Bitcoin Price Prediction based on Twitter feed Sentiment Analysis
ENSAE, Paris

El Eskander Marsit
ENSAE
eleskander.marsit@ensae.fr

1 Problem Framing

Cryptocurrencies is a decentralised and secure form of digital transactions which has taken the world by storm. At this stage of adoption, it's driven by young enthusiasts and early adopters who are very active in social media Bitcoin just like any other traded currency relies heavily on market speculation. Apart from the traditional sources of media namely news articles and television with the rise of Bitcoin we see a new source of information that is social media.

Earlier studies on this issue are based on historical stock prices. Stock market prices are largely fluctuating. The efficient market hypothesis states that financial market movements depend on news, events and product releases, quarterly fincancial results and all these factors will have a meaningful impact on a company's stock value (Pagolu, 2016).

And with the advent of social media, the information about public feelings and investors intentions are easily available in social media and namely Twitter.

2 Experiments Protocol

In this project we will be using tweets as input to a LSTM model in order to modelise and predict the Bitcoin Price movement.

More precisely, we will be using a Long Short Term Memory model having as input historical prices of Bitcoin as well as Tweets sentiments scores.

The sentiments scores are obtained with the VADER (Valence Aware Dictionary for Sentiment Reasoning) which relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores.

The data used for this project are Twitter feed about Bitcoin and cryptocurrencies in november 2017. The hourly historical bitcoin prices are retrieved through Yahoo Finance website.

Among preprocessing steps for the data, we performed a step of truncation to the tweet data to an hourly period and we calculated the mean of the sentiment score over this hourly period.

3 Results

After performing the basic ML of Train/test dataset validation, we obtained this graph highlighting both the actual bitcoin prices as well as the estimated ones.

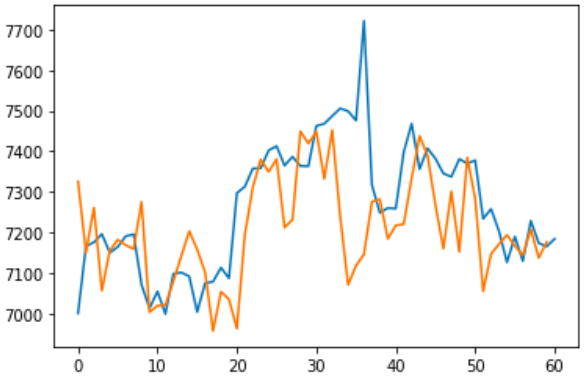

Figure 1: Actual vs Estimated

The Model gives fairly good results taking into account the complexity of this prediction. It has a test RMSE of 152.236.

4 Complexity

Before finally using VADER sentiment Analysis, we tried beforehand using Textblob package: Textblob NaiveBayesAnalyzer takes 3 seconds for each observation against 0.001 seconds for Vader. In fact, trusting the Vader Implementation of the sentiment analysis package it uses a Linear Algorithm $O(N)$ as it uses relies on a dictionary for the scores.

5 Discussion

In this work, we have considered only twitter data for analyzing people's sentiment which may be biased because not all the people who trade in stocks share their opinions on twitter. New platforms of social trading like eToro can provide a more accurate picture of the stock sentiment.
[1]

[1] https://github.com/SkanderMarsit/NLP-SentimentAnalysis

References

Venkata Sasank Pagolu. 2016. Sentiment analysis of twitter data for predicting stock market movements. *SCOPES*.