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

***Ayeni Trust Olamilekan***

Binary Classifier with Clean/Balanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | I used dataset of: 500 images   * Normal images contains: 250 images, with **Train/Test/Validation data** of: 150/50/50 respectively. * Pneumonic images contains: 250 images, with **Train/Test/Validation data** of: 150/50/50 respectively. |
| **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? | Each cells in the confusion matrix describes the correct and incorrect predictions summarized with count values and broken down by each class.  The values observed are stated in the diagram below:    Values are stated as:   1. **TP(True Positives)** = 83% 2. **FN(False Negatives)** = 17% of pneumonic cases are predicted as normal. 3. **FP(False Positives)** = No normal cases predicted as pneumonic. 4. **TN(True Negatives**) = 100%  * **The True Positive rate in the confusion matrix for the pneumonic class is: 83%** * **There is no False Positive rate for the normal class.** |
| **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 the amount of positive class predictions that actually belongs to the positive class. It evaluates the fraction of correct classified instances among the ones classified as positive. * Recall measures the number of positive class predictions made out of all positive examples in the dataset.   **PRECISION AND RECALL AT THRESHOLD OF 0.5**    **The models Precision and Recall are:** 0.90 and 0.90 respectively. |
| **Score Threshold**  When you increase the threshold what happens to precision? What happens to recall? Why? | **According to the image above an increase in the threshold causes the precision value to spike and a decline in the recall value, meaning that precision is directly proportional to the threshold and Recall is inversely proportional to the threshold.**  **WHY? It is because the number of false positives decreases and false negatives increases.** |

Binary Classifier with Clean/Unbalanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | I used dataset of: 399 images   * Normal images contains: 100 images, with **Train/Test/Validation data** of: 80/10/10 respectively. * Pneumonic images contains: 299 images, with **Train/Test/Validation data** of: 239/30/30 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 has been affected by the unbalanced data in the sense that the model is confused about the normal class making it 7% less likely to predict normal cases. Moreover, the FN in the clean/balanced confusion matrix gives us 17% for the pneumonia class and now it’s reduced to 7%. |
| **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 precision and accuracy in the image above shows a 5% increase from 90% in the first model to 95% at the same threshold of 0.5. |
| **Unbalanced Classes**  From what you have observed, how do unbalanced classed affect a machine learning model? | I have observed that the unbalanced classes gives me false representation of the models performance making it difficult for me to trust my models predictions, and this is due to the fact that Machine Leaning algorithms used for classifications are designed around the assumptions of equal number of examples for each classes. |

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 dirty data has brought down the True Positives and increased the amount of False Positives unlike my first model with equal and clean amount of data hereby causing the model to have a 0.5 confidence in the data trained. |
| **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 image above shows that the model precision and recall has drastically decreased from 90%(using clean/balanced classifier) to 60% despite using the same threshold of 0.5.  **SUMMARY OF ALL BINARY CLASSIFIERS**  **Classifiers Precision Recall**   1. Clean/balanced 90% 90% 2. Clean/unbalanced 95% 95% 3. Dirty/balanced 60% 60%  * Of all three binary classifiers it is the clean/unbalanced classifier that has the highest precision. * Of all three binary classifiers it is the clean/unbalanced classifier that has the highest recall. |
| **Dirty Data**  From what you have observed, how does dirty data affect a machine learning model? | The addition of 30% mislabeled data has reduced the confidence level of the model, making its confidence level drop to 50%. Thus, dirty data can drastically affect the performance of a model. |

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 3-class confusion matrix also known as a multi-class matrix is an N\*N dimensional table (Where N is 3) that evaluates the performance of a 3-class model.   * The model is most likely to confuse the “bacterial pneumonia” class. * The model is also most likely to get right the “normal and viral” classes right.   To improve the models accuracy I would:   1. Get more data as this would provide me with a plethora of view and versions for the dataset. 2. Identify the algorithms and data representations that performs better than my current results. 3. Getting the most out of algorithm tuning. 4. Combing predictions from multiple well-performing models 5. Increasing the threshold levels of the model also has a positive impact. |
| **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 models precision and recall values are 86.67% and 86.67% respectively   The values are calculated by using the formulas:  Precision = (TP+TN) / (Total number of points) OR  Precision = (Precision(virus) + Precision(bacterial) + Precision (normal)) / 3  Recall = (Recall(virus) + Recall(bacterial) + Recall(normal)) / 3 |
| **F1 Score**  What is this model’s F1 score? | The models F1 score is:  2 \* (precision \* recall) / precision + recall  Thus, F1 score = 2 \* (86.67 \* 86.67) / 86.67 + 86.67  F1 score = 86.67% |