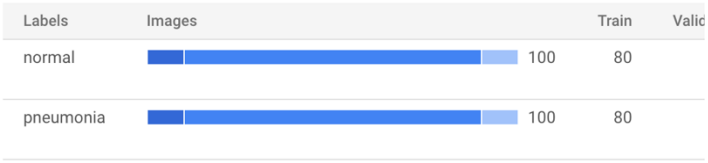
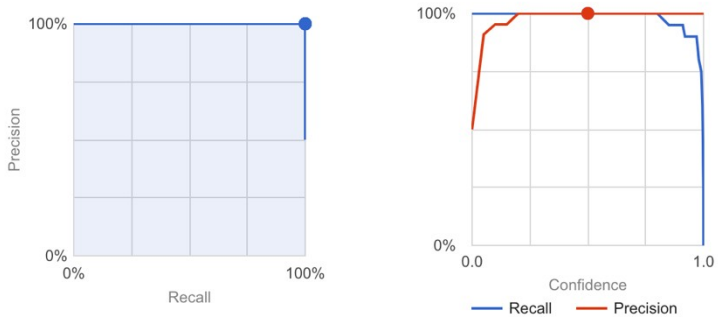
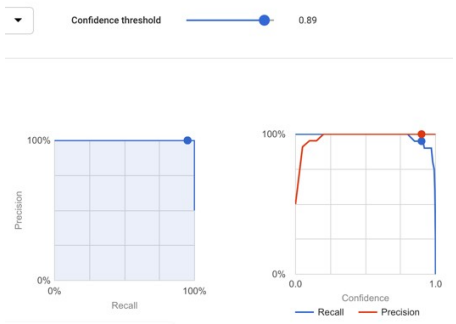
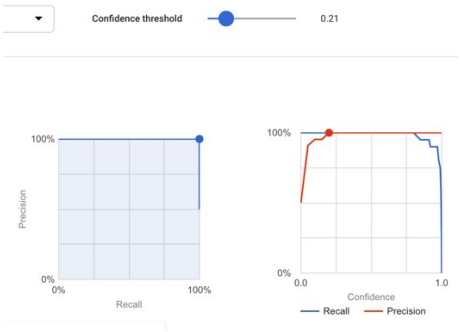
AutoML Modeling Report

Mohammed Almelabi

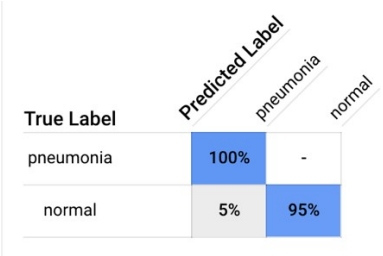
Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	 The data used in the model is 200 which were divided into two categries. 100 images of normal healthy lungs and 100 images of pneumonia. Out of the 200 images 160 were used for training while 10 normal ,10 pneumonia for validation and the same number as well 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 matrix shows all the predicted labels against the true labels. This matrix give us insight and where we need to improve our data to increase the model accuracy. False Positive for the normal label: 0% True Positive for the pneumonia label: 100%
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	 **Precision:** Measure the TP over the total of predictions, A high

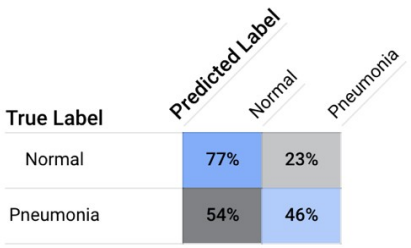
0.5)?	precision model produces fewer false positives. **Recall:** Measure the TP over the total of true predictions, A high recall model produces fewer false negatives. 
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	 **Increase in the threshold:** we notice that the precision increased and the recall decreased.  **decrease in the threshold:** we can notice the precision has been decreased while the recall increased. The score threshold has to do with the level of confidence that the model have to have to assign a category for testing. it is a tool to test the impact of different thresholds for all categories in the dataset. when the threshold score is low it means that the model will classify more images but will run the risk of misclassifying

	images in the process. On the other hand, when the score threshold is high, the model will classify images, and it will have a lower risk of misclassification.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	 The data used in the model is 537 in total. 156 images of normal healthy lungs and 318 images of lungs infected with pneumonia. Out of them 474 were used for training, (20,40) for validation and (20,40) for testing.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 The FP increased by 5% and the TP slipped to 95%. Because the model is trained on unbalanced data.
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)?	 Yes, it was affected.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	The model has a bias since we have trained it on an unbalanced data.

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.	 As shown in the figure above we see that we have more error in classification for this model.
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 the Precision and the Recall have decreased significantly, which mean this model has less accuracy. The binary classifier with the highest recall and precision is the first one "Clean and Balanced Model".
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Mislabeled data has impacted the model negatively, therefore, the model has performed poorly as expected.

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 model might confuse between the classes since viral penumonia and bactrial penmuonia are kinda simmilar. I believe the model will get the normal class right as it shown in the figure. However, we can also see that 10% of the viral penumoia was misclassified as normal.

What we can do to fix the model is:
1- Increase the trainig data.
2- Increase the score threshold which will increase the precision.

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)?

Total images	270
Test items	30
Precision	86.67%
Recall	86.67%

Similar to how we calculated the precision and recall for the binary classes.

Precision formula: TP/(TP+FP)
Recall formula: TP/ (TP+FN)
Sample for each class: 100

P= (7/10)+(10/10)+(9/10)/3= 0.8666667

R=(7/10)+(10/10)+(9/10)/3= 0.8666667

F1 Score
What is this model's F1 score?

F1= 2*(P*R)/(P+R)
= 2*(0.7512)/(1.617)
= 1.5024/1.617
= **92.91%**