

## Kurzfassung

### Stimmungsquantifizierung für den Preis von Bitcoin mit Deep Learning

Die Vorhersage der Aktienkurse spielt in der Geschichte eine wichtige Rolle. Wer den Kurs am genauesten vorhersagen kann, gewinnt. In der jüngeren Vergangenheit erfreuen sich externe Quellen, unter anderem Stimmungen in Zeitungen oder Blogs, großer Beliebtheit. Weiterhin schwächen kürzliche Veröffentlichungen die vollständige ökonomische Rationalität und weisen signifikant in die Richtung des Herdentriebes, engl. *Animal Spirits*. Inspiriert von dieser Kombination übertragen wir das Prinzip auf die neuartigen Systeme Bitcoin und den Microbloggingdienst Twitter.

Um dies zu bewerkstelligen laden wir 10,31 Millionen Tweets vom 10.01. bis 01.08.2018 herunter und speichern sie. Danach erstellen wir mittels Amazon Mechanical Turk unseren Trainingsdatensatz, den wir für das Training der Modelle verwenden. Im nächsten Schritt trainieren wir etablierte maschinelle Lernverfahren, wie Random Forest und SVM, und ziehen lexikalische Verfahren wie vaderSentiment und sentimentr heran. Der Hauptteil unserer Arbeit konzentriert sich auf die Erstellung und Verbesserung von bestehenden Deep Learning-Modellen. Das erste Modell HDLTex ist inspiriert von K. Kowsari und besteht ausschließlich aus KNN-Layern. Das zweite Modell stammt von Y. Kim und besteht aus einer Mischung aus KNN- und CNN-Layern. Beide Modelle dienen als Ausgangslage, die wir im Laufe der Thesis signifikant auf unser Szenario weiter anpassen.

Neben einer enormen Anzahl an Tuning-Parametern und zeitintensiven Trainingsphasen, können wir nachweisen, dass Deep Learning-Netze gegenüber den üblichen lexikalischen Verfahren und maschinellen Lernverfahren ebenbürtig sind. Außerdem können wir die Modelle auf unserem Fall signifikant verbessern, sodass unsere Deep Learning-Modelle mit unseren Top-Verfahren mithalten können. Entgegen der allgemeinen Meinung stellen wir fest, dass Deep Learning-Modelle auf wenig Daten durchaus gut mit einer Genauigkeit bis zu 85 % abschneiden können. Ebenfalls können wir behaupten, dass die lexikalischen Verfahren mit einem geringen Trainingsdatensatz als sehr gute Klassifikatoren mit einer Genauigkeit bis zu 85 % arbeiten.

Nach unserem Modellbau beschäftigen wir uns mit dem Nachweis einer Verbindung zwischen dem Bitcoin-Niveau und unserem Sentiment, indem wir mit unseren Modellen Sentiments erstellen. Die tageweise erstellten Sentiments verbinden wir mit dem zugehörigen Bitcoin-Niveau. Aus unserer Zeitreihe untersuchen wir den Sentimentindex und können damit einen hoch signifikanten Bezug zu Bitcoin bis zwei Tage in die Vergangenheit nachweisen. Daraus folgt, ist die Stimmung positiv, kann das Bitcoin-Niveau steigen und umgedreht.

Bernhard Preisler

Schlagwörter: *Deep Learning, CNN, KNN, Sentimentindex, Text Mining, Twitter, Bitcoin, Animal Spirit*

## Abstract

### Quantifying Sentiment for the Price of Bitcoin using Deep Learning

The prediction of share prices plays an important role in history. The one who can predict the price most accurately wins. In the recent past external sources, such as moods in newspapers or articles, have enjoyed great popularity. Furthermore, recent publications weaken complete economic rationality and point significantly in the direction of “animal spirits”. Inspired by this combination we transfer this principle to the novel systems Bitcoin and the microblogging service Twitter.

To accomplish this, we download and save 10.31 million tweets from the 10th of January till the 31st of August 2018. Afterwards we use Amazon Mechanical Turk to create our training dataset, which we use for training the models. In the next step, we train established machine learning methods, such as Random Forest and SVM, and use lexical methods such as vaderSentiment and sentimentr. The main part of our work focuses on the creation and improvement of existing Deep Learning models. The first model HDLTex is inspired by K. Kowsari and is build of exclusively by KNN layers. We take the second model from Y. Kim which consists of a mixture of KNN and CNN layers. Both models serve as a starting point, which we improve significantly our scenario in the course of the thesis.

In addition to an enormous number of tuning parameters and time-consuming training phases, we can demonstrate that Deep Learning networks are superior to the usual lexical and machine learning methods. Furthermore, in our case we can improve our models significantly, that the Deep Learning models can compete with our top procedures. Contrary to the general opinion we place that Deep Learning models can perform well with an accuracy till 85 % on a small amount of data. We can also say that the lexical procedures work with a small training data set as very good classifiers with an accuracy till 85 %.

After our model building we are concerned with proving a connection between the Bitcoin level and our sentiment index created by our models. We combine the sentiment created on a daily basis with the corresponding Bitcoin level. We take a sentiment index from our time series and can prove a highly significant reference to Bitcoin.

Bernhard Preisler

Key words: *Deep Learning, CNN, DNN, Sentimentindex, Text Mining, Twitter, Bitcoin, Animal Spirit*