

AutoML Modeling Report



<Nouf Al-Mutairi>

Binary Classifier with Clean/Balanced Data

Train/Test Split

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

For the first model we used 200 images for our dataset 100 were normal and 100 were Pneumonia 40 images were used for testing and 158 for training by the AWS Rekognition Platform.

Confusion Matrix

What do each of the sections 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 is a table that represent the prediction results and performance for the model. Shown the Confusion matrix for the first model below with 100% True Positive and 100% True negative

```
"ConfusionMatrix": [
  {
    "GroundTruthLabel": "normal",
    "PredictedLabel": "normal",
    "Value": 1.0
  },
  {
    "GroundTruthLabel": "normal",
    "PredictedLabel": "pneumonia",
    "Value": 0.0
  },
  {
    "GroundTruthLabel": "pneumonia",
    "PredictedLabel": "normal",
    "Value": 0.0
  },
  {
    "GroundTruthLabel": "pneumonia",
    "PredictedLabel": "pneumonia",
    "Value": 1.0
  }
]
```

Training Set			
TARGET \ OUTPUT	Normal	Pneumonia	SUM
Normal	20 50.00%	0 0.00%	20 100.00% 0.00%
Pneumonia	0 0.00%	20 50.00%	20 100.00% 0.00%
SUM	20 100.00% 0.00%	20 100.00% 0.00%	40 / 40 100.00% 0.00%

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 is the fraction of correct predictions (true positives) over all model predictions (true and false positives)¹. While Recall is the fraction of your test dataset labels that were predicted correctly¹. As for the first model at Score threshold of 0.5 we got a 100% for Precision and Recall.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	For the second model we used 300 images for our dataset 100 were normal and 200 were Pneumonia 60 images were used for testing and 238 for training by the AWS Rekognition Platform.																				
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix summary	<p>Confusion matrix is a table that represent the prediction results and performance for the model. Shown the Confusion matrix for the Second model below with 55% True Positive and 100% True negative</p> <pre>"ConfusionMatrix": [{ "GroundTruthLabel": "normal", "PredictedLabel": "normal", "Value": 0.55 }, { "GroundTruthLabel": "normal", "PredictedLabel": "pneumonia", "Value": 0.45 }, { "GroundTruthLabel": "pneumonia", "PredictedLabel": "normal", "Value": 0.0 }, { "GroundTruthLabel": "pneumonia", "PredictedLabel": "pneumonia", "Value": 1.0 }]</pre> <table><tr><th colspan="4">Training Set</th></tr><tr><th>TARGET \ OUTPUT</th><th>Normal</th><th>Pneumonia</th><th>SUM</th></tr><tr><th>Normal</th><td>11 18.33%</td><td>9 15.00%</td><td>20 55.00% 45.00%</td></tr><tr><th>Pneumonia</th><td>0 0.00%</td><td>40 66.67%</td><td>40 100.00% 0.00%</td></tr><tr><th>SUM</th><td>11 100.00% 0.00%</td><td>49 81.63% 18.37%</td><td>51 / 60 85.00% 15.00%</td></tr></table>	Training Set				TARGET \ OUTPUT	Normal	Pneumonia	SUM	Normal	11 18.33%	9 15.00%	20 55.00% 45.00%	Pneumonia	0 0.00%	40 66.67%	40 100.00% 0.00%	SUM	11 100.00% 0.00%	49 81.63% 18.37%	51 / 60 85.00% 15.00%
Training Set																					
TARGET \ OUTPUT	Normal	Pneumonia	SUM																		
Normal	11 18.33%	9 15.00%	20 55.00% 45.00%																		
Pneumonia	0 0.00%	40 66.67%	40 100.00% 0.00%																		
SUM	11 100.00% 0.00%	49 81.63% 18.37%	51 / 60 85.00% 15.00%																		
Precision and Recall How have the model's precision and recall been affected by the unbalanced data?	In the second model we got 55% for Precision and 100% for Recall.																				
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Since the data was skewed to the Pneumonia side the model had false negative cases (Normal predicted Pneumonia) because it has more examples on Pneumonia the model developed bias for 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 information.

The Confusion below shows a drop in the performance for the third model. The dirty data containing wrong labels resulted in the model predicting the wrong labels shown below with 70% True Positive and 95% True negative with 5% False positive.

```
"ConfusionMatrix": [
  {
    "GroundTruthLabel": "pneumonia",
    "PredictedLabel": "pneumonia",
    "Value": 0.95
  },
  {
    "GroundTruthLabel": "pneumonia",
    "PredictedLabel": "normal",
    "Value": 0.05
  },
  {
    "GroundTruthLabel": "normal",
    "PredictedLabel": "pneumonia",
    "Value": 0.3
  },
  {
    "GroundTruthLabel": "normal",
    "PredictedLabel": "normal",
    "Value": 0.7
  }
]
```

Training Set			
TARGET \ OUTPUT	Normal	Pneumonia	SUM
Normal	14 35.00%	6 15.00%	20 70.00% 30.00%
Pneumonia	1 2.50%	19 47.50%	20 95.00% 5.00%
SUM	15 93.33% 6.67%	25 76.00% 24.00%	33 / 40 82.50% 17.50%

Precision and Recall

How have the model's precision and recall been affected by the dirty data. Of the binary classifiers, which has the highest precision? Which has the highest recall?

In the second model we got 70% for Precision and 93.33% for Recall. Which performed poorly compared to the clean balanced data. With 70% for Precision and 93.33% for Recall in the dirty data against 100% for Precision and Recall in the clean balanced data it's clear that the highest measures were the clean balanced data measures.

Dirty Data

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

The dirty data confuses the model and creates contradictions in the training phase which results in misclassifications causing the poor performance provided by the model.

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 information.

Complete later

In the confusion matrix below we can see that in the 3 classes' classification case. The model confused the bacterial and viral pneumonia due to the low dataset size for each class. Since the model can get the normal class right almost every time. Moreover, we need to add more bacterial class data to help the model be familiar with the class characteristics, Due to the almost 10% mistake rate shown in the matrix below.

Training Set				
TARGET OUTPUT	Normal	Bacterial	Viral Pneumonia	SUM
Normal	20 33.33%	0 0.00%	0 0.00%	20 100.00% 0.00%
Bacterial	0 0.00%	19 31.67%	1 1.67%	20 95.00% 5.00%
Viral Pneumonia	0 0.00%	2 3.33%	18 30.00%	20 90.00% 10.00%
SUM	20 100.00% 0.00%	21 90.48% 9.52%	19 94.74% 5.26%	57 / 60 95.00% 5.00%

Precision and Recall

What are the model's precision and recall? How are these values calculated?

In the fourth model we can calculate the Precision and Recall by calculating each class value then divided them by 3 to get the overall measures. With 98.33% for precision and Recall, we can see that the 3 class model performed better than the dirty balanced and the unbalanced model but still needs improvement to pass the clean balanced model.

F1 Score

The F1 Score of this scenario would be 98%. which can

What is this model's F1 score?

be calculated by the equation below:

$$F1 = 2 \times \frac{P \times R}{P + R}$$

Where P and R are assigned to precision and Recall respectively.