# Deep Learning Project: Abstract

Andrea Cicchini, Alexandru Ionut Pascariu and Victor Plesco

## Contents

# 1. Introduction

By October 2021, the two largest cryptocurrencies, measured in terms of market capitalization, had a combined market value of 1562.3 billion dollars. Bitcoin alone made up nearly $1116.3 billion of this value. Given the significant value of these currencies, some people see value in them through use as actual currencies, while others view them as an investment opportunity. The result has been large swings in value of both currencies over short periods of time. During 2017 the value of a single Bitcoin increased 2000% going from $863 on January 9, 2017 to a high of $17.550 on December 11, 2017. The growth and interest were primarily caused by news stories which reported the unprecedented returns of cryptocurrencies, that subsequently attracted a type of gold rush. By eight weeks later, on February 5, 2018, the price of a single Bitcoin had been more than halved with a value of $7.964. Today, October 15, 2021, you can buy a single Bitcoin for a modest price of $59.315. The promising technology behind cryptocurrencies, the blockchain, makes its likely that cryptocurrencies will continue to be used in some capacity, and that their use will grow (Abraham et al., 2018) [1]. Simultaneously, current global regulations on cryptocurrencies are very limited, as cryptocurrencies are not yet acknowledged as a mature asset class. This regulatory void, in combination with the high popularity and lack of an institutional guarantor, makes the cryptocurrency market so volatile that it has even been called a "*wild west*".

The volatility in the value of cryptocurrencies is strongly fueled by news messages and posts on social media and means uncertainty for both investors, and people who wish to use them as a currency rather than an investment. This effect is further reinforced, as investors struggle to discover whether the posted information is true or false. Due to the relatively young age of the cryptocurrency market (Bitcoin was created in 2009) relative to fiat currencies such as the U. S. dollar (USD) or the Japanese Yen, traditional news outlets do not always timely report events, what has led to social media being a primary source of information for cryptocurrency investors. Specifically, micro-blogging website Twitter is a widely used source for cryptocurrency information. Not only does Twitter provide live updates on cryptocurrencies, it is also a rich source of emotional intelligence, as investors frequently express their sentiment (Kraaijeveld and De Smedt, 2020) [2].

Within the context of economics, Kaplanski and Levy (2010) [3] define sentiment as any misperception that can lead to mispricing the fundamental value of an asset. Sentiment can therefore make assets speculative, as according to Baker and Wurgler (2007) [4], and the crucial characteristic defining what makes some assets more speculative than others is "*the difficulty and subjectivity of determining their true value*". This is related to a fundamental psychological concept of (financial) markets which states that decision-making is driven by psychological factors and/or emotions and that therefore market behaviour is not always synonymous with the fundamental value of an asset. Investors can use this fundamental concept in a quest to profit from assets which, based on their sentiment, are either over- or undervalued (Peterson, 2016) [5].

# 2. Problem Statement

To start simple, online social network sentiment analysis for financial forecasting has been the talk of the town within the financial community for a decade so far. With transfer learning, using a hundred million parameters language model (e.g., BERT, RoBERTa, ect.) to classify the sentiment polarity of a document has become much simpler, by fine-tuning a single last neural network layer, and much more accurate. Nevertheless, the methods underneath the training process has remained the same. A decent large corpus of human pre-labeled documents is fed to a machine, which in turn tries to understand the mechanisms driving the cognitive process of defining something as positive or negative. How accurate can this be? Is positiveness objective?

A further doubt relates to the all well known debate: which came first, the chicken or the egg? The causality dilemma between a positive sentiment springing the increase of the price, and that of an increase of the price followed by a cascade of positive sentiment across the web is surely a question one would debate around. Guo and Li (2019) [6] discovered that the accuracy of the sentiment score for an upward trend outperforms that for a downward market, while in Pant et al. (2018) [7] Pearson's correlation coefficient between sentiment and price is found to be higher for the negative sentiment and corresponding fall in price than that for the positive sentiment and its corresponding increase in price. Two opposite results, for two poles representing the same data. Could it be that we're looking the right way but with the wrong approach?

We aim to address this question by proposing an alternative way of calculating the influence of natural language, within a context of online social networks, over the financial markets. To do this, we first start by changing the labeling of our data, and instead of letting humans interact with the polarity, we wish the system itself to do that. By accepting the hypothesis of social networks influencing the price of cryptocurrency assets, we label a tweet, with a pre-defined lag of time between its posting and the price it aims to influence, with a positive or negative numerical value based on the increase or decrease in price of the cited cryptocurrency in the next hour. How can this improve what was questioned so far? We hope that our machine will be able to understand the language anticipating the increase or decrease in price of a cryptocurrency and tell us in advance which investment strategy to adopt.

From a practical point of view the workflow will proceed as follows. Given a portfolio of cryptocurrencies randomly chosen among the first fifteen cryptocurrencies ordered by their market capitalization, their hourly prices and tweets citing them through cashtags are downloaded in an interval of four weeks. A pre-defined lag between the posting of the tweet and the price of the cryptocurrency it represents, serves as the labeling criteria: positive if the price has increased in the future, and negative if the price has decreased in the next timestamp. The corpus is then fed to a pre-trained language model (e.g., BERT, FinBERT, RoBERTa, ect.) to whom we add a linear layer who will output a numerical value representing the increase of decrease expected for that tweet to provoke. The results are subsequently compared to a more classical approach of sentiment analysis on a specific historical case.

# References


1. Abraham, Jethin; Higdon, Daniel; Nelson, John; and Ibarra, Juan (2018). Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. *SMU Data Science Review*: Vol. 1, No. 3, Article 1.

2. Kraaijeveld, Olivier; De Smedt, Johannes (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*: Volume 65, 101188.

3. Kaplanski, G.; Levy, H. (2010). Sentiment and stock prices: the case of aviation disasters. *Journal of Financial Economics*: Vol. 95, Issue 2, Pages 174-201.

4. Baker, M.; Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*: Vol. 21, No. 2.

5. Peterson, R. L. (2016). Trading on Sentiment: The Power of Minds Over Markets. *Wiley*.

6. Guo, Xinyi; Li, Jinfeng (2019). A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency. *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*.

7. Pant, D. et al. (2018). Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*: Pages 128-132.