Proposal for Final Project

Robert Wan

Overview

Bitcoin attracted immense attention in the last two years as it become popular among retail investors. With the creation of products such as Bitcoin Futures, intrinsic value aside, Bitcoin has become a feasible high-risk commodity for investment. Due to its high volatility, Bitcoin is a good product for algorithmic trading. For this project, we will attempt to predict the prices of Bitcoin using past price, popular indicators for technical stock analysis, and social media sentiment. Predicting price with a relatively high accuracy is the first step of forming strategies to algorithmically trade Bitcoin.

Research Questions

The primary goal of the project would be to predict the price of Bitcoin at the end of each day. However, predicting only the daily closing price might not be sufficient for good trading strategies because Bitcoin price is highly volatile even intra-day. As such, if time permits, we will attempt to predict the hourly closing price as well.

Cryptocurrency trading is a new field, and much of the research has been conducted by traders who kept their strategies confidential. Thus, there is not much publication on the factors that affect the price of Bitcoin. Nevertheless, we can hypothesize that Bitcoin price may be affected by three main factors: price in the last X days, technical indicators, and general sentiment. For this project, we will investigate all three factors. We will explore if price is a powerful predictor and look into whether some of the most popular indicators of stock analysis such as SMA, RSI, and momentum are effective at predicting price of Bitcoin. For sentiment of Bitcoin, we will use the number of mentions of Bitcoin on social media as a proxy – we will assume that the more discussion there is, the more favorable the sentiment.

Data

Some Bitcoin datasets are ready available online, such as on Kaggle and from businesses. However, they either lack granularity or don't capture all trades. Fortunately, we can pull data using APIs. For this study, we chose an API provided by CryptoCompare.com. The steps of pulling data is clearly laid out in documentation provided by the CryptoCompare, and additional instructions can also be found here. We will pull variables including time, OHLCV (open price, high price, low price, closing price, and volume), and number of mentions of Bitcoin. We will then compute the technical analysis indicators using the TTR package in R. For the first pass of the project, we'll look at hourly data for one year, which will give us approximately 8,760 rows of observations. If time permits, we will also look at price at 5-minute or 15-minute intervals, which will give us 100,000 or 35,000 observations.

Project Plan

There are two possible approaches to modeling. We can set up a time-series models to explore whether price, indicators, and social media frenzy with X lags are good predictors of current price. Alternatively, we can set up linear regression models to investigate if popular trading signals (e.g. short-term SMA crossing over long-term SMA) are good predictors. Without further investigation, it is difficult to decide which approach would work better or if they'll work at all, and we'll try out both approaches. Milestones and tentative timeline are as follows:

- Pull and clean data: **Tue, Nov. 2nd**
- Create indicators: **Sun, Nov. 7th**
- Build & validate models: **Fri, Nov. 19th**

- Finish the deck: **Sat, Nov. 21st**
- Finish the report: **Fri, Dec. 10th**