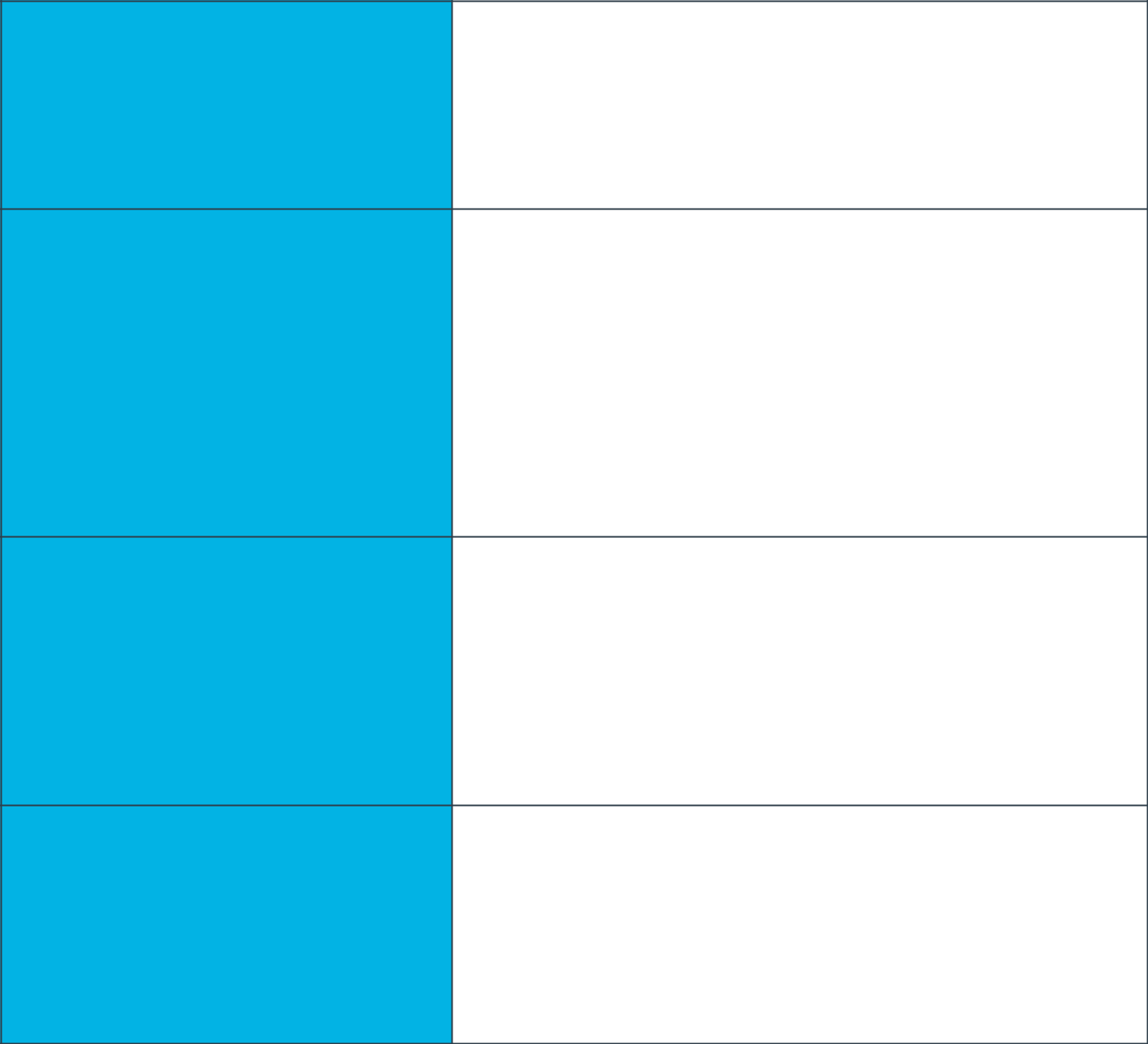
AutoML Modeling Report

*Gouthamsimha*



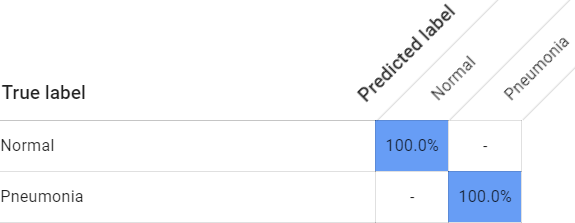
**Binary Classifier with Clean/Balanced Data**



There are 100 images of Pneumonia and 100 normal images. Out of 200, 160 were used for training, 20 were used for validating and 20 were used for testing

**Train/Test Split**

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



Confusion Matrix shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in orange). The 4 cells in Confusion Matrix are TP, FP, FN and TN.

For Normal, TP = 100, FP = 0.0;

For Pneumonia, FN = 0.0, TN = 100.0

**Confusion Matrix**

What do each of the cells in the

Confusion matrix describe? What ds

Values did you observe (include a

Screenshot?) What is the true

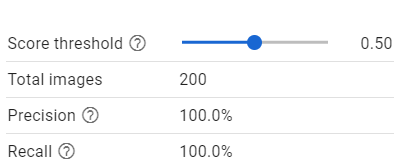
positive rate for the “pneumonia”

Class? What is the false positive

Rate for the “normal” class?

Precision tells us what portion of positive identifications are actually correct. Lower FP values gives higher Precision value.

Recall tells us, out of all positives how many did the machine identify correctly. Lower FN value gives higher Recall value.

This model achieved a Precision of 100% and Recall of 100%

**Precision & 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)?

When increasing threshold, we can see that Precision Increases and Recall decreases. We increase threshold to be more accurate of the prediction .When threshold increases we classify very less images but it has a very low risk of misclassifying the images.

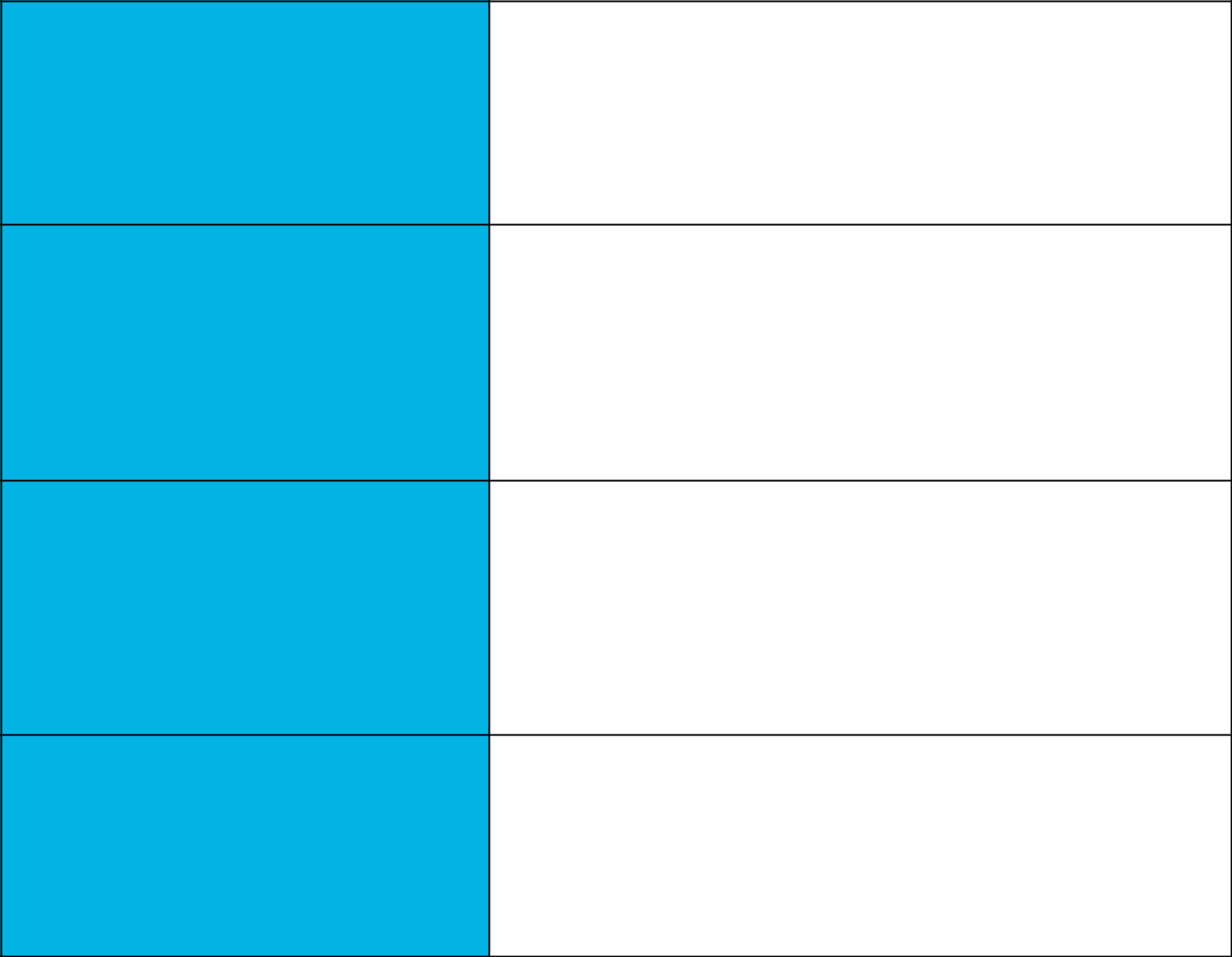
**Score Threshold**

When you increase the score

Threshold, what happens to Precision? What happens to recall?

Why?

**Binary Classifier with Clean/Unbalanced Data**



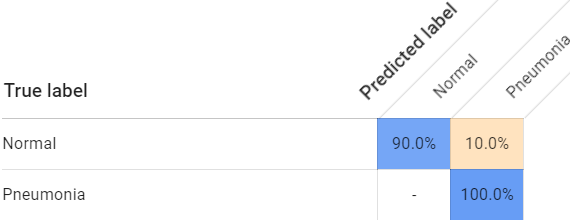
There are 100 normal images and 299 Pneumonia images. Out of these 319 were used for training, 47 for validating and 33 for testing

**Train/Test Split**

How much data was used for

training? How much data was used

for testing?

 Comparing to Balanced data, TP went down (90.0%). TN remains the same and False Positive = 10%

**Confusion Matrix**

How has the confusion matrix

been affected by the unbalanced

data? Include a screenshot of the

new confusion matrix.

For a Threshold of 0.5,

Precision = 96.97.

Recall = 96.97.

**Precision & 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.)

Unbalanced data will always bring bias to the model. Model will always be biased to the data which had more training. In this case we trained 100 normal images and 299 pneumonia images. I tested some images and the result is a bit biased towards Pneumonia as we trained more images of it.

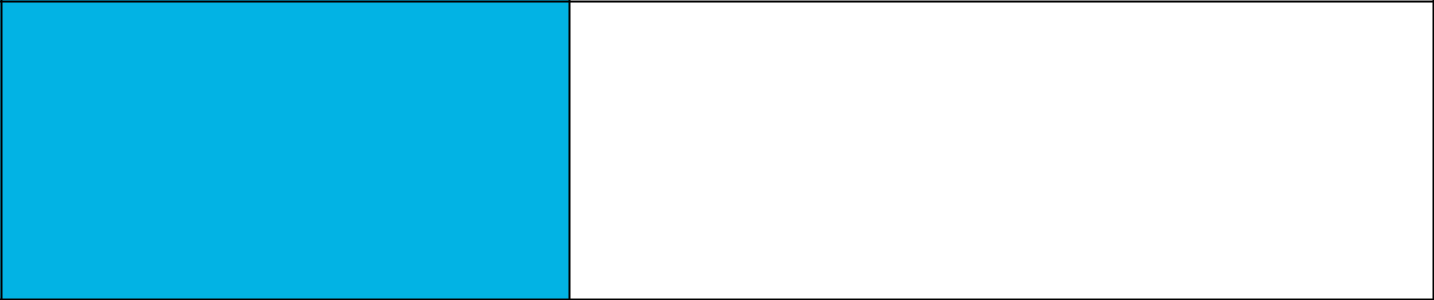
**Unbalanced Classes**

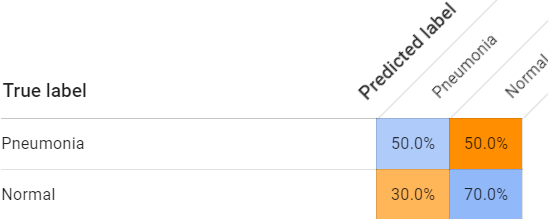
From what you’ve observed, how

do unbalanced classes affect a

machine learning model?

**Binary Classifier with Dirty/Balanced Data**



From the above fig. we can see that the confusion matrix is heavily affected by the Dirty data.

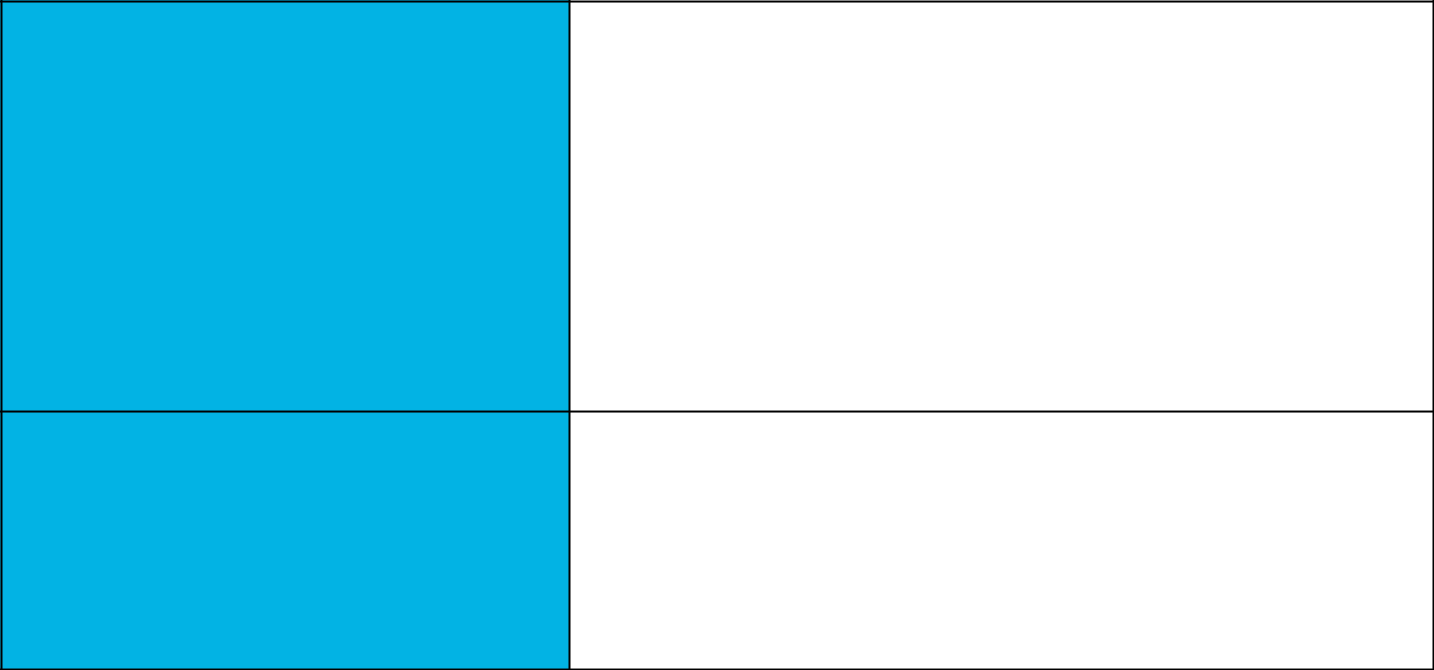
**Confusion Matrix**

How has the confusion matrix

been affected by the dirty data?

Include a screenshot of the new

confusion matrix.

**Precision & Recall**

The values of Precision and Recall are dropped very low due to the dirty data.

For 0.5 threshold,

Precision = 59.1 %

Recall = 59.1 %

Out of all these binary classifiers Clean/Balanced data has the highest 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 data affected the model very badly, due to the mix up of Pneumonia and Normal images the model was badly trained and it was unable to predict the desired output.

**Dirty Data**

From what you’ve observed, how

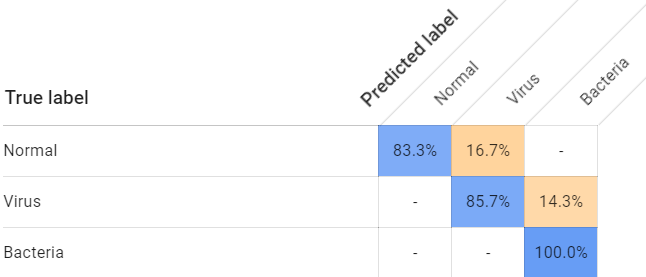
do dirty data affect a machine

learning model?

**3-Class Model**



The three classes are Normal, Bacteria and Virus. The Model is most likely to confuse with Normal and Virus classes. The model is most likely to get the Bacteria class right. The more the data the accurate the results are. So, I suggest with adding more amount of data we can reduce the confusion



**Confusion Matrix**

Summarize the 3-class confusion

matrix. What classes are the model

most likely to confuse? What

class(es) is the model most likely to

get right? What might you do to try

to remedy the model’s “confusion”?

Include a screenshot of the new

confusion matrix.

For a Threshold of 0.5,

At first these are calculated individually and then averaged.

Precision for Normal, Virus and Bacteria are 83.3, 85.7, 100 respectively. Similarly for recall, divide the value of TP by sum of the column. Recall values for N, V, B are 100, 83.69, 87.48 respectively.

Precision and Recall for whole model are 89.66 and 90.39respectively.

**Precision & Recall**

What are the model’s precision and

recall? How are these values

calculated? (Report the values for a

score threshold of 0.5.)

F1 Score = 2 \* (Precision \* Recall/ (Precision + recall)) = 0.8949

**F1 Score**

What is this model’s F1 score?