AutoML Modeling Report

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	• Training Data: 80% of the dataset was used for training. Given that each class (normal and pneumonia) had 100 images, this means 80 images per class were used for training. • Testing Data: 10% of the dataset was used for testing, which equates to 10 images per class. Note that 10% of the data is used for validation for early stopping method.
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?	• Percentage of actual normal cases correctly identified.- 90% • Percentage of normal cases incorrectly identified as Pneumonia.- 10% • Percentage of pneumonia cases incorrectly identified as normal. – 0% • Percentage of actual pneumonia cases correctly identified. – 100% • True Positive Rate for Pneumonia: 100% • False Positive Rate for Normal: 10% **Screenshot is attached in the zip file, please check!!**
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: Measures the accuracy of positive predictions, i.e., the percentage of positive identifications that were actually correct. • Recall: Measures the ability of the model to find all the positive cases, i.e., the percentage of actual positives that were correctly identified. Model's Precision and Recall: • Precision at 0.5 threshold: 95% • Recall at 0.5 threshold: 95%

Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	• Increasing the Threshold: • Effect on Precision: Generally, increasing the threshold improves precision because the model becomes more conservative in predicting positive cases, ensuring a higher confidence in its predictions. • Effect on Recall: Increasing the threshold typically decreases recall, as the model might miss some actual positive cases due to stricter criteria for predicting positives. • Reason: This behavior is because a higher threshold requires higher confidence for a prediction to be considered positive, potentially reducing false positives (hence higher precision) but might also miss out on true positives (hence lower recall).

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	• Training Data: 160 pneumonia, 80 normal images. • Testing Data: 20 pneumonia, 10 normal images. • Validation Data: 20 pneumonia, 10 normal images.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	• Observation: Unbalanced data doesn't seem to negatively impact the model's performance in this case. • Matrix: 100% accuracy for both classes. **Screenshot is attached in the zip file, please check!!**
Precision and Recall How have the model's precision	• Precision & Recall at 0.5 Threshold: Both are 100%.

and recall been affected by the unbalanced data (report the values for a score threshold of 0.5)?	• Effect of Unbalanced Data: Surprisingly, the model performed better. This could be due to the model being effectively trained to recognize the over-represented class (pneumonia) and maintaining good performance on the under-represented class (normal).
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	• General Impact: Unbalanced classes can lead to models being biased towards the majority class. • Observation in This Case: Despite the imbalance, the model showed exceptional performance, possibly because the distinguishing features of pneumonia in X-rays are pronounced enough for the model to learn effectively even with the imbalance. • **Note**: The unexpected increase in precision and recall could be attributed to the specific nature of the dataset, where the features distinguishing pneumonia might be significantly prominent, making it easier for the model to learn and generalize, even with unbalanced data.

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.	• Impact of Dirty Data: Significant performance reduction. Matrix Details: • Pneumonia Class Correctly Identified: 40% • Pneumonia Class Incorrectly Identified as Normal: 60% • Normal Class Correctly Identified: 80% • Normal Class Incorrectly Identified as Pneumonia: 20% **Screenshot is attached in the zip file, please check!!**
Precision and Recall How have the model's precision and recall been affected by the dirty data (report the values for a	• At 0.5 Threshold: Both precision and recall are 60%. • Comparison: This model shows the lowest

score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?	precision and recall among the three models. • Highest Performance: The unbalanced model had the highest precision and recall (both 100%).
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	• Effect on Model Performance: Dirty data significantly degrades model accuracy. The incorrect labels in the training set lead to confusion in learning, resulting in poor classification accuracy. • Observation: The drastic decrease in performance metrics highlights the critical need for clean, correctly labeled data in training machine learning models.

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.	Summary: • Bacteria-Pneumonia: Correctly identified 100% of the time. • Virus-Pneumonia: Correctly identified 60% of the time, 40% confused with Bacteria-Pneumonia. • Normal: Correctly identified 100% of the time. • Most Likely Confusion: Between Bacteria-Pneumonia and Virus-Pneumonia. • Best Performance: On Bacteria-Pneumonia and Normal classes. • Remedy: Enhance feature differentiation between Bacteria and Virus Pneumonia in the training data, possibly through more diverse examples or advanced preprocessing techniques. **Screenshot is attached in the zip file, please check!!**
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: 92.9% • Recall: 86.7% • Calculation Basis for precision and recall:

	P = TP / TP + FP R = TP / TP + FN • As far as mutli-class classification problem is concerned we can calculate the overall precision, overall recall by calculating the precision for each class individually and then dividing them by total number of classes. (Macro - Averaging) • Otherwise, we can sum the total number of TP, FP, FN and then apply in the above formula. • Note: The confusion matrix in Google AutoML (Vertex AI) remains unchanged when adjusting the threshold. This is expected behavior, as AutoML utilizes the argmax function for displaying the confusion matrix. However, recall and precision changes according to the threshold when sliding it. (refer link2) Nice/Helpful References: link1 link2
F1 Score What is this model's F1 score?	• F1 Score Calculation: The F1 score is the harmonic mean of precision and recall. • F1 Score = 2 × (precision × recall) / (precision + recall). • Model's F1 Score: Calculating using the given precision and recall at threshold 0.5 F1 score = 2(0.927 X 0.867) / (0.927 + 0.867) = **0.897**