A2.2 The Confusion Matrix in a Binary Classification Problem

For the purposes of the PhD thesis, ML is applied to a binary classification problem, with Class 0 for non-survivors and Class 1 for Survivors.

The confusion matrix plays a fundamental role in evaluating the performance of a machine learning model, especially in binary classification problems. It provides a detailed breakdown of how the model's predictions compare to the actual true labels. The confusion matrix consists of four main components:

1. True Positives (TP): The number of instances where the model correctly predicted the positive class (e.g., correctly identifying people with a disease as having the disease).
2. True Negatives (TN): The number of instances where the model correctly predicted the negative class (e.g., correctly identifying people without a disease as not having the disease).
3. False Positives (FP): The number of instances where the model incorrectly predicted the positive class when the true class was negative (e.g., incorrectly classifying healthy individuals as having a disease).
4. False Negatives (FN): The number of instances where the model incorrectly predicted the negative class when the true class was positive (e.g., incorrectly classifying individuals with a disease as healthy).

Definition: Specificity-->True Positive Rate
Sensitivity-->True Negative Rate

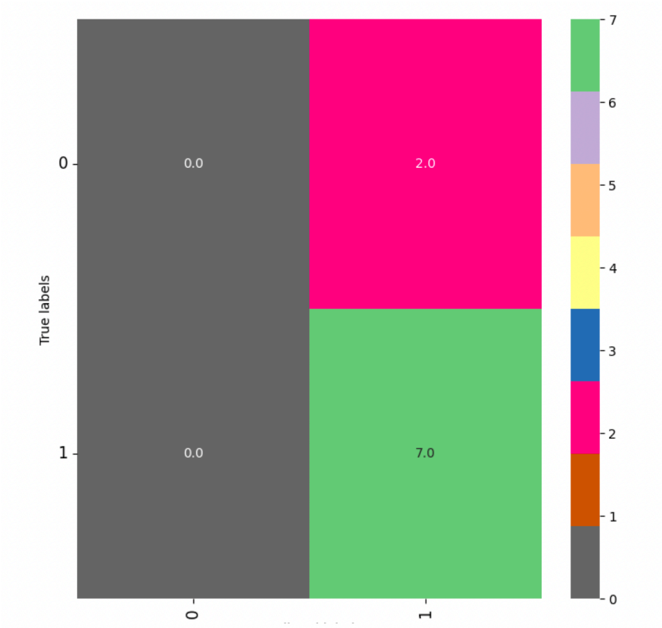
The confusion matrix allows one to compute various performance metrics to assess the model's effectiveness, including:

1. **Accuracy:** It measures the overall correctness of the model's predictions and is calculated as (TP + TN) / (TP + TN + FP + FN). However, accuracy may not be the best metric if the classes are imbalanced.
2. **Precision:** Also known as Positive Predictive Value, precision measures the

proportion of true positive predictions out of all positive predictions made by the model. It is calculated as TP / (TP + FP). Precision is valuable when minimising false positives, which is crucial.
3. **Recall:** Sensitivity or True Positive Rate (TPR), recall measures the proportion of true positive predictions out of all actual positives. It is calculated as TP / (TP + FN). Recall is important when identifying as many positives as possible is vital.
4. **F1-Score:** The F1-Score is the harmonic mean of precision and recall and is useful when there's a trade-off between precision and recall. It is calculated as 2 * (Precision * Recall) / (Precision + Recall).
5. **Specificity:** Specificity measures the proportion of true negative predictions from all actual negatives and is calculated as TN / (TN + FP). It is particularly relevant when minimising false negatives is critical.

By examining these metrics and analysing the confusion matrix, one can gain insights into the model's strengths and weaknesses. For example, a high-precision but low-recall model may be conservative in making positive predictions. In contrast, a high recall but low precision model may make more positive predictions but have more false positives. Choosing the right metric depends on the specific requirements of your problem and the associated costs or implications of false positives and false negatives.

A confusion matrix may be depicted as below.



This then corresponds to the following diagrammatic representation:-

The columns show the predicted data and the rows the true (actual data)

		[Predicted]	
		0	1
[True	0	TN	FP
Labels]	1	FN	TP

TN (True Negatives) is in the top-left cell (row 0, column 0).

FP (False Positives) is in the top-right cell (row 0, column 1).

FN (False Negatives) is in the bottom-left cell (row 1, column 0).

TP (True Positives) is in the bottom-right cell (row 1, column 1).

By examining these counts in the confusion matrix, one can calculate various evaluation metrics such as accuracy, precision, recall, and F1-score to assess the performance of a binary classification model. These metrics provide insights into how well the model correctly classifies instances and makes predictions.

Precision, Recall in Binary classification problems

When discussing matrix "A" generated from the classification problem, one must consider the crucial parameters such as precision and recall. Precision and recall are fundamental metrics in binary classification that shed light on a machine learning model's performance. The choice of which metric is most salient is contingent upon the nature of the problem being tackled and the inherent trade-offs deemed acceptable. Precision evaluates the correctness of positive predictions, determining the proportion of correctly predicted positive outcomes against all positive predictions. When high precision indicates a decrease in the occurrence of false positives. On the contrary, recall measures the model's capacity to identify all genuine positive cases correctly. A high recall indicates that the model can correctly identify positive cases, resulting in fewer false negatives. In making a choice between precision and recall, it is important to align with the specific requirements of the problem. Precision is prioritised for scenarios where the accuracy of positive predictions is paramount. However, recall becomes more crucial if the aim is to ensure no genuine positive case is overlooked. The F1-score is a metric that amalgamates precision and recall, aiming to strike a balance between them. One should select a metric that resonates with the implications of false positives and negatives in a given application.

To understand the relationship between matrix "A" and the confusion matrix "B", it's essential to delve deeper into the essence of a confusion matrix. The confusion matrix, typically categorised into four segments, offers a snapshot of the classification results: True Positives (TP) are instances where the model's prediction of a positive outcome aligns with the actual positive outcome; True Negatives (TN) signify correct predictions of negative outcomes; False Positives (FP) emerge when the model wrongly predicts positive outcomes for instances that are negative; and False Negatives (FN) indicate incorrect negative predictions for cases that are genuinely positive. The values represented in the confusion matrix are pivotal for performance evaluation. By leveraging these values, metrics like precision, recall, and the F1-score are computed to determine the model's efficacy in distinguishing data into appropriate classes. The 'support' parameter within the confusion matrix signifies the volume of data points for each class within the dataset. In this context, it highlights two instances corresponding to class 0 (non-survived) and seven instances associated with class 1 (survived) utilised to gauge the machine learning model’s performance.

Precision: Focuses on the accuracy of positive predictions and is less affected by the imbalance in the data.
Recall (Sensitivity): Measures the ability of the model to identify positive instances, regardless of the class imbalance correctly.
F1-Score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance in both classes.
Area Under the Receiver Operating Characteristic Curve (AUC-ROC): This metric evaluates the model's ability to discriminate between positive and negative instances across different probability thresholds.
Area Under the Precision-Recall Curve (AUC-PR): This metric focuses on the

The confusion matrix and unbalanced data in a binary classification problem
When one has unbalanced data in a binary classification problem, where one class significantly outnumbers the other, it can affect the values in the confusion matrix and make it challenging to interpret the model's performance accurately. In such cases, the values in the confusion matrix may be skewed, and the model's performance may not be adequately represented.

Unbalanced data can affect the confusion matrix:

1. True Negatives (TN): True negatives may be overrepresented because the majority class (the one with more samples) dominates the dataset. This can lead to a higher count of TN.
2. False Positives (FP): False positives may be relatively low compared to the true negatives because the model might hesitate to make positive predictions due to the class imbalance.
3. False Negatives (FN): False negatives can be quite high because the model may not correctly identify the minority class, leading to many instances of incorrect classification as the majority class.
4. True Positives (TP): True positives may also be low due to the difficulty of correctly identifying the minority class.

The imbalance in data can make metrics like accuracy less informative because a model that always predicts the majority class will still achieve relatively high accuracy while performing poorly on the minority class.

To address this issue, one may consider using alternative evaluation metrics less sensitive to class imbalance.

Some of these metrics include:

When working with unbalanced data, choosing the right evaluation metric that aligns with one’s specific goals and priorities for the problem at hand is crucial. Additionally, techniques such as resampling (oversampling the minority class or undersampling the majority class) and using different machine learning algorithms can help improve the model's performance on imbalanced datasets.