

Identify Fraud from Enron Dataset

Project summary

The goal of the project was to train a machine learning model to identify key persons in the Enron Fraud.

Overall description

I got many features of the Enron employees like salary bonus, stock options, etc. Besides that, I got some other features like the connection between employees through e-mail. There was one more feature, the so-called POI label. It designates that a person has a significant role in the fraud.

This dataset is suitable for supervised learning algorithms because it serves as a labeled dataset.

Benefit of ML on this project

If we could build a good model maybe we could prevent frauds at other companies, e.g., we create a model which can predict possible fraud based on money transfer.

Outlier identification and handling

Feature creation and selection

Selection of algorithm

Parameter tuning

Validation

Evaluation