

Assignment (Impact of Tweets)

As shown above I completed my assignment and it is written well how I performed all the tasks, from importing the data to predicting the correlated features. I have trained the data on models like Linear Regression, Decision Tree, Random Forests, and Artificial Neural Networks. Out of all the models Linear Regression have performed very well and the data given, is fitted perfectly to the model and predicted the best results.

Report:

The training error rates I obtained are:

Linear Regression:

```
Training data evaluation:
-----
MAE 390.49829714982894
MSE 39866179.92028137
RMSE 6313.967050934093
R2_squared 0.9963284216068522
```

Decision Tree:

```
Training data evaluation:
-----
MAE 9297.597976476207
MSE 543755944.42697
RMSE 23318.575094267017
R2_squared 0.949921397517997
```

Random forests:

```
Training data evaluation:
-----
MAE 214.1747494558196
MSE 4507666.964254432
RMSE 2123.126695290329
R2_squared 0.9995848548152204
```

Testing error rates:

Linear Regression:

```
Test data evaluation:
-----
```

```
MAE 311.18492577240653
MSE 5359971.045152342
RMSE 2315.1611272549353
R2_squared 0.9994681673873835
```

Decision Tree:

```
Test data evaluation:
-----
MAE 9125.75120292198
MSE 446548036.65582645
RMSE 21131.68324236918
R2_squared 0.9556921470297507
```

Random Forests:

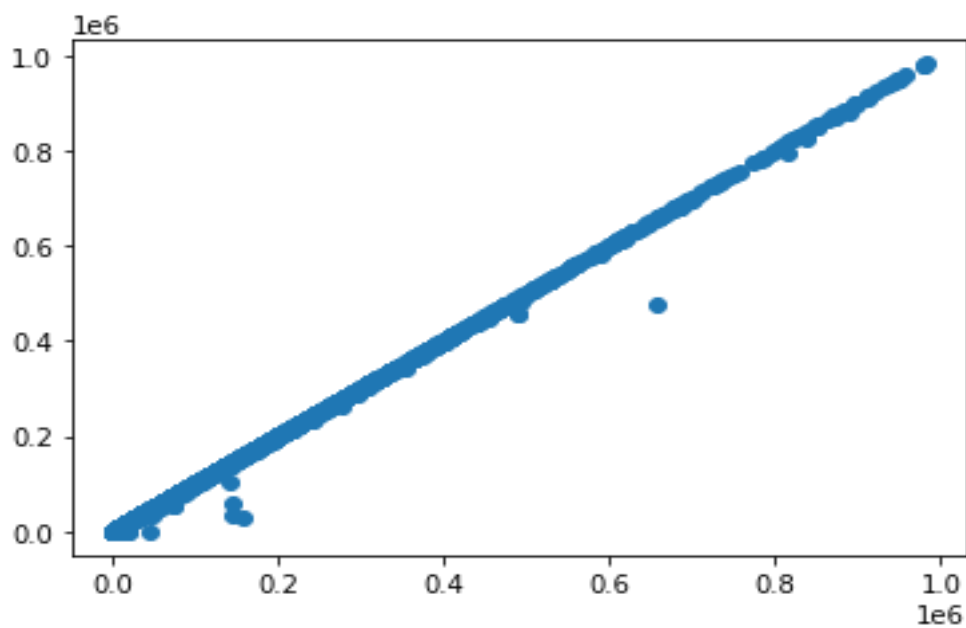
```
Test data evaluation:
-----
MAE 489.73985983942964
MSE 13605186.139588151
RMSE 3688.520860668698
R2_squared 0.99865005209379
```

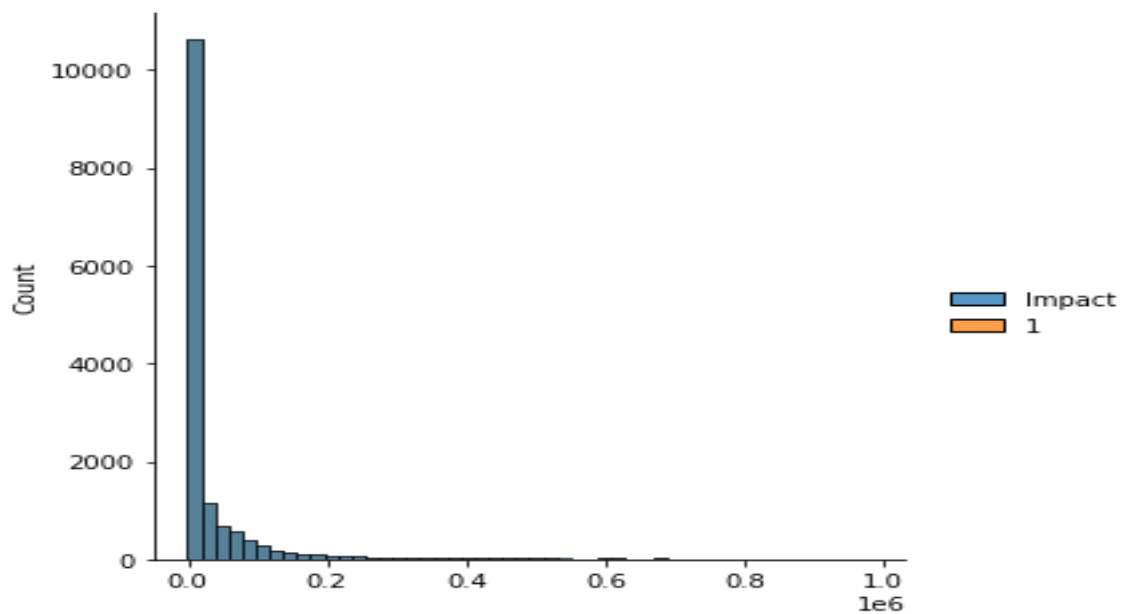
Artificial Neural networks:

```
Test data evaluation:
-----
MAE 901.4310343650818
MSE 10827562.999530697
RMSE 3290.5262496340456
R2_squared 0.9989256563011628
```

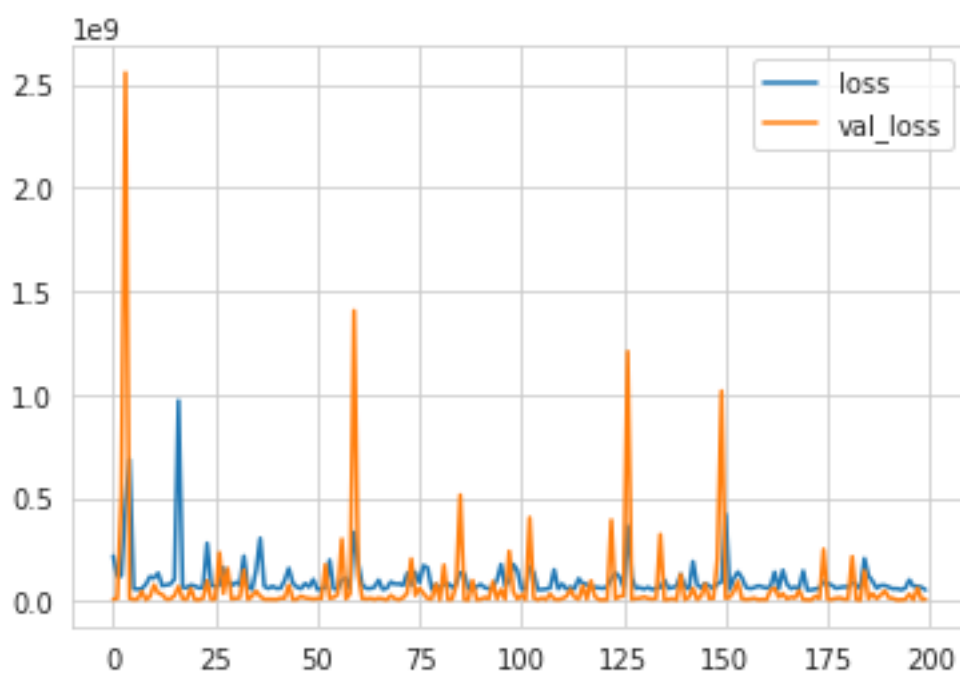
Graphs that show performance on both training and testing data.

Linear Regression:





Artificial Neural networks:



- The reason I got results is the data (features) given to the model are correlated to each other and the impact variable is highly dependent on these features, so the model trained well and given the best results. Given the accuracy almost high, i.e., R^2_score is nearly to one (0.9999) and the

Root mean square error (rmse is very low). These results shows it's the best.

- The model is pretty much fast as it is predicting the results in less than 5 seconds.
- I have chosen the learning rates to give the best loss without sacrificing the training time, I have started with, higher rates of 0.1 and tried for 0.01, 0.001 etc., and finally I have taken 0.0005 which is giving the best loss for my model.
- The best dropout rates for any model is 0.1 so, I take it for the best results.
- The Linear Regression algorithm performed best among all the models.
- I have defined the best model by predicting the R2_score. If R2_score is 1, it indicates the best, other than this the model has high rate of errors. So linear regression got best R2_score = 0.9999 which is nearly 1.