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

*Ishika Jain*

Binary Classifier with Clean/Balanced Data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | |  | | --- | | Pneumonia : 135 | | Normal : 138 | | Training : 246 | | Testing : 27 | | Total: 273 | |
| **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? | A **confusion matrix** is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.. The four cells represent TP, TN, FP, FN. The TP for pneumonia class is 92% and FP rate for normal class is 8% |
| **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)? | The ratio of correct positive predictions to the total predicted positives. Recall measures the proportion of positives that are correctly identified A high recall model produces fewer false negatives. While, a high precision model produces fewer false positives. The model achieved a precision of 96.3% and recall of 96.3%. |
| **Score Threshold**  When you increase the threshold what happens to precision? What happens to recall? Why? | On increasing the score threshold, the precision goes increases and recall decrease. Since by increasing threshold we want to be more confident in making prediction. Hence by increasing the score threshold, your model will classify fewer images but it will have lower risk of misclassifying the images. |

Binary Classifier with Clean/Unbalanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | A total of 398 images were present in dataset. Out of them 358 were used in training and 40 in test. |
| **Confusion Matrix**  How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix. | The TP % went up for pneumonia. Earlier all the normal images were classified correctly. But now there seems to be a slight dip in % to 91.7%. |
| **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)? | Both precision and recall went up to 94.3% |
| **Unbalanced Classes**  From what you have observed, how do unbalanced classed affect a machine learning model? | Unbalanced Classes make the data bias. Model will have a bias towards predicting the label that has more data in training. The results shows increase in the number of images for pneumonia but decreasing the number of normal images lead to increase in no of TP for pneumonia. |

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 you see from the confusion matrix, the model is totally confused in predicting and hence performs really bad. |
| **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 precision and recall values went to a very low value of 75%. The highest precision and recall values were observed with clean and unbalanced data. |
| **Dirty Data**  From what you have observed, how does dirty data affect a machine learning model? | ML model helps to find patterns among classes as the data is mixed up. Model find similar pattern in both the labels “pneumonia” and “normal”. Hence model performs 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. | Model is most likely to confuse with normal class and a little about virus. Model will not confuse about bacteria class as can be seen from confusion matrix.  We can add more images to each class as there only 50 images for each class now. |
| **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)? | Precision and recall are calculated foreach class individually and by taking an average of them you get precision and recall values for whole model. Precision for normal, virus and bacteria classes are 20%, 80%, 100%. Similarly for recall divide the value of TP by sum of the column. Hence precision and recall values for whole model are 66.7%(((20+80+100)/300\*)100) and 66.7% respectively. |
| **F1 Score**  What is this model’s F1 score? | Model’s F1 score is 0.667. |