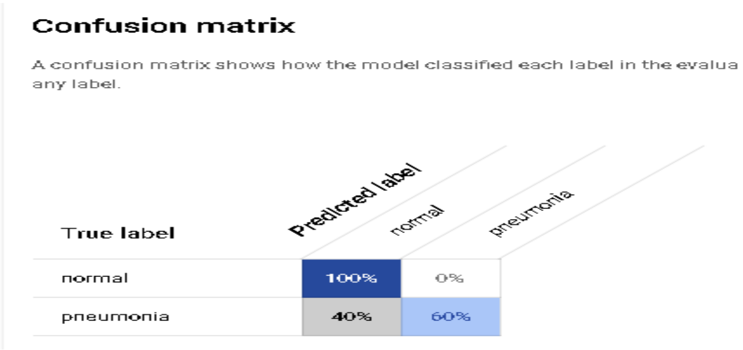
AutoML Modeling Report

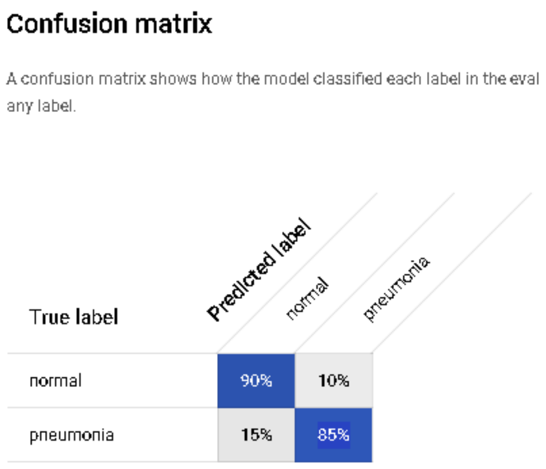
Abdulelah Almajed

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	200 images in total, 80% of them were used for training the model (160 images), 10% for validation (20 images), and 10% for testing (20 images).
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** A confusion matrix shows how the model classified each label in the evalua any label. True label | Predicted label: normal | pneumonia normal | 100% | 0% pneumonia | 40% | 60% • The confusion matrix shows the percentages of correct and incorrect predictions for each class. • The provided values indicate that all "normal" instances are correctly predicted, while 60% of "pneumonia" instances are correctly predicted, and 40% are incorrectly predicted as "normal." • The true positive rate (recall) for the "pneumonia" class is 60%. • The false positive rate for the "normal" class is 0%.
Precision and Recall What does precision measure? What does recall measure? What precision and recall did the model	**Precision** measures the proportion of true positive predictions among all positive predictions made by a model. In other words, it tells us how many of the instances predicted as positive by the model are actually

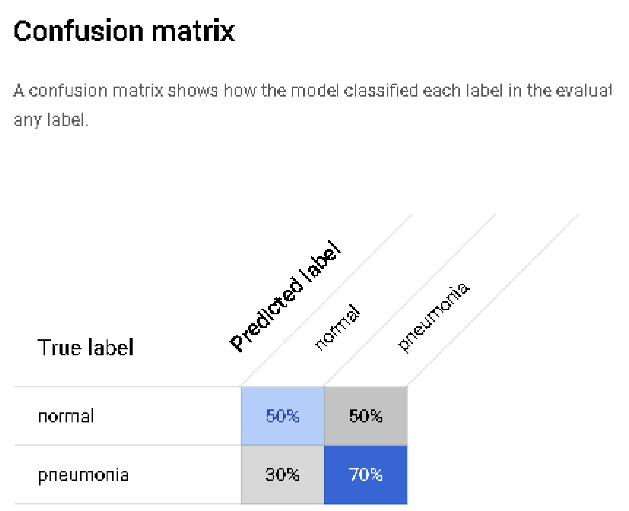
achieve (report the values for a score threshold of 0.5)?	positive. **Recall**, also known as sensitivity or true positive rate (TPR), measures the proportion of actual positive instances that are correctly predicted as positive by the model. It tells us how many of the actual positive instances the model is able to identify correctly. **Normal Class:** *Precision = 100% / (100% + 40%) = 100% / 140% = 71.43%* *Recall = 100% / (100% + 0%) = 100% / 100% = 100%* **Pneumonia Class:** *Precision = 60% / (60% + 0%) = 60% / 60% = 100%* *Recall = 60% / (60% + 40%) = 60% / 100% = 60%*
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	**Precision** tends to increase with a higher threshold. The model becomes more selective, requiring higher confidence to predict instances as positive. This reduces false positives and improves precision. **Recall** tends to decrease with a higher threshold. The model becomes more conservative, potentially missing some actual positive instances (false negatives). This reduces recall. So there is a trade-off between Recall and Precision. **Higher threshold** Favors precision, reduces false positives, but may miss some true positives. **Lower threshold** Favors recall, captures more true positives, but may include more false positives. Therefore, increasing the score threshold generally improves precision at the expense of recall, while decreasing the threshold improves recall at the expense of precision. The choice of threshold should align with the specific goals and priorities of the problem at hand.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	300 images in total, 80% of them were used for training the model (240 images), 10% for validation (30 images), and 10% for testing (30 images).
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 **Confusion matrix** A confusion matrix shows how the model classified each label in the eval any label. True label / Predicted label: normal, pneumonia normal: 90%, 10% pneumonia: 15%, 85% The unbalanced data has affected the confusion matrix by introducing a bias towards the majority "normal" class. This is evident from the higher sensitivity (90%) and lower false negative rate (10%) for the "normal" class compared to the "pneumonia" class, which has a lower sensitivity (85%) and a higher false negative rate (15%).
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)?	**Normal Class:** *Precision = 9 / (9 + 3) = 75%* *Recall = 9 / (9 + 1) = 90%* **Pneumonia Class:** *Precision = 17 / (17 + 1) = 94.4%* *Recall = 17 / (17 + 3) = 85%* The unbalanced case performs better for the "pneumonia" class compared to the balanced case. The precision for "pneumonia" is similar in both cases, but the recall is significantly higher in the unbalanced case (85%) than in the balanced case (60%). This means the model in the unbalanced case is better at correctly identifying actual "pneumonia" cases, which is crucial in medical diagnosis to minimize missed cases. The improved performance for the "pneumonia" class in the unbalanced case can be attributed to the larger

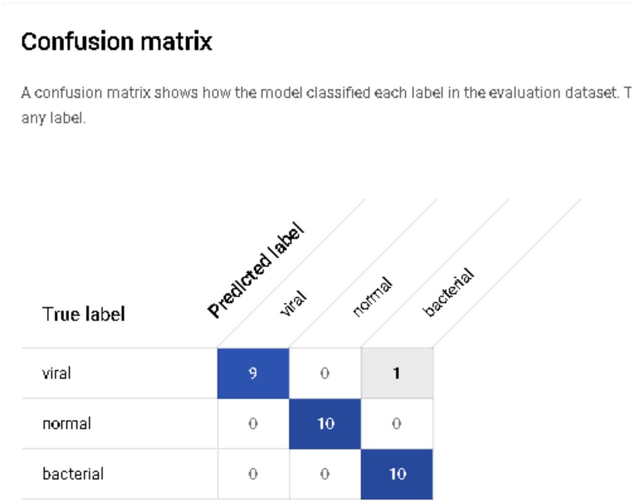
	number of "pneumonia" images (200) compared to the balanced dataset, allowing the model to learn more diverse features associated with the "pneumonia" class.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Unbalanced classes in a machine learning model can lead to **biased predictions** and **poor performance**, especially for the minority class. The model may be heavily influenced by the majority class, resulting in high accuracy for the majority class but low accuracy for the minority class. This can be problematic when the minority class is of greater importance or interest.

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.	 **Confusion matrix** A confusion matrix shows how the model classified each label in the evaluat any label. True label · Predicted label: normal · pneumonia normal · 50% · 50% pneumonia · 30% · 70% The introduction of dirty data, where 30% of the labels in each class were switched, has significantly affected the confusion matrix. The mislabeling has led to an increase in both **false positives** and **false negatives**, reducing the overall accuracy of the classifier to 42.86%. The presence of dirty data has degraded the classifier's ability to distinguish between "normal" and "pneumonia" cases, leading to a high number of misclassifications.
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	**Normal Class:** *Precision = 5 / (5 + 3) = 62.5%* *Recall = 5 / (5 + 5) = 50%* **Pneumonia Class:** *Precision = 7 / (7 + 5) = 58.3%* *Recall = 7 / (7 + 3) = 70%*

highest precision? Which has the highest recall?	the dirty data has reduced the precision and recall values for both classes. The "normal" class has the highest precision, while the "pneumonia" class has the highest recall.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data can negatively impact a machine learning model by **reducing accuracy**, **introducing biases**, and **causing overfitting**. In this use case, we specifically dealt with mislabeled data, a type of dirty data, by intentionally switching the labels of 30 images in each class. Mislabeled data can confuse the model during training, leading to reduced accuracy and poor performance. It's essential to identify and correct mislabeled instances to ensure the model learns from accurate examples and maintains high-quality performance.

3-Class Model

<table>
<tr>
<td>**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.</td>
<td>


Confusion matrix

A confusion matrix shows how the model classified each label in the evaluation dataset. T any label.

<table>
<tr><td>True label</td><td>Predicted label viral</td><td>normal</td><td>bacterial</td></tr>
<tr><td>viral</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td>normal</td><td>0</td><td>10</td><td>0</td></tr>
<tr><td>bacterial</td><td>0</td><td>0</td><td>10</td></tr>
</table>

The model performs well in distinguishing between the three classes, with high accuracy for each class.

1. **Normal class:**

- The model correctly predicted all 10 normal cases as normal.
- There were no misclassifications of normal cases as either viral or bacterial pneumonia.

2. **Viral pneumonia class:**

- The model correctly predicted 9 out of 10 viral
</td>
</tr>
</table>

	pneumonia cases. • It misclassified one viral pneumonia case as bacterial pneumonia. • There were no misclassifications of viral pneumonia as normal. **3. Bacterial pneumonia class:** • The model correctly predicted all 10 bacterial pneumonia cases as bacterial pneumonia. • There were no misclassifications of bacterial pneumonia as either normal or viral pneumonia. The model seems to be most likely to confuse viral pneumonia with bacterial pneumonia, as evidenced by the one misclassification. However, this confusion is minimal, and the model still demonstrates high accuracy in distinguishing between the two types of pneumonia. The model is most likely to get the normal and bacterial pneumonia classes right, as it achieves perfect classification for both these classes.
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)?	**1. Normal Class:** • *Precision = 10 / (10 + 0) = 1* • *Recall = 10 / (10 + 0) = 1* **2. Viral Class:** • *Precision = 9 / (9 + 0) = 1* • *Recall = 9 / (9 + 1) = 0.9* **3. Bacterial Class:** • *Precision = 10 / (10 + 1) = 0.91* • *Recall = 10 / (10 + 0) = 1* The model achieves high precision and recall for all classes, with perfect scores for the normal class. Viral pneumonia has perfect precision but slightly lower recall due to one false negative. Bacterial pneumonia has high recall but slightly lower precision due to one false positive.
F1 Score What is this model's F1 score?	**1. Normal class:** • *Precision = 1.0* • *Recall = 1.0* • *F1 = 2 * (1.0 * 1.0) / (1.0 + 1.0) = 1.0* **2. Viral class:** • *Precision = 1.0* • *Recall = 0.9* • *F1 = 2 * (1.0 * 0.9) / (1.0 + 0.9) ≈ 0.95* **3. Bacterial class:** • *Precision = 0.91* • *Recall = 1.0*

	• *F1 = 2 * (0.91 * 1.0) / (0.91 + 1.0) ≈ 0.95* **F1 Score = (1.0 + 0.95 + 0.95) / 3 ≈ 0.97** Therefore, the overall F1 score for this model is approximately 0.97, indicating a high level of performance across all three classes.