

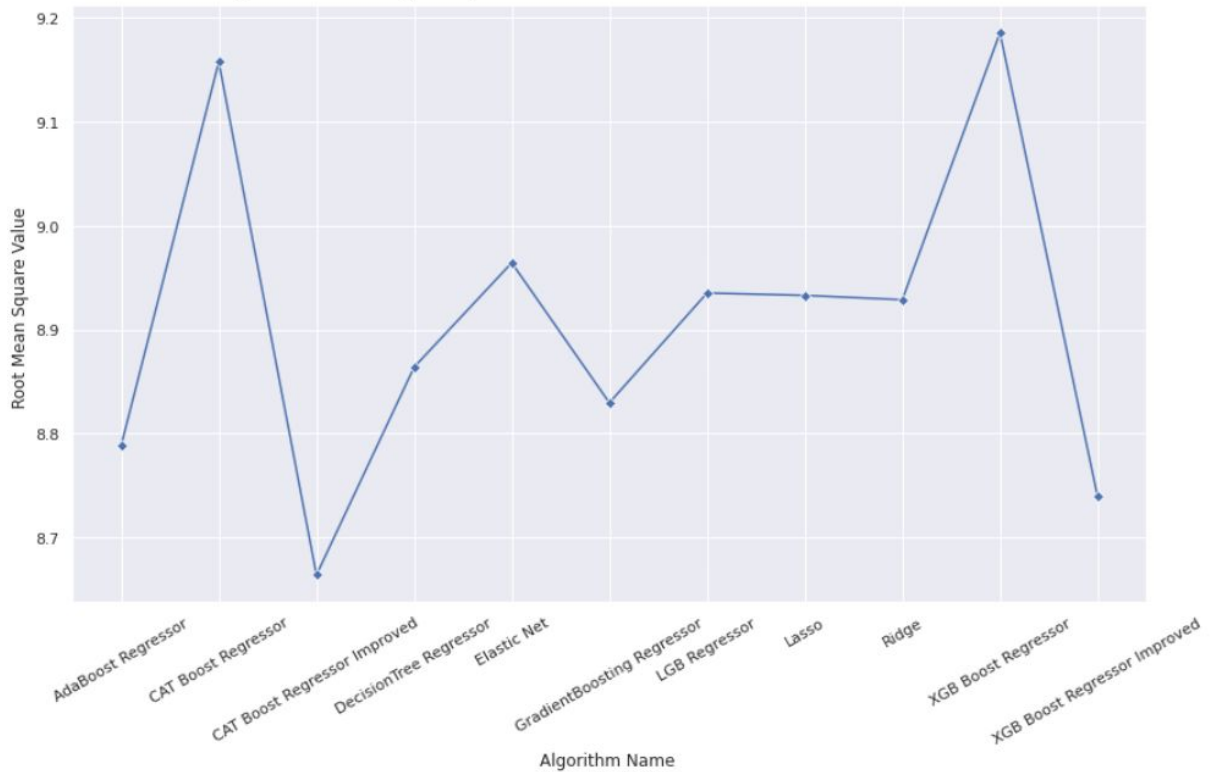
Solution Sheet

1. Which model have you used for probability prediction? Explain your model.

A. After analyzing different correlation to the prediction value and modelling different algorithms as below :-

	Algorithm Name	Algorithmic Modification	Root Mean Square Value
0	Lasso	NA	8.932825
1	Elastic Net	NA	8.964172
2	Ridge	NA	8.928631
3	DecisionTree Regressor	NA	8.863736
4	AdaBoost Regressor	NA	8.788214
5	GradientBoosting Regressor	NA	8.829472
6	LGB Regressor	NA	8.935110
7	CAT Boost Regressor	NA	9.157653
8	CAT Boost Regressor Improved	{'depth': 8, 'iterations': 100, 'learning_rate...	8.663731
9	XGB Boost Regressor	NA	9.185572
10	XGB Boost Regressor Improved	{'max_depth': 3, 'min_child_weight': 1}	8.739016

Table

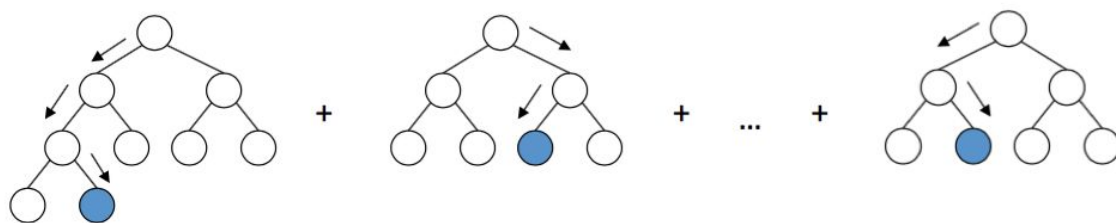


Line Chart

we got Cat boost as the best regressor with depth 8, iterations 100 and learning rate 0.03. These results are obtained by applying grid-search cv overall iterations possible.

Any regressor that has a strong bias resistant and false-positive detection mechanism suits the cause of deployment.

CatBoost :



The same decision base analysis is followed with changes such as direct input of categorical variables is allowed rather than in XGboost and Light GBM.

Iterations : -

iter	target	baggin...	depth
1	0.9481	3.174	7.113
2	0.9275	2.698	9.379
3	0.936	1.019	6.486
4	0.9592	5.0	6.0
5	0.9571	5.0	7.292
6	0.9592	5.0	6.0
7	0.9598	4.991	6.017
8	0.9598	4.992	6.009
9	0.9596	4.996	6.004
10	0.9589	4.998	6.038

Henceforth Catboost is perfectly fitting all criteria for deployment and that's why I have taken it as my regressor.

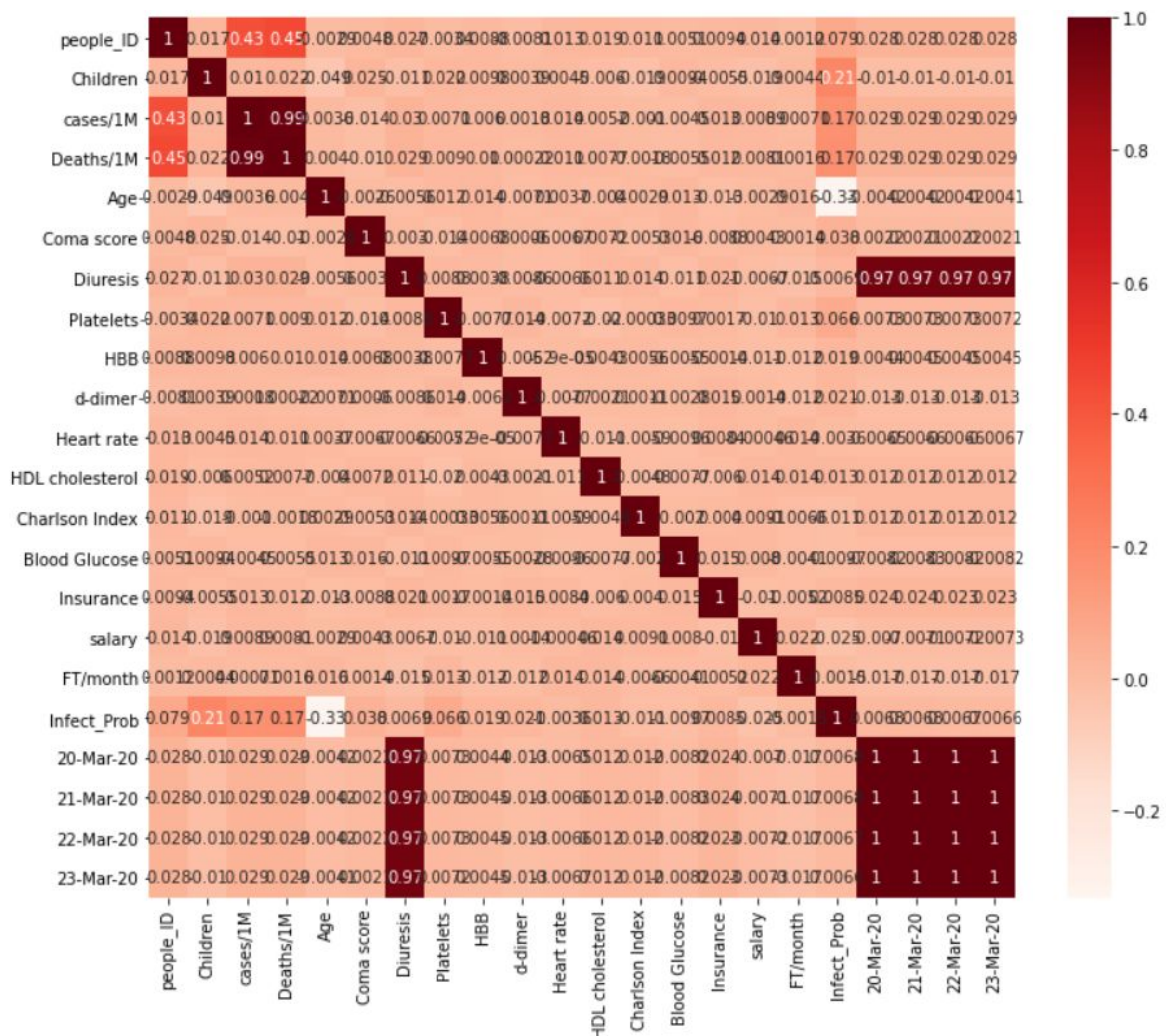
Do visit:

<https://github.com/fsrt16/Flickr-Hackathon-Coronavirus---a-pandemic-rehdreaseal-/blob/master/CoronaTreat.ipynb>

2. Which model have you used for Diuresis Time series prediction? Explain your model.

A. Time series modelling, first we combine both the database in order to see whether or not any correlation exists in the two.

Correlation :



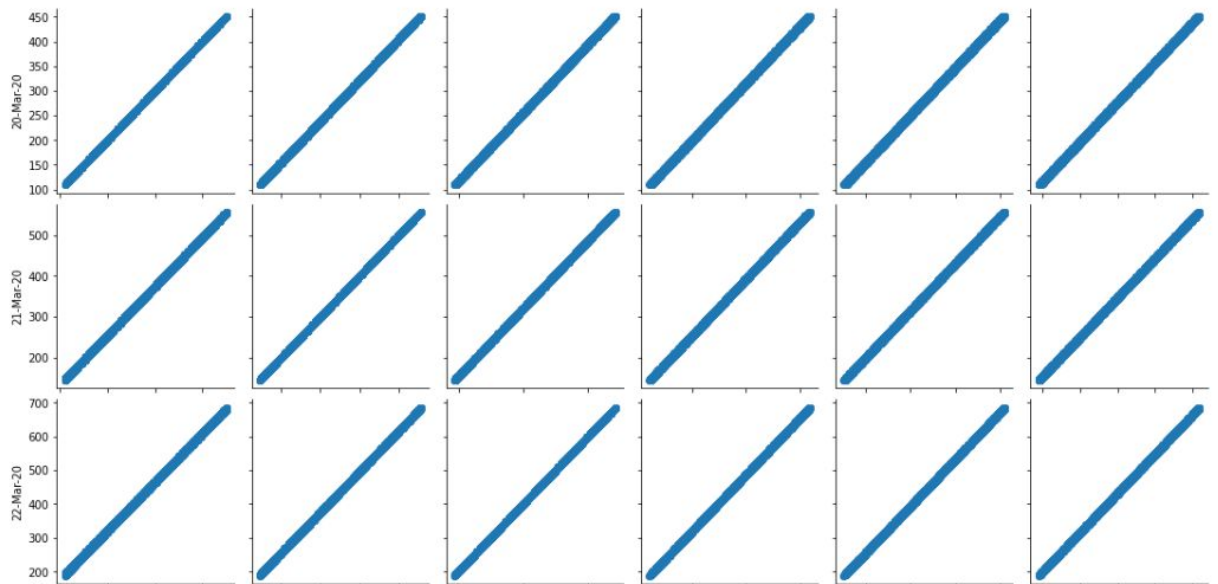
Thus the historical data of the person does not have any relation to the time series analysis, henceforth we can carry out with the only Timeseries data.

ML prediction on 26th march 2020 :

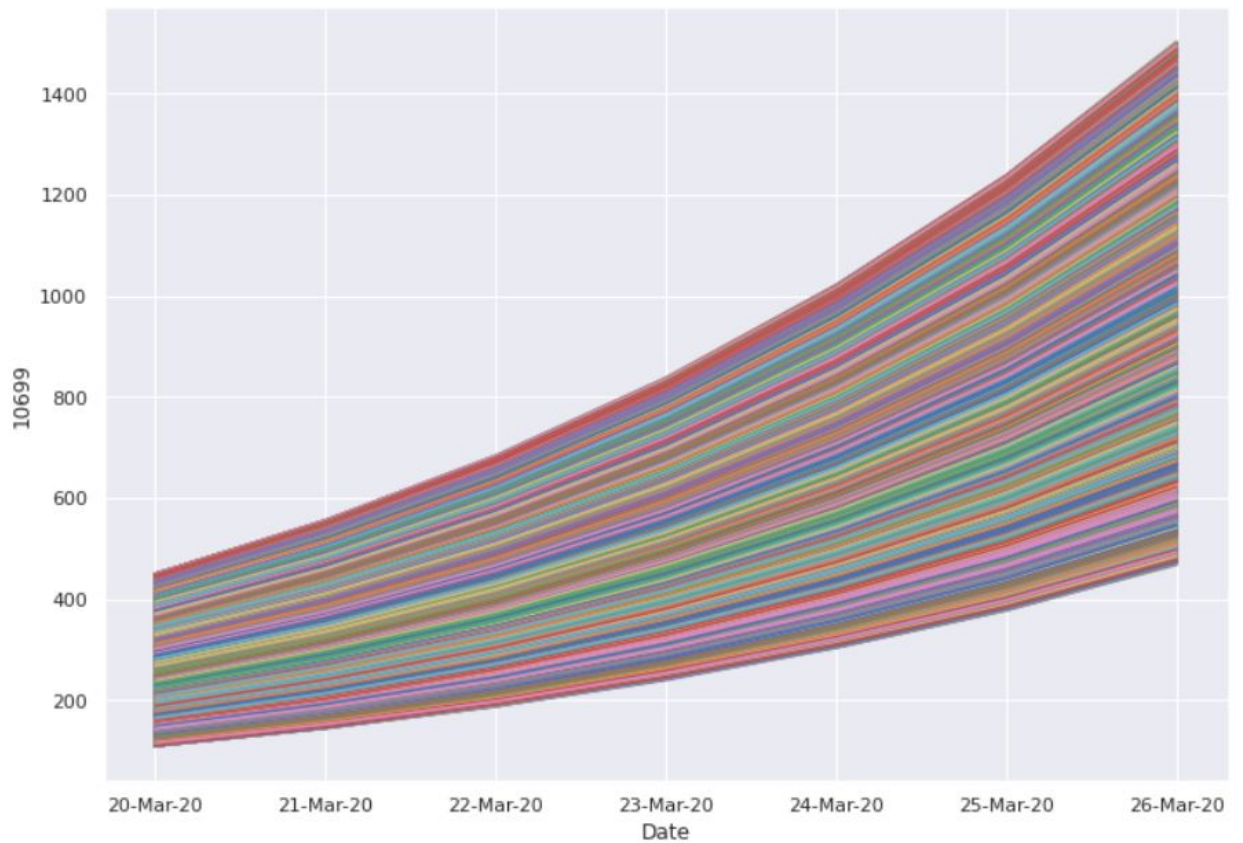
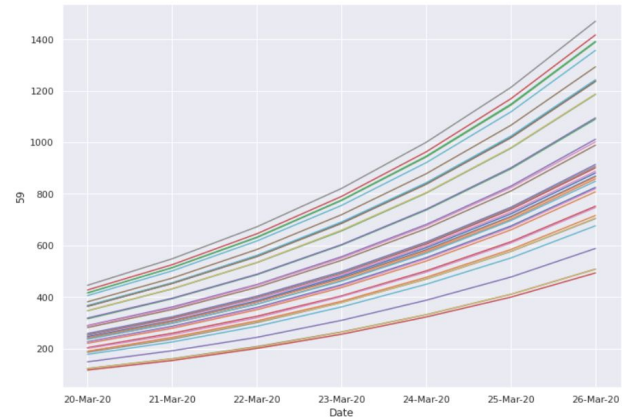
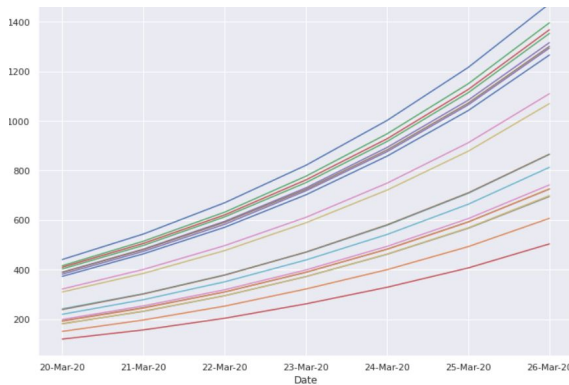
	Algorithm Name	Algorithmic Modification	Root Mean Square Value
0	Lasso	NA	3.841278
1	Elastic Net	NA	3.837064
2	Ridge	NA	0.974942
3	DecisionTree Regressor	NA	9.215991
4	AdaBoost Regressor	NA	1.408389
5	GradientBoosting Regressor	NA	8.493535
6	LGB Regressor	NA	104.897465
7	CAT Boost Regressor	NA	2.091381
8	CAT Boost Regressor Improved	{'depth': 10, 'iterations': 100, 'learning_rat...	3.450154
9	XGB Boost Regressor	NA	1.380501
10	XGB Boost Regressor Improved	{'max_depth': 9, 'min_child_weight': 3}	1.369699

This shows lasso, elastic net and Ridge works the best which also draws attention to the nature of the data which is Time Series hence simple models should substantiate better results which are seen so forth.

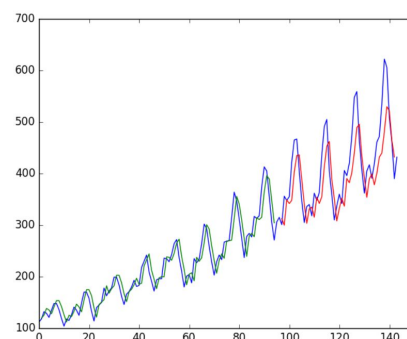
Below figure Strengthens our belief for the same Hypothesis:



Growth of 20 persons, 40 persons and all:

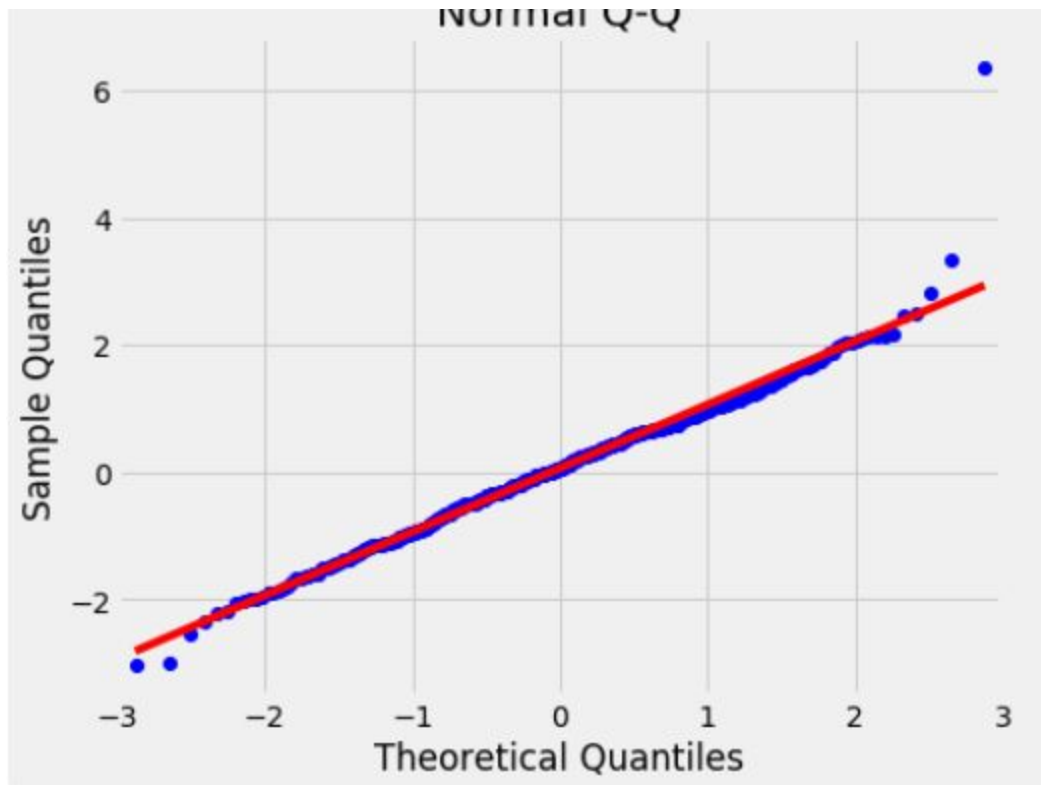


Thus this clear strip shows a window between which the Diuresis must lie, thus we are now motivated to use sliding window in LSTM and ARIMA models.



The figure shows LSTM learning pattern however the RMS value is 10,256 which is too high.

Arima rms is relatively less at 8,789 demonstrated below :



Finally, we are determined to suggest the use of Lasso regression because of :

1. RMS value of 3,125
2. Simple Algorithm for Time series is ideal
3. Biasness is reduced by alpha at 0.04.

Finally, we have taken prediction for 2020 march 27 and fed it to our earlier catboost model for results.

Do visit:

https://github.com/fsrt16/Flickr-Hackathon-Coronavirus---a-pandemic-rehdreaseal-/blob/master/Time_Series_Analysis.ipynb