**Stocks data report- Hagay Ringel (May 23, 2023)**

In this project I used a dataset related to stocks from Kaggle.

This project aimed to assess whether the factors of “Total debt” and “Total assets” have a relationship and influence on the binary variable “Buy worthy.”

The methodology is preparing the dataset including checking for nulls which we have. There are many ways to handle nulls. I used three techniques: Decreasing the nulls, dropping them, and changing them to 0, there are more other techniques but just for the purpose of getting the model done I used these three.

Then, for design reasons I renamed the columns of stock symbols and is the stock worth buying (yes or no) to a more readable name.

The next step was preparing the dataset to our machine learning model using the algorithm Logistic Regression, for the following reasons:

1. LR is an algorithm that fits binary classification problems like this.
2. LR is simple to understand.
3. LR can provide insight into the importance of different features in determining the outcome.

First, we split the data to training set and testing set on the ratio of 70% training and 30% testing.

Second, we made predictions based on the testing set.

Third, producing the confusion matrix.

The last step was creating a classification report which is basically the summary of our model.

According to our model these are the results:

A computer screen shot of a computer screen

Description automatically generated with low confidence

1. Precision- For class 0 (Should not buy), the precision is 0.00, indicating that no instances were correctly predicted as class 0. For class 1 (Should buy), the precision is 0.70, meaning that 70% of the instances predicted as class 1 were correct.
2. Sensitivity (Recall)- Measures the model's ability to correctly identify positive instances. Both class 0 and class 1 have a recall of 1.00, indicating that all instances of class 1 were correctly predicted as class 1, while class 0 was not correctly predicted at all.
3. F1-score is the mean of precision and recall, providing a balanced measure of the model's performance. In this case, the F1-score for class 0 is 0.00, while for class 1, it is 0.82.
4. Accuracy- The accuracy of the model is 0.70, indicating that it correctly predicted the class for 70% of the instances.

Note: This Project’s goal is to assess whether we can lean on factors such as total debt and total assets for buying stocks. According to the model we have a good (but not excellent) estimate as for the stock that we should buy as these factors correlated to the variable of buy worthy, but we don’t know if these factors have an impact of the stocks that we should not buy. I believe the poor results we received for class 0 (should not buy), are due to the lack of these stocks in our database, bias due to the null cleansing so it is possible that the nulls that I changed to 0 were under the classification of “Should not buy”.