

The repository contains three MATLAB files, MATLAB version 2024b. The functionality features are explained below:

- **MCS_code.m** – This code uses given parameter values to generate Monte Carlo Simulation (MCS) data for randomised investment and interest returns. The outputs are respectively presented in files ‘investment_growth.xls’ and annual_returns.xls’. The code can be used to generate any number of MCS data for different sets of parameter values (lines 5-8). The inputs are drawn from the SMEs and simulated to clone their behavioral data.
- **rank_NN.m** – Since original company data cannot be disclosed due to copyright (IP) protection, MSExcel simulation was used to perturb the original data around its mean and within 2 standard deviations, inputs drawn from a normal distribution. Such data, e.g. **Dataset6.xlsx** (discussed later) is used as an input to this code. This Matlab code (rank_NN.m) uses a Recurrent Neural Network to rank variables from the given dataset (e.g. Dataset6.xlsx) for which the first 14 columns represent independent variables and the last 7 columns represent variables dependent on the first 14 columns. Display relative contributions in percentages in descending order and save the output as an xls file. The code also ranks the top 5 independent variables for each of the top 5 dependent variables. In our enumeration, to train the network, we have used two hidden feedforward layers with 10 neurons each over 100 epochs with lower range and median ranges set at 0.01 and 0.9 respectively. All of these can be changed as per requirements.
- **rank_using_random_forest.m** – Using a reordered version of Dataset6.xlsx (e.g. Dataset2_RF.xlsx) as input, where both datafiles are identical apart from a reordering of columns based on company inputs on prioritized variables, this Matlab code uses random forest to rank variables from a dataset for which the first 14 columns represent independent variables and the last 7 columns represent dependent variables (reordered from Dataset6.xlsx). The relative contributions are displayed in percentages in descending order and saved as an xls file (together with plots). The code first ranks the top 5 dependent variables and then ranks the top 5 independent variables for each of these 5 dependent variables.

The repository also contains a pair of datasets, respectively for recurrent neural network evaluation (rank_NN.m) and random forest-based evaluation (rank_using_random_forest.m). The former set of data are housed under subfolder ‘NN DATASETS’ while the latter set are under subfolder ‘RF DATASETS’. Details below:

- **NN DATASETS** – Contains 6 sets of datafiles (first 14 columns represent independent variables, the following 7 columns represent variables dependent on the first 14), generated from the original dataset (undisclosed due to copyright) using Monte Carlo simulation. These datasets are used for recurrent neural network evaluation.
- **RF DATASETS** – Contains 6 sets of datafiles (first 14 columns represent independent variables, the following 7 columns represent variables dependent on the first 14), generated from the original dataset (undisclosed due to copyright) using Monte Carlo simulation. These datasets are used for random forest evaluation.