

Propuesta de Proyecto

Jhonny Stuart Lorenzo Rojas

12 Septiembre 2018

1 Nombre del proyecto

The Enron Email Dataset: *500,000+ emails from 150 employees of the Enron Corporation*

2 Proyecto a utilizar

El conjunto de datos a tratar fue obtenido por la Comisión Reguladora de Energía Federal durante la investigación del colapso de **Enron** a raíz de las malas prácticas realizadas por altos ejecutivos de dicha corporación, el cual contiene correos electrónicos, metadatos de los correos electrónicos, número de correos electrónicos enviados y recibidos por cada persona, la información financiera (salarios y operaciones realizadas con las acciones).

3 Objetivo del proyecto

Predecir, mediante técnicas de Machine Learning, personas implicadas en casos fraude o malas prácticas comerciales *-basándonos en los metadatos y la información financiera-* con la finalidad de reducir posibles complicaciones en auditorias y/o investigación criminal.

4 Artículos científicos relevantes

5 Propuesta tentativa de modelos a utilizar

KMeans. Empleado para separar por categorías a las personas implicadas y no implicadas en el fraude.