Do Tweets Have an Effect on Bitcoin

Abstract

The goal of the project was to find if tweets sentiment have an effect on bitcoin price, or if there is a relationship between tweets sentiment and bitcoin price

Design

The first dataset was acquired from kaggle The dataset originally contained more than 18M tweets, the original goal was to process the whole dataset, however there were few problems, also due to the time constraints, the data was limited to the year 2018 only. The Second dataset is from Yahoo Finance and contains bitcoin daily historical data for the year of 2018.

Finding factors that impact bitcoin price could help in predicting it

Data

The twitter dataset contains 699,787 tweets and 9 features each, 3 of which are numerical. The bitcoin data set had 365 entries and 6 features each, all of which are numerical

Algorithms

Feature Engineering

Mapped sentiment scores to each tweet, each score was composed of 3 values: Negative, Neutral, Positive

Model Evaluation and Selection

The training dataset consisted of the first 10 months of 2018, and all scores reported below were calculated with 3-fold cross validation on the training portion only. Predictions on the last 2 months holdout were limited to the very end, so this split was only used and scores seen just once.

Final random forest 3-fold CV scores: 4 features (4 numeric)

MAE: 3151.7475

MSE: 11544013.1256

RMSE: 3397.6482

R2: -8.1333

RMSLE: 0.6082

MAPE: 0.7930

Tools

Numpy

Pandas Hugging Face transformers

Pytorch

Plotly

AutoViz

SweetViz

Fasttext

PyCaret