AutoML Modeling Report

Chukwuemeka Ezumezu

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	80%, 10%, and 10% of the entire dataset were assigned to Train, Validation, and Test sets respectively. In each label, 160 images (i.e 80%) were used for training and 20 images (i.e 10%) were used for testing.
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?	- Under True Label are the labels of the classes, normal and pneumonia. Under Predicted Label, on the first row normal are the correctly predicted percentages as normal under normal column and percentage wrongly predicted as pneumonia under pneumonia column. Likewise, the wrongly predicted percentages as pneumonia under normal column and percentage rightly predicted as pneumonia under pneumonia column on the Pneumonia row. - Predictions for healthy and pneumonia images were 100% accurate (Screenshot, Clean_Balanced Model CM.jpg, was included in the folder). - The true positive rate for "pneumonia" class is 100% - The false positive rate for the "normal" class is 0%
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 correct predictions against the entire predictions made by a model, i.e true positive/ (true positive + false positive). - Recall measures correctly identified objects (like dogs) against all identification on the same objects (dogs), i.e true positive / (true positive + false negative) . - Model Precision = 100% - Model Recall = 100%
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	- When the threshold was increased precision remains the same while recall decreases. - Recall decreases because increasing threshold increases the percentage or proportion a model prediction will reach before a decision/classification is

	made. For instance, a model must reach a 0.6 threshold prediction of a healthy image before the image is classified as a healthy image. That will misclassify images with a lower threshold as false-negative causing a decrease in recall. While in general, raising the classification threshold reduces false positives, thus raising precision. Although increase precision with an increase in threshold is not always guaranteed. In this case, our precision is already 100% with no false negative.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	- 80% of the dataset was used for the Train set, i.e 360 images and - 10% for Validation and 10% for Test set (i.e 20 images for validation and 20 images for testing)
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	- More images were wrongly classified. In this case, 10% of healthy images were wrongly classified as pneumonia images while 7% of pneumonia images were wrongly classified as healthy images. The pneumonia class has less wrongly predicted image being that the class has more training data than the healthy class. - Screenshot, Clean_Unbalanced model CM.jpg, was included in the folder.
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)?	The model precision and recall were reduced as a result of the unbalanced dataset. Although the matrics are high, 92.5% for both precision and recall. - However, in the case of unbalanced classes, this metric can be misguiding, as high metrics don't always show prediction capacity for the minority class. - F1 metric can also be a good choice for the unbalanced classification scenarios.

Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Unbalanced classes decrease the precision and recall matric of the model. It can also make the model make wrong predictions because the predictions will lean towards the class with more training data (the model will predict more images of unseen test images to be the class with more training data), thereby confusing the model.

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 shows that the dirty data badly affected the model performance. 60% of healthy images were wrongly classified as pneumonia images. Also, though not badly affected as normal image classification, 20% of pneumonia images were classified as healthy images. This confusion is the result of having images of both classes mixed in a single class. - Screenshot, Confusion matrix (Dirty_Balanced Model CM.jpg, was included in the folder.
Precision and Recall How have the model's precision and recall been affected by the dirty data (report the values for a score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?	- Both classes' precision and recall were lowed by 40% when compared with the Clean/Balanced model. This is a result of having dirty data, in this case having both classes, in one class. - Clean/Balanced model has the highest precision and recall of 100% each. (Clean/Unbalanced model has precision and recall of 92.5% each while Dirty/Balanced Model has precision and recall of 60% each)
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data always badly affect models because the model will not learn properly as a result of having more than one class (dirty data) being trained as a single class.

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.	Predictions of the model were nearly perfect except for bacterial pneumonia, where 10% of bacterial pneumonia were classified as normal images and another 10% classified as viral pneumonia. - Both normal and viral pneumonia were mostly gotten right by the model with no misclassified images. - The model will need to be retrained focusing more on the most "confused" class (bacterial pneumonia). Like including more improved images of the class. - Screenshot, 3-Class Model CM.jpg, was included in the folder.
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)?	- The model's precision and recall are 93.33% each. - Precision and recall of each class need to be first calculated. For precision, we calculate correctly predicted normal divide by the total number of normal predictions. The same similar calculations are done for both bacterial and viral pneumonia. Then to find the overall precision of the model. The average of all the precisions will be taken. P(m) = (P(n) + P(b) + P(v)) / Num of class Similarly, for recall. The number of correctly predicted normal is calculated, divide by the total number of real (or ground truth positives) normal images. The same similar calculations are done for both bacterial and viral pneumonia classes. Then to find the overall recall value of the model. The average of all the recall values will be taken. R(m) = (R(n) + R(b) + R(v)) / Num of class
F1 Score What is this model's F1 score?	F1 = (2 * P * R) / (P + R) = 0.93