnung (EU) Nr 910/2014 vom 23. Juli 2014 ("eIDAS-VO") die gleiche Rechtswirkung wie ein handschriftlich unterschriebenes Dokument.</small>
Dieses Dokument ist digital signiert!
<small>Prüfinformation: Informationen zur Prüfung der elektronischen Signatur finden Sie unter: www.handy-signatur.at</small>



Creative Commons Lizenz

Das Urheberrecht der vorliegenden Arbeit liegt bei beim Autor. Sofern nicht anders angegeben, sind die Inhalte unter einer Creative Commons <„Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz“ (CC BY-NC-SA 4.0)> lizenziert.

Die Rechte an zitierten Abbildungen liegen bei den in der jeweiligen Quellenangabe genannten Urheber*innen.

Die Kapitel 2 bis 3 der vorliegenden Bachelorarbeit wurden im Rahmen der Lehrveranstaltung „Bachelor Seminar 1“ eingereicht und am 16.09.2022 als Bachelorarbeit 1 angenommen.

Kurzzusammenfassung: Korrelation von Social Sentiment und Kursentwicklungen an den Börsen

Mit zunehmender Digitalisierung und der vermehrten Nutzung von sozialen Medien ergibt sich die Möglichkeit die öffentlich zugänglichen Inhalte zu analysieren und Schlüsse hinsichtlich der Einstellung von Nutzern zu bestimmten Themen zu ziehen. Dieser Umstand ist auch im finanziellen Kontext, etwa im Bezug auf Aktienkurse, relevant. Das Ziel dieser Arbeit ist es zu beantworten, ob es einen Zusammenhang zwischen dem Social Sentiment von Tweets zu österreichischen, börsennotierten Unternehmen und deren Kursänderungen and der Börse gibt. Um diese Frage zu beantworten, wurde eine quantitative Untersuchung von Kursänderungen und Tweets, die mit solchen Unternehmen in Verbindung gebracht werden können, vorgenommen. Es wurden Tweets in einem Zeitraum von drei Jahren extrahiert und hinsichtlich ihres Sentiments nach einem lexikonbasierten Ansatz bewertet. Danach wurde versucht Korrelationen zwischen dem täglichen Sentiment und täglichen Kursänderungen zu finden. Das Ergebnis war, dass entgegen der Hypothese, keine statistisch signifikanten, linearen Korrelationen gefunden wurden.

Schlagwörter:

Social Sentiment, Sentimentsanalyse, Kursanalyse, Twitter-Sentiment, lexikonbasierter Ansatz

Abstract: Correlations between social sentiment and price movements in stock markets

With increasing digitalization and usage of social media there is a growing number of opportunities to analyse this publicly available data and draw conclusions concerning the attitude of users towards certain topics. This fact is also relevant in financial context. The goal of this thesis is to answer the question if there is a relation between the social sentiment of tweets targeting Austrian companies listed in the stock exchange and their fluctuation in stock price. In order to answer this question, a quantitative examination of stock price changes and tweets that can be related to those companies has been performed. Tweets within a time range of three years have been extracted and their sentiment was valued by applying a lexicon-based approach. Afterwards it has been tried to evaluate possible existing correlations. The result was that, against the hypothesis, no statistically significant, linear correlations were found.

Keywords:

Social sentiment, sentiment analysis, stock price analysis, twitter sentiment, lexicon-based approach

Inhaltsverzeichnis

1. EINLEITUNG	1
1.1 Problemstellung	1
1.2 Forschungsfrage und Ziel	1
2. AKTUELLER FORSCHUNGSSTAND	3
2.1 Definition Sentiment und Social Sentiment	3
2.2 Random-Walk-Theorie	4
2.3 Markteffizienzhypothese	4
2.4 Behavioral Finance und Homo Oeconomicus	5
2.5 Stimmungsanalyse	5
2.6 Sentimentsanalyse	6
3. METHODEN ZUR ERMITTLUNG DER DATEN	10
3.1 Auswahl eines Stichworts und Unternehmens	10
3.2 Bestimmung des Social Sentiments	11
3.3 Ermittlung des Kursverlaufs	13
3.4 Gegenüberstellung der Werte und Finden von Korrelationen	13
4. EMPIRISCHER TEIL DER ARBEIT	14
4.1 Bestimmung des Social Sentiment	14
4.1.1 Auswahl der Stichwörter	14
4.1.2 Abfrage der Twitter-Daten	15
4.1.3 Auswertung der Twitter-Daten	15
4.2 Ermittlung des Kursverlaufes	16
5. ANALYSE DER ERGEBNISSE	18
6. SCHLUSSFOLGERUNG	25
7. ZUSAMMENFASSUNG	26
8. AUSBLICK	27

9. ANHÄNGE	28
9.1 Programm zur Sentiment-Bestimmung	28
9.2 Programm zur Zusammenführung von Sentiment und Kursänderung	32
10. LITERATURVERZEICHNIS	35

1. Einleitung

1.1 Problemstellung

Das Prognostizieren von Kursentwicklungen eines bestimmten Unternehmens erweckt bei Anlegern großes Interesse, da es dazu genutzt werden kann die Gewinne durch eine, auf die Prognose angepasste, Anlagestrategie zu erhöhen. Als Basis für solche Prognosen werden oft Indikatoren und Signale verwendet, die als Basis den Kursverlauf an sich sowie Unternehmenszahlen mathematisch auswerten. Anhand dieser Daten kann etwa bestimmt werden, ob ein Wertpapier rein rechnerisch unter- oder überbewertet ist. Außerdem können bei Prognosen auch facheinschlägige Nachrichten und Veröffentlichungen des Unternehmens berücksichtigt werden, da auch diese in vielen Fällen den Kursverlauf kurzfristig beeinflussen. Die Zuverlässigkeit solcher Prognoseinstrumente hängt jedoch stark davon ab, wie rational sich die Mehrzahl der Anleger beim Handel tatsächlich verhalten. Es ist davon auszugehen, dass ein überwiegender Teil der Anleger irrational, auf Basis persönlicher oder öffentlicher Präferenzen und Meinungen, handelt. Soziale Medien bieten die Möglichkeit solche persönlichen Meinungen zu teilen und öffentlich zugänglich zu machen. Dadurch besteht die Möglichkeit diese geteilten Inhalte auszuwerten, um ein Sentiment daraus zu extrahieren und gegebenenfalls Zusammenhänge mit Kursverläufen zu erkennen. Dabei ergibt sich das Problem wie die große Menge an geteilten Inhalten zu durchsuchen bzw. zu filtern ist. Ein weiteres Problem stellt die Interpretation der geteilten Inhalte, einzeln und in ihrer Gesamtheit, sowie das Erkennen von Zusammenhängen mit dem Kursverlauf, dar.

1.2 Forschungsfrage und Ziel

Im Rahmen dieser Arbeit soll folgende Forschungsfrage, die sich aus der Problemstellung ergibt, beantwortet werden:

„Gibt es eine Korrelation zwischen dem Social Sentiment zum Stichwort A und der Kursentwicklung eines börsennotierten Unternehmens, welches mit dem Stichwort A in Verbindung gebracht werden kann?“

Die Hypothese ist, dass es zwischen dem Social Sentiment eines relevanten Stichwortes und dem Kursverlauf einer dazugehörigen Unternehmensaktie an der Börse, Korrelationen gibt.

Das Ziel dieser Arbeit ist es auf die Methode zur Auswahl eines passenden Stichwortes zu einem Wertpapier sowie zur Ermittlung des Social Sentiments aus den verfügbaren Inhalten zu dem gewählten Stichwort, in sozialen Medien einzugehen. Weiters soll ermittelt werden ob und welche Korrelationen zwischen dem ermitteltem Social Sentiment und dem Kursverlauf bestehen. Dazu werden im empirischen Teil der Arbeit mehrere börsennotierte Unternehmen ausgewählt und dazu passende Stichwörter ermittelt. Anhand der Stichwörter werden Daten des sozialen Netzwerkes „Twitter“ abgefragt, bereinigt und ausgewertet. Zur Bestimmung des Sentiments wird eine lexikonbasierte Methode verwendet. Die sich ergebenden Verläufe des Social Sentiments werden dann den einzelnen Kursverläufen gegenübergestellt, um Korrelationen zu identifizieren.

Um die Hypothese zu bewerten, wird im Rahmen dieser Arbeit im 2. Kapitel auf Begriffsdefinitionen, sowie bereits bestehende Lösungsansätze, die sich in der wissenschaftlichen Literatur finden, eingegangen. Im 3. Kapitel werden Methoden zur Auswahl einer Datenbasis von Kursverläufen und Stichwörtern erörtert. Weiters wird die anzuwendende Methode zur Ermittlung des Social Sentiments aus der Datenbasis heraus beschrieben. Im 4. Kapitel befindet sich der empirische Teil der Arbeit in dem die Methode praktisch angewandt wird. Im 5. Kapitel werden die gewonnenen Erkenntnisse zusammengefasst.

2. Aktueller Forschungsstand

2.1 Definition Sentiment und Social Sentiment

Die grundlegendste Beschreibung von Sentiment ist, dass es sich dabei um eine positive oder negative Einstellung oder Meinung zu einem bestimmten Thema oder einer bestimmten Sache handelt. Praktisch ist das Sentiment nicht binär, sondern verläuft von äußerst negativ über neutral bis hin zu äußerst positiv.¹ Unter Social Sentiment versteht man die, über soziale Medien, in Textform, zum Ausdruck gebrachte Einstellung zu einem Thema. Es handelt sich dabei nicht um das Sentiment eines einzelnen Individuums, sondern um die Bestimmung des Sentiments aus der Gesamtheit aller relevanten Inhalte, die von Benutzern geteilt wurden. Als Basis, zur Abfrage der auszuwertenden Daten, dienen alle sozialen Medien, die sich nach Schlüsselwörtern durchsuchen lassen. Dabei ergibt sich die Herausforderung ein passendes Schlüsselwort zu finden sowie die korrekte Bewertung und Gewichtung der einzelnen, zum Schlüsselwort passenden, Inhalte.²

Ein wichtiges Merkmal des Sentiments ist, dass es keine kurzfristigen Emotionen, sondern ausschließlich die Einstellung einer Person zu einem bestimmten Thema beschreibt. Damit grenzt sich etwa die Sentimentsanalyse klar von der Emotionsanalyse oder Stimmungsanalyse ab, bei der es darum geht, den allgemeinen Gemütszustand einer Person zu erfassen.³ Weiters ist es nicht das Ziel zu ermitteln, ob viel oder wenig zu einem Thema kommuniziert oder diskutiert wird, sondern wie die geteilten Inhalte insgesamt zu bewerten sind. Dadurch unterscheidet sich Social Sentiment klar von „Buzz“. Bei der „Buzz“-Analyse geht es ausschließlich darum wieviel Aufmerksamkeit ein bestimmtes Schlüsselwort in einem Zeitraum erhält bzw. wie oft es in Inhalten in sozialen Medien vorkommt.⁴

Die Analyse von Sentiment und die Interpretation von Meinungen zu bestimmten Themen ist nicht erst mit der Verbreitung moderner Technologien, wie etwa dem Internet, entstanden. Schon davor war die Sentimentsanalyse ein Teil der Forschung

¹ (Moshe Koppel, Jonathan Schler 2006)

² (Kim und Hovy 2004)

³ (Scherer 2003)

⁴ (Zhang, Fuehres und Gloor 2012)

in der Linguistik und Sprachverarbeitung. Erste Wissenschaftliche Texte zu diesem Thema sind bereits in den 1990er-Jahren erschienen. Seit dem Jahr 2000 hat die Forschung im Bereich des Social Sentiment stark zugenommen. Grund dafür sind einerseits die vielseitigen Einsatzmöglichkeiten im kommerziellen Bereich als auch die wachsende Datenbasis, besonders durch das Aufkommen sozialer Netzwerke im Internet.⁵ Ohne diese Datenbasis wäre zwar die punktuelle Analyse des Sentiments einzelner Personen oder Texte zu bestimmten Themen möglich, es ließe sich jedoch in der Praxis kein umfassendes oder gar globales Social Sentiment bestimmen. In den nachfolgenden Kapiteln wird davon ausgegangen, dass der Begriff „Sentiment“ sich immer auf das Social Sentiment bezieht.

2.2 Random-Walk-Theorie

Die Random-Walk-Theorie geht davon aus das Aktienkurse eine niedrige Volatilität haben und jede Kursschwankung unabhängig vom vorherigen Kursverlauf passiert. Damit würde sich der Preis einer Aktie immer gänzlich zufällig, in einer geringen Schwankungsbreite um den fundamental fairen Wert bewegen, ganz unabhängig davon ob neue relevante oder irrelevante Informationen, die die Aktie betreffen, zur Verfügung stehen.⁶

2.3 Markteffizienzhypothese

Ähnlich wie in der Random-Walk-Theorie wurde angenommen, dass Aktienkurse nicht vorhersehbar sind. Diese Annahme ist auch Teil der Markteffizienzhypothese, die beschreibt, dass Aktienkurse auf relevante verfügbare Informationen oder Nachrichten reagieren und diese unmittelbar „einpreisen“, dabei jedoch einen zufälligen Verlauf haben, da auch die Nachrichten nicht vorhergesagt werden können. Dadurch werden Prognosen für die Zukunft auf Basis von Nachrichten praktisch unmöglich, und die neuen Informationen, die das Wertpapier betreffen, sind zu jedem Zeitpunkt, bereits im Preis berücksichtigt.⁷

⁵ (Liu 2012)

⁶ (Walter 2000)

⁷ (Fama 1970)

2.4 Behavioral Finance und Homo Oeconomicus

Sowohl die Random-Walk-Theorie als auch die Markteffizienzhypothese gehen davon aus, dass alle Akteure rational und maximal gewinnorientiert handeln. Dazu stehen den Akteuren alle Informationen zur Verfügung und die Entscheidungen im Handel werden emotionslos getroffen. Die Akteure erfüllen damit theoretisch die Eigenschaften des „Homo Oeconomicus“, der lange Zeit die klassische Theorie im Finanzsektor darstellte. Neuere Forschungsergebnisse waren mit diesen theoretischen Annahmen jedoch nicht zu vereinen⁸ und es wurde versucht diese Anomalien mithilfe der Psychologie und Sozialwissenschaften zu erklären. Es konnte dabei gezeigt werden, dass sich die Finanzmärkte nicht vollständig rational verhalten und Anleger in vielen Fällen etwa auf ihre Intuition, die Angst vor einer Enttäuschung oder auf Vorurteile vertrauen.⁹

2.5 Stimmungsanalyse

In neueren Forschungsarbeiten wurde festgestellt, dass der Einfluss von Nachrichten auf den Kurs zwar nicht vorhersehbar ist, es jedoch Indikatoren gibt, die sich aus sozialen Medien extrahieren lassen, um Prognosen zu treffen. So wurde etwa die öffentliche Stimmung mithilfe von Twitter-Inhalten ausgewertet und festgestellt das eine Änderung der Stimmung auch eine zeitversetzte Änderung der Tendenz des „Dow Jones Industrial Average“-Indexes nach sich zieht. Dazu wurden etwa 8 Millionen Tweets analysiert und in Stimmungskategorien unterteilt. Die sich ergebende Gesamtstimmung wurde mithilfe eines neuronalen Netzwerkes in Relation zum Indexkurs gesetzt. Konkret wurde dabei die Stimmung der einzelnen Inhalte in die Kategorien „Calm“, „Alert“, „Sure“, „Vital“, „Kind“ und „Happy“ unterteilt.¹⁰ Hierbei ist zu beachten das die Arbeit sich auf einen Indexkurs bezieht, der auf mehreren einzelnen Aktienkursen basiert und explizit nur die gefühlsbetonte Stimmung der Beiträge im sozialen Netzwerk berücksichtigt wurde, das Sentiment wurde dabei nicht analysiert.

⁸ (Schiller 2003)

⁹ (Byrne und Brooks 2008)

¹⁰ (Bollen, Mao und Zeng 2011)

2.6 Sentimentsanalyse

Die besondere Herausforderung bei der Analyse des Sentiments stellt sich darin, den Text der Inhalte in sozialen Netzwerken zu quantitativ zu bewerten. Eine Studie aus dem Jahr 2017 hat dazu eine Sammlung an Wörtern erstellt und jedes einzelne Wort im Intervall von $[-1,1]$ nach positivem Sentiment absteigend bewertet. Bindewörter wurden aus den Texten entfernt und Emojis wurden, je nach Art, mit einem positiven oder negativen Textbaustein ersetzt. Zur Erstellung der Wortsammlung wurden vergangene Beiträge mehrerer Monate analysiert, um die gängigsten Begriffe herauszufiltern. Das Erstellen der Wortsammlung mit einschlägigen Begriffen, die im Finanzbereich verwendet werden, hat sich, im Vergleich zu bereits existierenden Sammlungen ohne Spezialisierung, als zuverlässiger erwiesen. Grund dafür ist, dass einige Begriffe, die im Zusammenhang mit Finanzen als positiv einzuordnen sind, im generellen Sprachgebrauch als negativ betrachtet werden können. Beispielsweise sind die Wörter „dominieren“ oder „Überschuss“ in der deutschen Wortsammlung „SentiWS“ als schwach negativ bewertet, obwohl sie im finanziellen oder unternehmerischen Kontext als sehr positiv angesehen werden können. Auch in der umfangreicheren englischsprachigen Wortsammlung „Harvard IV“ gibt es Wörter die, im Hinblick auf den finanziellen Kontext, unpassend klassifiziert sind. Beispiele dafür sind etwa die negativ klassifizierten Wörter „capital“ und „excess“.¹¹ Generell gelten auch sarkastische Formulierungen oder negierte Ausdrücke als problematisch hinsichtlich der korrekten Bestimmung des Sentiments. Ungenauigkeiten bei der automatisierten Auswertung von Texten lassen sich daher nicht gänzlich verhindern.¹²

Mithilfe des „lexikonbasierten Ansatzes“, der auf der erstellten Wortsammlung aufbaut, wurden 9 Millionen geteilte Kurztexte der sozialen Plattform „StockTwits“ in einem 30-Minuten-Intervall ausgewertet. Das Sentiment jedes einzelnen Beitrags wurde aus der Summe der einzelnen Bewertung jedes Wortes errechnet. Der so gewonnene Sentimentsverlauf wurde mit dem Intraday-Kursverlauf des „S&P 500“-Index verglichen. Daraus ergab sich die Beobachtung, dass die Ausprägung des Kurses im ersten 30-Minuten-Intervalls des Tages einen Einfluss auf die Marktentwicklung während des letzten 30-Minuten-Intervalls des gleichen Tages hat. Die genaue Ursache dafür konnte nicht abschließend geklärt werden, es wird vermutet das der „Intraday Momentum Effect“ damit zu tun haben könnte. Dieser Effekt wird auf das Trading-Verhalten von professionellen Daytradern zurückgeführt

¹¹ (Loughran und McDonald 2011)

¹² (Loughran und McDonald 2016)

und ist statistisch relevant.¹³ Weiters wurde festgestellt, dass es einen Zusammenhang zwischen dem Sentiment und dem Kursverlauf während des Tages gibt. Dieser Zusammenhang variiert jedoch an verschiedenen Tagen, so sind etwa Tage an denen wenige Nachrichtenpublikationen in Verbindung mit dem „S&P 500“-Index erscheinen, mehr an das Sentiment gekoppelt.¹⁴ Eine weitere Studie zu diesem Thema untersucht ebenfalls den „S&P 500“-Index, verwendet jedoch als Basis den „MarketPsych Index“, kurz TRMI, von Thomson Reuters, der das Sentiment darstellt. Der TRMI-Index wird nach einem patentierten System ermittelt und steht täglich sowie minütlich seit 1998 zur Verfügung. In den TRMI-Index fließen nicht nur Beiträge aus zahlreichen sozialen Netzwerken, sondern auch Berichte aus Zeitungen und Zeitschriften. Wie in der vorher beschriebenen Studie wird auch hier in einem Intervall von 30 Minuten gearbeitet, weshalb die minütlichen Daten des TRMI-Indexes jeweils zusammengefasst werden müssen. Es wurde darauf geschlossen, dass der Kursverlauf des Tages in Relation zum Sentiment steht, wobei die Relation in den letzten zwei Stunden des Tages stärker ist. Außerdem scheint die Relation in den Sommermonaten generell stärker zu sein als in Wintermonaten, was jedoch nicht genau begründet werden konnte. Weiters wurde auch in dieser Studie festgestellt, dass das erste 30-Minuten-Intervall einen Einfluss auf das letzte 30-Minuten-Intervall des Tages hat.¹⁵ Komplexere Methoden zur Ermittlung des Sentiments verwenden Machine Learning um Zusammenhänge zwischen Texten und Kursverläufen zu erkennen. Der Algorithmus wird dabei mit großen Datensätzen trainiert und getestet. Erkannt werden im Gegensatz zum lexikonbasierten Ansatz nicht nur bestimmte Wörter, sondern der Gesamtkontext des Inhaltes. Die Ergebnisse sind, je nach Art und Umfang der Trainingsdaten und Wortwahl der zu analysierenden Beiträge, sehr unterschiedlich und lassen sich nicht gut reproduzieren. Ob der Ansatz des maschinellen Lernens einen Vorteil gegenüber dem lexikonbasierten Ansatz hat ist, besonders im Hinblick auf die deutlich größere Komplexität, nicht endgültig geklärt. Es wird jedoch davon ausgegangen, dass ein fachspezifisches Lexikon mit relevanten Wörtern und Abkürzungen akkuratere Ergebnisse bei weniger Aufwand und Komplexität liefern kann.¹⁶ Eine aktuelle Studie aus dem Jahr 2019 verwendet zur Vorhersage des Kursverlaufes des „FTSE 100“-Index ebenfalls den lexikonbasierten Ansatz und bildet daraus den „Twitter Sentiment Score“. Zur Ermittlung dieses Wertes

¹³ (Gao, et al. 2014)

¹⁴ (Renault 2017)

¹⁵ (Sun, Najand und Shen 2015)

¹⁶ (Loughran und McDonald 2016)

wurden zuerst Tweets in einem bestimmten Zeitraum und mit einem bestimmten Keyword, in diesem Fall zum englischen Wort für Aktie, „stock“, gesammelt. Aus den Tweets heraus wurden Wörter, die am meisten darin vorkamen gesammelt. Diese Wörter wurden manuell hinsichtlich ihres enthaltenen Sentiments bewertet. Die Tweets wurden danach auf Basis der erstellten Wortsammlung bewertet und ergaben so das Sentiment. Dabei wurden ein zeitversetzter Zusammenhang zwischen Kursverlauf und Sentiment festgestellt. Besonders deutlich war der Zusammenhang bei steigendem Kurs. In weiterer Folge wurde ein Preisindikator erstellt, der als Signal für Anleger dienen kann. Der Indikator zeigt mit einer durchschnittlichen Trefferquote von etwa 67% und etwa 98% in Phasen mit steigendem Kurs an, ob der Markt in den nächsten 30 Stunden steigen oder fallen wird. Das Signal ist positiv bzw. sagt einen steigenden Kurs voraus, wenn der ermittelte „Twitter Sentiment Score“ einen bestimmten Wert, die „Baseline“ überschreitet. Das Signal ist negativ bzw. sagt einen fallenden Kurs voraus, wenn der „Twitter Sentiment Score“ einen bestimmten Wert unterschreitet. Die Auswahl dieses Schwellenwertes ist dabei entscheidend für die Zuverlässigkeit des Signals.¹⁷

Eine Studie aus dem Jahr 2020 beschäftigt sich mit dem Einfluss der H1N1- und COVID19-Pandemie auf das Social Sentiment und in weiterer Folge auf die Kurse verschiedener Indexes. Hierfür wurde ein lexikonbasierter Ansatz verwendet. Tweets wurden in den Zeiträumen von Juni bis Juli 2009 und Januar bis Mai 2020 gesammelt, bereinigt und mithilfe mehrerer verschiedener Wortsammlungen bewertet. Da nicht jede Wortsammlung das Sentiment im Intervall [-1, 1] ausgibt, mussten die Ergebnisse entsprechend umgerechnet werden. Die Indexkurse wurden mit dem Verlauf des Sentiments verglichen, um Zusammenhänge zu finden. Dabei wurde auch darauf geachtet, zeitversetzte Zusammenhänge zu untersuchen. Es wurde festgestellt, dass es Zusammenhänge zwischen dem Sentiment und den Indexkursen gibt, die teilweise mehrere Tage in die Vergangenheit oder in die Zukunft reichen. Im Vergleich zur H1N1-Pandemie war der Zusammenhang in den ersten Monaten der COVID19-Pandemie deutlich ausgeprägter. Das konnte damit erklärt werden, dass es während der H1N1-Pandemie weniger Beiträge auf Twitter dazu gab. Weiters wurde festgestellt, dass sich die maximale Zeitversetzung zwischen Sentiment und Kursverlauf im Vergleich zur H1N1-Pandemie, von maximal 15 auf maximal 10 Tage verkürzt hat, was darauf zurückgeführt werden könnte, dass soziale Medien in den letzten zehn Jahren noch mehr an Bedeutung gewonnen haben.¹⁸

¹⁷ (Guo und Jinfeng 2019)

¹⁸ (Valle-Cruz, et al. 2020)

Die Analyse von Sentiment als Grundlage für Entscheidungen, insbesondere im finanziellen und unternehmerischen Bereich, ist von hoher Relevanz und wird mit fortschreitender Digitalisierung immer wichtiger. Daher entstehen auch Kooperationen, wie etwa das 2015 unterzeichnete „Bloomberg and Twitter Data Licensing Agreement“, welches es dem Unternehmen Bloomberg erlaubt die Daten auf der Plattform Twitter umfassend zu analysieren und in seine eigenen Nachrichtenkanäle und Dashboards einzubinden. Mithilfe des von Bloomberg angebotenen „Professional Service“ können Anleger nicht nur den Kurs in Echtzeit, sondern auch die Anzahl an Tweets und das aktuelle Sentiment eines Börsenwertes im Auge behalten und in Entscheidungen mit einfließen lassen. Außerdem kann der Benutzer bei Änderungen in der Anzahl von Tweets oder dem Sentiment benachrichtigt werden sowie die relevantesten Tweets an sich im Auge behalten, was schnellere Reaktionen auf Änderungen am Markt ermöglicht.¹⁹

¹⁹ (Bloomberg L.P. 2015)

3. Methoden zur Ermittlung der Daten

Die bisherigen Erkenntnisse aus Kapitel 2 beziehen sich nicht auf einzelne Aktien, sondern auf Indexkurse. Der Fokus liegt zumeist auf dem Kurs- und Sentimentsverlauf innerhalb eines Tages. Außerdem werden ausschließlich englischsprachige Daten ausgewertet, da sich die untersuchten Indexes im englischsprachigen Raum befinden.

Um untersuchen zu können ob und wie sich das Social Sentiment in Relation zu dem Kurs eines dazugehörenden Unternehmens verhält, werden eine Menge an Daten benötigt. Zuerst muss ein Unternehmen, welches Teil des ATX ist, gewählt werden. Weiters müssen Stichwörter bestimmt werden, die zum Unternehmen passen. Als Nächstes müssen die Beiträge im sozialen Netzwerk anhand der gewählten Stichwörter gefiltert werden. Es muss manuell validiert werden, ob die gewonnenen Suchergebnisse tatsächlich auf das Unternehmen bezogen sind. Aus den Beiträgen heraus wird mithilfe eines lexikonbasierten Ansatzes das Sentiment im Intervall von einem Tag ermittelt. In einem nächsten Schritt muss der Kursverlauf des Unternehmens, ebenfalls im Intervall von einem Tag, ermittelt werden. Aus den sich ergebenden Zeitreihen kann ermittelt werden, ob die Bewegung des Kurses mit der Bewegung des Sentiments korreliert.

3.1 Auswahl eines Stichworts und Unternehmens

Postings in sozialen Netzwerken enthalten nicht unbedingt den Namen des jeweiligen Unternehmens, sondern die Kurzformen, den ISIN oder andere Schlagworte, die mit dem Unternehmen in Verbindung gebracht werden können. Konkret werden Posts des sozialen Netzwerks „Twitter“ analysiert. Dazu wird die erweiterte Suchfunktion des sozialen Netzwerks verwendet. Die Suche nach Stichwörtern verläuft dabei explorativ. Es werden unterschiedliche Wörter über die Suchmaske abgefragt und die Ergebnisse werden manuell evaluiert. Diese manuelle Evaluation besteht praktisch darin zu erkennen ob die einzelnen Beiträge, die sich aus der Suche ergeben auch tatsächlich etwas mit dem Unternehmen zu tun haben. Es wurden dazu bewusst mehrere Unternehmen verschiedener Branchen gewählt, um Aufschluss über die Reproduzierbarkeit des Ergebnisses bzw. die Unterschiede im Sentiment und Kursverlauf der verschiedenen Unternehmen am Ende der Arbeit zu bekommen.

Folgende Unternehmen sollen untersucht werden:

Unternehmen	ISIN	Branche	NACE ²⁰ / GICS ²¹
OMV AG	AT0000743059	Fossile Rohstoffe	B 06 / 10101010
Raiffeisen Bank International AG	AT0000606306	Banken	K 64 / 40101010
Andritz AG	AT0000730007	Maschinenbau	C 28 / 20106020
Voestalpine AG	AT0000937503	Metallproduktion	C 24 / 15104020

Tabelle 1 – Zu untersuchende Unternehmenswerte

3.2 Bestimmung des Social Sentiments

Als Basis für die Interpretation der Beiträge wird hierbei die deutsche Wortsammlung „SentiWS“ in der aktuellen Version 2.0 vom Oktober 2018 verwendet. Diese Wortsammlung ist frei zugänglich und wird stetig aktualisiert. Die Wortsammlung wird in Form von Textdateien zur Verfügung gestellt und enthält 3450 Wörter in ihrer Grundform, was insgesamt 34000 Wortformen ergibt. Enthalten sind nicht nur Adjektive sondern auch Nomen, Verben und Adverbien, die zur Bestimmung des Sentiments beitragen. Andere Wortsammlungen wie etwa „McDonald Financial Sentiment Dictionaries“ konzentrieren sich zwar auf Wörter mit finanziellem Kontext, was vorteilhaft ist, sind aber nur in englischer Sprache verfügbar.²² Eine simple Übersetzung der Wörter, ohne manuelle Überarbeitung und erneute Evaluation des Sentiments kommt nicht infrage, da sich Wortwahl und Sentiment von Sätzen in den unterschiedlichen Sprachen unterscheiden. Die „SentiWS“-Wortsammlung basiert auf dem „General Inquirer“, einer der ersten Wortsammlungen mit Sentimentsdaten. Diese Sammlung wurde übersetzt, manuell gefiltert und das Sentiment der einzelnen Wörter wurde im Zusammenhang mit der deutschen Sprache bestimmt. Außerdem wurden bei der Erstellung von „SentiWS“ auch zusätzliche Begriffe mit finanziellem Kontext eingearbeitet.²³

Der syntaktische Aufbau der einzelnen Wörter in den Textdateien sieht wie folgt aus:

WORT|WORTART SENTIMENT WORTFORM₁, ..., WORTFORM_n

²⁰ (EU NACE v2 2022)

²¹ (MSCI INC. 2018)

²² (Sardelich und Kazakov 2018)

²³ (Remus, Quasthoff und Heyer 2010)

Die Wortart gibt an, ob es sich um Nomen, Verben, Adverbien oder Adjektive handelt. Das Sentiment ist im Intervall [-1,1] auf vier Nachkommastellen genau angegeben, wobei 1 den positivsten und -1 den negativsten Wert darstellt. Für jedes Wort kann es eine beliebige, durch die Grammatik vorgegebene, Anzahl an Formen geben, die das Wort im Kontext eines Satzes annehmen kann.

Folgendes Beispiel zeigt den Datensatz für das Verb „abwerten“:

abwerten|VVINF -0.4521 abwertest,abgewertet,abwerttest,abzuwerten,
abwertete,abwerteten,abwerte,abwertetet,abwertet,abwertetest

Um das Sentiment auf täglicher Basis zu berechnen, werden alle Tweets des jeweiligen Tages, die aufgrund der Kriterien in Kapitel 3.1 gefunden werden, maschinell ausgewertet. Die Tweets werden von Emojis und Symbolen bereinigt um zu verhindern, dass deren übermäßige Verwendung das Ergebnis verfälscht. Die Vorgehensweise ist dahingehend gleich wie in der ersten Publikation aus Punkt 2.6.²⁴ Die einzelnen Tweets werden nach der Bereinigung, durch Beistriche getrennt, in eine Datei gepackt. Mithilfe eines Java-Programms werden die einzelnen Tweets durchgegangen und jedes einzelne Wort entsprechend der „SentiWS“-Wortsammlung bewertet. Wörter, die nicht in der Wortsammlung vorkommen, werden dabei nicht berücksichtigt. Das Sentiment eines Tweets (S_T) ergibt sich dabei aus dem Durchschnitt der Sentiments der einzelnen Wörter (w) des Tweets, wie die folgende Formel zeigt.

$$S_T = \frac{1}{n} \sum_{i=1}^n w_i$$

Das Sentiment des gesamten Tages (S_D) ergibt sich aus dem Durchschnitt der Sentiments aller Tweets des gleichen Tages, wie die folgende Formel zeigt.

$$S_D = \frac{1}{n} \sum_{i=1}^n S_{T_i}$$

²⁴ (Loughran und McDonald 2011)

3.3 Ermittlung des Kursverlaufs

Der Kursverlauf, der in Kapitel 3.1 definierten Unternehmen, kann lückenlos nachverfolgt und exportiert werden. Der Export erfolgt direkt vom Onlineportal der Wiener Börse im CSV-Format. Die täglichen Daten enthalten den Eröffnungspreis, das Tageshoch, das Tagestief sowie den Schlusspreis und die prozentuelle Differenz. Im Gegensatz zu den in Punkt 2.6 beschriebenen Untersuchungen ist keine Erfassung der Kursänderung während des laufenden Tages notwendig. Es wird lediglich auf die Kursänderung des gesamten Tages eingegangen, was eine eigene Vorgehensweise darstellt. Für die Gegenüberstellung der Kursentwicklung des jeweiligen Tages mit dem Sentiment des jeweiligen Tages ist die prozentuelle Differenz des Eröffnungs- und Schlusspreises ausreichend, die anderen Spalten werden deshalb für die weitere Verwertung verworfen.

3.4 Gegenüberstellung der Werte und Finden von Korrelationen

Aus den Analysen in Kapitel 3.2 und 3.3 wird, getrennt für jedes einzelne Unternehmen aus Kapitel 3.1, eine Zeitleiste gebildet. Zu jedem Tagesdatum ist die prozentuelle Entwicklung des Kurses sowie das ermittelte Sentiment bekannt. Mithilfe der Kreuzkorrelationsfunktion kann bestimmt werden ob und wann sich die Verläufe der beiden Datenreihen ähneln. Das Ergebnis ist der Korrelationskoeffizient, jedes einzelnen Tages, der innerhalb des Intervalls $[-1,1]$ liegt. Steigt also das Social Sentiment im exakt gleichen Ausmaß wie der Kurs, ergibt sich der Wert 1, bewegen sich die Werte in entgegengesetzte Richtungen ergibt sich ein Wert von -1. Als Hilfsmittel zur Berechnung bzw. zur Anwendung der Kreuzkorrelationsfunktion und zur grafischen Darstellung der Ergebnisse wird ein Tabellenkalkulationsprogramm, in diesem Fall Microsoft Excel, verwendet.

4. Empirischer Teil der Arbeit

Im folgenden empirischen Teil der Arbeit werden die in Kapitel 3 erörterten Lösungsvorschläge praktisch umgesetzt. Dazu werden zuerst Stichwörter zu den einzelnen zu untersuchenden Unternehmen ausgewählt. Zu diesen Stichwörtern werden alle Tweets abgefragt. Die Tweets werden dann mithilfe eines Java-Programms unter Verwendung der Senti-WS-Wortsammlung ausgewertet.

4.1 Bestimmung des Social Sentiment

4.1.1 Auswahl der Stichwörter

Da es sich um konkrete Unternehmen handelt, sind die Möglichkeiten der Stichwortauswahl beschränkt. Es gilt hierbei entsprechende Synonyme und insbesondere Abkürzungen zu finden, die zum Unternehmen passen und brauchbare Suchergebnisse liefern. Zusätzlich wird bei der Suche auf den zu betrachtenden Zeitraum von 3 Jahren, also den Zeitraum von 01.08.2019 bis 31.07.2022 eingeschränkt. Antworten auf Tweets werden ebenfalls einbezogen. Weiters wird auf deutschsprachige Inhalte gefiltert. Daraus ergibt sich eine Twitter-Suchabfrage, die in der Twitter-Suchmaske ausgeführt wird. Die Groß- und Kleinschreibung wird bei der Suche nicht berücksichtigt. Der Zeitraum von drei Jahren ist bewusst gewählt, um eine ausreichende Menge an Daten gewinnen zu können. Bei längeren Zeiträumen würde insbesondere der Twitter-Datenexport viel Zeit in Anspruch nehmen.

Unternehmen	Stichwörter	Suchabfrage	Anzahl Tweets
OMV AG	OMV, OEMV, ÖMV, AT0000743059	(omv, OR oemv, OR ömv, OR at0000743059) lang:de until:2022-08- 01 since:2019-08-01	19.638 Ø 26 / Tag
Raiffeisen Bank International AG	Raiffeisen, Raiffeisenbank, Raiffeisenkasse, Raiffeisenlandesbank, AT0000606306	(raiffeisen, OR raiffeisenbank, OR raiffeisenkasse, OR raiffeisenlandesbank, OR at0000606306) lang:de until:2022-08- 01 since:2019-08-01	11.806 Ø 16 / Tag
Andritz AG	Andritz, AT0000730007	(andritz, OR AT0000730007) lang:de until:2022-08-01 since:2019-08-01	5.107 Ø 7 / Tag

Voest Alpine AG	Voest Alpine, Vöest Alpine, Voest, Vöest, AT0000937503	(voest Alpine, OR vöest Alpine, OR voest, OR vöest, OR at0000937503) lang:de until:2022-08-01 since:2019- 08-01	9.012 Ø 12 / Tag
-----------------	--	--	---------------------

Tabelle 2 - Suchabfragen

4.1.2 Abfrage der Twitter-Daten

Da die Twitter-API nur Abfragen für den Zeitraum der letzten 7 Tage zulässt, ist es notwendig die Tweets mithilfe eines Skripts zu extrahieren. Das dazu angefertigte Skript wird im Browser ausgeführt, scrollt automatisch durch die Suchergebnisse und gibt die gefundenen Tweets im Textformat aus. Die so gewonnenen Ergebnisse werden in das Tabellenkalkulationsprogramm Microsoft Excel übertragen und für die automatisierte Auswertung im CSV-Format exportiert. Das CSV enthält eine Spalte für das Datum und eine Spalte für den Text des Tweets.

4.1.3 Auswertung der Twitter-Daten

Da die manuelle Auswertung der Tweets, aufgrund der großen Menge, sehr langwierig ist kommt hierzu ein eigens angefertigtes Programm²⁵ zum Einsatz. Das Programm nimmt den Pfad der im Punkt 4.1.2 exportierten Daten entgegen und liest diese ein. Weiters wird der Pfad zur SentiWS-Wortsammlung sowie die Ausgabepfade als Parameter an das Programm übergeben. Das Programm durchläuft alle Tweets sowie jedes einzelne Wort im Tweet und versucht ein passendes Wort in der Wortsammlung zu finden. Dazu werden im Vorfeld die Worte, Wortformen und das Sentiment aus der Wortsammlung ausgelesen und in einer Matrix bereitgehalten. Voraussetzung für die Zuordnung eines Wortes aus einem Tweet ist nicht ausschließlich Gleichheit mit einem Wort in der Wortsammlung. Um möglichst viele Wörter zuordnen zu können wird für jedes zu vergleichende Wortpaar die Ähnlichkeit mithilfe des Jaro-Winkler-Algorithmus errechnet. Dieser Algorithmus bestimmt den Bearbeitungsabstand, also die Anzahl an unterschiedlichen Zeichen, die die beiden zu vergleichenden Zeichenketten unterscheiden. Zusätzlich werden diese Unterschiede gewichtet, da meist Abweichungen am Beginn oder am Ende eines Wortes für die Bedeutung des Wortes ausschlaggebender sind. Je geringer dieser Abstand ist, desto höher ist die Ähnlichkeit der beiden Zeichenketten. Die

²⁵ Anhang 9.1 – Programm zur Sentiment-Bestimmung

Ähnlichkeit wird im Intervall [0,1] bestimmt wobei der Wert 1 für Gleichheit steht.²⁶ Um als Treffer zu gelten, muss dieser Wert größer als 0,97 sein. Eine Reduktion dieser Variable erhöht zwar die Fehlertoleranz, verfälscht dabei aber auch das Ergebnis. Konkrete Beispiele sind etwa die Wortpaare „beendet“ und „beendte“ (Ähnlichkeit 0,9714), „gültigen“ und „gültigen“ (Ähnlichkeit 0,9708), „ausgleichen“ und „ausgelichen“ (Ähnlichkeit 0,9818) oder „korrekt“ und „korrek“ (Ähnlichkeit 0,9714), die aufgrund von verdrehten Buchstaben, ohne die Verwendung des Algorithmus, nicht bewertet worden wären. Wird ein Wort gefunden, so wird der Wert des Sentiments für dieses Wort summiert. Das Sentiment eines Tweets ergibt sich dann aus der Summe der Sentiment-Werte geteilt durch die Anzahl der Wörter, also dem Durchschnitt. Wörter, die nicht gefunden werden können, erhalten ein neutrales Sentiment von 0 zugewiesen und werden nicht in der Wortanzahl mitgezählt, wodurch sie keine Auswirkungen auf das Sentiment des Tweets haben. Kann in einem Tweet kein Wort mit der Senti-WS-Wortsammlung abgeglichen werden, so erhält der Tweet ebenfalls ein neutrales Sentiment von 0. Das Sentiment eines Tages errechnet das Programm in weiterer Folge aus der Summe der Sentiment-Werte eines Tages geteilt durch die Anzahl der Tweets dieses Tages. Tweets mit einem Sentiment-Wert von 0 werden auch hier nicht mitgezählt und beeinflussen das Tagessentiment nicht. Die Ausgabe des Programms besteht sowohl aus den einzelnen Tweets mit Datum und berechnetem Sentiment als auch den errechneten täglichen Sentiment-Werten. Diese beiden Ausgaben werden als separate CSV-Dateien in den Ausgabepfaden gespeichert. Die täglichen Sentiments-Werte werden in weiterer Folge für die Gegenüberstellung mit den täglichen Kursänderungen genutzt.

4.2 Ermittlung des Kursverlaufes

Die Kursverläufe der einzelnen Unternehmen werden für den Zeitraum vom 01.08.2019 bis 31.07.2022 direkt von der Wiener Börse abgefragt und im CSV-Format exportiert. Die für die Untersuchung relevanten Spalten sind das Tagesdatum sowie die prozentuelle Zu- oder Abnahme des Kurses am jeweiligen Tag. Die restlichen Spalten werden verworfen. Für Tage, an denen die Wiener Börse geschlossen ist, gibt es keine Werte. Um die Gegenüberstellung im folgenden Teil zu vereinfachen, wird ein Programm²⁷ verwendet, welches die zusammengehörenden Werte jedes Tages in eine Zeile gruppiert. Als Eingabe dienen dem Programm die im vorherigen

²⁶ (Yaoshu, Jianbin und Wei 2017)

²⁷ Anhang 9.2 – Programm zur Zusammenführung von Sentiment und Kursänderung

Teil erstellte Datei mit den täglichen Sentiment-Werten sowie die in diesem Teil exportierte Datei. Die Ausgabe besteht aus einer Datei, die alle Werte chronologisch gruppiert enthält und im Tabellenkalkulationsprogramm Microsoft Excel geöffnet wird.

5. Analyse der Ergebnisse

Die aus der Sentiment-Analyse hervorgegangenen, täglichen, Werte werden den exportierten Kursänderungen des jeweiligen Tages gegenübergestellt. Folgende Diagramme veranschaulichen den Verlauf von Kursbewegung und Sentiment für jeden untersuchten Wert.

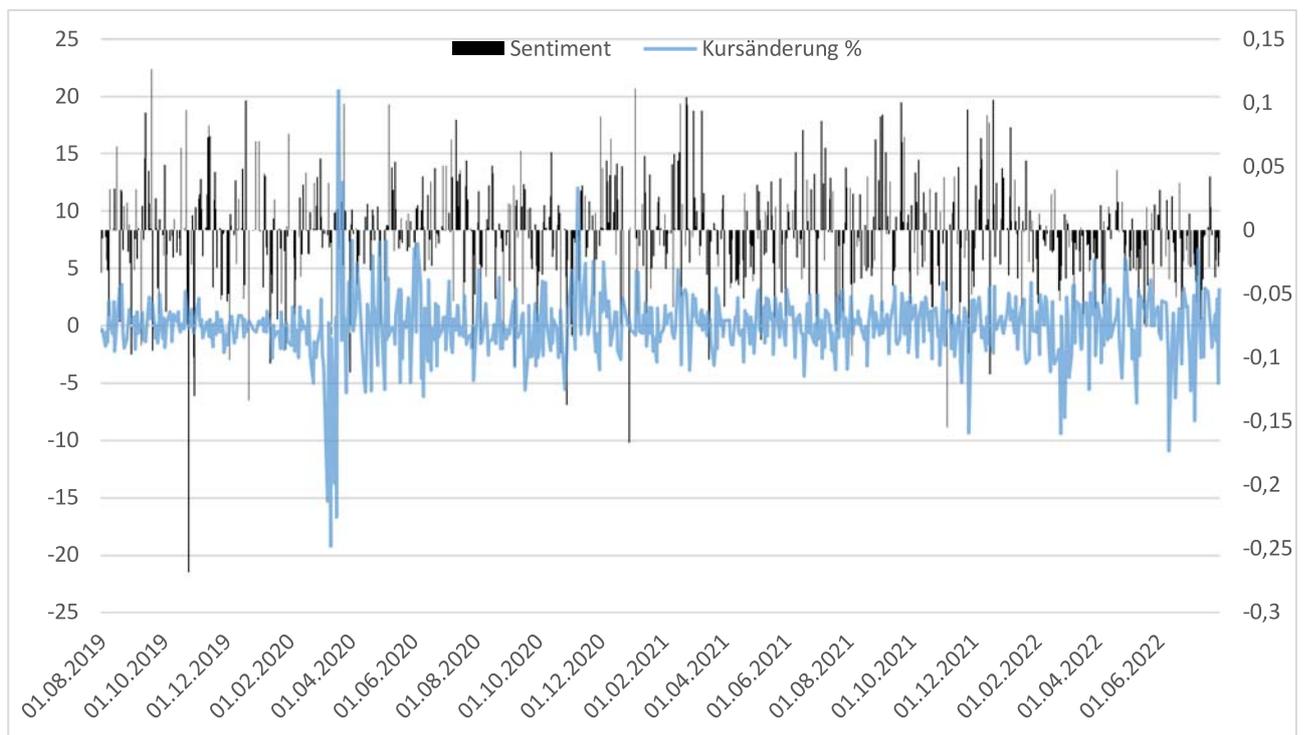


Abbildung 1 - Tabelle 3 - Sentiment und Kursänderung OMV AG

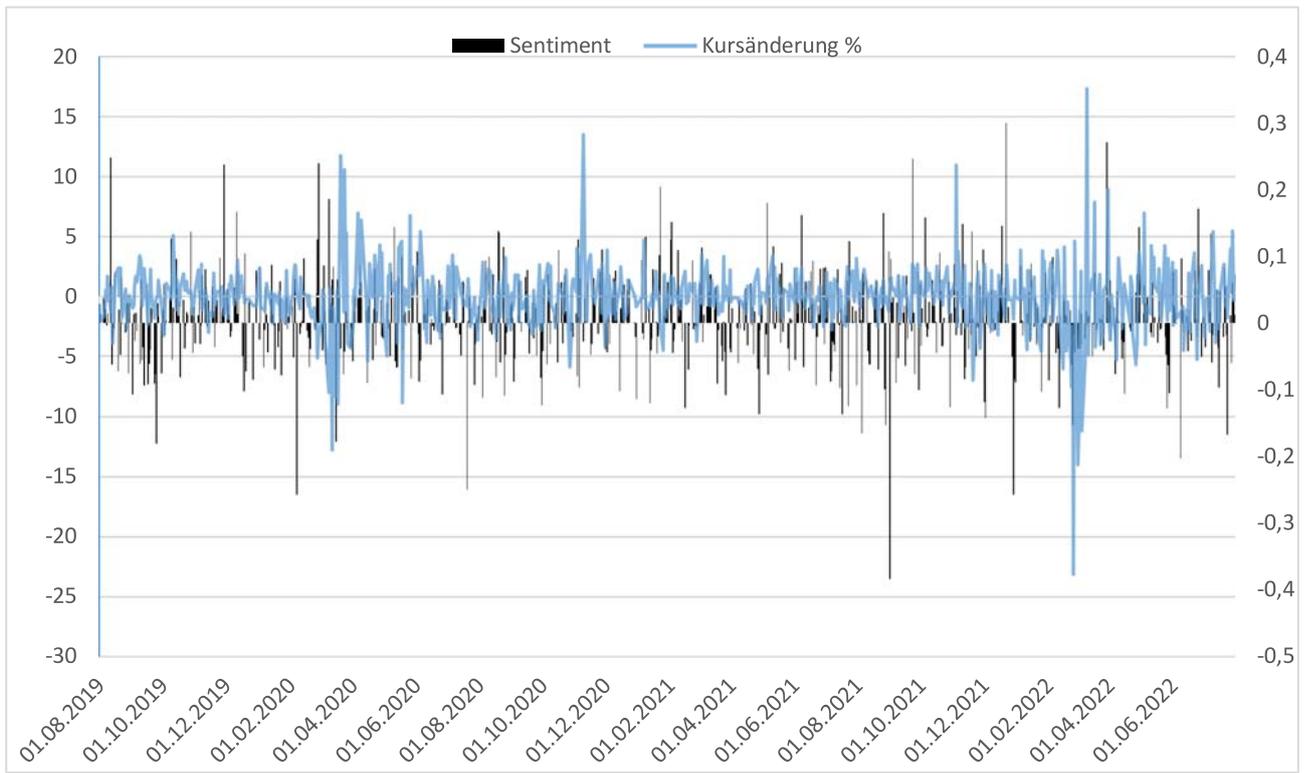


Abbildung 2 - Sentiment und Kursänderung Raiffeisen Bank International AG

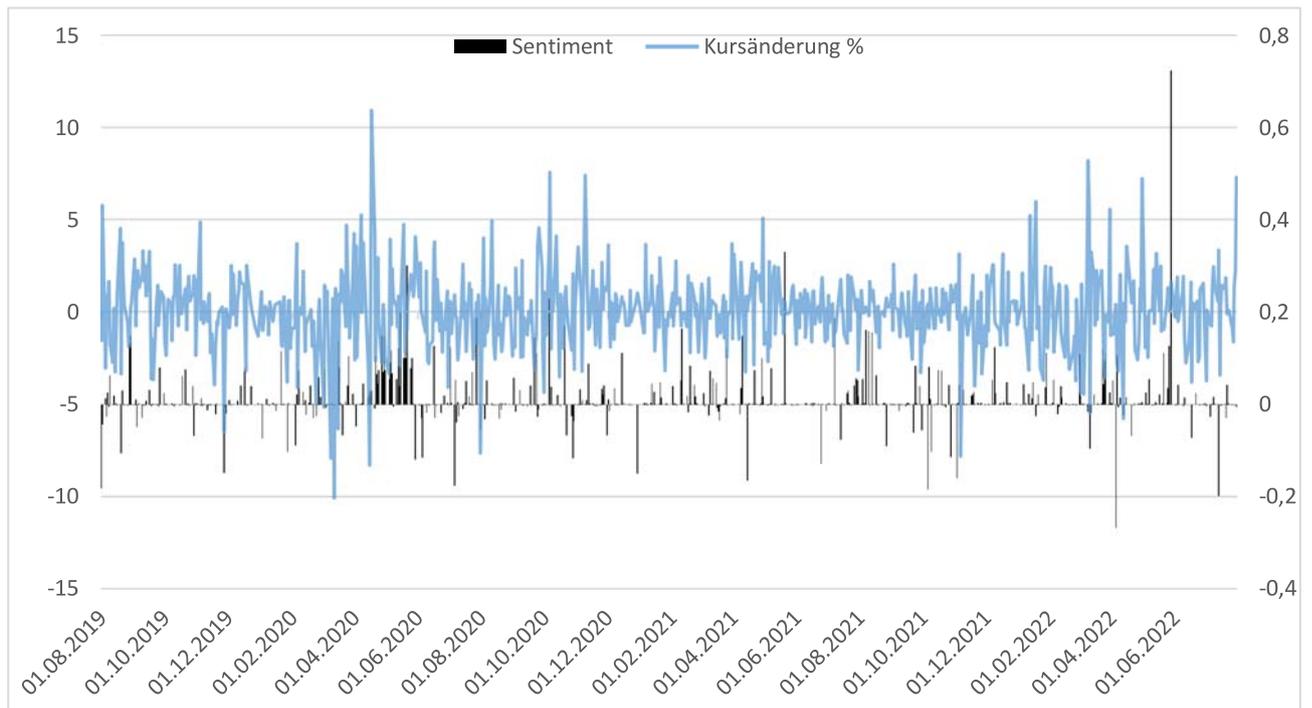


Abbildung 3 - Sentiment und Kursänderung Andritz AG

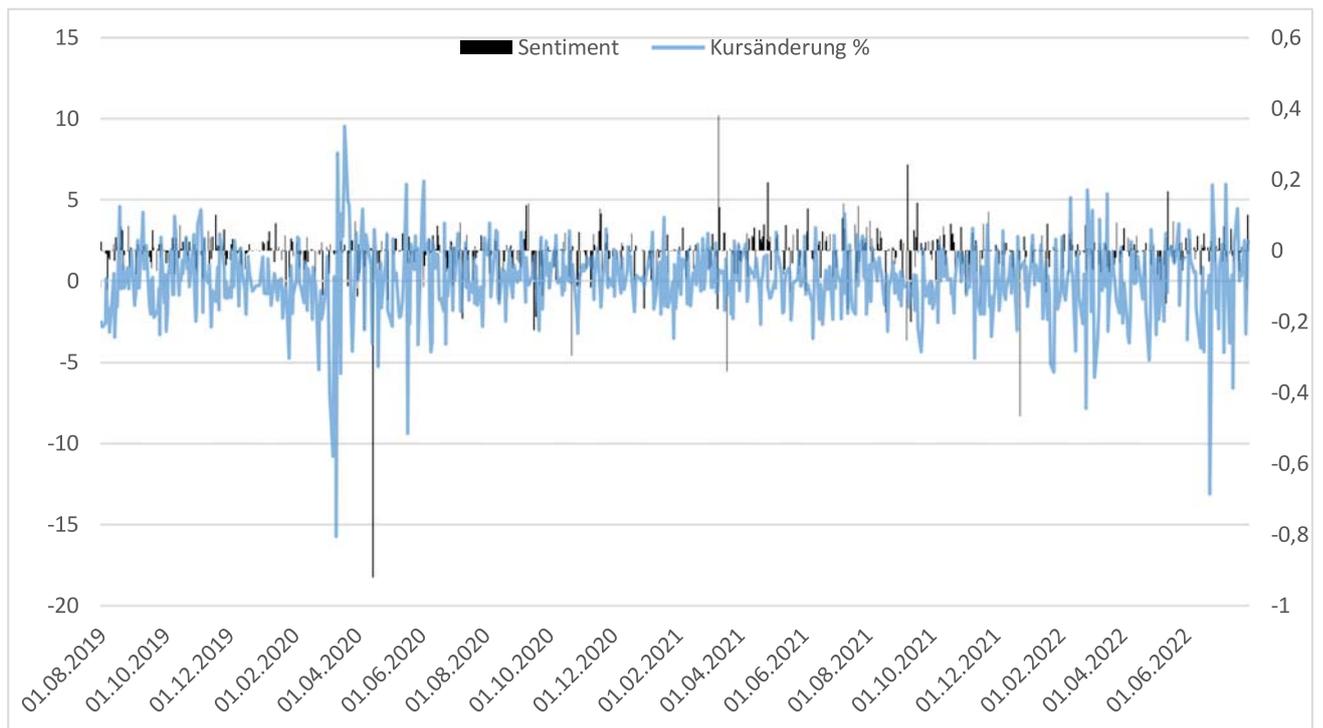


Abbildung 4 - Sentiment und Kursänderung voestalpine AG

In weiterer Folge gilt es zu bestimmen, ob eine Korrelation, also ein Zusammenhang zwischen den Variablen Kursänderung und Sentiment, besteht bzw. wie dieser ausgeprägt ist. Dazu wird die Pearson-Korrelationsfunktion²⁸ auf die beiden Spalten mit den Werten zur Kursänderung und zum Sentiment angewandt (K_0). Konkret wird dazu die „KORREL“-Funktion im Tabellenkalkulationsprogramm Microsoft Excel verwendet.²⁹ Die Korrelation nach Pearson setzt voraus, dass die zu untersuchenden Variablen linear zusammenhängen. Außerdem müssen die Werte möglichst normalverteilt und intervallskaliert sein.³⁰ Zusätzlich wird auch die, um ein bis fünf Tage, zeitversetzte Korrelation bestimmt ($K_1 - K_5$). Für die einzelnen Unternehmen ergeben sich dabei folgende Werte.

²⁸ (PennState Eberly College of Science - STAT 462 - Applied Regression Analysis 2018)

²⁹ (Zaiontz 2022)

³⁰ (Universität Zürich 2022)

Unternehmen	K ₀	K ₁	K ₂	K ₃	K ₄	K ₅
OMV AG	0,03226	0,02254	0,05796	0,05834	0,08925	0,04096
Raiffeisen Bank International AG	0,02240	0,01098	0,0063	0,02028	-0,0219	-0,0131
Andritz AG	0,00500	-0,01300	0,02476	-0,00181	0,00367	0,00831
Voestalpine AG	0,01777	0,00965	-0,0351	-0,0206	0,00213	-0,0086

Tabelle 3 – Korrelationskoeffizient K₀ – K₅

Anhand der Korrelationskoeffizienten ist ersichtlich, dass es keinen signifikanten Zusammenhang zwischen Kurs und Sentiment gibt. Das ändert sich auch durch die Berücksichtigung einer Zeitverzögerung nicht. Der Korrelationskoeffizient bewegt sich weiterhin um den Wert 0 herum, was darauf schließen lässt, dass es keine signifikante lineare Korrelation der Werte gibt, wie in der Hypothese behauptet. Zusätzlich werden die t- und p-Werte errechnet um die Nullhypothese, also die Annahme, dass es keinen Zusammenhang zwischen den Daten gibt und die Korrelationen nur durch Zufall in der Stichprobe bestehen, zu überprüfen.³¹ Der p-Wert gibt dabei die Wahrscheinlichkeit an, dass die Korrelation in der Stichprobe durch Zufall entstanden ist. Der t-Wert wird in Microsoft Excel mit der Formel „(L4*WURZEL(L7-2))/(WURZEL(1-L4^2))“ berechnet. Der p-Wert wird auf Basis des t-Werts mit der Funktion „T.VERT.2S(ABS(L5); L7-2)“ berechnet.³² Die Variable L4 steht dabei für den berechneten Korrelationskoeffizienten, L5 für den errechneten t-Wert und L7 für die Stichprobengröße. Dabei ergibt sich folgendes Ergebnis.

Unternehmen	t ₀ p ₀	t ₁ p ₁	t ₂ p ₂	t ₃ p ₃	t ₄ p ₄	t ₅ p ₅
OMV AG	0,88796 0,37485	0,61991 0,53551	1,59526 0,11107	1,60469 0,10898	2,45891 0,01416	1,12417 0,26165
Raiffeisen Bank International AG	0,61648 0,53776	0,30180 0,76288	0,17304 0,86267	0,55709 0,57763	-0,60123 0,54787	-0,35928 0,71949
Andritz AG	0,13749 0,89068	-3,60500 <u>0,00033</u>	0,68681 0,49282	0,68009 0,49665	-0,04967 0,96040	0,22789 0,81979
Voestalpine AG	0,48129 0,62490	0,26522 0,79091	-0,96373 <u>0,33549</u>	-0,56654 0,57119	0,05853 0,95334	-0,23557 0,81383

³¹ (JMP Statistical Discovery LLC 2022)

³² (education-wiki.com 2022)

Es gibt zwar Tage an denen Sentiment und Kursentwicklung gleichzeitig positiv oder negativ sind, wie folgendes Diagramm, am Beispiel der OMV AG, zeigt. Daraus lässt jedoch kein pauschaler Zusammenhang ableiten. Insgesamt haben in diesem Beispiel an 387 von 759 Tagen (50,98%) das Sentiment und die Kursänderung immerhin dasselbe Vorzeichen bzw. dieselbe Tendenz.

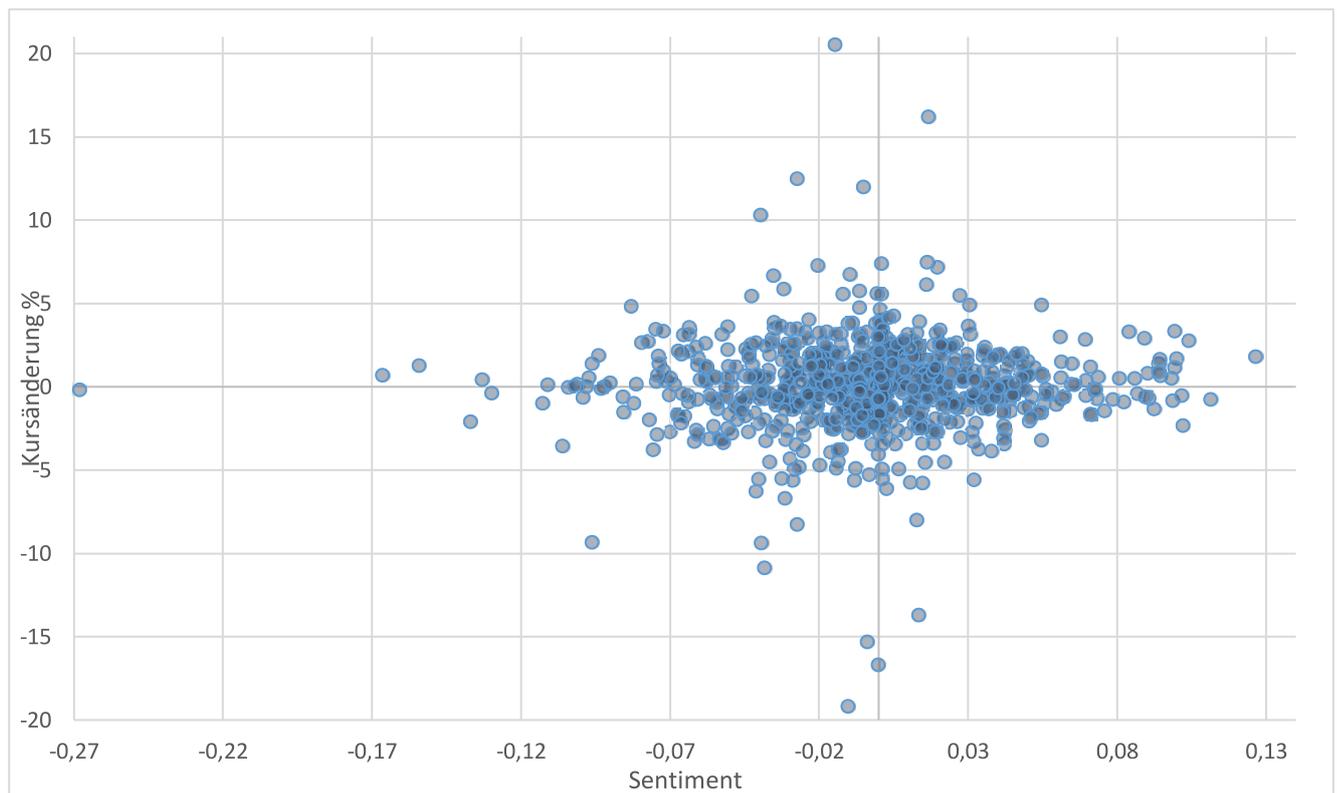


Abbildung 5 - Verteilung Kursänderung und Sentiment - OMV AG

Im Fall der OMV AG wurden durchschnittlich 26 Tweets pro Tag extrahiert, wovon durchschnittlich 6 Tweets vom Sentiment her neutral sind und die meisten keinen finanziellen Bezug haben. Beispielsweise hat der Kurs der OMV AG am 29.07.2022 um 3,29% zugenommen, was sich jedoch in den 137 Tweets dieses Tages, wenn man das Sentiment betrachtet, nicht widerspiegelt. Das Tagessentiment wird durch einige besonders kritische Beiträge negativ. Ein Gegenbeispiel ist etwa der 01.12.2020 an dem es nur vier Tweets gibt, von denen jedoch zwei einen finanziellen Bezug haben. Das Sentiment dieser Tage ist positiv und passt zur Kursentwicklung.

Ein ähnliches Ergebnis kommt auch für die Raiffeisen Bank International AG zustande, wobei hier 382 von 759 Tagen (50,32%) die gleiche Tendenz aufweisen. Bei der voestalpine AG sind es 371 von 759 Tagen (48,88%). Die geringste Anzahl an Tagen mit übereinstimmender Tendenz hat die Andritz AG mit 316 von 759 Tagen (41,63%). Dabei ist zu beachten, dass zur Andritz AG deutlich weniger Tweets gefunden wurden und viele Tage daher ein Sentiment von 0 aufweisen.

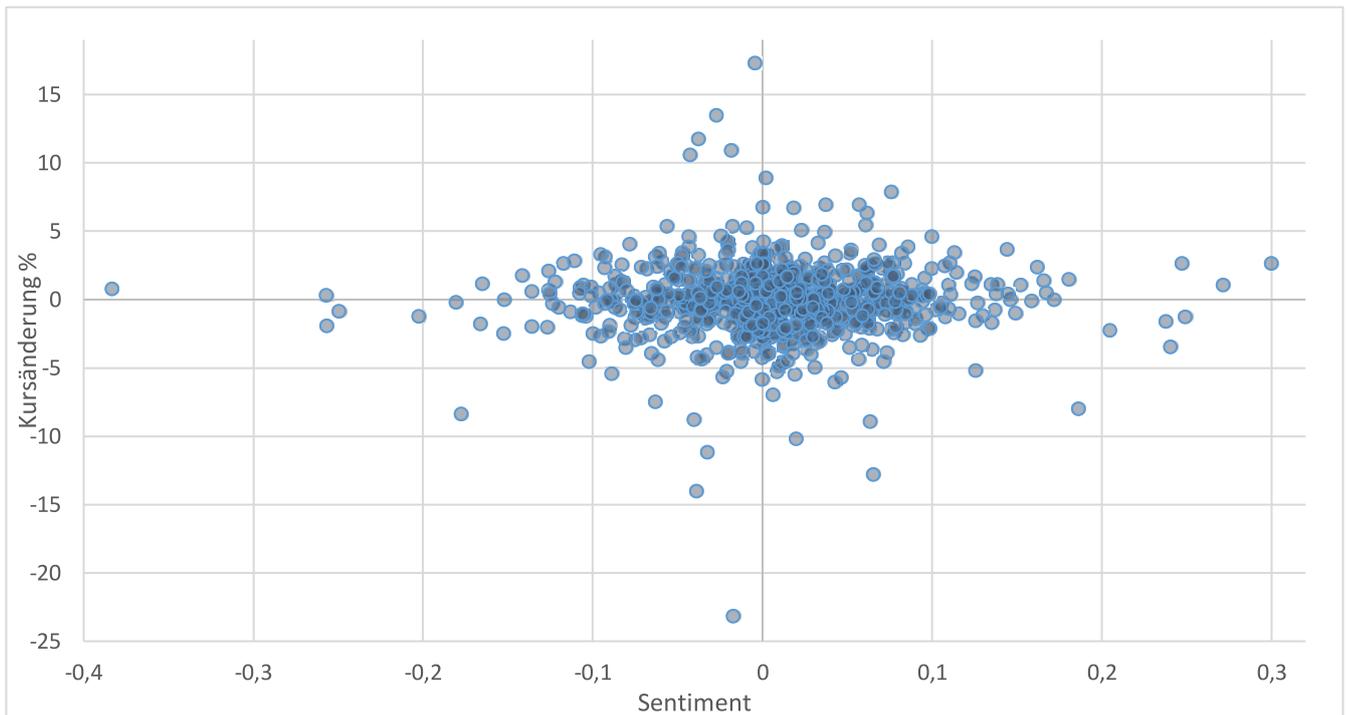


Abbildung 6 - Verteilung Kursänderung und Sentiment - Raiffeisen Bank International AG

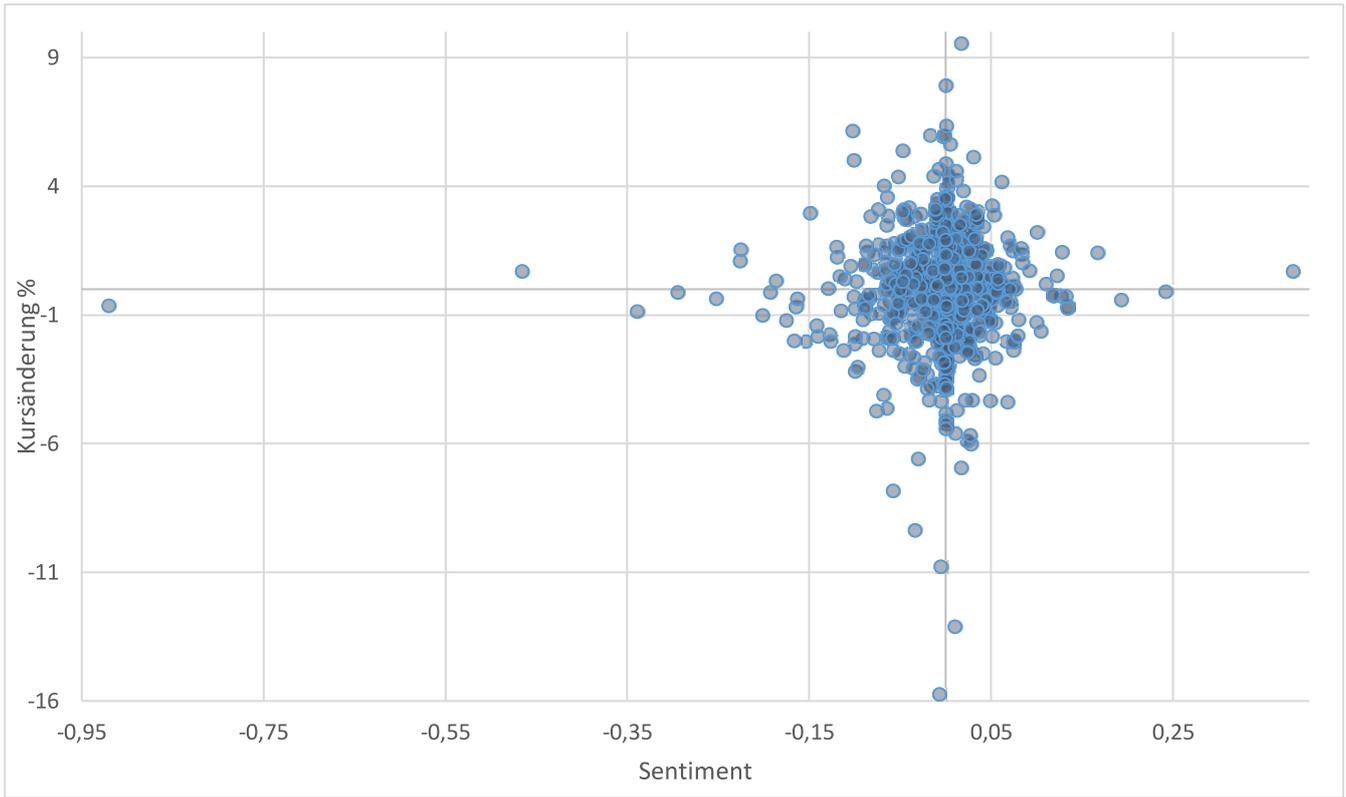


Abbildung 7 – Verteilung Kursänderung und Sentiment – voestalpine AG

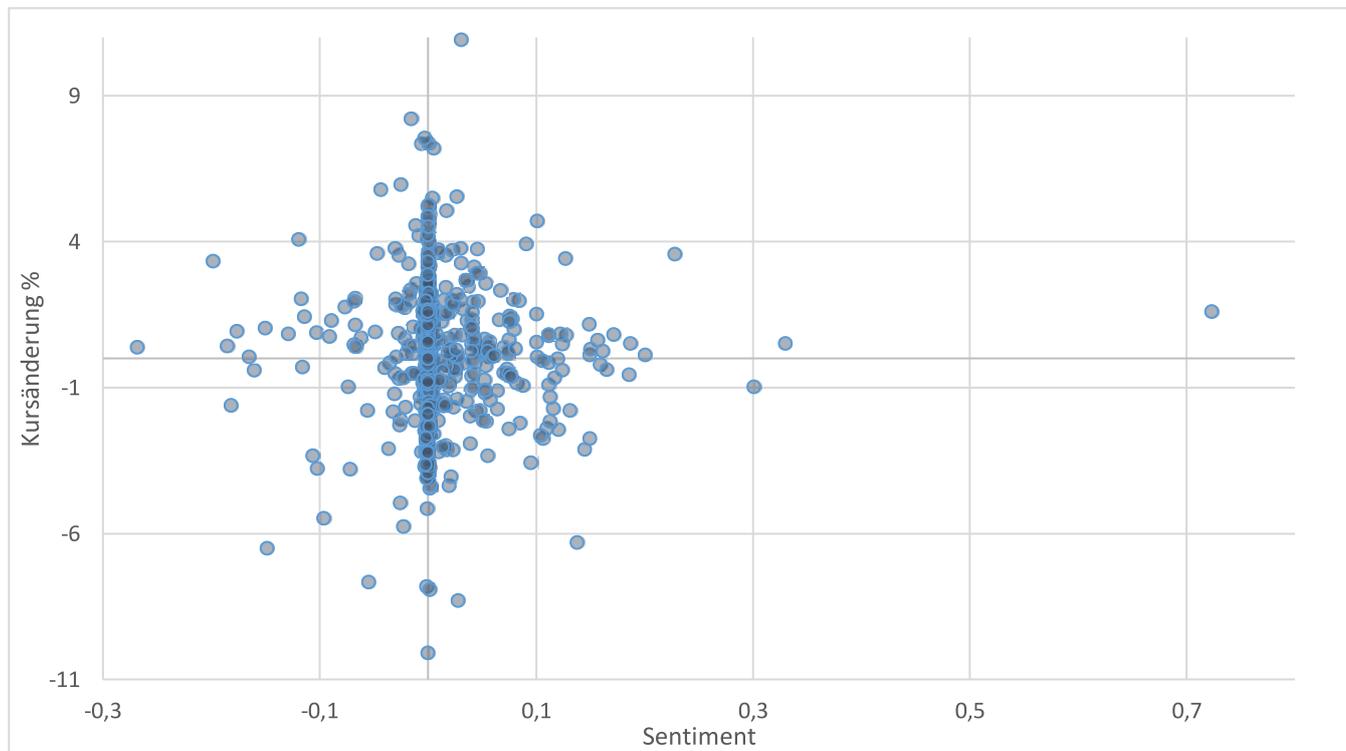


Abbildung 8 - Verteilung Kursänderung und Sentiment - Andritz AG

6. Schlussfolgerung

Wie im vorhergehenden Kapitel ersichtlich wurde, gibt es keinen signifikanten linearen Zusammenhang zwischen dem Kursverlauf und dem, lexikonbasiert errechneten, Social Sentiment einzelner Unternehmenswerte. Es lassen sich, auch unter Berücksichtigung von zeitversetzten Auswirkungen des Sentiments, keine signifikant starken Korrelationen finden. Es gibt zwar einzelne Zeitpunkte an denen Sentiment und Kurs die gleiche Tendenz aufweisen, was jedoch nicht für die Feststellung eines Zusammenhanges reicht.

Entgegen der aufgestellten Hypothese, dass es einen Zusammenhang zwischen der Kursentwicklung einzelner Unternehmen und dem Social Sentiment zu dem jeweiligen Unternehmen einen klaren Zusammenhang gibt, hat sich herausgestellt das dem nicht so ist. Daraus folgend kann die eingangs in Kapitel 1.2 gestellte Forschungsfrage grundsätzlich verneint werden. Es kann allerdings nicht ausgeschlossen werden, dass etwa Korrelationen zwischen dem ATX-Indexkurs oder einer Gruppe von Aktien und dem dazugehörendem Social Sentiment auf Twitter, Korrelationen bestehen. Auch kann nicht ausgeschlossen werden, dass Korrelationen zwischen der Kursänderung einzelner Unternehmen mit mehr relevanten Tweets, außerhalb des ATX-Indexes, und deren Social Sentiment besteht. Für einzelne Aktien des ATX-Indexes kann eine solche Korrelation jedoch nicht bestätigt werden.

7. Zusammenfassung

Ziel dieser Arbeit ist es zu untersuchen, ob es einen Zusammenhang zwischen dem Kursverlauf und dem Social Sentiment eines börsennotierten Unternehmens gibt. In etlichen Publikationen, auf die im Rahmen dieser Arbeit im Kapitel 2.6 verwiesen wird, wurde festgestellt, dass es zwischen dem lexikonbasiert ermittelten Sentiment und Kursverläufen von Aktien-Indexes, in verschiedenen zeitlichen Intervallen, eine Korrelation gibt. Als Herausforderung bestand dabei meist die korrekte Interpretation der Beiträge in den sozialen Netzwerken, die schließlich das Social Sentiment ergibt. Als zuverlässig bei der Analyse dieser Beiträge hat sich, auch in aktuellen Publikationen, stets der lexikonbasierte Ansatz erwiesen. Bei diesem werden die Textbeiträge aus den sozialen Netzwerken mithilfe von bestehenden oder eigens angefertigten Wortsammlungen Wort für Wort bewertet. Zur Beantwortung der Forschungsfrage wurde auch in dieser Arbeit das Social Sentiment zu vier börsennotierten, österreichischen Unternehmen extrahiert. Dazu wurden zuerst Stichwörter identifiziert, die bei der Suchanfrage in Twitter relevante Ergebnisse liefern. Um die Ergebnisse für diesen Zeitraum automatisiert verarbeiten zu können, wurden diese exportiert. Dazu war es notwendig ein Programm in JavaScript zu verfassen, welches die Beiträge mit Datum in Textform lieferte. Die Interpretation der Texte wurde mithilfe des lexikonbasierten Ansatzes unter Verwendung einer öffentlich zugänglichen, deutschsprachigen Wortsammlung erreicht. Dazu wurde ein Programm in Java erstellt, welches Wörter aus dem Text mit der Wortsammlung abgleicht und so jedem Wort und jedem Tweet ein Sentiment zuweist und das Sentiment jedes Tages zur weiteren Verwendung ausgibt. Weiters wurden die Kursverläufe der einzelnen Unternehmen exportiert und, mithilfe eines weiteren Java-Programmes, mit den Sentiment-Werten verknüpft. Die so gewonnenen Datenreihen wurden gegenübergestellt und visualisiert. Durch Ermittlung des Pearson-Korrelationskoeffizienten hat sich eine sehr schwache Korrelation gezeigt. Auch die Berücksichtigung von Zeitverschiebungen zwischen Sentiment und Kursänderung hat keine signifikante lineare Korrelation gezeigt. Es konnte lediglich gezeigt werden, dass es Tage gibt, an denen die Tendenz von Kursänderung und Sentiment übereinstimmen. Hervorzuheben ist hierbei, dass, in anderen, vergleichbaren Publikationen, auf die im Kapitel 2.6 eingegangen wird, keine einzelnen Unternehmenswerte betrachtet wurden, sondern stets mit umfangreichen Indexkursen und dementsprechend großer Datenbasis in englischer Sprache, mit deutlich umfangreicheren Wortsammlungen, gearbeitet wurde.

8. Ausblick

Insbesondere im Hinblick darauf, dass die Hypothese nicht bestätigt werden konnte, stellt sich die Frage ob es möglicherweise andere, zuverlässigere Methoden gibt, um das Sentiment zu einzelnen Börsenwerten zu bestimmen. Dabei geht es einerseits um die Datenbasis – möglicherweise ist Twitter, nicht optimal geeignet, weil nicht auf Finanzen spezialisiert. Andererseits wäre auch eine größere Wortsammlung oder eine aus den Tweets abgeleitete, manuell bewertete, Wortsammlung von Vorteil gewesen, deren Erstellung jedoch sehr viel Zeit in Anspruch genommen hätte. Es stellt sich auch die Frage, ob das Ergebnis ein anderes wäre, wenn die gleichen Tweets von einer künstlichen Intelligenz oder gar einzeln von Menschen auf ihr Sentiment hin bewertet worden wären.

Mit Fortschreiten der Digitalisierung eröffnen sich auch für die Sentiment-Analyse neue Möglichkeiten. So ist davon auszugehen, dass zukünftig eine künstliche Intelligenz in Kombination mit einem neuronalen Netzwerk und Machine Learning eine genauere Beurteilung von Texten zulässt als Wortsammlungen, auch wenn sich der lexikonbasierte Ansatz bisher bewährt hat. Der Begriff Social Sentiment umfasst nach heutigem Verständnis vor Allem die sozialen Medien, doch auch das könnte sich zukünftig ändern. Grundsätzlich kann nicht nur jedes veröffentlichte Wort, sondern auch jede sonstige im Internet gesetzte Aktion ein Sentiment enthalten. Daraus könnten sich in Zukunft äußerst komplexe und vielfältige Berechnungsmodelle für das Sentiment gegenüber Unternehmen und Organisationen aller Art ergeben.

9. Anhänge

9.1 Programm zur Sentiment-Bestimmung

```
package at.huk.sentiment;

import java.io.BufferedReader;
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.math.BigDecimal;
import java.math.RoundingMode;
import java.text.ParseException;
import java.text.SimpleDateFormat;
import java.util.ArrayList;
import java.util.Date;
import java.util.LinkedHashMap;
import java.util.List;
import java.util.Map;
import java.util.Map.Entry;
import java.util.stream.Collectors;

import org.apache.commons.lang3.StringUtils;
import org.apache.commons.lang3.tuple.Pair;
import org.apache.commons.text.similarity.JaroWinklerSimilarity;

import com.opencsv.CSVParser;
import com.opencsv.CSVParserBuilder;
import com.opencsv.CSVReader;
import com.opencsv.CSVReaderBuilder;
import com.opencsv.exceptions.CsvValidationException;

public class SentiWsInputAnalyzer {

    private static JaroWinklerSimilarity similarityCorrection = new JaroWinklerSimilarity();
    private static SimpleDateFormat exportDatumFormat = new SimpleDateFormat("dd.MM.yyy");

    public static void main(String[] args) throws CsvValidationException, IOException, ParseException {

        // EINGABEPFAD TWITTERDATEN
        String inputTwitterDatenPfad = args[0];

        // EINGABEPFAD SENTI-WS
        String inputSentiWsPfad = args[1];

        // AUSGABEPFADE TWITTER SENTIMENT PRO TWEET UND PRO TAG
        String outputTwitterSentimentProTweetPfad = args[2];
        String outputTaeglichesSentimentPfad = args[3];

        // EINLESEN DER TWITTER-DATEN
        Map<Date, List<String>> twitterDaten = LeseTwitterDaten(inputTwitterDatenPfad);

        // EINLESEN DER SENTI-WS-WORTSAMMLUNG
        Map<String, BigDecimal> sentiWsDaten = LeseSentiWsDaten(inputSentiWsPfad);

        // AUSWERTEN DES SENTIMENTS EINZELNER TWEETS
        Map<Date, List<Pair<String, BigDecimal>>> bewerteteTwitterDaten =
bewerteSentiment(twitterDaten, sentiWsDaten);

        // AUSWERTEN DES SENTIMENTS TAEGLICH
        Map<Date, BigDecimal> sentimentTaeglich =
erstelleSentimentVerlaufTaeglich(bewerteteTwitterDaten);

        schreibeAusgabeCSVBewerteteTwitterDaten(outputTwitterSentimentProTweetPfad,
        bewerteteTwitterDaten);

        schreibeAusgabeCSVSentimentTaeglich(outputTaeglichesSentimentPfad, sentimentTaeglich);
    }
}
```

```

    }

    private static void schreibeAusgabeCSVSentimentTaeglich(String outputTaeglichesSentimentPfad,
        Map<Date, BigDecimal> sentimentTaeglich) throws IOException {
        File fileAusgabePfadTglSentiment = new File(outputTaeglichesSentimentPfad);

        if (fileAusgabePfadTglSentiment.getParentFile().mkdirs()) {
            fileAusgabePfadTglSentiment.createNewFile();
        }

        FileWriter writer = new FileWriter(fileAusgabePfadTglSentiment);

        for (Entry<Date, BigDecimal> sentimentTgl : sentimentTaeglich.entrySet()) {
            writer.write(exportDatumFormat.format(sentimentTgl.getKey()) + ";"
                + (sentimentTgl.getValue().equals(BigDecimal.ZERO) ?
                    BigDecimal.ZERO.setScale(0)
                    : sentimentTgl.getValue())
                + System.LineSeparator());
        }

        writer.close();
    }

    private static void schreibeAusgabeCSVBewerteteTwitterDaten(String
        outputTwitterSentimentProTweetPfad,
        Map<Date, List<Pair<String, BigDecimal>>> bewerteteTwitterDaten) throws
        IOException {

        File fileAusgabePfadTglTweets = new File(outputTwitterSentimentProTweetPfad);

        if (fileAusgabePfadTglTweets.getParentFile().mkdirs()) {
            fileAusgabePfadTglTweets.createNewFile();
        }

        FileWriter writer = new FileWriter(fileAusgabePfadTglTweets);

        for (Entry<Date, List<Pair<String, BigDecimal>>> twitterTgl :
            bewerteteTwitterDaten.entrySet()) {

            for (Pair<String, BigDecimal> tweet : twitterTgl.getValue()) {
                writer.write(exportDatumFormat.format(twitterTgl.getKey()) + ";"
                    + (tweet.getValue().equals(BigDecimal.ZERO) ?
                        BigDecimal.ZERO.setScale(0) : tweet.getValue())
                    + ";" + tweet.getKey() + System.LineSeparator());
            }

        }

        writer.close();
    }

    private static Map<Date, BigDecimal> erstelleSentimentVerlaufTaeglich(
        Map<Date, List<Pair<String, BigDecimal>>> bewerteteTwitterDaten) {
        Map<Date, BigDecimal> sentimentTaeglichTemp = new LinkedHashMap<>();

        for (Entry<Date, List<Pair<String, BigDecimal>>> twitterTaeglich :
            bewerteteTwitterDaten.entrySet()) {
            BigDecimal summeSentimentTag = BigDecimal.ZERO.setScale(5);
            int anzahlTweetsTag = 0;

            for (Pair<String, BigDecimal> tweetSentiment : twitterTaeglich.getValue()) {
                summeSentimentTag = summeSentimentTag.add(tweetSentiment.getValue());
                anzahlTweetsTag++;
            }

            if (anzahlTweetsTag != 0) {
                sentimentTaeglichTemp.put(twitterTaeglich.getKey(),

                    summeSentimentTag.divide(BigDecimal.valueOf(anzahlTweetsTag), RoundingMode.HALF_UP).setScale(5,
                        RoundingMode.HALF_UP));
                System.out.println(twitterTaeglich.getKey() + "
                    +
                    summeSentimentTag.divide(BigDecimal.valueOf(anzahlTweetsTag), RoundingMode.HALF_UP)
                    .setScale(5,
                    RoundingMode.HALF_UP));
            }
        }
    }

```

```

        } else {
            sentimentTaeglichTemp.put(twitterTaeglich.getKey(),
BigDecimal.ZERO.setScale(0));
            System.out.println(twitterTaeglich.getKey() + " " +
BigDecimal.ZERO);
        }
    }
    return sentimentTaeglichTemp;
}

private static Map<Date, List<Pair<String, BigDecimal>>> bewerteSentiment(Map<Date, List<String>>
twitterDaten,
    Map<String, BigDecimal> sentiWsDaten) {
    Map<Date, List<Pair<String, BigDecimal>>> bewerteteTwitterDatenTemp = new
LinkedHashMap<>();
    for (Entry<Date, List<String>> twitterTag : twitterDaten.entrySet()) {
        List<Pair<String, BigDecimal>> bewerteteTweets = new ArrayList<>();
        for (String tweet : twitterTag.getValue()) {
            bewerteteTweets.add(Pair.of(tweet, bewerteTweet(tweet, sentiWsDaten)));
        }
        bewerteteTwitterDatenTemp.put(twitterTag.getKey(), bewerteteTweets);
    }
    return bewerteteTwitterDatenTemp;
}

private static BigDecimal bewerteTweet(String tweet, Map<String, BigDecimal> sentiWsDaten) {
    int anzahlWoerterMitSentiment = 0;
    BigDecimal summeSentiment = BigDecimal.ZERO;
    summeSentiment.setScale(5, RoundingMode.HALF_UP);
    for (String wort : tweet.split(" ")) {
        if (wort.length() > 2) {
            BigDecimal sentimentZuWort = BigDecimal.ZERO;
            // sentimentZuWort = sentiWsDaten.getOrDefault(wort.toLowerCase(),
            // BigDecimal.ZERO);
            // ----
            List<BigDecimal> passendesSentiment =
sentiWsDaten.entrySet().stream().filter(senti ->
senti.getKey().length()==wort.length() &&
similarityCorrection.apply(senti.getKey(), wort.toLowerCase()) >=
0.97).map(senti -> senti.getValue())
                .collect(Collectors.toList());
            if (!passendesSentiment.isEmpty()) {
                sentimentZuWort =
passendesSentiment.stream().reduce(BigDecimal.ZERO, BigDecimal::add)
                    .divide(BigDecimal.valueOf(passendesSentiment.size()), RoundingMode.HALF_UP);
            }
            // ----
            summeSentiment = summeSentiment.add(sentimentZuWort);
            if (!sentimentZuWort.equals(BigDecimal.ZERO)) {
                anzahlWoerterMitSentiment++;
            }
        }
    }
    System.out.println(summeSentiment + " - " + anzahlWoerterMitSentiment + " - " + tweet);
    return summeSentiment.equals(BigDecimal.ZERO) ? BigDecimal.ZERO
        :
summeSentiment.divide(BigDecimal.valueOf(anzahlWoerterMitSentiment).setScale(5, RoundingMode.HALF_UP),
        RoundingMode.HALF_UP);
}
}

```

```

private static Map<Date, List<String>> leseTwitterDaten(String inputTwitterDataPath)
    throws CsvValidationException, IOException, ParseException {

    Map<Date, List<String>> twitterDatenTemp = new LinkedHashMap<>();

    CSVParser parser = new CSVParserBuilder().withSeparator(';').build();
    CSVReader csvReader = new CSVReaderBuilder(new
FileReader(inputTwitterDataPath)).withCSVParser(parser).build();
    String[] zeile;
    SimpleDateFormat dateFormat = new SimpleDateFormat("MMMM dd, yyyy");

    // UEBERSCHRIFTENZEILE IN EINGABE-CSV UEBERSPRINGEN
    csvReader.skip(1);

    while ((zeile = csvReader.readNext()) != null) {
        twitterDatenTemp.putIfAbsent(dateFormat.parse(zeile[0]), new ArrayList<>());
        twitterDatenTemp.get(dateFormat.parse(zeile[0]))
            .add(StringUtils.normalizeSpace(zeile[1].replaceAll("[^A-Za-
zöäüß]", " ").trim().toLowerCase()));
    }

    return twitterDatenTemp;
}

private static Map<String, BigDecimal> leseSentiWsDaten(String inputSentiWsPath)
    throws FileNotFoundException, IOException {
    Map<String, BigDecimal> sentiWsDatenTemp = new LinkedHashMap<>();
    String zeile;

    File eingabePfad = new File(inputSentiWsPath);

    for (File eingabeDatei : eingabePfad.listFiles()) {

        BufferedReader br = new BufferedReader(new FileReader(eingabeDatei));
        while ((zeile = br.readLine()) != null) {
            String[] tabExtrakt = zeile.split(" ");
            String[] wortExtrakt = tabExtrakt[0].split("\\|");

            sentiWsDatenTemp.putIfAbsent(wortExtrakt[0].toLowerCase(),
                new BigDecimal(tabExtrakt[1]).setScale(5,
RoundingMode.HALF_UP));

            if (tabExtrakt.length > 2) {
                for (String wortform : tabExtrakt[2].split(",")) {
                    sentiWsDatenTemp.putIfAbsent(wortform.toLowerCase(),
                        new
BigDecimal(tabExtrakt[1]).setScale(5, RoundingMode.HALF_UP));
                }
            }
        }
    }

    return sentiWsDatenTemp;
}
}

```

9.2 Programm zur Zusammenführung von Sentiment und Kursänderung

```
package at.huk.sentiment;

import java.io.File;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.math.BigDecimal;
import java.text.ParseException;
import java.text.SimpleDateFormat;
import java.util.ArrayList;
import java.util.Date;
import java.util.LinkedHashMap;
import java.util.List;
import java.util.Map;
import java.util.Map.Entry;

import org.apache.commons.lang3.StringUtils;
import org.apache.commons.lang3.tuple.Pair;

import com.opencsv.CSVParser;
import com.opencsv.CSVParserBuilder;
import com.opencsv.CSVReader;
import com.opencsv.CSVReaderBuilder;
import com.opencsv.exceptions.CsvValidationException;

public class StockSentiCombiner {

    private static SimpleDateFormat exportDatumFormat = new SimpleDateFormat("dd.MM.yyy");

    public static void main(String[] args) throws CsvValidationException, IOException, ParseException {

        // EINGABEPFAD TWITTER-SENTIMENT
        String inputSentiTglPfad = args[0];

        // EINGABEPFAD BOERSE-DATEN
        String inputBoerseTglPfad = args[1];

        // EINGABEPFAD SENTI-WS
        String outputKombinierteDatei = args[2];

        // EINLESEN DER TGL SENTI-DATEN
        Map<Date, BigDecimal> sentiTglDaten = leseSentiTglDaten(inputSentiTglPfad);

        // EINLESEN DER BOERSE-DATEN
        Map<Date, BigDecimal> boerseTglDaten = leseBoerseTglDaten(inputBoerseTglPfad);

        // KOMBINIEREN DER WERTE
        Map<Date, Pair<BigDecimal, BigDecimal>> kombinierteTglDaten =
kombiniereAlleTglDaten(sentiTglDaten,
boerseTglDaten);

        schreibeGruppierteDatei(kombinierteTglDaten, outputKombinierteDatei);
    }

    private static void schreibeGruppierteDatei(Map<Date, Pair<BigDecimal, BigDecimal>>
kombinierteTglDaten,
        String outputKombinierteDatei) throws IOException {
        File fileAusgabePfadTglSentiment = new File(outputKombinierteDatei);

        if (fileAusgabePfadTglSentiment.getParentFile().mkdirs()) {
            fileAusgabePfadTglSentiment.createNewFile();
        }

        FileWriter writer = new FileWriter(fileAusgabePfadTglSentiment);

        for (Entry<Date, Pair<BigDecimal, BigDecimal>> kombTgl : kombinierteTglDaten.entrySet()) {
            writer.write(exportDatumFormat.format(kombTgl.getKey()) + ";"

```

```

        + (kombTgl.getValue().getKey() == null ?
BigDecimal.ZERO.setScale(0)
        :
kombTgl.getValue().getKey().toString().replace(".", ",")
        + ";")
        + (kombTgl.getValue().getValue() == null ?
BigDecimal.ZERO.setScale(0)
        :
kombTgl.getValue().getValue().toString().replace(".", ",")
        + System.LineSeparator());
    }
    writer.close();
}

private static Map<Date, Pair<BigDecimal, BigDecimal>> kombiniereAlleTglDaten(Map<Date, BigDecimal>
sentiTglDaten,
    Map<Date, BigDecimal> boerseTglDaten) {
    Map<Date, Pair<BigDecimal, BigDecimal>> kombinierteDatenTemp = new LinkedHashMap<>();
    for (Entry<Date, BigDecimal> sentiZeile : sentiTglDaten.entrySet()) {
        //ZEILE NUR UEBERNEHMEN WENN AUCH BOERSENDATENZUM TAG VORHANDEN
        if (boerseTglDaten.containsKey(sentiZeile.getKey())) {
            kombinierteDatenTemp.put(sentiZeile.getKey(),
Pair.of(sentiZeile.getValue(), BigDecimal.ZERO));
        }
        for (Entry<Date, BigDecimal> boerseZeile : boerseTglDaten.entrySet()) {
            if (kombinierteDatenTemp.containsKey(boerseZeile.getKey())) {
                BigDecimal sentiTgl =
kombinierteDatenTemp.get(boerseZeile.getKey()).getLeft();
                kombinierteDatenTemp.put(boerseZeile.getKey(), Pair.of(sentiTgl,
boerseZeile.getValue()));
            } else {
                kombinierteDatenTemp.put(boerseZeile.getKey(), Pair.of(null,
boerseZeile.getValue()));
            }
        }
    }
    return kombinierteDatenTemp;
}

private static Map<Date, BigDecimal> leseBoerseTglDaten(String inputBoerseTglPfad)
throws CsvValidationException, IOException, ParseException {
    Map<Date, BigDecimal> boerseTglTemp = new LinkedHashMap<>();

    CSVParser parser = new CSVParserBuilder().withSeparator(';').build();
    CSVReader csvReader = new CSVReaderBuilder(new
FileReader(inputBoerseTglPfad)).withCSVParser(parser).build();
    String[] zeile;
    SimpleDateFormat dateFormat = new SimpleDateFormat("dd.MM.yyy");

    // UEBERSCHRIFTENZEILE IN EINGABE-CSV UEBERSPRINGEN
    csvReader.skip(1);

    while ((zeile = csvReader.readNext()) != null) {
        boerseTglTemp.putIfAbsent(dateFormat.parse(zeile[0]),
            new BigDecimal(zeile[5].replaceAll("%", "").replaceAll(",",
".").trim()));
    }
    return boerseTglTemp;
}

private static Map<Date, BigDecimal> leseSentiTglDaten(String inputSentiTglPfad)
throws CsvValidationException, IOException, ParseException {
    Map<Date, BigDecimal> sentiTglTemp = new LinkedHashMap<>();

    CSVParser parser = new CSVParserBuilder().withSeparator(';').build();

```

```
        CSVReader csvReader = new CSVReaderBuilder(new
FileReader(inputSentiTglPfad)).withCSVParser(parser).build();
        String[] zeile;
        SimpleDateFormat dateFormat = new SimpleDateFormat("dd.MM.yyy");

        while ((zeile = csvReader.readNext()) != null) {
            sentiTglTemp.putIfAbsent(dateFormat.parse(zeile[0]), new BigDecimal(zeile[1]));
        }

        return sentiTglTemp;
    }
}
```

10. Literaturverzeichnis

- Bloomberg L.P. *Bloomberg*. 16. September 2015. <https://www.bloomberg.com/company/press/bloomberg-and-twitter-sign-data-licensing-agreement/> (Zugriff am 14. August 2022).
- Bollen, Johan, Huina Mao, und Xiao-Jun Zeng. *Twitter mood predicts the stock market*. Bloomington: Indiana University, 2011.
- Byrne, Alistair, und Mike Brooks. „Behavioral Finance: Theories and Evidence.“ *The Research Foundation of CFA Institute*, 2008: 1-10.
- education-wiki.com. *education-wiki.com*. 2022. <https://de.education-wiki.com/3221224-excel-p-value> (Zugriff am 30. September 2022).
- EU NACE v2. *umweltbundesamt.at*. 21. September 2022. https://secure.umweltbundesamt.at/edm_portal/redaList.do (Zugriff am 21. September 2022).
- Fama, Eugene F. „Efficient Capital Markets: A Review of Theory and Empirical Work.“ In *The Journal of Finance*, 383-417. New York: Wiley, 1970.
- Gao, Lei, Yufeng Han, Sophia Zhengzi Li, und Guofu Zhou. „Intraday Momentum: The First Half-Hour Return Predicts the Last Half-Hour Return.“ 2014.
- Guo, Xinyi, und Li Jinfeng. „A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency.“ *The 6th IEEE International Conference on Social Networks Analysis, Management and Security (SNAMS 2019)*. Granada, 2019.
- JMP Statistical Discovery LLC. *JMP Statistical Discovery LLC*. 2022. https://www.jmp.com/de_at/statistics-knowledge-portal/what-is-correlation/correlation-coefficient.html (Zugriff am 1. Oktober 2022).
- Kim, Soo-Min, und Eduard Hovy. „Determining the sentiment of opinions.“ *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, 2004: 1367.
- Liu, Bing. „Sentiment Analysis and Opinion Mining.“ Morgan & Claypool Publishers, 2012.
- Loughran, Tim, und Bill McDonald. „Survey, Textual Analysis in Accounting and Finance: A Survey.“ *Journal of Accounting Research* 54, 2016: 1187–1230.
- Loughran, Tim, und Bill McDonald. „When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks.“ 35-60. 2011.

- Moshe Koppel, Jonathan Schler. „The Importance of Neutral Examples for Learning Sentiment.“ *Computational Intelligence* 22(2):100-109, 5 2006.
- MSCI INC. „OFFICIAL TRANSLATIONS OF GICS STRUCTURE EFFECTIVE OCT 1, 2018.“ *msci.com*. 28. September 2018. https://www.msci.com/documents/1296102/11185315/GICS_map+2018_German.xlsx (Zugriff am 21. September 2022).
- PennState Eberly College of Science - STAT 462 - Applied Regression Analysis. *online.stat.psu.edu*. 2018. <https://online.stat.psu.edu/stat462/node/96/> (Zugriff am 2022. September 23).
- Remus, Robert, Uwe Quasthoff, und Gerhard Heyer. „SentiWS – a Publicly Available German-language Resource for Sentiment Analysis.“ *Proceedings of the 7th International Language Resources and Evaluation (LREC'10)*, 2010: 1168-1171.
- Renault, Thomas. „Intraday online investor sentiment and return patterns in the U.S. stock market.“ Paris, 2017.
- Sardelich, Marcelo, und Dimitar Kazakov. „Extending the Loughran and McDonald Financial Sentiment Words List from 10-K Corporate Filings using Social Media Texts.“ York, 2018.
- Scherer, Klaus R. „Vocal communication of emotion: A review of research paradigms.“ *Speech Communication* 40, 2003: 227–256.
- Schiller, Robert J. „From Efficient Markets Theory to Behavioral Finance.“ *Journal of Economic Perspectives, Volume 17, Number 1*, 2003: 83-104.
- Sun, Licheng, Mohammad Najand, und Jiancheng Shen. „Stock Return Predictability and Investor Sentiment: A High-Frequency Perspective.“ Norfolk, 2015.
- Universität Zürich. *uzh.ch*. 21. März 2022. https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/korrelation.html (Zugriff am 24. September 2022).
- Valle-Cruz, David, Vanessa Fernandez-Cortez, Asdrubal Lopez-Chau, und Rodrigo Sandoval-Almazan. „Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods.“ *Cognitive Computation* 14, Juli 2020: 372-387.
- Walter, Christian. „The Efficient Market Hypothesis, the Gaussian Assumption, and the Investment Management Industry.“ *SSRN Electronic Journal*, November 2000: 6-9.

Yaoshu, Wang, Qin Jianbin, und Wang Wei. „Efficient Approximate Entity Matching Using Jaro-Winkler Distance.“ *Web Information Systems Engineering – WISE 2017*. 2017. 231-239.

Zaiontz, Charles. *www.real-statistics.com - REAL STATISTICS USING EXCEL*. Juni 2022. <http://www.real-statistics.com/correlation/basic-concepts-correlation/> (Zugriff am 23. September 2022).

Zhang, Xue, Hauke Fuehres, und Peter A. Gloor. „Predicting Asset Value through Twitter Buzz.“ *Advances in Collective Intelligence and Soft Computing*. 2012. 23-34.