



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 7. 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 Motivation	1
1.1 Kryptowährungen	1
1.1.1 Bitcoin	1
1.1.2 Ethereum	1
1.1.3 Altcoins (Litecoin, Dogecoin)	1
1.2 Machine Learning und Data Mining	1
1.3 Cloud-Dienste und SaaS	2
2 Grundlagen	3
2.1 Data Mining (Prozess)	3
2.1.1 Ziele	3
2.1.2 Datenauswahl	3
2.1.3 Datenbereinigung	3
2.1.4 Datenreduktion (Transformaton)	3
2.1.5 Modellauswahl	3
2.1.6 Data-Mining (Datenanalyse)	3
2.1.7 Interpretation	3
2.2 Machine Learning	3
2.2.1 Supervised...	3
2.2.2 Unsupervised...	4
2.3 Kryptowährung(en)	4

Inhaltsverzeichnis

2.4	Saas	5
2.5	Microsoft Azure ML Studio	5
2.5.1	Allgemeine Beschreibung	5
2.5.2	Aufbau	5
2.5.3	Elemente	5
3	Daten	7
3.1	Kurse	7
3.2	Überschriften (Keggle)	7
3.3	andere Kurse/börsen	7
4	Durchführung	8
5	Related Work	9
6	Ausblick	10
7	Zitate	11
8	Bilder	12
9	Codelistings	13
10	Tabellen	14
	Literaturverzeichnis	16

Abbildungsverzeichnis

8.1	Domain cs.hm.edu	12
8.2	Anwendungsfalldiagramm	12
10.1	Anzahl der Threads im Verhältnis zu Laufzeit	14

Tabellenverzeichnis

10.1 Laufzeiten der einzelnen Module für die Domain Amazon.de	15
---	----

Listings

9.1	Starten einer Suche in DomainSearch.py	13
9.2	DNS Auflösung von cs.hm.edu	13

1 Motivation

1.1 Kryptowährungen

Bedeutung heutzutage; bisschen Geschichte; Ursprünge; -> Techniken dann später

1.1.1 Bitcoin

bekanntester

1.1.2 Ethereum

aufstreben, smart contracts

1.1.3 Altcoins (Litecoin, Dogecoin)

alternativen, warum hier nicht

1.2 Machine Learning und Data Mining

Was ist das; wozu nutzt man es; wo ist der Unterschied -> genaueres dann später

1.3 Cloud-Dienste und SaaS

Wiso Cloud Dienste Nutzen; Warum nicht nur lokal? (brauch ich diesen Teil?) oder einfach Azure auswählen

2 Grundlagen

2.1 Data Mining (Prozess)

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

2.1.1 Ziele

2.1.2 Datenauswahl

2.1.3 Datenbereinigung

2.1.4 Datenreduktion (Transformaton)

2.1.5 Modellauswahl

2.1.6 Data-Mining (Datenanalyse)

2.1.7 Interpretation

2.2 Machine Learning

2.2.1 Supervised...

Man weiß, nach was man sucht...

2 Grundlagen

Decision Tree

Neares Neighbour

Random Forest

SVM

2.2.2 Unsupervised...

K means

Hierarchical clustering

Neuronal networks

...

Man sucht nur cluster/gruppen/etc

2.3 Kryptowährung(en)

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

2.4 SaaS

2.5 Microsoft Azure ML Studio

2.5.1 Allgemeine Beschreibung

2.5.2 Aufbau

Projects

Experiments

Web Services

Notebooks

Datasets

Trained Models

Settings

2.5.3 Elemente

relevante auswählen

2 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

3 Daten

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

3.1 Kurse

börse 1, 2, Währungen

3.2 Überschriften (Keggle)

3.3 andere Kurse/börsen

dax, china!, dow jones ...

4 Durchführung

von Ziele bis Interpretation

5 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.

6 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.

7 Zitate

Behauptung und Quelle. (Sixt, 2017)

8 Bilder

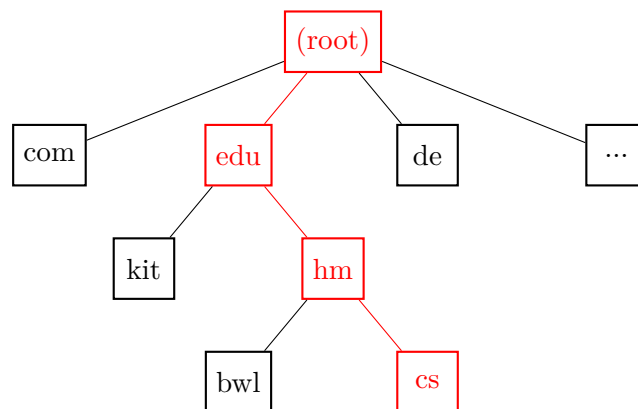


Abbildung 8.1: Domain cs.hm.edu

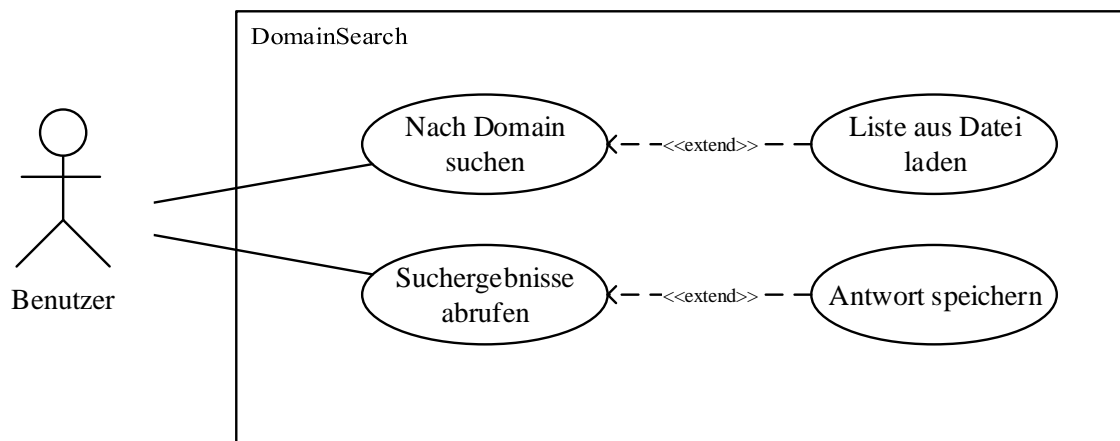


Abbildung 8.2: Anwendungsfalldiagramm

9 Codelistings

```
def perform_search(name):
    with threading.Lock():
        # Perform one search for the given domain.
        name = name.strip()
        out.print_search_started(name)

        db = database.db()
        searchID = db.insert_search(name)
        db.close_connection()

        sched = scheduler.Scheduler()
        sched.find_modules()
        sched.start_modules(searchID, name)
```

Listing 9.1: Starten einer Suche in DomainSearch.py

```
cs.hm.edu. 1800 IN NS ns.e-technik.fh-muenchen.de.
cs.hm.edu. 1800 IN NS ns.hm.edu.
cs.hm.edu. 72791 IN A 129.187.244.60
cs.hm.edu. 14260 IN MX 10 mailrelay2.lrz.de.
cs.hm.edu. 14260 IN MX 10 mailrelay1.lrz.de.
```

Listing 9.2: DNS Auflösung von cs.hm.edu

10 Tabellen

Anzahl Threads	Laufzeit in Sekunden
Sequenziell	531,27
2	260,67
5	106,77
10	69,02
15	70,53
20	66,48
50	62,94

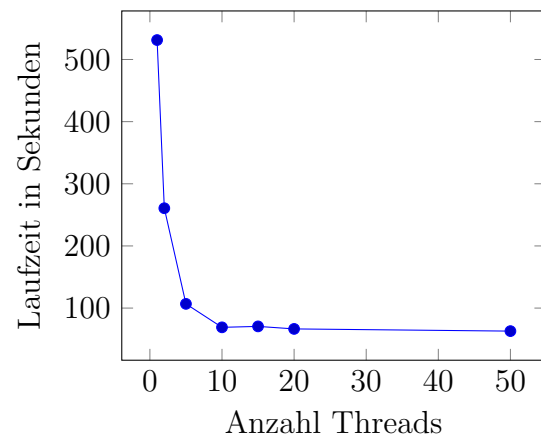


Abbildung 10.1: Anzahl der Threads im Verhältnis zu Laufzeit

Modul	Durchschnittslaufzeit in Sekunden
Blacklist_IPVoid	1,091
Blacklist_MXToolbox	—
DNSresolver	0,314
DomainAge	0,495
GeoIP	2,195
GoogleSafeBrowsingAPI	0,145
GoogleSearch	1,349
Nmap	4,836
RobotsTxt	0,284
SpellChecker	0,095
Traceroute	10,026
Typo	0,583
Whois	0,166
Alle Module	10,043

Tabelle 10.1: Laufzeiten der einzelnen Module für die Domain Amazon.de

Literaturverzeichnis

Sixt, E. (2017). *Bitcoins und andere dezentrale Transaktionssysteme*. Springer Fachmedien
Wiesbaden, Wiesbaden. DOI: 10.1007/978-3-658-02844-2.