



Case Study

Deep Learning:

Based on your performance in the recent screening interview for the position of Data Science Consultant at Quantiphi, we've decided to send you a case study to test your approach to problem solving when it comes to real world data.

Please find attached the Dog Breed Identification challenge. The primary intent is to examine your thought process and approach to building deep learning solutions. The model accuracy of stored data set is important but secondary. If there are any questions, please feel free to call/email. Ideally, it'd be good to have a preliminary working model in a week's time. However, in case you do need more time, let us know.

Here is an outline of the "Dog Breed Identification" case.

Pet insurance is a rapidly growing industry in the United States. Pet owners are beginning to purchase policies that require a monthly premium to de-risk the event of paying for an expensive medical condition or procedure for their pet down the line. As this space is quickly growing, there are several companies in the market, each looking to gain an advantage by benefiting consumers.

The goal of many insurers, to gain an advantage in the market, is to ease the quote process for customers purchasing pet insurance. In the current state, they need to supply information such as the pet's age, gender, and breed. We believe that all of this information can be determined from a picture. To prove this concept is possible we would like to be able to predict a dog's breed with just a photo.

A link is provided below to a training dataset with 20-75 images of 100+ dog breeds. With this data set, a model needs to be built that can ingest a picture of a dog and classify which breed it is most likely to be.

In addition, we would like to understand from a consulting perspective, what next steps to take and what additional data is needed to be able to send a quote to the prospective customer with minimal involvement from the customer.

Once again, the most important part of your solution is not the accuracy of the model. It will be the thought process around the model (and potential improvements you could make), readability of code, and business understanding of what else will be needed to make this a complete solution that issues the prospective customer a quote.

We encourage the use of Jupyter notebooks as long as it does not restrict any of your code. They tend to separate out code and analysis/explanation well and our easier for us to evaluate your overall body of work. Please submit your findings in the notebook format if possible.

Data set:

https://storage.googleapis.com/dog_breed_classifier/dogImagesComplete.zip