

### CatsVsDogsWeek\_8\_Write\_Up

This week we worked with an interesting problem, to correctly identify pictures of cats and dogs. It was a great data set to work with as it helped practice data cleaning and data manipulation skills again. The initial challenge was being able to streamline data from the dataset into an accurate data frame that housed all of the images. I created a few helper functions to help set checkpoints and output messages to showcase when work will be done. It was necessary to create these checkpoint messages, as for this week's assignment each query run would take a long time due to the nature of the data set.

The next few challenges was the hardware capability my laptop had. I thought I had enough space on my hard drive to be able to run these complex queries, but when I had multiple pages running and trying to run the model, it resulted in the jupyter notebook crashing due to running out of disk space. Due to this, I had to keep my models simple as well as the hyper parameter tunings but be informative to showcase the differences. Model 1 had the least amount of layers, and increased in model 2 as well. In model 3, I used the same model in model 2, yet I manipulated the epoch amount to showcase that this parameter was important in fine tuning the models.

The goodness of fit metrics showcased that the first model did not indicate a strong model, yet incrementally in model 2 and model 3, these metrics improved. Additionally, there was not much of a difference between model 2 and model 3, as the graphs and accuracy metrics do not differ a lot. I am interested to see how the models fared to the kaggle hidden test cases. The scores are below.

Submission and Description	Private Score	Public Score	Use for Final Score
<a href="#">model_3.csv</a> just now by <a href="#">Sahil Rangwala</a> <a href="#">add submission details</a>	0.64145	0.64145	<input type="checkbox"/>
<a href="#">model_2.csv</a> a few seconds ago by <a href="#">Sahil Rangwala</a> <a href="#">add submission details</a>	0.64262	0.64262	<input type="checkbox"/>
<a href="#">model_1.csv</a> a few seconds ago by <a href="#">Sahil Rangwala</a> <a href="#">add submission details</a>	0.59118	0.59118	<input type="checkbox"/>

It is interesting to note that the first model actually performed the best in comparison to the other models. To my understanding the test score is an error amount, thus the lower the score the better the model performed. This showed that the additional epochs in model 3 helped it in comparison to model 2. However, the incorporation of the additional layers brought down the overall robustness of both model 2 and 3. I am curious what the test cases test for specifically as this would help gain more insight on how model 1 performed better. Per the metrics, it did well enough but model 2 and 3 should have beat it. The model 2 and 3 could have accidentally introduced overfitting and brought down the quality of the model.