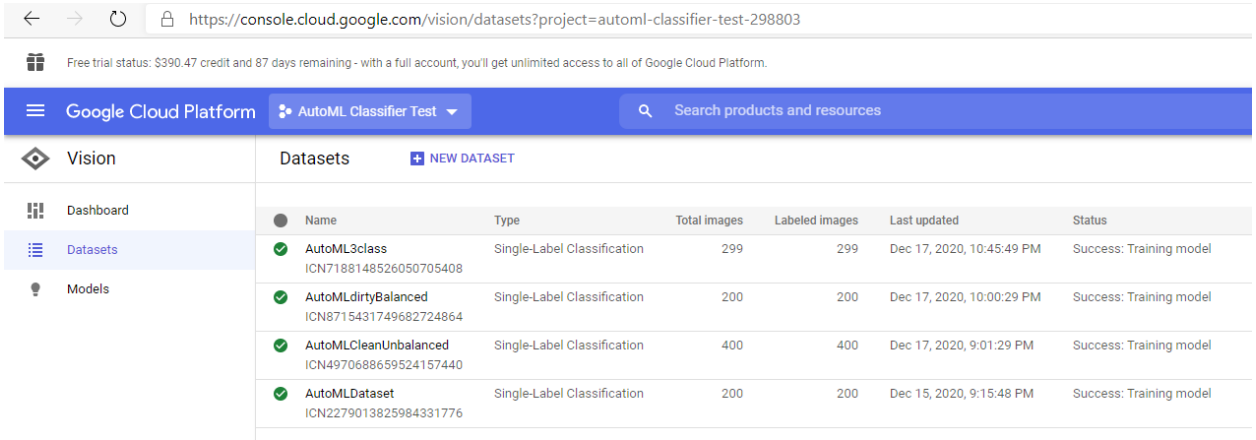
AutoML Modeling Report

Chromilo Amin | April 26, 2021 | Bertelsmann Technology Scholarship program

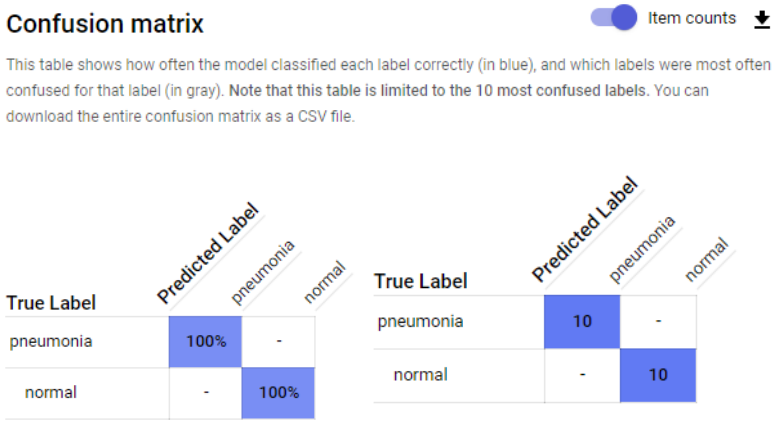
---

Contents

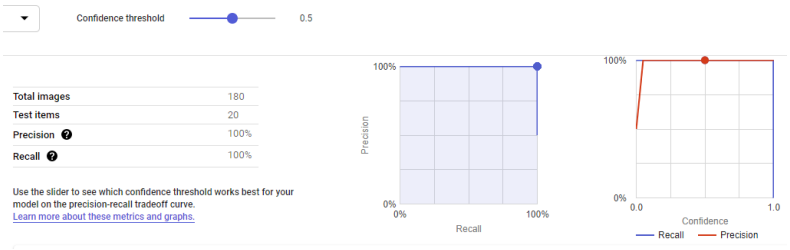
DataSets Used for the AutoML Classifier Test



Binary Classifier with Clean/Balanced Data

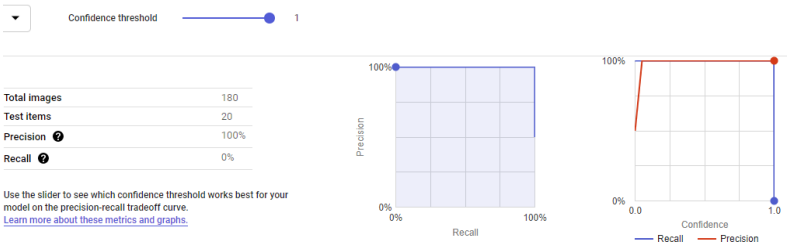
Train/Test Split How much data was used for training? How much data was used for testing?	100 images for normal and 100 images for pneumonia images were used in total. After training was run, 180 images are allocated to training and 20 images for testing.
Confusion Matrix What do each of the cells in 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?	 TP(pneumonia) = TP/Total Predicted = 10/10 = 1 FP(normal)=FP/Total Predicted = 0/10 = 0 The true positive rate for the pneumonia class is 100% and the false positive rate for the normal class 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 how likely the prediction is correct, i.e. TP/(TP+FP). A high precision model produces fewer false positives. At 0.5 score threshold, the model produced a precision value of 100% Recall measures how good the model is at identifying actual occurrences in the data, i.e. TP/(TP+FN). A high recall model produces fewer false negatives. At 0.5 score threshold, the model produced a recall value of 100%

		Predicated Labels		
		Pneumonia	Normal	
True Labels	Pneumonia	TP	FN	Total Pneumonia Set
	Normal	FP	TN	Total Normal Set
		Total Pneumonia Predicted	Total Normal Predicted	



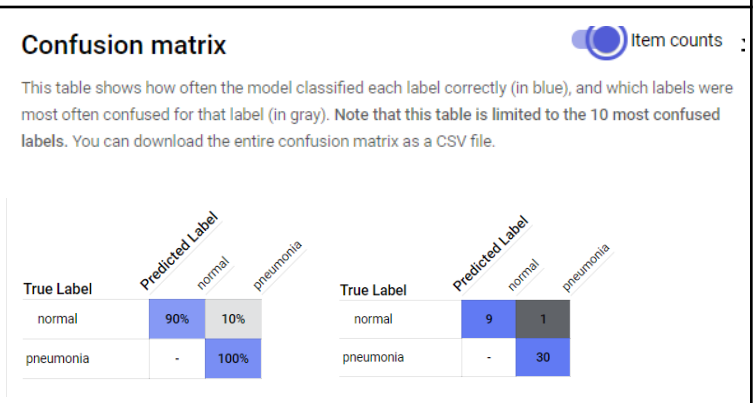
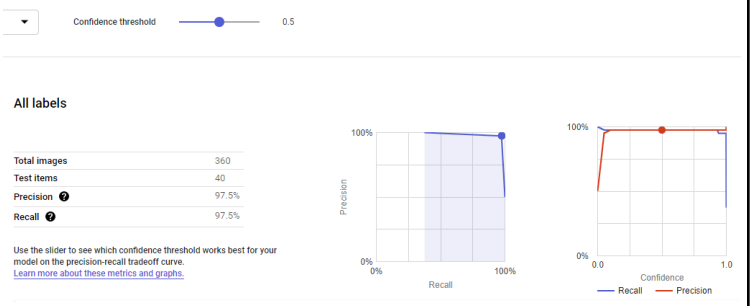
Score Threshold

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

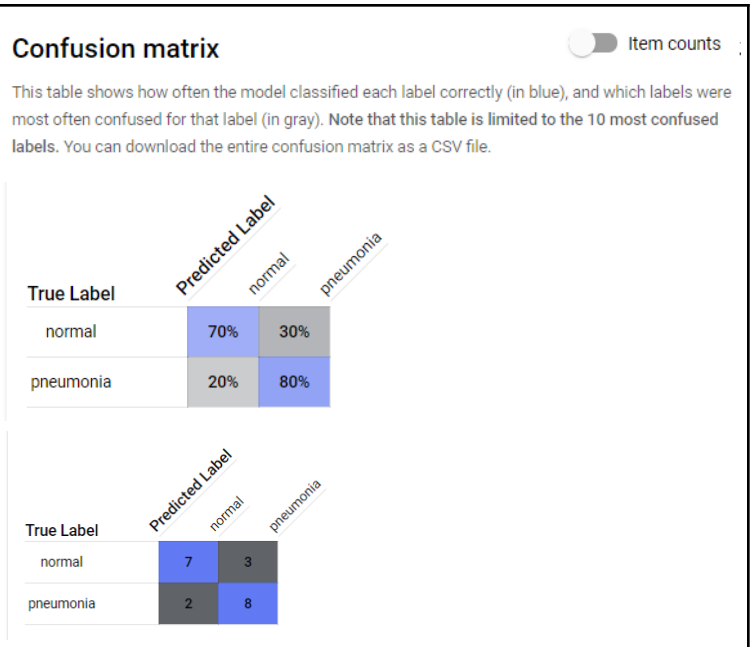


When I change the score threshold from 0.5 to 1, the Recall drops to 0% and Precision remains at 100%. This high confidence level means there is 0% chance of getting a false negative.

Binary Classifier with Clean/Unbalanced Data

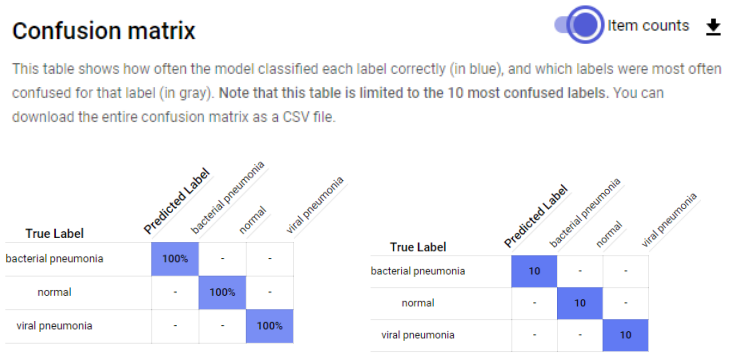
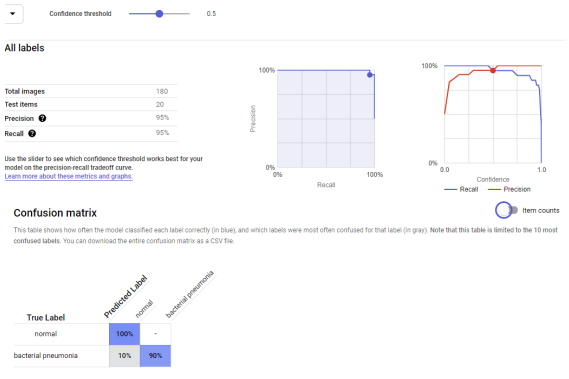
Train/Test Split How much data was used for training? How much data was used for testing?	100 images for normal and 300 images for pneumonia images were used in total. After training was run, 360 images are allocated to training and 40 images for testing.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 With unbalanced data favoring the pneumonia class, we see the matrix switch to classifying pneumonia in the “negative” column having either FN and TN values. There is a 10% false negative for the normal class and 90% true positive predictions. However, because there are a lot more images for the pneumonia class, the true negative detections is 100%, meaning there is 100% certainty that an image is not normal and that it is from a pneumonia patient.
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)?	 The model’s precision and recall given 0.5 score threshold is 97.5% for both.
Unbalanced Classes From what you have observed, how do unbalanced classes affect a machine learning model?	Having more data to train a specific class will skew the confidence level to that class.

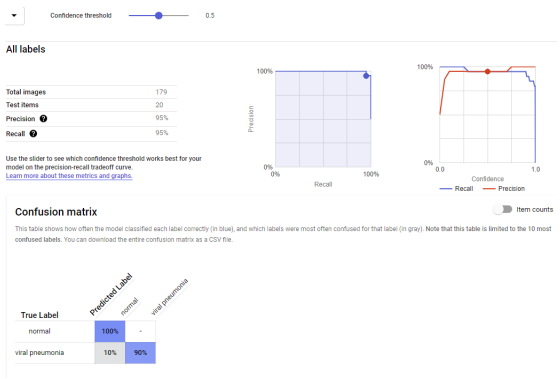
Binary Classifier with Dirty/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	70 images for normal are classified in the correct folder for training but the other 30 normal images were mixed in with the pneumonia folders. 70 images for pneumonia were also classified correctly in the folder but the other 30 pneumonia images were mixed in with normal images in the normal folder. Total images used is 100 each. After training was run, 180 images were allocated to training and 20 images for testing.
Confusion Matrix How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.	 For an equal split of datasets, there is a 70% chance of true positives for the normal class and 30% chance of a false negative, i.e. 30% chance of getting a prediction wrong to indicate an image belongs to the pneumonia class instead of being classified as normal. On the other hand, it seems there is a bigger percentage at 80% for the pneumonia class to get a true negative prediction that an image is classified correctly as not normal, i.e. coming from a pneumonia patient. Only 20% chance that it is a false positive.
Precision and Recall How have the model's precision and recall been affected by the	

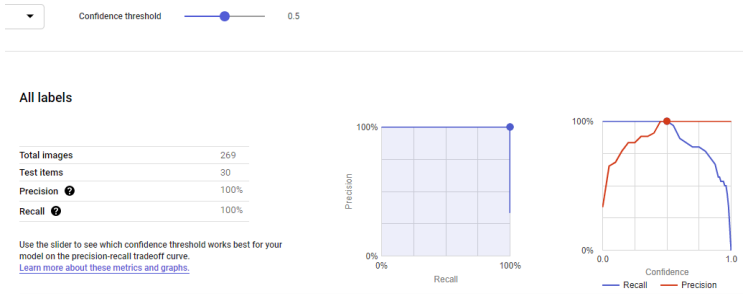
dirty data (report the values for a score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?	 At 0.5 score threshold, both precision and recall intersect to provide 75% value. However, increasing the score threshold increases the precision percentage and drops recall to 0%, meaning the false negatives are eliminated. . Of the binary classifiers, the highest precision is Clean/Balanced Data with 100% and also has the highest recall at 100%.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data introduces more FP and FN by about 20-30%. Dirty data has caused these multiple misclassifications.

3-Class Model

Train/Test Split How much data was used for training? How much data was used for testing?	300 total images were used for this model. 100 images allotted to the normal folder, 100 images with “virus” in the filename allotted to the “viral pneumonia” folder, and the last 100 images with “bacteria” in the filename allotted to the “bacterial pneumonia” folder.
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.	 The model will most likely get confused with the two pneumonia classes - Bacterial Pneumonia and Viral Pneumonia. The normal class is the most likely the model will get correct. To improve the confusion model, I would only keep one of the pneumonia classes with the normal class. The way this is setup with 3-classes, I don’t know what to classify as positives and negatives. Another way is to add more true “bacterial” images and have more training cycles. Here is the new confusion matrix: 



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)?



P_{Normal} = TP/(TP+FP) = 1/1+0.10 = 1/1.1 = 0.90
$P_{Bacterial}$ = TP/(TP+FP) = 1/1+0.10 = 1/1.1 = 0.90
P_{Viral} = TP/(TP+FP) = 1/1+0.10 = 1/1.1 = 0.90

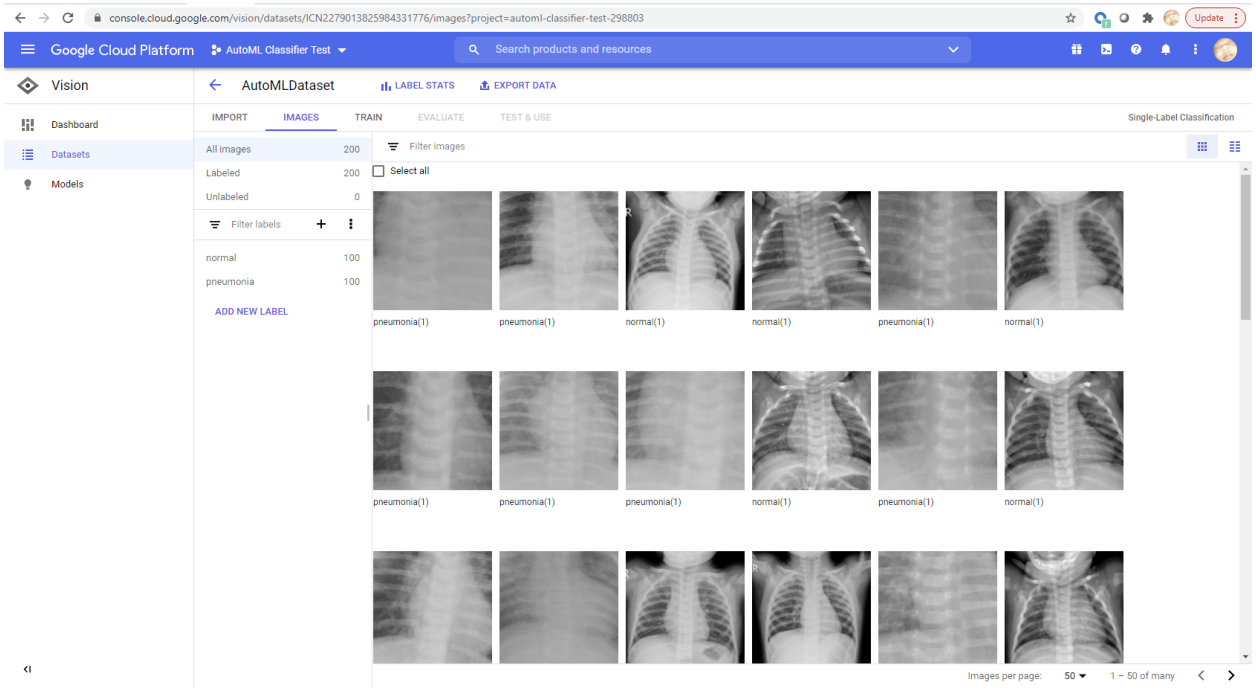
R_{Normal} = TP/(TP+FN) = 1/1+0.10 = 1/1.1 = 0.90
$R_{Bacterial}$ = TP/(TP+FN) = 1/1+0.10 = 1/1.1 = 0.90
R_{Viral} = TP/(TP+FN) = 1/1+0.10 = 1/1.1 = 0.90

F1 Score
What is this model's F1 score?

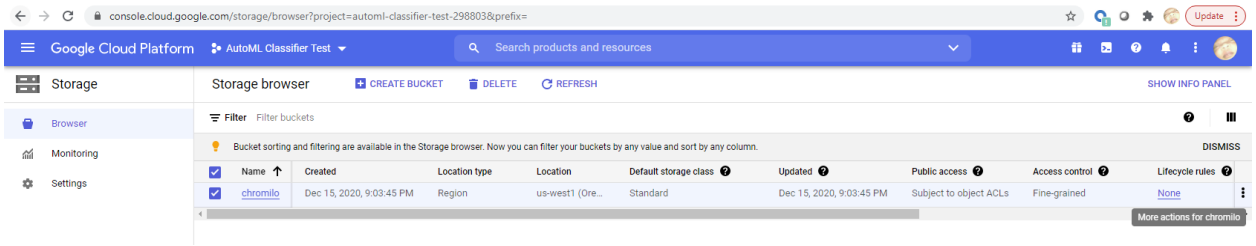
$P_{model} = (P_{Normal} + P_{Bacterial} + P_{Viral}) / 3 = 0.9+0.9+0.9/3=0.9$
$R_{model} = (R_{Normal} + R_{Bacterial} + R_{Viral}) / 3 = 0.9+0.9+0.9/3=0.9$

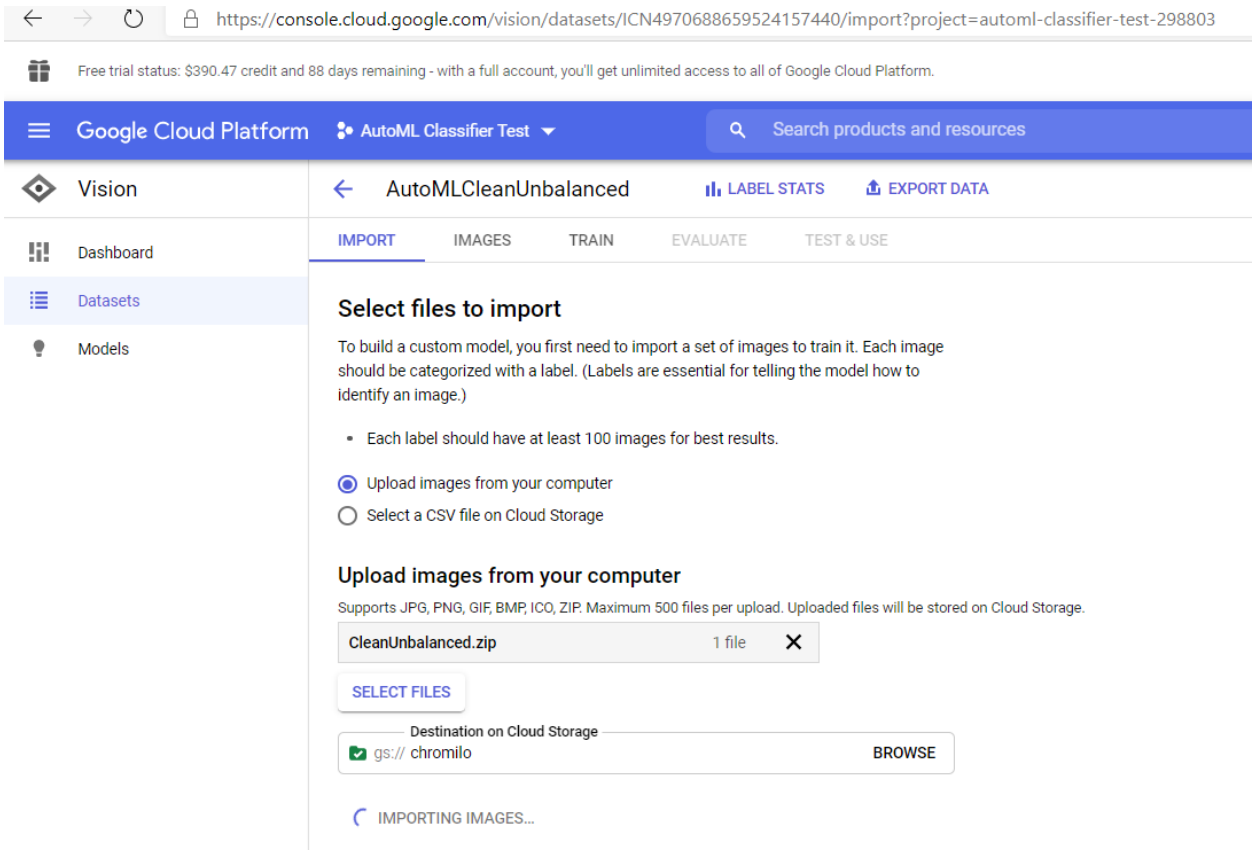
$F1 = (2 * (P_{model} * R_{model})) / (P_{model} + R_{model})$
$= (2*(0.9*0.9)/(0.9+0.9) = 2*0.81/1.8 = 0.9$

Appendix A – GCP AutoML Vision Classifier Test dataset

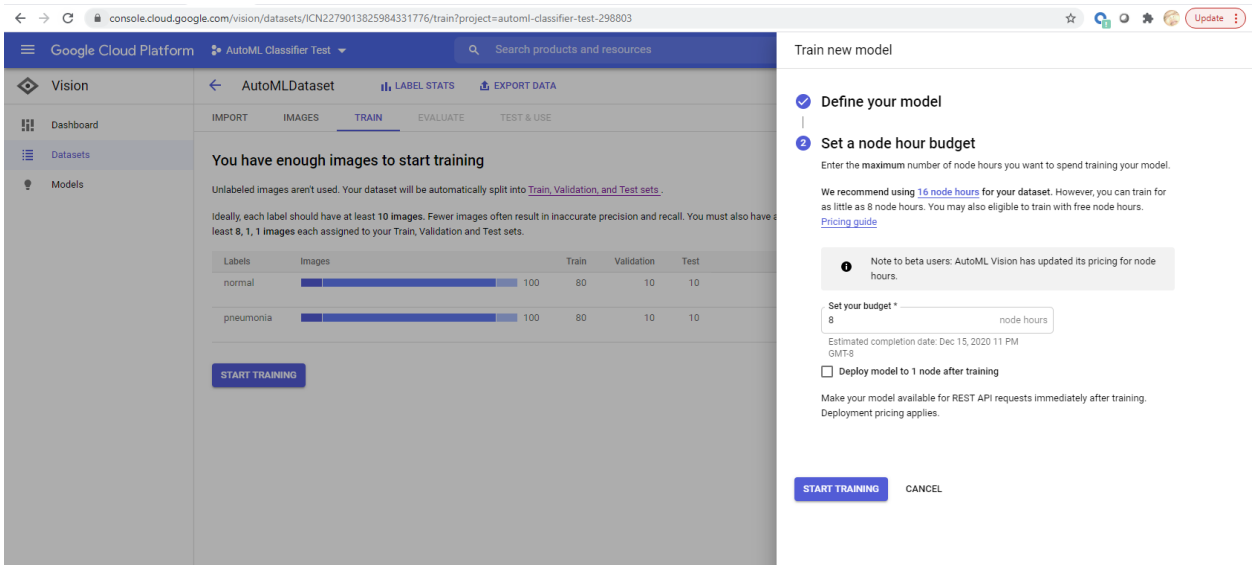


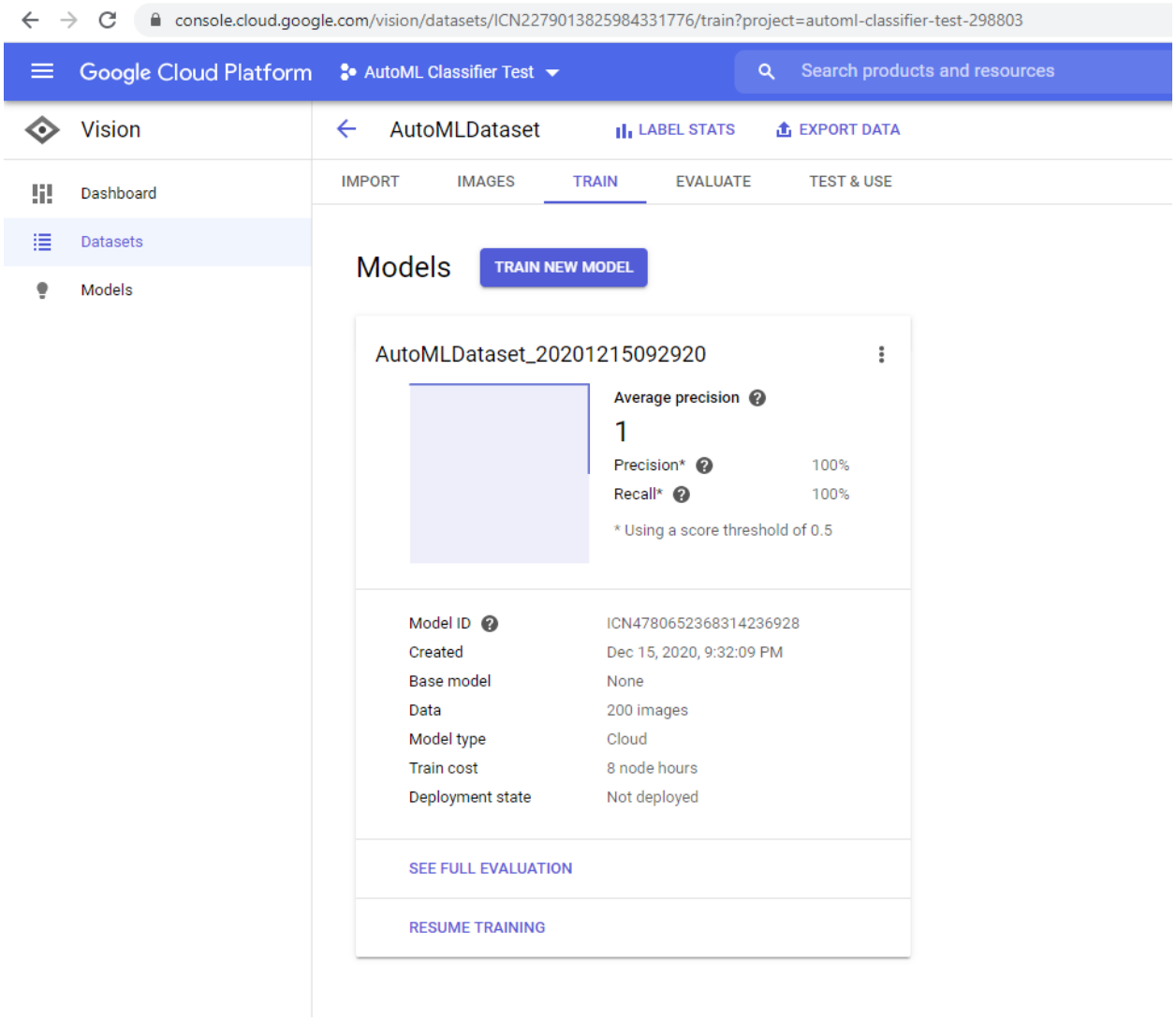
Appendix B – GCP Cloud Storage



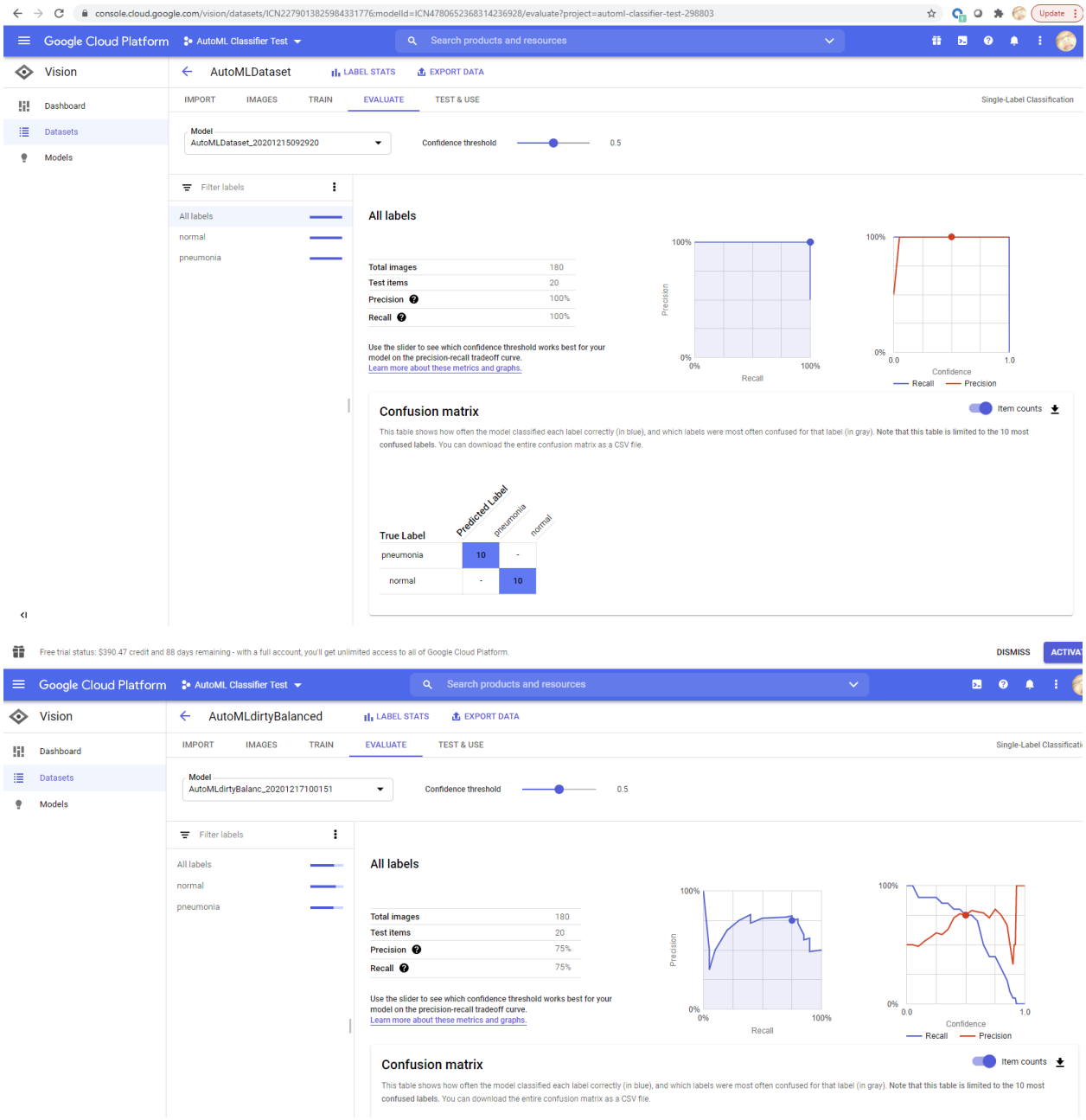


Appendix C – GCP AutoML Vision | Training

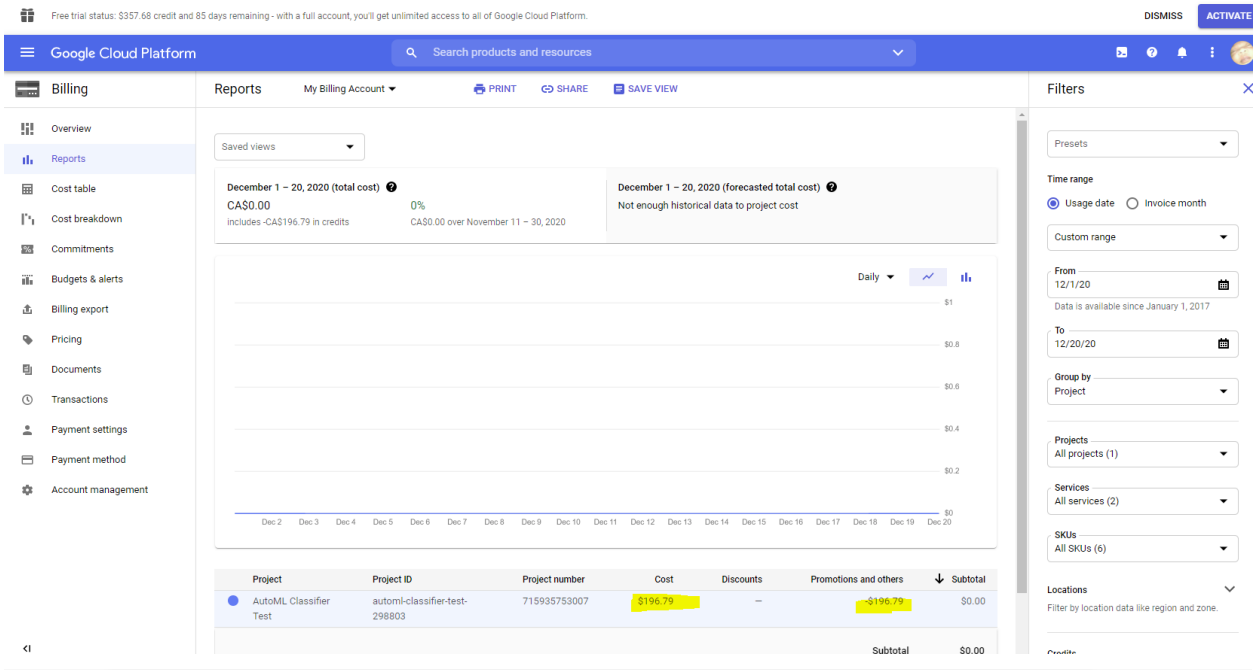


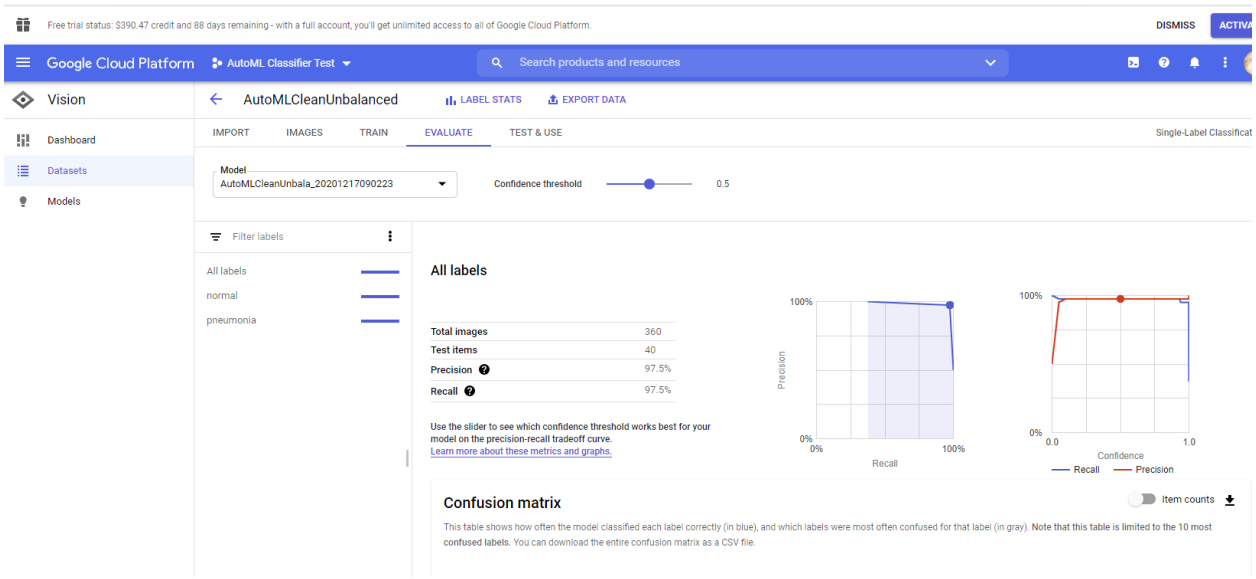


Appendix D – GCP AutoML Vision | Evaluation



Appendix E – GCP AutoML Vision | Costs


Free trial status: $357.68 credit and 85 days remaining - with a full account, you'll get unlimited access to all of Google Cloud Platform.
DISMISS
ACTIVATE
Google Cloud Platform
Search products and resources
Billing
Reports
My Billing Account
PRINT
SHARE
SAVE VIEW
Overview
Reports
Cost table
Cost breakdown
Commitments
Budgets & alerts
Billing export
Pricing
Documents
Transactions
Payment settings
Payment method
Account management
Saved views
December 1 – 20, 2020 (total cost)
CA$0.00
includes -CA$196.79 in credits
0%
CA$0.00 over November 11 – 30, 2020
December 1 – 20, 2020 (forecasted total cost)
Not enough historical data to project cost
Daily
$1
$0.8
$0.6
$0.4
$0.2
$0
Dec 2
Dec 3
Dec 4
Dec 5
Dec 6
Dec 7
Dec 8
Dec 9
Dec 10
Dec 11
Dec 12
Dec 13
Dec 14
Dec 15
Dec 16
Dec 17
Dec 18
Dec 19
Dec 20
Project
Project ID
Project number
Cost
Discounts
Promotions and others
Subtotal
AutoML Classifier Test
automl-classifier-test-298803
715935753007
$196.79
–
-$196.79
$0.00
Subtotal
$0.00
Filters
Presets
Time range
Usage date
Invoice month
Custom range
From
12/1/20
Data is available since January 1, 2017
To
12/20/20
Group by
Project
Projects
All projects (1)
Services
All services (2)
SKUs
All SKUs (6)
Locations
Filter by location data like region and zone.


Free trial status: $390.47 credit and 88 days remaining - with a full account, you'll get unlimited access to all of Google Cloud Platform.
DISMISS
Google Cloud Platform
AutoML Classifier Test
Search products and resources
Vision
Dashboard
Datasets
Models
AutoMLCleanUnbalanced
LABEL STATS
EXPORT DATA
IMPORT
IMAGES
TRAIN
EVALUATE
TEST & USE
Single-Label Classificat
Model
AutoMLCleanUnbala_20201217090223
Confidence threshold
0.5
Filter labels
All labels
normal
pneumonia
All labels
Total images 360
Test items 40
Precision 97.5%
Recall 97.5%
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.
100%
0%
Precision
Recall
Confidence
1.0
Recall
Precision
Confusion matrix
Item counts
This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray). Note that this table is limited to the 10 most confused labels. You can download the entire confusion matrix as a CSV file.