**Projekttitel: Aktienmärkte KO durch KI**

Teilnehmer (mit Alter): Wladimir Mikliaev (20)

Erarbeitungsort: Bad Essen

Projektbetreuer/in: --

Thema des Projekts: KI unterstützte Analyse der Finanzmärkte

Fachgebiet: Informatik

Wettbewerbssparte: Jugend forscht

Bundesland: Niedersachsen

Wettbewerbsjahr: 2020

## Kurzfassung

Die Programme mit künstlicher Intelligenz erobert immer mehr Industrien. Sie überarbeiten Bilder, übersetzen für uns Texte, steuern Autos. In meinem Projekt bin ich der Frage nachgegangen, ob es möglich ist mit Hilfe von KI die Zukunft zu vorhersagen. Die Aktienmärkte bieten dafür viele Daten, die den Stand der Gesellschaft wiederspiegeln.

Inhalt

[Einleitung 2](#_Toc58825506)

[Vorgehensweise, Materialien und Methoden 2](#_Toc58825507)

[Ergebnisse 3](#_Toc58825508)

[Ergebnisdiskussion 3](#_Toc58825509)

[Zusammenfassung 4](#_Toc58825510)

[Unterstützungsleistungen 5](#_Toc58825511)

## Einleitung

**Warum habt ihr gerade dieses Thema gewählt? Wie seid ihr auf das Thema   
gekommen?**

Ich interessiere mich für die Aktien seit einiger Zeit. Zum einen spiegeln die Aktien den Stand der Wirtschaft und der Gesellschaft. Zum anderen lerne ich bei der Analyse der Aktienkurse viel Neues in der Statistik und der Mathematik.

**Was ist das Ziel eurer Arbeit? Welche Hypothese(n)/Vermutung(en))wollt ihr mit eurer Arbeit überprüfen?**

In meiner Arbeit wollte ich untersuchen, ob man den Kurs einer Aktie anhand vieler anderer Kurse vorhersagen kann. Meine Annahme ist es, dass die Kurse einzelner Aktien voneinander nicht unabhängig sind. Unterschiedliche Ereignisse in der Welt müssen einen Einfluss auf die Aktienkurse haben. Diese Ereignisse sind z.B. Politische Entscheidungen, Umweltkatastrophen, Geschehnisse in der Wirtschaft, Mediale Ereignisse, Stimmungen in der Bevölkerung.

In den Zeiten von Big-Data, wo alles elektronisch erfasst wird, können diese Ereignisse elektronisch zeitnah ausgewertet und somit genutzt werden. Es gibt etliche Anbieter, die diese Daten erfassen, aufbereiten und meist für einen Entgelt anbieten. Die Abfrage dieser Daten aus dem Internet bedarf jedoch meist kostenpflichtige Zugänge und auf einen konkreten Aktienkurs zugeschnittene Aufbereitung.

Ich habe mir überlegt, dass alle Aktienkurse diese Ereignisse in irgendeiner Form mehr oder weniger widerspiegeln. Falls die Aktienkurse unterschiedlicher Aktien unterschiedlich schnell auf die Ereignisse reagieren, kann man die Kurse, die etwas langsamer reagieren, anhand der schneller reagierenden Aktienkurse vorhersagen.

**Was ist der aktuelle Stand der Forschung bzw. Stand der Technik im Zusammenhang mit eurem Thema?**

Ich habe nach meinem Ansatz recherchiert, habe jedoch nicht gefunden, ob jemand es angewandt hat. Im Allgemeinen werden die Methoden zur Aktienkursvorhersage selten im Detail offen gelegt. Schließlich verdienen diejenigen, die eine effektive Methode finden, mit dieser Methode Geld. Laut der Aussage von Dirk Müller, soll es gar keine mathematische Methode möglich sein.

Diese Aussage halte ich jedoch für übertrieben. Mit den Methoden der Künstlichen Intelligenz sollte es möglich sein, die Zusammenhänge in Daten, die von einem Menschen bewusst oder intuitiv erkannt werden können, sollten auch von einem KI-Algorithmus erkannt werden.

Was ist die Künstliche Intelligenz? Im Allgemeinen kann jede Art von automatischer Entscheidungen, als künstliche Intelligenz bezeichnet werden. Mit KI wird jedoch gemeint, dass das Treffen von Entscheigungen dem Computer nicht mit fest-programmierten Algorithmen beigebracht wird, sondern lernt der Computer von den Daten, Entscheidungen zu treffen. Dafür braucht man größere Datenmengen, so dass einem Datensatz eine eindeutige Entscheidung zugeordnet werden kann.

Wie die Daten zum anlernen der KI aussehen, kann man am einfachsten an dem sehr bekannten Iris-Beispiel verstehen. [IRIS]

Iris-Datensatz besteht aus 150 Messungen von 4 Attributen von den drei Arten der Schwertlilien. Gemessen wurden jeweils die Breite und die Länge des Kelch- sowie des Kronblatts. Für jeden Datensatz bestehend aus den 4 Werten wird die Art der Schwertlilie angegeben. Für jede Lilienart liegen 50 Datensätze vor. Die Tabelle unten zeigt einen Auszug aus dem Datensatz. Die Buchstaben A, B, C, D sind die Attribute der Pflanze (Breite und die Länge des Kelchblatts sowie des Kronblatts in Zentimeter). Die Attribute werden in der KI-Sprache als Features bezeichnet und die Werte von A, B, C, D bilden einen Feature-Vector. Die Antwort, in diesem Fall die Art der Blume nennt man in der KI-Sprache Class. Mit einem der Zahlreichen Klassifizierungsalgorithmen kann man mit diesen Daten ein Programm so anlernen, dass wenn wir ein Feature-Vector von einer unbekannten Blume nehmen, können wir die Art der Blume 100% erkennen. Im Internet findet man zahlreiche Beispiele rund um den Iris-Datensatz [IRIS Example].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **A** | **B** | **C** | **D** | **Art** |
| 5,1 | 3,5 | 1,4 | 0,2 | Setosa |
| 4,9 | 3 | 1,4 | 0,2 | Setosa |
| … | … | … | … | … |
| 7 | 3,2 | 4,7 | 1,4 | Versicolor |
| 6,4 | 3,2 | 4,5 | 1,5 | Versicolor |
| … | … | … | … | … |
| 5,8 | 2,7 | 5,1 | 1,9 | Virginica |
| 7,1 | 3 | 5,9 | 2,1 | Virginica |
| … | … | … | … | … |

Das Iris-Beispiel hat ein Feature-Vector von nur vier Werten. Die KI-Algorithmen sind jedoch fähig, die Datensätze mit tausenden Attributen zu analysieren. Das Python-Framework Sci-Kit Learn [SCI-KIT] bietet etwa Dutzend KI-Algorithmen.

Der Datenstz von den Aktiendaten enthält zu jedem Tag, an dem gehandelt wurde, den Wert der Aktie beim Öffnen und Schliessen der Börse sowie den höchsten und den tiefsten Wert der Aktie.

Zum Anwenden dieser Daten mit einem KI-Algorithmus, sollen die Daten anders formatiert werden. Das Feature-Vector soll die Aktienwerte mehrerer aneinander folgenden Tage enthalten. Die Antwort in dem Datensatz ist die Differenz zwischen dem letzten in dem Feature-Vector benutzten Wert und dem Wert des nächsten Tages.

## Vorgehensweise, Materialien und Methoden

Als erstes habe ich die Aktiendaten mit einem Python-Skript heruntergeladen. Das Skript ist in Github unter downloader.py abgelegt. Die wesentlichen Zeilen sind:

|  |
| --- |
| *import yfinance as yf #Importieren der Yahoo-Finance Bibliothek*  *#Historische Daten einer bestimmter Aktie (Variable t) herunterladen*  *hist = yf.download( t, start='2019-01-01', end='2020-10-13', interval="1d")*  *# Aktienwerte als CSV-Datei abspeichern*  *hist.to\_csv("%s/%s.csv" % (save\_path,t))* |

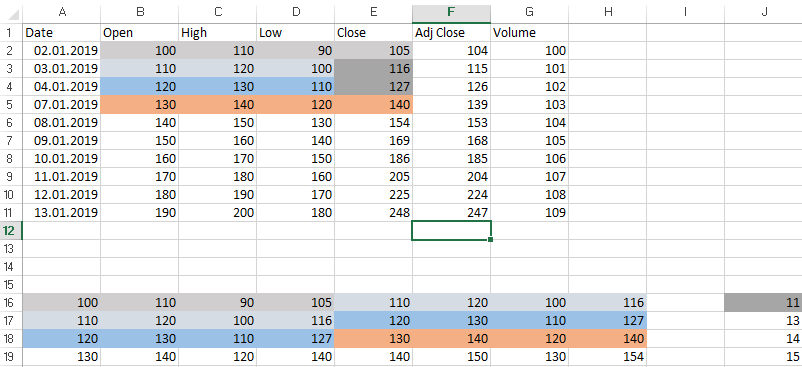
Jetzt, wo ich alle Aktienkurse beisamen als CSV-Dateien in einem Verzeichnis hatte, konnte ich alle Daten mit Pandas laden. Pandas [PANDAS] ist eine Python Bibliothek zum Arbeiten mit Tabellendaten. So eine Art Excel für die Konsole. Das Laden der CSV-Dateien mache ich mit dem Skript dataLoader.py. Mit Pandas geht das Lesen der CSV-Dateien mit einer Zeile. Die geladenen Pandas-Daten werden in einem Dictionary stock\_map abgelegt. Das Dictionary stock\_map enthält unter dem Schlüssel „Aktien-Kürzel“ die gesamten Aktiendaten.

|  |
| --- |
| *stockdata = pd.read\_csv("%s/%s.csv" % (load\_path,t))*  *stock\_map[t] = stockdata* |

Nun kommt der schwierigste Teil – Aufbau der Daten zum Anlernen und Testen der KI. Die Untere Abbildung zeigt zwei Tabellen. Die obere Tabelle enthält die Ausgangsdaten der Aktiendaten. Die untere Tabelle zeigt die Form, die zum Anlernen und Testen der KI geeignet ist. Es ist nur ein vereinfachtes Beispiel, in dem Als Feature-Vector die Daten von zwei aneinander folgenden Tagen genommen werden. Für das bessere Verständnis habe ich mit den Farben die Zeilen markiert. Die Werte von den Zeilen 2 und 3 kommen in die Zeile 16, die Werte von den Zeilen 3 und 4 kommen in die Zeile 17, usw.

Zu jedem Feature-Vector wird auch eine Antwort berechnet. Die Antwort die die Differenz des Wertes von dem Letzten Tag in dem Feature-Vector und dem Wert gleicher Spalte von dem darauf folgenden Tag. Zum Beispiel, die Antwort vom Feature-Vector in der Zeile 16 ist die Differenz zwischen den Zellen 4E und 3E. Die Differenz ist Positiv, wenn der Aktienkurs steigt und negativ, wenn der Aktienkurs fällt. Dieses Beispiel ist eine übertriebene Vereinfachung. Mit diesen Daten könnte man voraussagen, ob der Aktienkurs fällt oder steigt nur anhand der Kurse letzter zwei Tage von der gleicher Aktie. Für meine Idee brauche ich viele Tage von vielen Aktien.

In dem Skript dataBuilder.py habe ich die Umformung der Daten so generisch implementiert, dass ich die Anzahl der Aktien und die Anzahl der Tage frei wählen kann.



Die Daten werden anschließend in den Script modelBuilder.py normalisiert und in zwei Gruppen separiert. 75% aller zufällig Feature-Vektoren benutzt die KI zum Lernen, die Restliche 25% werden für das Überprüfen des Algorithmen verwendet. Anschließend berichtet die KI, wie Erfolgreich es war.

## Ergebnisse

Meine KI zum Vorhersagen der Aktienmärkte ist im Allgemeinen gelungen.

Nachdem alle Aktiendaten heruntergeladen worden sind, habe ich alle mögliche KI Algorithmen ausprobiert, die Ski-Kit anbietet. Der beste Algo war Support Vector Maschine, mit einer Erfolgsrate von . Dies würde zwar einen Trader nicht ermöglichen an den Mark sehr profitabel zu spekulieren, aber es hat gezeigt, dass ein KI allein durch die Kurse von vielen Aktien ermitteln kann, was wahrscheinlicher ist, eine positive, oder negative Entwicklung.

Bei den ersten Versuch hat die Künstliche Intelligenz gescheitert - das Erwartungswert aller Vorhersagen lag bei 50%. Das bedeutet, dass die KI nicht besser als ein Münzenwurf war. Das Problem wurde behoben, indem ich Aktien selektiert habe, die an der Gleichen Börse gehandelt wurden. Verschiedene Börsen haben verschiedene Regeln zum Schlusszeiten und umfassen verschiedene Länder sowie Wirtschaftssektoren. Das Begrenzen auf einen Markt macht alle Aktiendaten relevant zu einander.

Meine zweite Aufgabe war, zu ermitteln welche Zeitabschnitte für die KI am besten passen. Ich habe ermittelt, dass sich die Öffnungskurse, Schlusskurse und Hoch-Tiefpunkte von jedem Tag an der Börse sich gut eignen. Kürzere Zeitabschnitte überfordern den Prozessor und längere bereiten zu wenig Daten zu Verfügung.

Von allen KI-Modellen, sollte die das beste rausgesucht werden: Mit einem Abstand von ... der ist die SVM, das beste Klassifikator geworden.

## Ergebnisdiskussion

* Dieser Abschnitt sollte nicht länger als **zwei Seiten** sein. Hier interpretiert ihr eure Ergebnisse.
* Beantwortet dafür folgende Fragen:
  + Was ist gut gelaufen? Was ist schlecht gelaufen?
  + Worauf musstet ihr im Verlauf eurer Projektarbeit verzichten?
  + Wie passen eure Ergebnisse zu dem, was ihr über euer Thema gelesen oder gehört habt? Stimmt ihr mit euren Ergebnissen der allgemeinen Meinung bzw. der Meinung von Wissenschaftlern zu?
  + Habt ihr durch eure Experimente und Ergebnisse neue Ideen bekommen, wie ihr das Projekt weiterentwickeln könnt?
  + Habt ihr Fehler gemacht oder hat etwas nicht funktioniert? Dann beschreibt auch das. Aus Fehlern lernt man schließlich. Erklärt, was ihr in Zukunft an der Vorgehensweise ändern oder verbessern könntet.
  + Welche Folgen kann eure Entdeckung, Erfindung oder Forschung für die Gesellschaft, den Arbeits­platz, die Wissenschaft oder Menschen haben?
  + Welche Forschungslücke konntet ihr möglicherweise schließen?

## Zusammenfassung

* Am Schluss greift ihr auf maximal einer Seite die Forschungsfrage bzw. das Projektziel nochmal auf: Wie lautet eure Antwort auf die Forschungsfrage? Habt ihr euer Ziel erreicht?

## Quellen- und Literaturverzeichnis

1. Iris Datensatz, <https://en.wikipedia.org/wiki/Iris_flower_data_set>, besucht am 07.01.2021
2. Iris Example with Python, <https://towardsdatascience.com/exploring-classifiers-with-python-scikit-learn-iris-dataset-2bcb490d2e1b>, besucht am 07,01,2021
3. Sci-Kit Learn, https://scikit-learn.org/stable/auto\_examples/classification/plot\_classifier\_comparison.html, besucht am 07,01,2021
4. Pandas, <https://pandas.pydata.org/>, besucht am 07,01,2021

Ihr müsst alle Quellen und Unterstützungsleistungen nennen, die ihr für euer Projekt verwendet und in Anspruch genommen habt. Quellen sind z. B. Internetseiten, Fachzeitschriften und Bücher. Alle Angaben werden jeweils alphabetisch nach Nachnamen sortiert. Benutzt dafür euer Rechercheprotokoll. Wie ihr die unterschiedlichen Quellen richtig angebt, findet ihr in der nachfolgenden Box.

**Quellenangabe bei Internetseiten**

* Genaue URL (Webadresse) und Datum des Seitenaufrufs
* Verfasser/Verantwortlicher der Seite, Titel und Thema des Inhalts

*Beispiel: Stiftung Jugend forscht e. V., schriftliche Arbeit und weiterführende Informationen,* [*http://jugend-forscht.de/teilnahme/ablauf/schriftliche-arbeit.html*](http://jugend-forscht.de/teilnahme/ablauf/schriftliche-arbeit.html)*, besucht am 12.07.2018.*

**Quellenangabe bei Büchern**

* Vorname und Nachname des Autors
* Buchtitel
* Erscheinungsort
* Erscheinungsjahr
* Seitenangabe

*Beispiel: Andrea Gruß, Ute Hänsler: „Knallraketen und Gummigeister“, Frankfurt/Main 2007, S.10*

**Quellenangabe bei Zeitschriften**

* Zusätzlich zu den Angaben wie bei Büchern werden der Name der Zeitschrift, die Nummer der Ausgabe und die Seitenangabe des Artikels angegeben.

*Beispiel: Susanne Strunk: „Prozesse und Wirkungen der Teilnahme an Schulwettbewerben“ in Die Deutsche Schule, Zeitschrift für Erziehungswissenschaft, 104. Jahrgang, Heft 2, 2012, S. 150*

**Quellenangabe bei Fotos**

Quellenangaben für Fotos werden im Allgemeinen direkt unter das Foto gesetzt:

* Agentur, Unternehmen oder Institution
* Vorname und Nachname des Fotografen

*Beispiel: Foto: Agentur Krummen, Robert Schnappschuss*

## Unterstützungsleistungen

Die Angabe von Unterstützungsleistungen ist als Textblock am Ende der schriftlichen Arbeit unter einer eigenen Überschrift aufzuführen. Hierbei sind Personen, Unternehmen und Institutionen zu nennen, die euch unterstützt haben*.* Wie ihr Unterstützungsleistungen richtig angebt, findet ihr in der nachfolgenden Box.

**Angabe von Unterstützungsleistungen**

Folgende Angaben sind erforderlich:

* Vorname und Nachname des persönlichen Unterstützers
* Funktion und Berufsbezeichnung
* Name der Institution/des Unternehmens
* Art der Unterstützung, z. B. Durchführung von Messungen oder Programmtestläufen, Erstellung von Modellen, Korrektur von Texten, Beratung bei der Themenwahl, Bereitstellung von Geräten und Materialien

*Beispiel: Dr. Maria Mathus, Informatikerin, Simsen AG, Düsseldorf, hat mich bei der Erstellung meiner Arbeit unterstützt. Sie hat mir den Test des Programms auf einem Großrechner ermöglicht und mich bei der Themenwahl beraten…*