

Seminar

“Neueste Trends in Big Data Analytics”

Thema: Stock Market Prediction

Teilnehmer: Clemens Becker

Abgabedatum: 30.03.2018

Betreuer: Dr. Julian Kunkel

Inhaltsverzeichnis

1. Finanzmarktprodukte	3
1.1 Aktien	3
1.2 Anleihen.....	3
1.3 Derivate	4
1.3.1 Futures (Terminkontrakte)	4
1.3.2 Optionen.....	4
2. Zeitreihen	5
3. ARIMA Modell	6
4. Zusammenfassung Stock Market prediction using Financial News Articles on Ho Chi Minh Stock Exchange.....	6
4.1 Vorgehensweise	7
4.2 Ergebnisse.....	9
5. Zusammenfassung Automated Bitcoin Trading via Machine Learning Alogrithms	9
5.1 Vorgehensweise	10
5.2 Ergebnisse.....	11
6. Eigene Meinung und Zukunftsaussicht	11
7. Literaturverzeichnis.....	13

1. Finanzmarktprodukte

1.1 Aktien

Aktien sind Wertpapiere, mit denen ein Anleger Anteile an einer Aktiengesellschaft erwerben kann. Voraussetzung dafür ist, dass ein Unternehmen sich entscheidet Aktien auszugeben. Dabei wird der Gesamtwert des Unternehmens berechnet und dann der Aktienpreis, anhand der Anzahl der Aktien die ausgegeben werden sollen, ermittelt. Der Vorgang der Ausgabe wird Emission genannt. Falls ein Unternehmen, im Laufe der Jahre, neues Kapital benötigt, können neue Aktien ausgegeben werden. Dies nennt man eine Kapitalerhöhung.¹ Der vorhergehend erwähnte Aktienkurs bezeichnet den Preis, welchen man für eine Aktie zahlen muss. Da Aktien einen Preis haben, werden sie am Aktienmarkt gehandelt. Der Kurs bestimmt sich hier alleinig aus Angebot und Nachfrage. Ein Börsenmakler bringt diese beiden ins Gleichgewicht, indem er den Aktienpreis so anpasst, dass Angebot und Nachfrage für eine Aktie im Gleichgewicht sind.² Es werden am Aktienmarkt jedoch nicht nur einzelne Aktien gehandelt. Es gibt sogenannte Indizes, welche Entwicklungen bestimmter Sektoren abbilden. In solchen Indizes, werden viele Aktienkurse zusammengenommen. Veränderungen verschiedener Unternehmenskurse lassen so den Index schwanken und bilden die angesprochenen Veränderungen ab. Beispiele wären der DAX (Deutscher Aktienindex) oder der amerikanische Dow Jones. Der DAX bildet die Aktienindizes der 30 größten, deutschen Unternehmen ab und zeigt somit die Lage der deutschen Wirtschaft.³ Ein weiteres Finanzprodukt, welches Aktien beinhaltet, sind Fonds. Diese dienen dazu das Risiko der Aktienanlage zu minimieren. Man legt viele Aktien zusammen, welche zu einer bestimmten Gruppe gehören. Ein Beispiel wäre, dass man durch einen Fond in viele Unternehmen der Nahrungsmittelindustrie investiert, anstatt in nur ein Unternehmen dieser Branche.⁴

1.2 Anleihen

Ein anderer Weg für Unternehmen Kapital zu erhalten sind Anleihen. Anleihen sind in ihrer Funktionsweise wie Kredite. Investoren leihen dem Unternehmen Geld und erhalten dafür von dem Unternehmen eine Anleihe. In dieser verspricht das Unternehmen, das geliehene Geld innerhalb einer bestimmten Zeitperiode zurückzuzahlen und dies mit einem festgelegten Zinssatz. Unternehmen nehmen diese Möglichkeit gerne wahr, da es im Vergleich, zu Krediten einer Bank, Vorteile birgt. Die Unternehmen können die Konditionen für die Anleihen, anders als bei Krediten, selbst bestimmen. Dies ermöglicht ihnen eine günstigere Kapitalakquise, als wie über einen Bankkredit möglich wäre.⁵

¹ Vgl.: <https://de.wikipedia.org/wiki/Aktie#Aktiengattungen>

² Vgl.: <https://de.wikipedia.org/wiki/B%C3%B6rsenkurs>

³ Vgl.: <http://www.aktien-kaufen-fuer-anfaenger.de/was-ist-der-dax/>

⁴ Vgl.: <http://www.aktien-kaufen-fuer-anfaenger.de/definition-was-sind-fonds-was-ist-ein-investmentfonds/>

⁵ Vgl.: <https://www.finanzfluss.de/was-sind-anleihen/>

1.3 Derivate

1.3.1 Futures (Terminkontrakte)

Futures oder Terminkontrakte sind „börsengehandelte Formen eines unbedingten Termingeschäfts.“⁶ Sie sind charakterisiert durch einen genau bestimmten Vertragsgegenstand, eine bestimmte Menge und Qualität des Vertragsgegenstandes, einen festgelegten Kaufpreis, einen fixen Erfüllungszeitpunkt und die Art der Erfüllung. Ihre Besonderheit ist dabei, dass sie standardisiert sind und deshalb, wie schon eingehend erwähnt, an der Börse gehandelt werden können.⁷ Ein Beispiel für einen solchen Börsenhandel soll nachfolgend erklärt werden. Wir nehmen an, dass im Januar ein Kontrakt für 100 Unzen Gold geschlossen wird, welcher im Dezember fällig wird. Der aktuelle Goldpreis beträgt zu diesem Zeitpunkt 360\$. Der Preis für das Termingold beträgt 400\$. Der Unterschied von 40\$ ergibt sich daraus, dass derjenige, der das Gold bis zum Erfüllungszeitpunkt lagert, Versicherungs- und Lagerkosten erhält. Weiterhin ist in dieser Gebühr, ein Zinssatz inkludiert, da das Gold auch schon jetzt gewinnbringend verkauft und dieser Ertrag anlegt werden könnte. Dies wird durch den Zinssatz ausgeglichen. Insgesamt hat der Handel also einen Wert von 40.000\$. Der Investor muss eine Garantiezahlung von 5-10% bei Vertragsschluss leisten, dies ist aber deutlich weniger als wenn der das Gold komplett kaufen würde. Er kontrolliert mit nur einem Bruchteil des Wertes somit das komplette Gold. Gehen wir nun ein paar Monate voran und gehen davon aus, dass nach einem halben Jahr der Preis für Gold steigt. Der Investor hat nun die Möglichkeit bis Dezember zu warten und das Gold liefern zu lassen oder seinen Gewinn zum aktuellen Zeitpunkt zu realisieren. Er tut dies, indem er einen exakt gleichen Kontrakt, zu dem welchen er im Januar abgeschlossen hat, an einen Interessenten verkauft. Der Goldwert ist in unserem Beispiel auf 400\$ gestiegen. Er würde somit den Goldpreis im Kontrakt auf 440\$ ansetzen, da er die gleiche Gebühr verlangt, wie er sie im ersten Kontrakt bezahlt hat. Dieser Handel hat folglich einen Gesamtwert von 44.000\$ und der Investor hat sich 4000\$ Gewinn gesichert.⁸

1.3.2 Optionen

Optionen werden in Call und Put Optionen unterteilt. Der Unterschied zu Terminkontrakten liegt darin, dass der Käufer der Option keine Pflicht hat den Handel auszuführen, sondern sich nur ein Recht sichert. Call Optionen beschreiben im Kern, dass man sich bereit erklärt nach einer bestimmten Zeit, etwas für einen bestimmten Wert und in einer bestimmten Menge zu verkaufen. Für das Einräumen einer solchen Option erhält der Anbieter eine Prämie.⁹ Wir gehen in einem Beispiel davon aus, dass sich ein Händler die Option für 10 Tonnen Getreide zu 200€ sichern will, da er davon ausgeht, dass der Getreidepreis in den nächsten Monaten stark steigen wird. Er kauft sich eine Kaufoption für 10 Tonnen Getreide zu 200€ und zahlt dem Anbieter eine anteilige Prämie. Falls der Getreidepreis nun wie erwartet steigt und zum Ausübungszeitpunkt bei 250€ pro Tonne liegt, ist der Verkäufer verpflichtet dem Händler 10 Tonnen zu 200€ zu verkaufen. Falls der Getreidepreis nicht steigt, sondern auf 195€ pro Tonne fällt, wird der Händler seine Option nicht einlösen, sondern auf sie verzichten. In diesem Szenario macht der Anbieter der Option Gewinn, da er die Prämie behalten kann und das Getreide immer noch mit einem minimalen Verlust verkaufen kann, welcher die Prämie nicht übersteigt.¹⁰ Bei einer Put Option erklärt man sich bereit, etwas nach einer bestimmten Zeit, für einen bestimmten Preis und in einer bestimmten Menge zu kaufen. Auch hier erhält man für das Einräumen einer solchen

⁶ Wikipedia, 30.03.2018, <https://de.wikipedia.org/wiki/Terminkontrakt>

⁷ Vgl.: <https://de.wikipedia.org/wiki/Terminkontrakt>

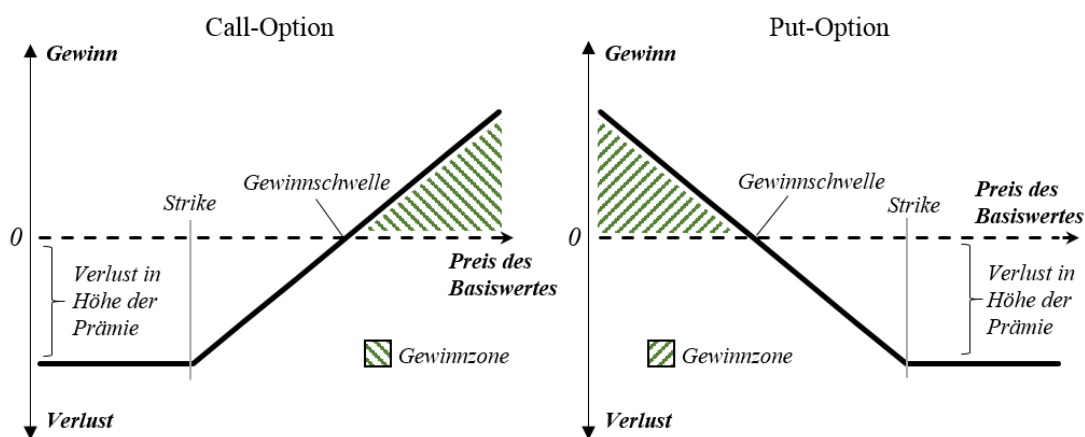
⁸ Vgl.: <https://www.boerse.de/grundlagen/eurex/Was-ist-ein-Futures-Kontrakt-6>

⁹ Vgl.: <https://www.lynxbroker.de/wissen/erfolgreich-handeln-mit-optionen/grundlagen-von-optionen/>

¹⁰ Vgl.: <https://de.wikipedia.org/wiki/Kaufoption>

Option eine Prämie.¹¹ Im gleichen Beispiel von oben nehmen wir nun an, dass ein Händler sich gegen ein Absacken des Getreidepreises, welchen er erwartet, absichern möchte. Er kauft sich deshalb eine Verkaufsoption, in welcher er sich die Möglichkeit sichert 10 Tonnen Getreide für 200€ pro Tonne zu verkaufen. Falls der Getreidepreis nun wie erwartet fällt und zum Beispiel einen Wert von 160€ erreicht, kann der Händler seine Option einlösen und dem Verkäufer der Option 10 Tonnen Getreide für 200€ pro Tonne verkaufen. Falls der Getreidepreis aber gleichbleibt, steigt oder nur geringfügig fällt, macht der Verkäufer Gewinn. Der Händler wird nämlich bei einem gleichbleibenden oder steigenden Getreidepreis sein Gut nicht unter Wert verkaufen, somit hat der Verkäufer die Prämie als Gewinn. Bei einem leicht fallenden Getreidepreis gleicht die Prämie den geringen Verlust aus.¹² Optionen werden gerne von großen Unternehmen oder Hedgefonds für Aktien in Anspruch genommen. Indem sie sich Put Optionen auf bestimmte Aktien sichern, können sie Schwankungen derer aussitzen und deren Kursentwicklung beobachten. Wenn die Aktie auf einem niedrigen Kurs verharrt, können sie zum festgelegten Wert verkaufen und falls die Aktie wieder steigt nach einem Einbruch ist ihr einziger Verlust die Prämie die gezahlt werden musste. Es ist meistens so, dass solche Hedgefonds nicht 100 oder 1000 Stück einer Aktie halten, sondern meistens 1.000.000 oder mehr. Diese sind auf einmal schwer zu verkaufen und selbst wenn dies gelingen würde, wäre es unmöglich bei einem erneut steigenden Kurs die gleiche Anzahl, zum vorherigen Verkaufskurs, erneut zu erwerben.¹³

SO FUNKTIONIEREN CALL- UND PUT-OPTION



14

2. Zeitreihen

Zeitreihen sind Abfolgen gemessener Datenpunkte in Abhängigkeit zu einem bestimmten Zeitraum. Mit ihnen werden Daten über Wetterbeobachtungen, Börsenkurse und ähnliches grafisch dargestellt. Weiterhin werden sie verwendet um zukünftige Prognose, durch die Auswertung vergangener Daten, zu tätigen. Auf Zeitreihen wirken regelhafte, sowie zufällige Ursachen ein. Dies heißt, dass sie periodisch (saisonal) variieren oder langfristigen Trends folgen.¹⁵

¹¹ Vgl.: <https://www.lynxbroker.de/wissen/erfolgreich-handeln-mit-optionen/grundlagen-von-optionen/>

¹² Vgl.: <https://de.wikipedia.org/wiki/Verkaufsoption>

¹³ Vgl.: <https://www.boerse.de/grundlagen/boerseneinfuehrung/Was-sind-Futures-und-Optionen-10>

¹⁴ So funktionieren Call- und Put-Optionen, Dr. Manuel Kay und Sara Zinnecker, <https://www.finanztip.de/anleihen/options/>

¹⁵ Vgl.: <https://de.wikipedia.org/wiki/Zeitreihenanalyse>

3. ARIMA Modell

Es ist das Auto Regressive Integrated Moving Average Modell und eines der Standardmodelle zur Analyse von Zeitreihen. Der Nutzen besteht in einer Beschreibung von Datenreihen in einer Zeitreihenanalyse. Es teilt sich auf in einen Autoregressiven Teil und einen gleitenden Mittelwertbeitrag (Moving Average), welche mit der ersten Ableitung der Zeitreihe kombiniert werden (Integrated). Im Autoregressiven Teil versucht man die Messwerte durch vorangegangene Beobachtungen, anhand von linearen Modellen zu beschreiben. Der Teil des gleitenden Mittelwertbeitrags geht von Schätzungsfehlern bei den Zeitserien aus. Diese Schätz- oder Vorhersagefehler werden bei der Schätzung der nächsten Werte in der Zeitserie einberechnet. Durch die Kombination dieser beiden Modelle entsteht das ARMA-Modell. Anschließend erfolgt eine zusätzliche Differenzierung und Integration und dadurch entsteht das ARIMA-Modell. Dieses Modell wird verwendet, wenn man einen Trend aus einer Zeitreihe herausfiltern möchte.¹⁶

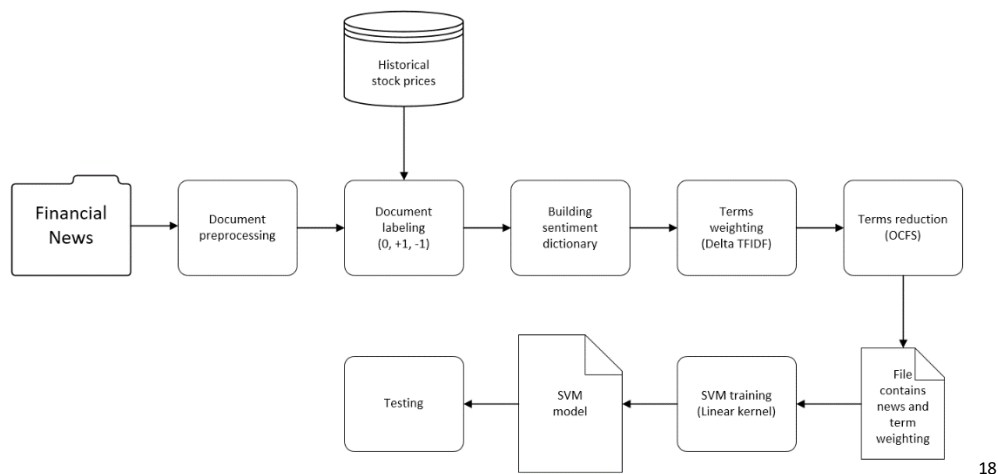
4. Zusammenfassung Stock Market prediction using Financial News Articles on Ho Chi Minh Stock Exchange

Das Ziel der Forscher war es ein Modell zu bilden, welches in der Lage ist anhand von Aktiennews und -preisen die Entwicklung des vietnamesischen Aktienmarktes vorherzusagen. Im Vorfeld gab es zwei Herangehensweisen zur Einordnung von Dokumenten. Herangehensweise 1 beinhaltete, dass ein Experte den Artikel las und diesem eine Kennzeichnung gab, ob er sich positiv oder negativ auf den Aktienpreis auswirken würde. Das Problem war, dass dieser Ansatz zwar eine hohe Erfolgsquote hatte, aber ein Experte nur eine bestimmte Anzahl von Artikeln pro Tag lesen und kategorisieren konnte. Die zweite Herangehensweise war, dass ein Artikel automatisch gekennzeichnet wurde anhand seiner Auswirkung auf den Aktienpreis. Die Problematik bei dieser Herangehensweise bestand darin, dass es viele Faktoren gibt, weshalb sich ein Aktienpreis ändern kann und deshalb viele Artikel falsch kategorisiert wurden. Die Forscher verfolgten einen neuen Ansatz, welcher die Nutzung eines sentiment-Dictionarys beinhaltete. In diesem wird jedes Wort einzeln geprüft ob es einen positiven oder negativen Effekt auf den Artikel hat.¹⁷

¹⁶ Vgl.: <http://www.reiter1.com/Glossar/ARIMA.htm>

¹⁷ Vgl.: Duc Duong, Toan Nguyen, Minh Dang: Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.1-2

4.1 Vorgehensweise



18

Zuerst erfolgte ein Preprocessing der Dokumente. In diesem Schritt wurden alle Dokumente im HTML-Format gesammelt, von HTML-Kommandos gesäubert und der Inhalt wurde in Reintext exportiert. Die nun vorliegenden Artikel wurden Satz für Satz, mit Hilfe eines Programms in einzelne Worte zerlegt. Es werden aus diesen Dokumenten anschließend nur die relevanten Wörter extrahiert und somit alle Stopwörter entfernt. Stopwörter wären zum Beispiel Konjunktionen, Artikel und ähnliches. Durch dieses Vorgehen wird die Effizienz erhöht und die Anzahl der gebrauchten Systemressourcen verringert sich.¹⁹

Darauffolgend wurde festgelegt, in wie viele Klassen die Artikel eingeteilt werden sollen. In den vorherigen Herangehensweisen wurden 2 oder 3 Klassen verwendet. Diese waren gut, schlecht oder neutral. Die Forscher entschieden sich für eine Klassifizierung in 3 Klassen bei ihrem Ansatz, um alle Richtung des Aktienmarktes abbilden zu können. Man nutzt die vorhandenen klassifizierten Artikel als Datenbasis.²⁰

Anschließend erfolgte der Aufbau des sentiment dictionary. Zuerst wurde ein vietnamesisches Wörterbuch im Rein Text heruntergeladen und mit Hilfe eines Programms jedem Wort seine Klasse zugeordnet. Von diesen kategorisierten Wörtern wurden nur die Verben und Adjektive weiterbenutzt. In den vorliegenden Artikeln wurden nun alle vorkommenden Wörter gezählt und geschaut, ob sie im Wörterbuch vorkommen. Mit Hilfe von Formeln wurde berechnet, ob das eben überprüfte Wort positiv oder negativ ist. Alle Wörter aus dem Wörterbuch die in keinem Artikel vorkamen wurden aus dem Wörterbuch entfernt, um die Rechenzeit zu verkürzen.²¹

Der nächste Schritt beinhaltet das Term weighting. Es soll unterschieden werden, ob die Worte in guten oder schlechten Nachrichten vorkommen. Normalerweise wird zu Wortkategorisierung die TF-IDF (Term frequency-inverse document frequency) Methode verwendet. Diese unterscheidet jedoch nicht zwischen positiven und negativen Nachrichten, sondern berechnet die Wortwerte anhand der

¹⁸ An overview of our stock trend prediction process, Duong et al. Dang, <https://dl.acm.org/citation.cfm?id=3457619>

¹⁹ Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.2

²⁰ Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.2

²¹ Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.2

Häufigkeit ihres Auftretens im Text und in wie vielen Dokumenten. Dadurch ist keine aussagekräftige Kategorisierung möglich. Die Forscher verwendeten deshalb in ihrem Ansatz die Delta TF-IDF Methode. In dieser wurden aus den Artikeln zwei Untermengen gebildet. Eine Untermenge enthielt alle positiven Artikel und die andere alle negativen. Für jedes Wort wurden die TF-IDF Werte in beiden Untermengen bestimmt. Anschließend wurde der Wert für die negativen Artikel von dem Wert für die positiven abgezogen. Aus der Differenz ist erkennbar, ob das Wort positiv oder negativ zu kategorisieren ist.²²

I *think* this phone is *good*. [POS]

I *think* this phone is *bad*. [NEG]

This is a *good* phone. [POS]

It has a very *bad* user experience [NEG]



$$\text{Delta TFIDF} = \text{TFIDF_POS} - \text{TFIDF_NEG}$$

Term frequency: 1

Document frequency: 2

23

Im Anschluss folgt die Term reduction. Dieser Schritt diente dem Zweck, die Anzahl der zu untersuchenden Wörter zu verringern, um die benötigte Rechenzeit zu verkürzen. Es wurde dafür die OCFS-Methode benutzt, welche von allen verfügbaren Dimensionreduzierungsalgorithmen die besten Ergebnisse lieferte. Die Methode geht grundlegend so vor, dass zuerst für jede Kategorie von Trainingseinheiten ein Mittelwert bestimmt wird. Anschließend wird für alle Trainingskategorien ein Mittelwert bestimmt und dann für jedes einzelne Wort eine Wertung bestimmt. Schlussendlich werden nur K-Worte mit den höchsten Wertungen ausgewählt und weiterverwendet.²⁴

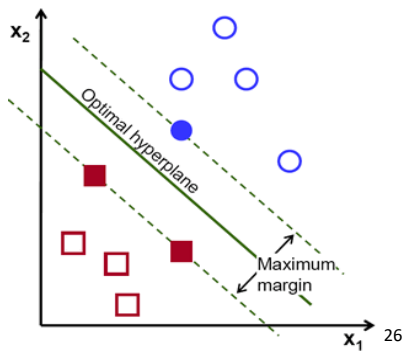
Nun benutzte man die Support Vector Machine als Machine Learning Methode, um die Finanznewsartikel zu klassifizieren. In dieser Methode wird nur binär gearbeitet, das bedeutet die Artikel werden nur in positiv und negativ unterteilt. Man stellt die Artikel in einem Diagramm dar und kann sehen, dass die positiven und negativen Artikel zwei Gruppen in diesem bilden. Die Punkte die sich von diesen beiden Gruppen am nächsten liegen, sind die Extrempunkte der jeweiligen Klasse. Ihre eingezeichneten Vektoren sind die sogenannten Supportvektoren. Zur Trennung der beiden Gruppen zeichnet man zwischen den beiden Supportvektoren eine sogenannte Hyperlane ein. Diese sollte so gezeichnet sein, dass sie die größtmögliche Spanne zwischen den beiden Supportvektoren bildet. Dies dient dem Zweck bei der Einordnung neuer Artikel so wenig Fehler wie möglich zu erhalten.²⁵

²² Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.2-3

²³ How does Delta TF-IDF work?, Hady Elsahar, <https://www.quora.com/How-does-Delta-TF-IDF-work>

²⁴ Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.3

²⁵ Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.3



Der letzte Schritt ist die Data Preparation. Mit Hilfe eines Webcrawlers sammelten die Forscher Artikel von bekannten vietnamesischen Finanzseiten. Diese Artikel betrafen jedoch nur den VN30. Dies ist der vietnamesische Aktienindex und ähnlich dem deutschen DAX. Die täglichen Aktienkurse wurden manuell im gleichen Zeitrahmen wie die Artikel gesammelt. Dabei wurden die Artikel in drei Kategorien eingeteilt, um zu überprüfen, ob die Vorhersageergebnisse von der Zeit und Anzahl der Artikel abhingen. Die drei Zeitspannen waren 4, 8 und 12 Monate.²⁷

4.2 Ergebnisse

Man verglich zuerst die Ergebnisse der drei Kategorisierungsmethoden. Es fand ein Vergleich zwischen TF-IDF, Delta TF-IDF und Delta TF-IDF kombiniert mit dem sentiment dictionary statt. Die Auswertung zeigte, dass die letzte Methode die besten Ergebnisse in der Kategorisierung der gesammelten Artikel lieferte. Die Forscher gaben die Artikel, welche sie von Mai 2014 bis April 2015 gesammelt hatten in ihr Modell, dabei wurden diese, wie schon erwähnt, in unterschiedliche Zeitspannen aufgeteilt. Die Genauigkeit zum realen Kursverlauf lag bei 78,9%. Weiterhin stellte sich dabei heraus, dass die Zeitspanne von 4 Monaten am genauesten ist. Längere Zeitspannen bedeuten mehr Artikel und dadurch ein größeres Rauschen in den Daten, welches eine genaue Vorhersage erschwert. Da die Forscher zufrieden mit ihrem Ergebnis waren, versuchten sie mit ihrem Modell auch die Kurse einzelner Aktien vorherzusagen. In diesem Versuch variierte die Genauigkeit zwischen 60-90%. Dies könnte mit der Anzahl der Artikel für jede einzelne Aktie zusammenhängen. Jedoch war selbst ein Ergebnis von 60% immer noch besser, als den Aktienkurs einfach zu raten, was eine Chance von 50% hervorbringen würde.²⁸

5. Zusammenfassung Automated Bitcoin Trading via Machine Learning Algorithms

Die Forscher unterteilten ihr Vorgehen in zwei Phasen. In der ersten Phase versuchte man tägliche Trends im Bitcoinmarkt zu verstehen und besser zu identifizieren. Weiterhin war das Ziel optimale Eigenschaften herauszufinden, welche gegeben sein müssen, um bestimmte Trends einzuleiten. In der zweiten Phase konzentrierten sich die Forscher lediglich auf den Bitcoinpreis. Man wollte hier Daten über den Bitcoinpreis in 10 Minuten und 10 Sekunden Abständen auswerten. Es sollte versucht werden Preisvoraussagen auf unterschiedlichen Detail- und Rauschleveln vorzunehmen. Bei

²⁶ What is a SVM?, OpenCV, https://docs.opencv.org/2.4/_images/optimal-hyperplane.png

²⁷ Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.3

²⁸ Vgl.: Duong et al. Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, 2016, S.3-5

einer solchen Vorhersage handelt es sich um ein binomielles Problem, da bestimmt werden soll, ob der Bitcoinpreis steigen oder fallen wird.²⁹

5.1 Vorgehensweise

Der erste Schritt war die Data Collection. Man benötigte zwei Sets von Daten, für jede Phase eines. Im ersten Set waren der tägliche Bitcoinpreis und 26 zusätzliche Eigenschaften über das Bitcoinnetzwerk und den Bitcoinmarkt enthalten. Durch die Erhebung der Daten aller 24 Stunden wurde das Rauschen minimiert, welches durch höhere Detailtiefen und minutiöse Volatilität zustande kommt. Die Daten wurden für die letzten 5 Jahre gesammelt. Im zweiten Set waren die Bitcoinpreise in 10 Sekunden und 10 Minuten Intervallen enthalten. Die Daten in 10 Minuten Intervallen erhielten die Forscher über die API eines großen Bitcoinhändlers. Über die gleiche API wurden mit Hilfe eines Webscrapers die Daten in 10 Sekunden Intervallen gesammelt. Es wurden so von den Forschern 120.000 einzigartige Preise gesammelt.³⁰

Der nächste Schritt bestand aus der Feature Selection. Aus den 26 vorliegenden Eigenschaften wurden 16 ausgewählt für das Modell zur Vorhersage. Die Eigenschaften wurden manuell von den Forschern aus Basis ihrer aktuellen Forschung ausgewählt, im Bezug darauf wie ihre Signifikanz für das zu lösende Problem sei. Beispiele für solche Eigenschaften wären der Verdienst für den Miner, die Schwierigkeit einen neuen Block zu finden oder die Rechenrate des Bitcoinnetzwerkes. Die vorliegenden Daten und Eigenschaften wurden genutzt um einen binären Klassifikationsalgorithmus zu erstellen, welcher die Zeichen bestimmen sollte, die für die tägliche Preisänderung verantwortlich sind. Das Trainingsset für dieses Modell erhielt 70% der Daten und das Testset 30%.³¹

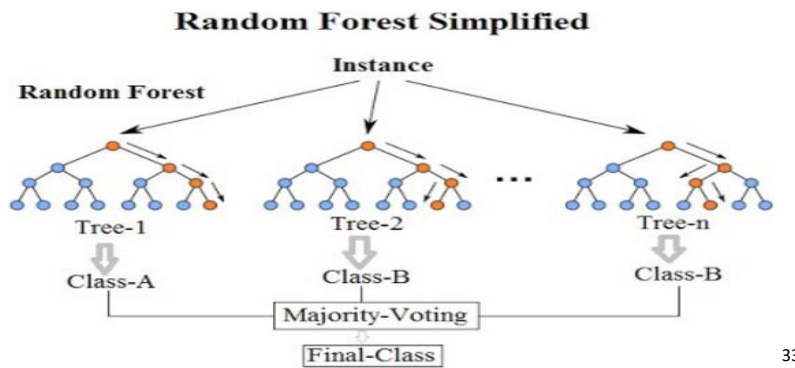
Der letzte Schritt war die Time Series Equation. Um genauere Einblicke zu bekommen wurden in diesem Schritt auch die 10 Sekunden Intervall Daten analysiert. Die vorliegenden Eigenschaften konnten hier jedoch nicht verwendet werden, da sie entweder schlecht oder gar nicht dokumentiert waren in diesem Zeitintervall. Die Forscher kreierte drei Zeitreihen für 30, 60 und 120 Minuten vorausgehend dem aktuellen Datenpunkt. Auf jedes dieser einzelnen Datensets wurde ein Random Forest angewendet. Dadurch erhielt man unterschiedliche lineare Modelle von den einzelnen Zeitreihen. Ein Random Forest funktioniert grundlegend so, dass zufällige Subdatensätze und Features ausgewählt werden. Aus diesen werden zufällige Entscheidungsbäume gebaut, welche ein Ergebnis liefern. Aus der Summe aller dieser Ergebnisse, bildet man einen Mittelwert, welcher sich darauf bezieht, welche Antwort man durch den Random Forest erhalten wollte. In unserem Beispiel ließ man von jedem die Preisänderung vorhersagen. Wenn diese Ergebnisse kombiniert wurden, konnte man längerfristige Preisentwicklungen vorhersagen. Nachdem das Prozedere jeweils für die 10 Sekunden und 10 Minuten Intervall Daten durchlaufen wurde, stellte sich heraus, dass die 10 Sekunden Daten besser für schnellen Handel geeignet sind. Die 10 Minuten Intervalle waren eher dafür geeignet größere Preisänderungen zu erkennen.³²

²⁹ Vgl.: Isaac Madan, Shaurya Saluja, Aojia Zhao, Automated Bitcoin Trading via Machine Learning Algorithms, 2015, S.1

³⁰ Vgl.: Madan et al., Automated Bitcoin Trading via Machine Learning Algorithms, 2015, S.2

³¹ Vgl.: Madan et al., Automated Bitcoin Trading via Machine Learning Algorithms, 2015, S.2

³² Vgl.: Madan et al., Automated Bitcoin Trading via Machine Learning Algorithms, 2015, S.3



33

5.2 Ergebnisse

Es wurden verschiedene Kategorisierungsmodelle in der ersten Phase verwendet und getestet, wie gut sie die Bitcoinpreisveränderung vorhersagen können. Die Support Vector Machine, welche schon eher erklärt wurde, hatte Probleme Vorhersagen zu treffen, da sie versucht Muster in den Daten zu finden. Die Zeitabstände waren mit 24 Stunden jedoch sehr groß und somit war es schwer diese Trends klar zu erkennen. Der Random Forest lieferte eine gute Genauigkeit, aber eine schlechte Präzision. Präzision beschreibt, wie nah die gelieferten Ergebnisse aneinander liegen und Genauigkeit, wie nah jedes Ergebnis am echten Wert liegt. Durch die zufälligen Entscheidungsbäume gab es eine hohe Fluktuation, was zu wenig Präzision führte. Wurde jedoch der Mittelwert aus den Ergebnisse gebildet lag er nah an den echten Entwicklungen und hatte somit eine hohe Genauigkeit. Der Forscher konnte die Zeichen für den täglichen Preiswandel mit 98,7% Wahrscheinlichkeit vorhersagen. In der zweiten Phase zeigte sich, dass die 10 Sekunden Intervall Daten ungenauer als die 10 Minuten Intervalle waren. Die 10 Minuten Daten waren besser geeignet, weil die Forscher kein Hochgeschwindigkeitstrading durchführen wollten, sondern vorausschauendes, längerfristiges Trading. Dies wurde durch die längeren Abstände in den Daten möglich, da hierdurch Trends besser erkennbar wurden. Zudem ist es wahrscheinlicher in diesen Abständen gewinnbringend zu handeln, da der Vorhersagerahmen für 10 Sekunden zu gering ist, bedingt durch die Verzögerung von Transaktionen mit Bitcoin. Die normale Entwicklung des Bitcoinpreises konnte relativ gut vorausgesagt werden, jedoch war es nicht möglich starke Spitzen und Abfälle vorherzusehen.³⁴

In der Zukunft wollten die Forscher ihren Ansatz weiter ausbauen. Der Plan beinhaltet, dass aus Untersets der Daten Muster herausgefunden werden. Pro Unterset sollen nur 100 Daten enthalten sein, aus denen ein Muster bestimmt wird. Diese Muster werden anschließend in Gruppen eingeordnet. Möchte man nun einen neuen Datenpunkt berechnen, werden die 100 neusten Daten vor diesem gesammelt und aus ihnen ein Unterset gebildet, welches man auf sein Muster untersucht. Für dieses Muster bestimmt man die Gruppe und wendet das Modell nur auf die ausgewählte Gruppe an, anstatt auf alle Daten.³⁵

6. Eigene Meinung und Zukunftsaussicht

Dieses Thema ist sowohl von einem wissenschaftlichen, als auch wirtschaftlichen Standpunkt sehr interessant. Wissenschaftlich, weil viele verschiedene Modelle und Ansätze kombiniert werden. Es ist noch nicht eine Best Practice gefunden in diesem Forschungsbereich. Jedes Forscherteam hat eine

³³ Random Forest Simplified, Venkata Jagannath, https://d2wh20haedxe3f.cloudfront.net/sites/default/files/random_forest_diagram_complete.png

³⁴ Vgl.: Madan et al., Automated Bitcoin Trading via Machine Learning Algorithms, 2015, S.3-5

³⁵ Vgl.: Madan et al., Automated Bitcoin Trading via Machine Learning Algorithms, 2015, S.5

neue Idee, wie es die vorhandenen Modelle kombinieren und verwenden möchte. In jedem Paper werden andere Ziele und Ideen verfolgt, zudem werden immer andere Datensätze verwendet. Verschiedene Datensätze entstehen daraus, dass verschiedene Produkte vorausgesagt werden sollen. Im Moment reichen diese eher von einzelnen Aktien bis zu Aktienindizes. Später halte ich es aber auch für sehr gut möglich, dass komplexere Finanzprodukte vorhergesagt werden, sobald die grundlegenden Modelle eine zufriedenstellende Genauigkeit erreichen. Es ist deshalb meiner Meinung nach schwierig, die einzelnen Ansätze zu vergleichen. Die Entwicklungen in diesem Forschungsgebiet bleiben spannend, gerade mit neuen Forschungsergebnisse im Bereich der neuronalen Netze und künstlichen Intelligenz. Möglicherweise können auch Erkenntnisse aus diesen Vorhersagemodellen unterstützend in neuronalen Netzen und künstlicher Intelligenz wirken und somit die Forschungsfelder voneinander profitieren lassen. Zudem ist es ein internationales Thema, da überall auf der Welt Finanzmärkte bestehen und jeder daran Interesse hat diese gewinnbringend vorherzusagen, weshalb die Entwicklung und Forschung stärker nach vorn getrieben werden sollte, als bei nationalen Forschungsthemen. Wirtschaftlich ist es spannend, da es eine neue Komponente in den Finanzmarkt bringt. Bisher wurden Rechner im Aktienhandel zum Großteil für Hochgeschwindigkeitstrading eingesetzt. Dies bildet jedoch nur eine Zeitspanne von Millisekunden bis Sekunden ab und beruht nur auf einem Zahlenaspekt. Durch die neuen Analysemodelle wird jedoch die soziale Komponente wieder mehr integriert. Da hier, durch die Analyse der öffentlichen Meinungen und Berichterstattungen eher die mittelfristigen Entwicklungen im Fokus stehen. Verschiedene Ansätze verfolgen auch hier unterschiedliche Gewichtungen dieser Quellen. Einige Forscherteams konzentrieren sich komplett darauf anhand von Social Media Posts Aktienkurse zu antizipieren, andere versuchen es anhand von Newsartikeln, wie ich schon eher dargestellt habe. Fragen werfen sich in dem Aspekt auf, dass wenn Forscherteams an solchen Modellen arbeiten, die mehr oder minder gut funktionieren, große Unternehmen ebenfalls schon solche Modelle entwickelt haben könnten. Solche Unternehmen würden nicht preisgeben, wenn sie ein Modell entwickelt hätten, welches sehr gut für die Aktienmarktvorhersage funktionieren würde. Logischerweise würden sie zuerst einen solchen Profit einstecken und sich den Vorteil vor ihren Konkurrenten sichern. Eine weitere Frage wäre, ob solche Modelle die auf Newsartikeln oder der öffentlichen Medienberichterstattung nicht durch gezielte Nachrichten manipuliert werden können. Die Medien werden heutzutage von wenigen, einflussreichen Personen kontrolliert, da diesen meist große Konglomerate aus verschiedenen Medienanstalten gehören. Diese könnten bestimmte Meldungen ausgeben und andere zurückhalten um die Modelle zu manipulieren und somit den Markt für ihren eigenen Vorteil zu lenken. Zudem bleibt die Frage, ob solche Modelle und vorausgesagten Trends nicht teilweise eine self fulfilling prophecy sind. Wenn Ergebnisse solcher Analysen der Öffentlichkeit zugänglich wären, würden die Voraussagen sich wahrscheinlich von selbst erfüllen, da viele Personen dem Vorausgesagten folgen würden. In Folge dessen würde Aktienkurse von ganz allein steigen, da viele Leute die Aktie kaufen würden. Auch hier besteht die Möglichkeit für Manipulation, da falsche Vorhersagen publik gemacht werden könnten und somit die Entwicklungen am Aktienmarkt beeinflusst werden würden. Wenn man dieses Thema jedoch von einem objektiven Standpunkt betrachtet, ohne die Vermutung der möglichen Manipulation, wird es meiner Meinung nach schwer bleiben eine hundertprozentige Vorhersage der Aktienkurse zu treffen, da Emotionen eine große Rolle im Markt spielen und man diese nicht allgemein durch News oder Social Media Beiträge vorhersagen kann. Menschliche Emotionen bleiben hoffentlich auch in der Zukunft etwas, was nicht von Computern komplett vorhergesagt werden kann. Alles in allem ist es ein sehr spannendes Thema, gerade von einem wissenschaftlichen Standpunkt, da es noch sehr jung ist und viele neue Entdeckungen und Ideen entwickelt werden. Es werden jedoch, wie schon vorher erwähnt, auch Fragen aufgeworfen über die Unfehlbarkeit des Systems, welche in der Zukunft, mit der weiteren Entwicklung der Modelle, beantwortet werden müssen.

7. Literaturverzeichnis

Internetquellen:

- Wikipedia, „Aktie“, <https://de.wikipedia.org/wiki/Aktie#Aktiengattungen>, 12.03.2018
- Wikipedia, „Börsenkurs“, <https://de.wikipedia.org/wiki/B%C3%B6rsenkurs>, 12.03.2018
- Alexander Wolf, „Was ist der DAX?“, <http://www.aktien-kaufen-fuer-anfaenger.de/was-ist-der-dax/>, 12.03.2018
- Alexander Wolf, „Definition: Was sind Fonds? Was ist ein Investmentfonds?“, <http://www.aktien-kaufen-fuer-anfaenger.de/definition-was-sind-fonds-was-ist-ein-investmentfonds/>, 12.03.2018
- Finanzfluss, „Was sind Anleihen? Unternehmens- und Staatsanleihen einfach erklärt“, <https://www.finanzfluss.de/was-sind-anleihen/> 12.03.2018
- Wikipedia, „Terminkontrakt“, <https://de.wikipedia.org/wiki/Terminkontrakt>, 30.03.2018
- Boerse.de, „Was ist ein Futures-Kontrakt?“, <https://www.boerse.de/grundlagen/eurex/Was-ist-ein-Futures-Kontrakt-6>, 30.03.2018
- Lynx, „Grundlagen von Optionen“, <https://www.lynxbroker.de/wissen/erfolgreich-handeln-mit-optionen/grundlagen-von-optionen/>, 27.03.2018
- Wikipedia, „Kaufoption“, <https://de.wikipedia.org/wiki/Kaufoption>, 27.03.2018
- Wikipedia, „Verkaufsoption“, <https://de.wikipedia.org/wiki/Verkaufsoption>, 27.03.2018
- Boerse.de, „Was sind Futures und Optionen?“, <https://www.boerse.de/grundlagen/boerseneinfuehrung/Was-sind-Futures-und-Optionen-10>, 27.03.2018
- Wikipedia, „Zeitreihenanalyse“, <https://de.wikipedia.org/wiki/Zeitreihenanalyse>, 26.03.2018
- Reiter1.com, „ARIMA Zeitreihenanalyse“, <http://www.reiter1.com/Glossar/ARIMA.htm>, 26.03.2018

Textquellen:

- Duc Duong, Toan Nguyen, Minh Dang, 2016, Stock Market Prediction using Financial News Articles on Ho Chi Minh Stock Exchange, <https://dl.acm.org/citation.cfm?id=2857619>, 28.12.2017
- Isaac Madan, Shaurya Saluja, Aojia Zhao, 2015, Automated Bitcoin Trading via Machine Learning Algorithms, <http://cs229.stanford.edu/proj2014/Isaac%20Madan,%20Shaurya%20Saluja,%20Aojia%20Zhao,Automated%20Bitcoin%20Trading%20via%20Machine%20Learning%20Algorithms.pdf>, 30.12.2017

Bildquellen:

- Dr. Manuel Kay und Sara Zinnecker, So funktionieren Call- und Put-Optionen, <https://www.finanztip.de/anleihen/optionen/>, 27.03.2018
- Duc Duong, Toan Nguyen, Minh Dang, An overview of our stock trend prediction process, <https://dl.acm.org/citation.cfm?id=3457619>, 26.03.2018
- Hady Elsahar, How does Delta TF-IDF work?, <https://www.quora.com/How-does-Delta-TF-IDF-work>, 25.03.2018
- OpenCV, What is a SVM?, https://docs.opencv.org/2.4/_images/optimal-hyperplane.png, 25.03.2018
- Venkata Jagannath, Random Forest Simplified, https://d2wh20haedxe3f.cloudfront.net/sites/default/files/random_forest_diagram_complete.png, 25.03.2018