AutoML Modeling Report

<Sung Jun Eun>

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	In total 200 data, 180 were used for training and 20 were used 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?	Upleft cell describes ratio that model predicted image as a normal and it was actually normal. Downleft cell describes ratio that model predicted it as normal but it was actually pneumonia. Upright cell describes ratio that model predicted image as pneumonia but it was actually normal image. Downright cell describes ratio that model predicted image as pneumonia and it was actually pneumonia. For normal images, model correctly predicted them for 100% accurate. For pneumonia images model predicted correctly for 90% and made wrong prediction for 10%. True positive rate for 'pneumonia' is 90% which is value of downright cell. False positive rate for 'normal' is 10% which is value of downleft cell.
Precision and Recall What does precision measure? What does recall measure? What precision and recall did the model achieve (report the values for a	Precision is the proportion of true positives of total predicted positives. Recall is the proportion of true positive of total actual positives. This model got 95% for both precision and recall value.

score threshold of 0.5)?	
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	Precision increases and recall decreases. As we increase the threshold, we give a model a higher and stricter standard for prediction. It makes model to predict less. Because, precision uses total prediction for denominator, precision value increases. However, as recall value uses correct prediction for numerator, recall value decreases.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	In total 400 data, 360 images are used for training and 40 images were used for testing.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	At the first model, it had 10% of wrong prediction for 'pneumonia' images. However, as we increased the pneumonia data, the model became more accurate and eventually got 97% accuracy. However, at the first model, it predicted normal images perfectly, but now it predicts only 90% correct.
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)?	Precision and recall value are both 93.33%. Which is decreased from the first case.
Unbalanced Classes From what you have observed, how do unbalanced classed affect	As the dataset becomes unbalanced, it affects model to be biased and make more mistakes compare to model when we trained with balanced dataset. The model may

a machine learning model?	become more accurate to the label with more data, but overall accuracy will decline.

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.	For the pneumonia image, model could only predict 50% accurately and for other 50%, it predicted normal image. For the normal image, model could only predict 70% accurately.

		Predicted Label	pneumonia	normal
True Label				
pneumonia			50%	50%
normal			30%	70%

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?	Precision and recall value have decreased to 60%. The highest precision and recall value is from the first model which is trained by clean and balanced data.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Because of the 30 mislabeled images in each label, the accuracy of the model has significantly decreased. Dirty data obstructs the model to recognize each label's pattern and can lead to huge problem.

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	For the bacterial pneumonia image, model predicted 91% accurately and wrongly predicted 9% of images into normal images. For the viral pneumonia image, model predicted 80% accurately and wrongly predicted 20% of images into bacterial pneumonia. For the normal images, model predicted 100% accurately. Viral pneumonia was the most confusing class for model to

"confusion"? Include a screenshot of the new confusion matrix.	predict. Normal images were easiest for model to predict. Maybe adding more data of viral pneumonia would help model to understand about this class more accurately and eventually make more accurate prediction about this class. **Predicted Label** | True Label	bacterial pneumonia	viral pneumonia	normal
---	---	---	---
bacterial pneumonia	91%	-	9%
viral pneumonia	20%	80%	-
normal	-	-	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)?	Model's precision is 90% and recall value is 84.38%. In case of precision, it is calculated by dividing sum of the ratio of accurate prediction of total prediction with number of classes. Sum of the ratio is 91/111 + 80/80 + 100/109 = 2.74. Number of classes is 3. So, precision value is 2.74/3 = 0.9124= 91.24%. In case of recall, it is calculated by dividing sum of the ratio of accurate prediction of actual occurrence with number of classes. Sum of the ratio is 91/100 + 80/100 + 100/100 = 2.71. Number of classes is 3. So, recall value is 2.71/3 = 90%.
F1 Score What is this model's F1 score?	0.906 = 90.6%