Machine Learning for High Frequency Bitcoin Price Prediction

Chris Bynum

"High Frequency"

Co-location on a sophisticated exchange

- **microseconds**

HTTP to a bitcoin exchange

- **milliseconds / seconds**

Dataset

Recorded live

1. Snapshot of open orders every second

   - 4 weeks: **2.4 million** observations

   - Each observation: matrix of buy / sell orders

2. Record of all executed transactions

Order Book

Bids			Asks		
Price	Amount	Cumulative	Cumulative	Amount	Price
231.24	6.87	6.87	6.93	6.93	231.32
231.23	0.05	6.93	6.94	0.01	231.42
231.02	16.00	22.93	34.42	27.47	231.43
231.01	0.50	23.43	43.88	9.46	231.44
231.00	81.45	104.89	83.88	40.00	231.46
230.95	0.02	104.91	93.89	10.01	231.50
230.78	1.64	106.55	94.39	0.50	231.57
230.77	1.63	108.18	107.14	12.74	231.60
230.76	0.01	108.19	123.60	16.46	231.68
230.74	0.11	108.31	123.61	0.01	231.70
230.65	1.04	109.36	136.69	13.07	231.76
230.64	98.28	207.64	136.70	0.01	231.80
230.62	13.58	221.22	136.71	0.01	231.90
230.60	1.83	223.06	150.45	13.73	231.92
230.59	10.00	233.06	170.45	20.00	231.96
230.51	1.00	234.06	215.45	45.00	231.97
230.50	6.00	240.06	215.57	0.11	232.00
230.46	13.18	253.25	216.63	1.06	232.07

Midpoint Price

	Bids			Asks	
Price	**Amount**	**Cumulative**	**Cumulative**	**Amount**	**Price**
231.24	6.87	6.87	6.93	6.93	**231.32**
231.23	0.05	6.93	6.94	0.01	231.42
231.02	16.00	22.93	34.42	27.47	231.43
231.01	0.50	23.43	43.88	9.46	231.44
231.00	81.45	104.89	83.88	40.00	231.46
230.95	0.02	104.91	93.89	10.01	231.50
230.78	1.64	106.55	94.39	0.50	231.57
230.77	1.63	108.18	107.14	12.74	231.60
230.76	0.01	108.19	123.60	16.46	231.68
230.74	0.11	108.31	123.61	0.01	231.70
230.65	1.04	109.36	136.69	13.07	231.76
230.64	98.28	207.64	136.70	0.01	231.80
230.62	13.58	221.22	136.71	0.01	231.90
230.60	1.83	223.06	150.45	13.73	231.92
230.59	10.00	233.06	170.45	20.00	231.96
230.51	1.00	234.06	215.45	45.00	231.97
230.50	6.00	240.06	215.57	0.11	232.00
230.46	13.18	253.25	216.63	1.06	232.07

Objective

- Engineer features from orders and trades

- Predict midpoint price 30 seconds in the future

Features

23 total, including:

- Market width

- Measures of order book volume imbalance

- A measure of whether buyers or sellers were recently more aggressive
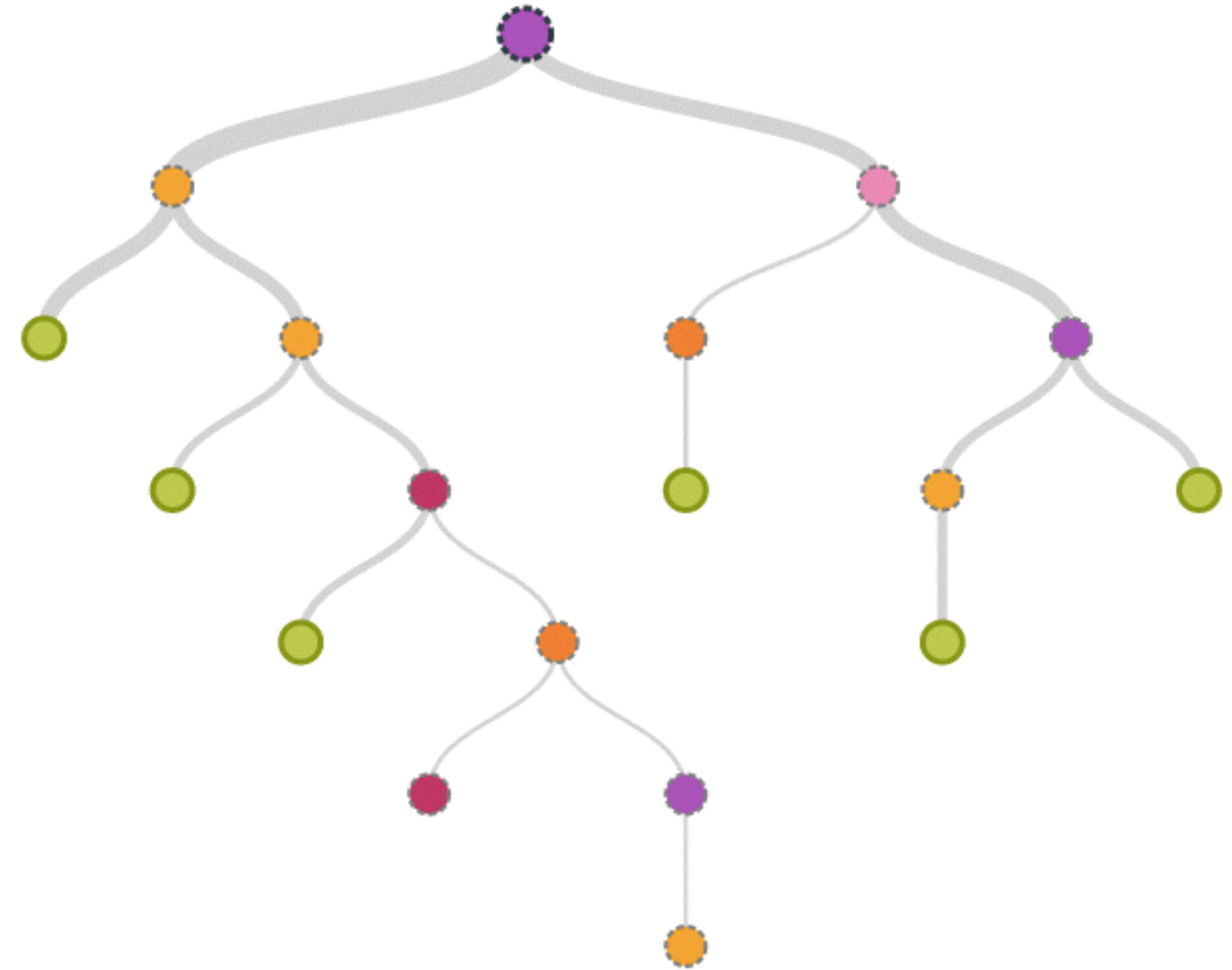
- Other features on trade volumes and prices

The Model

Decision Tree

- Find interactions

Random Forest for experimentation

- Can be parallelized
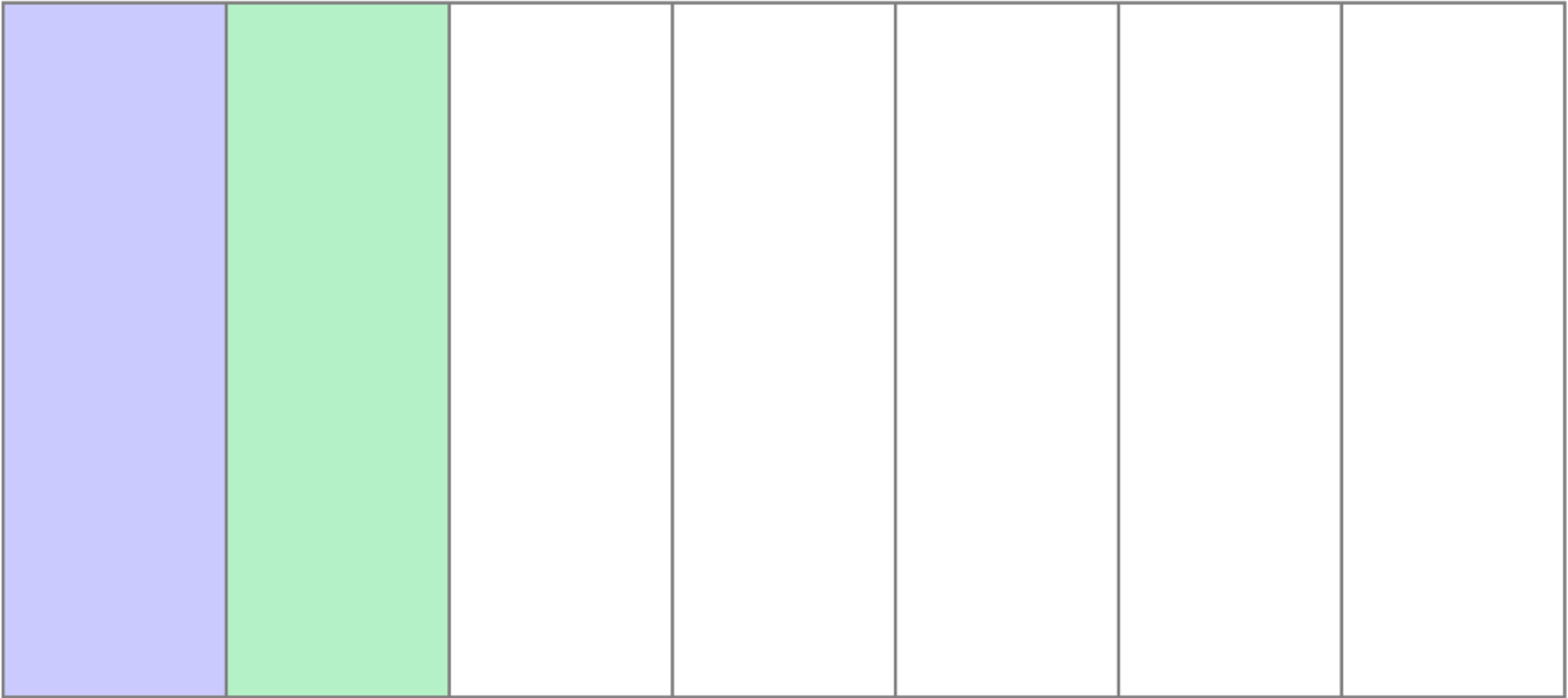
- Not highly sensitive to parameters

Gradient Boosting for final model

Model Validation
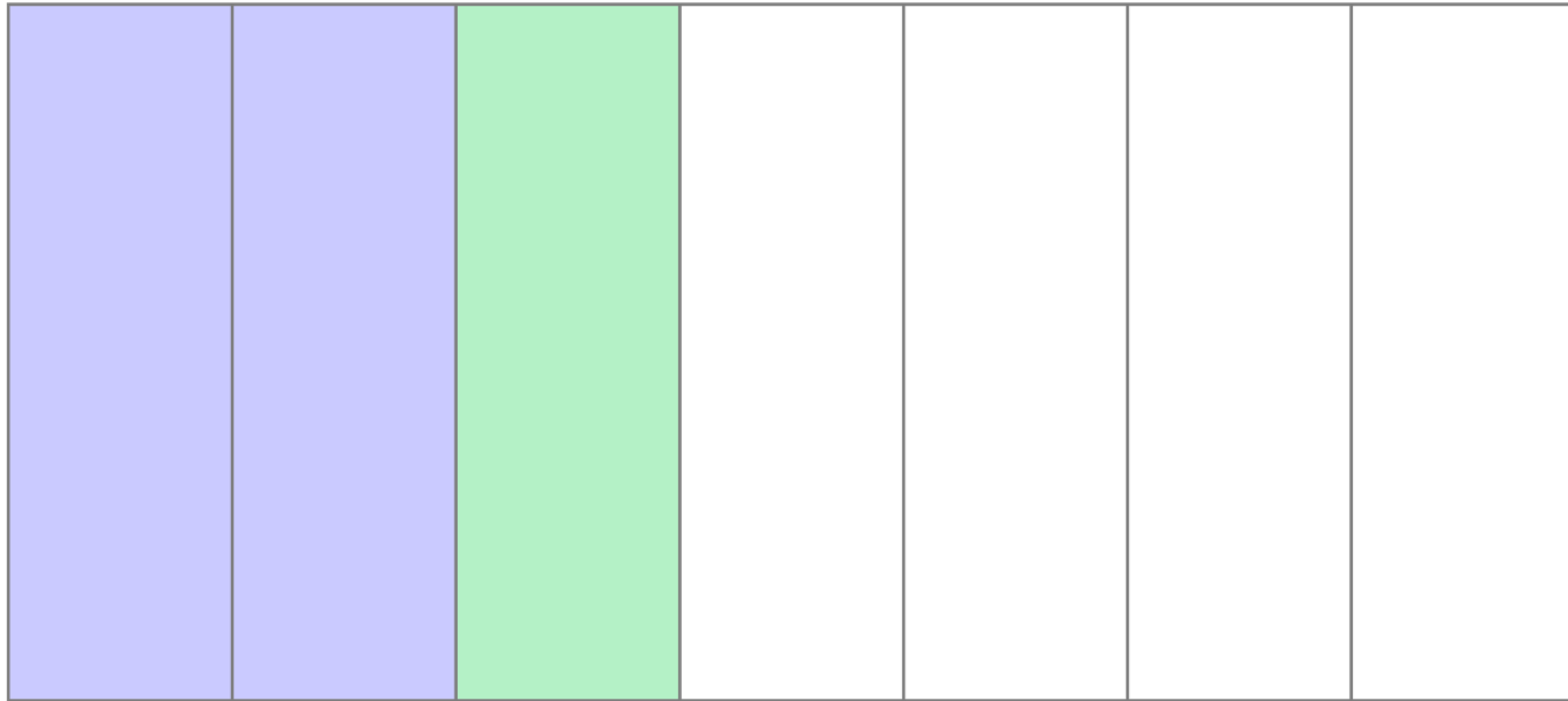
Model Validation

Model Validation
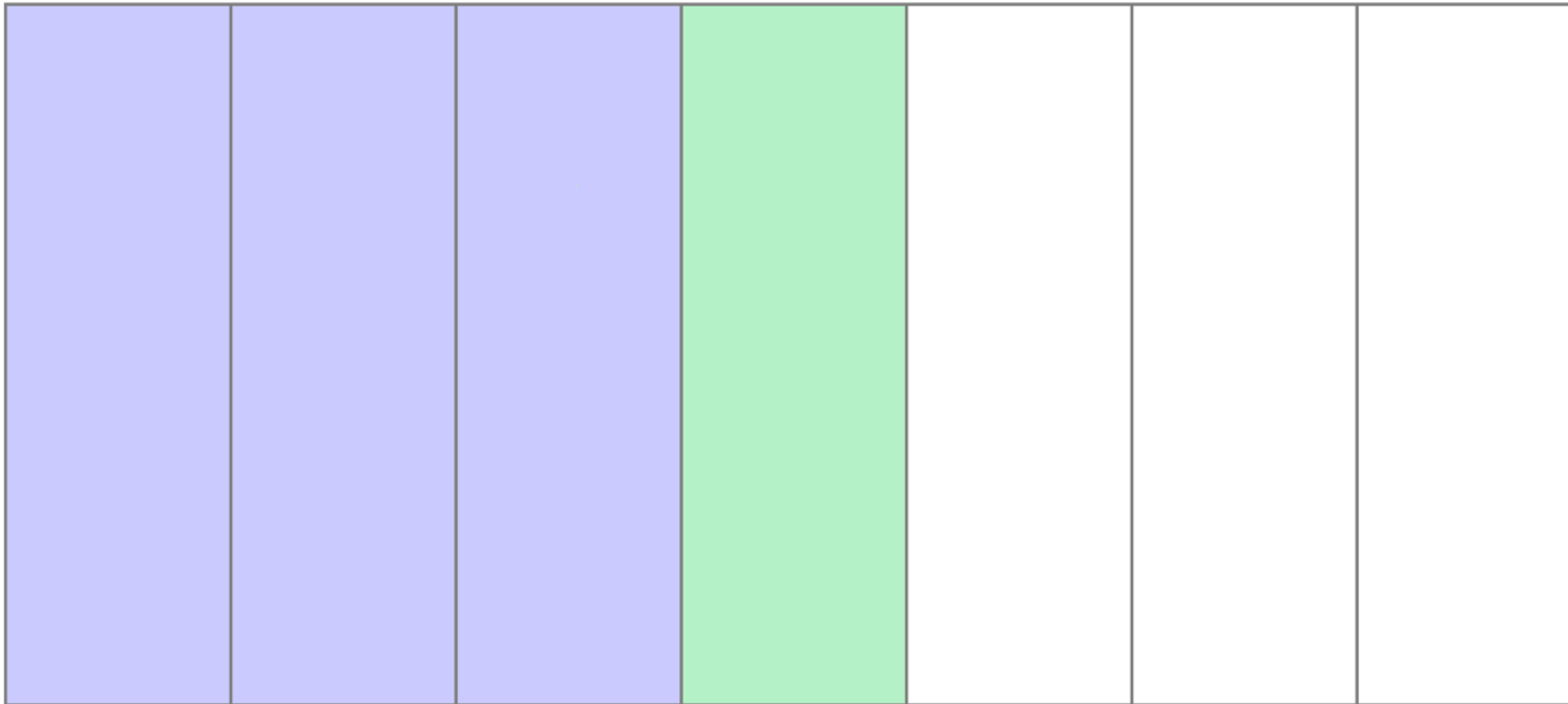
Train   **Test**

Model Validation
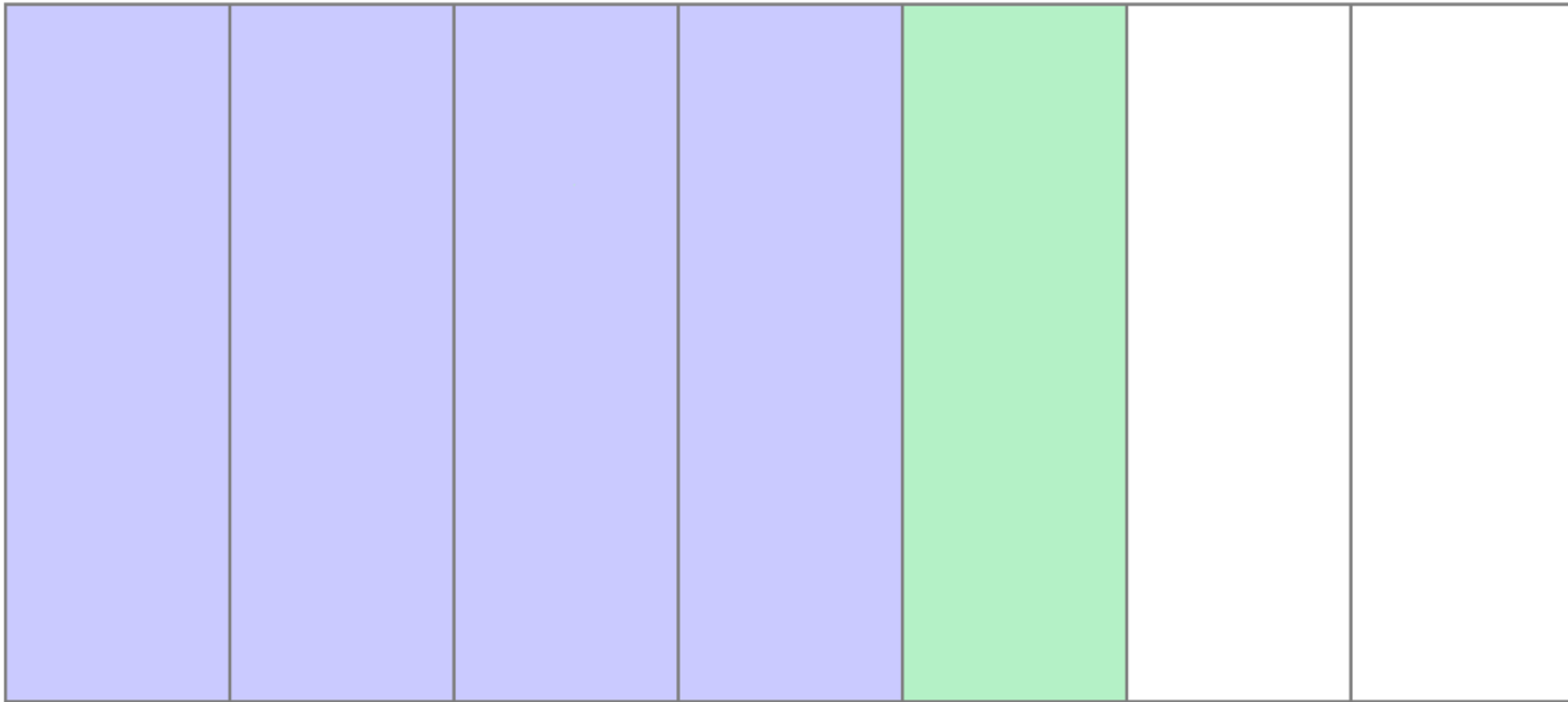
Train **Test**

Model Validation
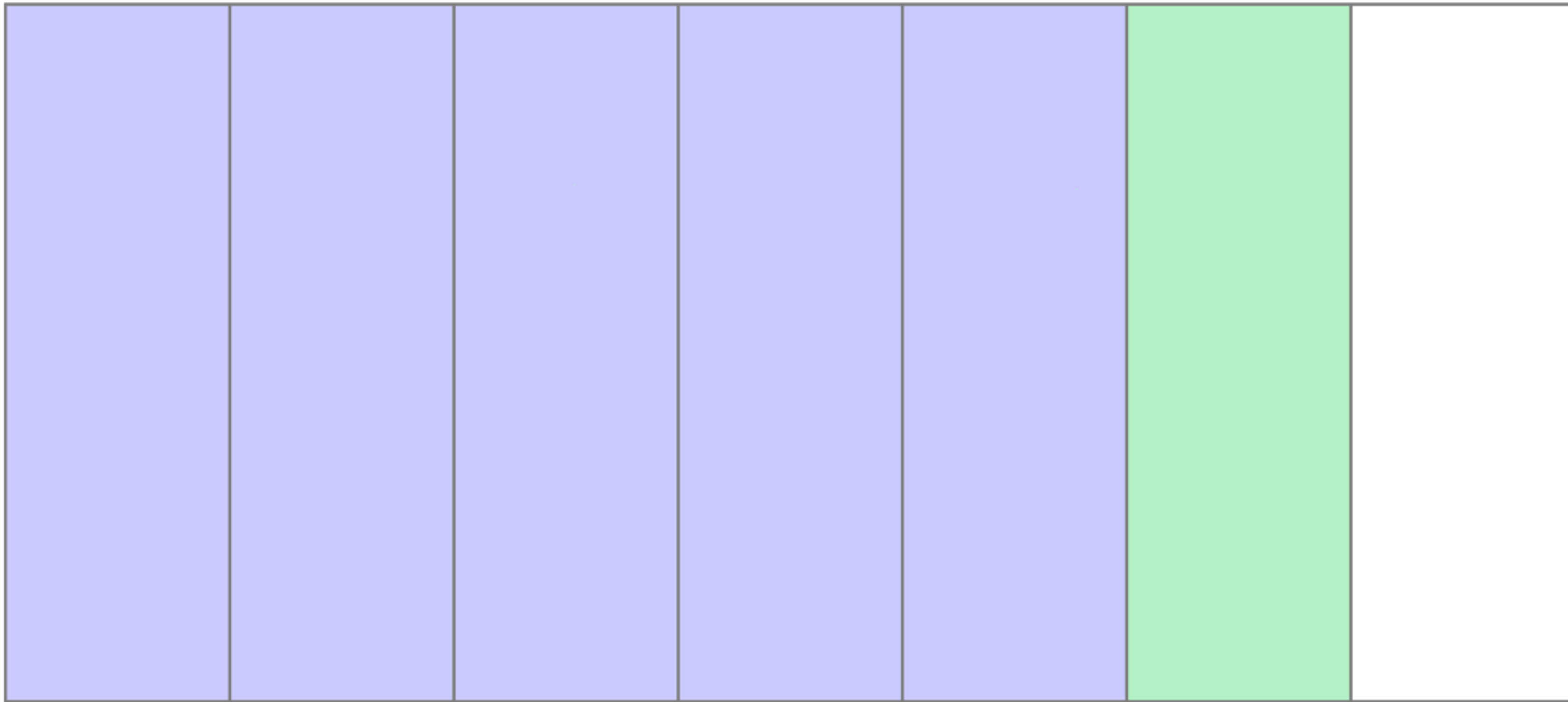
Model Validation

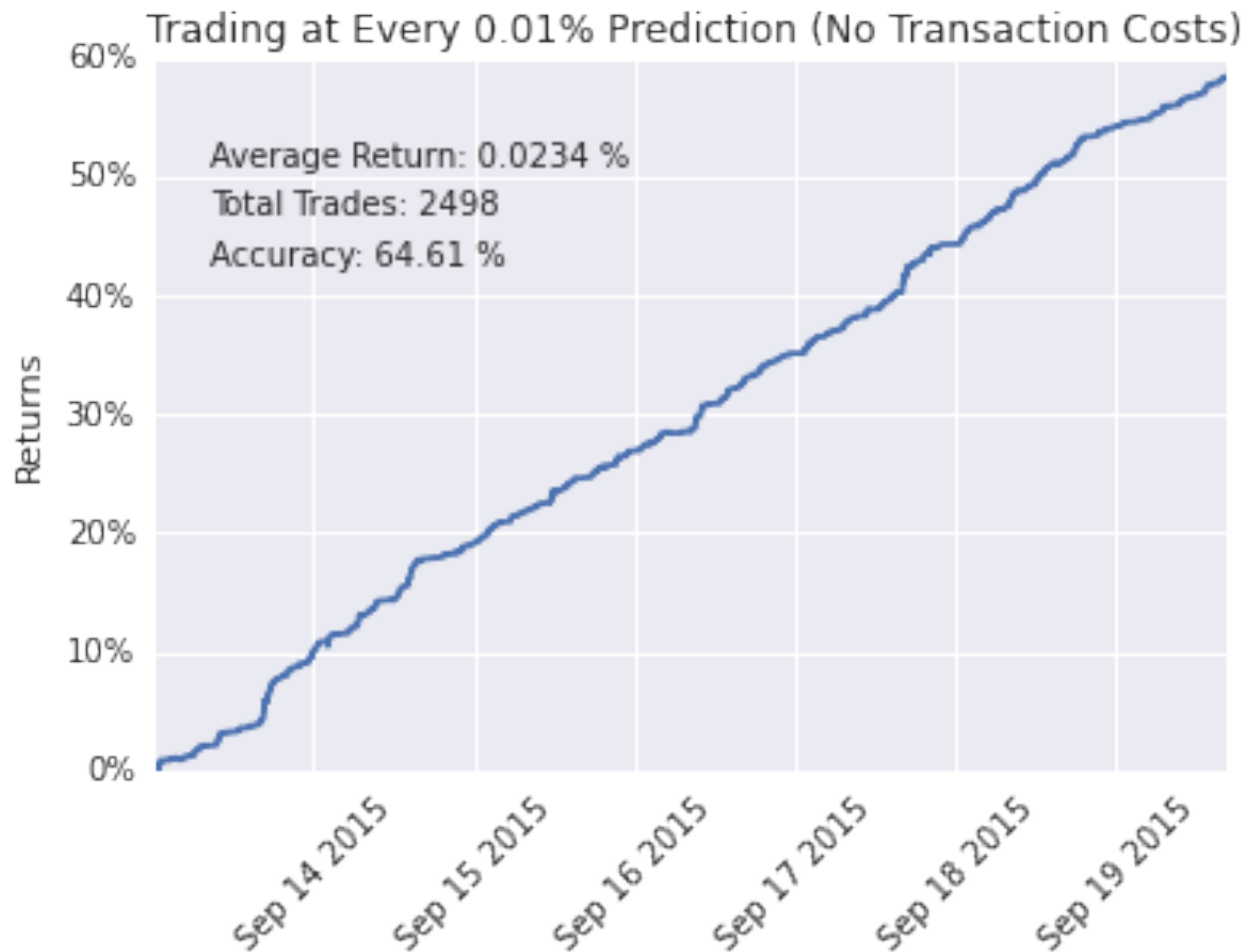# Model Validation
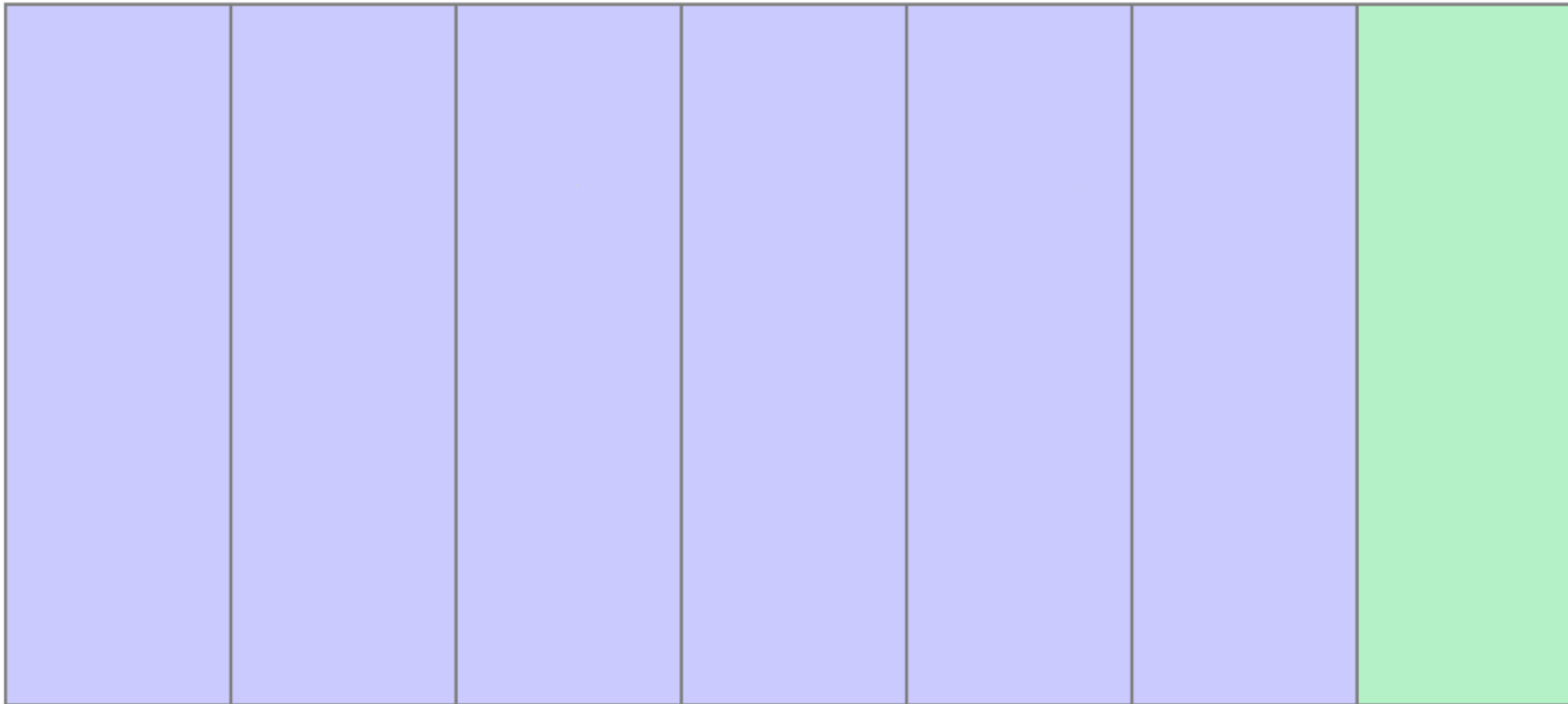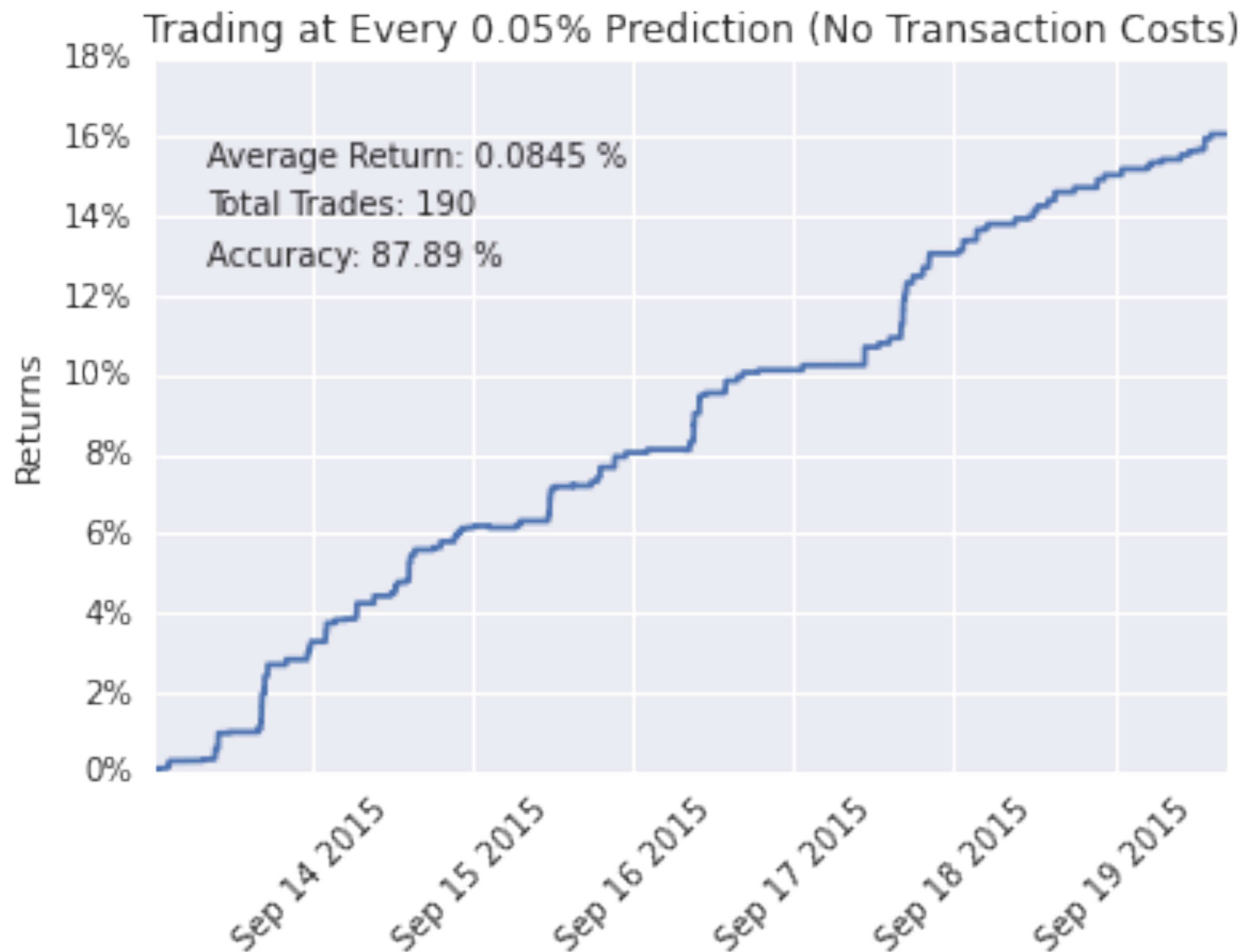
Results

	R-Squared
Linear Regression	0.0332
Random Forest	0.0744
Gradient Boosting	0.0846

Theoretical Trading Strategy

Theoretical Trading Strategy

Trading at Every 0.05% Prediction (No Transaction Costs)

Average Return: 0.0845 %
Total Trades: 190
Accuracy: 87.89 %

Next Steps

- Keep collecting data

- Incorporate findings into a trading system

bitpredict.io