Predicting
Dramatic Drawdowns
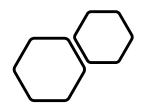# in Bitcoin

Group One

Henry Schrader, Amanullah Afzali, Justin Sparkman
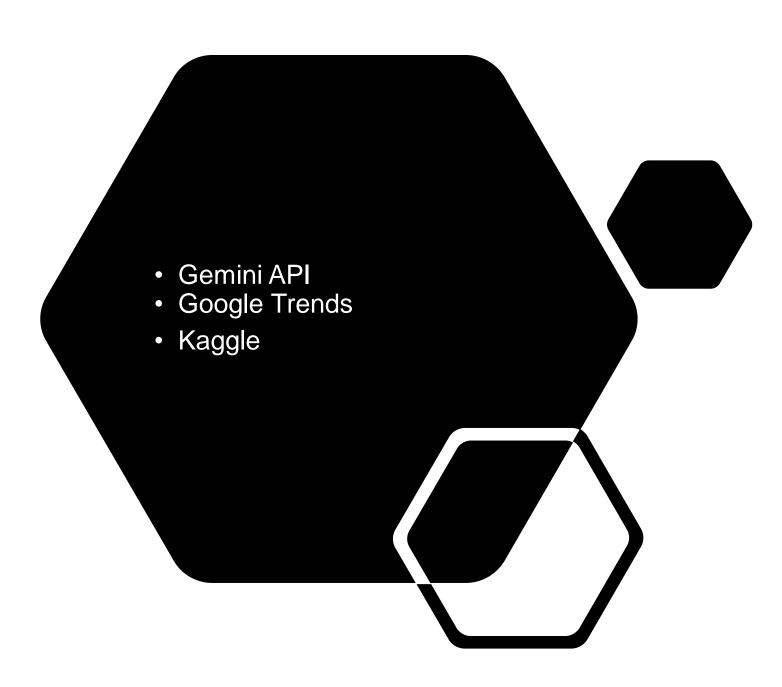
Summary

We used quantitative analysis to predict significant BTC price drawdowns exceeding 20% in a 24 hour period. We trained a LSTM Neural Network with a combination of historical price, volatility, volume and sentiment data to develop a predictive model.

Approach

- Time series analysis BTC hourly price forecasting using Hodrick-Prescott Filter and ARMA, ARIMA, GARCH models.

- 20% drawdown chosen to provide sufficient examples for deep learning neural net training.
  - BTC trades 24 / 7
  - hourly price and volume data utilized
  - 59x 20% drawdowns since 10/2016
  - 33x 20% drawdowns since 1/2017 (dataset used)

- Sentiment analysis 16million historical tweets about Bitcoin using NLP.

Data Sources

- Gemini API
- Google Trends
- Kaggle

Python Libraries

Nltk
Tensorflow
Sklearn
Keras
Hvplot
Pandas
Matplotlib
Numpy

Seaborn
Wordcloud
Holidays
Whatthelang

- Data preparation and sentiment analysis for 16million historical tweets about Bitcoin


```
tweet_sentiment_df.describe()
```
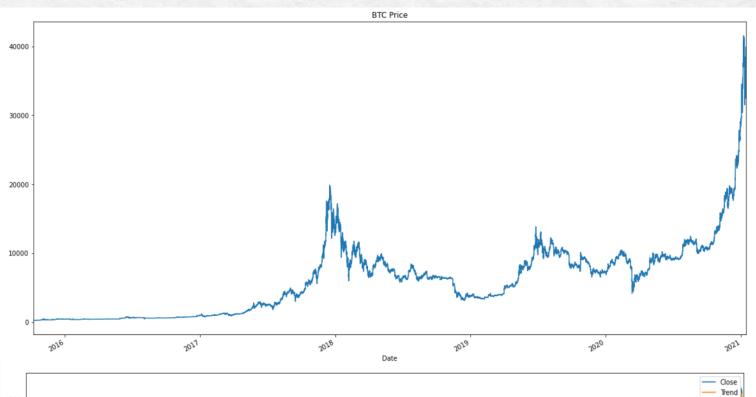
	tweet_compound	tweet_pos	tweet_neu	tweet_neg	tweet_sent
count	3.933184e+06	3.933184e+06	3.933184e+06	3.933184e+06	3.933184e+06
mean	1.295208e-01	8.934871e-02	8.779006e-01	3.243905e-02	2.297675e-01
std	3.605812e-01	1.471210e-01	1.635316e-01	8.024080e-02	6.728844e-01
min	-9.983000e-01	0.000000e+00	0.000000e+00	0.000000e+00	-1.000000e+00
25%	0.000000e+00	0.000000e+00	7.940000e-01	0.000000e+00	0.000000e+00
50%	0.000000e+00	0.000000e+00	9.420000e-01	0.000000e+00	0.000000e+00
75%	3.919000e-01	1.460000e-01	1.000000e+00	0.000000e+00	1.000000e+00
max	9.957000e-01	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00

TIME SERIES ANALYSIS AND FORECASTING

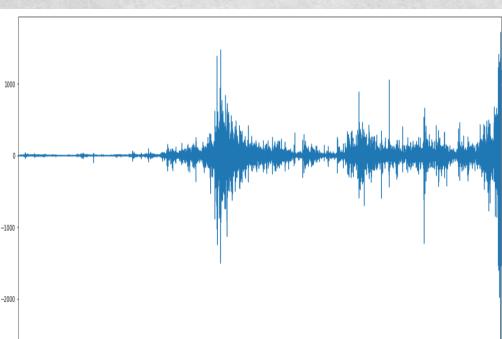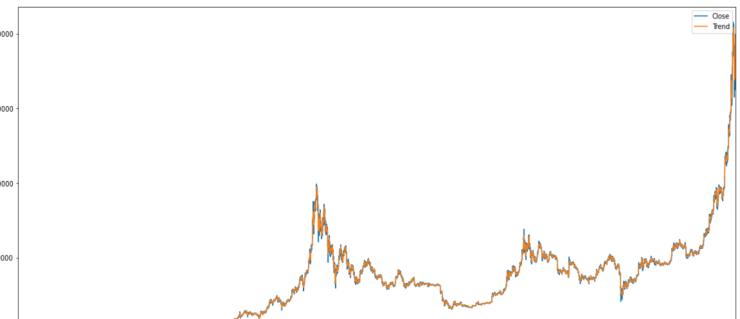Decomposing the Close price into a trend and noise using Hodric-Prescott Filter

Forecasting Returns and Close prices using ARMA and ARMA models
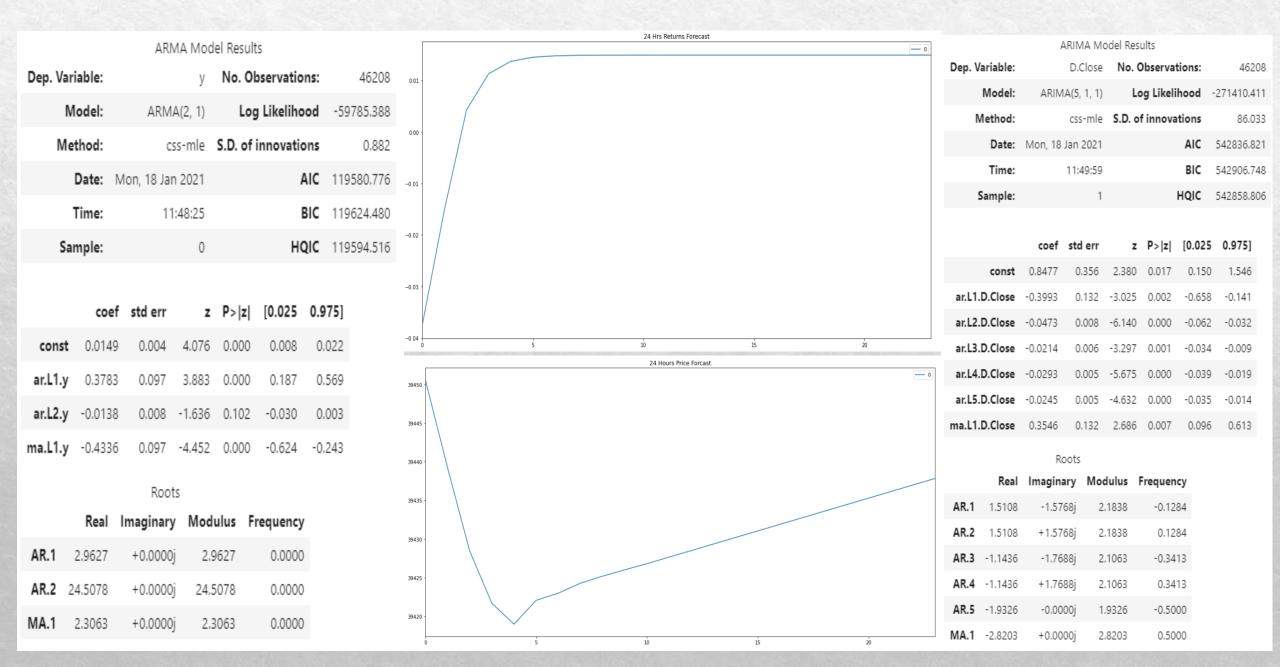
Volatility Forecasting with GARCH

Decomposing the Close price into a trend and noise using Hodric-Prescott Filter
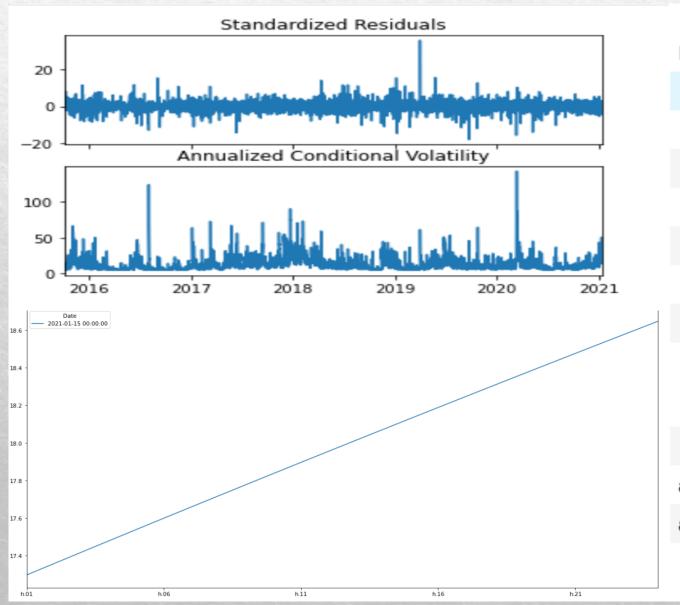
Forecasting 24 hours Returns and Prices using ARMA and ARMA models

ARMA Model Results

Dep. Variable:		y	No. Observations:		46208
Model:		ARMA(2, 1)	Log Likelihood		-59785.388
Method:		css-mle	S.D. of innovations		0.882
Date:	Mon, 18 Jan 2021		AIC		119580.776
Time:	11:48:25		BIC		119624.480
Sample:		0	HQIC		119594.516

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0149	0.004	4.076	0.000	0.008	0.022
ar.L1.y	0.3783	0.097	3.883	0.000	0.187	0.569
ar.L2.y	-0.0138	0.008	-1.636	0.102	-0.030	0.003
ma.L1.y	-0.4336	0.097	-4.452	0.000	-0.624	-0.243

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	2.9627	+0.0000j	2.9627	0.0000
AR.2	24.5078	+0.0000j	24.5078	0.0000
MA.1	2.3063	+0.0000j	2.3063	0.0000

ARIMA Model Results

Dep. Variable:		D.Close	No. Observations:		46208
Model:		ARIMA(5, 1, 1)	Log Likelihood		-271410.411
Method:		css-mle	S.D. of innovations		86.033
Date:	Mon, 18 Jan 2021		AIC		542836.821
Time:	11:49:59		BIC		542906.748
Sample:		1	HQIC		542858.806

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.8477	0.356	2.380	0.017	0.150	1.546
ar.L1.D.Close	-0.3993	0.132	-3.025	0.002	-0.658	-0.141
ar.L2.D.Close	-0.0473	0.008	-6.140	0.000	-0.062	-0.032
ar.L3.D.Close	-0.0214	0.006	-3.297	0.001	-0.034	-0.009
ar.L4.D.Close	-0.0293	0.005	-5.675	0.000	-0.039	-0.019
ar.L5.D.Close	-0.0245	0.005	-4.632	0.000	-0.035	-0.014
ma.L1.D.Close	0.3546	0.132	2.686	0.007	0.096	0.613

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.5108	-1.5768j	2.1838	-0.1284
AR.2	1.5108	+1.5768j	2.1838	0.1284
AR.3	-1.1436	-1.7688j	2.1063	-0.3413
AR.4	-1.1436	+1.7688j	2.1063	0.3413
AR.5	-1.9326	-0.0000j	1.9326	-0.5000
MA.1	-2.8203	+0.0000j	2.8203	0.5000

24 hours Volatility Forecasting with GARCH

Zero Mean - GARCH Model Results			
Dep. Variable:	Close	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.000
Vol Model:	GARCH	Log-Likelihood:	-46145.8
Distribution:	Normal	AIC:	92299.6
Method:	Maximum Likelihood	BIC:	92334.6
		No. Observations:	46208
Date:	Mon, Jan 18 2021	Df Residuals:	46204
Time:	11:50:10	Df Model:	4

Volatility Model
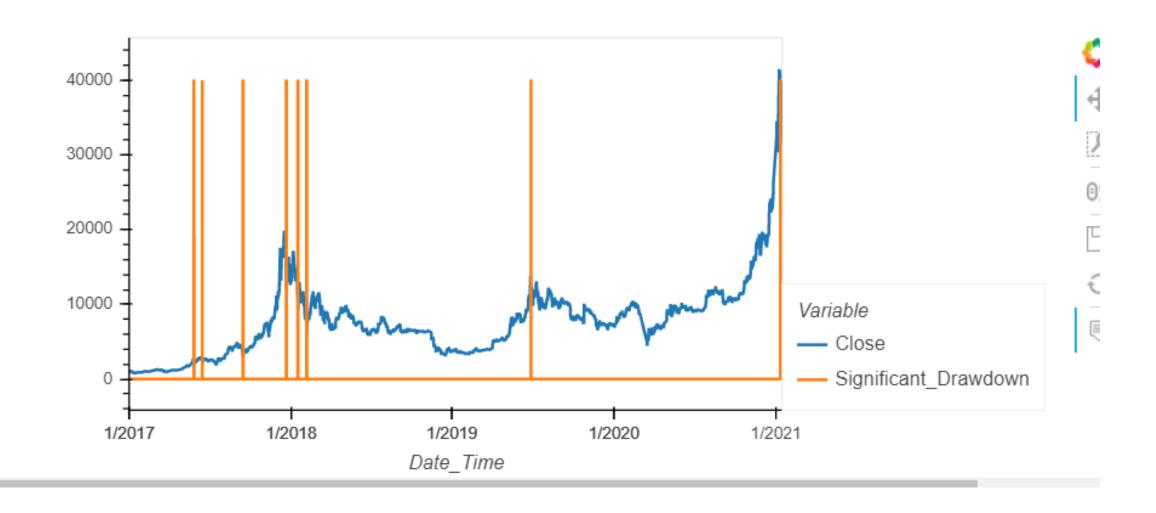
	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	8.3654e-03	5.732e-03	1.459	0.144	[-2.870e-03,1.960e-02]
alpha[1]	0.1004	1.710e-02	5.873	4.290e-09	[6.689e-02, 0.134]
alpha[2]	1.4613e-08	5.014e-02	2.914e-07	1.000	[-9.828e-02,9.828e-02]
beta[1]	0.8996	4.012e-02	22.425	2.239e-111	[0.821, 0.978]

Predict BTC 20% drawdown

- Data preparation using raw data from Gemini, Google trends, Kaggle

- Multiple features generated from price and volume data

- LSTM RNN model training and testing

- Plotting close vs actual/predicted drawdown

Predict BTC 20% drawdown

Model feature generated from price volume and google trends

```
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   Close                 35513 non-null   float64
 1   Volume                35513 non-null   float64
 2   Date                  35513 non-null   object
 3   Weekday               35513 non-null   float64
 4   Hour                  35513 non-null   float64
 5   US_Holiday            35513 non-null   float64
 6   US_Market_Open        35513 non-null   float64
 7   Trail24hr_MaxClose    35513 non-null   float64
 8   Trail12Wk_MaxClose    35513 non-null   float64
 9   Trail52Wk_MaxClose    35513 non-null   float64
 10  Trail24hr_CloseRatio  35513 non-null   float64
 11  Trail12Wk_CloseRatio  35513 non-null   float64
 12  Trail52Wk_CloseRatio  35513 non-null   float64
 13  Hr_Return             35513 non-null   float64
 14  Trail24hr_Return      35513 non-null   float64
 15  Trail24hr_Std         35513 non-null   float64
 16  Trail12Wk_Return      35513 non-null   float64
 17  Trail12Wk_Std         35513 non-null   float64
 18  Trail52Wk_Return      35513 non-null   float64
 19  Trail52Wk_Std         35513 non-null   float64
 20  Trail24hr_MaxVol      35513 non-null   float64
 21  Trail12Wk_MaxVol      35513 non-null   float64
 22  Trail52Wk_MaxVol      35513 non-null   float64
 23  Trail24hr_VolRatio    35513 non-null   float64
 24  Trail12Wk_VolRatio    35513 non-null   float64
 25  Trail52Wk_VolRatio    35513 non-null   float64
 26  Vol_PctDelta          35513 non-null   float64
 27  Bitcoin Trend         35513 non-null   float64
 28  Cryptocurrency        35513 non-null   float64
 29  crypto all time high  35513 non-null   float64
 30  crypto drop           35513 non-null   float64
 31  btc halving           35513 non-null   float64
 32  Significant_Drawdown  35513 non-null   int64
dtypes: float64(31), int64(1), object(1)
memory usage: 9.2+ MB
```

TensorFlow Keras RNN

- Utilize 30-40% dropout, 6-12 window size (0.25,0.5 day)
- Batch sizes of 1 represent 1.5-2hr/epoch
- Sigmoid activation for binary output stage
- Binary crossentropy loss
- Model run result examples:

```
# Evaluate model performance with test data

saved_model.evaluate(X_test, y_test)
```

```
333/333 [==============================] - 7s 20ms/step - loss: 0.0026 - accuracy: 0.9997
[0.0026045350823551416, 0.9997183680534363]
```

```
# Training time!
model.fit(X_train, y_train, epochs=5, shuffle = False, batch_size=12, verbose=1)

Epoch 1/5
2072/2072 [==============================] - 182s 83ms/step - loss: 0.1453 - accuracy: 0.9820
Epoch 2/5
2072/2072 [==============================] - 168s 81ms/step - loss: 0.0175 - accuracy: 0.9987
Epoch 3/5
2072/2072 [==============================] - 165s 80ms/step - loss: 0.0157 - accuracy: 0.9987
Epoch 4/5
2072/2072 [==============================] - 314s 151ms/step - loss: 0.0146 - accuracy: 0.9987
Epoch 5/5
2072/2072 [==============================] - 404s 195ms/step - loss: 0.0147 - accuracy: 0.9987
<tensorflow.python.keras.callbacks.History at 0x1d675c28dc0>
```

CHALLENGES

- Getting sufficient historical data from a trusted source.

- Having Problem scaling and testing model with more than one feature

QUESTIONS