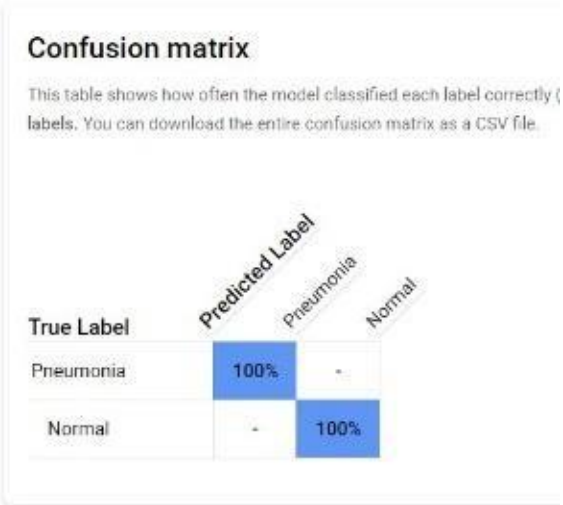
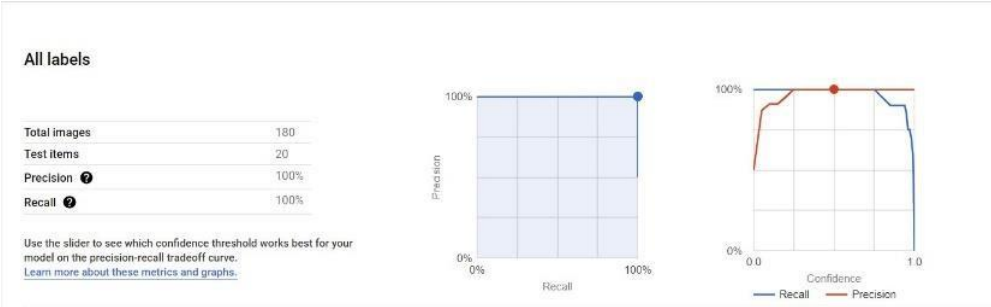
AutoML Modeling Report

Mohamed Fazal Mustafa

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	There was an automatic split of 80-10-10 where 80% of the data was used for training, 10% for validating and 10% for testing. 160 images were used as training data, 20 as validation and 20 as testing data.
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 top left cell denotes True Positive (pneumonia is present and the model has predicted pneumonia), top right denotes False Negative (pneumonia is present but the model has predicted normal), down left denotes False Positive (pneumonia is not present but the model has predicted pneumonia) and down right denotes True Negative (pneumonia is not present and the model has predicted normal). The true positive rate for ‘pneumonia’ is 100% and the false positive rate for ‘normal’ 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)?

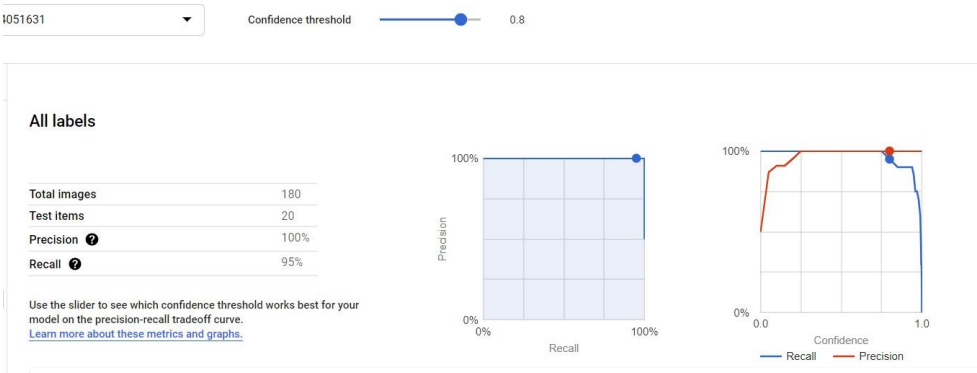


Precision gives a measure of what proportion of positive identifications are actually correct. It is calculated as the number of true positive divided by the number of total positives predicted (true positive + false positive).

Recall gives a measure of what proportion of actual positives were identified correctly. It is calculated as the number of true positives divided by the sum of true positives and false negatives.

The model has 100% Precision and 100% Recall values.

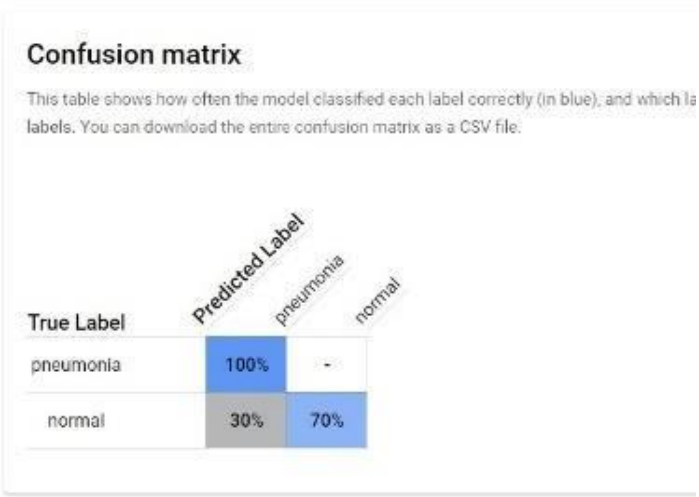
Score Threshold
When you increase the threshold what happens to precision? What happens to recall? Why?



On increasing the threshold from 0.5 to 0.8, the value of precision remains the same while the value of recall reduces from 100% to 95%.

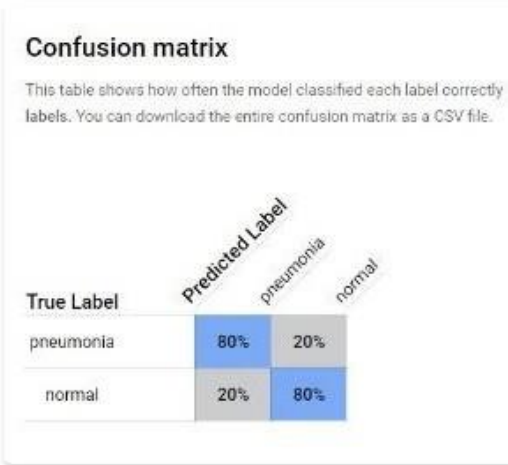
The score threshold is a measure of level of confidence of a model to classify to an image. If the value of threshold is low, the model classifies more images but has a higher risk of misclassifying while if the threshold is high, the model classifies fewer images but has a lower risk of misclassifying.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	The AutoML automatically splits the data into 80-10-10, where 80% of the images are used as training data, 10% as validation data and 10% as testing data. 320 images were used as training data, 40 images as validation and 40 images as testing data.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 In the confusion matrix of clean and unbalanced data, the false positive are 30% and the true negatives are 70% whereas in the model trained using clean and balanced data, the false positives were 0% and the true negatives were 100%. The normal class has some negatives due to the model being biased towards Pneumonia 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)?	 The precision and recall values have reduced from 100% in the clean and balanced model to 92.31% in this model. Unbalanced data has resulted in reduction of precision and recall values.
Unbalanced Classes From what you have observed, how do unbalanced classes affect a machine learning model?	Unbalanced classes negatively affect the performance of the classification model. It reduces the precision and recall values and impacts the confusion matrix. If the dataset contains unbalanced data, it will affect the model's performance and the model will be skewed towards a particular outcome.

Binary Classifier with Dirty/Balanced Data

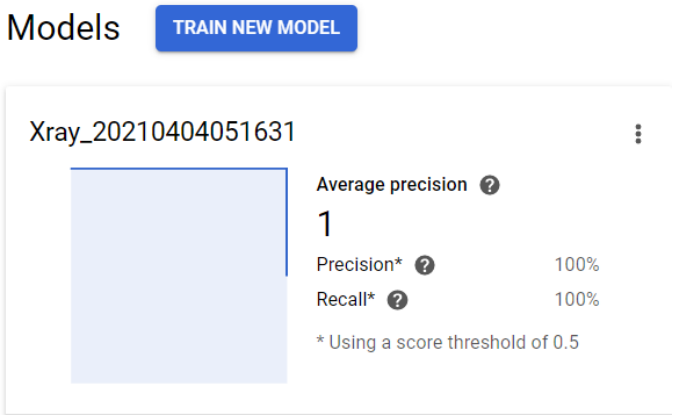
Confusion Matrix
How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.



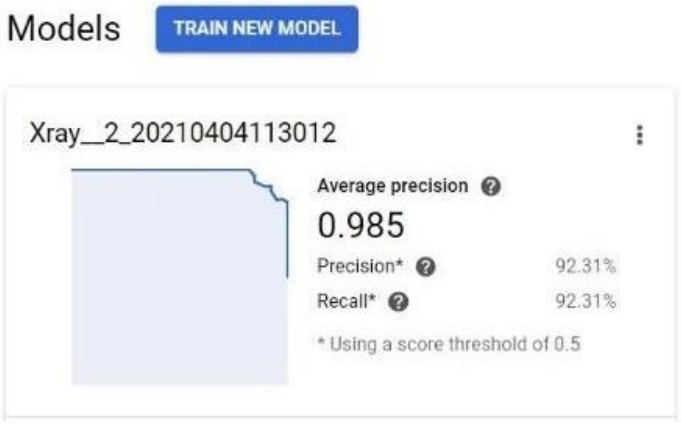
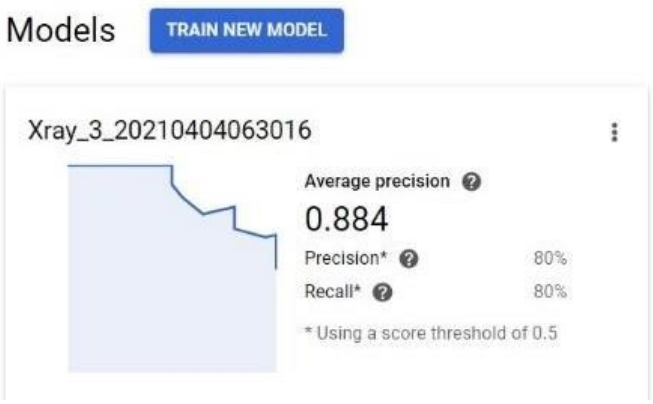
20% pneumonia images were misclassified as normal and 20% of normal images were misclassified as pneumonia.

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?

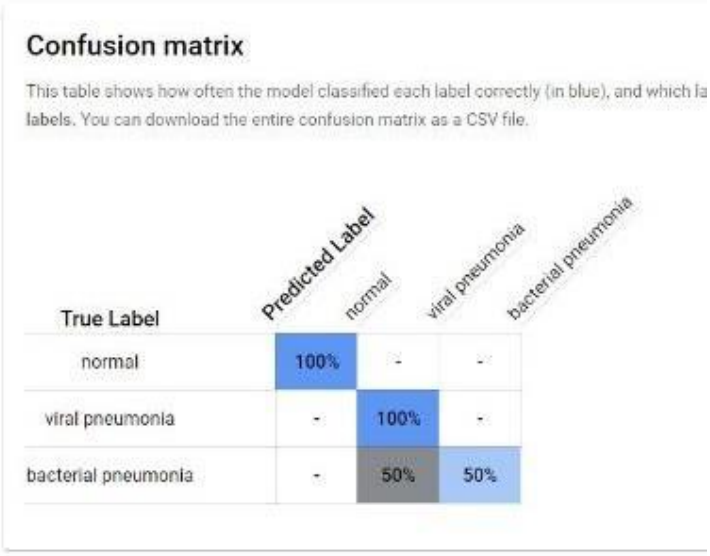
The precision of the dirty and balanced dataset was 80% and the recall was also 80%.
For the model trained on **clean and balanced data**:
Precision: 100%
Recall: 100%



For the model trained on **clean and unbalanced data**:
Precision: 92.31%

	Recall: 92.31%  And for the model trained on **dirty and balanced data**: Precision: 80% Recall:80%  The highest precision and recall (with the score threshold set at 0.5) were both of the clean and balanced dataset which was 100%.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	If the dataset contains dirty data, it significantly lowers the precision and recall score, thus negatively affecting the model's performance. It also impacts the confusion matrix.

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 is more likely to confuse the bacterial pneumonia x-rays. It classified the normal and viral pneumonia images at 100%. The model is most likely to classify the normal images and viral pneumonia images correctly. Increasing the size of the dataset such that the data for each label remains balanced, will be able to remedy this problem.
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)?	

Confusion matrix

Item counts

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray). **Note that this table is limited to the 10 most confused labels.** You can download the entire confusion matrix as a CSV file.

True Label \ Predicted Label	normal	viral pneumonia	bacterial pneumonia
normal	10	-	-
viral pneumonia	-	10	-
bacterial pneumonia	-	5	5

The precision for the first model was 83.33% and the recall was 83.33%. It gives us the measure of how many times the model is correct whenever it classifies any image. Recall gives us the measure of how many times the model correctly classifies the images.

For the **normal class**,
Precision= number of correctly predicted normal class = 10 = 1
All predicted normal class 10

In this model, whenever the model predicts that the x ray is normal, it is correct 100% of the time.

Recall= number of correctly predicted normal class = 10 = 1
Total number of actual normal images 10

In this model, the model correctly classifies 100% of the normal cases.

For the **bacterial pneumonia class**,
Precision= number of correctly predicted bacterial pneumonia = 5 = 0.5
All predicted bacterial pneumonia 10

In this model, whenever the model predicts that the x ray is bacterial pneumonia, it is correct 50% of the time.

Recall= number of correctly predicted bacterial pneumonia = 5 = 1
Total number of actual bacterial pneumonia images 10

In this model, the model correctly classifies 50% of the bacterial pneumonia cases.

For the **viral pneumonia class**,
Precision= number of correctly predicted viral pneumonia = 10 = 0.66
All predicted viral pneumonia 15

	In this model, whenever the model predicts that the x ray is viral pneumonia, it is correct 66.67% of the time. Recall= number of correctly predicted viral pneumonia = 10 = 1 Total number of actual viral pneumonia images 10 In this model, the model correctly classifies 100% of the viral pneumonia cases.
F1 Score What is this model's F1 score?	The F1 score is calculated as: $F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$ Precision for this model= 83.33%= 0.83 Recall for this model =83.33%= 0.83 F1 score= (2*0.83*0.83)/(0.83+0.83) = 0.83 Therefore the F1 score of the 3 class classifier model is 0.83.