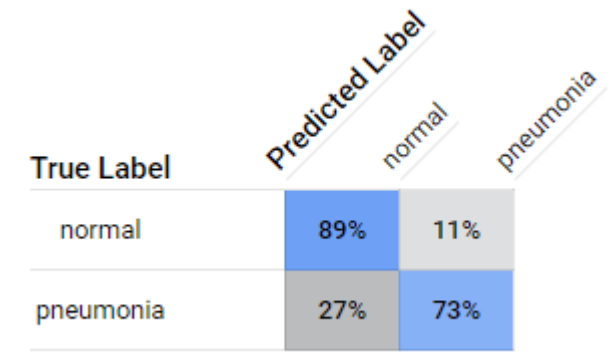
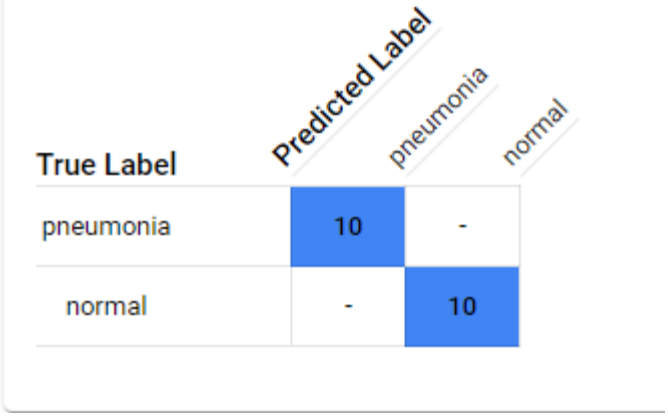
AutoML Modeling Report

Mario Sorgente

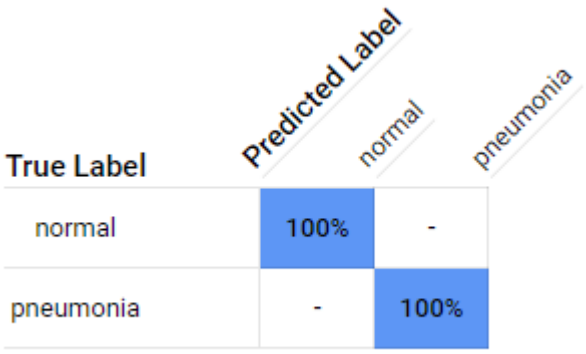
Binary Classifier with Clean/Balanced Data

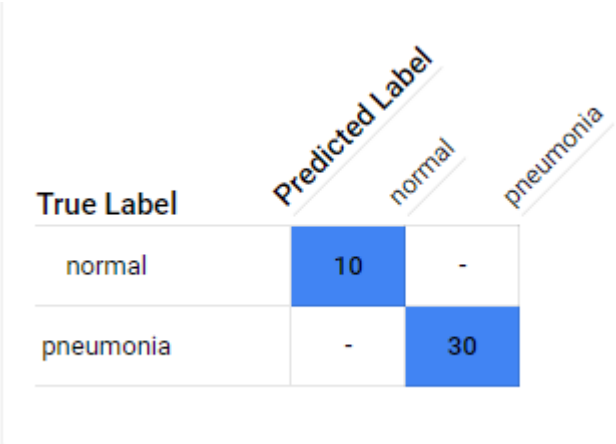
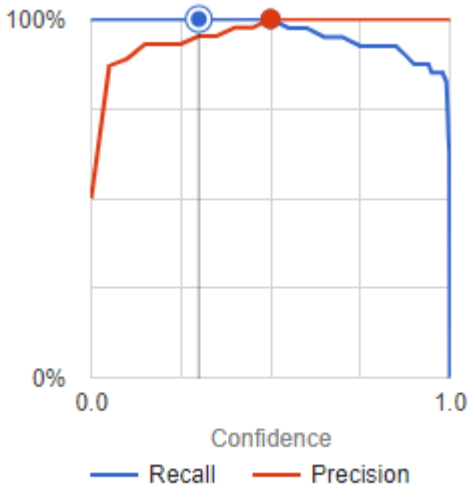
Train/Test Split How much data was used for training? How much data was used for testing?	The model was trained with 100 images per category, of which 10 used for testing and 10 for validation. It is important to split the amount, to ensure the model is tested on images it has never "seen". Overall: Trained: 180 Test: 20
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?	

	  A confusion matrix is a summary of prediction results on a classification problem. In the picture above, we can see that for 10 images with real pneumonia, the model correctly predicted 10 out of 10. Same for the normal category. The cells of the confusion matrix describe the combination of true positive, true negative, false positive and false negative. We can see that in this model true positive and true negative are present only. **True positive rate for pneumonia:** 100% **False positive rate for normal:** 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: answers the question "when the model makes a prediction, how likely the prediction is correct?" Recall: answers the question "how good is the model at identifying actual occurrences of objects in the data" [whether or not the model recognizes the objects]. **Precision: 100%** **Recall: 100%**
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	For a value of the threshold greater or equal to 0.83, the precision remains the same, while the recall drops of 10%. **Precision: 100%**

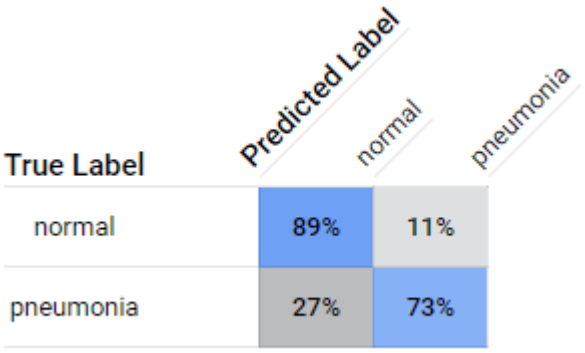
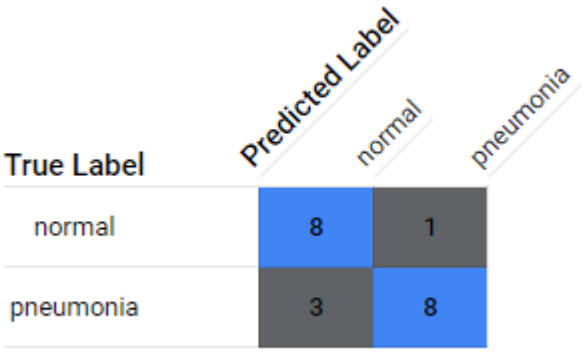
	Recall: 90% The score threshold refers to the level of confidence the model must have to assign a category to a test item. Increasing it, we are asking the model to be very precise in assigning the correct category. Consider the Recall the parameter that tells us, from all the test examples that should have had the label assigned, how many were actually assigned the label. With a higher threshold we are classifying fewer images, that's why recall is lower.

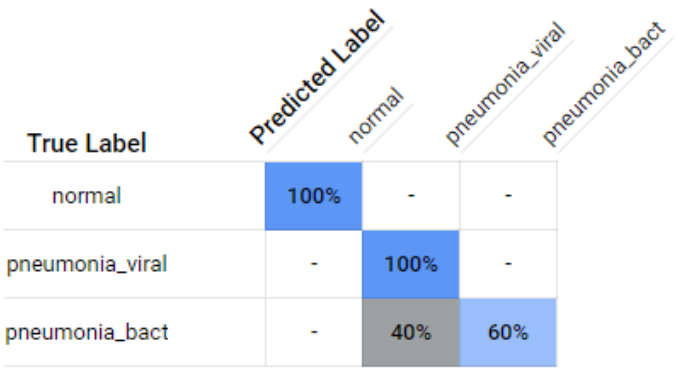
Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	For both categories, the model used 360 images for training 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.	

	 At first glance, it seems that the confusion matrix is not affected by the unbalance. After all the TP and TN in % are the same. What changes is the confidence of the model. In fact, we can observe that the item counts for pneumonia is 200% higher compared to the normal label.
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)?	For a threshold of 0.5 (that is 50% confidence), the precision and the recall are not affected. However it is clear that slightly changing the threshold (confidence) will impact the two parameters, since the model is highly unbalanced. 
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	The confidence model is highly affected, especially tweaking a different threshold or activation function. The model is most likely to be biased towards predicting the class with the most images (Pneumonia)

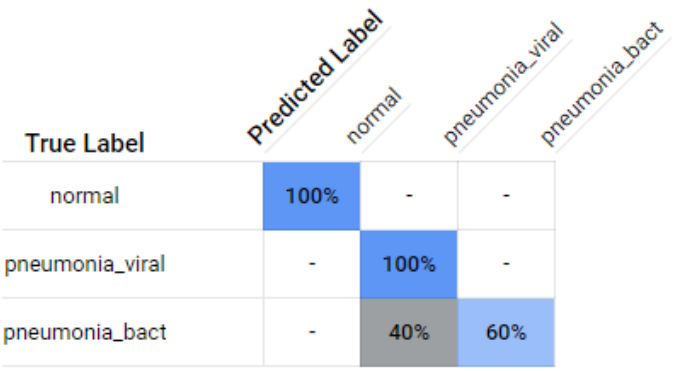
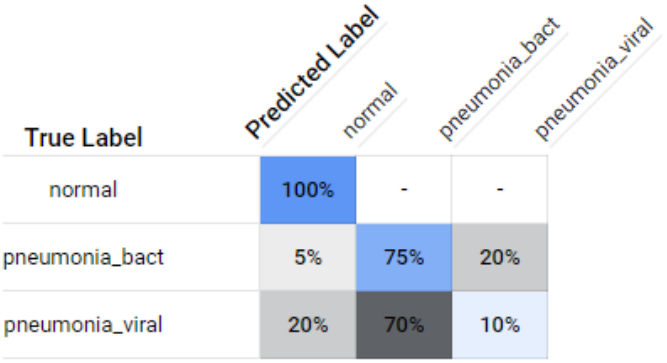
Binary Classifier with Dirty/Balanced Data

Confusion Matrix How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.	  As we can see, we have now all the four cells activated. True/false positive, True/False negative. Of the 20 images tested, the distribution is the below: TP:8 TN:8 FN:3 FP:1
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?	For a threshold of 0.5. **Precision: 80%** **Recall: 80%** We observe a drop of 20% in both parameters. Clean+balanced

	 binary classifier has the highest precision and recall
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	It induces the model to produce much more mispredicted label, decreasing precision and recall. Moreover, the selection of the threshold is much more delicate, depending on the performance goal 

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.	 Including a third class – with the distinction between bacterial and viral pneumonia - it becomes trickier for the model to be 100% confident. From the confusion matrix, we can see that the pneumonia_bacterial class is the one that most likely will be mispredicted (confused) and should be improved with training data. The model will most likely get right normal and pneumonia_viral classes. To improve the model, we could increase the number (or quality) of data for all the three classes (e.g. using 300 images per class). Warning: increasing the pneumonia_bacteria data only, would unbalance the model, decreasing precision and recall, as per the example reported below: 
Precision and Recall What are the model's precision and recall? How are these values	**Precision:** 86.67% **Recall:** 86.67%

calculated (report the values for a score threshold of 0.5)?	Using 30 images for testing P_normal = TP/(TP+FP) = 10/(10+0) = 100% P_bact = TP/(TP+FP) = 6/(6+4) = 60% P_viral = TP/(TP+FP) = 10/(10+0) = 100% P_aver = (P_normal+P_bact+P_viral)/3 = 86.67 R_normal = TP/(TP+FN) = 10/(10+0) = 100% R_bact = TP/(TP+FN) = 6/(6+4) = 60% R_viral = TP/(TP+FN) = 10/(10+0) = 100% R_aver = (P_normal+P_bact+P_viral)/3 = 86.67
F1 Score What is this model's F1 score?	F1 = 2*(P*R)/(P+R) = 86.67%