**Financial Reporting Dataset**

The included datasets include synthetic data from 4 quarterly reports, where some of the datasets are correct and some contain errors. In this example the number of fields is limited, but it is desirable to find machine learning techniques to automatically understand how to validate arbitrary sets of data, some of which may possibly include many fields. For example, an algorithm might note that a given column always has data between 0 and 1, likely indicating that the value is a rate, while in another instance it might note a correlation relationship between different columns of data.

* Training datasets = Report 1 & 2 represents accurate dataset.
* Test & validation dataset = Report 3 & 4 contain errors.

|  |  |
| --- | --- |
| Field | Description |
| rating | Rating is a categorical variable designated in letters (“A”,”B”,”C”,”D”,”E”). Ratings are assigned to each obligor and correspond to a 1 year probability of default. |
| id | Sequence id of data record |
| reporting\_date | Date the report was generated |
| assets | Assets are reported on a company's balance sheet, and they are bought or created to increase the value of a firm or benefit the firm's operations. |
| profits | Profit is an economic indicator that calculates net income |
| maturity\_date | Maturity date refers to the date on which the principal and interest associated with a debt security must be repaid |
| state | Headquarter state of the obligor |
| pd | Probabilities of default |
| liabilities | Liability is a the financial debt or obligations that arise during the course of a company’s business operations |
| assets\_short\_term | Short term asset is an asset that is to be sold, converted to cash, or liquidated to pay for liabilities within one year |
| assets\_long\_term | Assets that are not intended to be turned into cash or be consumed within one year of the balance sheet date |
| equity | By definition equity is equal to assets less liabilities |
| sales | Total sales in given quarter |
| employees | Total number of employees |