



# KURSVORHERSAGE VON KRYPTOWÄHRUNGEN MIT AZURE MACHINE LEARNING

FORECASTING PRICES OF CRYPTOCURRENCIES USING AZURE MACHINE  
LEARNING

ABSCHLUSSARBEIT

ZUR ERLANGUNG DES AKADEMISCHEN GRADES  
MASTER OF SCIENCE

VORGELEGT VON

SEBASTIAN LISCHEWSKI

GEBOREN AM 08.08.1991 IN ROSENHEIM  
MATRIKELNUMMER: 04326912

MÜNCHEN, DEN 26. JULI 2017

Prüfer: Prof. Dr. PATRICK MÖBERT, Hochschule München

# Erklärung

Hiermit erkläre ich, dass ich die Bachelorarbeit selbständig verfasst, noch nicht anderweitig für Prüfungszwecke vorgelegt, keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

Ort, Datum

Unterschrift

test

# Zusammenfassung

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

# Inhaltsverzeichnis

Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
Listings	VIII
1 Einführung zum Thema	1
1.1 Thema der Arbeit . . . . .	1
1.2 Bitcoin als Vorreiter der Kryptowährungen . . . . .	1
1.2.1 Bitcoin . . . . .	2
1.2.2 Ethereum . . . . .	2
1.2.3 Altcoins (Litecoin, Dogecoin) . . . . .	2
1.3 Machine Learning, Data Mining, Data Analysis und Data Science . . . . .	2
1.4 Cloud-Dienste und SaaS . . . . .	4
2 Vorgehen und Ziele	5
3 Grundlagen	6
3.1 Data Mining (Prozess) . . . . .	6
3.1.1 Ziele . . . . .	6
3.1.2 Datenauswahl . . . . .	6
3.1.3 Datenbereinigung . . . . .	6
3.1.4 Datenreduktion (Transformaton) . . . . .	6
3.1.5 Modellauswahl . . . . .	6
3.1.6 Data-Mining (Datenanalyse) . . . . .	6
3.1.7 Interpretation . . . . .	6

## *Inhaltsverzeichnis*

3.2	Machine Learning . . . . .	6
3.2.1	Supervised... . . . .	6
3.2.2	Unsupervised... . . . .	7
3.3	Kryptowährung(en) . . . . .	7
3.4	SaaS . . . . .	8
3.5	Microsoft Azure ML Studio . . . . .	8
3.5.1	Allgemeine Beschreibung . . . . .	8
3.5.2	Aufbau . . . . .	8
3.5.3	Elemente . . . . .	8
4	Einflüsse	10
5	Daten	11
5.1	Kurse . . . . .	11
5.2	Überschriften (Keggle) . . . . .	11
5.3	andere Kurse/börsen . . . . .	11
6	Durchführung	12
7	Fazit	13
8	Related Work	14
9	Ausblick	15
	Literaturverzeichnis	16

# Abbildungsverzeichnis

# Tabellenverzeichnis

# Listings



# 1 Hinführung zum Thema

## 1.1 Thema der Arbeit

In der vorliegenden Arbeit werden Einflussfaktoren auf den Kurs von ausgewählten Kryptowährungen gesucht und der Grad des Einflusses evaluiert. Dies geschieht mit dem Ziel herauszufinden, ob sich die Kursschwankungen der digitalen Währungen voraussagen lassen und wenn ja, in welchem Maße. Im nachfolgenden Kapitel wird auf die Motivation hinter der Analyse eingegangen. Das genaue Vorgehen und die Ziele werden in Abschnitt 2 erläutert.

## 1.2 Bitcoin als Vorreiter der Kryptowährungen

Geld online von einem Teilnehmer direkt zu einem Anderen senden, ohne dabei (Transaktions-)Gebühren für einen zwischengelagerten Finanz-Dienstleister zahlen zu müssen, ist der Gedanke hinter dem „Peer-To-Peer Electronic Cash System“ (Nakamoto, 2008) Bitcoin. Obwohl es Teilnehmern ohne Aufwand möglich ist, dem Netzwerk beizutreten oder es wieder zu verlassen, ist es solange unangreifbar, solange ein Angreifer nicht dauerhaft über mehr Rechenkapazität verfügt, als das komplette restliche Netzwerk. (Nakamoto, 2008) Ob das Bitcoinnetzwerk wirklich absolute Anonymität gewährt, wird stark kritisiert. (Reid and Harrigan, 2013; Androulaki et al., 2013). In der Tat werden beim Nutzen des Netzwerk jedoch keine persönlichen Informationen an ein Kreditinstitut (wie PayPal, Paydirekt, ApplePay oder Masterpass) weitergegeben. Diese Argumente (Kostenreduktion, Sicherheit und Anonymität) sorgen für Interesse an der digitalen Währung (auch hier gibt es Kritiker, die den Bitcoin als Investition und nicht als Währung bezeichnen) (Baur et al., 2015). Nicht zu vernachlässigen ist an dieser Stelle auch das Interesse der Industrie an „Smart Contracts“ (Dannen, 2017, S. 10), die beispielsweise im Bereich des Internet of Things Anwendung finden. (Christidis and

## 1 Hinführung zum Thema

Devetsikiotis, 2016)

Neben Bitcoin hat sich deshalb zusätzlich eine Vielzahl an anderen sogenannten Kryptowährungen entwickelt. Die Währungen mit dem größten Marktvolumen sind Bitcoin(1.2.1) und Ethereum(1.2.2)(Wood, 2014).(Brandt, 2017; CoinMarketCap, 2017) Daneben gibt es noch sogenannte Altcoins (aus dem Englischen: alternative coin(Bajpai, 2014))(1.2.3). Mittlerweile umfassen diese 664 Bitcoin-Alternativen.(CoinDesk, 2017). Obgleich die tatsächliche Nutzung der Kryptowährungen sehr gering ist (1% der Befragten in Deutschland(TSYS, 2016)), steigt das Interesse an Kryptowährungen(WikiTrends, 2017; GoogleTrends, 2017).

TODO: irgendwas zu Technik später oder so?

TODO: brauche ich diese Punkte hier wirklich?

### 1.2.1 Bitcoin

bekannteste cryptowährung; bekannt als vorreiter; medien etc.; auch hier statistika

### 1.2.2 Ethereum

aufstreben, smart contracts wie angesprochen; nicht nur copycat

### 1.2.3 Altcoins (Litecoin, Dogecoin)

alternativen, warum hier nicht betrachtet; nur kopien...

## 1.3 Machine Learning, Data Mining, Data Analysis und Data Science

Die Themen Machine Learning, Data Mining, Data Analysis und Data Science stellen zwar verwandte Begriffe dar, sind jedoch alles eigene Felder.

Der Begriff Machine Learning gehört in der Informatik und Mathematik zur Familie der Künstlichen Intelligenz.(Kim, 2017, S. 2; Swamynathan, 2017, S. 54). Es kann als „Sammlung

## 1 Hinführung zum Thema

von Algorithmen und Techniken“ verstanden werden, die „genutzt werden, um Computersysteme zu erstellen, die aus Daten lernen, um Vorhersagen zu erstellen“. (Swamynathan, 2017, S. 53; eigene Übersetzung) Bekannte Anwendungen aus dem Alltag sind Empfehlungssysteme oder Spamerkennungen. (Swamynathan, 2017, S. 53)

Data Mining beschreibt den Prozess aus einer gewaltigen Menge an Daten die „richtigen Daten“, zur „richtigen Zeit“ für die „richtigen Entscheidungen“ (Swamynathan, 2017, S. 61; eigene Übersetzung) zu gewinnen. Für diesen Prozess haben sich im Laufe Zeit drei Frameworks herauskristallisiert (Swamynathan, 2017, p.69)

- Knowledge Discovery Databases (KDD) process model
- Cross Industrial Standard Process for Data Mining (CRISP – DM)
- Sample, Explore, Modify, Model and Assess (SEMMA)

Neben Schnittpunkten zu Künstlicher Intelligenz, Machine Learning und der Statistik, befasst sich Data Mining ebenfalls mit Datenbanksystemen. (Ramasubramanian and Singh, 2017, S. 4)

Eng verwandt mit dem Data Mining ist die Datenanalyse (engl. Data Analysis; in der Industrie auch Business Analytics (Swamynathan, 2017, S. 58)). Sie wird benutzt um (Hertle, 2016, S. 2)

1. Messdaten zu verstehen,
2. Gesetzmäßigkeiten zu extrahieren und
3. die Zukunft vorherzusagen.

Dazu bedient sich der deskriptiven Statistik, der explorativen Datenanalyse (engl. Explorative Data Analysis; EDA) und der induktiven Statistik. (Hertle, 2016, S. 17)

Um

- den Anstieg der Datenmengen in der Datenanalyse,
- die Veränderung im Aussehen der Daten (unstrukturiert oder semi-strukturiert statt strukturiert) und
- die Wandlung Semantik der zugrundeliegenden Daten (Daten liegen in Markup-Sprachen vor und enthalten zusätzliche Informationen)

## *1 Hinführung zum Thema*

darzustellen, hat sich der Begriff Data Science entwickelt.(Dhar, 2013) Er versucht die geänderten Anforderungen der heutigen Datenanalyse abzubilden.

Wie anfänglich erwähnt, sind alle genannten Begriffe miteinander verwandt. Das Gewinnen von Erkenntnissen aus Daten, um beispielsweise die Zukunft vorherzusagen, nennt sich Data Analysis. Werden die Daten aus verschiedensten Datenbanken oder Datawarehouses gewonnen, spricht man von Data Mining. Handelt es sich dabei um Informationen unterschiedlicher Struktur und große Datensätze, so befindet man sich im Bereich der Data Science. Der inhärente Erkenntnisgewinn dieser Verfahren kann von menschlicher Seite kommen oder durch Machine Learning geschehen.

Projekte wie Googles DeepMind(Limited, 2017), IBMs Watson((IBM), 2017) oder Sprachassistenten wie Siri, Alexa und Bixby zeigen, dass großes Interesse an Machine Learning und Data Science herrscht. Deshalb haben sich auch ganze Berufsfelder wie „machine learning engineer“, „data engineer“ oder „data scientist“(Ramasubramanian and Singh, 2017, S. 1) gebildet.

### **1.4 Cloud-Dienste und SaaS**

Wieso Cloud Dienste Nutzen; Warum nicht nur lokal? (brauch ich diesen Teil?); einfach vorstellen und dann azure ml studio

## **2 Vorgehen und Ziele**

# **3 Grundlagen**

## **3.1 Data Mining (Prozess)**

Data mining Prozess beschreiben und die Stufen von Zeile etc. Bis zur Interpretation

### **3.1.1 Ziele**

### **3.1.2 Datenauswahl**

### **3.1.3 Datenbereinigung**

### **3.1.4 Datenreduktion (Transformaton)**

### **3.1.5 Modellauswahl**

### **3.1.6 Data-Mining (Datenanalyse)**

### **3.1.7 Interpretation**

## **3.2 Machine Learning**

### **3.2.1 Supervised...**

Man weiß, nach was man sucht...

### *3 Grundlagen*

**Decision Tree**

**Neares Neighbour**

**Random Forest**

**SVM**

#### **3.2.2 Unsupervised...**

**K means**

**Hierarchical clustering**

**Neuronal networks**

...

Man sucht nur cluster/gruppen/etc

### **3.3 Kryptowährung(en)**

Bitcoin,ethereum, litecoin, dogecoin; auswahl hier nur 1/2

## 3.4 SaaS

## 3.5 Microsoft Azure ML Studio

### 3.5.1 Allgemeine Beschreibung

### 3.5.2 Aufbau

Projects

Experiments

Web Services

Notebooks

Datasets

Trained Models

Settings

### 3.5.3 Elemente

relevante auswählen



### *3 Grundlagen*

**Saved Datasets**

**Data Transformation Conversations**

**Data Transformation**

**Data Input and Output**

**Feature Selection**

**Machine Learning**

**OpenCV Library Models**

**Python Language Model**

**R Language Model**

**Statistical Functions**

**Text Analysis**

**Time Series Anomaly Detection**

**Web Service**

## 4 Einflüsse

aus paper und mehr suchen

welchen einfluss hier; im nächsten teil dann: wie kann man das repräsentieren, welche daten gibt es da und kann man das abbilden?

beispiele: regierungen und regionen (usa, china, EU) → Gesetze

bitcoin-eigene dinge (volumen, umschlag, miner? etc.)

öffentlichkeit (twitter, zeitungen, blogs, domains im web)

natürliche Ressourcen (Öl, Gold, Silber, Diamanten w/e)

Financial Stress Index (FSI)

HIER PAPER NOCHMAL: \* Economic Drivers \* Transaction Drivers \* Technical Drivers \* Interest \* Safe Haven \* Influence of China

# 5 Daten

Welche Daten Brauche ich, wo kriege ich sie her, was steht drin, beschreibung, features etc.

## 5.1 Kurse

börse 1, 2, Währungen

## 5.2 Überschriften (Keggle)

## 5.3 andere Kurse/börsen

dax, china!, dow jones ...

# 6 Durchführung

von Ziele bis Interpretation

## 7 Fazit

## 8 Related Work

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

## 9 Ausblick

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

# Literaturverzeichnis

- Androulaki, E., Karame, G. O., Roeschlin, M., Scherer, T., and Capkun, S. (2013). Evaluating User Privacy in Bitcoin. In *Financial Cryptography and Data Security*, Lecture Notes in Computer Science, pages 34–51. Springer, Berlin, Heidelberg.
- Bajpai, P. (2014). Altcoin.
- Baur, D. G., Lee, A. D., and Hong, K. (2015). Bitcoin: Currency or Investment? SSRN Scholarly Paper ID 2561183, Social Science Research Network, Rochester, NY.
- Brandt, M. (2017). Infografik: Die Top 10 der Kryptowährungen.
- Christidis, K. and Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4:2292–2303.
- CoinDesk (2017). Anzahl der Altcoins weltweit in ausgewählten Monaten von Dezember 2015 bis September 2016.
- CoinMarketCap (2017). Ranking der größten virtuellen Währungen nach Marktkapitalisierung im Juli 2017 (in Millionen US-Dollar).
- Dannen, C. (2017). *Introducing Ethereum and Solidity*. Apress, Berkeley, CA. DOI: 10.1007/978-1-4842-2535-6.
- Dhar, V. (2013). Data Science and Prediction. *communications of the acm*, vol. 56 no. 12:10. doi:10.1145/2500499.
- GoogleTrends (2017). GoogleTrends Vergleich: Bitcoin, Ethereum, Cryptocurrency.
- Hertle, J. (2016). Datenanalyse - Vorlesung Master, Hochschule München, SS 2016.
- (IBM), I. B. M. C. (2017). IBM Watson.



## Literaturverzeichnis

- Kim, P. (2017). *MATLAB Deep Learning*. Apress, Berkeley, CA. DOI: 10.1007/978-1-4842-2845-6.
- Limited, D. T. (2017). DeepMind.
- Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System Bitcoin: A Peer-to-Peer Electronic Cash System.
- Ramasubramanian, K. and Singh, A. (2017). *Machine Learning Using R*. Apress, Berkeley, CA. DOI: 10.1007/978-1-4842-2334-5.
- Reid, F. and Harrigan, M. (2013). An Analysis of Anonymity in the Bitcoin System. In Altshuler, Y., Elovici, Y., Cremers, A. B., Aharony, N., and Pentland, A., editors, *Security and Privacy in Social Networks*, pages 197–223. Springer New York. DOI: 10.1007/978-1-4614-4139-7\_10.
- Swamynathan, M. (2017). *Mastering Machine Learning with Python in Six Steps*. Apress, Berkeley, CA. DOI: 10.1007/978-1-4842-2866-1.
- TSYS (2016). Kennen oder nutzen sie virtuelle Währungen wie Bitcoin?
- WikiTrends (2017). Compare popularity of Bitcoin, Cryptocurrency, Ethereum on Wikipedia | Wiki Trends.
- Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper*, 151.