AutoML Modeling Report

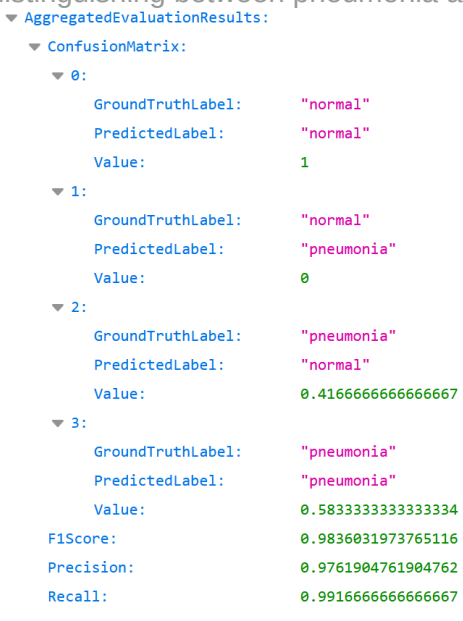
<Aseel Meshal AlKhadaidi>

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	There were **99** images for normal class, and **99** for the pneumonia class. **80%** of them combined were for training which is **158** (**79** from normal class and **79** from pneumonia), and **20%** for testing which is **40** (**20** from normal class and **20** from pneumonia).
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?	1. **True Positives (TP)**: The number of pneumonia Images that were correctly predicted as the pneumonia class. 2. **False Positives (FP)**: The number of normal Images that were wrongfully predicted as the pneumonia class. 3. **True Negatives (TN)**: The number of normal Images that were correctly predicted as the normal class. 4. **False Negatives (FN)**: The number of pneumonia Images that were wrongfully predicted as the normal class. AggregatedEvaluationResults: ConfusionMatrix: 0: GroundTruthLabel: "normal" PredictedLabel: "normal" Value: 0.9 1: GroundTruthLabel: "normal" PredictedLabel: "pneumonia" Value: 0.1 2: GroundTruthLabel: "pneumonia" PredictedLabel: "normal" Value: 0 3: GroundTruthLabel: "pneumonia" PredictedLabel: "pneumonia" Value: 1 F1Score: 1 Precision: 1 Recall: 1 The **TPR**(True positive rate) is **100%** which is calculated as TP/TP+FN and this means that the model correctly

	identified all pneumonia cases as pneumonia class, while **FPR**(False positive rate)about **10%** of normal cases were Identified incorrectly as pneumonia class.
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)?	• **Recall:** tells you how well the model identifies actual positive cases. • **Precision**: tells you how many of the predicted positive cases were actually correct. • **Threshold of 0.5:** **Precision Calculated as:** TP/TP+FP = 1/1+0.1= 0.909 = 90.9%. or given as 1. **Recall Calculated as:** TP/TP+FN=1/1+0=1=100% given as 1. **F1 Score:** given as 1. **This indicates that for a threshold of 0.5 the model performed extremely well with precision and recall both achieving perfect scores, which indicates a perfect model evaluation.**

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	There were 99 images for normal class\cases, and 299 for the pneumonia class\cases. 80% of them were for training (318 Images) and 20% for testing (80 Images).
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix summary	The imbalance in the data, with more pneumonia cases than normal cases, likely led to the model focusing more on pneumonia detection, causing a higher error rate when distinguishing between pneumonia and normal cases. 

Precision and Recall How have the model's precision and recall been affected by the unbalanced data?	**The precision:** Despite the imbalance, the precision is still high, meaning that when the model predicts pneumonia, it is mostly correct. This is likely because there were very few false positives **The recall:** It is also very high, indicating that the model is very good at detecting pneumonia cases, which can be attributed to the large number of pneumonia cases in the training data. **Although the imbalanced data didn't affect the precision and recall but it did result in some misclassification of the pneumonia cases as normal.**
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	• There was bias towards the pneumonia class, which indicates that the normal class learning is less effective. • There could be lower recall for normal class, which means it might not be well detected because the model wasn't trained on more examples (Not enough).

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.	• There were 130 images for normal class (that includes 30 images of pneumonia cases), and 130 for the pneumonia class (that includes 30 images of normal cases) 80% of them were for training(208 Images) and 20% for testing (52 Images). • Dirty data has caused a big increase in false positives, where normal cases are mistakenly identified as pneumonia. However, the model's accuracy in identifying pneumonia cases remains high. This shows that dirty data can have a greater impact on normal cases, causing uneven performance across the different classes.

<table>
<tr>
<td></td>
<td>
<pre>
AggregatedEvaluationResults:
  ConfusionMatrix:
    0:
      GroundTruthLabel: "normal"
      PredictedLabel: "normal"
      Value: 0.7307692307692307
    1:
      GroundTruthLabel: "normal"
      PredictedLabel: "pneumonia"
      Value: 0.2692307692307692
    2:
      GroundTruthLabel: "pneumonia"
      PredictedLabel: "normal"
      Value: 0
    3:
      GroundTruthLabel: "pneumonia"
      PredictedLabel: "pneumonia"
      Value: 1
  F1Score: 0.8864285714285713
  Precision: 0.875
  Recall: 0.9038461538461539
</pre>
</td>
</tr>
<tr>
<td>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?</td>
<td>

Precision: 0.875, impacted by dirty data, causing more false positives in pneumonia predictions.
Recall: 0.9038, slightly lower due to mislabeled data, but still strong at identifying pneumonia cases.
Highest Precision: Pneumonia predictions at 0.875, accurate in predicting pneumonia.
Highest Recall: Pneumonia predictions at 0.9038, effectively detecting most pneumonia cases.

</td>
</tr>
<tr>
<td>Dirty Data
From what you have observed, how does dirty data affect a machine learning model?</td>
<td>

Increased misclassifications: Dirty data can confuse the model, leading to more false positives or false negatives.
Reduced confidence: It makes it harder for the model to learn correct decision boundaries, reducing precision and recall.
Bias in the model: Dirty data can cause bias, especially if one class is more affected than the other, skewing predictions.

</td>
</tr>
</table>

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.

Save Copy Collapse All Expand All Filter JSON

```
AggregatedEvaluationResults:
  ConfusionMatrix:
    0:
        GroundTruthLabel:   "bacterial_pneumonia"
        PredictedLabel:     "bacterial_pneumonia"
        Value:              0.95
    1:
        GroundTruthLabel:   "bacterial_pneumonia"
        PredictedLabel:     "normal"
        Value:              0
    2:
        GroundTruthLabel:   "bacterial_pneumonia"
        PredictedLabel:     "viral_pneumonia"
        Value:              0.05
    3:
        GroundTruthLabel:   "normal"
        PredictedLabel:     "bacterial_pneumonia"
        Value:              0
    4:
        GroundTruthLabel:   "normal"
        PredictedLabel:     "normal"
        Value:              1
    5:
        GroundTruthLabel:   "normal"
        PredictedLabel:     "viral_pneumonia"
        Value:              0
    6:
        GroundTruthLabel:   "viral_pneumonia"
        PredictedLabel:     "bacterial_pneumonia"
        Value:              0.25
    7:
        GroundTruthLabel:   "viral_pneumonia"
        PredictedLabel:     "normal"
        Value:              0.05
    8:
        GroundTruthLabel:   "viral_pneumonia"
        PredictedLabel:     "viral_pneumonia"
        Value:              0.7
  F1Score:                  0.9657894736842105
  Precision:                0.9833333333333334
  Recall:                   0.9499999999999998
```


Summary of confusion matrix:

- **Bacterial Pneumonia**:

1. Correctly classified as bacterial pneumonia: **95%**.
2. Incorrectly classified as viral pneumonia: **5%**.
3. Incorrectly classified as normal: **0%**.

- **Normal**:

1. Correctly classified as normal: **100%**.
2. Incorrectly classified as bacterial pneumonia or viral pneumonia: **0%**.

- **Viral Pneumonia**:

1. Correctly classified as viral pneumonia: **70%**.
2. Incorrectly classified as bacterial pneumonia:

	25%. 3. Incorrectly classified as normal: **5%**. • **Most Likely to Confuse**: The model seems to confuse **viral pneumonia** with **bacterial pneumonia** (25%). This suggests that the features between these two classes are harder for the model to distinguish. • **Most Likely to Get Right**: The model gets the **normal** class perfectly correct (100%). It might also get it right for bacterial pneumonia since it's 95%. • **What would I do to make the model confusion less**? Increase the complexity of the examples of viral pneumonia to make the model less confusing.
Precision and Recall What are the model's precision and recall? How are these values calculated?	**First I calculated the Precision for each class by (TP/TP+FP) and it resulted in:** **Normal class Precision:** 1 **Viral pneumonia class Precision:** 0.95 **bacterial pneumonia class Precision:** 1 **Then I calculated the Recall for each class by (TP/TP+FN) and it resulted in:** **Normal class Precision:** 1 **Viral pneumonia class Precision:** 0.7 **bacterial pneumonia class Precision:** 0.95 **Lastly, I calculated the precision for all classes by combining all their values and dividing by number:** **(1+0.95+1)/3= 0.9833** **The Recall for all classes by combining all their values and dividing by their number :** **(1+0.7+0.95)/3= 0.8833**
F1 Score What is this model's F1 score?	**I calculated the F1 Score for each class by (2*[Precision*Recall / Precision + Recall]) and it resulted in:** **Normal class F1 Score:** 1 **Viral pneumonia class F1 Score:** 0.81 **bacterial pneumonia class F1 Score:** 0.97 **Finally, I calculated the F1 Score for All classes by combining all their values and dividing by number:** **(0.97+0.81+1)/3=0.93**