Udacity Machine Learning Engineer Nanodegree

Cryptocurrency Signal Generation Capstone Project

Ronen Shohat

Description

Asset managers have been utilizing algorithmic models to generate investment decisions for decades, using both real-time and historical data. With the introduction of the first cryptocurrency, Bitcoin, in 2009 along with its emergence in popularity in 2017 as major cryptocurrencies hit all-time highs, the growth of crypto asset managers has emerged as a new industry. These firms are trying to implement similar quantitative investment strategies found in traditional asset classes to build predictive models for cryptocurrency prices.

Problem Statement

Although the technology of cryptocurrencies has substantial potential, its volatile price swings have deterred many investors from entering the market and has seen other investors suffer substantial losses.

Datasets and Inputs

CryptoCompare API- price and volume data for BTC/USD prices as well as other major cryptocurrencies (LTC, ETH, XRP, BCH, XMR)

Blockchain website for data related to bitcoin market structure/fundamentals

Label will be based on the week over week percentage price of BTC/USD. There will be 3 classes:

- Downward (WoW change was less than -5%)
- Stable (WoW change was between -5% and 5%)
- Upward (WoW change was greater than 5%)

There are 100 features (including transformations) in the dataset, with 558 daily data points ranging from 2017-08-30 to 2019-03-10

The list of features are:

['close',

'high',

'low',

'open',

'volumefrom',

'volumeto',

'intraday_range',

'dod_change',

'rolling_price_vol_7',

'rolling_price_vol_30',

'rolling_DoD_vol_7',

'rolling_DoD_vol_30',

'close_ETH',

'high_ETH',

'low_ETH',

'open_ETH',

'volumefrom_ETH',

'volumeto_ETH',

'intraday_range_ETH',

'dod_change_ETH',

'rolling_price_vol_7_ETH',

'rolling_price_vol_30_ETH',

'rolling_DoD_vol_7_ETH',

'rolling_DoD_vol_30_ETH',

'close_LTC',

'high_LTC',

'low_LTC',

'open_LTC',

'volumefrom_LTC',

'volumeto_LTC',

'intraday_range_LTC',

'dod_change_LTC',

'rolling_price_vol_7_LTC',

'rolling_price_vol_30_LTC',

'rolling_DoD_vol_7_LTC',

'rolling_DoD_vol_30_LTC',

'close_XRP',

'high_XRP',

'low_XRP',

'open_XRP',

'volumefrom_XRP',

'volumeto_XRP',

'intraday_range_XRP',

'dod_change_XRP',

'rolling_price_vol_7_XRP',

'rolling_price_vol_30_XRP',

'rolling_DoD_vol_7_XRP',

'rolling_DoD_vol_30_XRP',

'close_BCH',

'high_BCH',

'low_BCH',

'open_BCH',

'volumefrom_BCH',

'volumeto_BCH',

'intraday_range_BCH',

'dod_change_BCH',

'rolling_price_vol_7_BCH',

'rolling_price_vol_30_BCH',

'rolling_DoD_vol_7_BCH',

'rolling_DoD_vol_30_BCH',

'close_XMR',

'high_XMR',

'low_XMR',

'open_XMR',

'volumefrom_XMR',

'volumeto_XMR',

'intraday_range_XMR',

'dod_change_XMR',

'rolling_price_vol_7_XMR',

'rolling_price_vol_30_XMR',

'rolling_DoD_vol_7_XMR',

'rolling_DoD_vol_30_XMR',

'block_size',

'cost_pct_volume',

'cost_per_transaction',

'mining_difficulty',

'transaction_volume_usd',

'transaction_volume',

'tera_hashes',

'confirm_time',

'mempool_count',

'mempool_size',

'mining_revenue',

'sent_btcusd',

'sent_buybitcoin',

'sent_bitcoin',

'sent_cryptocurrency',

'wallet_users',

'trans_ex_long',

'trans_ex_popular',

'transactions_per_block',

'transactions_total',

'trans_per_day',

'unique_addresses',

'output_volume',

'num_bitcoins',

'usd_volume',

'mining_fees_usd',

'trans_per_second',

'unspent_transactions']

Solution Statement

This project is going to try to build predictive signals on whether a specific cryptocurrency will increase in price, decrease in price, or remain stable. As a classification problem, we will test several supervised learning algorithms to optimize predictive performance.

Potential Algorithms to use:

- Random Forest
- Gradient Boosting
- XGBoost
- SVM (RBF and sigmoid kernels)
- Decision Trees

An ensemble algorithm would be a preferred method as global and local interpretation of the results are very transparent in determining which features drove the prediction

Benchmark Model

To benchmark our model performance, we will backtest the algo's investment decisions to calculate a P&L amount in USD, then compare to a passive strategy of just buy-and-hold.

Evaluation Metrics

Apart from traditional metrics like precision and recall, the main measure that will be used is the USD gained from implementing the algorithm's price predictions, after considering trading costs. We will also assess traditional performance metrics like precision and recall, while considering the fact that 1 step misclassifications (ie upward vs stable) should be penalized less than 2 step classifications (ie downward vs upward). This will determine whether this algorithm provides better price prediction than simply buying and holding the asset.

Project Design

- Collecting/Structuring Data from various APIs
- Data Preprocessing/Cleaning
- Feature engineering
- Model testing/training
- Deciding between regression and classification
- Picking optimal model based on continuous P&L backtests
- Model tuning with gridsearch
- Final conclusion