AutoML Modeling Report

<Ying Shan>

Binary Classifier with Clean/Balanced Data

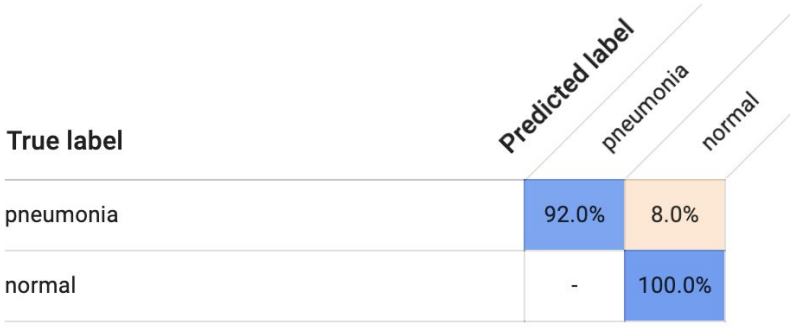
Train/Test Split

How much data was used for training? How much data was used for testing?

There are a total of 398 data, 199 for normal and 199 for pneumonia. 44 data was used for testing and 354 data used for training

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?

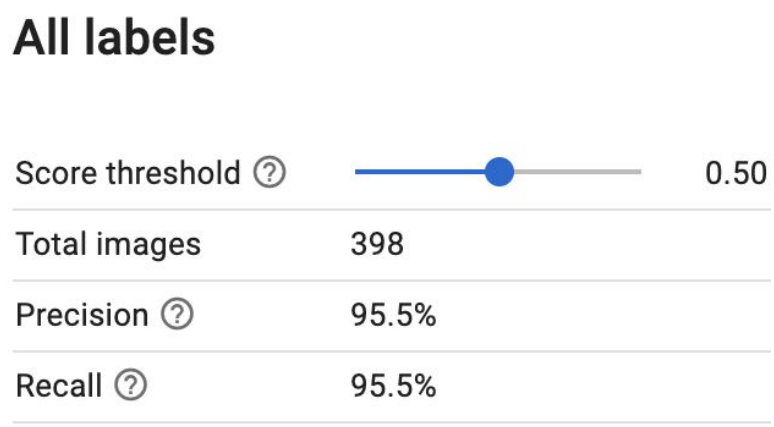


Confusion matrix Illustrate all the predicted labels relative to all the true labels. Each cell in the matrix represents percentage of prediction.

The confusion matrix shows that 92% of true pneumonia is classified as pneumonia and 8% of true pneumonia is classified as normal. Also, 100% of true normal is classified as normal. The true positive rate for pneumonia is 92%; the false positive rate for normal class is 0%

Precision & 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)?



The precision measure the percentage of correct predictions against total number of predictions; The recall measure the percentage of correctly identified instances against the total number of instances.

As score threshold of 0.5, Precision is 95.5%; Recall is 95.5%

.

Score Threshold

When you increase the score threshold, what happens to precision? What happens to recall? Why?

As I increase the score threshold, the precision increases, and recall decreases. Since the score threshold refers to the level of confidence the model must have to assign a category to a test item. Higher score threshold will lower the
In order to ensure a high threshold, the model retains data that identifies relatively high correct rates and removes data that identifies relatively low accuracy. This leads to an increasing rate or correct predictions, but the less data is identified.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split

How much data was used for training? How much data was used for testing?

There are a total of 400 data, 100 for normal and 300 for pneumonia. 43 data was used for testing and 357 data used for training

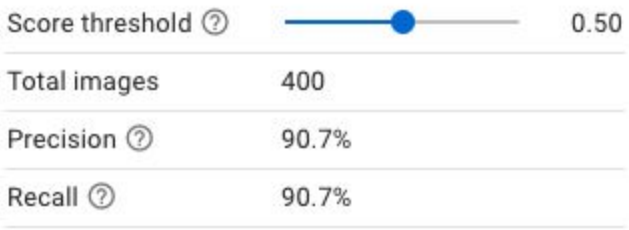
Confusion Matrix

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

As more pneumonia data added into the training data set, the correct prediction rate of pneumonia is increased, which leads to a higher true positive rate for pneumonia.

Precision & 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.)



After added 200 pneumonia data to the training model, while kept 100 normal data. Both precision and recall of the model has increased from 83.3% to 90.7%

Unbalanced Classes

From what you've observed, how do unbalanced classes affect a machine learning model?

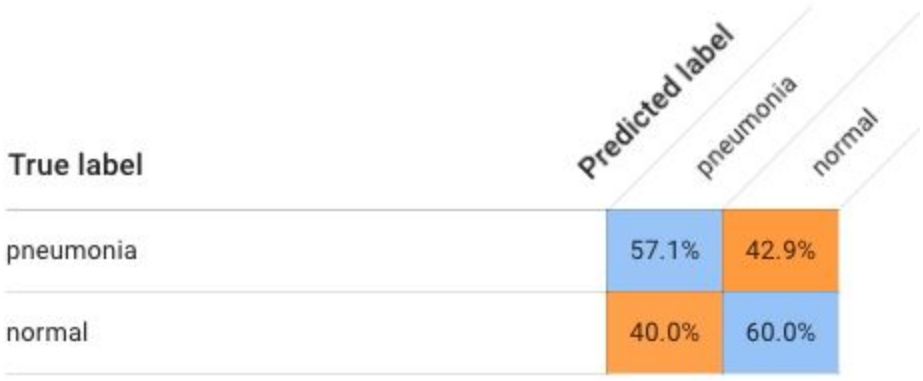
Add more pneumonia data will increase the true positive rate for pneumonia. Based on my training result, that 100% true normal is already classified as normal. Unbalanced data (more pneumonia data than normal data) will increase both precision and recall of a machine learning model. The label with more data will tend to have higher precision as well as recall.

In addition, I manually predict 10 images for both normal and pneumonia (not used for training) on the model that is trained by unbalanced data
10 normal images are all correctly be predicted. Two images shows relatively low recognition probability (0.675 and 0.854). Others are higher than 0.9 (three >= 0.98)
10 pneumonia images are all correctly be predicted as well. All images shows fairly high recognition probability (Nine images has probability of 1, and 1 image has probability of 0.997.)

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.

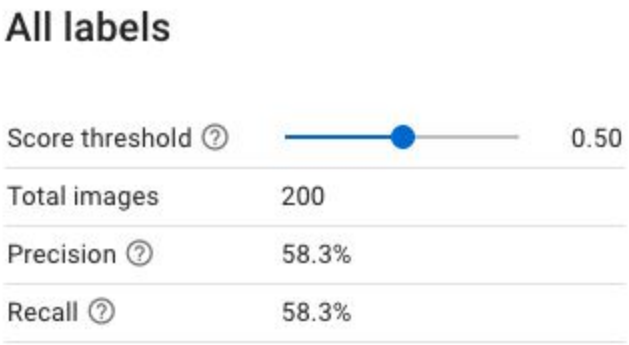


The confusion matrix shows significant difference between dirty data and clean data. Prediction rate for both pneumonia and normal are significantly reduced. Each cell in the matrix represents percentage of prediction.

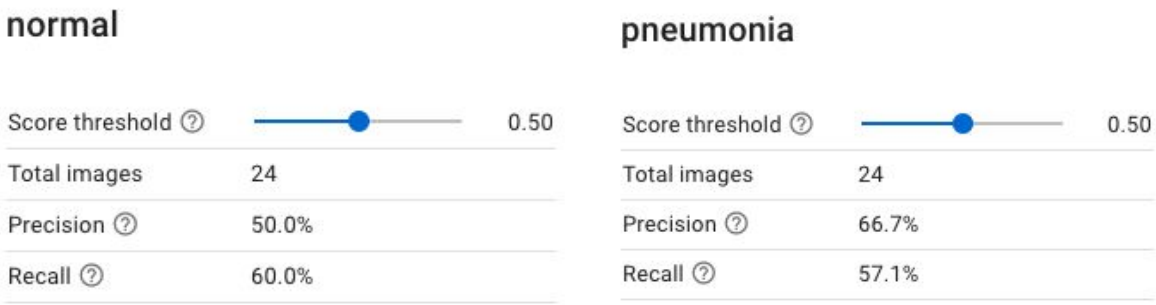
This time, the confusion matrix shows that only 57.1% of true pneumonia is classified as pneumonia and 42.9% of true pneumonia is classified as normal. In addition, 40% of true normal is classified as pneumonia.

Precision & 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?



Dirty data decrease both precision and recall of the model. At score threshold of 0.5, the model only shows 58.3% of precision, and 58.3% of recall.



Of the binary classifiers, pneumonia has the highest precision, and normal has highest recall.

Dirty Data

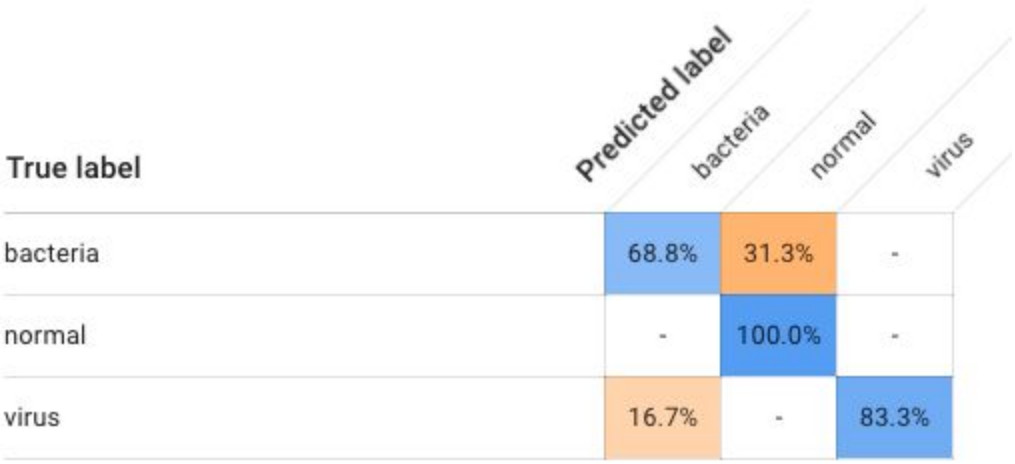
From what you've observed, how do dirty data affect a machine learning model?

Dirty data will decrease both precision and recall of the model. Dirty data may leads to a bad performance of an image classification modal.

3-Class Model

Confusion Matrix

Summarize the 3-class confusion matrix. What classes are the model most likely to confuse? What class(es) is the model most likely to get right? What might you do to try to remedy the model's "confusion"? Include a screenshot of the new confusion matrix.



The confusion matrix shows that 68.8% of true bacteria pneumonia is classified as bacteria pneumonia, and 31.3% of true bacteria pneumonia is classified as normal. Also, 100% of true normal is classified as normal. In addition, 83.3% of true virus pneumonia is classified as virus pneumonia, and 16.7% of true virus pneumonia is classified as bacteria pneumonia.

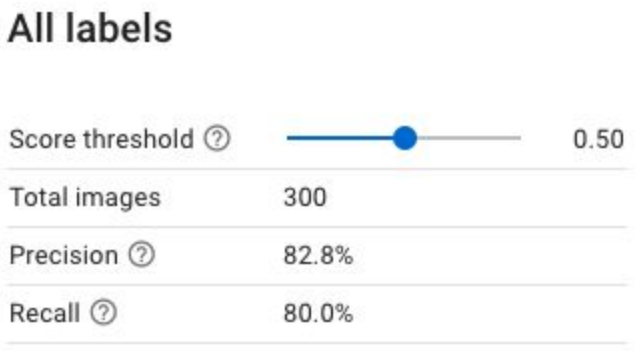
Bacteria and virus are the model most likely to confuse.

Normal is the model most likely to get right.

In order to remedy the model's confusion, I would like to add 200 more data to both bacteria and virus.

Precision & Recall

What are the model's precision and recall? How are these values calculated? (Report the values for a score threshold of 0.5.)



At score threshold of 0.5, the model shows 82.8% of precision, and 80.0% of recall.

Precision is equals the number of correct predictions divided by the total predictions, which illustrate the accuracy of prediction task. [true positive / (true positive + false positive)]

Recall is equals the number of correct predictions divided by the total number of labeled instances, which illustrate the percentage of data that the model identified (ground truth). [true positive / (true positive + false negative)]

F1 Score

What is this model's F1 score?

F1 = 2 * precision * recall / (precision + recall)
F1 = 2 * 0.828 * 0.8 / (0.828 + 0.8)
F1 = 0.814
This model's F1 score is 0.814