

Hung Lam, Phung Ly



10000

- A method standing out for its prediction speed and accuracy,
 particularly with large and complex datasets
- From Kaggle competitions to machine learning solutions for business, this algorithm has produced the best results.
- Gradient boost algorithm helps us minimize bias error of the model



Why preprocessing data importance need to do first?

- Convert raw data into suitable form
- Store raw data in the file -> refine or preprocessing data before
 passing to the machine learning model for the training purpose so
 that the machine learning can understand data and give better
 performance



10000

Dataset Example

btc_usd

Date	Open	High	Low	Close	Adj Close	Volume
2014-09-17	465.864014	468.174011	452.421997	457.334015	457.334015	21056800
2014-09-18	456.859985	456.859985	413.104004	424.440002	424.440002	34483200
2014-09-19	424.102997	427.834991	384.532013	394.795990	394.795990	37919700
2014-09-20	394.673004	423.295990	389.882996	408.903992	408.903992	36863600
2014-09-21	408.084991	412.425995	393.181000	398.821014	398.821014	26580100
2014-09-22	399.100006	406.915985	397.130005	402.152008	402.152008	24127600
2014-09-23	402.092010	441.557007	396.196991	435.790985	435.790985	45099500
2014-09-24	435.751007	436.112000	421.131989	423.204987	423.204987	30627700
2014-09-25	423.156006	423.519989	409.467987	411.574005	411.574005	26814400
2014-09-26	411.428986	414.937988	400.009003	404.424988	404.424988	21460800
2014-09-27	403.556000	406.622986	397.372009	399.519989	399.519989	15029300
2014-09-28	399.471008	401.016998	374.332001	377.181000	377.181000	23613300
2014-09-29	376.928009	385.210999	372.239990	375.467010	375.467010	32497700



	Open I	High	Low	Close	Adj Close	Volume	RSI_14	WT_10	WT_11	CCI_20	CCI_2	ADX_20	ADX_2	RSI_9	RSI_2	Candlestick Shape	Candlestick Pattern	Price Change	200 MA	Primary Trend I	RSI	Direction Tar	irget
	254.291	255.257996	251.100006	253.697006	253.697006	12493500	47.50830043739130	-17.696192532818300	-15.342915251404800	-25.13223939620700	-66.6666666666410	15.552763085324500	55.83198819462160	48.62406668398010	72.67039292746330	Bearish	Doji	-0.625	310.8431702350000	Bearish	47.50830043739130	-1.0	0
	260 721008	261 798004	254,574997	255 492004	255 492004	20034200	48 78580768473200	-27.922424190684700	-27 922424190684700	36 62734522436380	-66 6666666667510	14.313296639830900	79 03673441061480	50 158854225976900	43 46633560403260	Bearish	None	-5 105987999999970	309.01475013000000	Bearish	48.78580768473200	-10	0
_			252.205002			18467400			-38.159809518234500	3.606224063134580		13.637446814381300				17.2001200	None		308.30667014500000	-	47.1855987530671	-1.0	0
_	253.063995						41.956140742332800		-74.28268571921200								None		307.48726020000000		11.956140742332800		-1
-							41.14588379951530		-80.24269306777770								None		306.71153510500000		41.14588379951530	-1.0	
	243.694						36.819869108723400					13.49377594360140					None		305.88113509500000		36.819869108723400	-1.0	-1
			234 175003				37 26823229002610		-86.62/79/94596820 -86.97417935737640			1011001110011001110					Doil	0.4799960000000000			37.26823229002610	-1.0	-1
_																	-						-1
_	236.535004	EUT.TETUUT	E00.404000	200.100	200.100		37.032965811161300	00.04010010412140	-88.34876815412740	104.72004200010100	00.000000000000000000000000000000000000	10.000000000000000000000000000000000000	00.00210040000020	01.400 IEE00 IOIE IOO	0.00100000123201	Dourism	Doji	-0.39900199999999600	000.040000E4000000	DOM: III	37.032965811161300	-1.0	-1
_							30.936419461115700			-210.38507454532000							None		303.01474525000000		30.936419461115700	-1.0	-1
_							28.560137559175700		-93.76360939897410								None		302.0884152950000		28.560137559175700	-1.0	-1
_							33.310409005336900		-83.48544472979290								None		301.20998031500000		33.310409005336900	-1.0	-1
_			223.917007				37.82548796801580			-111.47959085103300							None		300.46694030500000		37.82548796801580	-1.0	-1
	222.852997	224.315994	220.876007	223.356003	223.356003	12939000	35.24831229847610	-76.39534998458470	-81.25731631716690	-107.60268947829300	-66.6666666666780	17.687194782267500	34.42657973147400	32.614855997631200	38.47242313966970	Bullish	Doji	0.47399899999999200	298.8860752900000	Bearish	35.24831229847610	-1.0	-1
	222.612	226.350998	221.977005	224.626007	224.626007	18364700	37.02994006150550	-64.23275949152260	-69.66494321252840	-78.81065135201560	66.6666666666320	18.285353837579700	22.396789697736200	35.95952577689410	66.96360274619740	Bullish	None	2.0260009999999800	297.32877045000000	Bearish	37.02994006150550	-1.0	-1
	224.619995	235.268997	224.300995	235.268997	235.268997	24978000	47.16587223363090	-11.487970637139400	-18.385478938253000	-34.31725016341630	66.6666666666640	17.933426558861100	55.777553391244900	52.542705257850100	94.92346652036230	Bullish	None	10.642990000000000	296.70755547000000	Bearish	47.16587223363090	-1.0	-1
	235.602005	237.908997	233.475998	234.175995	234.175995	23847900	46.34090614514110	-17.293626299158100	-17.293626299158100	-10.045783631749500	66.6666666666670	17.426340948806400	74.32161674952990	51.01642901925850	80.86631509149700	Bearish	None	-1.0930020000000100	296.23410546000000	Bearish	46.34090614514110	-1.0	-1
	234.052994	236.475006	233.199005	236.462006	236.462006	17036000	48.37472670221830	-6.7033774994615600	-6.7033774994615600	-3.5156790543719800	66.6666666666670	16.96036998754550	77.13722550385630	54.15020531041020	88.1851224199285	Bullish	Engulfing	2.286011000000000	295.81386544000000	Bearish	48.37472670221830	-1.0	-1
	235.970001	236.304993	229.932999	231.268005	231.268005	21448700	44.26910927227630	-34.48075661494210	-30.7652752138088	-17.19988010499440	-66.6666666666670	16.70470454240960	53.67869324268570	46.53964172183140	32.205712468884200	Bearish	None	-5.194001000000010	295.3198104150000	Bearish	44.26910927227630	-1.0	-1
	231.235001	232.561005	226.337006	226.389999	226.389999	13957200	40.76949432010770	-67.62757448927050	-59.807897146390800	-40.478252453550600	-66.66666666666670	16.660373363293900	61.981073746819800	40.522312703649400	14.690050043847500	Bearish	None	-4.878006000000000	294.7708253500000	Bearish	40.76949432010770	-1.0	-1
	226,410004	226.944	214.873993	219.429993	219.429993	28943700	36.35359818868200	-80.22140564855130	-80.22140564855130	-100.76360911903400	-66.6666666666670	17.170942560480300	77.80232888940410	33.557756495980300	5.756295506518680	Bearish	None	-6.960005999999999	294.1032753050000	Bearish	36.35359818868200	-1.0	-1
	219.429001	233.304993	218 022995	229 285995	229.285995	38574000	45.376369300387500	-37 434341231284300	-37.434341231284300	-35.00653066660940	66.66666666666670	17.20409505035570	40.62548591907670	47.8395886098554	65 38201791398650	Bullish	Engulfing	9.856002000000020	293 4245752750000	Bearish (15.376369300387500	-1.0	-1
_	228 968994						43.086494501365200		-52.32905972145700								None	-3.430999000000000			13.086494501365200	-1.0	-1
	225 501002	227 020002					43.05444404576240		-52.53308399685980			17.26551003785000					Doji	-0.046997000000000460			43.05444404576240	-1.0	-4
-			232.078995					-30.313117848260600				15.70586821678420					None		290.55989007000000		18.506972127897100	-1.0	-1
_			229.231003						-47.44412785272750			13.066786339846200					None	-6.33900400000000000000			45.97783520237530	-1.0	-1
	229.662003				237.334		45.97783520237530 52.09490891362490		-47.44412785272750 -23.420718792592800			12.420053662639600						7.552002000000020			45.97783520237530 52.09490891362490	1.0	-1
_																	Engulfing						- 0
_							49.52727762647720					11.303511652319200					None	-4.735000999999980			49.52727762647720	-1.0	. 0
			236.794998				50.712695991246100		-53.034181316491900			10.311965500395600					Doji	0.6759950000000000		-	50.712695991246100	0.0	0
			236.009995				49.915298557886800					9.333363330713630					None		279.81145014500000		49.915298557886800	-1.0	0
_	200.001000	204.101001	Londinosi	201.041000	2011041000	14241000	44.698102485673300	30.300004341330040	-99.18680803152570	101:01:000104E1EE00	00.0000000000000	S. GOOGE GOODE OF GUECOS	11100004002101220	40.23042010040000	0.040440E04704000	Doderon.	None	111010000000000000000000000000000000000	278.71869517000000	DOM: III	14.698102485673300	-1.0	0
	231.889999	234.682999	231.841995	234.018005	234.018005	15499400	47.36956713485020	-82.32551618272330	-82.32551618272330	-87.09226629916200	66.666666666880	8.876553534284570	58.135384686319400	45.635624717394200	57.10175807997130	Bullish	Engulfing	2.0709990000000000	278.2600402300000	Bearish	47.36956713485020	-1.0	0
	234.016006	236.242004	233.835007	235.343994	235.343994	15108900	49.06609534323200	-70.3289746760198	-71.67757046232140	-51.20998245174640	66.6666666666670	8.574023941023790	52.33506060762620	48.9287536333746	74.64757771286290	Bullish	None	1.3259890000000200	277.80730021000000	Bearish	49.06609534323200	-1.0	-1
	237.104004	238.242004	235.692001	237.115997	237.115997	16425000	50.31025136661860	-47.84963993687240	-47.84963993687240	-26.118603800778900	-66.6666666666510	7.6415525591670600	56.48374364220050	50.143445510863100	30.02400862391020	Bullish	Doji	0.00599599999981910	275.33311022000000	Bearish	50.31025136661860	0.0	0
	237.065002	238.636002	236.695007	237.283005	237.283005	18837000	50.54029379567840	-46.20602153568400	-46.20602153568400	-16.47746250103930	66.6666666666670	7.352665677133210	37.926997270769900	50.577875544249300	36.93845518459940	Bullish	Doji	0.16700800000001000	274.7920852100000	Bearish	50.54029379567840	0.0	0
	237.257004	237.824005	236.651993	237.408005	237.408005	13829600	50.72417294607860	-44.97582717284790	-44.97582717284790	-17.883308441580200	-66.66666666666670	7.073472145497100	30.56892111432010	50.93784813406410	45.064254958830000	Bullish	Doji	0.125	274.1628051950000	Bearish	50.72417294607860	0.0	0
	237.091995	237.093002	232.046005	233.345001	233.345001	14098600	44.51912162594240	-86.9210929690641	-85.17160908216080	-106.71389718516200	-66.6666666666670	6.813377202391180	74.34388018582890	39.27126597760000	2.6338012367309000	Bearish	None	-3.7509920000000000	272.8419150750000	Bearish	44.51912162594240	-1.0	0
	233,134995	233.251999	229.542007	230.190002	230.190002	14730800	40.366893894024000	-94.7893573410723	-94,7893573410723	-188.84206475242700	-66,66666666666730	7.134392853750740	86,24777709566110	32.676939000144300	1.0448864444435500	Bearish	None	-3.1549990000000000	271.8750600600000	Bearish 4	10.366893894024000	-1.0	0
_							32.78508826591410			-280.58714511266500					0.2765748813115290		None	-7,264006999999999			32.78508826591410	-1.0	0
_	222.893997		222,419006				37.77040151823230			-224.15649081507400				31.961325694838300	36,98179087434940	Bullish	None	2.876999000000010	270.0259401150000		37.77040151823230	-1.0	-1
-		227 404007	223 929993	225.873993	225.873993	17752400	37.892825301771700			-166.02651404149000							Doil	0.07099900000000050			37.892825301771700	-1.0	-1
_			224.054001				36 217688370473900			-150.09717444773200		10.549717033363300					None	-1.54998800000000100			36.217688370473900	-1.0	.1
-			223,179001			1-1120100	37 42468521879830	02.00. 102.00 12. 100	-78 91584987010090			11.312474043991300					None	0.6279910000000000		acara.	37.42468521879830	-1.0	-1
-	225.595993		222.651993				35,431896323658100			-129.8943426453440 -106.08727661985500							None	-2.7380059999999800			35.431896323658100		-1
-																						2000	-1
-			222.839005							-61.233060718876500		13.187959497186400					Engulfing	5.608001000000000			45.69088615435520	-1.0	-1
-			227.929001			28353100			-50.92738828778670								None	0.5590059999999820			46.60165349167080	-1.0	-1
_			228.009995			15904800	40120000100104000		-37.21150244806080								Doji		264.23197023000000		46.23566798154650	-1.0	-1
-			228.766998			14416000	47.85916613812510		-19.276149967284200			13.352245764285100					None	0.902007999999995			47.85916613812510	-1.0	-1
	229.705002	231.057007	229.313004	229.981995	229.981995	14017700	48.37470708191530	-12.7901274168013	-12.44514789938070	-6.875835506976490	66.6666666666670	13.236554345013800	83.41886986632550	51.94602482636500	88.09370416876980	Bullish	Doji	0.2769930000000050	262.8080352200000	Bearish	48.37470708191530	-1.0	-1
	249.427994	252.108002	244.126999	249.007004	249.007004	30980200	70.15918594506780	-27.12624147783740	-23.066957268644100	187.77695685962400	-66.6666666666670	15.329917267940400	73.4140084460773	78.39348234479650	79.04725146104450	Bearish	Doji	-0.2769930000000050	258.85094027	Bearish	70.15918594506780	0.0	0
	248.720993	257.173004	248.580994	257.063995	257.063995	34742900	68.66970894687030	-0.6485542984623840	-0.6485542984623840	131.61320249973200	66.6666666666670	20.65538785175230	86.76135563378550	72.79117351252860	89.76639198333150	Bullish	Engulfing	8.05299299999999	252.181510255	Bullish	68.66970894687030	1.0	0
	257.036011	267.867004	255.945999	263.071991	263.071991	44533800	72.66705690355290	-17.43514353265730	-17.43514353265730	202.5146725604740	66.6666666666640	22.196692727235000	93.01106276322900	77.63074161979110	95.03340414740730	Bullish	None	6.007996000000050	251.73416025000000	Bullish	72.66705690355290	1.0	0
	263.345001	265.171997	255.774002	258.621002	258.621002	27029800	65.95344013005430	-33.619381631131700	-33.619381631131700	160.97910085641900	-66.6666666666670	23.637442541873400	93.86320705861890	67.60824605373970	53.9168378638743	Bearish	None	-4.45098900000005	251.290385225	Bullish	65.95344013005430	1.0	0
	258 552002	261 631012	254 115997	255 412003	255 412003	21551900	61 53908038979650	-45 287620728951400	-45 287620728951400	120 2933072293640	-66.66666666666530	24 775790360122500	64 55088157265200	61 19973333285730	33 20298936930940	Bearish	None	-3 2089989999999	250 80928529500000	Bullish	61 53908038979650	1.0	1



Imagine we have a dataset that contains information about the price of a cryptocurrency, in this case, Bitcoin. We want to predict whether the price will go up, down, or stay the same. To do that, we need to analyze different features in the dataset.

First, we load the dataset and select the features that we think are important for prediction, such as the closing price, candlestick shape, candlestick pattern, 200-day moving average, primary trend, RSI (Relative Strength Index), and direction.

Since some of these features are categorical (e.g., candlestick shape and pattern), we encode them into numerical values so that our model can understand them.

Next, we scale the numerical features (closing price, 200-day moving average, and RSI) using a process called standardization. This ensures that all features are on a similar scale, making it easier for the model to learn.

We then split the data into training and testing sets. The training set is used to train our model, while the testing set is used to evaluate its performance.

Now comes the exciting part. We create a Gradient Boosting model, which is a machine learning algorithm known for its ability to make accurate predictions. We want to find the best configuration for this model, so we use a technique called grid search to explore different combinations of hyperparameters (settings) and select the ones that give us the best results.

Once we have the best model, we make predictions on the testing set and even predict the target for tomorrow based on the features of the last day in the dataset.

We evaluate the performance of our model by comparing the predicted values with the true target values in the testing set. This gives us metrics like precision, recall, and F1-score, which tell us how well our model is performing.

Finally, we save our best model and export the testing set with predictions and true target values to a CSV file for further analysis if needed.

In summary, this script shows how we preprocess and analyze data using machine learning techniques to predict the future price movement of a cryptocurrency.

of ripple

10000

MODEL RESULT

	70% ~			:=	⊞,		ī	e e	®	A	9	2				.			(2)
View	Zoom			Add Category	-		_			t Text S	_					Collaborate			Organize
_				Add Category	Pivot lable		1111	sert rabi	e Citai	t lext s	паре п	ieuia	Comment			Collaborate	G .		Organize
+	Sheet	1																Sheet	
Close		200 MA	RSI	Direction Candlestick St	hape_Bearish Candlestick Sha	e_Bullish Candlestick Patter	Doji Candlestick	Pattern_Engulfin	g Candlestic	k Pattern_Hamme	r Candlestick	Pattern_No	one Primary Trend_B	earish Primary Tr	end_Bullish Prediction	True_Target			
0.42		0.40645511373346400	2.1848695083347900	1.0	1	0	11	8	0		0		0	0	-11	1 1	Sheet Name		
		-0.7100712168917870		1.0	0	1	0		0		0		1	0	1	1 0			
	124451965962600	0.4117062713796880 -0.5623254325833110		1.0	0	1	0		1 0		0		0	0	0	0 0	Sheet 1		
		-0.5523254325833110	-0.6911160850321600	1.0	0	1	0		1		0		0	0	- 1	1 0			
3.	81615436115630	2.1185879729573900	0.1773131499991670	1.0	0	1	0		0		0		1	0	- 1	1 1			
1.2	75856401673600	1.976893158182560	-0.6327556370065160	-1.0	1	0	1		0		0		0	1	0 -	1 -1	Background		
	98497621927300	2.139036021228460	0.44721436002727700	1.0	1	0	1		0		0		0	0	- 1	1 1	Daonground		
1000000	34034038963080	-0.3947597122508920 0.5321282344812290	-1.187416644937760 -0.74927617737258	-1.0 0.0	0	1 0	1		0		0		0	0	0 -	1 -1			
	20938113543228	-0.7004507786112340		1.0	1	0	1		0		0		0	0		0 1			
	41371885307100	0.8584833361014630	-0.7674735223129450	-1.0	1	0	0		0		0		1	1	0 -	1 -1	Du	plicate Sheet	
1700000		-0.8499210747200720	-1.1363112427855800	-1.0	0	1	0		0		0		1	1	0 -	1 -1			
		-0.37467150707346800	-1.0188513565525900	0.0	0	1	1		0		0		0	0	1	0 0		elete Sheet	
	554732602816670 501845093319530	-0.8000374282681060 -0.8502176655472350	1.456529246502900 0.7798477368563080	1.0	0	1	1		0		0		0	0	1	1			
	22919144874100	-0.8481025959823470	-0.6940150048787030	-1.0	1		0		0		0		1	1					
	98459895816800	1.7663920753046000	-1.274756416023020	-1.0	0	1	0		1		0		0	1	0 -	1 -1			
	10496040663300	2.343080691151780	-0.6250758006932180	-1.0	0	1	0		0		0		1	1	0	0 -1			
	48581622758190	1.5870175805130600	-1.7565656569845400	-1.0	1	0	1		0		0		0	1	0 -	1 -1			
	773960502345560 516436092546800	-0.761705853932684 -0.286931548481423	0.22377986542306300	1.0	0	1 0	0		0		0		0	0	0 -	1 1			
		-0.286931948481423		1.0	0	1	0		0		0		1	0	1 .	1 1			
-0.5	550919192993010	-0.5685421562760160	0.9631281531991270	1.0	0	1	0		1		0		0	0	1	1 1			
	112847830139180	-0.8322485142478640	-0.9515381259043140	0.0	0	1	1		0		0		0	0	1	0 0			
	752626673288230		-0.07485241503518750	1.0	0	1	0		0		0		1	0	- 1	0 1			
	583966186931850 506926830512110	-0.4838128896720180 -0.8423075530826870	-1.0452692524296900	-1.0 1.0	0	0	0		0		0		1	1	0 -	1 1			
	803031289334000	-0.286541823499108	0.7184161925075730	1.0	0	1	0		1		0		0	0	- 1	1 1			
-0.6	38299193832750	-0.4323791641761960	-1.5529727119393000	-1.0	0	1	1		0		0		0	1	0 -	1 -1			
	13210389816920	-0.3056243394844030	-1.4572137066944400	-1.0	1	0	1		0		0		0	1	0 -	1 -1			
		-0.04350075119681220 -0.8499310576313970	1.168226612007100 -0.8147764786925110	1.0	0	1	0		1		0		0	0	1	1 1			
	823719135587630 880684485395310	-0.8499310576313970 -0.8478510728001500	-0.8147764786925110 1.1590473487273300	-1.0	0	1	0		0		0		1	0	0 -	1 -1			
	.08754855771064		-0.07892682745814420	1.0	0	1	0		0		0		1	0	1	0 1			
	48620363807872	2.0661473963320800	0.791234480442009	1.0	1	0	1		0		0		0	0	-1	1 1			
		-0.22663106212417200	0.7915671261856340	1.0	0	1	0		0		0		1	0	1	1 1			
	990905857022270	0.4183425394290000 -0.826270189501002	0.37772908895918800	1.0	0	1 0	0		0		0		3	0	1	1 0			
	040874379095110	-0.826270189501002	1.214819145683230	1.0	0	1	0		0		0		1	0	- 1				
	341465892751300	-0.3400039147699810	-0.3711077298100280	-1.0	1	0	1		0		0		0	1	0	0 0			
-0.04	183503440517600	-0.20489899693985200	1.3279945199660100	1.0	0	1	0		0		0		1	0	1	1 1			
	065355228183910	1.6330486820608600	-1.433248353915050	-1.0	1	0	0		0		0		1	1	0 -	1 -1			
	043342574698940 746135769590760	-0.7556496741416410 -0.1751516978872150	0.10588076590526800	1.0	0	1	0		1		0		0	0	- 1	1 1			
	27666334802900	0.7970566676451680	1.2347123186356200 -0.5672019717563100	-1.0	0	1	1		0		0		0	1		1 1			
-0.26	860154177727200	-0.28691778155680600	0.2940262672957380	1.0	0	1	0		0		0		1	0	1	1 1			
-0.6	546342171354450	-0.4469374089391190	-1.5761491068520100	-1.0	0	1	o		0		0		1	1	0 -	1 -1			
	137555265839740	-0.8390872989570620	1.757260454528140	1.0	0	1	0		0		0		1	0	-1 -	1 1			
	R38378978864700 R88906968293280	2.083084344372850 -0.395841921208332	0.6412276337153290 -0.5003658291037130	1.0	0	1 0	0		1		0		0	0	1 0 -	1 1			
	788906968293280 157113480599170	-0.396841921208332 -0.8183592903655300	-0.5003658291037130 -0.10507762353977100	-1.0	1	0	0		0		0		0	0	1	1 1			
	140042642892900	2.1232787059586400	-0.7538597612855050	-1.0	1	0	0		0		0		1	1	0	0 0			
-0.8	10783109504230	-0.8382387484972550	1.4707535063896800	1.0	0	1	-1		0		0		0	0	-1	1 1			
	99885952310800	-0.8502217599635960	1.081924853720220	1.0	0	1	0		0		0		1	0	1	1 1			
		-0.24124902614907400	0.4832299824580390	1.0	0	1	0		0		0		1	0	1	1 1			
	922055358978820 991277065755480	-0.08675158445090810	-0.2771783907138130 0.05876888653548880	1.0	1	0	0		0		1		0	0	1	1 1			
		-0.8313467601291850 -0.8499269094881400	-0.571058406851718		1	0	0		0		0		1	1	0 -	1 0			
	43993272047970			1.0	0	1	0		0		0		1	0	1	1 1			

