# Audit risk Prediction

## Problem statement:

Predict the fraudulent firm on the basis of the present and historical risk factors.

- perform data exploration, preprocessing and visualization
- implement classification model using sklearn library
- evaluate the model using appropriate performance metrics
- develop the Audit Risk  Prediction system.

## About Dataset:

building a classification model that can predict the fraudulent firm on the basis of the present and historical risk factors. The information about the sectors and the counts of firms are listed respectively as Irrigation (114), Public Health (77), Buildings and Roads (82), Forest (70), Corporate (47), Animal Husbandry (95), Communication (1), Electrical (4), Land (5), Science and Technology (3), Tourism (1), Fisheries (41), Industries (37), Agriculture (200).

Attribute Information:

Many risk factors are examined from various areas like past records of audit office, audit-paras, environmental conditions reports, firm reputation summary, on-going issues report, profit-value records, loss-value records, follow-up reports etc. After an in-depth interview with the auditors, important risk factors are evaluated and their probability of existence is calculated from the present and past records.

Data Set Characteristics: Multivariate

Number of Instances:777