Cryptocurrency Price Prediction using Social Media Sentiment

Battach Marouane
CentraleSupelec
marouane.battach@student-cs.fr

Oussalah Anouar
CentraleSupelec
anouar.oussalah@student-cs.fr

1. Motivation and Problem Definition

These last couple of years, cryptocurrencies are more and more attracting people, both from public and private sectors. Indeed, banks, investors and many other companies introduce cryptocurrencies to financial markets and create very disruptive technologies using these currencies. In addition, Salvador and Central Africa adopted Bitcoin as their official currency, respectively in 2021 and 2022. Therefore cryptocurrencies are taking a paramount position in our modern societies, and this trend will certainly strenghten in the coming years. However, these currencies are very volatile, which may be problematic for a country who can see its wealth half cut if the bitcoin price decreases by half from a day to another. As a result, predicting price of bitcoin becomes very crucial in this context, and accuracy of prediction also.

As already stated, Bitcoin (BTC) price is very volatile, and one of the factors that can have an impact on its price is social media mood. As a matter of fact, when Elon Musk announced that Tesla cars could be bought in BTC, its price increased by 2%. However, two months later, Musk tweeted that Bitcoin was rapidly increasing use of fossil fuels for bitcoin mining, so Tesla cars won't be bougt in cryptocurrencies. Bitcoin price dropped about 5% in the first minutes after Musk's announcement.

1.1. Related work

There is actually a previous work which showed the existing correlation between Twitter mood and stock market [3]. In addition, similarly to what we're trying to do, a group of Stanford students worked on predicting 3 cryptocurrencies price using news and social media [2]. However, they used some basic algorithm for prediction, wehereas as presented in the following section, we will use more complex and state of the art approaches.

2. Methodology

The main objective of this project is to answer the question of how to predict the reaction of Cryptocurrency Price to social media and news article, which are rich in valuable information [3].

Apart from the mood information, the cryptocurrency Prices are affected by many factors such as microeconomic and macroeconomic factors. However, this project only focuses on how the mood information from social media can be used to predict cryptocurrency price movements.

In order to investigate whether social media sentiment, as expressed in large-scale collections of daily Twitter posts, can be used to predict the cryptocurrency market, we develop a forecasting model to predict Bitcoin price movement using information from social media and historical information. Our methodology will consist of the following steps: data Collection, data Preprocessing, building a sentiment classifier, building a forecasting model, then make predictions and evaluating our model.

The methodology that will be used in this project builds on several techniques[3] [4]

The raw Bitcoin values are first fed into the preprocessor to obtain the processed values. At the same time, the tweets are fed to the sentiment analysis classifier which outputs mood values for the three mood classes for each day (positive, neutral, negative). The proportion of these moods will allow us to calculate a mood index This index and the processed Bitcoin values are then fed to our forecasting hybrid neural network model which uses CNN+LSTM.

Figure 1 shows a brief flow diagram of our methodology.

2.1. Twitter sentiment

Performing Natural Language Processing (NLP) on data extracted from social media is not a straightforward task. In order to perform our classification problem, we will use BERT to train a text classifier. Specifically, either we will take the pre-trained BERT model, add an untrained layer of neurons on the end, and train the new model for our classification task, or we will use it for feature extraction, we start with a pretrained model and only update the final layer weights from which we derive predictions.

This fine-tuning or feature extraction procedure was shown to achieve state of the art results with minimal task-specific adjustments [1]. Rather than implementing custom

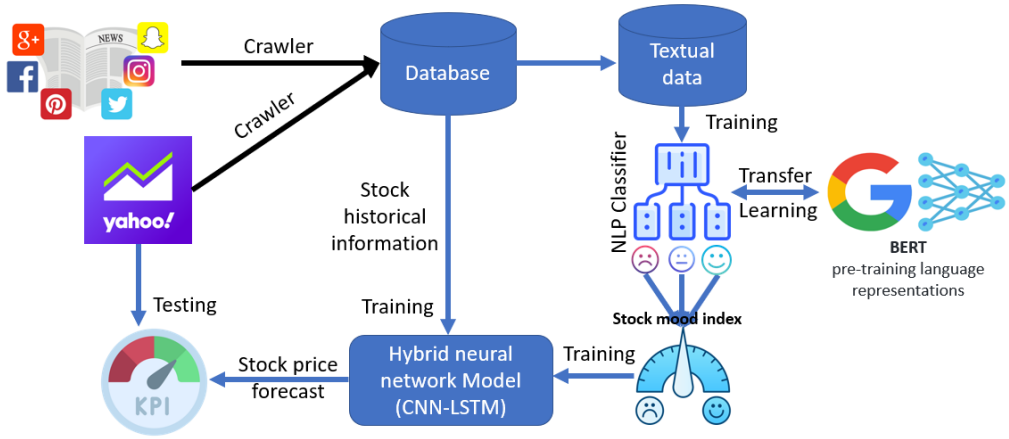

Figure 1. Brief flow diagram for Cryptocurrency Price Prediction using Social Media Sentiment.

model from scratch, using BERT is shown to be a better (or at least equal) alternative.

2.2. Forecasting method

Further to traditional econometric techniques, we conduct analyses of return forecast using a deep learning (DL) methodology and a hybridization of two deep leaning approaches.

In this project we exploit the greater pattern recognition power of CNN and build a hybrid model combining CNN and LSTM (CNN-LSTM) [4]. The input data is first processed by a convolutional layer and then summarised by a pooling layer. The information is then forwarded to an LSTM layer that processes it and generates the output. The LSTM model was proposed as an effort to overcome limitations of recurrent neural networks, such as the exploding and vanishing gradients problem.

The input variables of the forecasting model are the historical information of bitcoin values and the daily mood index.

2.3. Data

Financial information regarding Bitcoin are gathered from Yahoo Finance. The data includes the open, close, high and low values for a given period that we will determine later.

In regards to the social media platform, we will collect daily twitter post synchronised with financial information and whose content is related to the topic of Cryptocurrencies. Obviously, we can not process all the data related to cryptocurrencies due to their huge volume. Therfore, we will limit ourselves to a few hashtags to collect the data and we will hope that il will represent the public mood.

To build our NLP classifier, we will need an external dataset to train it.

3. Evaluation

As explained above, we will have 2 DL models that we will need to evaluate.

3.1. NLP Classifier

Since it's a classification problem, we can use the well known metrics (confusion matrix, accuracy, f1-score) depending on the distribution of labels (balanced dataset or not).

3.2. Hyprid neural network model : CNN-LSTM

Here, our goal is to predict the price of the Bitcoin. In order to assess the accuracy of our predictions, we can compute the RMSE, MAE or MAPE, based on day to day predictions on the test set.

References

[1] Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019. 1

[2] Eric Nielsen Connor Lamon and Eric Redondo. Cryptocurrency price prediction using news and social media sentiment. 1

[3] Huina Mao Johan Bollen and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011. 1

[4] Juan C Reboredo and Andrea Ugolini. The impact of twitter sentiment on renewable energy stocks. *Energy economics*, 76:153–169, 2018. 1, 2