



Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Machine Learning Prognosen in der Finanzwirtschaft: Zuverlässig oder doch nur ein aufwendiger Münzwurf?

Bachelorthesis Zusammenfassung

Hochschule für Wirtschaft und Recht Berlin

Eingereicht von: Molitor, Mick

E-Mail: s_molitor19@hwr-berlin.de

Matrikelnummer: 77211771147

Erstgutachter/in: LL.M. Sebastian Block

Zweitgutachter/in: Prof. Dr. Martin Uzik

Berlin, 05.09.2022

Inhaltsverzeichnis

| | |
|--|---|
| 1. Einleitung | 3 |
| 2. Zielstellung und Aufbau der Arbeit | 3 |
| 3. Methode | 4 |
| 4. Ergebnisse | 4 |
| 5. Schlussfolgerung | 8 |
| 6. Literaturverzeichnis | 9 |
| 7. Abbildungsverzeichnis | 9 |
| 8. Tabellenverzeichnis | 9 |

1. Einleitung

Der Begriff „Maschinelles Lernen“ bezeichnet ein Unterkonzept der Künstlichen Intelligenz (KI). Er wurde 1959 von einem IBM-Mitarbeiter namens Arthur Samuel geprägt. Maschinelles Lernen beschreibt Techniken, die es Computern ermöglichen, in Datensätzen Muster zu erkennen, diese zu verstehen und an künstliche intelligente Systeme weiterzugeben (Gogas and Papadimitriou, 2021).

Die Implementierung von maschinellem Lernen bringt relevante Vorteile für alle Industrien mit sich. Aber besonders in der Finanzwirtschaft ist davon auszugehen, dass das maschinelle Lernen zum alltäglichen Begleiter von Analysten, Vermögensverwaltern und Aktienhändlern wird (Tatsat et al., 2020, p. 1)., etwa beim Handel mit Aktien und Wertpapieren.

Besonders bei Aktien ist der Reiz, solche Systeme zur Analyse des Marktes zu nutzen, groß. Sie könnten sich als nützliches Instrument in der Kundenberatung erweisen. Denn die scheinbar zufälligen Schwankungen von Aktientrends halten manche Investoren vom Handel mit Aktien ab. Was wäre aber wenn diese Schwankungen auf einmal vorhersagbar wären? Wenn ihnen entgegengewirkt werden könnte?

2. Zielstellung und Aufbau der Arbeit

Diese Arbeit versucht dem Leser den Begriff des maschinellen Lernens näher zu bringen und anhand einer konkreten Anwendung zu illustrieren. Dies erfolgt aufgrund der Aktienkursvorhersage von 16 DAX 30 Unternehmen. Somit wird versucht, ein solches maschinelles Lernmodell von der Konstruktion bis hin zur Ergebnisinterpretationen darzustellen.

3. Methode

Für die literarischen Grundlagen wurde Fachliteratur in Form von Büchern, Fachzeitschriften, Internetseiten, Zeitungen und Studien analysiert und zitiert.

Für den empirischen Teil der Arbeit wurden Aktienkursdaten von der Internetseite Yahoo! Finance genutzt. Die Zeitspanne der geladenen Aktienschlusspreise für die 16 DAX30 Unternehmen waren identisch. Für den Trainingszeitraum, dem Zeitraum, aus welchem die Modelle lernten, wurden die Aktienschlusspreise vom 01.01.2018 bis zum 31.12.2021 heruntergeladen. Für den Testzeitraum, den Vergleichszeitraum, wurden die Aktienkursschlusspreise vom 01.01.2022 bis zum 31.03.2022 heruntergeladen. Diese Daten wurden dann in eine Python 3.10 Umgebung hereingeladen und weiterverarbeitet. Die Daten wurden mithilfe der Python Module pandas_datareader und Yfinance in die Python Umgebung hereingeladen. Die Weiterverarbeitung erfolgte durch verschiedene Python Module (pandas, numpy, matplotlib, sklearn).

4. Ergebnisse

Im ersten Teil der Arbeit wurde das Kernprinzip und die Funktionalität der künstlichen Intelligenz und des maschinellen Lernens dargestellt. Dabei wurde zuerst die Herkunft des maschinellen Lernens erläutert und dann die wichtigsten Grundbegriffe definiert und die Unterschiede aufgezeigt. Man hat den allgemeinen Vergleich vom maschinellen Lernen zu der wissenschaftlichen Disziplin der Statistik dargestellt. Daraufhin wurden die verschiedenen Arten des maschinellen Lernens beschrieben und mit Beispielen veranschaulicht. Schlussendlich wurde eine Verbindung zwischen maschinellem Lernen und der Finanzwirtschaft dargestellt und die Nützlichkeit von bereits existierenden Systemen aufgewiesen.

Im zweiten Teil der Arbeit ging es darum ein Anwendungsfall für ein maschinelles Lernsystem zu entwickeln und die Ergebnisse mit realen Zahlen/Fakten zu vergleichen. Dafür wurden drei maschinelle Lernsysteme programmiert. Dabei handelte es sich um eine lineare Regression, ein ARIMA Modell und ein LSTM-Modell. Neben der eigentlichen Aktienkursvorhersage durch die verschiedenen Algorithmen wurden ebenfalls die theoretischen Grundlagen dargestellt.

Die drei Modelle wurden auf Trainingsdaten trainiert (Aktienchlusspreise vom 01.01.2018-31.12.2021) und sollten dann Vorhersagen für die drei kommenden Monate treffen. Diese Vorhersagen wurden dann mit Testdaten (Aktienchlusspreise vom 01.01.2022-31.03.2022) verglichen. Die Abweichung von den durch die Modelle vorhergesagten Werten und den realen Aktienpreisen wurde durch den Root Mean Squared Error dargestellt. Der Root Mean Squared Error (Wurzel des mittleren quadratischen Fehlers) ist eine Fehlermaßzahl oder Fehlerkriterium, welches die Zuverlässigkeit und die Genauigkeit eines Modells darstellen kann.

In der Tabelle 1 wurden für die drei verschiedene Modelle die entsprechenden Trainingsfehler und die Testfehler dargestellt. Aus dieser Tabelle kann herausgelesen werden welches Modell, für den entsprechenden Aktienkurs, die besten Ergebnisse geliefert hat. Je niedriger der Root Mean Squared Error ist, desto näher war die Vorhersage an dem realen Aktienkurspreis.

Erkennbar ist, dass in der Regel der Trainingsfehler niedriger ist als der Testfehler. Dies resultiert daraus, dass die Modelle auf den Trainingsdaten trainiert wurden und so bei Vorhersagen anhand von bekannten Daten das Verhalten der Aktienkurse "wiedererkennen" und bei den Vorhersagen anwenden können. Die Testfehler sind meist weniger akkurat, da es sich hierbei um Vorhersagen für einen längeren Zeitraum handelt, für welchen keine realen Daten während der Vorhersage vorliegen.

| Symbol | Firmenname | Lineare Regression | | ARIMA | | LSTM | |
|----------------|--------------------------|--------------------|---------|--------|---------|--------|---------|
| | | RMSE | RMSE | RMSE | RMSE | RMSE | RMSE |
| | | Train | Test | Train | Test | Train | Test |
| 1COV.DE | Covestro AG | 13,9175 | 11,6568 | 2,9040 | 5,3570 | 1,4269 | 4,5397 |
| ALV.DE | Allianz SE | 17,3243 | 18,7119 | 6,7289 | 14,0662 | 5,0743 | 13,3399 |
| BAS.DE | BASF SE | 9,5406 | 9,7979 | 3,0563 | 6,5710 | 2,5959 | 7,8641 |
| BAYN.DE | Bayer Aktiengesellschaft | 8,8178 | 14,8345 | 3,4300 | 7,5471 | 1,7038 | 4,7986 |

| | | | | | | | |
|----------------|---|---------|---------|--------|---------|--------|---------|
| BMW.DE | Bayerische Motoren Werke Aktiengesellschaft | 11,9086 | 17,4515 | 2,9844 | 8,8591 | 2,0500 | 8,2711 |
| DBK.DE | Deutsche Bank AG | 2,1184 | 3,1252 | 0,6026 | 1,5549 | 0,3275 | 2,5193 |
| DHER.DE | Delivery Hero SE | 14,3765 | 74,0097 | 2,3593 | 44,1485 | 3,6913 | 60,6284 |
| DPW.DE | Deutsche Post AG | 7,0386 | 5,3546 | 1,5109 | 20,7226 | 1,0650 | 8,2534 |
| DTE.DE | Deutsche Telekom AG | 1,0199 | 0,6286 | 0,5026 | 0,6659 | 0,2497 | 0,7592 |
| HFG.DE | HelloFresh SE | 11,7938 | 33,3767 | 1,4039 | 19,9806 | 3,0091 | 28,3259 |
| IFX.DE | Infineon Technologies AG | 5,2225 | 4,3479 | 0,9093 | 8,4605 | 0,8103 | 6,2137 |
| RWE.DE | RWE AG | 2,4798 | 1,2879 | 0,7418 | 1,8473 | 0,8933 | 2,2359 |
| SAP | SAP SE | 11,2385 | 26,9178 | 4,2797 | 21,5692 | 3,4511 | 24,7713 |
| SIE.DE | Siemens Aktiengesellschaft | 14,1792 | 11,2689 | 4,0701 | 19,1080 | 3,3609 | 11,6015 |
| VOW3.DE | Volkswagen AG | 25,0058 | 20,4631 | 6,2682 | 17,4702 | 6,3238 | 23,7377 |
| ZAL.DE | Zalando SE | 13,3496 | 34,0689 | 1,9801 | 14,8475 | 3,1760 | 27,5526 |

Tabelle 1 Ergebnisse Kursvorhersage (Train/Test) – eigene Darstellung

In den Abbildungen 1-3 werden die Kursvorhersagen der verschiedenen Modelle für die Covestro AG dargestellt. Diese visualisieren die Fehlermaße aus der Tabelle 1 und zeigen die Vorhersage im Vergleich zu den realen Aktienkurspreisen. Für alle anderen Aktienkurse und

deren Vorhersagen wurden ähnliche Abbildungen erstellt.

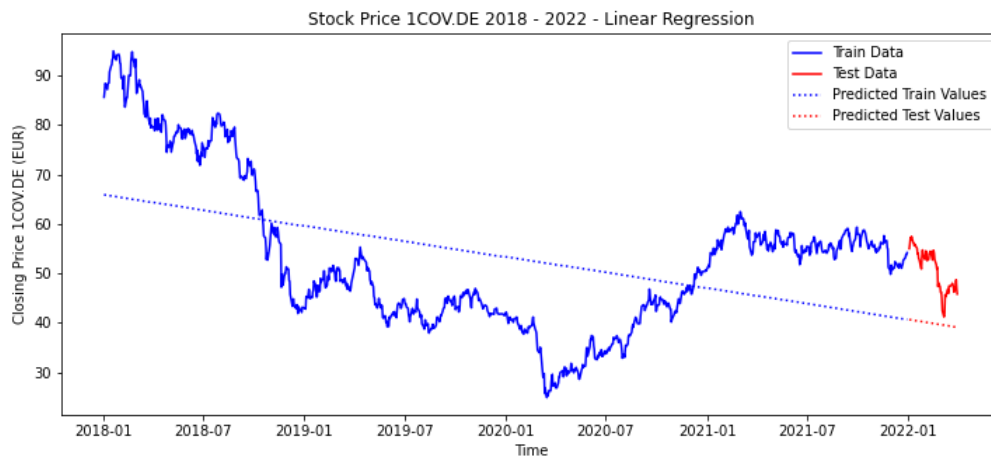


Abbildung 1 1COV.DE - Lineare Regression - eigene Darstellung

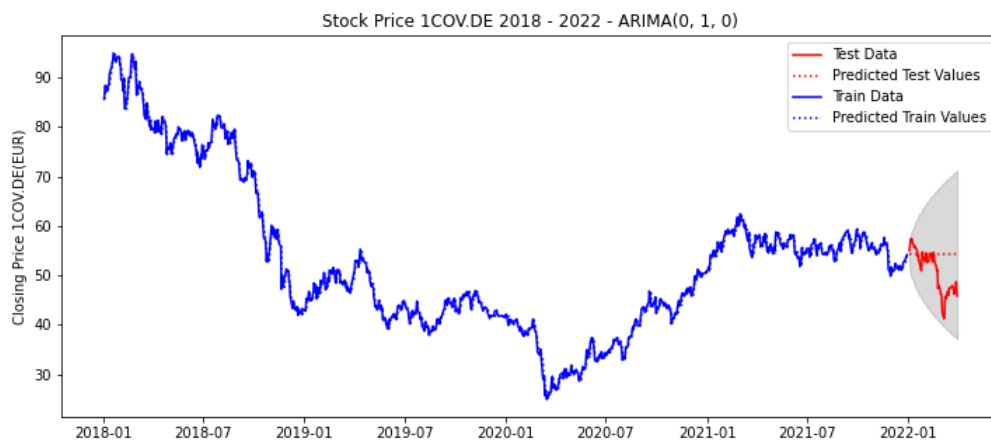


Abbildung 2 1COV.DE - ARIMA Modell - eigene Darstellung

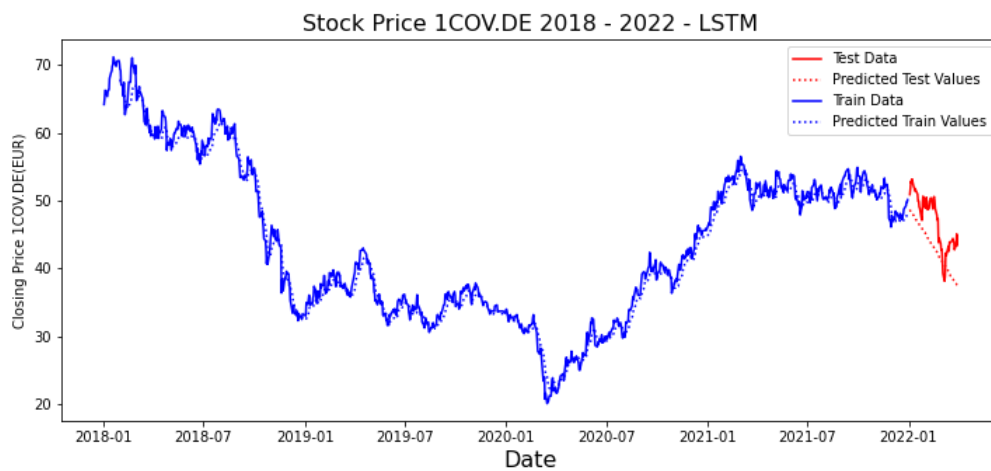


Abbildung 3 1COV.DE - LSTM Modell - eigene Darstellung

Das Experiment der Aktienkursvorhersage ermöglichte, dass folgende Schlussfolgerungen

aus den verschiedenen Modellen entnommen werden konnte:

Die lineare Regression scheint nicht besonders geeignet zu sein, um Vorhersagen auf Basis einer Zeitreihe zu geben, da lediglich der Trend des Trainingszeitraums, im Testzeitraum, wiedergegeben wird. Zur Trendbestimmung ist die lineare Regression jedoch nützlich und kann den Trend eines Aktienkurses über einen gegebenen Zeitraum gut wiedergeben.

Das ARIMA Modell scheint das wohl zuverlässigste Modell zu sein. Im großen Vergleich mit den anderen Modellen hat dieses im Durchschnitt den niedrigsten $RMSE_{Train}$ sowie $RMSE_{Test}$. Die Resultate des Modells sind sehr solide. Die vom Modell erstellten Konfidenzintervalle konnten in fast allen Fällen den tatsächlichen Kurs der Aktie gut einfangen.

Die Vorhersagen des LSTM-Modells auf Basis der Trainingsdaten waren in den meisten Fällen sehr nah an den eigentlichen Aktienkurspreisen. Die Vorhersage anhand von eigens vorhergesagten Daten (Testzeitraum) war jedoch nicht besonders aussagekräftig, da bei Wiederholungen der Vorhersagen, sich die Ergebnisse mit jedem Durchlauf änderten und die Vorhersagen so oftmals eher zufällig erschienen.

5. Schlussfolgerung

Noch scheint maschinellen Lernsystemen die Fähigkeit zu fehlen, die Zukunft anhand weniger Zeitreihendaten vorherzusagen. Ihr Potenzial ist aber unverkennbar: Die Vorhersagen sind ausschließlich auf der Grundlage historischer Aktienkurse entstanden und weisen zum Teil trotzdem eine erstaunlich hohe Präzision auf. Ob die Vorhersagen tatsächlich von erlernten Mustern innerhalb der Datensätze geleitet wurden oder doch nur Zufallstreffer waren, lässt sich schwer sagen.

Spannend ist jedoch, dass die maschinellen Lernsysteme erst am Anfang ihrer Entwicklung stehen. Durch die Integration weiterer Datenquellen könnten Vorhersagen wesentlich genauer werden. Neben Zahlen-, könnten in Zukunft auch Ton-, Zeichen- und Bilddaten einen Einfluss auf die Prognosen nehmen, indem neben den Zeitreihendaten der Aktien- und Währungskurse

auch Nachrichtentexte und -bilder auf nützliche Informationen hin untersucht und in die Vorhersage mit einfließen würden.

Wie viel Zeit solche Entwicklungen in Anspruch nehmen werden ist genauso schwer zu beantworten, wie die Frage, wie präzise die Prognosen dann sein werden. Fakt ist jedoch, dass in wenigen Jahren riesige Fortschritte in der Perfektionierung dieser Lernsysteme erzielt werden konnten, und dass täglich weiter an ihnen gearbeitet und geforscht wird. So entstehen immer wieder neue Algorithmen und Lernverfahren, die neue Einsatzmöglichkeiten von maschinellem Lernen in Aussicht stellen. Ob die Aktienkursvorhersage dazugehört, wird und was in diesem Bereich in den kommenden Jahren zu erwarten ist, wird zu beobachten sein.

6. Literaturverzeichnis

- Gogas, P., Papadimitriou, T., 2021. Machine Learning in Economics and Finance. *Comput. Econ.* 57, 1–4. <https://doi.org/10.1007/s10614-021-10094-w>
- Tatsat, H., Puri, S., Lookabaugh, B., 2020. Machine learning and data science blueprints for finance: from building trading strategies to robo-advisors using Python, First edition. ed. O'Reilly Media, Sebastopol, CA.

7. Abbildungsverzeichnis

| | |
|---|---|
| Abbildung 1 ICOV.DE - Lineare Regression - eigene Darstellung | 7 |
| Abbildung 2 ICOV.DE - ARIMA Modell - eigene Darstellung | 7 |
| Abbildung 3 ICOV.DE - LSTM Modell - eigene Darstellung | 7 |

8. Tabellenverzeichnis

| | |
|--|---|
| Tabelle 1 Ergebnisse Kursvorhersage (Train/Test) – eigene Darstellung..... | 6 |
|--|---|