# Data Science Test

## 2. ML methodology

### **Question 2.1**

**A new classifier model identifies bad bonds in the financial market for a hedge fund. Bad bonds can have devastating effects and must be avoided in the portfolio. 0.01% of all bonds fall into this category and our model has an accuracy of 99.99%. Is this ML model doing a good job? Why?**

I’m not sure if it’s good. If the accuracy is this good, and we have a highly imbalanced data like this, we should see other metrics to see the whole picture (for example a confusion matrix to see all the four outcomes). Or Precision, Recall, or the F1 score, which is an avarage of the previous two.

If we simply label every bond as not bad bonds (good bond), we could easily get the 99.99% accuracy, which is misleading.

**Fill in the empty confusion matrix below with a possible concrete outcome if there are 100,000 bonds in the market.**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Actual | |
|  |  | Bad | Not bad |
| Predicted | Bad | *10* | *0* |
| Not bad | *10* | *99980* |

### Question 2.2

**On a logistic regression model with binary outcome in {0,1} that is optimized with stochastic gradient descent you have to tune hyperparameters**

* **learning rate**
* **L2 regularization**
* **batch size**
* **threshold value: the predicted probability above which we assign 1**

**Choose 3 metrics that you can use to compare the trained model and decide which one is the best for this use case. Explain why.**

- Confusion matrix

- F1 Score

- AUC-ROC

I would use AUC-ROC, because that’s the most complex, and can distinguish between classes, and calculate with TPR and FPR.