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

*Munirah Mohammed Aldossary*

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 5242 images for training data.  I used 582 test items, which represent more than 5% of the training data. (%5=262.1) |
| **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 first column represents the true label and the two on the right represent the predicted label for normal and pneumonia.  The first cell represents 87% which is true positive.  The second cell represents 13% which is false positive.  The third cell represents 3% which is false negative.  The fourth cell represents 97% which is true negative.  The true positive rate for pneumonia is 97%  The false positive rate for the normal class is 13%  Chart, waterfall chart  Description automatically generated |
| **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 precision measurement is the value of positive predictions, and it is equal an average of 94.16%.  The average precision is 0.992 which is 0 < 0.992 < 1  The recall measurement is the number of true positives over true positives adding to false negatives, and the relationship between the number of false and the recall is inverse relationship. |
| **Score Threshold**  When you increase the threshold what happens to precision? What happens to recall? Why? | When the score threshold is 0.5  The precision is 94.16%  The recall is 94.16%  When the score threshold is 0.9  The precision is 97.89%  The recall is 87.8%  The precision value increases and the recall decreases, because when the threshold value increase the positive predictions increase and the negative predictions decrease. |

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 100 images of normal x-ray image  And I used 300 of pneumonia x-ray images  I used 357 data total images and 40 test images |
| **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 zero false negative prediction for the normal label and the false positive prediction of the pneumonia label has decreased.  Chart, waterfall chart  Description automatically generated |
| **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 values have increased a little to 95%, because the total false predictions have decreased so the equation will give me more clear and higher than before. |
| **Unbalanced Classes**  From what you have observed, how do unbalanced classed affect a machine learning model? | When there are unbalanced classes the class that has smallest number of items will have more true positive percent, and the higher will have less false negative values. |

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 false negative and positive predictions increased in both classes.  Chart, waterfall chart  Description automatically generated |
| **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 and recall have decreased to 70%.  Class 1:  Precision = = = 0.8 x 100  Precision= 80%  Recall = = = 0.6 x 100  Recall=60%  Class 2:  Precision = = = 0.6 x100  Precision= 60%  Recall = = = 0.75 x 100  Recall= 75%  The highest precision between the binary classifiers is the clean/ unbalanced data which was 95%  And the highest recall value between the binary classifiers is clean/ unbalanced data which was 95% |
| **Dirty Data**  From what you have observed, how does dirty data affect a machine learning model? | When there are a dirty data the false positive and negative prediction will increase because the data is not clear for the machine learning model. So, we should clean the data and try to have correct input items. |

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 is a matrix of 3 classes and does not have the true positive/ negative or false positive/negative values, but we can find it by using the determinant of each class.  The most class that the model has confused in is 67% or the normal class as viral pneumonia.  The most classes were correct are viral pneumonia and bacterial pneumonia with 83%.  I would minimize the false negative values by adding class width or under-sampling the negative.  Waterfall chart  Description automatically generated |
| **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)? | From the statics we are giver that the precision of this matrix is equal to 73.33%, to find  And the recall is equal to 61.11%, to find these values:  True Positive = 17%  True Negative = 17% + 83% + 17% + 83% = 200%  False Positive = 0% + 0% = 0%  False Negative = 67% + 17% = 84%  Class 1:  TP1 = 83% = 0.83  TN1 = 17%+17%+83%+0% =117%= 1.17  FP1 = 0% +17% = 17% = 0.17  FN1 = 67% +17% = 84% = 0.84  Precision= = = 0.83  Recall= = = 0.497  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Class 2:  TP2 = 17% = 0.17  TN2 = 83%+17%+17%+83% =200% = 2  FP2 = 67% +17% = 84% = 0.84  FN2 = 0% +0% = 0% = 0  Precision= = = 17/101 = 0.2  Recall= = = 1  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Class 3:  TP3 = 83% = 0.83  TN3 = 83%+17%+67%+0% =167% = 1.67  FP3 = 0% +17% = 17% = 0.17  FN3 = 17% +17% = 34% =0.34  Precision= = = 0.83  Recall= = = 83/117 = 0.71  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  The Micro-average:  Total TP = 0.83 +0.17 +0.83 = 1.83  Total FP = 0.17 + 0.84 + 0.17 = 1.18  Total FN = 0.84 + 0 + 0.34 = 1.18  Micro-average of precision =  (Total TP )/(Total TP +Total FP)  = (1.83)/(1.83 +1.18)  = 0.61 = 61%  Micro-average of recall =  (Total TP)/(Total TP +Total FN)  = (1.83)/(1.83+1.18)  =0.61 = 61% |
| **F1 Score**  What is this model’s F1 score? | Class 1:  F1 score = 2 \*  F1 score = 2 \*  F1 Score = 0.62  Class 2:  F1 score = 2 \*  F1 Score = 0.29  Class 3:  F1 score = 2 \*  F1 Score = 0.76  \_\_\_\_\_\_\_\_\_  Micro F1 score:  Micro F1 score = Micro-average of precision = Micro  -average of recall = 61%  \_\_\_\_\_\_\_\_\_  Macro F1:  Macro F1= F1 score for class 1 + F1 score for class 2 + F1 score for class 3  Macro F1 = 0.62 + 0.29 + 0.76 = 1.67 |