Predicting Pneumonia in Chest X-Rays

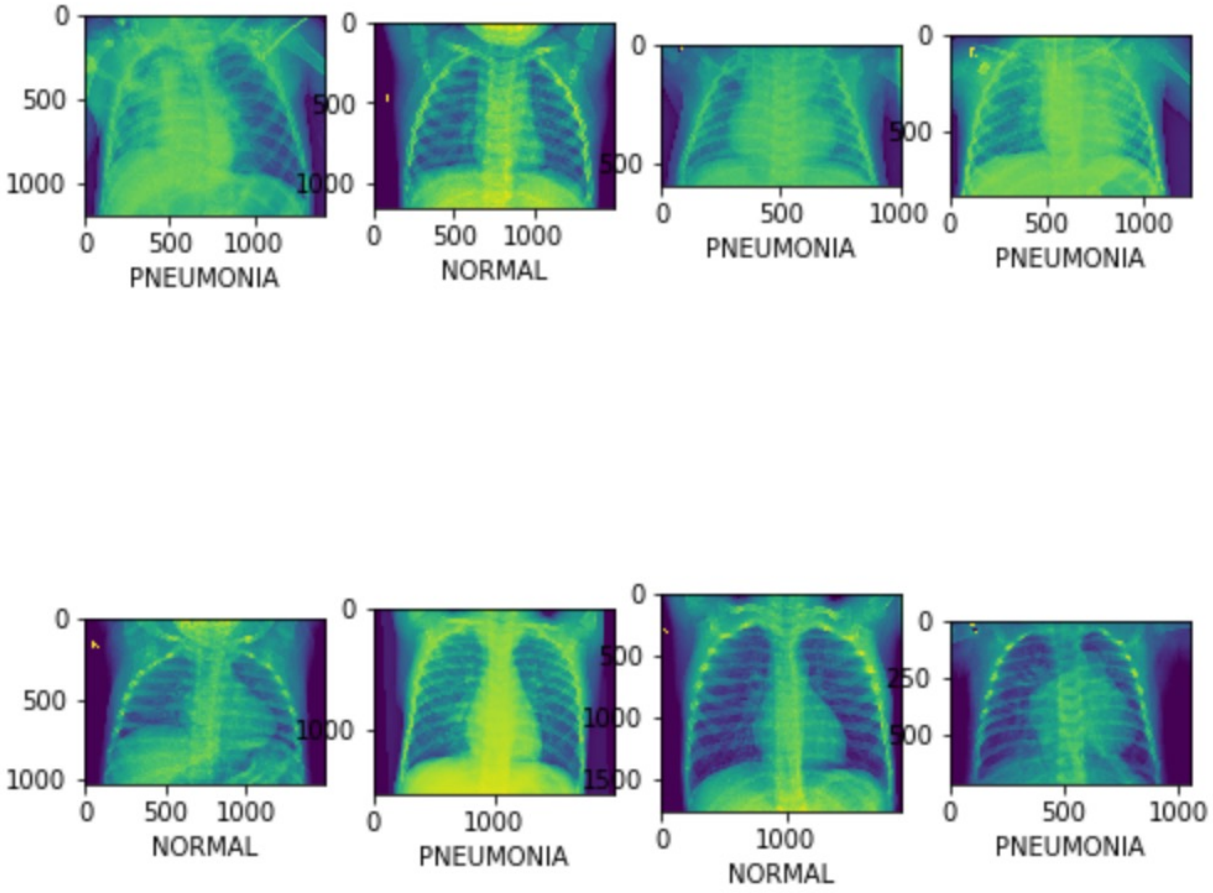
Caitlin Snyder
Flat Iron - Data Science
Module 4

Research Question

Given a set of patient chest x-rays, can we accurately predict pneumonia cases?

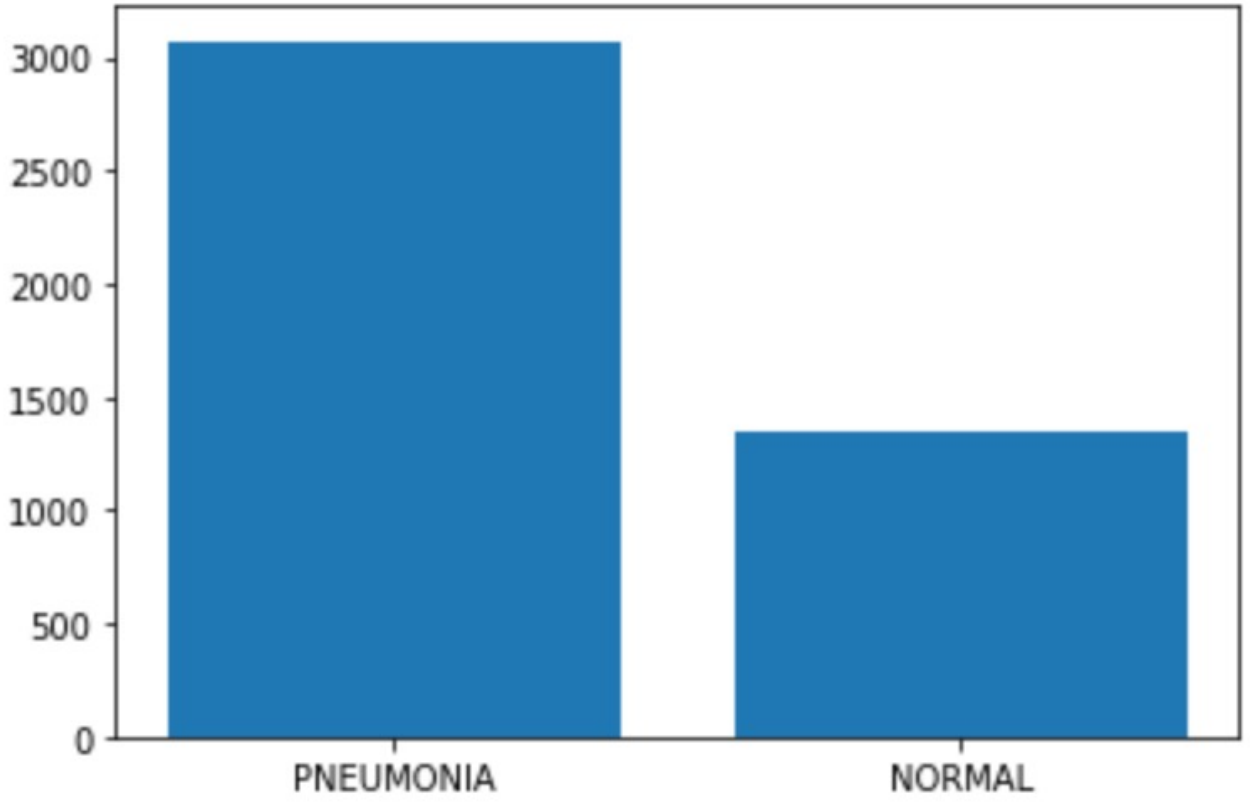
Iteration:

SageMaker Configuration



Iteration:

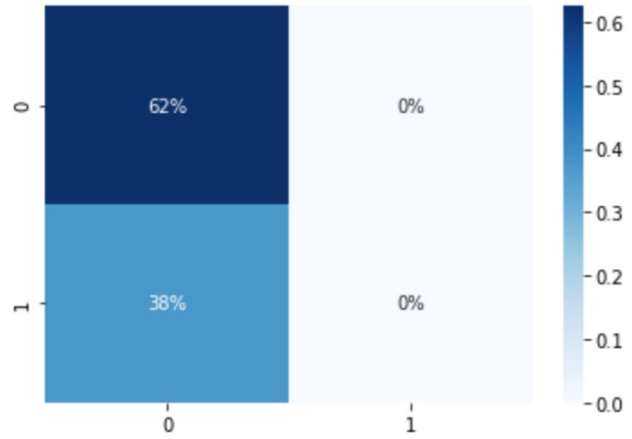
Weighting



Iteration:

Parameter Hypertuning & Weighting

0.625 %



36.38 %

Inaccurate but extremely reliable!

Results

- Despite spending about 40 hours iterating over the model training parameters and using SMOT and the RandomOver/UnderSamplers, I cannot achieve a model that predicts anything other than one single outcome class
- In other words: This model should not be relied on for diagnostic purposes.

Take-aways

- SageMaker is a useful tool for integrating cloud-based computing resources when faced with local memory constraints.
- Any model is inherently biased in its ability to predict outcomes based on the data fed into it.

Future avenues for exploration

- Would different weighting techniques yield greater accuracy?
- What types of patient data (ie, sex, smoker) are most clearly represented in the images?

Sources

- https://www.coursera.org/learn/image-classification-sagemaker/ungradedLti/fsCCK/image-classification-with-sagemaker
- https://mxnet.apache.org/versions/1.8.0/api/faq/recordio.html
- https://docs.aws.amazon.com/sagemaker/latest/dg/image-classification.html
- https://sunjackson.github.io/2018/09/13/84e31f8847e234adb7a5c3eadb253a14/
- https://github.com/aws/amazon-sagemaker-examples/issues/1885
- https://medium.com/@dtuk81/confusion-matrix-visualization-fc31e3f30fea

Thank you!