

AutoML Modeling Report



<Chidimma Dominic Nweke>

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	The data was split in the ratio of 8:1:1, 8 for train, 1 for test, and 1 for validation. Therefore, out of the 200 data points (100 for each class), 160 was used for training, 20 for testing and the remaining 20 for validation.
Confusion Matrix What do each of the cells in the confusion matrix describe? What values did you observe (include a screenshot)? What is the true positive rate for the "pneumonia" class? What is the false positive rate for the "normal" class?	Each cell (bound by the true and predicted class row and column respectively) in the confusion matrix simply describes the number of predictions made by the model. The values observed are as follows: 90%, 10%, 10%, and 90% (screenshot is attached below). The true positive rate for the "pneumonia" is 90% while the false positive rate for the "normal" class is 10%. (NB: Values are reported for a score threshold of 0.5)
Precision and Recall What does precision measure? What does recall measure? What precision and recall did the model achieve (report the values for a score threshold of 0.5)?	The precision is a metric that measures the percentage of correct predictions against the total number of predictions. Recall measures the percentage of correctly identified instances against the total possible instances. The precision and recall achieved by the model on a score threshold of 0.5 are both 0.9. i.e., 90%.
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	Increasing the threshold to a value of 0.77 increases the precision with no effect on the recall. Further increase of the threshold increases the precision but recall starts to reduce. This is so because as the threshold increases, the model is prompted to prioritize the positive class and as the threshold increases over the optimum point, the model emphasizes more on the positive class thereby neglecting the ground truth labels, thus, reducing the recall.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	358 data points were used for training while 42 data points were used for testing.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	The confusion matrix showed that the model made more prediction of the "pneumonia" class than the "normal" class (screenshot is attached below)
Precision and Recall How have the model's precision and recall been affected by the unbalanced data (report the values for a score threshold of 0.5)?	On a score threshold of 0.5, the precision and the recall were not greatly affected as their values were both 88.1%, which was not very different from the model trained with a balanced data. Increasing the threshold up to 0.62 prompted an increase in the precision with no changes in the recall. while further increasing the threshold will cause only the precision to increase and the recall to decrease.
Unbalanced Classes From what you have observed, how do unbalanced classes affect a machine learning model?	The unbalanced classes have introduced a form of bias in the dataset. Thus, causing the model to favor the class with the higher number of data points.

Binary Classifier with Dirty/Balanced Data

Confusion Matrix How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.	The confusion matrix was greatly affected by the dirty. A striking observation is that no data point was classified as the "Normal" class (a screenshot of the confusion matrix is attached below).
Precision and Recall How have the model's precision and recall been affected by the dirty data (report the values for a	The precision and recall of the model were affected too. On a score threshold of 0.5, the precision and recall were observed to be both 71.43% and this was the optimum threshold. Any further increase resulted in a

score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?	<p>slight increase in the precision and a drastic decrease in the recall.</p> <p>Of the three binary classifiers, the one trained with clean and balanced data scored the values of 94.74% and 90% in precision and recall respectively.</p>
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	<p>From my observations, the dirty data caused the model to develop a bias against the “normal” class. Therefore, training a machine learning model on dirty and mislabeled data will prompt the same or similar behavior, causing the model to make poor decisions.</p>

3-Class Model

Confusion Matrix Summarize the 3-class confusion matrix. Which classes is the model most likely to confuse? Which class(es) is the model most likely to get right? Why might you do to try to remedy the model’s “confusion”? Include a screenshot of the new confusion matrix.	<p>From the confusion matrix, the classes which the model most likely confused were the “viral_pneumonia” and “bacterial_pneumonia” classes.</p> <p>The class which the model most likely got right was the “normal” class.</p> <p>The confusion is mostly due to the striking resemblance in the two pneumonia classes, (afterall, they are both showing pneumonia). This confusion can be reduced or remedied by adding more data points in both the two pneumonia classes to enable the model learn and identify more features and differences in the two classes, while maintaining the balance in the data by adding more data points of the normal class (a screenshot of the confusion matrix is attached below).</p>
Precision and Recall What are the model’s precision and recall? How are these values calculated (report the values for a score threshold of 0.5)?	<p>The values for the model’s precision and recall on a score threshold of 0.5 are both 93.55%. These values are calculated as follows:</p> <p>Precision: this is calculated by dividing the model’s correct predictions by its total predictions and then the taking the percentage.</p> <p>Mathematically: $\frac{\text{correct predictions}}{\text{total predictions}} \times 100\%$</p> <p>Recall: this is calculated by dividing the model’s correct predictions by the total possible ground truth and then taking the percentage.</p> <p>Mathematically: $\frac{\text{correct predictions}}{\text{ground truth}} \times 100\%$</p>
F1 Score What is this model’s F1 score?	<p>The model’s F1 evaluated using the mathematical expression: $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ is 93.55%</p>

True Label	Predicted Label	
	PNEUMONIA	NORMAL
PNEUMONIA	90%	10%
NORMAL	10%	90%

Screenshot_1: Confusion Matrix for First Model (Balanced Clean Data).

True Label	Predicted Label	
	PNEUMONIA	NORMAL
PNEUMONIA	91%	9%
NORMAL	20%	80%

Screenshot_2: Confusion Matrix for Second Model (Unbalanced Clean Data)

True Label	Predicted Label	
	PNEUMONIA	NORMAL
PNEUMONIA	94%	6%
NORMAL	100%	-

Screenshot_3: Confusion Matrix for Third Model (Dirty data)

True Label	Predicted Label		
	NORMAL	BACTERIAL_PNEUMONIA	VIRAL_PNEUMONIA
NORMAL	100%	-	-
BACTERIAL_PNEUMONIA	-	90%	10%
VIRAL_PNEUMONIA	-	9%	91%

Screenshot_4: Confusion Matrix for Fourth Model (Three Class Model)

