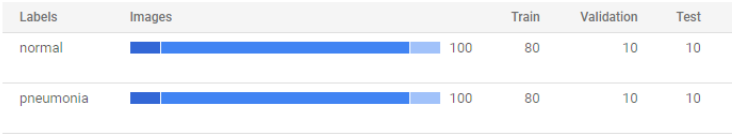
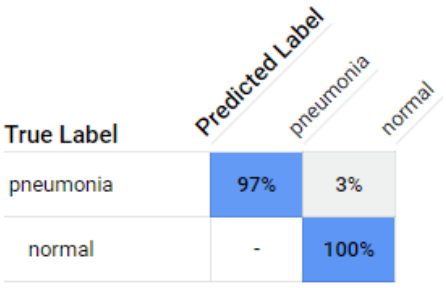
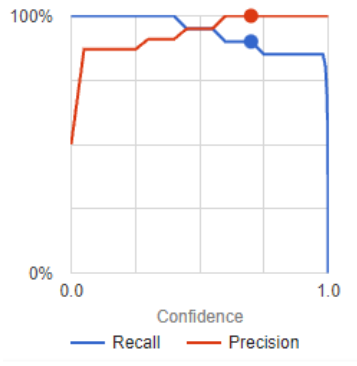
AutoML Modeling Report

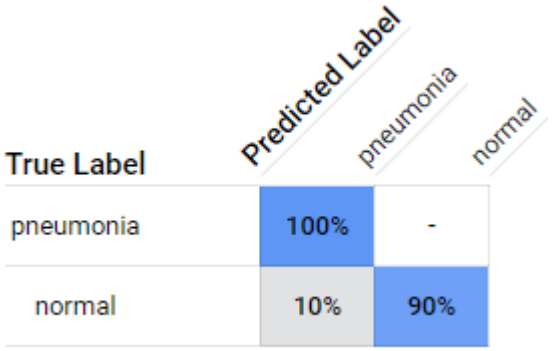
Jorge Luiz Figueira da Silva Junior

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	A total of 200 instances were used as a dataset. The split into training, testing, and validation sets is given in the chart below. 
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 Confusion Matrix provides an effective measure of the classification model by showing the number of correct classifications (True Positives, True Negatives) versus incorrect classifications (False Positive, False Negative. True positives are the correct classification of pneumonia, which is 97%. False Positives are cases in which the true label is negative (“normal” class) however the model classified it as positive (“pneumonia” class). As noted in the confusion matrix, not occurred these cases, so the rate is not specified. False Negatives are cases in which the true label is positive (“pneumonia”) however classified as negative (“normal”). The False Negative rate in this model is 3%. Finally, True Negatives are the cases where the true label is negative (“normal”) and the predicted label is classified as negative. The True Negative rate is 100%.

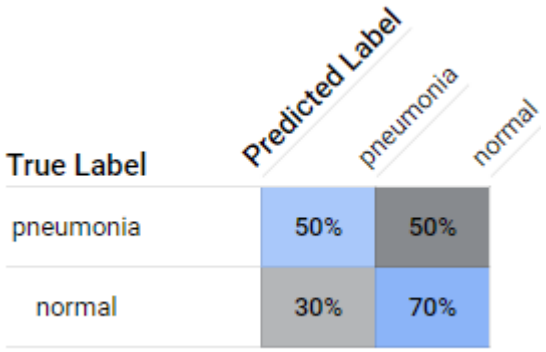
	
Precision and Recall What does precision measure? What does recall measure? What precision and recall did the model achieve (report the values for a score threshold of 0.5)?	Precision measures the ratio of true positives to the predicted values. Recall measures the ratio of true positives to actual positives. **Precision** 95% **Recall** 95%
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	A higher threshold results in an increase in precision (because the model never makes a prediction unless it is extremely sure) but the recall (the percentage of positive examples that the model gets right) decreases. * Using a score threshold of 0.7 **Precision** 100% **Recall** 90% 

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	The split into training, testing, and validation sets is given in the chart below.  Train / Validation / Test 80 / 10 / 10 239 / 30 / 30
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	Regarding the clean/balanced model, there was a significant increase in the True Positive rate (correct classifications of “Pneumonia”) and a drop in the True Negative rate (correct classifications of “normal”). This is because a bias was introduced in the model. As there are more images of the “pneumonia” class, therefore, the model will learn more patterns from this class, and the effect of this is the one visualized in the confusion matrix, a tendency to predict the class with more images. 
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)?	There was a slight improvement in precision and model recall.

	unbalanced_163353_20211006021646  Average precision 0.993 Precision* 97.5% Recall* 97.5% * Using a score threshold of 0.5
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Balancing the data is as important if not more to produce an unbiased model. As the results suggest, there is a tendency for the model to classify the examples as 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.	The model obtained in these configurations had a drastic drop in performance. 
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?	This model, when compared to the model with balanced data and even with the model with unbalanced data, had the worst performance with 60% of precision and recall.

	dirty_balanced_20211005080153  Average precision 0.558 Precision* 60% Recall* 60% * Using a score threshold of 0.5 For this experiment, the Binary Classifier with Clean/Unbalanced Data had the highest precision and recall.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	One of the most common issues with training data is using mislabeled data. Models learn from patterns extracted from training data. As equal patterns are seen on both labels, the model had difficulties in data separability, resulting in performance poorly.

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.	From the confusion matrix it is observed that the model makes confusion between the "viral" and "bacterial" classes. Still, the model was able to predict all instances of the "viral" class. The model also got all the correct predictions of the normal "class". According to the literature, the classification of subcategories is still quite challenging. In this sense, for this model I would unify the labels "viral" and "bacterial" as being solely "pneumonia".

	
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)?	three_classes_20211005085111  Average precision 0.945 Precision* 89.66% Recall* 86.67% * Using a score threshold of 0.5 Precision and Recall rates can be obtained from: $Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$ $Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$
F1 Score What is this model's F1 score?	F1 = 80.14% $F1 = 2 * \frac{precision * recall}{precision + recall}$