

# CRYPTOCURRENCY AND MACHINE LEARNING

- ZEEL PATEL

- NIHARIKA DALSANIA

# AIM OF PROJECT

1. OBSERVE THE TREND OF CHANGING PRICES OF BITCOIN
2. CALCULATE STATISTICS
3. PREDICT THE FUTURE PRICES

# WHY USE RNN?

- IN TIME SERIES ANALYSIS, WE NEED TO KEEP SOME TRACK OF THE HISTORY
- REGRESSION CANNOT DO THIS
- ARTIFICIAL NEURAL NETWORK WILL NOT WORK HERE BECAUSE IT IS JUST A FEEDFORWARD NEURAL NETWORK, WHICH IS NOT A GOOD OPTION TO USE BECAUSE WE CANNOT HAVE LOOPS IN IT SO WE CANNOT FEED THE CURRENT PREDICTED PRICES BACK INTO THE MODEL TO BE USED AS HISTORY LATER
- HENCE, WE USE RECURRENT NEURAL NETWORK

# WHY ONLY LSTM?

REASON FOR SPECIFICALLY USING A SPECIAL KIND OF RNN (LSTM)

- LSTM CELL CAN REMEMBER PATTERNS FOR A CERTAIN AMOUNT OF TIME
  1. "REMEMBER" IMPORTANT INFORMATION
  2. "FORGET" IRRELEVANT INFORMATION
- LSTM UNITS INCLUDE A 'MEMORY CELL' THAT CAN MAINTAIN INFORMATION IN MEMORY FOR LONG PERIODS OF TIME (NOT POSSIBLE IN SIMPLE RNN DUE TO VANISHING GRADIENT)

# DATASET

- NEARLY 30 DIFFERENT CHARACTERISTICS OF BITCOIN ON [BITCOIN.COM](https://bitcoin.com) (LIKE PRICE, BLOCKCHAIN SIZE, MARKET CAP, ETC.)
- THIS DATA HAS BEEN COLLECTED SINCE JULY 2010.

# FEATURE REDUCTION

- OF THE SEVERAL METHODS AVAILABLE FOR FEATURE SELECTION, WE FOUND OUR DATASET TO CONTAIN THE FOLLOWING :
  1. FEATURES WITH MISSING VALUES
  2. COLLINEAR FEATURE
- SOLUTION: REPEAT PREVIOUS VALUES AND REMOVE FEATURES WITH HIGH CORRELATION. 20 OUT OF 35 FEATURES SELECTED

# THE MODEL

- A THREE-LAYER LSTM
- ADAMOPTIMIZER AS THE OPTIMIZATION FUNCTION (ADAPTIVE LEARNING RATE OPTIMISATION FUNCTION, USED SPECIALLY FOR NEURAL NETWORKS)
- MEAN SQUARED ERROR AS THE LOSS FUNCTION (MODEL WILL TRAIN ITSELF BY ATTEMPTING TO MINIMIZE THE MEAN SQUARED ERROR)
- ACTIVATION FUNCTION: LINEAR AND SIGMOID
- THE MODEL USES KERAS'S SEQUENTIAL MODEL WITH BIDIRECTIONAL LSTM LAYERS.

# THE PARAMETERS

- 3 LSTM LAYERS
- 20 FEATURES(ABOVE 95% CORRELATION. 16 FEATURES FOR ABOVE 90% CORRELATION)
- 20 % DROPOUT AT EACH LAYER IN ORDER TO REDUCE OVERFITTING
- WINDOW SIZE OF 50 DAYS
- BATCH SIZE OF 1024 WINDOWS
- 100 EPOCHS WHILE TRAINING
- A TOTAL OF 278,419 TRAINABLE PARAMETERS THROUGHOUT ALL OF ITS LAYERS



THE RESULTS...

(EXPERIMENTS TO DECIDE THE OPTIMAL PARAMETERS)

BATCH SIZE = 1024, 3 LAYERS, WINDOW SIZE = 50

<u>NO. OF FEATURES</u>	<u>NO. OF EPOCHS</u>	PRECISION(%)	MSE
16	100	56.756	0.06983
<u>20</u>	<u>100</u>	<u>63.0437</u>	<u>0.057224</u>
16	150	60.36	0.063298
20	150	58.9147	0.05639

BATCH SIZE = 1024, 3 LAYERS

<u>WINDOW SIZE</u>	PRECISION(%)	MSE
25	56	0.0178668
<u>50</u>	<u>63.0437</u>	<u>0.057224</u>
75	62.698	0.19518

# 3 LAYERS

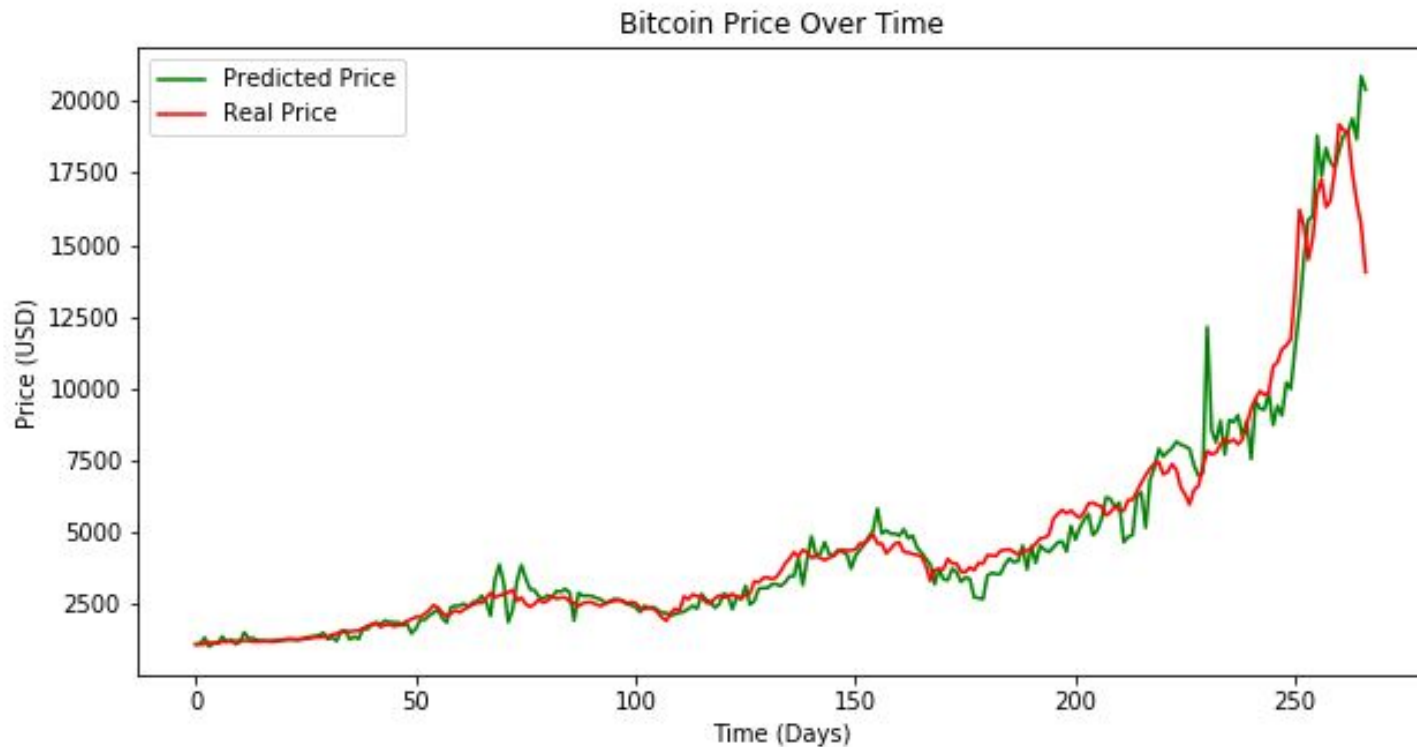
<u>BATCH SIZE</u>	PRECISION(%)	MSE
750	60.683	0.06718
<u>1024</u>	<u>63.0437</u>	<u>0.057224</u>
1250	60.4026	0.237162

<u>NO. OF LAYERS</u>	PRECISION(%)	MSE
2	62.6984	0.05638
<u>3</u>	<u>63.0437</u>	<u>0.057224</u>
4	60	0.10865

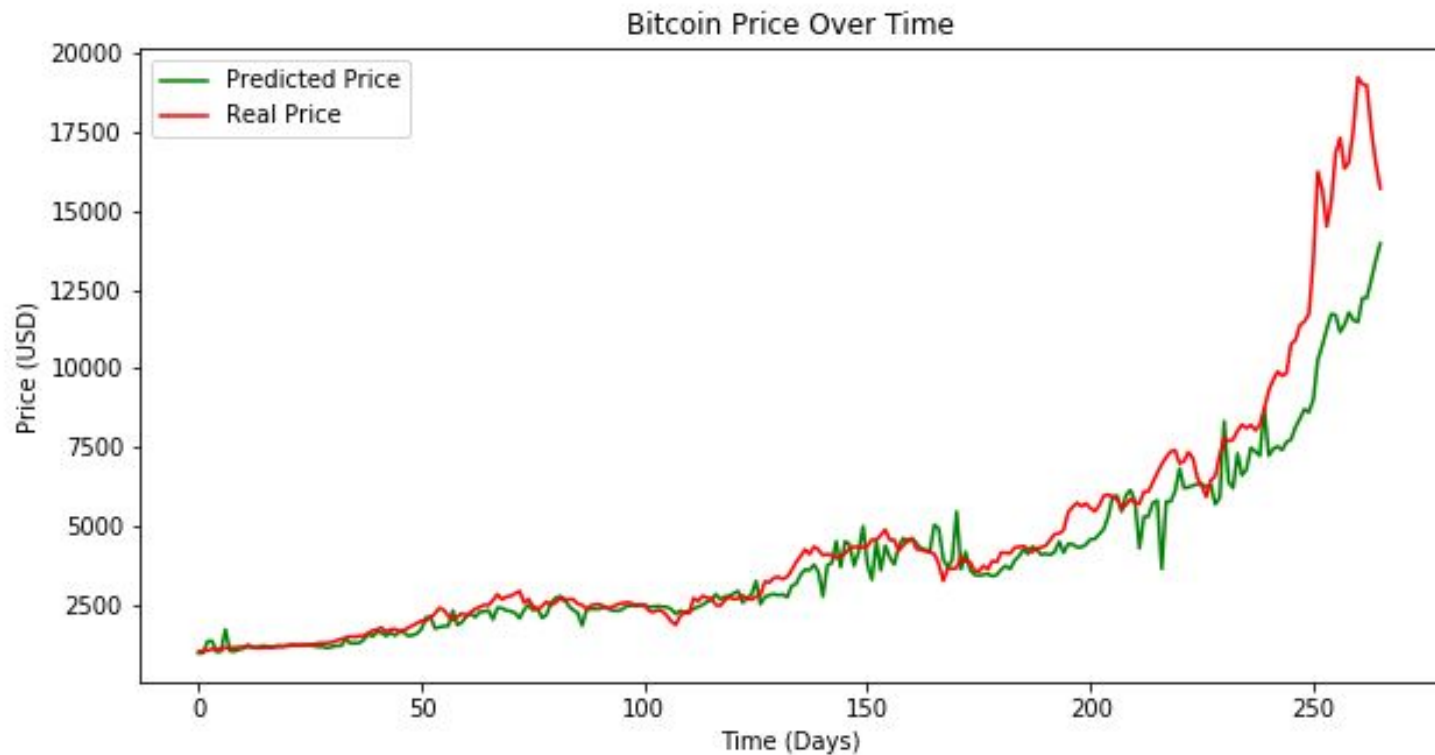
# COMPARING SIGMOID & LINEAR ACTIVATION FUNCTION

ACTIVATION FUNCTION	PRECISION(%)	MSE
LINEAR	63.0437	0.057224
SIGMOID	67.647	0.09616

# THE FINAL PREDICTION WITH DECIDED PARAMETER VALUES



# PREDICTION WITH SIGMOID ACTIVATION FUNCTION





# CONFUSION MATRICES

	ACTUAL YES	ACTUAL NO
PREDICTED YES	46	22
PREDICTED NO	122	77

SIGMOID ACTIVATION

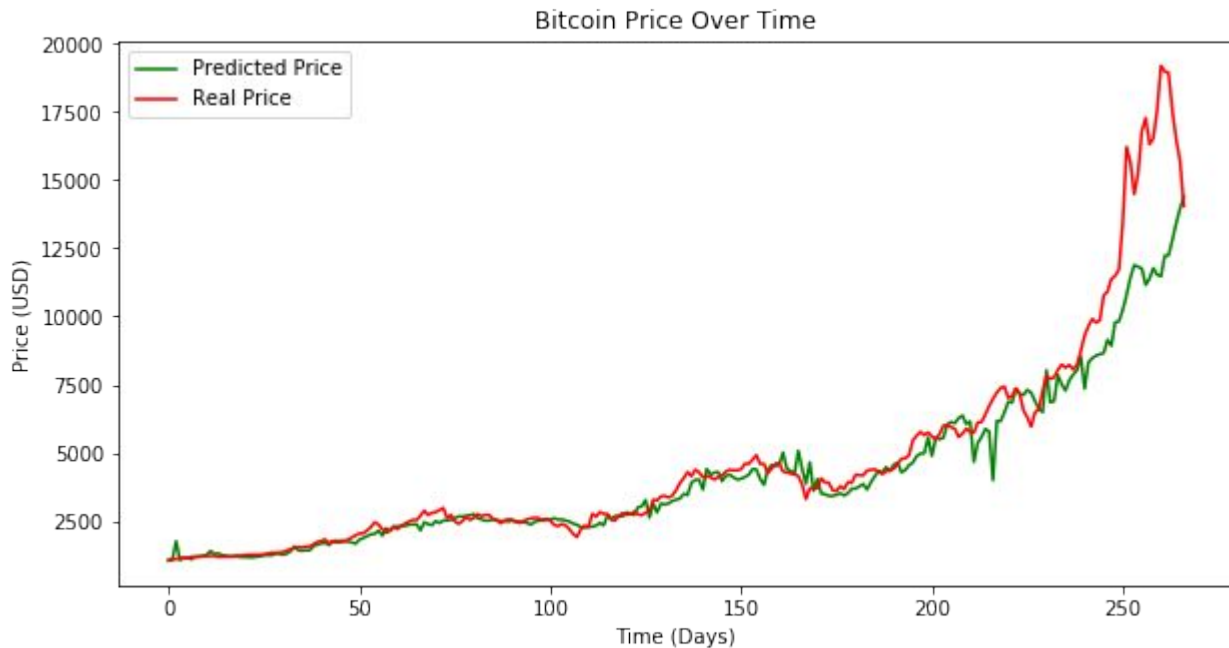
	ACTUAL YES	ACTUAL NO
PREDICTED YES	70	48
PREDICTED NO	98	51

LINEAR ACTIVATION

# WHY LINEAR ACTIVATION BETTER THAN SIGMOID?

- EVEN THOUGH THE PRECISION WAS BETTER WHEN WE USED SIGMOID FUNCTION, WE STILL SAY THE LINEAR FUNCTION IS A BETTER CHOICE
- THIS IS BECAUSE PRECISION IS CALCULATED BY JUST USING THE TRUE AND FALSE POSITIVES, BUT OUR TARGET IS NOT JUST TO PREDICT THE INCREASE OR DECREASE, BUT TO PREDICT THE ACTUAL BITCOIN PRICES
- SIGMOID MAKES LESSER ERRORS WHILE DECIDING INCREASE OR DECREASE, BUT IT DOES NOT GIVE A GOOD PREDICTION OF THE ACTUAL VALUE AS SEEN FROM TREND IN GRAPH
- HENCE LINEAR FUNCTION IS BETTER

# ACTIVATION FUNCTION = TANH



PRECISION : 60.97%

MSE : 0.071039