Bitcoin Prediction with 5 Deep Learning Models

APS 1052 Final Project 2021 Spring

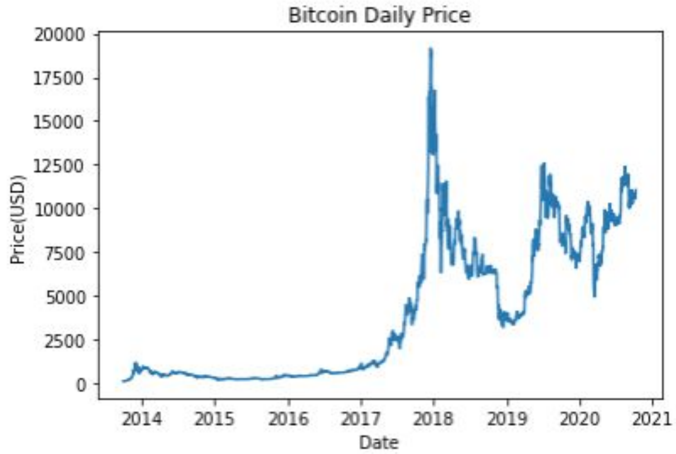
Xiang Cao

Project Goal

- ❖ Predict Bitcoin daily price by using 5 deep learning models
- ❖ Compare performance of LSTM, RNN, WaveNet, Seq2Seq LSTM, GRU
- ❖ Add more features to better predict Bitcoin daily price
- ❖ Establish trading strategies based on model predictions(long only)
- ❖ Evaluate different trading strategies with buy & hold strategy using CAGR, Sharpe and White Reality Check
- ❖ Select best model for trading Bitcoin
- ❖ Determine whether new features are useful to predict Bitcoin daily price

Original DataSet

- Number of Sample: 2556
- Date: from 2013-10-01 to 2020-10-09
- Date Interval: daily
- Number of Features: 16 (including Bitcoin price)



Date	Closing Price (USD)	active_addresses	hash_rate	btc_left	total_addresses	difficulty	total_fees	fed_assets	GLD	IYE	SLV	SPY	TLT	UUP	NYFed_inflation	Google_popularity
2013-10-01	123.65499	89218	1309351515956620	9220600.0	18656301	639173596179762000	39.139746	3.747387e+06	124.589996	47.430000	20.41	169.339996	105.800003	21.610001	1.738484	3
2013-10-02	125.45500	105303	1307159450402430	9215825.0	18708039	639173596179762000	39.923033	3.747387e+06	127.059998	47.540001	20.92	169.179993	105.959999	21.530001	1.738484	3
2013-10-03	108.58483	89993	1452700445873280	9210850.0	18751619	639173596179762000	36.602946	3.748998e+06	127.180000	47.099998	20.92	167.619995	105.790001	21.510000	1.738484	3
2013-10-04	118.67466	77370	1283409529949880	9206350.0	18791964	639173596179762000	24.772795	3.750609e+06	126.529999	47.520000	20.93	168.889999	105.709999	21.610001	1.738484	3
2013-10-05	121.33866	64961	1602048841926960	9200875.0	18823649	639173596179762000	26.062425	3.752220e+06	126.529999	47.520000	20.93	168.889999	105.709999	21.610001	1.738484	3

Dataset:Bitcoin Features

- ❖ Closing Price (USD): closing price of bitcoin in USD
- ❖ Active addresses: total number of unique address that either sent or received bitcoin
- ❖ Hash rate: measure of computing power in the network
- ❖ Bitcoins left: number of bitcoins left to be mined
- ❖ Difficulty: measure of how difficult it is to mine a bitcoin
- ❖ Total fees: total transaction fees per day, representing volume

Dataset : Finance Features

- ETFs:
  - GLD: SPRD Gold Shares track the price of gold bullion
  - IYE: tracks an index composed of US energy sector equities
  - SPY: S&P 500 tracks the performance of 500 large companies listed on US stock exchanges
  - TLT: tracks an index composed of US treasury bonds with remaining maturity of greater than 20 years.
  - UUP: tracks the relative performance of the USD.
- Economics:
  - Fed assets: federal reserve balance sheet

Deep Learning Models

Our team compares the following 5 models that specialize in sequence prediction. Those RNN models are trained on time-series data containing a 3rd temporal dimension.

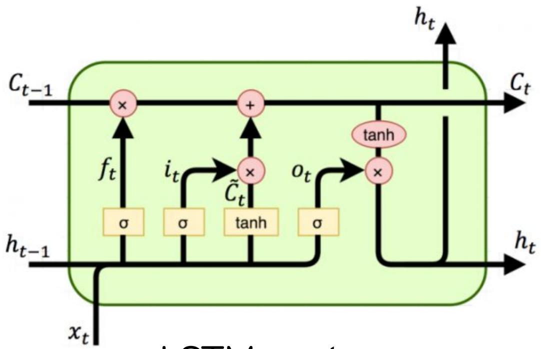
- Original LSTM model
  - Based on S10.Miscellaneous\LSTMBitcoin\core\model.py
- Simple RNN
- WaveNet
- Sequence to Sequence LSTM
  - Modified based on S10.Miscellaneous\GeronLSTM\Geron_mod_LSTM_SPY.py and S10.Miscellaneous\LSTMStockPrice\03_multivariate_stacked_lstm.py
- GRU with CNN
  - also Seq-to-Seq model, modified based on the same above python files

Common Model Parameters Setting

- Look Back Window: 20 time steps
- Future Number of Time Step to Predict: 1 (except sequence-to-sequence models predict multiple future time steps at every timestep)
- Training Data: 64%, Validation data: 16%, Test data: 20%
- Number of Epoch: 40, Batch Size: 16
- Callbacks in training: Early stopping, adaptively reduce Learning rate.
- Model uses MSE loss, and Adam optimizer

Original LSTM model

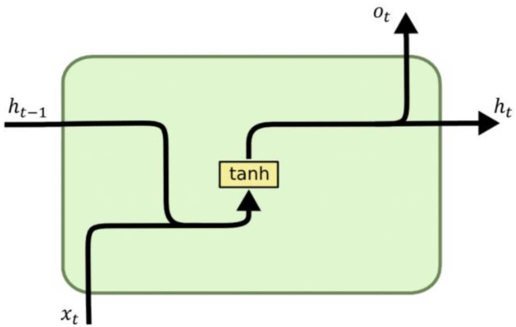
- Model layer structure:
  - Layer 1: LSTM with 100 units
  - Layer 2: Dropout with 0.05 rate
  - Layer 3: LSTM with 100 units
  - Layer 4: LSTM with 100 units
  - Layer 5: Dropout with 0.05 rate
  - Layer 6: Dense with 1 unit and linear activation



LSTM neutron

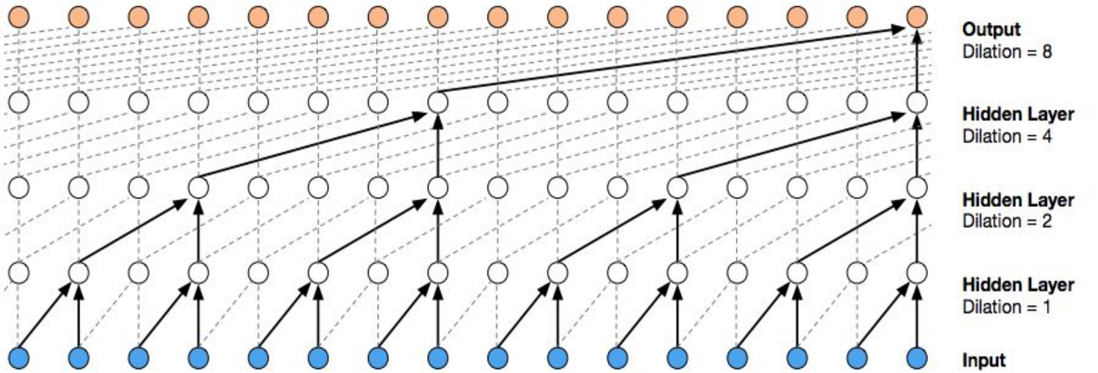
Simple RNN

- Model layer structure:
  - Layer 1: SimpleRNN with 100 units
  - Layer 2: SimpleRNN with 100 units
  - Layer 3: Dropout with 0.05 rate
  - Layer 4: Dense with 1 unit and linear activation



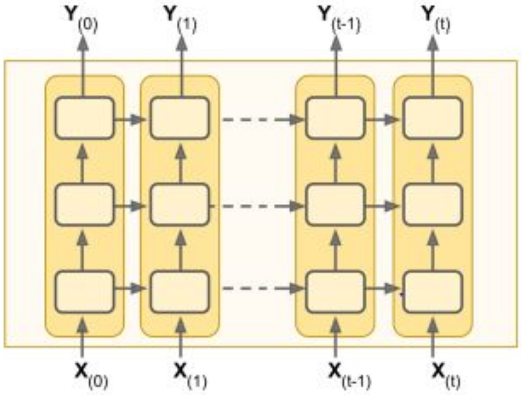
RNN neutron

WaveNet



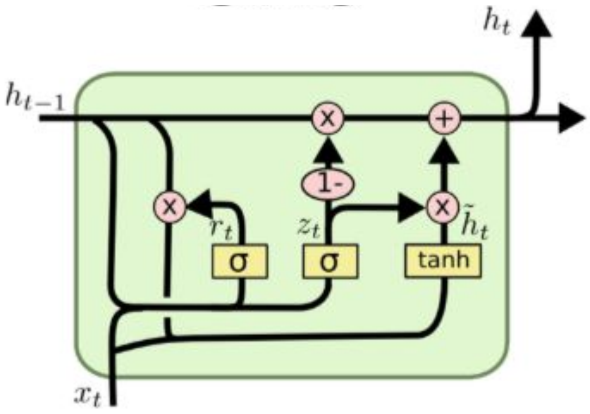
- ❖ Published by Oord et al in 2016
- ❖ Stacking 4 1D Convolutional layers
- ❖ Doubling dilation rate (1-2-4-8) for long-term patterns
- ❖ Repeat this dilation block 3 times for total of 12 1D-CONV layers
- ❖ Capable of handling very long sequence and perform great on audio time-series data

Sequence to Sequence LSTM



- Sequence to Sequence Prediction
  - Look Back Window: 20 time steps
  - Future number of Time Step to Predict: 4
  - Note: The model look back 20 time steps, and predicts 4 future time steps at every time step, and our team only picked the prediction on the 21st time step
- The ability of producing multiple outputs at every time step enables seq2seq model to be more stable and faster to train
- Model layer structure:
  - Layer 1: LSTM with 200 units, dropout=0.05, and recurrent_dropout=0.1
  - Layer 2: LSTM with 200 units
  - Layer 2: Dropout with 0.05 rate
  - Layer 4: LSTM with 200 units
  - Layer 5: Dropout with 0.05 rate
  - Layer 6: TimeDistributed with Dense layer of 4 units

GRU with CNN (Seq2Seq)



GRU neutron

- ❖ Gated Recurrent Unit is similar to LSTM but without output gate
- ❖ Use CONV1D layer to help learn longer patterns
- ❖ Sequence to Sequence Prediction
  - ➢ Look Back Window: 20 time steps
  - ➢ Future number of Time Step to Predict: 4
- ❖ Model layer structure:
  - ➢ Layer 1: Conv1D with 20 filters, kernel size 20 and stride 2
  - ➢ Layer 2: Gate Recurrent Unit (GRU) layer with 20 units
  - ➢ Layer 3: GRU layer with 20 units
  - ➢ Layer 4: Time Distributed Layer of 4 sequences output

Performance Comparison on Original Dataset

	Market (Buy and hold Bitcoin)	Original LSTM Model	RNN	WaveNet	Sequence to Sequence LSTM	GRU
Compound Annual Grow Rate (CAGR)	19.9%	52.2%	42.8%	**69.6%**	44.9%	56.0%
Sharpe Ratio	0.6	1.2	1.1	**1.6**	1.0	1.1
White Reality Check (p-value)		0.060	0.087	0.057	0.018	**0.014**

Additional Features: Commodity futures (COT)

- ❖ We added additional features that report how commodity traders are trading Bitcoin.
- ❖ With the combination of these 5 features, we can track the percentage of long and short positions among traders

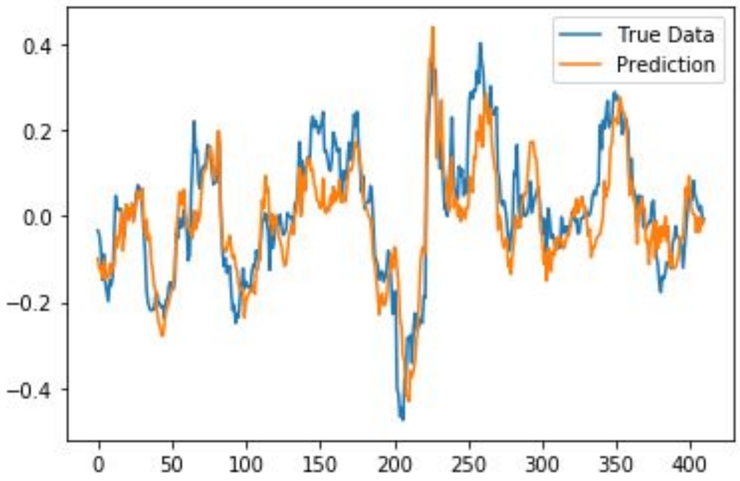
Commodity(COT) Features:	The COT reports can be used to follow influential traders in the commodities markets
Open_Interest_All	Total of all futures and/or option contracts that are not yet terminated or offsetted by a transaction, delivery, or exercise
Tot_Rept_Positions_Long_All	Total number of long positions hold by all reporting traders
Tot_Rept_Positions_Short_All	Total number of short positions hold by all reporting traders
NonRept_Positions_Long_All	Number of long positions hold by non reporting traders (individual speculators)
NonRept_Positions_Short_All	Number of short positions hold by non reporting traders (individual speculators)

Additional Features: Related Stock Price

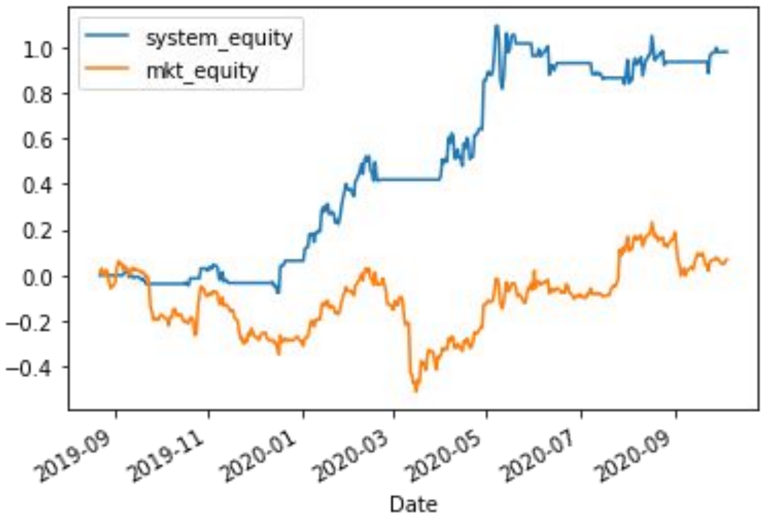
❖ We also added related stock prices that reflect the status of Nasdaq tech, fintech, and Bitcoin mining company.

Related Stock Features:	
QQQ	QQQ tracks the performance of 100 nonfinancial stocks listed on NASDAQ, heavily tilted to megacap tech stocks
FINX	FINX tracks companies on the leading edge of the emerging financial technology sector
RIOT	Stock of a bitcoin mining company which engages in the provision cryptocurrency mining computers

Results - Prediction vs Ground Truth & System Performance



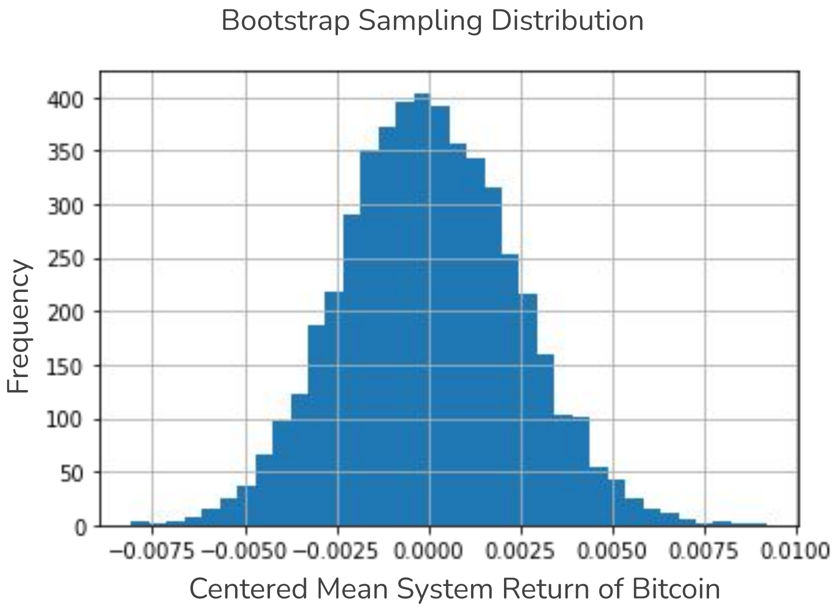
The WaveNet model trained on augmented dataset (24 features) predict closely with ground truth on the 40% testset (Sep 2019 - Oct 2020)



Using predicted result of Wavenet, the system achieved 51.9% CAGR comparing to Market’s 4.2% CAGR.

White's Reality Check

- ❖ Evaluate whether the model performance is a result of data bias
- ❖ We calculate the probability p-value using detrend price and percentile of the data series
- ❖ We use the threshold of 0.1 on p-value:
  - ➢ If p-value >0.1, we do not reject the null hypothesis
  - ➢ If p-value <0.1, we reject the null hypothesis and conclude the obtained results are not due to chance



Final Result

Dataset/ Model	Market (Buy and hold	Original LSTM Model	RNN	WaveNet	Sequence to Sequence with LSTM	GRU
Original Dataset (2013-2020, 16 features)	CAGR:19.9% Sharpe: 0.6	CAGR:52.2% Sharpe:1.2 p-value: 0.060	CAGR:42.8% Sharpe:1.1 p-value: 0.087	**CAGR:69.6% Sharpe:1.6** p-value: 0.057	CAGR:44.9% Sharpe:1.0 p-value: 0.018	CAGR:56.0% Sharpe:1.1 p-value: 0.014
Original (2017-2020, 16 features)	CAGR:4.2% Sharpe:0.4	CAGR:48.3% Sharpe:1.4 p-value: 0.032	CAGR:13.5% Sharpe:0.5 p-value: 0.176	**CAGR:62.7% Sharpe:1.5** p-value: 0.023	CAGR:45.0% Sharpe:1.1 p-value: 0.062	CAGR:50.2% Sharpe:1.3 p-value: 0.053
COT&Stock: (2017-2020, 24 features)	CAGR:4.2% Sharpe:0.4	CAGR:49.1% Sharpe:1.4 p-value: 0.116	CAGR:26.6% Sharpe:0.9 p-value: 0.166	**CAGR:51.9% Sharpe:1.5** p-value: 0.087	CAGR:37.5% Sharpe:1.0 p-value: 0.071	CAGR:38.5% Sharpe:1.0 p-value: 0.064

Discussion

- The dataset size is limited by COT features which are only available after December 2017. The relative small size of dataset would directly affect the the quality of the mapping function approximated by neural networks and the model performance.
- RNN model performance is slightly worse than the original LSTM model on the original dataset. After adding COT&stock features, RNN model performance becomes even worse, which may due to the smaller dataset size.
- Theoretically, seq-to-seq approach has the advantage of stabler gradient and faster training, which may lead to better performance. Yet in this case, seq-to-seq models generally give higher rmse.
- Moreover, Seq-to-Seq models'(LSTM & GRU) performances are always worse than the original LSTM model(Seq-to-Vector) before or after adding COT&stock features, however, Seq-to-Seq models perform better than RNN model(Seq-to-Vector) after adding COT&stock features
- Due to randomness in training and small sample size, a relative worse system sharpe or CAGR does not necessarily disqualify that model.
- WaveNet model performance always exceeds the original LSTM model before and after adding COT&stock features, possibly due to its capability of learning from very long sequence such as historical event years ago. **Thus our team chooses WaveNet to be the best model for Bitcoin prediction.**

Challenge: Loss function vs real-world performance

- Deep learning models are trained on regression loss function RMSE, calculating the difference between prediction price and actual price of Bitcoin
- Smaller loss doesn't necessarily lead to bigger return in real-world trading
- **Challenge: How to design the loss function to learn maximizing real-world trading return ?**

Original 2013-2020 dataset	**Market (Buy and hold Bitcoin)**	**Original LSTM Model**	**RNN**	**WaveNet**	**Sequence to Sequence LSTM**	**GRU**
RMSE		0.061	**0.054**	0.129	0.642	1.034
CAGR	19.9%	52.2%	42.8%	**69.6%**	44.9%	56.0%
Sharpe ratio	0.6	1.2	1.1	**1.6**	1.0	1.1

Conclusion

- Enlarged dataset with COT&stock features generally gives us better result than the original dataset subset.
- WaveNet has the better performance than the other 4 RNN based deep learning models, and is thus recommended for Bitcoin prediction.

Future Work

- As mentioned in challenges, the model will perform better in real world trading, if we can custom define a loss function to train the models that learns how to maximize CAGR and Sharpes, instead of minimizing the difference between actual price and prediction