| AutoML Modeling Report |  |
| --- | --- |

*Mohammed Tayor*

Binary Classifier with Clean/Balanced Data

| **Train/Test Split**  How much data was used for training? How much data was used for testing? | Default settings of Vertex AI training are used, which correspond to a random split of 80% training , 10% validation, and 10% testing, i.e. images are 160 in train , 20 in validation, and 20 in test. |
| --- | --- |
| **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? | Source of definitions as supplied by Udacity course material: [link](https://github.com/fa-ahmad/AIPMND/blob/master/ND088-MLPMND-Model%20Evaluation-Cheat%20Sheet%20.pdf).  TP: True positive - Correctly predicted positive  FN: False negative - Incorrectly predicted negative  FP: False positive - Incorrectly predicted positive  TN: True negative - Correctly predicted negative  In our first model the confusion matrix looks like this:    What is the true positive rate for the “pneumonia” class?  100%  What is the false positive rate for the “normal” class?  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 as a real number between 0 and 1 measures how good are positive predictions of a model, and calculated by the ratio between true positives to predicted positives (sum of true positives and false positives) * Recall measures how the model can accurately identify the relevant data. It is calculated by the ratio of true positives to actual positives(sum of true positive and false negative)   For our first model:    Precision: 100% or 1  Recall: 100% or 1 |
| **Score Threshold**  When you increase the threshold what happens to precision? What happens to recall? Why? | If threshold is increased then the precision also increases while recall decreases, at the limit of a 1 threshold the precision tends to go to 1 and recall falls to 0.    This is due to the fact that we increase the criteria for model prediction so we have more “strict” rules (higher threshold) for model predictions, hence less false positives, more precise predictions, and lower recall. |

Binary Classifier with Clean/Unbalanced Data

| **Train/Test Split**  How much data was used for training? How much data was used for testing? |  |
| --- | --- |
| **Confusion Matrix**  How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix. | With all models max train node hours set to 10 and early stopping activated in Vertex AI we get a model more prone to predict pneumonia instead of actual normal since the imbalance is geared towards pneumonia. This should increase the false negatives, however in this training result we can see that no normal case has been incorrectly classified. A hypothesis to explain this would be the relatively small sample size of both train and test. However the correct pneumonia predictions decrease. |
| **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)? | Both precision and recall decrease, which is logical due to imbalance in data distribution. |
| **Unbalanced Classes**  From what you have observed, how do unbalanced classed affect a machine learning model? | An imbalance dataset would affect the bias of the model towards the higher sample class. |

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 predictions are more incorrect, meaning more false positives and false negatives, decreasing the precision and accuracy of the model. |
| --- | --- |
| **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? | Both precision and recall decreased in comparison to the first model of balanced clean data.  The first model is the one with the highest precision and recall, which confirms the role of clean balanced data more than quantity of samples. |
| **Dirty Data**  From what you have observed, how does dirty data affect a machine learning model? | Dirty data decreases the accuracy , precision , and recall of a machine learning model. It has a bad impact on the quality of its predictions. |

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. | Virus class is the most likely to get confused as only 60% gets accurately classified. Normal class is the least likely class to be misclassified or the most likely to get right.  To decrease the model's false predictions we can increase the dataset size , or increase the threshold score. |
| --- | --- |
| **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)? | Both precision and recall are calculated for each class in the same way as described in the first model above. The overall or total precision and recall of this new multiclass model is calculated by averaging the ones for each class. |
| **F1 Score**  What is this model’s F1 score? | [Source](https://github.com/fa-ahmad/AIPMND/blob/master/ND088-MLPMND-Model%20Evaluation-Cheat%20Sheet%20.pdf).  In our case:  F1 = 2\* (0.821\*0.767) / (0.821+0.767)= 0.793 = 79.3% |