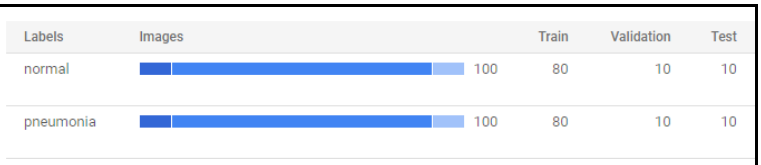
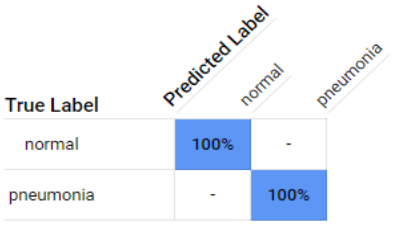
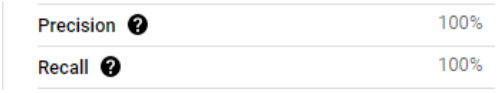
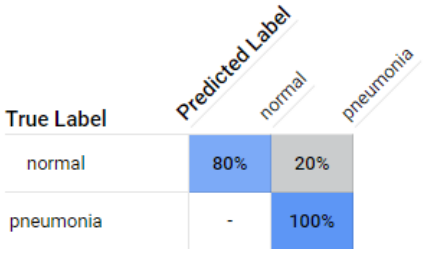
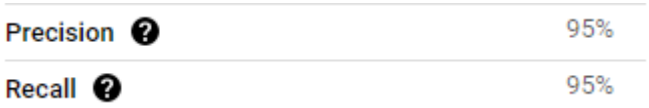
AutoML Modeling Report

<Lejaah Alowayidh>

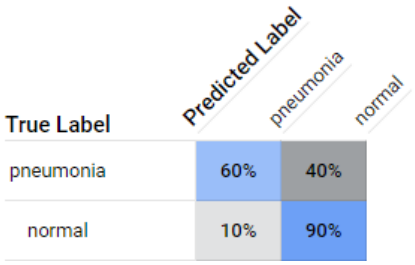
Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	 Out of 200 images, 160 were used for training, 20 for validation, and 20 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?	 The “True label” describe the total number of positive inputs in the data set, which equal (TP+FN). The “Predicted label” describe the total number of negative inputs in the data set, which equal (TN+FP). The true positive rate (TP) for the “pneumonia” class was 100%. on the other hand, the “normal” class had 0% false positive rate (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 correct predictions against total number of predictions, while recall measures the percentage of correctly identified instances total possible instances. At 0.5 threshold score, the model achieved 100% for both precision and recall.
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	By increasing the threshold, Precision remains at 100% while recall decrease gradually to 0%. This happen because the model is classifying less images in order to get more accurate predictions, which eventually results in more false negative values.

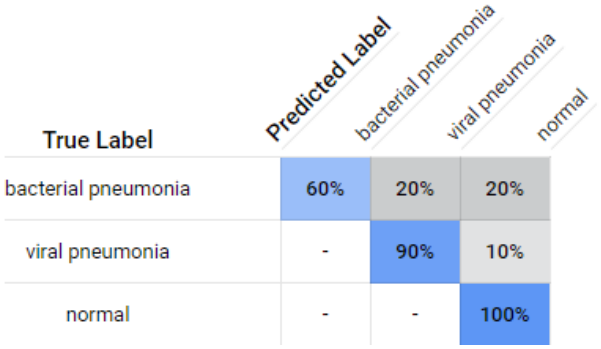
Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	 Out of 400 images, 320 were used for training, 40 for validation, and 40 for testing.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 Adding more “pneumonia” images resulted in missing a 20% of “normal” images (FN), and correctly predicting an only 80% of them (TN).
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)?	 At 0.5 threshold score, the model’s precision and recall decreased to 95%.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	As observed, adding more “pneumonia” in the training process causes bias in machine learning model, which resulted in classifying more “pneumonia” images; the more “class A” you put, the more “class A” you get.

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.	 Dirty data causes misclassification. TP (60%) and TN (90%) decreased, while both FP (40%) and FN (10%) decreased.
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?	 The model's precision and recall decreased to 75% at 0.5 threshold score. The binary classifier with clean/balanced data has the highest precision and recall (100%).
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data causes confusing to machine learning model through the training process, which eventually lead to misclassification.

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.	 the 3-class confusion matrix is a 3x3 table which consist of 3 true labels and 3 predicted labels. is the model most likely to confuse "Bacterial Pneumonia" and probably "Viral Pneumonia" as well, and most likely to get right "Normal" images. Apparently, the model is confusing Bacterial Pneumonia with Viral Pneumonia, increasing the number of the images in the training process could result in more accurate results. In conclusion, adding more images could help remedy the model's confusion.
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)?	Precision 85.71% Recall 80% At 0.5 threshold score, the model's precision was (85.71%) and an (80%) recall. In a 3-class matrix, precision and recall are calculated by calculating each class's initially, then their average is taken.
F1 Score What is this model's F1 score?	The model's F1 score is 82.75% (0.827566) $$F1 = \frac{2 * Precision * Recall}{(Precision + Recall)}$$