The **problem statement:** the purpose of this project is to use the machine learning techniques of a regression decision tree and an artificial neural network to understand which stock features are the most crucial to predict a stock's future nominal return upon its sale. The variables investigated include the volatility of the stock at the time of its purchase, the return on equity ratio, the ESG ranking, the return on asset ratio, current ratio(current assets to current liabilities) of the company, the net annual profit, and the earnings per share ratio of the company that issued the bought stock. The accuracy of the regression decision tree(RDT) and the Artificial neural network(ANN) will be measured using mean squared error(MSE). The greater this value is, the less accurate the predictive model is. In addition, the project is also meant to compare the accuracy of the predictions of future nominal return made by a regression decision tree or the artificial neural network.

1.) An artificial neural network model and a regression decision tree will be made using all the variables mentioned and additional necessary variables The dataset used will be broken up into a training dataset and a validation dataset. The neural network dataset will be trained using the training dataset and validation dataset. Once, the model is thoroughly trained, it will be fed the testing dataset. This mean squared error value that results from applying the model on the testing data will be the value used to measure the accuracy of the model. This model will be called the base model.
2.) Then, copies of these models will be made. Each copy will be modified to remove one of the previously mentioned variables. All other aspects of the model will remain the same. After the model is trained and run the MSE of the model will be recorded.
3.) The third step is to rank the models by accuracy using the mean squared error. The MSE resulting from the base model is compared to that of the other altered copies. The higher the MSE of the copied model is   compared to the base model, the more significant the variable that the copied model is missing. If the MSE is less than that of

the base model, this indicates that the variable does not contribute to the accuracy of the prediction model.

Ranking of variables according to neural network:

1.) Current ratio: MSE: 0.0014

2.) Roe ratio MSE: 8.9251*10^-4

3.) Esg_ranking: 8.0753:*10^-4

4.) Volatility_buy: 8.0753*10^-4

5.) Net profit: 7.5917*10^-4

6.) More accurate without these variables below

7.) Eps: 7.1554*10^-4

8.) Roa: 6.6753*10^-4

Rank according to RDT:

1.) Volatility: 8857.498*10^-4

2.) ESG: 8760.6 *`10^-4

3.) ROA: 301.1*10^-4

4.) ROE 0.03070199999999998

5.) Normal: 295.98 *10^-4

6.) Current : 180.7639 *10^-4

7.) Netprofit: 176.61*10^-4

8.) EPS: 37.59 *10&-4

Analysis of the Data:

We see that the rankings made by the regression tree models differ from that of the Artificial neural network. Our next step is to decide which model we can trust by comparing the performance of both types of models for the dataset that was used to train and test them.
In order to compare the accuracy of the RDT and the ANN methods, we can only compare RDT model with the ANN models with the same input variables.(the models missing the same variable).

Upon observation, we can establish that all neural network models outperform their Rregression decision tree counterparts in terms of accuracy.
This tells us that the ranking of variables established by the artificial neural network is more accurate than the ranking indicated by the regression decision tree.

The ANN's ranking indicates that Volatility of the stock at the point of buying, the ESG ranking and the Re

If we observe the MSE values of the RDT models and the MSE of the ANN, the
In order to understand whether the RDT model or the ANN model is more accurate,

- analyzed a Kaggle dataset posted by Imanol Recio Erquicia with over 350k stock trades to explore which stock feature (current ratio, net yearly profit) is most effective in predicting future nominal return using machine learning techniques(regression decision tree and neural network)
    - compared effectiveness between regression decision tree and neural network model to assess the more accurate prediction technique using means squared error

We use a kaggle dataset of over 250k

When observing the results of the neural network models and the regression tree models, we can compare models with the same variables used as inputs. (models that exclude the same variable).
This approach helps us realize that the neural network predicts with far greater accuracy than its regression tree counterpart when using the same variables as inputs.

We can than conclude that the neural network models are the more reliable in assessing hte significance of each stock feature when trying to predict the future nominal return.

Both models indicate that EPS is the least helpful variable

After assessing the Mean squared errors of the neural network models and the MSE of the tree models, we can tell that the neural network's models have a much greater accuracy than the regression tree even when using the same inputs.

This indicates that the results of the neural network are a lot more reliable in understanding the significance of the each variable in predicting the nominal return