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

*Jude Anlasun*

Binary Classifier with Clean/Balanced Data

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for training? How much data was used for testing? | There is total of 200 images of which 100 images are labeled as “normal” and the other 100 labeled as “pneumonia”. 80% of the images was used for training. 10% was used for Validation and the remaining 10% 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? | A confusion matrix is a table that describes the performance of a classificaiton model on a set of test data for which the true values are known.  From the diagram above the different cells labeled (TP, FN, FP and FN) of the confusion matrix can be descriped based on the model as follows:   * **TP – True Positive:** cases predicted as pneumonia and were actually pneumonia. * **TN – True Negetive:** cases predicted as normal and were actually normal. * **FP – False Positive:** cases predicted as pneumonia, but were actually normal. * **FN – False Negetive:** cases predicted as normal but were actually pneumonia.   The values observed in the actual predictions can be seen in the below screenshot:    **True Positive Rate (pneumonia)** = TP/(TP+FN)  = 10 /(10+0) \*100 = **100%**  **False Positive Rate (normal)** = FP/(FP+TN)  = 0/(0+10) \* 100 = **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 often the correctly predicted observation is correct. It is expressed as the ratio of the correctly predicted positive observations to the total positive observations.  Recall measures how truly relevant results are return. It is expressed as a ratio of the number of positive class predictions to all positive observations in the dataset.  The model at a score threshold of 0.5 recorded a Precision of 100% and a Recall at 100%. |
| **Score Threshold**  When you increase the threshold what happens to precision? What happens to recall? Why? | Precision is already at 100%, increasing the score threshold decreases the recall however, the precision is still maintained at a 100%.  Assuming that the precision wasn’t already at 100% increasing the score threshold would have decreased recall score with an increase in precision score. This happens because when the score threshold is increased precision increases which produces fewer false positives with a high confidence of lower misclassification. |

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. |  |
| **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)? |  |
| **Unbalanced Classes**  From what you have observed, how do unbalanced classed affect a machine learning model? |  |

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. |  |
| **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? |  |
| **Dirty Data**  From what you have observed, how does dirty data affect a machine learning 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. |  |
| **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)? |  |
| **F1 Score**  What is this model’s F1 score? |  |

References:

[1] https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/