Capstone Project Proposal

1. What is the problem you want to solve?

   The aim of this project is to label image data.

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

   The client needs this kind of classification for his mobile application, they want that user can organize photos on their mobile phone into different categories. For example: I want every picture with a dog.

3. What data are you going to use for this? How will you acquire this data?
   For this project I will use the Caltech256 Dataset (http://www.vision.caltech.edu/Image_Datasets/Caltech256/) it contains 30607 images in 256 categories, for example bats, bears, dogs and camels are categories.

4. In brief, outline your approach to solving this problem (knowing that this might change later)?

   - Get overview about the data
     - How do the images look?
     - How many images do I have per class?
     - Do all images have the same size?
     - Check differences in the RGB - distribution of the different classes.
   - Rescaling the images if they differ in size
   - Training different models
     - SVM (Support Vector Machine)
     - K - nearest
     - CNN (Convolutional Neural Network)
   - Compare the performance of the models
   - Compare the models with the benchmarks at the caltech website

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

   - IPyhton notebooks, containing the analysis
   - Paper, describing the analysis