

AutoML Modeling Report



<Fatimah Hamed Aljuhani>

Binary Classifier with Clean/Balanced Data

Train/Test Split

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

We generally use 80% from labeled data for training data, the other 20% for testing data
The total number of data = 200
Data used for training = 160
Data used for testing = 40

Total images	180
Test items	20
Precision ?	100%
Recall ?	100%

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 has 2 values: positive and negative
Columns represent: actual value
Rows represent: predicted value
Inside the matrix there are: TP, TN, FP, FN

	Predicted positive	Predicted negative
Actual positive	TP	FN
Actual negative	FP	TN

- The values I observe is: 100% true positive (TP) pneumonia "and it is true"
- 100% False positive (FP) normal "and it is false"

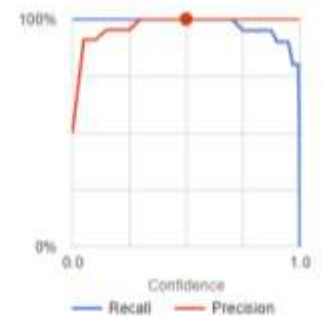
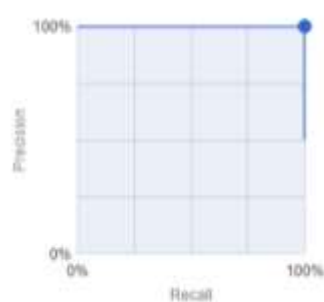
True Label	Predicted Label	
	pneumonia	normal
pneumonia	100%	-
normal	-	100%

Confusion matrix

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 and recall are two common measures we use when evaluating a model, both of them will measure the model such that we can understand how the model performs for an individual class, as well as how it performs across classes
- Recall is the number of relevant documents retrieved by a search divided by the total number of existing relevant documents, while precision is the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search.
- Precision = True Positive / True Positive + False positive
Recall = True Positive / True positive + False positive
- When confidence threshold = 0.5
Precision = 100%
Recall = 100%




Score Threshold

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

If I increase the threshold value: precision will increase
But Recall decreases.
And visa versa if I decrease the threshold value the
recall increase but precision decrease
Because the model has to have level of confidence to
assign a category risk of misclassifying images

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	Total number of data =400 Data used for training =320 Data used for testing= 80											
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	<p>Confusion matrix has changed.</p> <ul style="list-style-type: none">• 97% true positive(TP) pneumonia (in blue)• 70% true negative (TN) normal (in blue)• Some label confused 3% FN and 30% FP (gray)  <table><tr><th rowspan="2">True Label</th><th colspan="2">Predicted Label</th></tr><tr><th>pneumonia</th><th>normal</th></tr><tr><th>pneumonia</th><td>97%</td><td>3%</td></tr><tr><th>normal</th><td>30%</td><td>70%</td></tr></table>	True Label	Predicted Label		pneumonia	normal	pneumonia	97%	3%	normal	30%	70%
True Label	Predicted Label											
	pneumonia	normal										
pneumonia	97%	3%										
normal	30%	70%										
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)? Confidence threshold = 0.5 Precision=90% Recaal=90% It is high but it is not perfect because it is not 100%, and the model will produce less FP and less FN												



Unbalanced Classes

From what you have observed, how do unbalanced classes affect a machine learning model?

unbalanced classes make challenge for predictive modeling because most of the machine learning algorithms used for classification assumed that of an equal number of items for each class. This results in models that have poor predictive performance, specifically for class with less number of items

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.

Class1: 70% pneumonia + 30% normal
 Class2: 70% normal + 30% pneumonia
 This matrix shows most of model classified label correctly(blue) and some labels confused(geay).

- 50% true positive(TP) pneumonia
- 70% true negative (TN) normal
- Some label confused 50% FN and 30% FP

True Label	Predicted Label	
	pneumonia	normal
pneumonia	50%	50%
normal	30%	70%

Precision and Recall

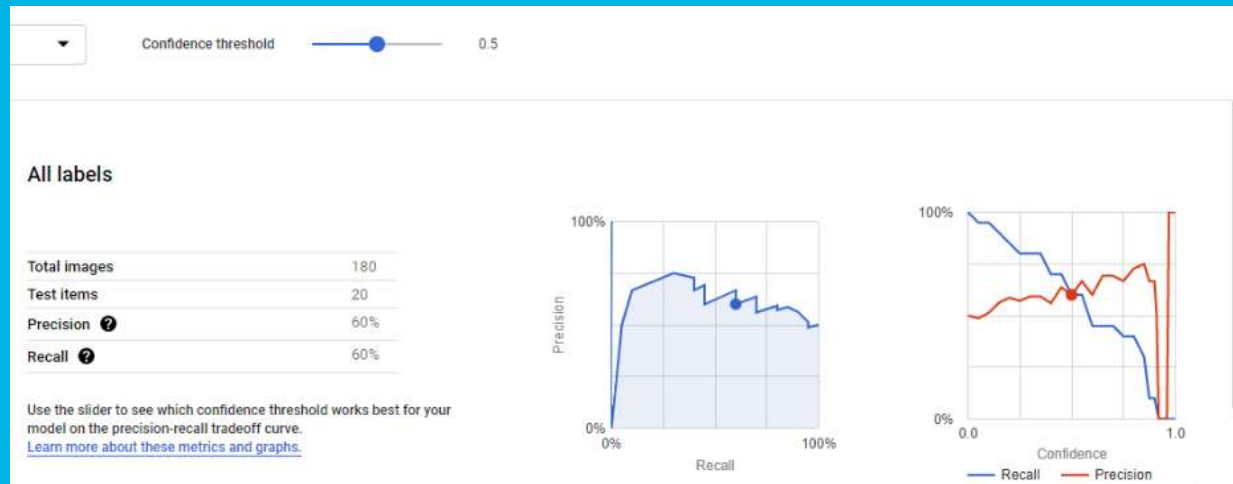
How have the model's precision and recall been affected by the dirty data (report the values for a score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?

Confidence threshold = 0.5

Precision=60%

Recall=60%

With the cleanbalanced classification data has the highest precision(100%) and highest recall(100%)



Dirty Data

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

Dirty data affects negatively on data mining and machine learning results.

It is affect data quality and accuracy of results.

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.

It is has 3 classes

- 1- 100 normal items from normal
- 2- 100 bacterial pneumonia items from pneumonia
- 3- 100 viral pneumonia items from pneumonia

The confusion matrix shows that the model highly classified (normal) labels 100%

Viral pneumonia label was predicated 70% and most often confused for viral pneumonia labels 30%
bacterial pneumonia label was predicated 80% and most often confused for bacterial pneumonia labels 20%

True Label \ Predicted Label	normal	viral PNEUMONIA	bacterial PNEUMONIA
normal	100%	-	-
viral PNEUMONIA	-	70%	30%
bacterial PNEUMONIA	10%	10%	80%

A high confusion implies incorrect predictions which may indicate inconsistent labels a way to remedy this confusion is to add more additional training data is a great remedy for the confusion also add more images

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)?

Precision = 86.21%

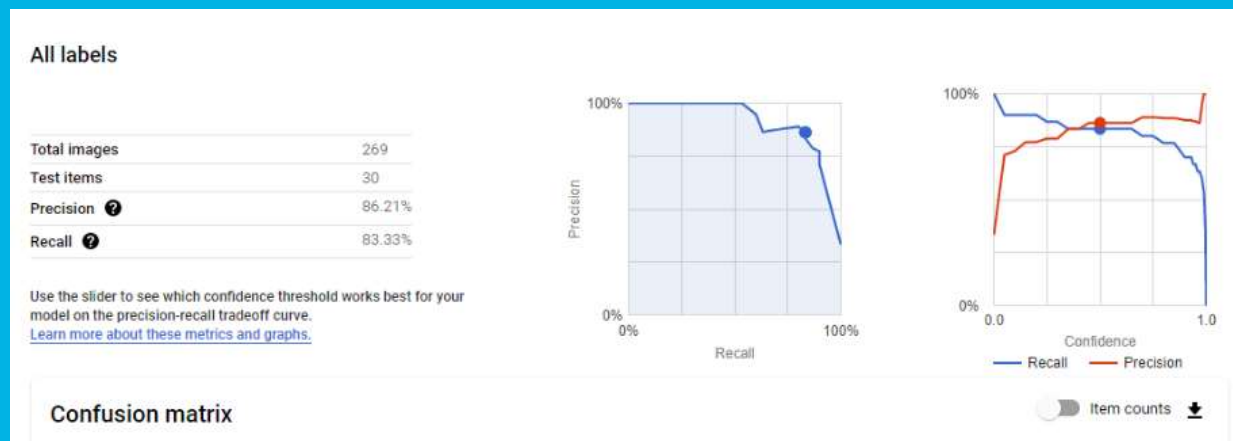
Recall = 83.33

Precision = $TP / (TP + FP)$

Recall = $TP / (TP + FN)$

Average precision = $(\text{cleanbalanced} + \text{cleanunbalanced} + \text{Dirtybalanced} + \text{threeclassmodel}) / 4$

Average precision = $(1 + 0.90 + 0.60 + 0.847) / 4 = 0.5$



F1 Score

What is this model's F1 score?

F1 score combines the precision and recall to produce an overall performance measure model

$$F1 = (2 * P_{\text{model}} * R_{\text{model}}) / (P_{\text{model}} + R_{\text{model}})$$

$$F1 = 2 * (P * R) / (P + R)$$

$$= 2 * (86.21 * 83.33) / (86.21 + 83.33)$$

$$= 84.74$$

$$= 0.84$$

	$= 1$ If F1 score in range from 0 to 1 being perfect precision in recall
--	---