AutoML Modeling Report

<Rahul Sarkar>

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	Training Items: 180 Test Items: 20
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?	The confusion matrix mainly has 4 components: True Positive, False Positive, False Negative and True Negative. True Positive is when positive class samples are detected correctly. True Negative is when negative class samples are detected correctly. False Positive is when negative class sample is predicted as positive. False Negative is when positive class sample is predicted as negative T.P=9, T.N=10, F.P=0, F.N=1 True Positive Rate for pneumonia = 0.9 False Positive Rate for normal = 0 **Note**: *screenshots are provided in the last page*
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 tells us how many of the positive class predictions **actually** belong to positive class. Recall tells us how many of the positive class instances were correctly predicted out of the actual positive class instances. Precision=1, Recall=0.9 at threshold=0.5
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	If threshold is increased, precision increases but recall decreases and vice versa if threshold is decreased. The reason is when threshold is increased, we are basically increasing the cutoff for the positive class and thus the false positive decreases leading to high precision and false negative increases thus leading to

	less recall. And when threshold is reduced, false positives increase and false negative reduces as most of the entries are classified as positive by the model due to lower threshold, thus leading to high recall and low precision.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	Training Items: 360 Testing: 40
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	Since the data was unbalanced with more pneumonia class instances, the model has learnt to predict pneumonia cases very well. Therefore, it has zero false positives. But due to less dataset of normal class, the model misclassified one of the images wrongly.
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)?	T.P=29,F.P=0,F.N=1,T.N=10 Precision=1, Recall=0.97 Since the model learns to identify all the positive classes very well, it has zero false positive cases, thus achieving 100% precision. But since the negative class instances are very less in the training dataset, the false negative cases have reduced, thus leading to higher recall value.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Unbalanced dataset causes the model to learn more about the majority label in the dataset. Thus, the model often predicts the minor class instance as major class instance, thus lacking the power to identify the minor class data. This is what happened when we provided more instances of pneumonia than normal. The model predicted pneumonia cases correctly, but it also classified one image wrongly.

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.	As the training dataset had dirty data, the model was not able to learn correctly. This resulted in large number of false positive and false negative cases. Both precision and recall reduced due to wrong predictions.

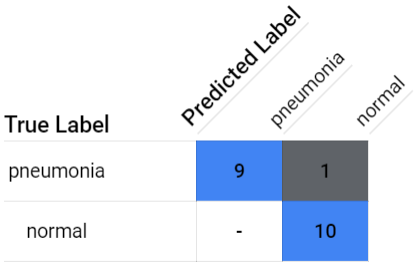
How have the model's precision and recall been affected by the dirty data (report the values for a score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?	As both the false positive and false negative cases increased, precision and recall reduced. Precision=0.6, Recall=0.6 Among all the binary classifiers till now, the first classifier with clean data and the second classifier with unbalanced data had the highest precision of 1 and the second classifier with unbalanced data had the highest recall value of 0.97.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data doesn't help the model to learn class specific features, which leads to wrong prediction, thus leading to low precision and recall. It results in large number of false positive and false negative cases.

3-Class Model

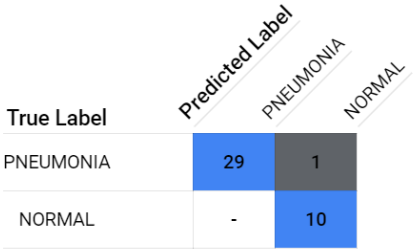
Confusion Matrix Summarize the 3-class confusion matrix. Which classes is the model most likely to confuse? Which class(es) is the model most likely to get right? Why might you do to try to remedy the model's "confusion"? Include a screenshot of the new confusion matrix.	Precision of viral class is 0.77 and that of normal and bacterial is 1. Recall of viral and normal is 1 and that of bacterial is 0.7. The model learnt to identify normal class with 0 false positives and false negatives, but got confused between viral and bacterial class for 3 test instances. Thus, the model is likely to predict normal class correctly with high accuracy. The remedies can be as follows: • Use better quality images for the bacterial and viral instances so that model learns easily. • Increase dataset size for all 3 classes so that the model can learn from more data. • Try to include data from diverse sample set so that the dataset is representative of actual population.
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 = Average (Precision of normal + Precision of viral + Precision of bacterial) = 0.923 Recall = Average (Recall of normal + Recall of viral + Recall of bacterial) = 0.90
F1 Score What is this model's F1 score?	F1 score = Harmonic mean of precision and recall F1 score = 0.911

Screenshots of Confusion Matrix

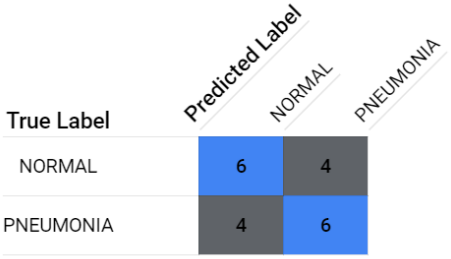
- For clean and balanced dataset:



- For clean and unbalanced dataset:



- For dirty dataset:



- For multi class dataset: