Submit a zip file per group containing implementation code, description and presentation slides. If individual implementations vary, create a zip for each member.

**Code and Description (3 marks):** The 'description' can be a PDF or MS WORD file providing instructions about how to run your code, using which tool, what was the system you used (OS, processor speed) and what versions of other software is needed. Code means your program code that you developed or extended for the project. YOU MUST identify the development you did for the project. If you used a GUI based tool, provide screen shots of what you did and explain every step.

**CISC 452 Pet Adoption Prediction**

**Code Explanation**

**Code Requirements**

* Python
  + Version 3.6 (TensorFlow compatible)
* TensorFlow
  + pip install tensorflow in command prompt
  + Can install tensorflow GPU with
    - pip install tensorflow-gpu
  + Used as the back-end for Keras
* NumPy
  + pip install NumPy
  + Used mainly for data preprocessing
* Pandas
  + pip install pandas
  + For reading csv files. It handles csv files with commas in content
* Scikit-learn
  + pip install scikit-learn
  + Used for metric functions, normalizing, splitting data into equal distribution training and testing sets
* csv
  + Comes with python installation
  + Used to read csv files that do not contain commas in content
* json
  + Comes with python installation
  + Used to read json files
* glob
  + Comes with python installation
  + Used for building an iterable of file names given wildcard directory

**Running the Code**

* Ensure that all required libraries are installed
* Run the python file in idle or another editor
* You can change the running parameters by modifying the call to main() at the bottom of the python script. In its current state, the parameters are
  + Parameters include
    - String dataDirectory
      * The directory that points to the data folder
    - Boolean preprocessed
      * Whether or not the training data and testing data has already been pre-processed. This is used to speed up computation time. It only needs to be set to False if the pre-processing strategy has changed, or if the model is being deployed. (new data needs to be pre-processed).
    - Boolean retrain
      * Whether or not to retrain the network.

**System Specs**

* The model was created, trained, and tested on the system specification below.
* Intel i7 4770k @ 4.2Ghz
* 2x GTX 780 3gb
* 16gb RAM @ 1866Mhz

Note that only one GPU is utilized because the training procedure is not parallelized.