AutoML Modeling Report

Luluh Alyahya

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	This scenario, we have 200 images, includes 100 images for pneumonia and 100 images for normal cases. Of these, that contains 2 training labels, 158 training images, 2 test labels, 40 test image.
Confusion Matrix What do each of the sections 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?	A confusion matrix that is mean a tabular layout that visualizes the performance of a classification algorithm, true positive (TP), false negative (FN), true negative (TN), and false positive (FP).
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)?	Precision measures the percentage of correctly identified positive cases out of all the cases predicted as positive. Recall, however, measures the percentage of actual positive cases that were correctly identified. As the threshold value increases, precision generally improves, while recall tends to decline. This happens because a lower threshold results in the model classifying more images as positive, which can lead to a higher number of misclassifications. Conversely, with a higher threshold, the model becomes more selective, classifying fewer images as positive and thereby reducing the likelihood of misclassifications.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	This scenario, the number of total images is 300 used, includes 200 images for pneumonia and 100 images for normal cases. Of these, that contains Training dataset: 2 labels, 238 images, and Testing dataset: 2 labels, 60 images
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix summary	Because now i have unbalanced data so adding 100 more for pneumonia images to make it totally 200 pneumonia images and 100 images for normal images, the confusion matrix changed notably. Specifically, the number of true positives (TP) increased, indicating more accurate identification of pneumonia cases. However, the number of true negatives (TN) decreased, meaning the model became less accurate at identifying normal images correctly.
Precision and Recall How have the model's precision and recall been affected by the unbalanced data?	Indeed, both precision and recall values shifted after modifying the dataset in this scenario. With a threshold value of 0.5, both precision and recall have increased compared to the previous situation.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	The results clearly show that an imbalanced dataset introduces bias into the machine learning model. This means the model is more likely to classify input objects into the class that is more prevalent during training. Here, the model tends to classify more images as 'pneumonia' because there are more pneumonia images in the training set.

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 information.	In this case I alter the dataset and modifying the dataset to include balanced but mislabeled 'dirty' data, the confusion matrix deteriorated significantly. The performance of the ML model declined across all prediction aspects, clearly demonstrating the impact of incorrect and mislabeled data. This poor data quality caused the model to make erroneous predictions.

Precision and Recall How have the model's precision and recall been affected by the dirty data. Of the binary classifiers, which has the highest precision? Which has the highest recall?	The precision and recall values changed again after adjustments were made to the dataset for this case. At a threshold of 0.5. Additionally, a comparative analysis revealed that the 'ML Model trained on clean and unbalanced data' yielded the highest precision and recall scores.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	This leads to various misclassifications, as the inherent inaccuracies in the data contribute to the model's suboptimal performance.

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 information.	This **Confusion Matrix** presented here illustrates the classification outcomes for all three categories, akin to a 2 × 2 matrix, to evaluate the effectiveness of the machine learning model. It shows that the model frequently confuses the 'normal' and 'virus' categories, while it reliably identifies the 'bacteria' category. To enhance performance and minimize this confusion, increasing the number of training images for each category could enable the model to better recognize patterns for accurate predictions.
Precision and Recall What are the model's precision and recall? How are these values calculated?	Finally In this scenario involving the 3-class confusion matrix, the precision and recall values are obtained at a threshold score of 0.5. Furthermore, in the 3 × 3 scenario, the individual precision and recall values for each class are first calculated, followed by averaging these values. This approach allows for the determination of the overall precision and recall metrics.
F1 Score What is this model's F1 score?	The F1 score is a measure of a model's accuracy that considers both precision and recall. It provides a balance between the two metrics, especially when there is an uneven class distribution. The F1 score ranges from 0 to 1, where a higher score indicates better performance. In this case, the F1 score of the model is 0.913 $F1 = 2 \times \frac{P \times R}{P + R}$