In []: #Installing/Updating libraries pip install numpy pip install pandas pip install pyplot pip install seaborn pip install sklearn pip install keras pip install tensorflow Bitcoin forecast using LSTM In [33]: #Data manipulation libraries import numpy as np import pandas as pd #Data visualisation libraries from matplotlib import pyplot as plt import seaborn as sns Now, pulling Bitcoin data from Yahoo finance(Historical Data) In [34]: #only 370 instances bitcoinPath2= "https://raw.githubusercontent.com/DamianWasTaken/Python/main/BTC-USD.csv" #Used an user's Github BTC-USD upload instead of downloading and uploading it again from 1/10/2013 to 17/04/202 bitcoinPath = "https://raw.githubusercontent.com/Milisha-gupta/Status-Neo/main/BTC USD 2013-10-01 2021-04-17-Co btcData = pd.read csv(bitcoinPath) btcDataSorted = btcData.sort values('Date') #confirming date btcDataSorted.head() Out[34]: Currency Date Closing Price (USD) 24h Open (USD) 24h High (USD) 24h Low (USD) 0 BTC 2013-10-01 123.65499 124.30466 124.75166 122.56349 BTC 2013-10-02 125,45500 125.75850 1 123.65499 123.63383 83.32833 2 BTC 2013-10-03 108.58483 125.45500 125.66566 108.58483 3 BTC 2013-10-04 118 67466 118.67500 107.05816 118,67466 121.93633 4 BTC 2013-10-05 121.33866 118.00566 #confirming end date btcDataSorted.tail() In [35]: #Closing price will be the target of the prediction, as such it needs to be separated from the rest of the data btcPrice = btcDataSorted[['Closing Price (USD)']] #Setting the plot's parameters and plotting price plt.plot(figsize = (15,9))plt.plot(btcPrice) plt.xticks(range(0, btcDataSorted.shape[0], 50), btcDataSorted['Date'].loc[::50], rotation=45) plt.title('Bitcoin Price', fontsize=18, fontweight='bold') plt.xlabel('Date', fontsize=18) plt.ylabel('Close price(USD)', fontsize=18) plt.show() **Bitcoin Price** 60000 50000 40000 30000 Close 20000 10000 Date In [6]: #Looking at the instances and data type btcPrice.info() <class 'pandas.core.frame.DataFrame'> Int64Index: 2755 entries, 0 to 2754 Data columns (total 1 columns): # Column Non-Null Count Dtype -----______ O Closing Price (USD) 2755 non-null float64 dtypes: float64(1) memory usage: 43.0 KB **Preparing Data** In [36]: #Normalising data, essentially putting all variables on the same scale #so that certain disproportionate variable's values don't influence the weights too much from sklearn.preprocessing import MinMaxScaler minMax = MinMaxScaler() #default value for scaling is 0-10, scaling should be adjusted depending on the activat normalisedData = minMax.fit transform(btcPrice.values) normalisedData array([[2.38307842e-04], Out[36]: [2.66771807e-04], [0.00000000e+00], [9.93354779e-01], [1.00000000e+00], [9.78161833e-01]]) Splitting data into training and testing, this needs to be done so as to evaluate the results obtained from testing the network that was trained on the training data set by feeding it data that it has never seen before, a much more effective and reliable way of splitting the data for this is bootstraping, the % depends entirely on how much data you have, if it's hundreds of thousands to millions of instances, apply a ratio of about 2-5% Testing and 95-98% Training, if less around 20-80% should do the trick, obviously this is dependent on research question and data at hand In [49]: def univariate data(dataset, start index, end index, history size, target size): data = [] labels = []start index = start index + history size if end index is None: end index = len(dataset) - target size for i in range(start index, end index): indices = range(i-history size, i) data.append(np.reshape(dataset[indices], (history size, 1))) labels.append(dataset[i + target size]) return np.array(data), np.array(labels) #These values represent days, essentailly we're using 5 days #of time searies events to learn paterns in the data and predict the furute target past history = 5 future target = 0 #80-20% ratio split trainSplit = int(len(normalisedData)* 0.8) #Using univariate data function to split the data into Train and Test x Train, y Train = univariate data(normalisedData, trainSplit, past history, future target) x Test, y Test = univariate data(normalisedData, trainSplit, None, past history, future target) LSTM Model Building This is where we can do adjustments that will have the major impacts in efficiency, by adjusting the Hyper parameters, whilst there are a few theories into hyper parameters selection, it's highly dependent on the Data and use case, as such deciding empirically usually brings good results In [50]: import keras from keras.models import Sequential from tensorflow.keras.optimizers import Adam from keras.layers import Dense, LSTM, LeakyReLU, Dropout #set the number or neurons, learning rate, function, optimiser function, loss function, batch and epochs, essen num units = 64learning Rate = 0.0001 activation function = 'sigmoid' adam = Adam(learning_rate=learning_Rate) loss function = 'mse' batch size = 5num epochs = 250#Calling on sequential and putting everything together model = Sequential() model.add(LSTM(units = num units, activation = activation function, input shape = (None, 1))) model.add(LeakyReLU(alpha = 0.5)) model.add(Dropout(0.1)) model.add(Dense(units = 1)) model.compile(optimizer = adam, loss=loss function) In [16]: model.summary() Model: "sequential 4" Layer (type) Output Shape Param # 16896 1stm 4 (LSTM) (None, 64) leaky re lu 1 (LeakyReLU) 0 (None, 64) dropout 1 (Dropout) (None, 64) 0 (None, 1) dense 1 (Dense) Total params: 16,961 Trainable params: 16,961 Non-trainable params: 0 Training model In [51]: history = model.fit(x Train, y Train, validation split = 0.1, batch size = batch size, epochs = num epochs, shuffle=False Epoch 1/250 Epoch 2/250 Epoch 3/250 Epoch 4/250 Epoch 5/250 Epoch 6/250 Epoch 7/250 Epoch 8/250 Epoch 9/250 Epoch 10/250 Epoch 11/250 Epoch 12/250 Epoch 13/250 Epoch 14/250 Epoch 15/250 Epoch 16/250 Epoch 17/250 Epoch 18/250 Epoch 19/250 Epoch 20/250 Epoch 21/250 Epoch 22/250 Epoch 23/250 Epoch 24/250 Epoch 25/250 Epoch 26/250 Epoch 27/250 Epoch 28/250 Epoch 29/250 Epoch 30/250 Epoch 31/250 Epoch 32/250 Epoch 33/250 Epoch 34/250 Epoch 35/250 Epoch 36/250 Epoch 37/250 Epoch 38/250 Epoch 39/250 Epoch 40/250 Epoch 41/250 Epoch 42/250 Epoch 43/250 Epoch 44/250 Epoch 45/250 Epoch 46/250 Epoch 47/250 Epoch 48/250 Epoch 49/250 Epoch 50/250 Epoch 51/250 Epoch 52/250 Epoch 53/250 Epoch 54/250 Epoch 55/250 Epoch 56/250 Epoch 57/250 Epoch 58/250 Epoch 59/250 Epoch 60/250 Epoch 61/250 Epoch 62/250 Epoch 63/250 Epoch 64/250 Epoch 65/250 Epoch 66/250 Epoch 67/250 s: 0.0055 Epoch 68/250 Epoch 69/250 Epoch 70/250 Epoch 71/250 Epoch 72/250 Epoch 73/250 Epoch 74/250 Epoch 75/250 Epoch 76/250 Epoch 77/250 Epoch 78/250 Epoch 79/250 Epoch 80/250 Epoch 81/250 Epoch 82/250 Epoch 83/250 Epoch 84/250 Epoch 85/250 Epoch 86/250 Epoch 87/250 Epoch 88/250 Epoch 89/250 Epoch 90/250 Epoch 91/250 Epoch 92/250 Epoch 93/250 Epoch 94/250 Epoch 95/250 Epoch 96/250 Epoch 97/250 Epoch 98/250 Epoch 99/250 Epoch 100/250 Epoch 101/250 Epoch 102/250 Epoch 103/250 Epoch 104/250 Epoch 105/250 Epoch 106/250 Epoch 107/250 Epoch 108/250 Epoch 109/250 Epoch 110/250 Epoch 111/250 Epoch 112/250 Epoch 113/250 Epoch 114/250 Epoch 115/250 Epoch 116/250 Epoch 117/250 Epoch 118/250 Epoch 119/250 Epoch 120/250 Epoch 121/250 Epoch 122/250 Epoch 123/250 Epoch 124/250 Epoch 125/250 Epoch 126/250 Epoch 127/250 Epoch 128/250 Epoch 129/250 Epoch 130/250 Epoch 131/250 Epoch 132/250 Epoch 133/250 Epoch 134/250 Epoch 135/250 Epoch 136/250 Epoch 137/250 Epoch 138/250 Epoch 139/250 Epoch 140/250 Epoch 141/250 Epoch 142/250 Epoch 143/250 Epoch 144/250 Epoch 145/250 Epoch 146/250 Epoch 147/250 Epoch 148/250 Epoch 149/250 Epoch 150/250 Epoch 151/250 Epoch 152/250 Epoch 153/250 Epoch 154/250 Epoch 155/250 Epoch 156/250 Epoch 157/250 Epoch 158/250 Epoch 159/250 Epoch 160/250 Epoch 161/250 Epoch 162/250 Epoch 163/250 Epoch 164/250 Epoch 165/250 Epoch 166/250 Epoch 167/250 Epoch 168/250 Epoch 169/250 Epoch 170/250 Epoch 171/250 Epoch 172/250 Epoch 173/250 Epoch 174/250 Epoch 175/250 Epoch 176/250 Epoch 177/250 Epoch 178/250 Epoch 179/250 Epoch 180/250 Epoch 181/250 Epoch 182/250 Epoch 183/250 Epoch 184/250 Epoch 185/250 396/396 [==: Epoch 186/250 Epoch 187/250 Epoch 188/250 Epoch 189/250 Epoch 190/250 Epoch 191/250 Epoch 192/250 Epoch 193/250 Epoch 194/250 Epoch 195/250 Epoch 196/250 Epoch 197/250 Epoch 198/250 Epoch 199/250 Epoch 200/250 Epoch 201/250 Epoch 202/250 Epoch 203/250 Epoch 204/250 Epoch 205/250 Epoch 206/250 Epoch 207/250 Epoch 208/250 Epoch 209/250 Epoch 210/250 Epoch 211/250 Epoch 212/250 Epoch 213/250 Epoch 214/250 Epoch 215/250 Epoch 216/250 Epoch 217/250 Epoch 218/250 Epoch 219/250 Epoch 220/250 Epoch 221/250 Epoch 222/250 Epoch 223/250 Epoch 224/250 Epoch 225/250 Epoch 226/250 Epoch 227/250 Epoch 228/250 Epoch 229/250 Epoch 230/250 Epoch 231/250 Epoch 232/250 Epoch 233/250 Epoch 234/250 Epoch 235/250 Epoch 236/250 Epoch 237/250 Epoch 238/250 Epoch 239/250 Epoch 240/250 Epoch 241/250 Epoch 242/250 Epoch 243/250 Epoch 244/250 Epoch 245/250 Epoch 246/250 Epoch 247/250 Epoch 248/250 Epoch 249/250 Epoch 250/250 In [53]: loss = history.history['loss'] val_loss = history.history['val_loss'] epochs = range(len(loss)) plt.figure(figsize = (15,9))plt.plot(epochs, loss, 'b', label = 'Training Loss') plt.plot(epochs, val_loss, 'r', label = 'Validation Loss') plt.title("Training and Validation Loss") plt.legend() plt.show() Training and Validation Loss Training Loss Validation Loss 0.04 0.03 0.02 0.01 0.00 Ò 50 100 150 200 250 In [58]: original = pd.DataFrame(minMax.inverse_transform(y_Test)) predictions = pd.DataFrame(minMax.inverse_transform(model.predict(x_Test))) sns.set(rc={'figure.figsize':(11.7+2,8.27+2)}) ax = sns.lineplot(x=original.index, y=original[0], label='Test Data', color='royalblue') ax = sns.lineplot(x=predictions.index, y=predictions[0], label='predictions', color='tomato') ax.set_title('Bitcoins Price', size = 14, fontweight='bold') ax.set_xlabel('Days', size=14) ax.set_ylabel('Cost (USD)', size = 14) ax.set_xticklabels('', size=10) [Text(-100.0, 0, ''),Out[58]: Text(0.0, 0, ''), Text(100.0, 0, ''), Text(200.0, 0, ''), Text(300.0, 0, ''), Text(400.0, 0, ''), Text(500.0, 0, ''), Text(600.0, 0, '')] Bitcoins Price Test Data predictions 60000 50000 40000 Cost (USD) 20000 10000 Days