AutoML Modeling Report

*Anastasia Chueva*



**Binary Classifier with Clean/Balanced Data**

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for  training? How much data was used   for testing? | 600 images were used total for the training: 300 normal images and 300 pneumonia images. 58 test images. |
| **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 cell on the crossing of “normal” row and column “normal” has a value of 100% and signifies that all images were correctly identified in the normal class.  The cell on the cross between row “pneumonia” and column “pneumonia” of value 100% signifies that 100% of images were correctly identified in the pneumonia class.  TP rate for pneumonia class 100%  FP rate for pneumonia class 0% |
| **Precision & 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 number of correct predictions vs a total number of predictions. Recall measures the percentage of correctly identified predictions.  The precision rate for the model is 100%  The recall rate for the model is 100% |
| **Score Threshold**  When you increase the score  threshold, what happens to  precision? What happens to recall?  Why? | Lowering the threshold decreases precision as seen on below screenshot    Increasing the threshold decreasing recall as seen on below screenshot    Changing the threshold for the predicted probability affects the performance of the algorithm. For example, the true positive rate, or recall, is 0 if we set the threshold as 1, as no images are classified as pneumonia/normal.    But having a recall of 1 is not necessarily good, as a model which classifies everything as spam has recall equal to 1, but also very low precision, as there will be a lot of false positives. |

**Binary Classifier with Clean/Unbalanced Data**

|  |  |
| --- | --- |
| **Train/Test Split**  How much data was used for  training? How much data was used  for testing? | 400 images were used total for the training: 300 pneumonia images and 100 normal images. 38 test images were used by the model. |
| **Confusion Matrix**  How has the confusion matrix  been affected by the unbalanced  data? Include a screenshot of the  new confusion matrix. | Due to data being unbalanced for normal class it introduced FPs and FNs within each class (pneumonia and normal). |
| **Precision & 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 decreased by ~3% |
| **Unbalanced Classes**  From what you’ve observed, how  do unbalanced classes affect a  machine learning model? | Using the lower amount of images for one class compared to another, affects negatively precision for the class that has a lower amount of data ( in my case for the normal class the precision is 85.7%) and decreases recall for the class with a higher number of images ( for pneumonia class recall is 96.9%) |

**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. | 200 images were used total for the training: 100 pneumonia images and 100 normal images. 19 test images were used by the model.    Using dirty data to train the model introduced a large amount of FPs and FNs |
| **Precision & 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? | Precision and recall have decreased by 32% when using dirty data. Among binary classifiers, best precision attained using balanced dataset (when I was using a clean dataset with even among of images within each class) |
| **Dirty Data**  From what you’ve observed, how  do dirty data affect a machine  learning model? | Using dirty data affects negatively precision and recall for both classes (labels) by lowering it on average to 68%.  It introduces FNs and FPs in each class as well. |

**3-Class Model**

|  |  |
| --- | --- |
| **Confusion Matrix**  Summarize the 3-class confusion  matrix. What classes are the model  most likely to confuse? What  class(es) is the model most likely to  get right? What might you do to try  to remedy the model’s “confusion”? | A screenshot of a cell phone  Description automatically generated  Average precision of the model is 0.95.  “Normal” class has 100% TP and no FP or FNs, in other words, all images used for testing were correctly identified.  “Viral Pneumonia” has 100% of precision and 85.7 % of recall. The model provided some FNs.  “Bacterial Pneumonia” has 88.9% of precision and 100% of recall. The model provided some FNs.  Two classes that model most likely to confuse are bacterial and viral pneumonia.  The model is most likely to get the right “normal” class as it provided 100% of recall and 100% precision on test data.  To remedy the model’s “confusion” we would need to add more images in each class (viral and bacterial pneumonia) to train the model. |
| **Precision & Recall**  What are the model’s precision and  recall? How are these values  calculated? (Report the values for a  score threshold of 0.5.) | Model’s precision is 97.4%  Models recall is 94.9%    My assumptions is that those values are calculated by following method and adjusted for number of images ( i.e. test dataset=39, where full dataset=300) and AUC (area under the curve).  Model’s precision and recall calculated as an average precision and recall for all 3 classes + some adjustment for number of images in the full dataset:  Bacterial Pneumonia  precision =TP/(TP +FP) = 8/(8+1)=0.888  recall =TP/(TP +FN) = 8/(8+0)=1  Viral Pneumonia  Precision =Tp/(TP +FP) = 8/(8+0)=1  recall=Tp/(TP +FN) = 8/(8+2)=0.8  Normal  Precision norm=Tp/(TP +FP) = 8/(8+0)=1  Recall norm=Tp/(TP +FN) = 8/(8+0)=1  Model precision (using 39 test images) = Pvp+Pbp+Pnorm= (0.888+1+1)/3 =0.96  Model recall (using 39 test images)= Rvp+Rbp+Rnorm= (0.8+1+1)/3 =0.94 |
| **F1 Score**  What is this model’s F1 score? | F1 score = 2\*precision\*recall/(precision+recall)  F1 score= 2\*0.94\*0.96/(0.96+0.94)=0.95 |