

Inwieweit nimmt Elon Musk durch sein Verhalten auf der Social-Media-Plattform X (ehem. Twitter) Einfluss auf den Kursverlauf von Tesla und Dogecoin?

Eine Analyse und Bewertung der Big Data Bandits
(Daniel Linck, Daniel Weißenberger, David Helwich, Jonas Sigmund)

3. Dezember 2023

Inhaltsverzeichnis

1 Einleitung

Da gerade das digitale Zeitalter geprägt von dem Einfluss großer Social-Media-Plattformen ist, haben die Big Data Bandits beschlossen, im Rahmen der Klausuraufgabe diese Extremen genauer zu betrachten. Unter dem Aspekt, dass Elon Musk wohl zu den populärsten, einflussreichsten und dominantesten Personen des öffentlichen Lebens zählt, nominiert ihn insbesondere sein “einzigartiges” Korrespondenzverhalten gepaart mit seinen wirtschaftlichen Interessen für eine genauere Untersuchung. Im folgenden Abstract findet eine Beantwortung und Beurteilung der Fragestellung im Rahmen der gestellten Aufgabe statt. Dafür analysieren wir die entsprechenden Datensätze mit einem für die Aufgabe entworfenem Python Script. Dieses kann, aufgrund seiner hohen Modularität, auch in der Zukunft zur Beantwortung ähnlicher Fragestellungen verwendet werden. Weitere Details bezüglich des Codes können der [code.main.ipynb](#) entnommen werden.

2 Fragestellung

Die Fragestellung: “Inwieweit nimmt Elon Musk durch sein Verhalten auf der Social-Media-Plattform X (ehem. Twitter) Einfluss auf den Kursverlauf von Tesla und Dogecoin?” wurde in mehreren Iterationsschritten optimiert. Dies ist insbesondere dem Zeitraum geschuldet, in welchem die Auswertung abgeschlossen sein musste. Somit war es uns zwar nicht möglich, eine allgemeinere Fragestellung über den Einfluss von Privatpersonen auf den Aktienmarkt unter der Berücksichtigung verschiedener Personen zu beantworten. Jedoch konnten wir nun zusätzliche Legalitätsaspekte, sowie konkrete Empfehlungen für ein profitables Investieren als Privatperson mit in unsere Ausarbeitung nehmen.

3 Datensätze

Die Datensätze entstammen der Community-Plattform [Kaggle](#). Im Folgenden verwenden wir einen Datensatz, der alle Posts (“Tweets”) des Tesla-CEOs Elon Musk beinhaltet: [Posts](#), sowie zwei weitere Datensätze, die den Kursverlauf von [Dogecoin](#) und [Tesla](#) umfassen. Alle Datensätze vermitteln einen seriösen und schlüssigen Eindruck, was unter anderem an den fehlerfreien und plausiblen Werten liegt. Zudem wurde der Tesla-Datensatz an den Aktiensplit angepasst und unter Verwendung der offiziellen Yahoo Finance API erstellt. Für die Auswertung werden insbesondere die Zeilen “date” und “raw_close” der Stock-Datensätze, sowie die Spalten “datetime” und “text” des Post-Datensatzes betrachtet.

3.1 Data Understanding

3.1.1 Datendichte



Abbildung 1: Datendichte

Die erste Grafik zeigt die gesamte Zeitspanne (04.06.2010 - 12.10.2023) aller gesammelter Daten aus den drei Datensätzen. Dabei beschreibt der rote Anteil alle fehlenden Tage des jeweiligen Datensatzes. Die Regelmäßigkeit der fehlenden Daten im Tesla-Datensatz kann auf die geschlossenen Tage der Aktienbörse zurückgeführt werden; demzufolge, weist der Dogecoin-Datensatz auch keine fehlenden Datensätze auf, da die Kryptomärkte keiner Schließzeit unterliegen. Die prozentuale Anzahl der fehlenden Tage kann der zweiten Grafik entnommen werden. Darüber hinaus lässt sich aus dem Schaubild die Aussage ableiten, dass das Korrespondenzverhalten von Elon Musk auf der Plattform X über die Zeit zunimmt. Zusammen mit der eineinhalbjährigen Pause nach seinem ersten Tweet am 04.06.2010 lässt dies die Vermutung offen, dass Musk erst nach und nach das Potenzial der Plattform erkannte, bis er diese schließlich am 04.04.2022 –unter fragwürdigen Umständen– aufkaufte. Die Auswertung der fehlenden Tage wird durch die Methode

`get_density(df:pd.DataFrame)` ermöglicht, welche zunächst eine Liste aller Tage erstellt, die im jeweiligen Datensatz vorhanden sein sollten. Im Folgenden wird ein neuer Datensatz erstellt, der alle Tage, sowie den entsprechenden booleschen Wert (1 vorhanden / 0 fehlend) für jeden Tag des zu untersuchenden Datensatzes enthält. Dieser resultierende Datensatz kann nun mit einem Eventplot skizziert werden und bietet die Möglichkeit, den Prozentsatz der fehlenden Tage zu ermitteln.

```
def get_density(df:pd.DataFrame):
    start_date = min(df["date"]).replace(hour=0, minute=0, second=0, microsecond=0)
    end_date = max(df["date"]).replace(hour=0, minute=0, second=0, microsecond=0) + pd.Timedelta(days=1)

    all_dates = pd.DataFrame({"date": pd.date_range(start_date, end_date, freq='D')})
    all_dates["exists"] = all_dates["date"].isin(df["date"]).astype(int)
    return all_dates
```

3.2 Data Preperation

Da die Datensätze schon sehr gut strukturiert sind, ist es nicht nötig eine erweiterte Data Preperation durchzuführen. Optional könnten die ungenutzten Spalten entfernt werden, um Rechenzeit und Speichergröße zu verringern, jedoch ist beides unerheblich –und würde die Weiterentwicklung erheblich einschränken.–

4 Lösungsansatz

4.1 Hilfsfunktionen

4.1.1 Intervalselection

Die Methode `check_filter()` erkennt, ob in einem Text gewisse Schlüsselwörter aus einer Filter-Liste enthalten sind. Dies ermöglicht das Filtern der Posts anhand der eventuell im Tweet enthaltenen Schlüsselwörter (bspw: “Dogecoin”). Dabei kann in Abhängigkeit der Variable `hit`, entweder nach dem Vorhandensein (`hit=True`) oder dem Fehlen (`hit=False`) der Schlüsselwörter sortiert werden. Die Methode `get_interval` generiert einen Datensatz-Ausschnitt, der ausgehend von einem Startdatum die nächsten vorhandenen n Tage aus einem Stock-Datensatz enthält. Die Anzahl der Tage wird durch den Parameter `interval_length` gegeben und beträgt standardmäßig den Wert 5 sozusagen einer Arbeitswoche. Es wird außerdem berücksichtigt, dass das Startdatum nicht im Stock-Datensatz enthalten ist. In diesem Fall nimmt das Unterprogramm das nächste vorhandene Datum. Ausgehend von dem gefundenen Datum, werden der Start-Index sowie der End-Index des Datensatz-Ausschnittes berechnet.

4.1.2 Kursverlauf

4.2 Hauptschleife

Die Hauptfunktion umfasst die grundlegende Schleife zur Auswertung der Datensätze in Bezug auf die Fragestellung. Dafür wird ein neuer Datensatz “influence” erstellt, der für jeden Tag, an dem ein Post mit den entsprechenden Filterkriterien verfasst wurde, das Datum, eine Liste aller Posts, die Anzahl dieser Posts sowie den prozentualen Unterschied der dafür vorgesehenen Hilfsfunktionen enthält. So iteriert diese Schleife zunächst über alle Tweets, die im Datensatz Posts enthalten sind. Im Anschluss wird nun durch die Methode `check_filter` überprüft, ob die aktuelle Zeile den Filteransprüchen gerecht wird. Trifft dies nicht zu, so wird diese Zeile übersprungen. Ansonsten wird das Datum aus der aktuellen Zeile ausgelesen. Im Folgenden wird nun überprüft, ob der aktuelle Tag schon im Datensatz “influence” enthalten ist. Was passiert, wenn an einem Tag mehrere Posts vorliegen? Trifft dies zu, so wird dieser nur um den aktuellen Posttext ergänzt, sowie die Anzahl der Posts an diesem Tag inkrementiert. Ist dieser Tag noch nicht im Datensatz “influences” enthalten, so wird für den aktuellen Tag zunächst das Intervall des Stock-Datensatzes bestimmt. Ist dies nicht möglich, so wird eine Warnung ausgegeben und die Schleife fährt fort. (Dies ist jedoch wie in ?? beschrieben zu vernachlässigen). Wenn das Intervall erfolgreich erstellt wurde, werden nun alle Berechnungen auf das Intervall angewandt. Schlussendlich werden die Werte unter diesem Tag in den Datensatz “influences” eingefügt. Hat die Schleife alle Posts abgearbeitet, kann dieser Datensatz zurückgegeben werden. Zudem wird der prozentuale Anteil der im “influence” enthaltenen Posts von allen Posts ermittelt und zurückgegeben.

```
def get_influence(stock:pd.DataFrame, posts:pd.DataFrame=Posts, filter_list:list=[], hit:bool=False,):
    # find a common start date
    start_date = max(posts['date'].min(), stock['date'].min())
```

```

# adjust Posts to startdate
posts = posts[posts['date'] >= start_date]
# adjust Stock to startdate
stock = stock[stock['date'] >= start_date]

influence:pd.DataFrame = pd.DataFrame(columns=['date', 'posts', 'count_posts', 'trend', 'absolute']
)

j = 0
old_date = None
for i, post in posts.iterrows():
    if not check_filter(str(post['text']), filter_list, hit=hit):
        continue
    else:
        date = pd.to_datetime(post['date']).normalize()
        if not date == old_date:
            interval = get_interval(date, stock)
            if interval is None:
                print("No Data Available")
                continue
            cp_trend, model = get_trend(interval)
            cp_absolute = get_absolute(interval)
            influence.loc[j] = [date] + [[post['text']]] + [0] + [cp_trend] + [cp_absolute]
            old_date = date
            j += 1
        else:
            influence.loc[j-1, 'posts'].append(post['text'])
            influence.loc[j-1, 'count_posts'] += 1
return [influence, j/i*100]

```

4.3 Erzeugung

Filter.

5 Auswertung

5.1 Posts

5.1.1 Verteilung

Dieses Diagramm zeigt die Verteilung der entsprechenden Schlüsselwörter in allen Posts des zu betrachtenden Zeitraums. Dabei ist es wichtig zu beachten, dass der zu betrachtende Zeitraum zusammen mit den Aufzeichnungen des Kursverlaufs beginnt. Demzufolge wird beispielsweise die Verteilung der Schlüsselwörter für Dogecoin nur für den Zeitraum des Dogecoin-Datensatzes erstellt. Interessant ist hierbei die Tatsache, dass Elon Musk fast doppelt so viel über Tesla twitert als über Dogecoin. Insbesondere wenn man sich die kombinierten Schlüsselwörter über den Zeitraum des Dogecoin-Datensatzes (da Dogecoin-Datensatz < Tesla-Datensatz) (3. Kuchendiagramm) näher anschaut, ist deutlich zu erkennen, dass selbst hier Tesla einen Anteil von 28.8% besitzt (da 41.1%-12.3%). Dies ist immer noch mehr als doppelt so viel als die 12.3% der Dogecoin-Posts. Dies könnte daran liegen, dass Dogecoin von Elon Musk mehr als eine Art Neben- oder Spaßprojekt angesehen wird; seine Verbundenheit zu Tesla ist auch schon alleine durch seine Position als CEO gegeben.

5.1.2 Verlauf

Dieses Diagramm zeigt den Tag (x-Achse), sowie die Anzahl der Posts an diesem Tag (y-Achse). Auch hier sieht man die in ?? aufgeworfene These, dass das Korrespondenzverhalten von Elon Musk im Verlauf der Zeit zunimmt. Zudem wird diese Vermutung nun auch durch die Anzahl der Posts an einem Tag gestützt. Denn dieser ist zu entnehmen, dass Elon Musk nicht nur an mehreren Tagen, sondern vor allem auch deutlich häufiger pro Tag tweetet. Dieses Verhalten kann insbesondere seit der Übernahme der Plattform am 04.04.2022 beobachtet werden. Während im Sommer 2019 ein Großteil seiner Posts die Schlüsselwörter von Tesla beinhalteten, postet Elon Musk mittlerweile schon fast genau so viel über Dogecoin wie über Tesla. Das könnte ein Indiz dafür sein, dass Dogecoin für Musk zurzeit an Interesse gewonnen hat.

5.1.3 Zusammenhänge

5.2 Trend

5.2.1 Verlauf

Die analysierten Diagramme zeigen den Einfluss von Elon Musks Posts auf die Kursentwicklung von Dogecoin und Tesla. Bezüglich Dogecoin fällt auf, dass Musks Tweets nur geringfügige Auswirkungen zu haben scheinen, da die prozentualen Veränderungen im Vergleich zu den Trends ohne das Stichwort “Doge” oft relativ gering sind. Es gibt sogar Situationen, in denen die Trends ohne Musks Tweets deutlich stärkere Veränderungen aufweisen als mit. Im Fall von Tesla hingegen ist erkennbar, dass Musks Tweets einen deutlich stärkeren Einfluss auf den Aktienverlauf haben. Insbesondere ab 2018 ist eine Zunahme seiner Tweets zu beobachten. Die Diagramme zeigen Spitzen, bei denen ein Tweet den Trend sowohl positiv als auch negativ beeinflusst. Dennoch verläuft der Trend größtenteils ähnlich, ob mit oder ohne einen entsprechenden Tweet. Dies legt nahe, dass Musk zwar in der Lage ist, den Aktienkurs von Tesla durch seine Tweets zu beeinflussen, aber dass auch andere, uns unbekannte Faktoren eine bedeutende Rolle spielen. In den meisten Fällen scheint sich die Aktie so zu verhalten, als ob Musks Tweets den Markt nicht wesentlich beeinflussen würden. Dennoch gibt es bestimmte Punkte, an denen Musk in der Lage ist, den Trend der Aktie stark zu beeinflussen. Dabei bleibt jedoch unklar, ob diese Veränderungen gezielt geplant sind oder eher Zufälle sind.

5.2.2 Durchschnitt

In diesen Grafiken werden Tesla und Dogecoin anhand des durchschnittlichen Aktientrends verglichen. In beiden Fällen wird einmal der Trend genommen ohne, dass Elon Musk das Wort “Tesla” oder “Doge” in seinen Tweets genutzt hat und einmal, wenn der Post die jeweiligen Filterwörter beinhaltet. In der linken Grafik wird auf den Dogecoin Trend geschaut, bei Tweets ohne die Verwendung des Wortes “Doge” und mit der Verwendung des Wortes “Doge”. Hier kann man gut sehen, dass der durchschnittliche Trend ohne die Verwendung von Wörtern, die in Verbindung mit Dogecoin stehen bei 1.08 liegt. Sobald Elon Musk diese Filterwörter in seinem Tweet verwendet, liegt dieser Trend nur noch bei 1.02. Daraus kann man hier ziehen, dass Elon Musk durch die Verwendung des Wortes “Doge” in seinen Tweets einen Einfluss auf den Trend der Aktie Dogecoin haben kann. In diesem Fall ist sein Einfluss in den nächsten 5 Tagen negativ. In der rechten Grafik wird auf die Tesla Aktie geschaut, hier wird nach Verwendung des Wortes “Tesla” in seinen Tweets geschaut. Hier liegt der Wert ohne die Verwendung des Wortes bei 1.01 und wenn er es verwendet bei 1.02. Dadurch kann man bei der rechten Grafik ziehen, dass es bei der Tesla Aktie keine wirklichen Auswirkungen auf den allgemeinen Trend der Aktie gab.

5.3 Absolute

5.3.1 Verlauf

Nun haben wir die absoluten Werte in % von Dogecoin und Tesla nach eine Post von Musk. Bei dem ersten Diagramm erkennt man, dass die Tweets über Dogecoin sogar weniger Einfluss haben als, wenn Musk einfach nur so twitterte, daraus lässt sich schließen, dass bei Dogecoin, Musks Einfluss so gering ist, dass er praktisch gar nicht vorhanden ist, oder zumindest irrelevant ist. Bei Tesla jedoch gibt es einige Punkte bei denen sich die Aktie stark veränderte, –aufgrund– bzw. nach einem Tweet von Musk. Jedoch sieht man auch das die Aktie auch ohne einen solchen Tweet sich oft stark veränderte, und somit kann man sagen das Musk zwar einen gewissen Einfluss hat, aber diesen nicht oft nutzt. Das könnte z. B. daran liegen, dass es eventuell illegal ist, oder die Menschen sonst erkennen würden, wann er manipuliert und das dann für sich ausnutzen würden.

5.3.2 Durchschnitt

In dieser Grafik schauen wir uns die Auswirkung auf die absoluten Werte der beiden Aktien an, hier kann man direkt auf den ersten Blick einen etwas größeren Unterschied erkennen. Bei Dogecoin ist der absolute Wert bei 2.03, wenn Elon Musk das Wort nicht in seinen Tweets erwähnt hat, aber wenn er es benutzt hat, steigt der absolute Wert auf 2.35. Hieraus lässt sich schließen, dass er mit Sicherheit etwas Einfluss darauf hat, auch wenn es nicht viel ist. Genau dasselbe mit der Tesla Aktie. Hier liegt der absolute Wert bei 1.08, wenn er es in seinen Tweets nicht erwähnt, aber wenn es erwähnt wird, steigt der absolute Wert auf 1.63 an. Hier ist sein Einfluss sogar noch etwas stärker als bei Dogecoin.

6 Gesamtbewertung / Erwartung

6.1 Die Legalität

Die Legalität und die Auswirkung von Elon Musks Tweets auf den Aktienmarkt sind ein kompliziertes Thema. Normalerweise müssen Äußerungen insbesondere von einflussreichen Personen, Managern und Geschäftsführern von börsenorientierten Unternehmen in dem Rahmen der geltenden Gesetze und Vorschriften erfolgen. Die SEC (Securities and Exchange Commission) hat spezifische Regeln für Offenlegungen durch Direktoren börsennotierter Unternehmen. Der Zweck dieser Regeln besteht darin, sicherzustellen, dass relevante Informationen fair und gleichzeitig für alle Marktteilnehmer zugänglich sind. Musk und Tesla standen in der Vergangenheit unter Beobachtung der SEC, insbesondere aufgrund von Elon Musks Tweets über Tesla. Einige von Musks Tweets wurden als potenzielle Verstöße gegen die SEC-Regeln angesehen, was zu Rechtsstreitigkeiten und Vergleichen mit der SEC führte. Es ist wichtig zu beachten, dass die Rechtslage in solchen Fällen weitgehend von den konkreten Umständen und den geltenden Gesetzen abhängt. Musk und sein Unternehmen müssen sich noch mit der öffentlichen Kommunikation und ihren Auswirkungen auf den Aktienmarkt befassen.

7 Fazit

Anhand der Erkenntnisse, die wir aus den jeweiligen Auswertungen erlangt haben, können wir die genaue Fragestellung beantworten. Bezüglich der Legalität der Einflussnahme von Elon Musk auf den Aktienmarkt kann man sagen, dass er sich in einer Grauzone bewegt. Grundsätzlich ist es nur im Rahmen der Gesetze möglich, für Personen, die sehr viel Einfluss haben, Äußerungen zu tätigen. Besonders auf Social-Media-Plattformen wie X (ehemals Twitter) ist eine Einflussnahme auf den Aktienmarkt gesetzlich nicht erlaubt. Elon Musk ist eine dieser einflussreichen Personen und daher hat auch er die Pflicht, sich an diese Vorschriften zu halten, jedoch nutzt er den gegebenen Spielraum hin und wieder aus.

Welchen Einfluss seine Äußerungen auf Social Media schlussendlich wirklich haben, konnten wir ebenfalls herausfinden. Anhand unserer Beobachtungen, dass der Trend und die absoluten Werte der Aktie Tesla sich immer wieder durch Musk beeinflussen lassen, schließen wir, dass Musk durchaus den Kursverlauf von Tesla relativ stark beeinflusst. Jedoch nur meist gemäßigt und nur selten stark. Dieses Verhalten lässt sich mit der Legalität begründen, denn so überschreitet Musk nicht die Grenze.

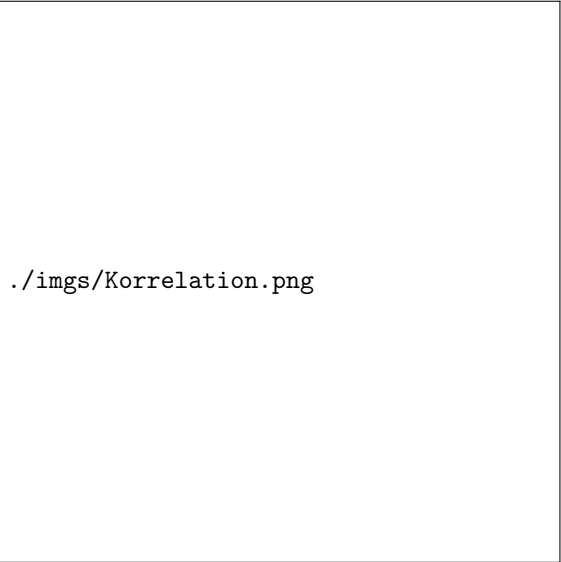
Auf Dogecoin hat Musk zwar weniger Einfluss, aber auch hier übt er diesen immer wieder aus. Um als Privatperson auf solch einen Einfluss besser vorbereitet zu sein, könnte man nun das Programm noch so erweitern, dass es eine Zukunftsvorhersage berechnet und ausgibt. Aufgrund dieser Vorhersagen kann man dann seine Investments besser planen und durchführen. Aus zeitlichen Gründen war uns nur die Auswertung möglich, wie der Trend sich allgemein verändert, wenn Musk etwas mit Tesla oder Dogecoin postet. Aufgrund dieser Auswertung kann man nicht automatisch nach einem neuen Post sagen, wie der Kurs sich nun verhalten wird. Eine Programmerweiterung könnte, wenn man genauer anschaut, was gepostet wird, dieses Problem beheben und fundierte Vorhersagen erstellen.

./imgs/Verteilung.png

Abbildung 2: Verteilung

./imgs/Verlauf.png

Abbildung 3: Datendichte



./imgs/Korrelation.png

Abbildung 4: Korrelation

Der wichtigste Aspekt, der der Korrelationstabelle des Einfluss-Datensatzes entnommen werden kann, ist die Unabhängigkeit zwischen der Spalte “count_posts” und den Werten “trend” und “absolute”. Dies zeigt, dass die Anzahl der Posts (Tweets) an einem Tag keinen Einfluss auf den Verlauf der beiden Wertpapiere hat. Die Korrelationstabelle wurde unter Verwendung der Bibliothek Seaborn aus der Vereinigung der Schlüsselwort freien Einfluss-Datensatz erstellt.

```
influences_corr = pd.concat([
    influences_t_d[['count_posts', 'trend', 'absolute']],
    influences_t_t[['count_posts', 'trend', 'absolute']]
]).corr()
sns.heatmap(influences_corr, annot=True, fmt=".2f")
plt.show()
```

./imgs/Trend.png

Abbildung 5: Trend

./imgs/Trend_Durchschnitt.png

Abbildung 6: Trend Durchschnitt

./imgs/Absolut.png

Abbildung 7: Absolut

./imgs/Absolut_Durchschnitt.png

Abbildung 8: Absolut Durchschnitt