AutoML Modeling Report

Osama Alsubaie

Table of Contents:

Definition:

Training Dataset	The sample of data used to train the machine learning model by feeding it input data and the corresponding target output (labels).
Validation Dataset	The validation dataset is used during the training process to guides the model's learning process, also it's help in optimize the model's parameters and prevent overfitting.
Testing Dataset	A separate subset of the data that the model has never seen during training. This data helps you estimate how well your model will perform in the real world and whether it's overfitting or underfitting.
Model threshold	The model threshold is a value that determines the decision boundary for classifying instances in a binary classification problem. If the model's output is greater than or equal to the threshold, the instance is classified as the **positive class.** If the model's output is less than the threshold, the instance is classified as the **negative class.**
Confusion Matrix	Grid which shows all the predicted labels relative to all true labels. Four cells are present in the confusion matrix: 1. True Positive (TP) 2. True Negative (TN) 3. False Positive (FP) 4. False Negative (FN)
Model precision	Measures the percentage of correct predictions against total number of predictions.
Model recall	Measures the percentage of correctly identified instances against total possible instances.
F1 Score	Metric that combines both precision and recall into a single value, offering a balanced view of the model performance.

Binary Classifier with Clean/Balanced Data

Train/Test Split
How much data was used for training? How much data was used for testing?

Total Images in the data set is 198 [99 normal, 99 pneumonias]
80% of the images were set **for training**, and **10% for testing**,**10% for validation**.

Number of **Training** Images = 158
Number of **Validation** Images = 20
Number of **Testing** Images = 20

Confusion Matrix
What do each of the sections 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?

Four cells are present in the confusion matrix [TP, TN, FP, FN]:

True Positive (TP)	Actual pneumonia images correctly classified as pneumonia.
True Negative (TN)	Actual normal images correctly classified as normal.
False Positive (FP)	Actual normal images classified as pneumonia.
False Negative (FN)	Actual pneumonia images classified as a normal.

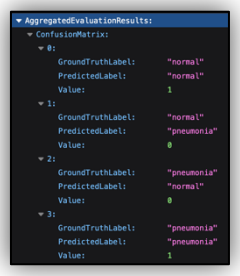
The confusion matrix:

As percentage:

	Pneumonia	Normal
Pneumonia P	100%	0%
Normal N	0%	100%

As number of images:

	Pneumonia	Normal
Pneumonia P	20	0
Normal N	0	20



As can be seen from the **image & the confusion matrix** above:
For ***Normal*** images:
100% of the images (20 in total) were predicted as ***normal (TN).***

For ***Pneumonia*** images:
100% of the images (20 in total) were predicted as ***Pneumonia (TP).***

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 what portion of positive identifications are actually correct. A high precision model produces fewer false positives.
On the other hand, recall tells us what portion of actual positives was identified correctly.
A high recall model produces fewer false negatives.

The model achieved 100% Precision and 100% recall.

Normal class:

Precision	TP/(TP+FP) = 20/(20+0) = 1
Recall	TP/(TP+FN) = 20/(20+0) = 1
F1 score	(2*Precision*Recall)/ (Precision + Recall) = 2/2 = 1
Threshold	0.52 for normal, 0.66 for Pneumonia

Binary Classifier with Clean/Unbalanced Data

Train/Test Split
How much data was used for training?
How much data was used for testing?

Total Images in the data set is 298 [99 normal, 199 pneumonias]
80% of the images were set **for training**, and **10% for testing**,**10% for validation**.

Number of **Training** Images = 238
Number of **Validation** Images = 20
Number of **Testing** Images = 40

Confusion Matrix
How has the confusion matrix been affected by the unbalanced data?
Include a screenshot of the new confusion matrix summary

Four cells are present in the confusion matrix [TP, TN, FP, FN]:

True Positive (TP)	Actual pneumonia images correctly classified as pneumonia.
True Negative (TN)	Actual normal images correctly classified as normal.
False Positive (FP)	Actual normal images classified as pneumonia.
False Negative (FN)	Actual pneumonia images classified as a normal.

The confusion matrix:

As percentage:

	Pneumonia	Normal
Pneumonia P	0%	100%
Normal N	15%	85%

As number of images:

	Pneumonia	Normal
Pneumonia P	40	0
Normal N	3	17

As can be seen from the **image & the confusion matrix** above:
For ***Normal*** images:
85% of the images (17 in total) were predicted as ***Normal (TN)***. 15% of the images (3 in total) were predicted as ***Pneumonia (FP)***.

For ***Pneumonia*** images:
100% of the images (40 in total) were predicted as ***Pneumonia (TP)***.

Precision and Recall
How have the model's precision and recall been affected by the unbalanced data?

Precision tells us what portion of positive identifications are actually correct. A high precision model produces fewer false positives.
On the other hand, recall tells us what portion of actual positives was identified correctly.
A high recall model produces fewer false negatives.

The model achieved 93% Precision and 100% recall.

Evaluation Metrix:

Accuracy	(TP+TN)/(TP+TN+FP+FN) = 57/60 = 0.95
Precision	TP/(TP+FP) = 40/(40+3) = 0.93
Recall	TP/(TP+FN) = 40/(40+0) = 1
F1 score	(2*Precision*Recall)/ (Precision + Recall) = 1.86/1.93 = 0.96

Unbalanced Classes
From what you have observed, how do unbalanced classed affect a machine learning model?

The model become **bias toward the unbalanced class** (Pneumonia for our case).
Also, **poor generalization happened to the minority class** (normal for our case).

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 information.

Four cells are present in the confusion matrix [TP, TN, FP, FN]:

True Positive (TP)	Actual pneumonia images correctly classified as pneumonia.
True Negative (TN)	Actual normal images correctly classified as normal.
False Positive (FP)	Actual normal images classified as pneumonia.
False Negative (FN)	Actual pneumonia images classified as a normal.

The confusion matrix:

As percentage:

	Pneumonia	Normal
Pneumonia P	90%	10%
Normal N	60%	40%

As number of images:

	Pneumonia	Normal
Pneumonia P	18	2
Normal N	12	8

As can be seen from the **image & the confusion matrix** above:
For ***Normal*** images:
40% of the images (8 in total) were predicted as ***Normal (TN)***. 60% of the images (12 in total) were predicted as ***Pneumonia (FP)***.

For ***Pneumonia*** images:
90% of the images (18 in total) were predicted as ***Pneumonia (TP)***. 10% of the images (2 in total) were predicted as ***normal (FN).***

The confusion matrix is worse than the clean/balanced and the clean/unbalanced data cases, and that's because of the dirty data the training algorithm cannot learn the patterns in the data well enough.

Precision and Recall
How have the model's precision and recall been affected by the dirty data. Of the binary classifiers, which has the highest precision? Which has the highest recall?

Precision tells us what portion of positive identifications are actually correct. A high precision model produces fewer false positives.
On the other hand, recall tells us what portion of actual positives was identified correctly.
A high recall model produces fewer false negatives.

The model achieved 60% Precision and 90% recall.

Evaluation Metrix:

Accuracy	(TP+TN)/(TP+TN+FP+FN) = 26/40 = 0.65
Precision	TP/(TP+FP) = 18/(18+12) = 0.6
Recall	TP/(TP+FN) = 18/(18+2) = 0.9
F1 score	(2*Precision*Recall)/ (Precision + Recall) = 1.08/1.5 = 0.72

The highest precision and recall went for the **Clean/Balanced Data**, while the worst went for the **unclean/balanced data.**

Dirty Data
From what you have observed, how does dirty data affect a machine learning model?

The dirty data increases the errors for both classes and make the model struggle to find patterns among classes, the accuracy for the model went down as a result. Also, the results are unreliable because of the dirty data in the test and evaluation sets.

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 information.

The confusion matrix:

As percentage:

	normal	bacterial_p[1]	viral_p[1]
normal	100%	0%	0%
bacterial_p[1]	0%	100	0%
viral_p[1]	15%	15%	70%

1 = the letter p is stand for pneumonia

As number of images:

	normal	bacterial_p[1]	viral_p[1]
normal	20	0	0
bacterial_p[1]	0	20	0
viral_p[1]	3	3	14

1 = the letter p is stand for pneumonia

The ***viral_pneumonia*** class is most likely to be **the confused class,** the ***normal*** and ***bacterial_pneumonia are*** most likely to be get right.
We can add more data for all classes to try remedy the model confusion [equal number of images should be added for each class in order to keep the dataset balanced], this might help in resolving the model confusion.

Precision and Recall
What are the model's precision and recall? How are these values calculated?

Precision tells us what portion of positive identifications are actually correct. A high precision model produces fewer false positives.
On the other hand, recall tells us what portion of actual positives was identified correctly.
A high recall model produces fewer false negatives.

Evaluation Metrix:

For the ***normal*** class:

Precision	TP/(TP+FP) = 20 / (20 + 0) = 1
Recall	TP/(TP+FN) = 20 / (20 + 0) = 1
F1 score	(2*Precision*Recall)/ (Precision + Recall) = 2/2 = 1

For the ***bacterial_pneumonia*** class:

Precision	TP/(TP+FP) = 20 / (20 + 0) = 1
Recall	TP/(TP+FN) = 20 / (20 + 0) = 1
F1 score	(2*Precision*Recall)/ (Precision + Recall) = 2/2 = 1

For the ***viral_pneumonia*** class:

Precision	TP/(TP+FP) = 14 / (14 + 6) = 0.70
Recall	TP/(TP+FN) = 14 / (14 + 6) = 0.70
F1 score	(2*Precision*Recall)/ (Precision + Recall) = 0.98/1.40 = 0.70

We take the Macro-Averaging or the Micro-Averaging of all the classes evaluation matrix in order to get the Precision and Recall for the model.

We will take the Macro-Averaging which is the average of the sum of [Precision and Recall and F1] for all the model:
Precision= (1+1+0.7)/3 = 0.9
Recall = (1+1+0.7)/3 = 0.9

F1 Score
What is this model's F1 score?

Macro-Averaged F1-score = (1.00 + 1.00 + 0.70) / 3 = 0.90