



# STATISTICAL MACHINE LEARNING FOR BITCOIN PREDICTIONS

June 23rd, 2021

Presented by Haiman Wong, Yunting Chiu

# Roadmap



Problem + Motivation

01

02

THE DATA

Methodology

03

04

Results + Discussion



# 01

## Problem + Motivation

---

Employ statistical machine learning techniques and analyses to predict the direction of Bitcoin stocks and identify strategies for curating the predictions with the highest accuracies

Bitcoin 24h  
\$34,476.44 -3.36%

Ethereum 24h  
\$2,124.86 -4.26%

XRP 24h  
\$0.730449 -6.74%

Cardano 24h  
\$1.39 -0.85%

Dogecoin 24h  
\$0.272287 -5.46%

Data > Bitcoin

# Bitcoin BTC

CoinDesk 20

PRICE

24 HOUR % CHANGE

MARKET CAP

VOLUME (24H)

**\$34,476.44**

▼ **-3.36%**

**\$646.05B**

**\$32.19B**

USD ▼

Linear

Log



EXPORT DATA



06/19/2021 to 06/20/2021

1h

12h

**1d**

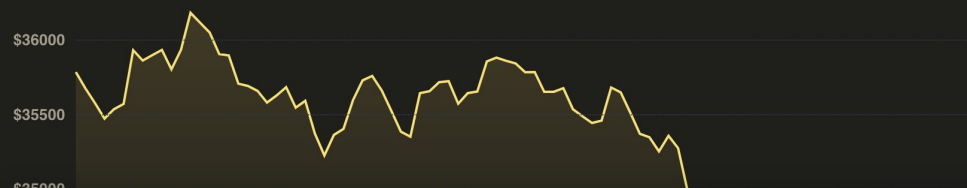
1w

1m

3m

1y

all



# CoinDesk Dataset

	Currency	Date	Closing_Price	Open_Price	24h High (USD)	24h Low (USD)	Lag1	Lag2	Lag3	Lag4	Lag5	Return_Today	Direction_Today
1	BTC	2013-10-01	123.6550	124.3047	124.7517	122.56349	NA	NA	NA	NA	NA	-0.522643318	Down
2	BTC	2013-10-02	125.4550	123.6550	125.7585	123.63383	-0.522643318	NA	NA	NA	NA	1.455671138	Up
3	BTC	2013-10-03	108.5848	125.4550	125.6657	83.32833	1.455671138	-0.522643318	NA	NA	NA	-13.447188235	Down
4	BTC	2013-10-04	118.6747	108.5848	118.6750	107.05816	-13.447188235	1.455671138	-0.522643318	NA	NA	9.292117508	Up
5	BTC	2013-10-05	121.3387	118.6747	121.9363	118.00566	9.292117508	-13.447188235	1.455671138	-0.522643318	NA	2.244792612	Up
6	BTC	2013-10-06	120.6553	121.3387	121.8522	120.55450	2.244792612	9.292117508	-13.447188235	1.455671138	-0.522643318	-0.563159343	Down
7	BTC	2013-10-07	121.7950	120.6553	121.9917	120.43199	-0.563159343	2.244792612	9.292117508	-13.447188235	1.455671138	0.944566643	Up
8	BTC	2013-10-08	123.0330	121.7950	123.6402	121.35066	0.944566643	-0.563159343	2.244792612	9.292117508	-13.447188235	1.016462088	Up
9	BTC	2013-10-09	124.0490	123.0330	124.7835	122.59266	1.016462088	0.944566643	-0.563159343	2.244792612	9.292117508	0.825794705	Up
10	BTC	2013-10-10	125.9612	124.0490	128.0168	123.81966	0.825794705	1.016462088	0.944566643	-0.563159343	2.244792612	1.541455393	Up
11	BTC	2013-10-11	125.2797	125.9612	126.4367	124.11383	1.541455393	0.825794705	1.016462088	0.944566643	-0.563159343	-0.541039794	Down
12	BTC	2013-10-12	125.9275	125.2797	126.0365	123.12966	-0.541039794	1.541455393	0.825794705	1.016462088	0.944566643	0.517115069	Up
13	BTC	2013-10-13	126.3833	125.9275	126.8965	125.27650	0.517115069	-0.541039794	1.541455393	0.825794705	1.016462088	0.361978122	Up
14	BTC	2013-10-14	135.2420	126.3833	135.2420	126.25500	0.361978122	0.517115069	-0.541039794	1.541455393	0.825794705	7.009357959	Up
15	BTC	2013-10-15	133.2033	135.2420	135.6922	130.39275	7.009357959	0.361978122	0.517115069	-0.541039794	1.541455393	-1.507416447	Down
16	BTC	2013-10-16	142.7633	133.2033	143.6048	132.94933	-1.507416447	7.009357959	0.361978122	0.517115069	-0.541039794	7.176997752	Up
17	BTC	2013-10-17	137.9233	142.7633	145.6010	136.01466	7.176997752	-1.507416447	7.009357959	0.361978122	0.517115069	-3.390226328	Down
18	BTC	2013-10-18	142.9517	137.9233	144.3110	137.53966	-3.390226328	7.176997752	-1.507416447	7.009357959	0.361978122	3.645742892	Up
19	BTC	2013-10-19	152.5518	142.9517	152.6685	142.47949	3.645742892	-3.390226328	7.176997752	-1.507416447	7.009357959	6.715675775	Up
20	BTC	2013-10-20	160.3388	152.5518	167.0800	152.51350	6.715675775	3.645742892	-3.390226328	7.176997752	-1.507416447	5.104494649	Up
21	BTC	2013-10-21	164.3150	160.3388	164.3315	159.89666	5.104494649	6.715675775	3.645742892	-3.390226328	7.176997752	2.479848456	Up
22	BTC	2013-10-22	177.6333	164.3150	177.7117	163.87149	2.479848456	5.104494649	6.715675775	3.645742892	-3.390226328	8.105371275	Up
23	BTC	2013-10-23	188.2972	177.6333	188.8238	177.11399	8.105371275	2.479848456	5.104494649	6.715675775	3.645742892	6.003282154	Up
24	BTC	2013-10-24	200.7017	188.2972	202.3433	185.81683	6.003282154	8.105371275	2.479848456	5.104494649	6.715675775	6.587725487	Up
25	BTC	2013-10-25	180.3550	200.7017	201.3332	158.72216	6.587725487	6.003282154	8.105371275	2.479848456	5.104494649	-10.137763684	Down

## 02. THE DATA

Variable	Class	Description
Currency	character	Bitcoin or BTC
Date	date	From 2013-10-01 to 2021-06-13
Closing_Price	double	The price at the end of a trading day (24hr)
Open_Price	double	The price at the beginning of a trading day (24hr)
24h High (USD)	double	The day's highest price
24h Low (USD)	double	The day's lowest price
Lag1	double	Percentage return for previous day
Lag2	double	Percentage return for 2 days previous
Lag3	double	Percentage return for 3 days previous
Lag4	double	Percentage return for 4 days previous
Lag5	double	Percentage return for 5 days previous
Return_Today	double	Percentage return for today
Return_Direction	factor	A factor with levels Down and Up indicating whether the market had a positive or negative return on a given day

# DATA DICTIONARY

## 03. Methodology

---



**DECISION TREE**



**RANDOM FOREST**



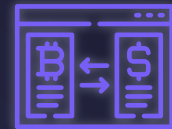
**LOGISTIC REGRESSION**



**LDA**



**QDA**



**KNN**

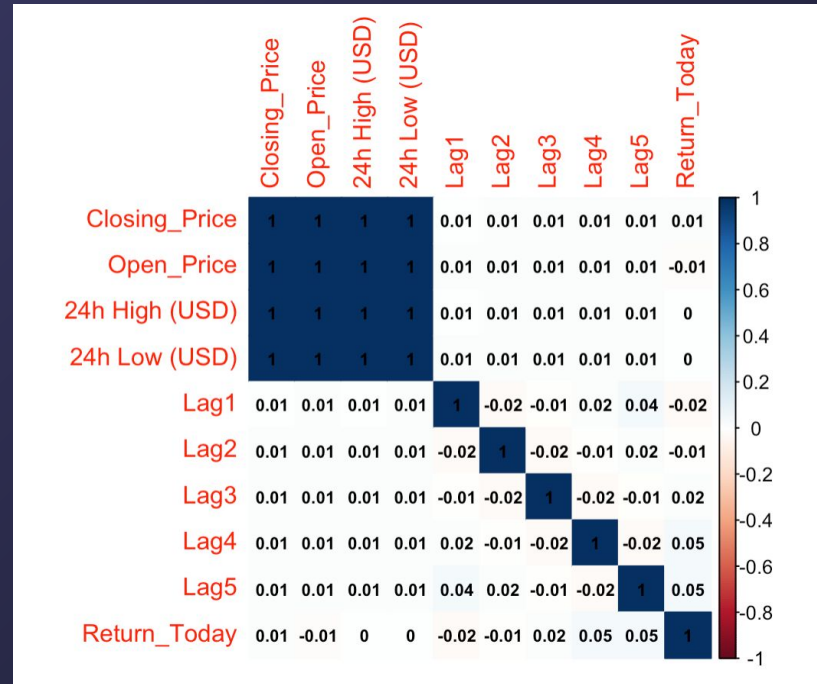
# 04. Results

Variable	Date	Closing_Price	Open_Price	24h High (USD)	24h Low (USD)	Lag1	Lag2	Lag3	Lag4	Lag5	Return_To day
Min	2013-10-06	120.7	120.7	121.8	120.4	-27.08	-27.08	-27.08	-27.08	-27.08	-27.08
1st Q.	2015-09-07	468.4	467.7	478.0	454.4	-1.32	-1.30	-1.31	-1.30	-1.30	-1.31
Median	2017-08-09	3259.1	3233.5	3343.4	3157.2	0.16	0.16	0.16	0.16	0.16	0.16
Mean	2019-08-09	6834.6	6820.8	7032.0	6588.5	0.30	0.30	0.30	0.30	0.30	0.30
3rd Q.	2019-07-12	8697.1	8694.5	8905.4	8381.3	1.96	1.96	1.96	1.96	1.96	1.96
Max.	2021-06-13	63346.8	63562.7	64801.8	62094.6	35.85	35.85	35.85	35.85	35.85	35.85



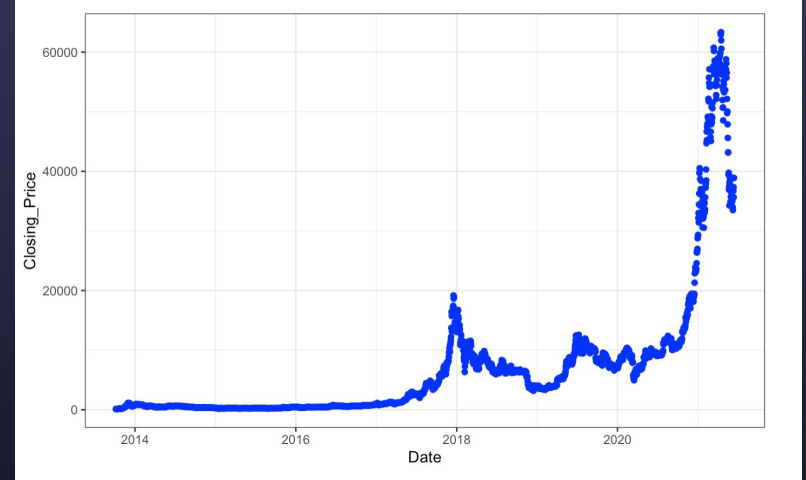
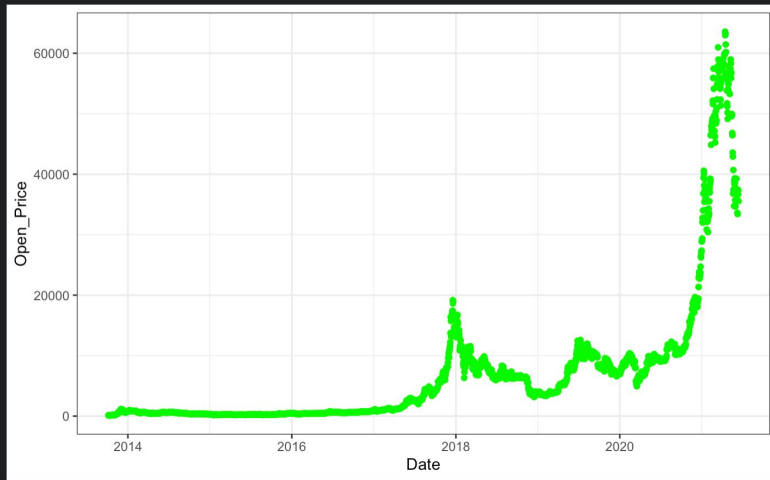
# 04. Results

Variable	Closing_Price	Open_Price	24h High (USD)	24h Low (USD)	Lag1	Lag2	Lag3	Lag4	Lag5	Return_Today
Closing_Price	1.000	0.999	0.999	0.999	0.010	0.012	0.012	0.014	0.011	0.013
Open_Price	0.999	1.000	0.999	0.999	0.011	0.010	0.012	0.013	0.011	-0.014
24h High (USD)	0.999	0.999	1.000	0.999	0.010	0.010	0.012	0.013	0.011	0.000
24 Low (USD)	0.999	0.999	0.999	1.000	0.013	0.0123	0.012	0.014	0.011	0.003
Lag1	0.010	0.011	0.010	0.013	1.000	-0.024	-0.010	0.016	0.044	-0.023
Lag2	0.012	0.010	0.010	0.012	-0.021	-1.000	-0.024	-0.010	0.017	-0.009

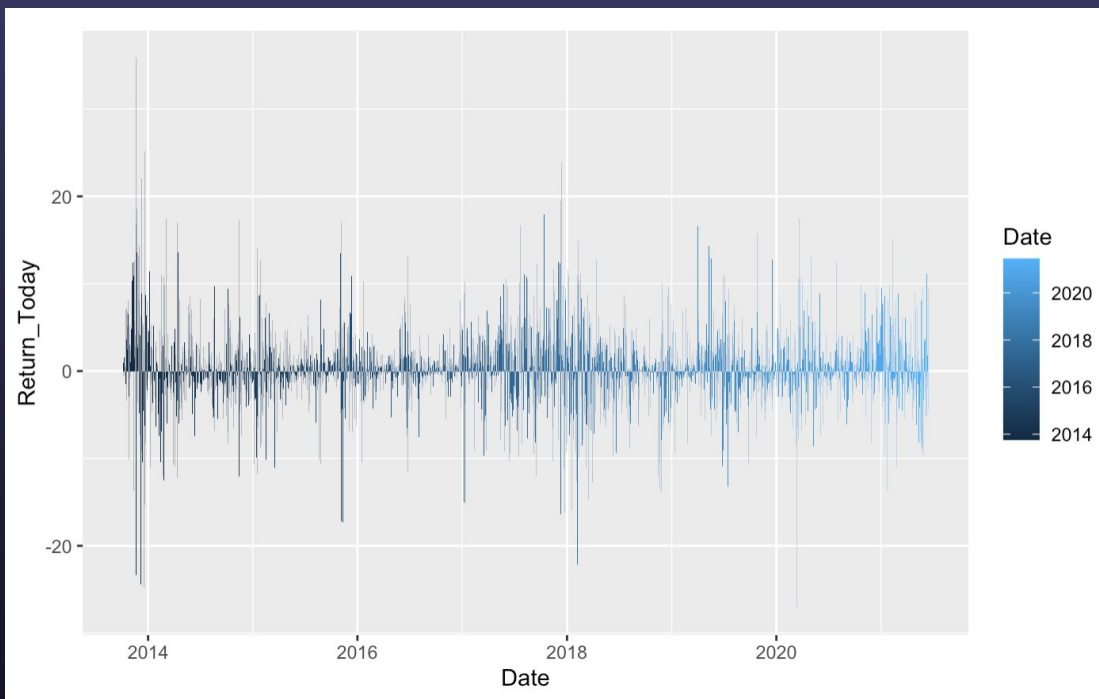


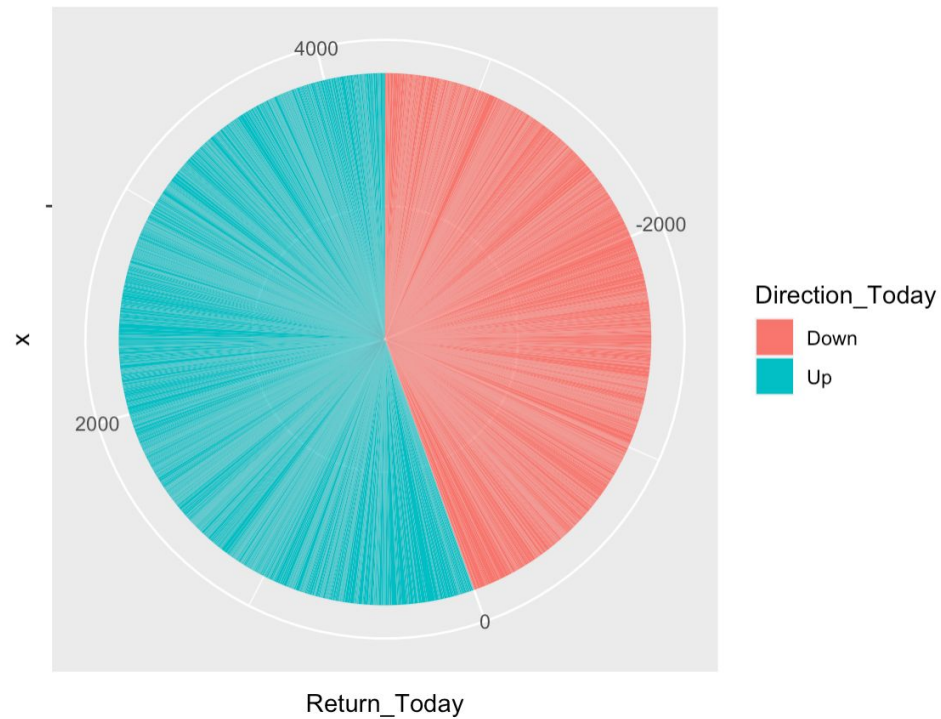
# CORRELATION PLOT

# BITCOIN TO THE MOON??



# BARPLOT FREQUENCIES





# PIE CHART

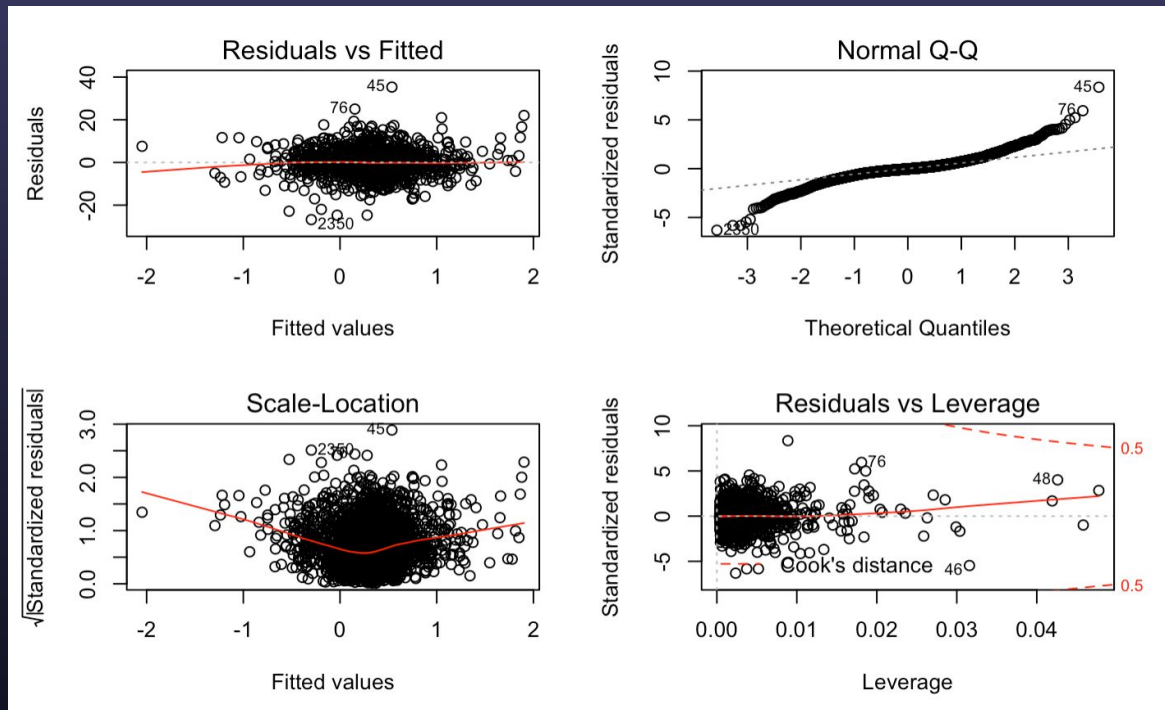
# 04. RESULTS

	Estimate	Std. Error	t-value	p-value	Residual Standard Error	Multiple R-Squared	Adjusted R-squared	F-Statistic	Overall p-value
<b>M1 Intercept</b>	0.30557	0.08052	3.795	0.000151	4.256 on 2806 degrees of freedom	0.0005078	0.0001516	1.426 on 1 and 2806 DF	0.2326
<b>M1 Lag1</b>	-0.02255	0.01889	-1.194	0.232566					
<b>M2 Intercept</b>	0.301610	0.80539	3.745	0.000184	4.257 on 2806 degrees of freedom	8.164e-05	-0.0002747	0.2291 on 1 and 2806 DF	0.6322
<b>M1 Lag2</b>	-0.009037	0.018881	-0.479	0.632238					
<b>M3 Intercept</b>	0.29437	0.08052	3.656	0.000261	4.257 on 2806 degrees of freedom	0.0002343	-0.000122	0.6576 on 1 and 2806 DF	0.4175
<b>M3 Lag3</b>	0.01528	0.01884	0.811	0.417479					
<b>M4 Intercept</b>	0.28518	0.8045	3.545	0.000399	4.253 on 2806 degrees of freedom	0.002138	0.001782	6.011 on 1 and 2806 DF	0.01428
<b>M4 Lag4</b>	0.04616	0.01883	2.452	0.014277					
<b>M5 Intercept</b>	0.28564	0.08045	3.551	0.000391	4.253 on 2806 degrees of freedom	0.002048	0.001692	5.758 on 1 and 2806 DF	0.01648
<b>M5 Lag5</b>	0.04523	0.01885	2.400	0.016482					

## 04. RESULTS

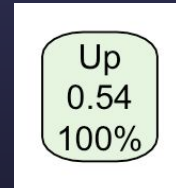
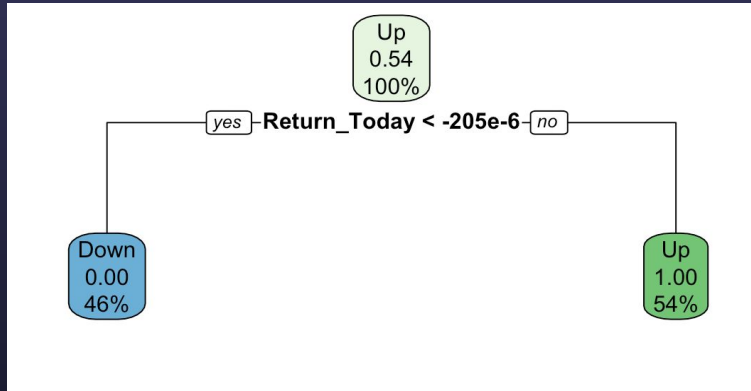
	Estimate	Std. Error	t-value	p-value	Residual Standard Error	Multiple R-Squared	Adjusted R-squared	F-Statistic	Overall p-value
<b>Intercept</b>	0.276195	0.081172	3.403	0.000677	4.249 on 2802 degrees of freedom	0.005297	0.003522	2.984 on 5 and 2802 DF	0.01084
<b>Lag1</b>	-0.025475	0.018883	-1.349	0.177409					
<b>Lag2</b>	-0.009517	0.018859	-0.505	0.613849					
<b>Lag3</b>	0.016422	0.018823	0.872	0.383055					
<b>Lag4</b>	0.047975	0.018825	2.548	0.010874					
<b>Lag5</b>	0.047803	0.018860	2.535	0.011313					

# MULTIPLE REGRESSION PLOT

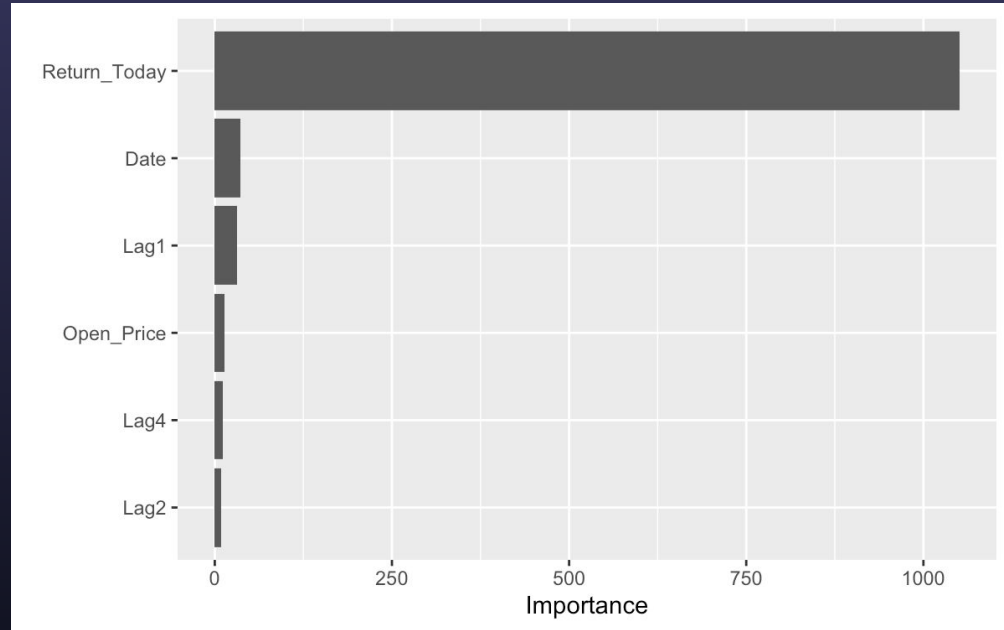




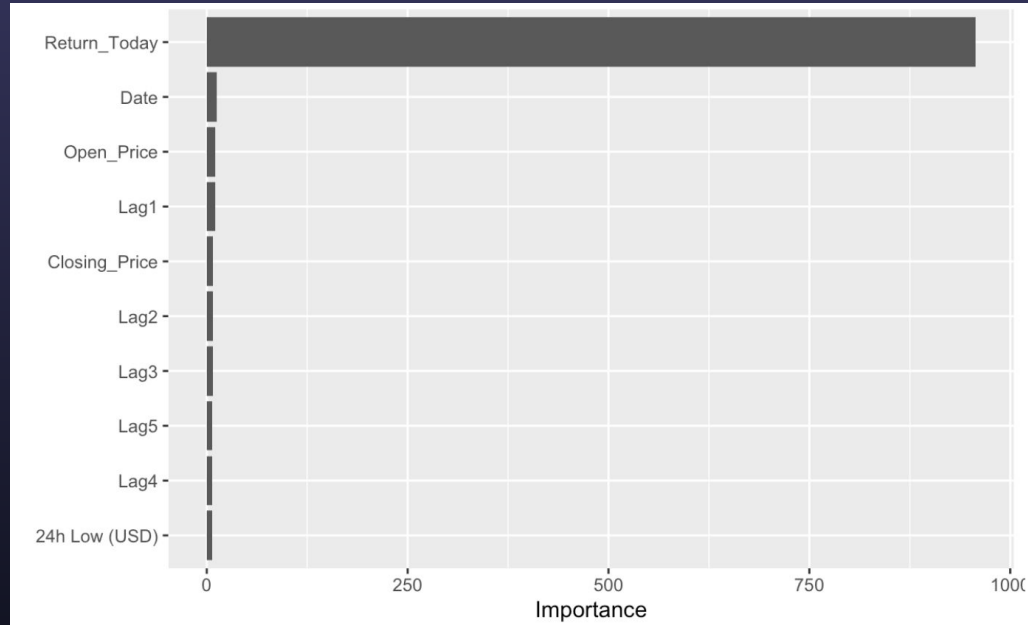
# DECISION TREE MODELS



# DECISION TREE VARIABLE IMPORTANCE



# RANDOM FOREST VARIABLE IMPORTANCE



# DATA SPLITTING

```
set.seed(1000)
BTC_split <- initial_split(BTC_tidied)
BTC_train <- training(BTC_split)
BTC_test  <- testing(BTC_split)
```

# LOGISTIC REGRESSION COEFFICIENTS TABLE

```
## # A tibble: 6 x 6
```

##	term	estimate	std.error	statistic	p.value	sign_level
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
## 1	(Intercept)	0.115	0.0444	2.59	0.00968	Yes
## 2	Lag1	-0.0273	0.0104	-2.63	0.00845	Yes
## 3	Lag2	0.0179	0.0102	1.75	0.0797	No
## 4	Lag3	0.0205	0.0103	1.98	0.0480	Yes
## 5	Lag4	-0.00550	0.0105	-0.526	0.599	No
## 6	Lag5	0.00650	0.0102	0.634	0.526	No

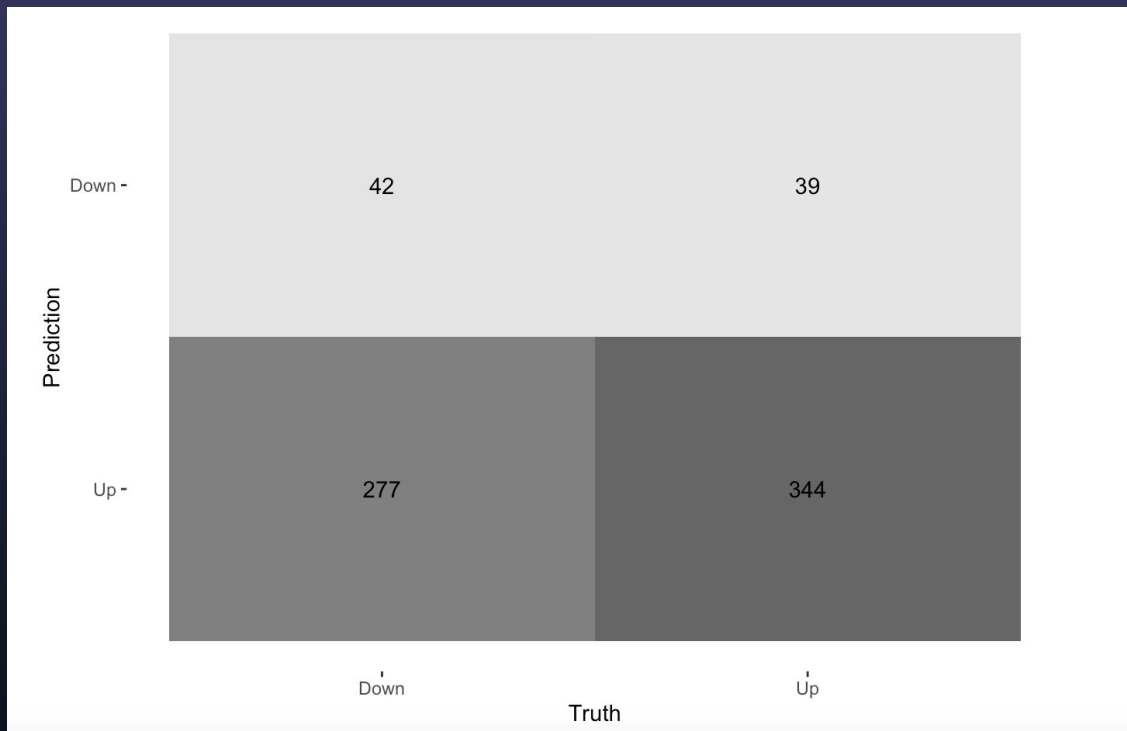
# LOGISTIC REGRESSION

## Lag1 to Lag5

Prediction	Truth	
	Down	Up
Down -	44	48
Up -	275	335

# LOGISTIC REGRESSION

## Lag1 and Lag3



# LDA MODEL

## Lag1 and Lag3





# QDA MODEL

## Lag1 and Lag3

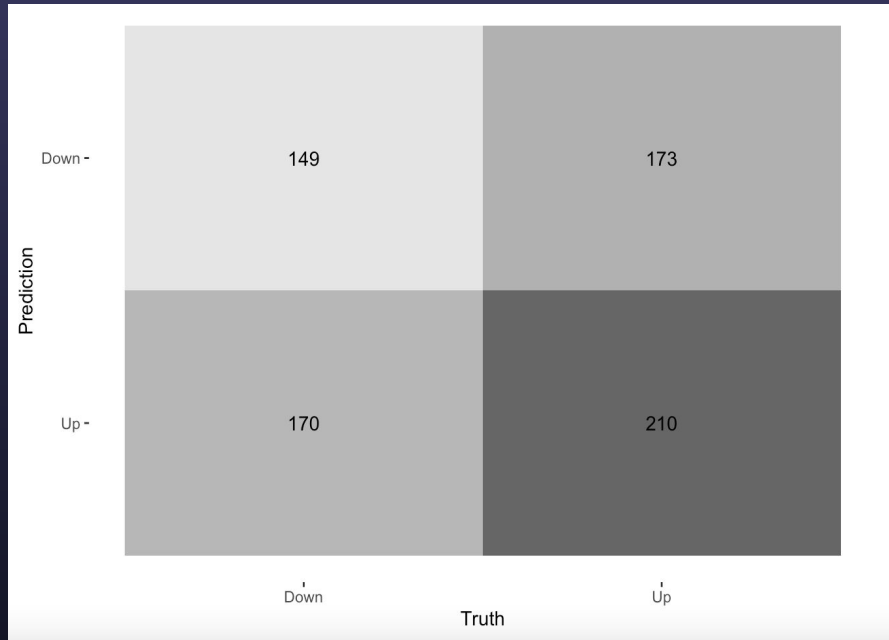
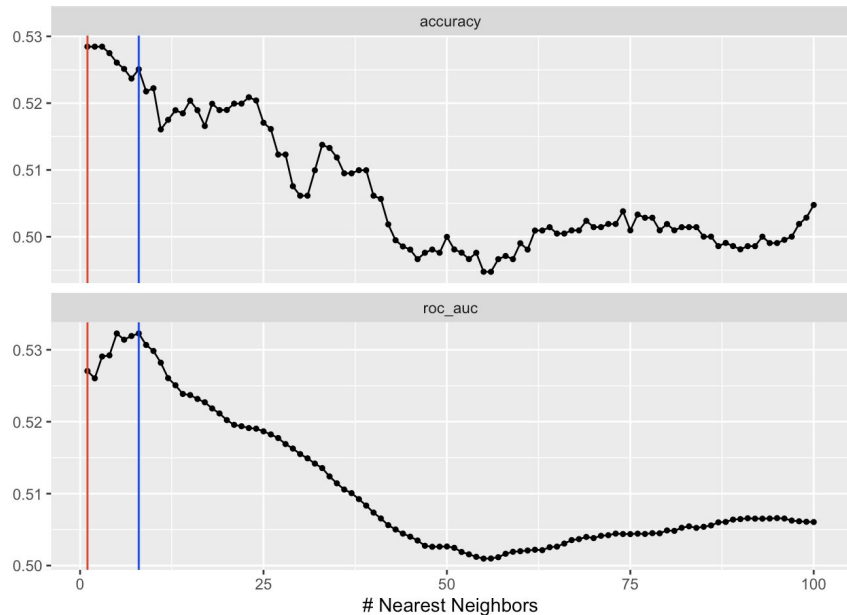
Prediction	Down	Up
	Truth	Truth
Down -	188	203
Up -	131	180

# CENTERING & SCALING in KNN

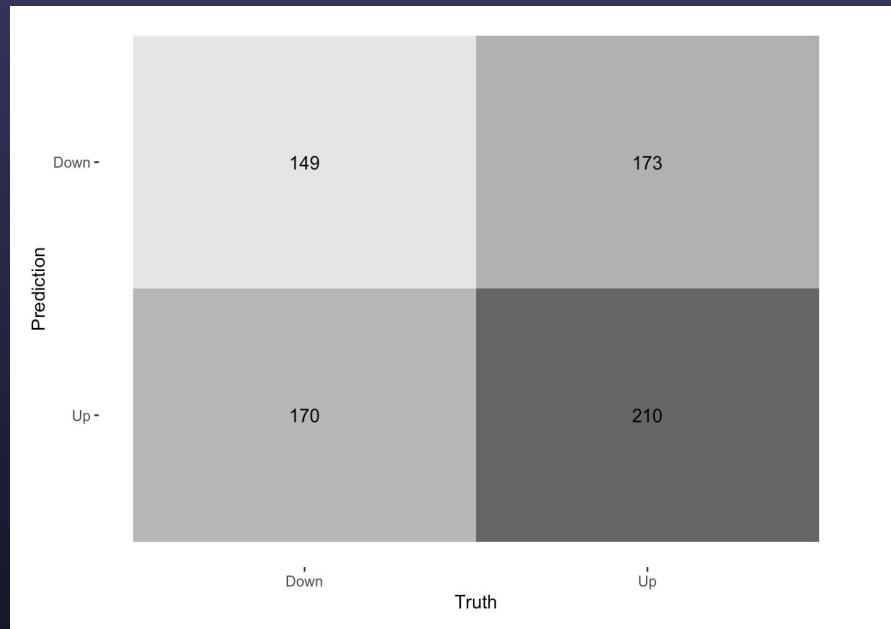
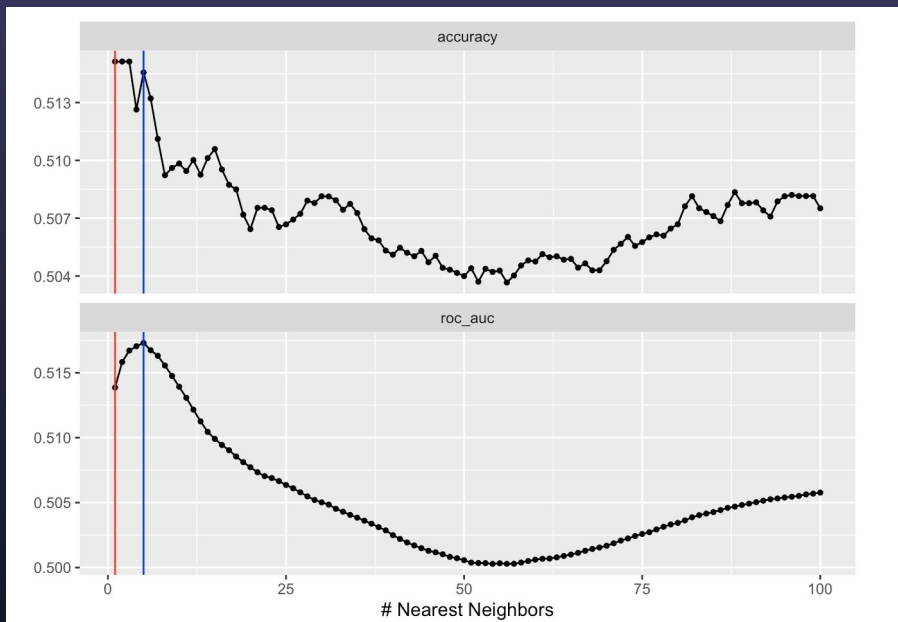
```
rec_norm <- recipe(Direction_Today ~ Lag1 + Lag3, data = BTC_train) %>%  
  step_normalize(all_numeric_predictors())  
rec_norm
```

```
## Data Recipe  
##  
## Inputs:  
##  
##      role #variables  
## outcome      1  
## predictor      2  
##  
## Operations:  
##  
## Centering and scaling for all_numeric_predictors()
```

# Cross Validation in KNN Model



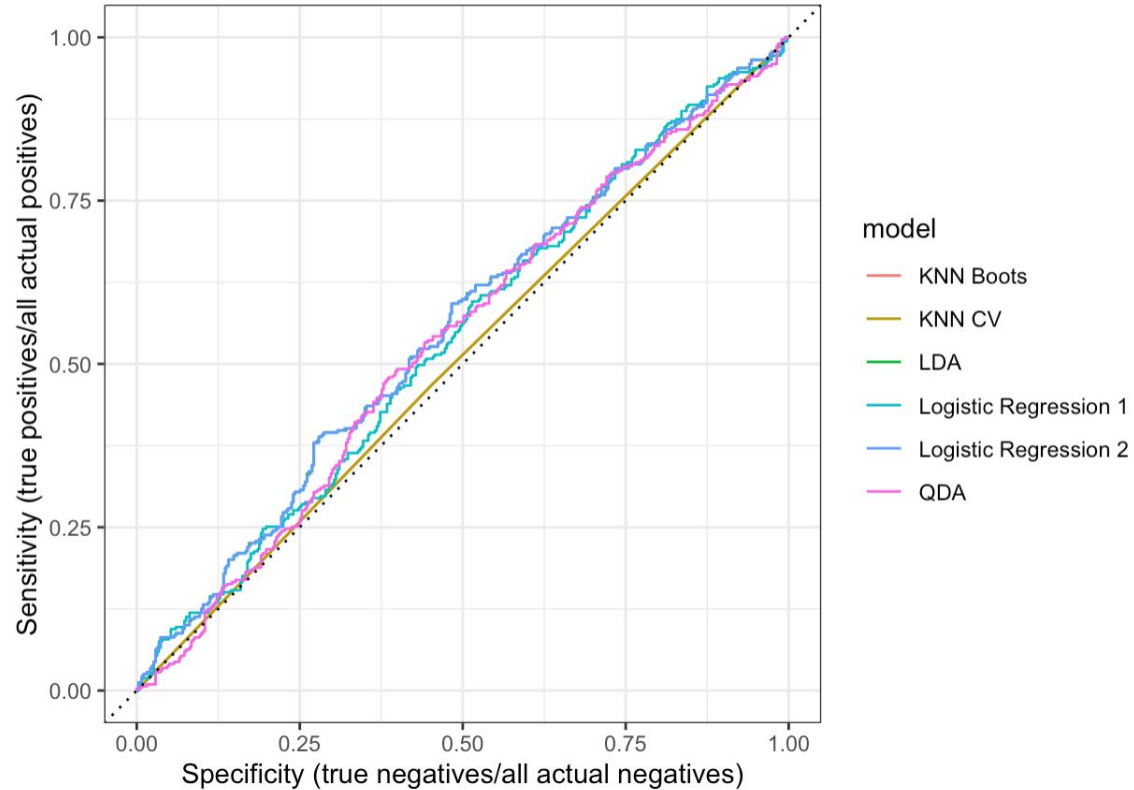
# Bootstrapping in KNN Model



## 04. COMPARATIVE RESULTS

```
## # A tibble: 6 x 4
##   .metric .estimator .estimate description
##   <chr>    <chr>         <dbl> <chr>
## 1 accuracy binary         0.550 LR Model with Lag1 and Lag3
## 2 accuracy binary         0.547 LDA Model with Lag1 and Lag3
## 3 accuracy binary         0.540 LR Model with Lag1 to Lag5
## 4 accuracy binary         0.524 QDA Model with Lag1 and Lag3
## 5 accuracy binary         0.511 KNN Model with Lag1 and Lag3 (CV)
## 6 accuracy binary         0.511 KNN Model with Lag1 and Lag3 (Bootstrap)
```

# COMPARATIVE ROC CURVES



## 05. DISCUSSION

---

Opportunities and recommendations for future work and research include:

- ❖ Research emerging cryptocurrencies and evaluate if and how existing cryptocurrency trend prediction strategies can be applied effectively
- ❖ Adopt an interdisciplinary research and experimental approach that aims to synthesize across disciplines to better understand the factors that influence investor behavior and relate these factors back to cryptocurrency trend predictions
- ❖ Extract and analyze how early cryptocurrency trends relate to the current cryptocurrency trends we observe now and identify ways this can lead our cryptocurrency trend prediction efforts in the near future

# Thank you!

## Questions?

**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

**Please keep this slide for attribution**