Earnings Management

Instructions:

- You have to use Python/R for the complete analysis. Use Jupyter notebook/RStudio for documentation and python/R code.

Earnings management occurs when managers use their judgment in financial reporting and in structuring transactions to alter financial reports to mislead stakeholders about the underlying economic performance of the company. As on December 2016, the total number of listed companies in Indian stock exchange was approximately 5622. Securities and Exchange Board of India (SEBI) reported that approximately 3.14% of the Indian companies are involved in earnings manipulation.

We will be using Earnings Manipulation data (fraud_data.csv) in this exercise. Refer the **Exhibit 1** to understand the feature list. Use the data and answer the below questions.

1. Load the dataset in Jupyter Notebook/RStudio.
2. Build a correlation matrix between all the numeric features in the dataset. Report the features, which are correlated at a cut-off of 0.70. What actions will you take on the features, which are highly correlated?
3. Create a new data frame with the numeric features and categorical features as dummy variable coded features. Which features will you include for model building and why?
4. Split the data into training set and test set. Use 80% of data for model training and 20% for model testing.
5. Use sklearn package (Python)/Caret Package (R) to build a random forest model using `Status` as a dependent variable and all other features as independent variable. Report the model performance on the test set.
6. Python Specific: Use sklearn model selection module to fine-tune the model parameters of random forest model. Report the model performance on the test set.
7. R Specific: Use caret package to fine-tune the model parameters of random forest model. Report the model performance on the test set.

Exhibit 1

Sl. No.	Name of Variable	Variable Description
1	Company ID	Unique Identifier
2	DSRI	Days' Sales in Receivables Index (DSRI): A large increase in receivable days might suggest accelerated revenue recognition to inflate profits.
3	GMI	Gross Margin Index (GMI): A deteriorating gross margin sends a negative signal about a firm's prospects and creates an incentive to inflate profits.
4	AQI	Asset Quality Index (AQI): An increase in long term assets (for example, the capitalisation of costs), other than property plant and equipment, relative to total assets indicates that a firm has potentially increased its involvement in cost deferral to inflate profits.
5	SGI	Sales Growth Index (SGI): High sales growth does not imply manipulation but high growth companies are more likely to commit financial fraud because their financial position and capital needs put pressure on managers to achieve earnings targets. If growth firms face large stock prices losses at the first indication of a slowdown, they may have greater incentives to manipulate earnings.
6	DEPI	Depreciation (DEPI): A falling level of depreciation relative to net fixed assets raises the possibility that a firm has revised upwards the estimated useful life of assets, or adopted a new method that is income increasing.
7	SGAI	Sales, General and Administrative Expenses (SGAI): Analysts might interpret a disproportionate increase in SG&A relative to sales as a negative signal about a firm's future prospects, thereby creating an inventive to inflate profits.
8	ACCR	Accruals to Total Assets (ACCR): Total accruals are calculated as the change in working capital (other than cash) less depreciation relative to total assets. Accruals, or a portion thereof, reflect the extent to which managers make discretionary accounting choices to alter earnings. A higher level of accruals is, therefore, associated with a higher likelihood of profit manipulation.
9	LEVI	Leverage Index (LEVI): Leverage is measured as total debt relative to total assets. An increase in leverage creates an incentive to manipulate profits in order to meet debt covenants.
10	Status	Manipulator – Yes, Non Manipulator – No