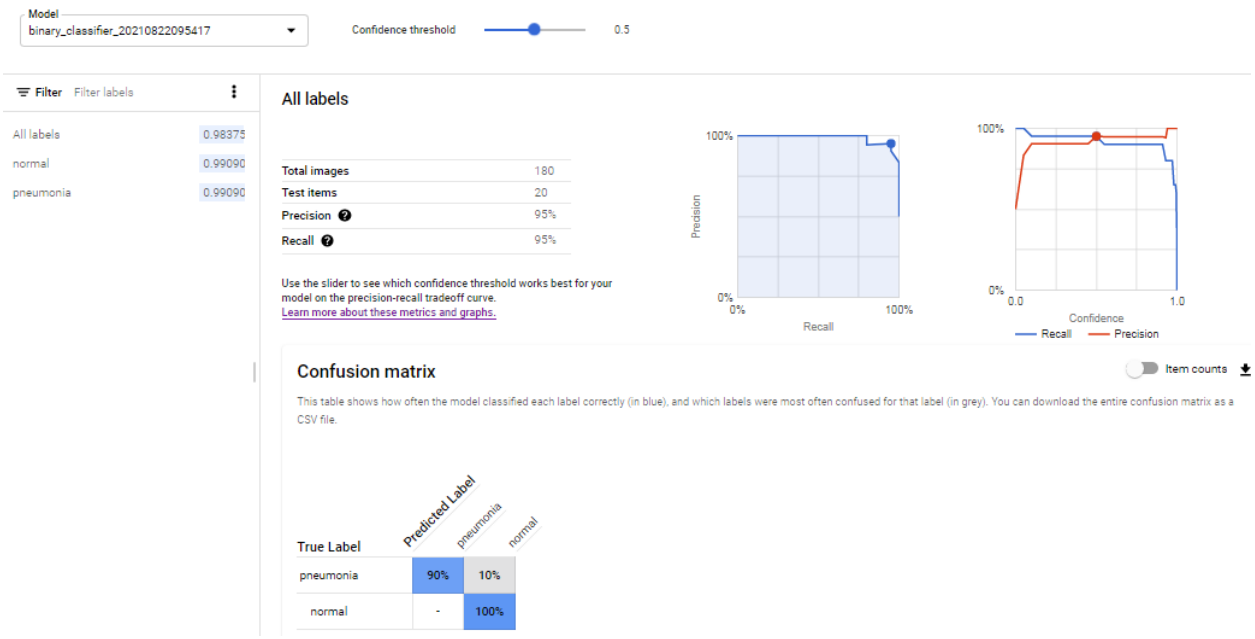
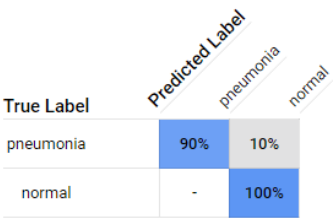
AutoML Modeling Report

Mohammed Al-Dhabaab

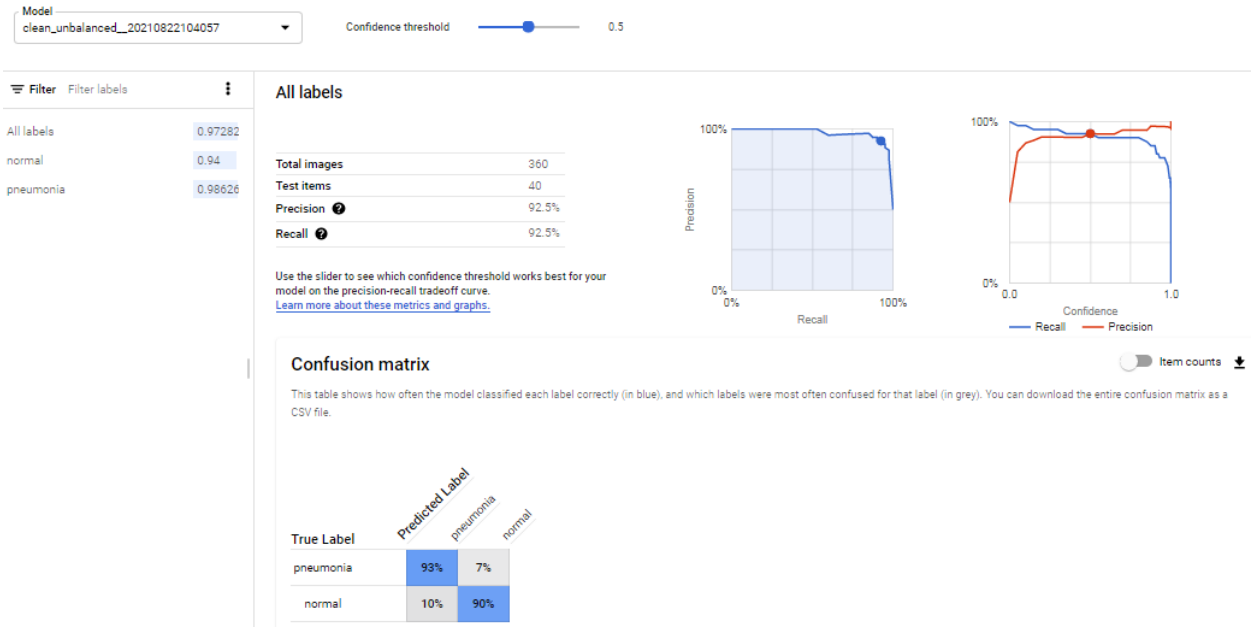
Binary Classifier with Clean/Balanced Data



Train/Test Split How much data was used for training? How much data was used for testing?	**200 labeled data (images) has been used. 100 for pneumonia x-ray images of children have been splitting between 79 trained images, 11 validation and 10 for testing. Also, 100 for normal x-ray images of children has been splitting between 79 trained images, 11 validation and 10 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?	**Confusion matrix shows us that true pneumonias were classified 90% of the time in the pneumonia category as well as 10% were false negative classified. On the other hand it shows us that true negative normal was classified 100% while 0% was false positive.** 
Precision and Recall What does precision measure? What does recall measure? What	**Precision is the ratio between the True Positives and all the Positives ($\frac{TP}{TP+FP}$), and will be the higher when the amount of false positives is low. In our model at**

precision and recall did the model achieve (report the values for a score threshold of 0.5)?	**0.5 threshold the precision achieved 95%.** **Recall is the measure of our model correctly identifying True Positives defined as the number of true positives over true positives plus false negatives ($\frac{TP}{TP+FN}$) and will be higher when the number of false negatives is low. In our model at 0.5 threshold the recall achieved 95%.**
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	**When increasing the threshold the precision increase (Positive relationship), while recall decrease (Inverse relationship). That is because the FP decreased results in increasing the precision at its equation: $\text{precision} = \frac{TP}{TP+FP}$. On the other side, FN Increased result in decreasing the recall at its equation: $\text{recall} = \frac{TP}{TP+FN}$.**

Binary Classifier with Clean/Unbalanced Data



Train/Test Split

How much data was used for training? How much data was used for testing?

400 labeled data (images) has been used. 300 for pneumonia x-ray images of children have been splitting between 239 trained images, 31 for validation and 30 for testing. Also, 100 for normal x-ray images of children has been splitting between 79 trained images, 11 validation and 10 for testing.

Labels	Images	Train	Validation	Test
normal	100	79	11	10
pneumonia	300	239	31	30

Confusion Matrix

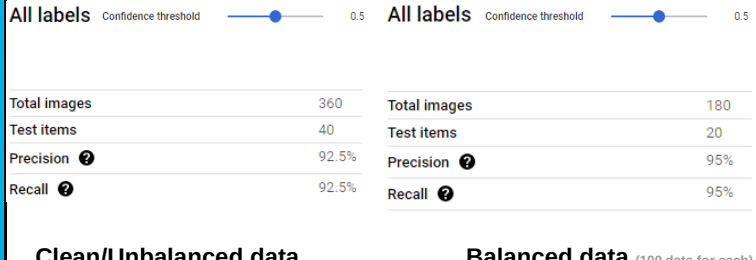
How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.

Due to unbalanced data (unequally number of data for each category) that result the model tend to bias towards pneumonia category, because it has 200 data more than normal category.

True Label	Predicted Label: pneumonia	Predicted Label: normal
pneumonia	93%	7%
normal	10%	90%

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)?

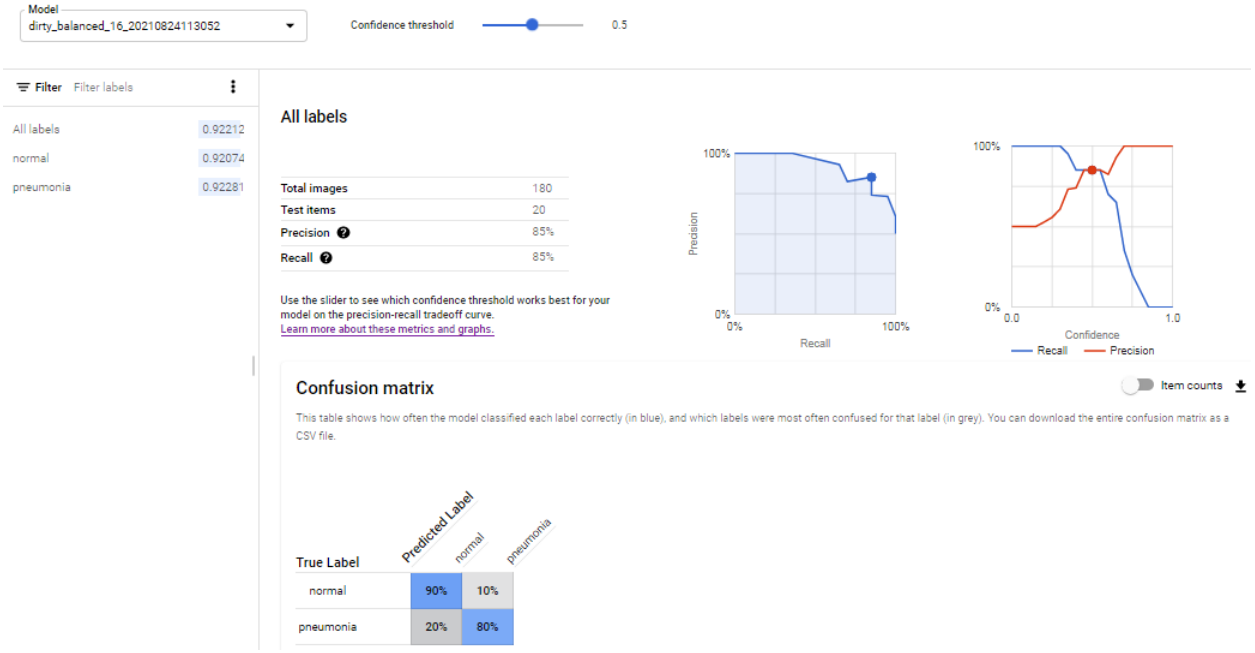


Clean/Unbalanced data **Balanced data** (100 data for each)

In our model unbalanced data at 0.5 threshold, the precision decreased to 92.5% and the recall decreased to 92.5%. Due to the unbalanced data, the model's accuracy has decreased.

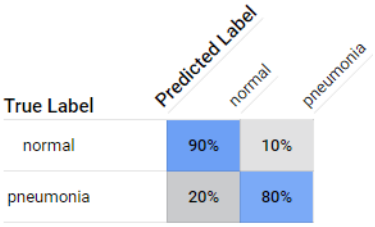
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	**Unbalanced classes is one of the common issue in the phase of training data, the model will learn significantly more about the highest classes with data and will tend to bias towards them. Leaving the lowest classes with lose at training them as well.**

Binary Classifier with Dirty/Balanced Data

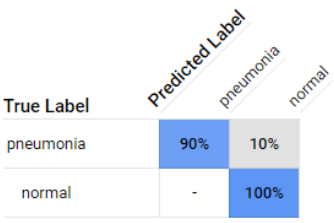


Confusion Matrix

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



Dirty/balanced data

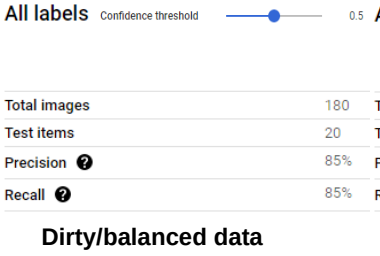


Balanced data

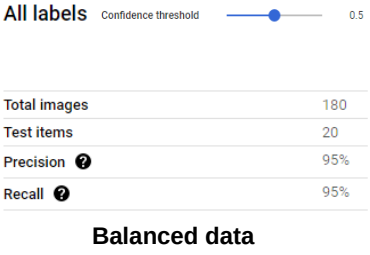
As we can see the model performance inaccuracy the false positive had increased 20%, as well as true negative, had decreased, due to data manipulation by switch the labels of 30 images in each class that means 30% of the data are manipulated (mislabeled).

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?



Dirty/balanced data



Balanced data

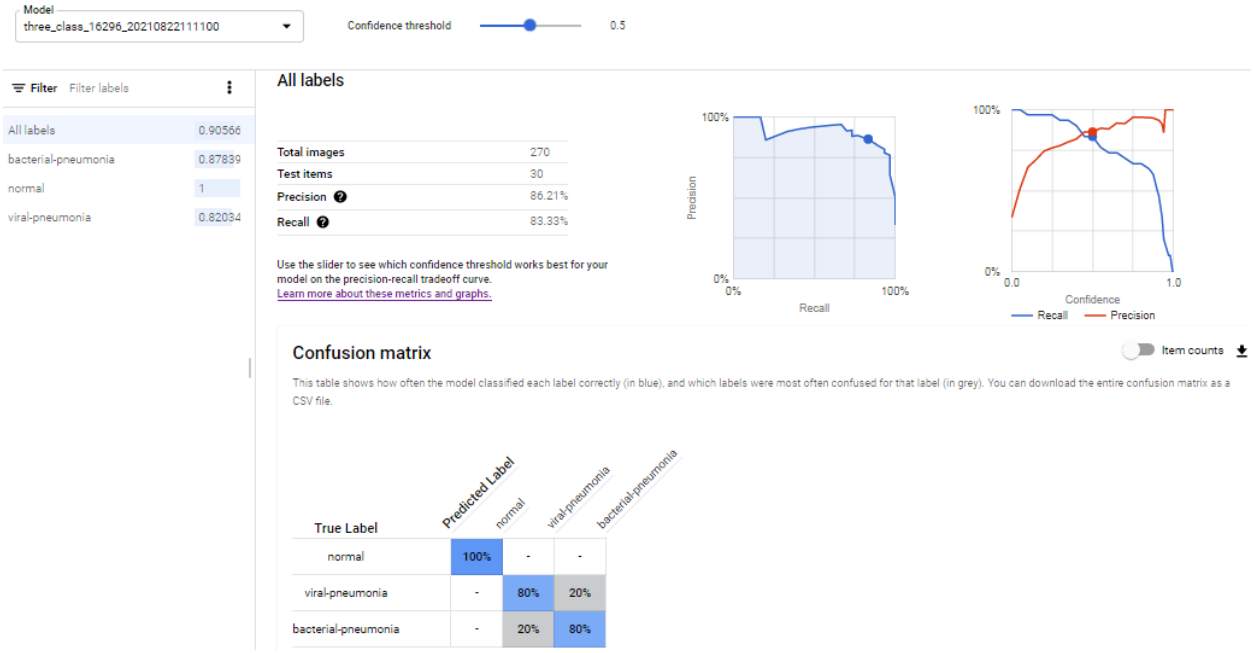
As figures show us that precision's dirty/balanced data and recall has decreased 10% which means our model's accuracy has decreased to predict from our classes (pneumonia and normal). While balanced data has the highest accuracy (precision and recall) 10% more.

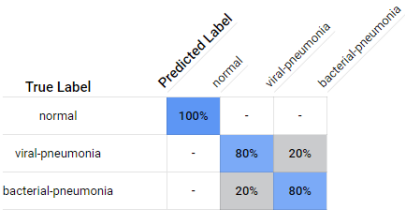
Dirty Data

From what you have observed, how does dirty data affect a machine learning model?

From my knowledge, dirty data affected our model at an unacceptable performance like its accuracy had decreased which means it predicts some data correctly and some data wrongly.

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 will be confuse with two similar classes bacterial-pneumonia and viral-pneumonia, because the confusion matrix show us that these two classes have 20% of false negative for each one. On the other side normal class has 100% true positive that means the model performed well with this class. We can optimize our confusions with adding more data to increase the accuracy, thus less confusion. Also, There are more advanced improvements methods, Such as multiple algorithms, Algorithm Tuning and more... (According to Analytics Vidhya website)** 
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)?	**In our model 3 classes at 0.5 threshold the precision achieved 86.21% and the recall achieved 83.33%. The precision value had calculated through AutoML from Google by summing all precision for each class over their number:** $$\text{Precision} = \frac{P_{\text{Bacterial pneumonia}} + P_{\text{Viral pneumonia}} P_{\text{Normal}}}{3}$$ **And recall value the same equation as well:** $$\text{Recall} = \frac{R_{\text{Bacterial pneumonia}} + R_{\text{Viral pneumonia}} R_{\text{Normal}}}{3}$$
F1 Score What is this model's F1 score?	**There are multiple ways to do that:** **Short one:** **The AutoML from Google had calculated the precision and recall for all classes at 0.5 threshold:**

Precision = 86.21%
Recall = 83.33%
The model's F1 score:

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

$$F1 = \frac{2 \times 0.8621 \times 0.8333}{0.8621 + 0.8333} = 0.84..$$

Long one:

First we need to calculate F1 score for each class.

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

The AutoML from Google had calculated the precision and recall for each class at 0.5 threshold:

- **Bacterial-pneumonia class with 77.78% precision and 70% recall. The F1 score:**

$$F1(\text{Bacterial pneumonia}) = \frac{2 \times 0.7778 \times 0.7}{0.7778 + 0.7} = 0.73..$$

- **Viral-pneumonia class with 80% precision and 80% recall. The F1 score:**

$$F1(\text{Viral pneumonia}) = \frac{2 \times 0.8 \times 0.8}{0.8 + 0.8} = 0.8$$

- **Normal class with 100% precision and 100% recall.**

$$F1(\text{Normal}) = \frac{2 \times 1 \times 1}{1 + 1} = 1$$

Now we can to calculate model's F1 score, By summing each F1 score over their number:

$$F1(\text{model}) = \frac{0.73 + 0.8 + 1}{3} = 0.84$$

This way help us to know more about our model where it performances well (class) and where not.