# Enron POI Identifier Report

## Introduction

The project explores algorithmic classifiers with the objective of finding persons of interest, POI, from public Enron financial and email data.

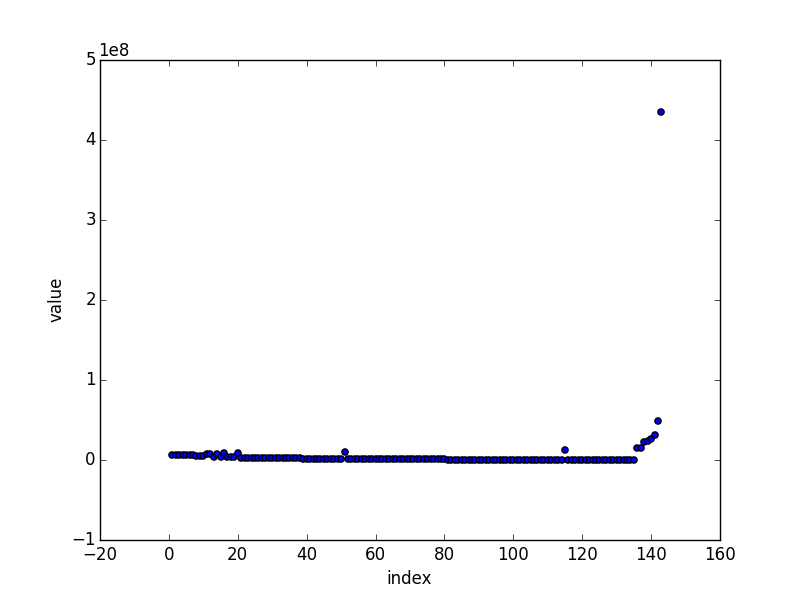
## The Enron Data

The data contains 20 features on 145 persons with a subset of 18 marked with the additional POI flag. POI are those who were indicted, reached a settlement, involved in plea deals with the government, or testified in exchange for prosecution immunity. Some features did not contain data for every person.

Various supervised machine-learning algorithms are explored with this data to find the optimal classifier for the POI flag based on selected features.

## Removing Outlier Data

There was one obvious outlier in the data—the total combined values for all persons. The figure below shows the data with outlier at the top of the chart for the feature “total\_payments.” This outlier was removed for testing.



### Feature Transformation

### Feature Selection—First Round

To select initial features, two processes were tried: PCA and decision tree importance.

**PCA**

The eigenvalues for all features of the first PCA component is analyzed. Four features are chosen in the final selection:

* total\_payments: 0.68172
* loan\_advances: 0.51338
* total\_stock\_value: 0.40506
* exercised\_stock\_options: 0.29293

**Decision Tree Importance**

A decision tree is analyzed with all possible combination of up to 4 features. The sum of importances of each feature is recorded. See task\_6.py. The importances of features with the best ten performing set of features are:

* exercised\_stock\_options, 5.4425149427154835
* fraction\_to\_shared\_with\_poi, 2.7504467362191813
* total\_payments, 1.2691611397988012
* long\_term\_incentive, 0.42577529949934501
* total\_stock\_value, 0.11210188176719012

In addition to the four features from PCA, the fraction\_to\_shared\_with\_poi feature is chosen for the final list of features.

### Algorithm Selection and Tuning

## Analysis Validation and Performance

## Discussion

## Conclusions

{5: {'accuracy': 0.8595413333333335,

'precision': 0.46938363632254854,

'recall': 0.40944},

10: {'accuracy': 0.8928266666666669,

'precision': 0.6068773603690502,

'recall': 0.5570799999999999},

15: {'accuracy': 0.8933066666666665,

'precision': 0.6093022134758322,

'recall': 0.55692},

20: {'accuracy': 0.8940280000000002,

'precision': 0.6130747194836178,

'recall': 0.55634},

25: {'accuracy': 0.8941346666666665,

'precision': 0.6135506874808744,

'recall': 0.5566},

30: {'accuracy': 0.8942359999999998,

'precision': 0.6141462020311311,

'recall': 0.5562699999999999},

35: {'accuracy': 0.8937466666666667,

'precision': 0.6130932798865925,

'recall': 0.5505399999999998},

40: {'accuracy': 0.8800106666666666,

'precision': 0.5770199186534902,

'recall': 0.37493999999999983},

45: {'accuracy': 0.8720866666666669,

'precision': 0.5395876905598371,

'recall': 0.27713999999999994},

50: {'accuracy': 0.8728360000000002,

'precision': 0.5458359508619689,

'recall': 0.27559999999999996}}