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

*NOURAH ALHASSAN*

Binary Classifier with Clean/Balanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | The total of data used in the model is: 200, 100 images of normal healthy lungs and 100 images of lungs infected with pneumonia. Out of them 160 were used for training, (10 normal ,10 pneumonia) for validation and (10,10) 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? | The grid shows all the predicted labels against the true labels. And it’s a tool to give insight and help determine where we need to improve our data to increase the model accuracy.  **False Positive** (FP) for the normal label: 0%  **True Positive** (TP) for the pneumonia label: 100% |
| **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:** Measure the TP over the total of predictions, A high precision model produces fewer false positives.  **Recall:** Measure the TP over the total of true predictions, A high recall model produces fewer false negatives. |
| **Score Threshold**  When you increase the threshold what happens to precision? What happens to recall? Why? | **Increase in the threshold:** we notice that the precision increased and the recall decreased.    **decrease in the threshold:** we notice that the precision decreased and the recall increased.  The score threshold relates to the level of confidence the model must have to assign a category for testing. it is a tool to test the impact of different thresholds for all categories in the dataset. when the threshold score is low, the model will classify more images but will run the risk of misclassifying images in the process. when the score threshold is high, the model will classify fewer images, and it will have a lower risk of misclassification. |

Binary Classifier with Clean/Unbalanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | The total of data used in the model is: 537, 156 images of normal healthy lungs and 318 images of lungs infected with pneumonia. Out of them 474 were used for training, (20,40) for validation and (20,40) for testing. |
| **Confusion Matrix**  How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix. | The FP increased by 5% and the TP slipped to 95%. Because the model is trained more on pneumonia than for normal cases. |
| **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)? | Yes, it was affected as shown in the figure above. |
| **Unbalanced Classes**  From what you have observed, how do unbalanced classed affect a machine learning model? | The model has a bias since we have more pneumonia cases than normal cases so the model tends to classify normal cases as pneumonia because it has more pneumonia data. |

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. | As shown in the figure above we see that we have more error in classification for this 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 the Precision and the Recall have decreased significantly, which mean this model has less accuracy.  The binary classifier with the highest recall and precision is the first one “Clean and Balanced Model”. |
| **Dirty Data**  From what you have observed, how does dirty data affect a machine learning model? | The dataset is highly corrupted by mislabeled data so the model has performed poorly. |

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. | The model is more likley to confuse the classes “viral penumonia” and “bactrial penmuonia”. Its more likely that it will get the normal class right as you can see in the figure above it got all the normal cases right. Hover we can also see that 10% of the viral penumoia was misclassyfied as normal.  What we can do to fix the model is:   1. Increase the trainig data uniformity. 2. Increase the score threshold which will increase the precision. |
| **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)? | Similar to how we calculated the precision and recall for the binary classes.  Precision formula: TP/(TP+FP)  Recall formula: TP/ (TP+FN)  Sample for each class: 100  P= (7/10)+(10/10)+(9/10)/3= 0.8666667  R=(7/10)+(10/10)+(9/10)/3= 0.8666667 |
| **F1 Score**  What is this model’s F1 score? | **F1**= 2\*(P\*R)/(P+R)  = 2\*(0.7512)/(1.617)  = 1.5024/1.617  = **92.91%** |