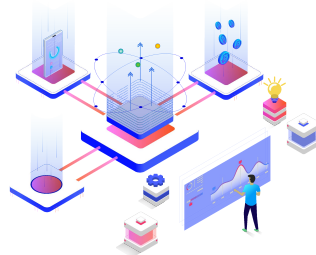


Bitcoin RNN modeling



Objective

- Design and train the RNN networks using Bitcoin time-series with daily data
- Predict the Bitcoin closing price for the next day so that the lowest validation MAE between true and predicted values is achieved
- Perform out-of-sample validation on the test set



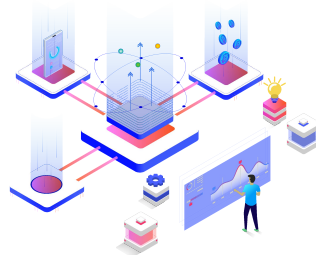
Model

- Dataset was limited to Bitcoin data
- RNN models were tested (LSTM & GRU & SimpleRNN)
- Models with multiple layers were also tested
- Hyperparameters were tuned by hand

$$\begin{aligned} \text{close}_t = & \text{close}_{t-1\dots t-n} + \text{open}_{t-1\dots t-n} + \text{high}_{t-1\dots t-n} + \text{low}_{t-1\dots t-n} + \text{volume}_{t-1\dots t-n} \\ & + \\ & \text{market}_{t-1\dots t-n} + \text{close_ratio}_{t-1\dots t-n} + \text{spread}_{t-1\dots t-n} \end{aligned}$$

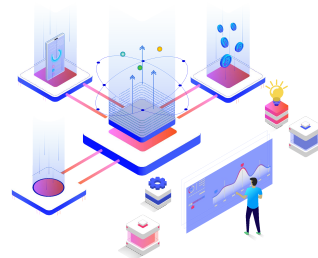
Data preprocessing

- Calculating percentage changes
 - Close, Open, High, Low, Volume, Market, Spread
- Creating lags
- Replacing infinite values with zeros
- Splitting to training/test sets (1000/200 days of data)
- Min-max scaling
- Changing dimensions

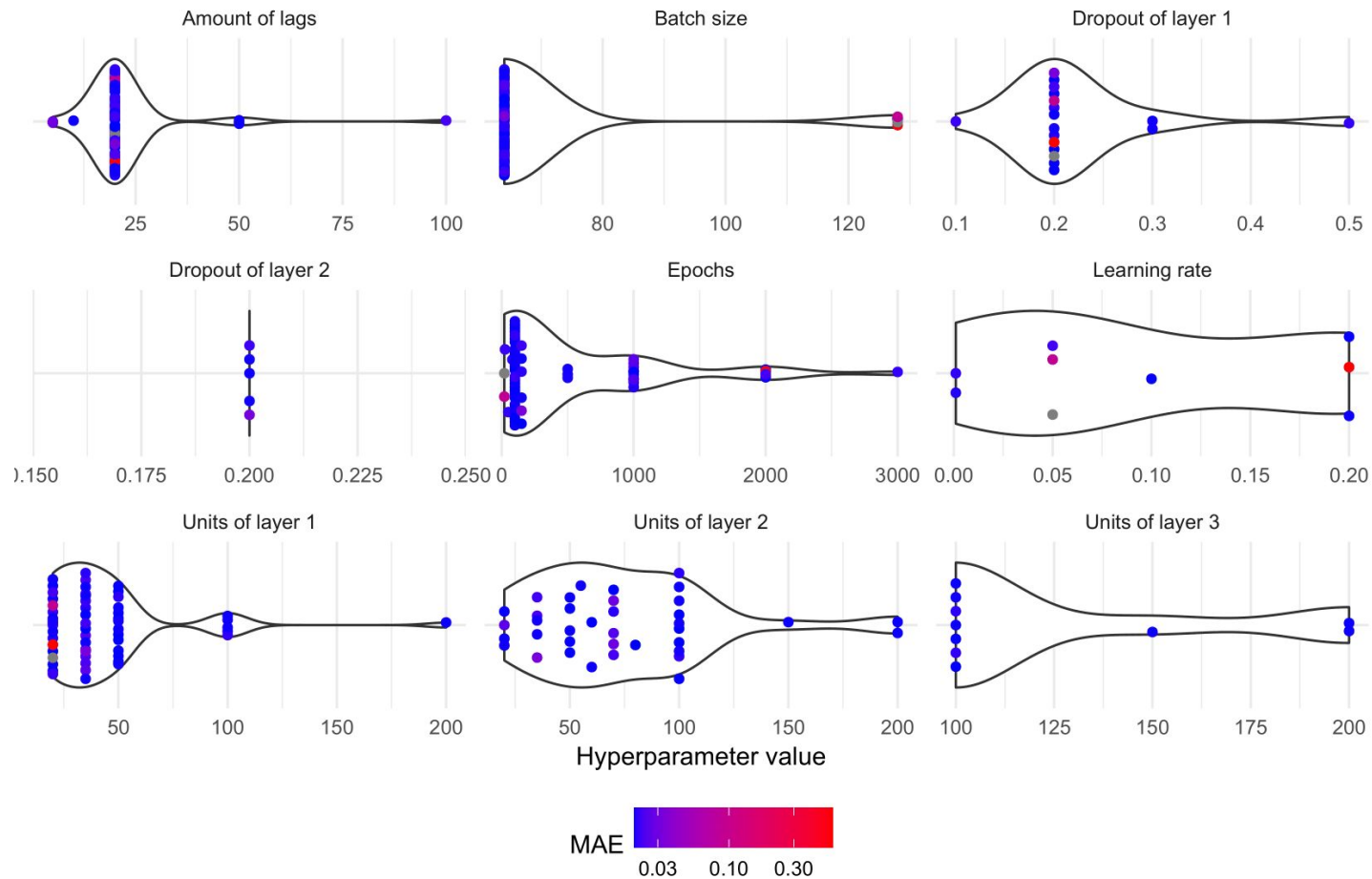


Tested hyperparameters

- Model type: {LSTM, GRU, SimpleRNN}
- Layers:
 - Number of layers: {1, 2, 3}
 - Types of other layers: {LSTM, GRU, SimpleRNN}
 - Units: 20 – 200
 - Activation functions: {tanh, relu, sigmoid}
 - Dropouts: 0 – 0.5
- Epochs: 20 – 3000
- Batch size: 64 – 128
- Optimizer: {adam, SGD, RMSProp}
- Loss function: {mse, mae}
- Amount of lags: 5 – 100



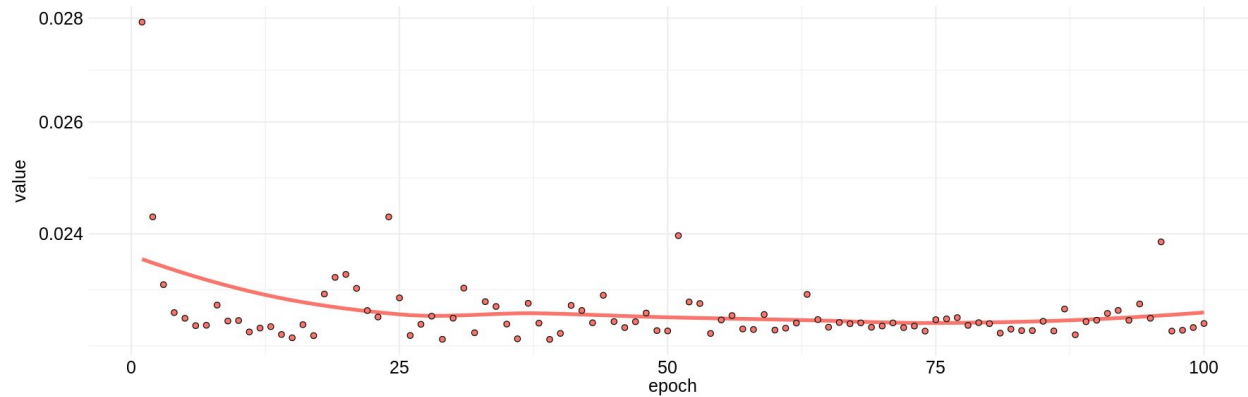
Tested hyperparameters



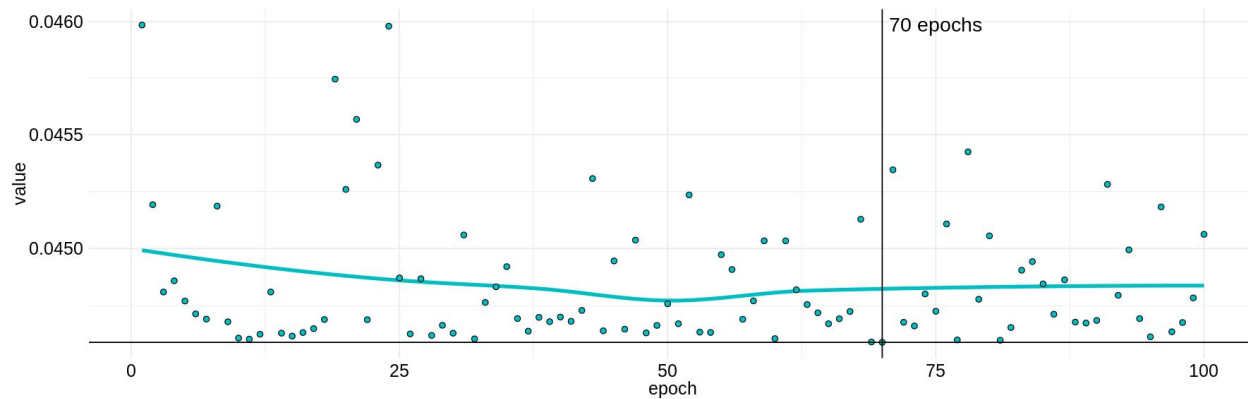
Model fit

Best model as measured by test set accuracy

Training



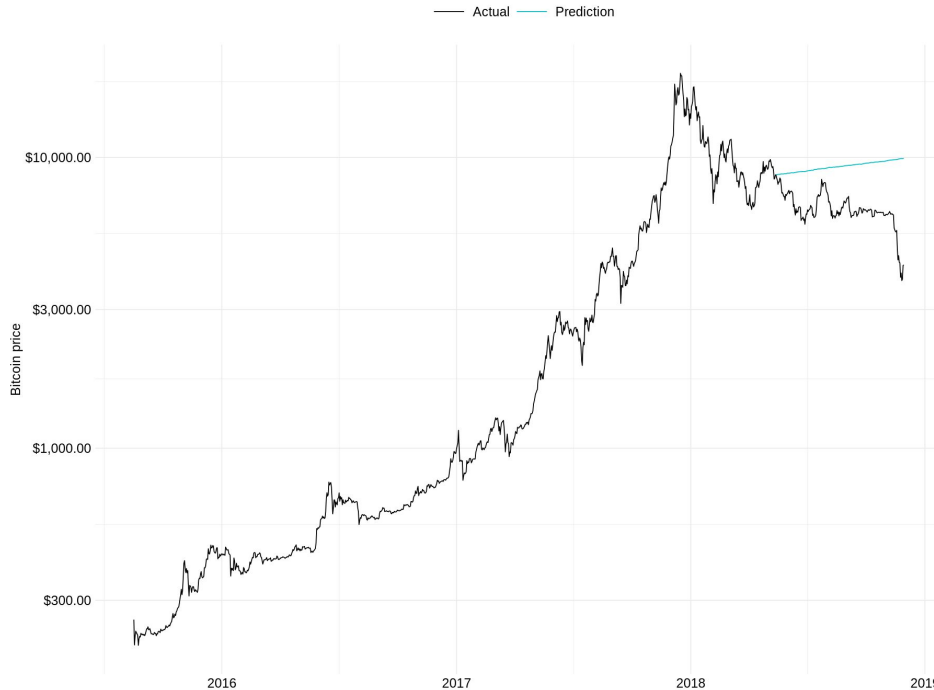
Validation



Min validation loss/MAE: 0.00351/0.04459

Resulting model and performance

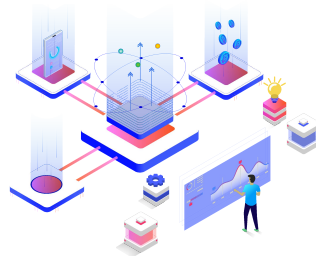
Actuals and predictions



MAE: \$2656
MAPE: 43%
R-squared: -42.7%

Model
Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 20)	1760
dense (Dense)	(None, 1)	21
Total params: 1,781		
Trainable params: 1,781		
Non-trainable params: 0		



Hyperparameters of the best model

- Model type: **LSTM**
- Layers:
 - Number of layers: **1**
 - Types of other layers: -
 - Units: **20**
 - Activation function: **tanh**
 - Dropout: **0.2**
- Epochs: **100**
- Batch size: **64**
- Optimizer: **adam**
- Loss function: **mse**
- Amount of lags: **20**



Conclusion

Difficult to predict prices based on historical data

Data leakage problem in the beginning

Potential improvements

Automated hyperparameters tuning

More features

Consideration of microeconomic factors, including (social) media

