Analysis on Ethereum Market Cap

Block.

Analysis on Ethereum Market Cap

CONTENTS

fluctuating Ethereum value

Our goal

Our purpose is to make it easy and reliable for those who are interested in Ethereum, the second largest cryptocurrency in the world, to understand and help their decision-making process.

Problems

There are so many factors that compose the market cap of the cryptocurrency, and this is the reason by we cannot intuitively get meaningful insight from ethereum market cap.

Related works

Some research has been done on large Bitcoin transactions such as Huobi. This has helped to gain a better understanding of the transactions. But for our topic, no meaningful research has done.

Exploring , Cleaning , PreProcessing, Dimension Reduction

	tx_date	tx_count	Tx_volume_Ether	tx_count_ERC20	Tx_volume_ERC20	Volume($)	Market Cap($)	DAU	CUM
0	2015-10-30	7941	1.209372e+06	3.0	2.000000e-15	2429200	77401817	7501	7501
1	2015-10-31	7557	2.764179e+05	0.0	0.000000e+00	673892	68163368	7076	14577
2	2015-11-01	6915	1.455244e+05	0.0	0.000000e+00	588913	78530263	6516	21093
3	2015-11-02	6558	3.518903e+05	1.0	5.000000e-16	1145200	73654327	6223	27316
4	2015-11-03	7399	5.951720e+05	30.0	8.001248e-07	1907690	75434114	6973	34289

figure1 : Dataset Target variable y : Market Cap ($)

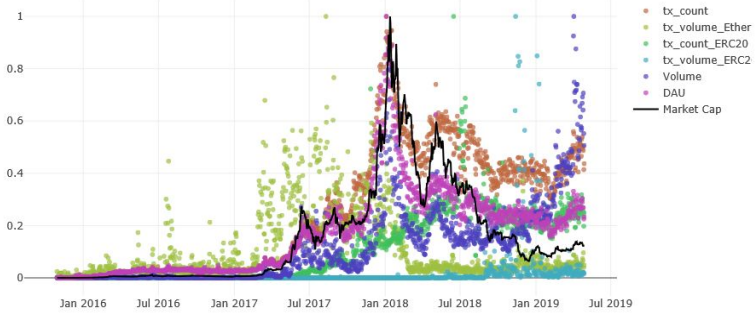


figure2 : feature scatter plots

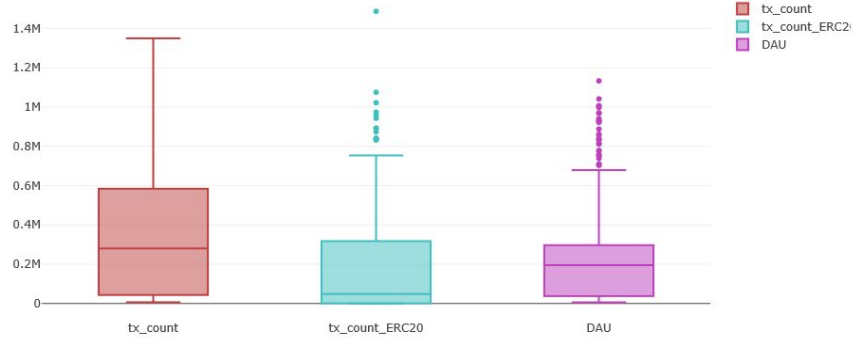


figure3 : Box plot for checking distribution

Linear Regression Model

x_lst	Coef	Score
DAU	1.327946e+05	0.848060
tx_count	7.853991e+04	0.815805
Volume($)	1.784916e+01	0.700871
Tx_volume($)	5.828936e+00	0.433478
tx_count_ERC20	8.718415e+04	0.361827
CUM	1.976944e+02	0.271924
Tx_volume_Ether	7.874096e+02	0.033819
Tx_volume_ERC20	-3.055525e-53	0.001385
Tx_volume_ERC20($)	-7.225363e-56	0.000188

- DAU recorded the highest R2 score to predict Market Cap
- tx_count, volume followed
- Under tx_volume($), there are no meaningful variables
- PLS regression with top 3 variables = 0.832 R2 score

OLS regression with top 3 variables

OLS Regression Results

Dep. Variable:	Market Cap($)	R-squared:	-98618.601
Model:	OLS	Adj. R-squared:	-98889.286
Method:	Least Squares	F-statistic:	-364.3
Date:	Thu, 07 Feb 2019	Prob (F-statistic):	1.00
Time:	10:10:11	Log-Likelihood:	-26885.
No. Observations:	1096	AIC:	5.378e+04
Df Residuals:	1093	BIC:	5.379e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
DAU	9.58e+04	7453.474	12.854	0.000	8.12e+04	1.1e+05
tx_count	1.586e+04	3870.728	4.097	0.000	8262.901	2.35e+04
Volume($)	0.4014	0.615	0.653	0.514	-0.805	1.608

Omnibus:	206.302	Durbin-Watson:	0.183
Prob(Omnibus):	0.000	Jarque-Bera (JB):	704.581
Skew:	0.895	Prob(JB):	1.01e-153
Kurtosis:	6.496	Cond. No.	4.09e+04

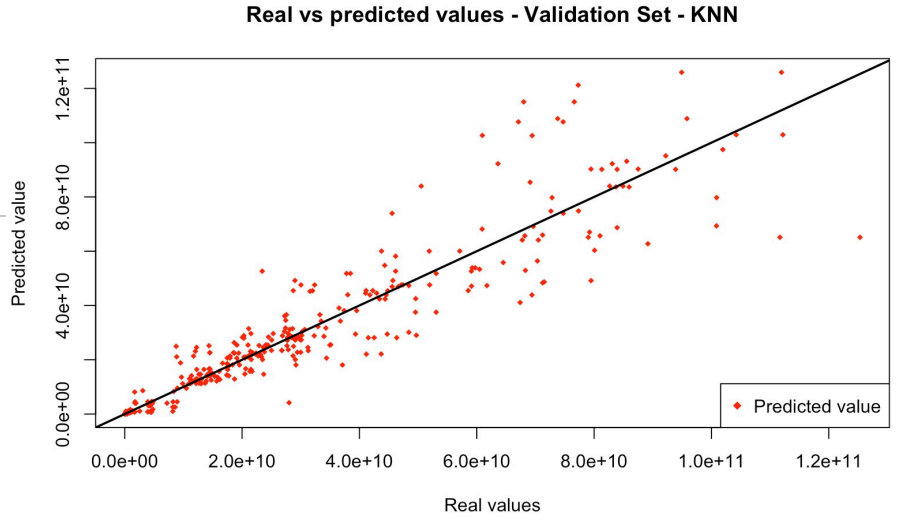
02 **Method**

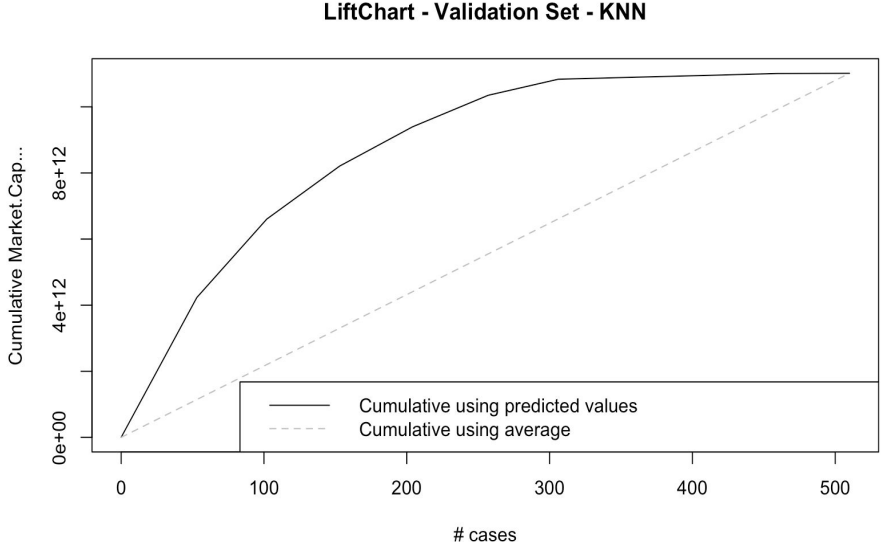
K Nearest Neighbors model (KNN)

	k	accuracy
1	1	33876264586
2	2	33876264585
3	3	33876264584
4	4	33876264584
5	5	33876264586
6	6	33876264585
7	7	33876264588
8	8	33876264587
9	9	33876264586
10	10	33876264587

Best k = 3

According to these two graphs, the KNN method applied to our problem in order to predict the Market Cap seems to perform well.





Neural Net (NN)

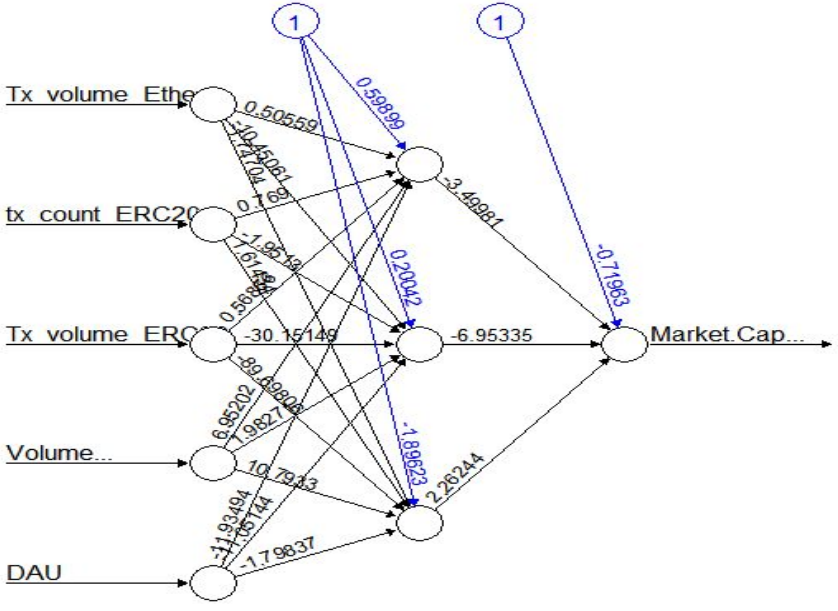
Decision : How many Node for hidden layer??

K : from 1 to 5(number of input layer's node)

best : k = 3

k	mse
1	7.513717e+19
2	5.738568e+19
3	5.037425e+19
4	5.061420e+19
5	5.303013e+19

Result [hidden node = 3, hidden layer =1]



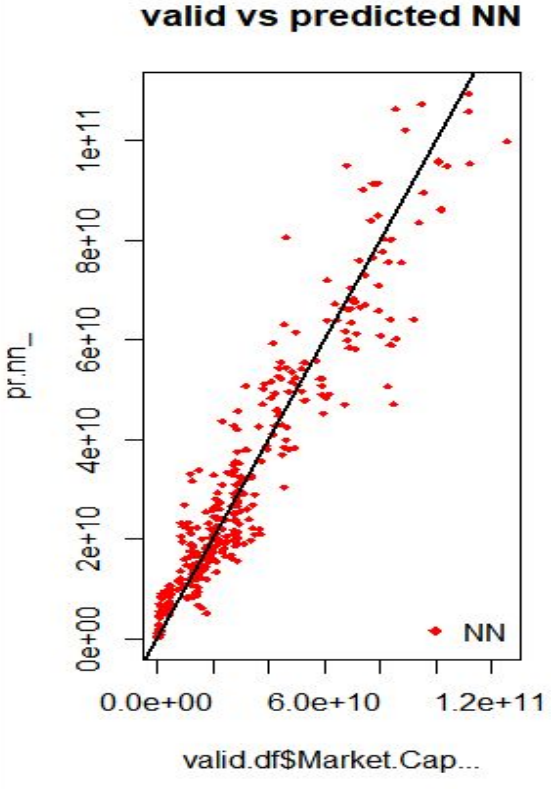
Selecting the proper value of K(number of hidden node) is important **to prevent overfitting**

Generally NN predicts well, but it is **hard to find the correlation between input and output nodes.**

NOTE : Our model's input node is the same as previous model (LM, KNN)'s predictors

Neural Net (NN)

Evaluate : Show very high prediction compared other model!



Prediction : applying MSE

1. For using NN, all input and output nodes must be Normalized to [0,1] .

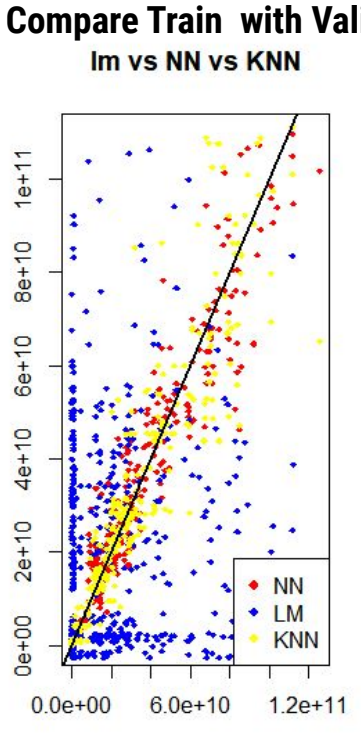
2. For Prediction We convert Output node 'Market.Cap' to its Original Value

3. Apply MSE.

4. Red Dot : Predicted Value for output node
   Line : real value (validation set)

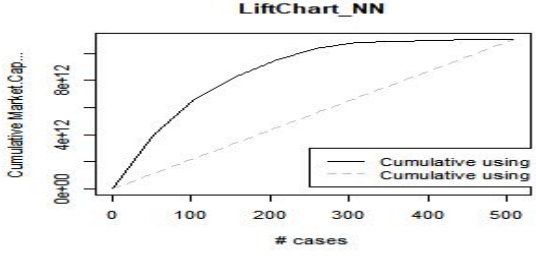
Comparison : We can show the difference through Line graph and Lift Chart

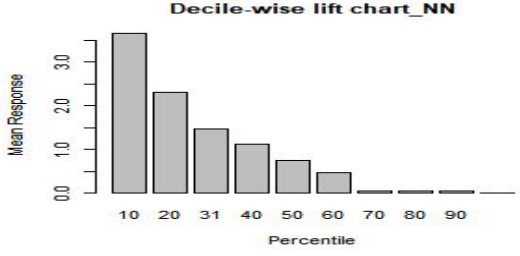
1. Line Graph

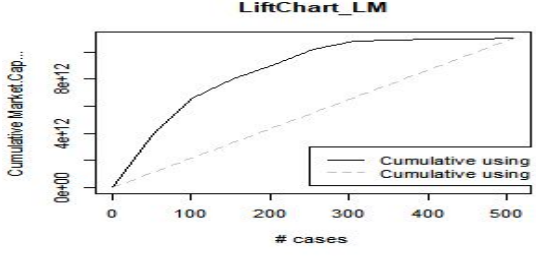


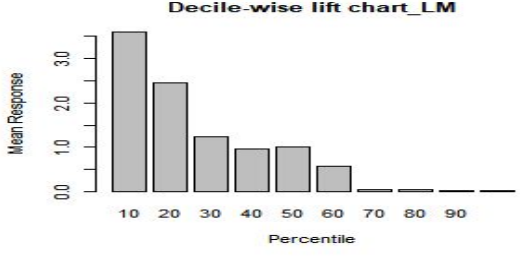
2. lift chart and decile-wise lift chart

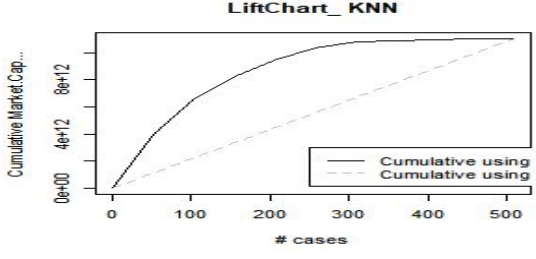
compare prediction value of validation set with real value of validation set

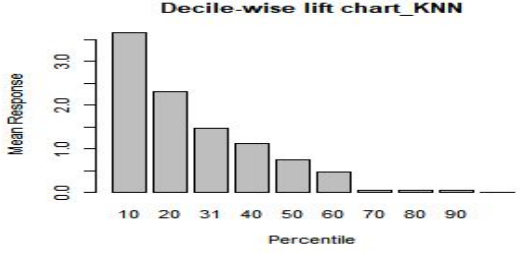












Ensemble

- The three models that we have obtained can be used together in order to predict the Target variable Market Cap. We can combine the predictions by **taking the average.**

  → Then the final model is the “super-model” created from KNN, NN and Linear Regression model

- There are some reasons why we can use ensemble by taking the average.

1. All Three model are sharing the same data set = each model’s data quality is equal.
2. There is no big differences in performance (Check Lift chart and Decile-wise lift chart in “Result”), which also means that there is no risk of outliers’ existence.

- For more elaborate model, we can calculate weighted average, weights based on each accuracy

- In conclusion, our purpose, which was to help the users of Ethereum in their decision-making process, is in a way achieved. The final model that we have built is quite reliable.