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

***Rohit Roy Chowdhury***

Binary Classifier with Clean/Balanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | 80% data was used for training & from rest 20%, 10% data used for validation and 10% data was used 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 Confusion Matrix represents how the model predicted Pneumonia cases & normal cases correctly in the Blue box & wrongly in the Grey box.  Here, the True Positive for Pneumonia class is 100% and the False Positive of 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 of the model is going to be correct.  Recall tells us, how good the model is at identifying the actual presence of particular objects in the data (i.e. – weather or not it can recognize the cases).  At threshold score of 0.5, the precision & recall both were 100%. |
| **Score Threshold**  When you increase the threshold what happens to precision? What happens to recall? Why? | After increasing the threshold to 0.88, the recall value decreased to 90% from 100%, because by increasing the threshold value we want to be more confident with our model and it will classify less x-rays to get lower risk of miss-classified x-rays. |

Binary Classifier with Clean/Unbalanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | For training 319 x-ray images were taken & for validation and testing 40, 40 x-rays images were taken respectively. |
| **Confusion Matrix**  How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix. | The confusion Matrix wasn’t affected much due to unbalanced data, though I initially expected the number of false positive & false negatives to increase. |
| **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)? | For a score threshold of 0.5, the precision & recall are both 100%, since there are no error present. |
| **Unbalanced Classes**  From what you have observed, how do unbalanced classed affect a machine learning model? | Since unbalanced dataset introduces bias in the model, here with Google’s AutoML it predicted the model correctly with precision value of 100%. However, after increasing the threshold value to 0.67, the re-call value decreases to 92.5% as shown in the screenshot below: |

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. | It affects drastically to the model, increasing False Positive cases. As a conclusion I can say due to the dirty data, it became very difficult for the model to classify the cases correctly. |
| **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? | The model’s precision & recall values went to as low as 55% using a score threshold of 0.5.  From the binary classifiers, both the Balanced & Un-balanced Clean datasets has the highest Precision & Recall. |
| **Dirty Data**  From what you have observed, how does dirty data affect a machine learning model? | From what I observed, the AutoML model struggles with finding the patterns to classify them, which negatively affects the model accuracy. |

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. | From the confusion matrix, we can conclude that, the model is most likely to confuse between Viral & Bacterial Pneumonia cases. The Normal class is most likely to get right. And, I think ore data from each class will prevent the model’s confusion. |
| **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)? | The model’s Precision is reported as 0.84 & recall value is 0.83.  Precision & Recall are calculated as:   * Precision = True Positive/(True Positive + False Positive) * Recall = True Positive/(True Positive + False Negative)   \*Took the average of the sum of precision & recall from each class. |
| **F1 Score**  What is this model’s F1 score? | This model’s F1 score is: 83.3.  \*calculated F1 score by:  F1 Score = [ 2\*Precision\*Recall / (Precision + Recall)] |