**Project Overview**

In 2000, Enron was one of the largest companies in the United States. By 2002, it had collapsed into bankruptcy due to widespread corporate fraud. In the resulting Federal investigation, a significant amount of typically confidential information entered into the public record, including tens of thousands of emails and detailed financial data for top executives. In this project, I will play detective, and put my new skills to use by building a person of interest identifier based on financial and email data made public as a result of the Enron scandal. To assist me in my detective work, the data combined with a hand-generated list of persons of interest in the fraud case, which means individuals who were indicted, reached a settlement or plea deal with the government, or testified in exchange for prosecution immunity.

**Summary**

The goal of this project is to develop an algorithm that could have identified people of interest related to its eventual financial disaster. This project will use machine learning with supervised learning since we know the outcome of guilty parties. The data used in this project is comprised of many financials and emails from dozens of former Enron employees. Financial data includes categories such as salary, bonus, exercised stock options, etc. The emails data is all the available email traffic to or from a former employee.

Two outliers were identified and removed: *TOTAL* and *THE TRAVEL AGENCY IN THE PARK (TAP)*. Both were removed mainly because they were not people. TOTAL is obviously not relevant for determining specific people engaged in financial misconduct. TAP was removed as we are not focused on potential companies of financial misconduct—only people. TAP was 50% owned by Sharon Lay, Ken Lay’s sister. TAP received commissions for managing travel books for Enron.[[1]](#footnote-1) While the family connection may be suspicious, it is outside the scope of the project.

**Features**

Top 8 features and scores using SelectKBest:

#1 exercised\_stock\_options with score 24.8150797332

#2 total\_stock\_value with score 24.1828986786

#3 bonus with score 20.7922520472

#4 salary with score 18.2896840434

#5 deferred\_income with score 11.4584765793

#6 long\_term\_incentive with score 9.92218601319

#7 restricted\_stock with score 9.21281062198

#8 total\_payments with score 8.77277773009

**Algorithm**

**Parameter Tuning**

**Validation**

**Metrics**

1. Texas agency takes a huge hit from Enron's fall. March 5, 2002. http://www.travelweekly.com/Travel-News/Travel-Agent-Issues/Texas-agency-takes-a-huge-hit-from-Enron-s-fall [↑](#footnote-ref-1)