WIDS Project
Enron Investigation

Himanshu Rajpoot | 190260024 | Jan 26, 2022

Mentors >> Payal Choudhary and Akash Chodankar

Enron Scandal

About Enron

- Enron Corporation was an American energy, commodities and services company based in Texas. It was founded in 1985 by Kenneth Lay.
- Fortune named Enron "America's Most Innovative Company" for six successive years.
- The Company declared itself bankrupt in october, 2001 after being publicized.
- Company's stock price achieved a value of US dollar 90.75 per share in 2000 which plummeted to less than 1 dollar by the end of 2001.
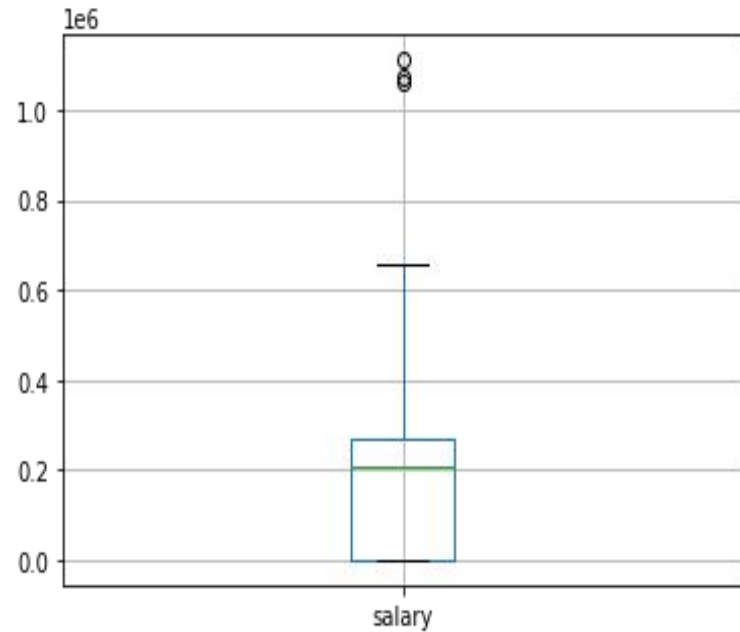
Objective

The Enron's emails and financial dataset contain the information which can help us predict the person of interest involved in Enron Scandal.
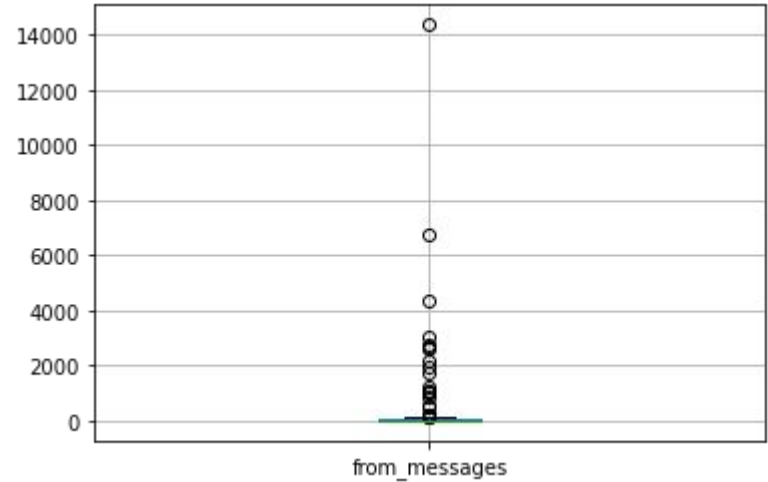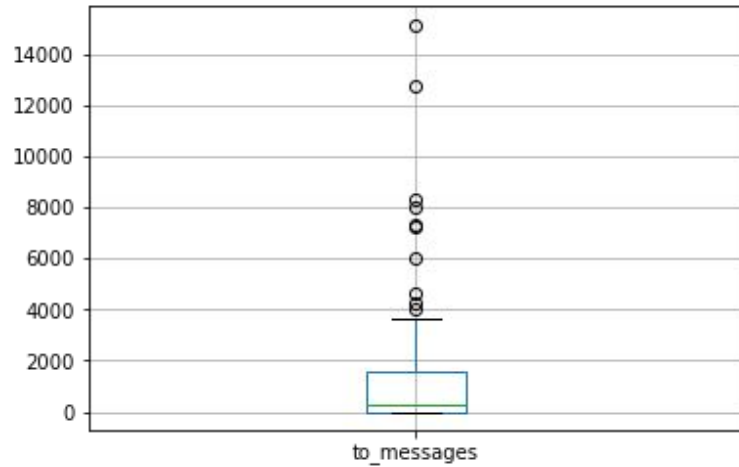
I will first do some EDA on the dataset and then Apply various ML models and finally compare their results.
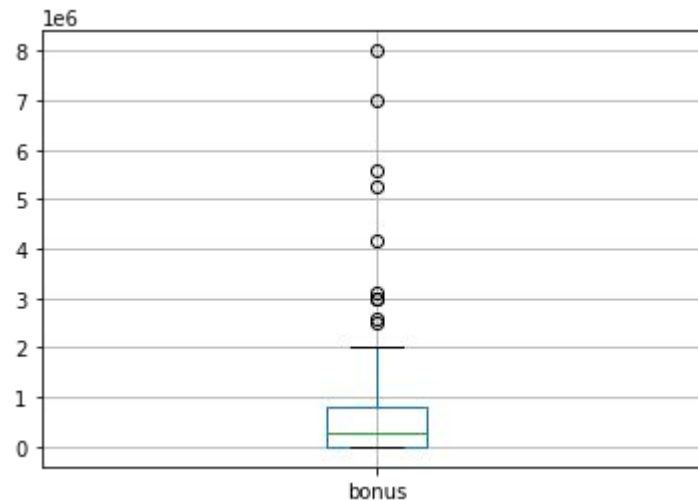
Outliers:

- LAY KENNETH L
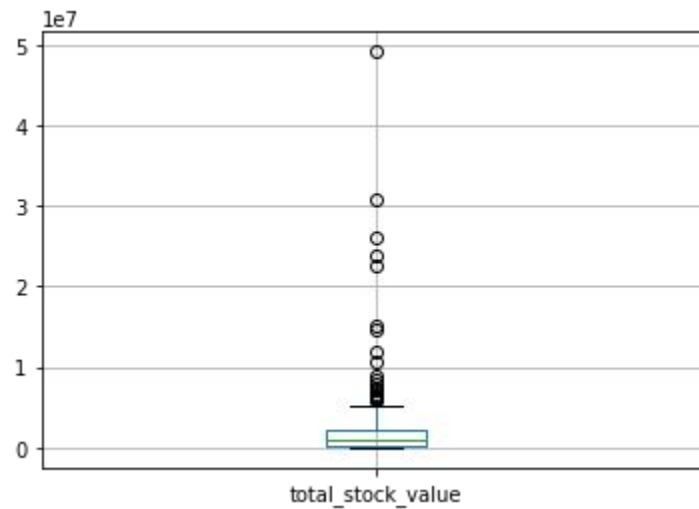- SKILLING JEFFREY K
- FREVERT MARK A

Highest mail sent and received by:

'BECK SALLY W', 'KAMINSKI WINCENTY J', 'KEAN STEVEN J', 'LAVORATO JOHN J'
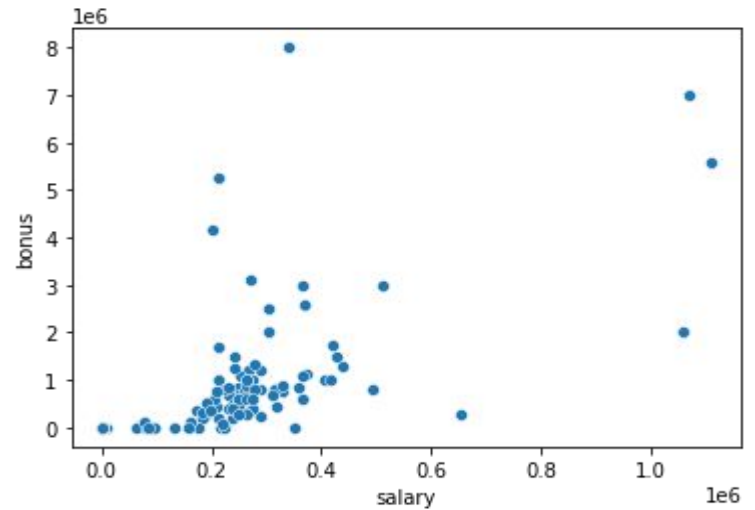
Outliers in bonus:

'KITCHEN LOUISE','LAVORATO JOHN J','BELDEN TIMOTHY N','ALLEN PHILLIP K'

Anomalies in stock values:

'LAY KENNETH L', 'HIRKO JOSEPH', 'SKILLING JEFFREY K', 'RICE KENNETH D'
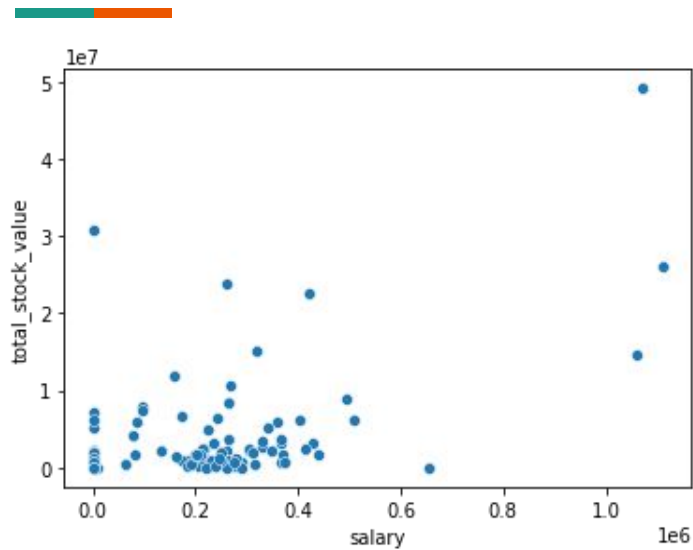
Outliers in the above plots:

'LAY KENNETH L', 'RICE KENNETH D', 'SKILLING JEFFREY K', 'HIRKO JOSEPH', 'PAI LOU L'

POI in Enron Dataset

'LAVORATO JOHN J': more bonus than CEO , low salary

'LAY KENNETH L': Highest stock and extreme outlier in plots like salary and stock_values

'PAI LOU L' : less salary High stock_value

'RICE KENNETH D': less salary High stock_value

'KEAN STEVEN J':   highest mail sent and received

'HIRKO JOSEPH': 2nd Highest stock, outlier in plot of salary vs stock_value

'SKILLING JEFFREY K': high bonus,3rd highest stock

'BELDEN TIMOTHY N': 4th person with huge bonus and low salary

'ALLEN PHILLIP K': 5th person with huge bonus and low salary

ML Models

SVM using auto grid-search

Best Parameters are: {'C': 10, 'degree': 3, 'kernel': 'linear'}, best f1 : 0.3

Test Classification Result:

	precision	recall	f1-score	support
0.0	0.91	0.89	0.90	44
1.0	0.17	0.20	0.18	5

accuracy			0.82	49
macro avg	0.54	0.54	0.54	49
weighted avg	0.83	0.82	0.82	49

test auc-roc: 0.5431818181818181

training Score: 0.7000000000000001

RandomForest

	precision	recall	f1-score	support
0.0	0.91	0.93	0.92	44
1.0	0.25	0.20	0.22	5
Test accuracy			0.86	49
macro avg	0.58	0.57	0.57	49
weighted avg	0.84	0.86	0.85	49

test auc-roc: 0.5659090909090909
training Score: 1.0

GNB

	precision	recall	f1-score	support
0.0	0.91	0.91	0.91	44
1.0	0.20	0.20	0.20	5
Test accuracy			0.84	49
macro avg	0.55	0.55	0.55	49
weighted avg	0.84	0.84	0.84	49

test auc-roc: 0.5545454545454545
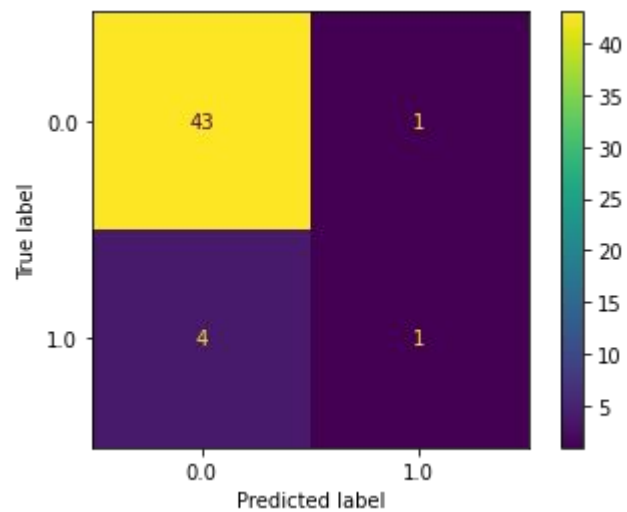training Score: 0.8556701030927835

SVM

	precision	recall	f1-score	support
0.0	0.91	0.98	0.95	44
1.0	0.50	0.20	0.29	5
Test accuracy			0.90	49
macro avg	0.71	0.59	0.62	49
weighted avg	0.87	0.90	0.88	49

test auc-roc: 0.5886363636363636
training Score: 0.9278350515463918

Confusion Matrix

DecisionTreeClassifier

	precision	recall	f1-score	support
0.0	0.90	0.80	0.84	44
1.0	0.10	0.20	0.13	5
Test accuracy			0.73	49
macro avg	0.50	0.50	0.49	49
weighted avg	0.82	0.73	0.77	49

test auc-roc: 0.4977272727272727
training Score: 1.0

AdaBoost

	precision	recall	f1-score	support
0.0	0.93	0.91	0.92	44
1.0	0.33	0.40	0.36	5
accuracy			0.86	49
macro avg	0.63	0.65	0.64	49
weighted avg	0.87	0.86	0.86	49

test auc-roc: 0.6545454545454545
Score: 1.0

Confusion Matrix
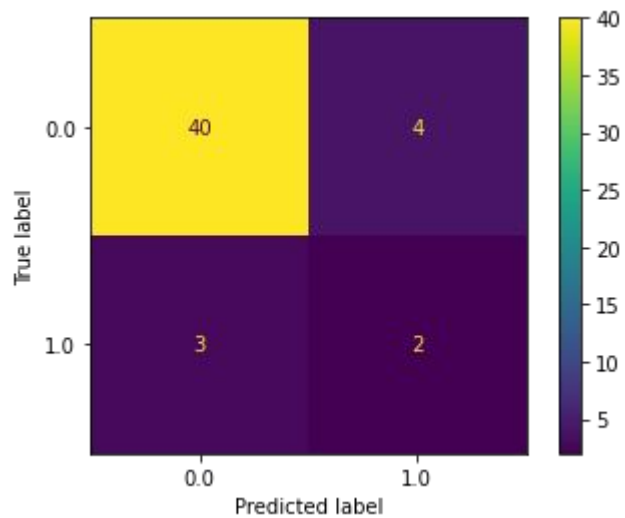
Stochastic Gradient Boosting

Cross_validation_score = 0.8566666666666667

	precision	recall	f1-score	support
0.0	0.93	0.95	0.94	44
1.0	0.50	0.40	0.44	5
Test accuracy			0.90	49
macro avg	0.72	0.68	0.69	49
weighted avg	0.89	0.90	0.89	49

test auc-roc: 0.6772727272727272
Training Score: 1.0

Confusion Matrix
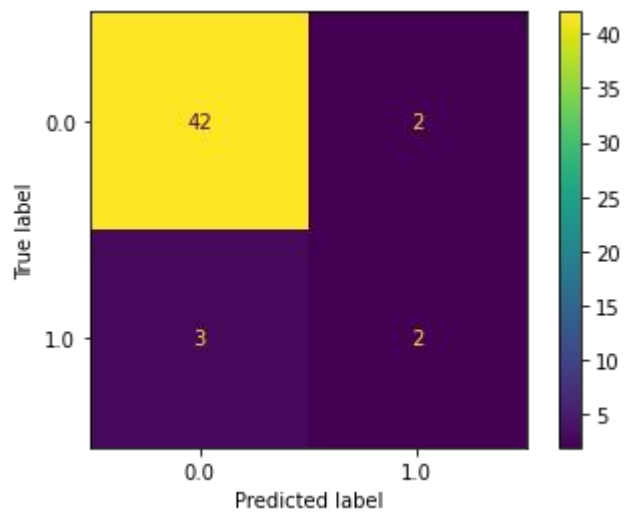
Finally

The best results are given by "Stochastic Gradient Boosting Classification Model" with

Test accuracy of 90% and 68% AUC-ROC score.

For more information do visit the code: Here