

Cryptocurrency

A dataset contains day-level price and volume information of 8 cryptocurrencies with a total of 13384 data samples.

Dataset Snapshot								
NATURE OF CONTENT Date, trading price (open, high, low, close price), trading volume, stock ticker and day of the week (0, 1, 2, 3, 4, 5, 6 representing Monday to Sunday).								
BREAKDOWN-BY INSTANCE					NOTES			
Total instances		13384			Stock data is collected daily from 2018-01-02 to 2022-08-01 on all trading days. The recommend split is [0.8,0.1,0.1] for training, validation and testing respectively.			
Training		10712						
Validation		1336						
Testing		1336						
Total cryptocurrencies		8						
Instances per stock		1673						
EXAMPLES OF ACTUAL DATA POINT								
	date	open	high	low	close	volume	tic	day
0	2018-01-02	13625.000000	15444.599609	13163.599609	14982.099609	16846600192	BTC-USD	1
1	2018-01-02	0.008873	0.009598	0.008551	0.009145	89424096	DOGE-USD	1
2	2018-01-02	772.346008	914.830017	772.346008	884.443970	5783349760	ETH-USD	1
3	2018-01-02	228.990005	263.625000	228.990005	255.684006	1237949952	LTC-USD	1
4	2018-01-02	1.041660	1.196800	1.034600	1.160180	108186000	XEM-USD	1

Publisher & Licenses & Access		
PUBLISHER CoinMarketCap		LICENSE PERMISSIONS You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.
DIRECT LINKS TO DATASET link		
ACCESS Open Access	ACCESS COST Free	LICENSE TYPE(S) CC0 1.0

Motivations & Use	
DATASET PURPOSE The dataset was created to provide representative data of cryptocurrency trading for research in various quantitative trading tasks by selecting the most influential cryptocurrencies.	
INTENDED USE CASES <ul style="list-style-type: none">Algorithmic tradingPortfolio Management	EXTENDED USE <ul style="list-style-type: none">Intraday tradingHigh frequency trading

Collection
DATA SOURCE Retrieved from Kaggle
DATA COLLECTION We download the data from Kaggle following this link: https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory

Preprocessing								
<h3>INDICATOR ADJUSTMENT</h3> <p>The raw data consists of 8 indicators, which are date, open, high, low, close, adjcp, volume and tic. Our dataset uses adjusted close price (adjcp) to replace original close price because it is considered as a more accurate measure of cryptocurrency’s value.</p>								
<h3>DATA CLEANING</h3> <p>Firstly, all the NaN terms are dropped, and it is observed that some of the cryptocurrencies are lack of data (number of instances less than 1673). In order to maintain consistency, data of these cryptocurrencies are filtered out.</p>								
<h3>FEATURE GENERATION</h3> <p>We generate 11 temporal features to describe the financial markets. z_{open}, z_{high}, z_{low} represent the relative values of the open, high, low prices compared with the close price at current time step, respectively. z_{close} represents the relative values of the closing prices compared with time step t-1. z_{dk} represents a long-term moving average of the adjusted close prices during the last k time steps compared to the current close prices. The detailed calculation formulas are as follow:</p> <table><tr><th>Features</th><th>Calculation Formula</th></tr><tr><td>$z_{open}, z_{high}, z_{low}$</td><td>$z_{open} = open_t / close_t - 1$</td></tr><tr><td>$z_{close}, z_{adj_close}$</td><td>$z_{close} = close_t / close_{t-1} - 1$</td></tr><tr><td>$z_{d_5}, z_{d_10}, z_{d_15}$ $z_{d_20}, z_{d_25}, z_{d_30}$</td><td>$z_{d_5} = \frac{\sum_{i=0}^4 adj_close_{t-i} / 5}{adj_close_t} - 1$</td></tr></table>	Features	Calculation Formula	$z_{open}, z_{high}, z_{low}$	$z_{open} = open_t / close_t - 1$	z_{close}, z_{adj_close}	$z_{close} = close_t / close_{t-1} - 1$	$z_{d_5}, z_{d_10}, z_{d_15}$ $z_{d_20}, z_{d_25}, z_{d_30}$	$z_{d_5} = \frac{\sum_{i=0}^4 adj_close_{t-i} / 5}{adj_close_t} - 1$
Features	Calculation Formula							
$z_{open}, z_{high}, z_{low}$	$z_{open} = open_t / close_t - 1$							
z_{close}, z_{adj_close}	$z_{close} = close_t / close_{t-1} - 1$							
$z_{d_5}, z_{d_10}, z_{d_15}$ $z_{d_20}, z_{d_25}, z_{d_30}$	$z_{d_5} = \frac{\sum_{i=0}^4 adj_close_{t-i} / 5}{adj_close_t} - 1$							

Maintenance & Status		
STATUS Actively Maintained	FIRST RELEASE 08/2022	CURRENT VERSION 1.0