

Prognose von Aktienkursen mit maschinellen Lernverfahren

Maximilian Schepp

Technische Hochschule Mittelhessen

Fachbereich Mathematik, Naturwissenschaften und
Datenverarbeitung
Wilhelm-Leuschner-Str. 13
61169 Friedberg (Hessen)
maximilian.schepp@mnd.thm.de

Prof. Dr. Harald Ritz

Technische Hochschule Mittelhessen

Fachbereich Mathematik, Naturwissenschaften und
Informatik
Wiesenstraße 14
35390 Gießen
harald.ritz@mni.thm.de

Kategorie

Masterarbeit

Schlüsselwörter

Machine Learning, maschinelles Lernen, Neuronale Netze, Text Mining, Sentimentanalyse, technische Analyse, Web Scraping, Python

Zusammenfassung

Im Zuge der Digitalisierung und im Zeitalter von Industrie 4.0 werden immer mehr Daten generiert. Beispielsweise sorgen mit Sensoren ausgestattete Fertigungsanlagen für eine große Anzahl an Daten, die wichtige Erkenntnisse für die Optimierung und Automatisierung des Fertigungsprozesses in produzierenden Unternehmen liefern. Generell bieten Kunden-, Log- und Sensordaten enormes Informations- und Wissenspotential, das jedoch erst durch die richtigen Vorhersagemodelle ausgeschöpft werden kann. In diesem Zusammenhang eröffnen Algorithmen und Modelle aus dem Bereich der künstlichen Intelligenz sowohl Unternehmen als auch Privatpersonen völlig neue Möglichkeiten und Chancen, um aus den großen Datenmengen gewinnbringende Informationen für komplexe Sachverhalte zu extrahieren.

Ein komplexes Themengebiet, das außerdem eine große Menge an Daten für Auswertungs- und Analyse Zwecke aufweist, ist die Prognose von zukünftigen Aktienkursen. Bislang ist wenig darüber bekannt, welche Ergebnisse ein Modell erzielt, das zusätzlich zu ausgewerteten Twitter- und Börsennachrichten andere Faktoren klassischer Bewertungstechniken für Aktienkurse einbezieht, um die Anzahl möglicher Einflussfaktoren zu erhöhen. Ein Beispiel für solche klassischen Bewertungstechniken stellen Indikatoren der technischen Analyse der Finanzmärkte dar, die aus den Zeitreihen der Aktienkursdaten berechnet werden.

Im Rahmen dieser Masterarbeit soll daher beantwortet werden, ob ausgewählte Indikatoren der technischen Analyse mit den Ergebnissen einer Sentimentanalyse von Twitter- und Börsennachrichten kombiniert werden können, um ein Modell des maschinellen Lernens zu

erstellen, mit dem Gewinne am Aktienmarkt erzielt werden können.

Für die Erstellung des Modells wurden Twitter- und Börsennachrichten, sowie historische Aktienkursdaten zu den Unternehmen des Nasdaq 100-Index gesammelt. Die Twitter- und Börsennachrichten wurden für jedes Unternehmen gruppiert und einer Sentimentanalyse unterzogen, wodurch für jeden Tag des Untersuchungszeitraum vom 28.10.2019 bis zum 31.01.2020 jeweils ein Sentiment für die Twitter- und Börsennachrichten pro Unternehmen vorlag. Die gesammelten Daten wurden anschließend in eine einheitliche Struktur zusammengeführt. Durch die Datenzusammenführung konnten die Daten im weiteren Verlauf untersucht werden, um erste Informationen bezüglich der Verteilung und Korrelation der einzelnen Features zu gewinnen. Die im Rahmen der Datenuntersuchung gewonnenen Erkenntnisse wurden im nächsten Schritt für die Vorbereitung der Daten eingesetzt, um den Trainingsprozess des Modells zu optimieren. Die vorbereiteten Daten wurden schließlich für das Training eines vorwärtsgerichteten neuronalen Netzes verwendet, das darauf trainiert wurde, geeignete Kauf- und Verkaufspunkte für ausgewählte Aktien zu identifizieren.

Auf Basis der vorliegenden Testdaten erzielt das erstellte Modell bei einem Startkapital von \$19.000 in einem Zeitraum von 3 Monaten einen Gewinn von \$1.155,83, was einer Rendite von 6,08% entspricht. Die zentrale Erkenntnis ist, dass das erstellte Modell Ergebnisse liefert, mit denen sich vereinzelte, jedoch keine garantierten Gewinne am Aktienmarkt erzielen lassen. Weiterhin werden in der Arbeit Punkte genannt, mit denen sich die Prognosen des Modells weiter verbessern lassen. Die Ergebnisse einer Sentimentanalyse von Twitter- und Börsennachrichten, in Kombination mit den gewählten Indikatoren der technischen Analyse der Finanzmärkte, stellen daher eine solide Datenbasis für die Identifizierung von geeigneten Kauf- und Verkaufspunkten für Aktien dar.