AutoML Modeling Report

<Your Name Here>

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	 80% data is used for training, 20% is for testing
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 shows the difference between predicted label and actual label. In this case, it's a 2*2 matrix to show if desired categories (pneumonia and normal) are being identified correctly. The value demonstrates if the model is classifying test images correctly. **Confusion matrix** True positive rate for the pneumonia is 100%. False positive rate for normal is 0.
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 for all test samples which have been labeled, how many are supposed to be labeled in the way. Recall measures for all test samples which should have been labeled in the way, how many are actually labeled. With threshold of 0.5. Precision = 1, Recall = 1
Score Threshold When you increase the threshold	When increase threshold to 1, Precision = 1, Recall = 0. When threshold is 1 which is the highest, the model will label no

|	
what happens to precision? What happens to recall? Why?	image to lower risk of misclassifying images. So recall changes to 0. Confidence threshold ━━━● 1 **All labels** | Total images	180
Test items	20
Precision ❓	100%
Recall ❓	0% | Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs.
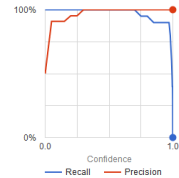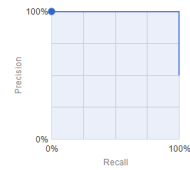
Binary Classifier with Clean/Unbalanced Data

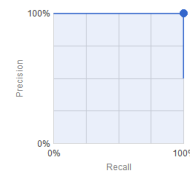|	
Train/Test Split How much data was used for training? How much data was used for testing?	90% data is for training, 10% is for testing **All labels** 	Total images	360
Test items	40
Precision ❓	100%
Recall ❓	100%	
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	**All labels** 	Total images	360
Test items	40
Precision ❓	100%
Recall ❓	100% | Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs. **Confusion matrix** | True Label \ Predicted Label	pneumonia	normal
---	---	---
pneumonia	100%	-
normal	-	100%	 With unbalanced classes, there is no change to confusion matrix. Normally, we would expect to see a bias towards the crowded class but this did not happen in my case. Model performance relies on many factors as well as crowdedness. But in this case, unbalanced data doesn't have impact to confusion matrix.
Precision and Recall How have the model's precision and recall been affected by the	As per previous explanation, no change to precision and recall.

unbalanced data (report the values for a score threshold of 0.5)?	
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	In this test, unbalanced classes have no impact to model performance.

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.	**All labels** Total images 180 Test items 20 Precision 90% Recall 90% Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs. **Confusion matrix** Model starts to misclassify images. TN changes to 80%. FP changes to 20%.
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?	Model precision and recall change to 90%. Pneumonia has 100% precision versus 83.33% of normal. **pneumonia** Total images 180 Test items 0 Precision 100% Recall 80% Normal has 100% recall versus 80% of pneumonia. **normal** Total images 180 Test items 0 Precision 83.33% Recall 100%
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data has destabilized model. It demonstrates the concept "garbage in garbage out" scenario. If the model is trained with dirty data, the model would have issue to validate test data. In this case, it reduces accuracy.

3-Class Model

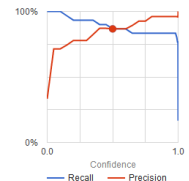|	
Confusion Matrix Summarize the 3-class confusion matrix. Which classes is the model most likely to confuse? Which class(es) is the model most likely to get right? Why might you do to try to remedy the model's "confusion"? Include a screenshot of the new confusion matrix.	Model is more likely to get normal class right. Model has certain confusion to differentiate viral pneumonia and bacterial pneumonia. In this case, virus pneumonia has low accuracy. A few things can be considered to improve model confusion. 1) Medical staff can be involved for some feedback like any more symptoms could be identified as input. 2) More viral and bacterial data can be added like increasing from 100 to 200. 3) Consider demographic of input data, may differentiate by gender and age group when build model. **All labels** 	Total images	270
Test items	30
Precision ⓘ	86.67%
Recall ⓘ	86.67% | Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs. **Confusion matrix** | True Label	normal	viral pneumonia	bacterial pneumonia
normal	100%	-	-
viral pneumonia	-	90%	10%
bacterial pneumonia	-	-	100%	
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)?	Overall model precision = 88.6%, recall = 86.67%. The calculation is to get each precision and recall from every class then take the average. Normal: precision = 8/8=100%, recall = 8/(8+2) = 80% Viral pneumonia: precision = 8/(8+1) = 88.89%, recall = 8/(8+2) = 80% Bacterial pneumonia: precision = 10/(10+3)= 76.92%, recall = 10/10 = 100% So precision = (100%+88.89%+76.92%)/3 = 88.6% Recall = (80% + 80% + 100%)/3 = 86.67%
F1 Score What is this model's F1 score?	F1 = 2*88.6%*86.67%/(88.6%+86.67%) = 0.87