

Read me

Files

Filename	Description
streamTwitterData.py	live streams Twitter, saving tweets with token, <i>bitcoin</i>
settings.py	settings for streamTwitterData.py including token stings
private.py	private user key and password for Twitter *not included in upload
utilities.py	functions commonly used during development
technicalAnalysis.py	functions to calculate technical indicators
generateTrainingData.py	merges raw twitter data and bitcoin OHLCV data, performs sentiment analysis and calculates technical indicators, saving resulting training data as .csv
indicatorOptimization.py	determines optimal periods for technical indicators
featureSelection.py	creates histograms to view distributions and apply feature transformations to skewed data
ANN.py	ANN model applied to training data, saves resulting predictions of test dataset
marketSim.py	simulation of applying model to real world market including fees and slippage
dataVisualization.R	create plots of statistical analysis and other exploratory data analysis visualizations

Libraries

Python Libraries	Description
pandas	dataframes and aggregating data
numpy	array and numeric computations
matplotlib	plotting
seaborn	plotting
os	managing file directories
re	string manipulation
time	computational time profiling
csv	reading .csv files
tabulate	fancy visualization of pandas dataframes in python console
tweepy	stream twitter data
vader	sentiment analysis
textblob	sentiment analysis
nltk	natural language processing
keras	neural network architecture construction and implementation
scikit learn	logistic regression, scaling and centering

R Libraries	Description
ggplot	plotting
midasr	timeseries column shifting
pals	color palettes
viridis	color palettes
psych	geometric mean
xlsx	open excel files