This proposal should include these sections:

## The Team

Kerry Gip

Kerry.gip@ucdenver.edu

Individual Contributor – Support Vector Machine and Convoluted Neural Networks

Odyssey Villagomez

Odyssey.villagomez@ucdenver.edu

Individual Contributor - K Nearest Neighbor and Convoluted Neural Networks

Yash More

Yash.more@ucdenver.edu

Individual Contributor – Artificial Neural Networks and Gradient Decent Backpropagation and Confusion Matrix for Evaluation

## Problem Statement and Background

Recognizing handwritten digits is a common deep learning problem that is useful for students to learn as an introduction into deep learning methods and evaluation.[1] In addition to academic interest, handwriting recognition can also be useful for commercial reasons such as reading bank checks and sorting postal mail. This digitrecognition task is supplied and evaluated by Kaggle. The task it to scan pixel images of handwritten digits and identify them as a number between 0 – 9. The success of our digit-recognition deep learning model will be based on the accuracy of the digit we predict compared to its actual value. Current existing models have been able to achieve excellent results, with a prediction error of less than 1%.[1]

This is an interesting project to work on because there are multiple applications to automating digital recognition. One application is to authenticate logins, where the user must enter the digital number displayed on the screen. We would like to see if we can 'hack' captcha recognition software, by being able to artificially recognize the digits displayed in an image. Verifying the values in the CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) has been a part of our online daily lives for most websites that require an account. Being able to see how CAPTCHA works and what is needed to either automatically enter in digits or how difficult it will be to crack will be part of the discovery process as we progress throughout this competition.

### The Data Source(s) You Intend to Use

The data for this project will be provided through Kaggle. The Modified National Institute of Standards and Technology (MNIST) provides datasets for computer vision, and we will be using their dataset of handwritten images.[2] The competition that we are participating in provides the dataset needed for us to do the competition. We will be getting the test and train datasets from them to provide our analysis

### Describe how you plan to obtain the data, or how you got it if you already have it.

The data will be provided by Kaggle and the MNIST Database of Handwritten Digits and has the training set of 60,000 samples and test set of 10,000 samples.[3] It is subset of a larger set available from MNIST. These digits have been size-normalized and centered in fixed-size images.

### Give a summary of the cleaning/joining of data that you expect to do up front.

Because the datasets will be given to us in full, we will have no need to clean or join data. We will have a train dataset and a test dataset in csv format.

### Goals of Your Analysis

This competition is evaluated based on the accuracy of our predictions with a higher accuracy indicating a better model .  The goal of the competition is to take every handwritten image from the test set and write a program that can accurately determine what that digit is.

We will be learning multiple classification algorithms to apply to this problem as well as computer vision. Since this is a competition, we need to improve our accuracy scores with every iteration until the competition closes. A measurable goal is to have an accuracy score of over 90% and an error rate comparable to the current existing models of less than 1%.

### Description of Data Analysis Tools You Plan to Use

The main tool that we will be using is Jupyter notebook. We will track each other's progress using Github, with its ability to manage version control. The dataset will be downloaded through Kaggle where the data has already been cleaned and processed. We will be training the dataset and then multiple regression and algorithm techniques that will be applied, such as K-Nearest Neighbors and Support Vector Machine (SVM). We will also be applying other machine learning and computer vision algorithms to measure different accuracy rates. They can include Principal Component Analysis(PCA), Linear/Logistic Regression, Adaboost( Decision Trees), Gradient Descent, Neural Networks or Convoluted Neural Networks. We will evaluate the other algorithms to determine which will be the best ones to use to give the highest measure of accuracy.

### Describe the Data Products Your Project Will Produce

The results will include how well we predicted the handwritten numbers using a regular accuracy score and MSE (Root Mean Squared Error).

## References

1. Brownless, Jason. Handwritten Digit Recognition using Convolutional Neural Networks in Python with Keras. Published June 27, 2016. https://machinelearningmastery.com/handwritten-digit-recognition-using-convolutional-neural-networks-python-keras/

2. Kaggle Digit Recognizer. Learn computer vision fundamentals with the famous MNIST data. https://www.kaggle.com/c/digit-recognizer/overview/description

3. Lecun, Yann, Cortes, Corinna, Burges, Christopher J.C. THE MNIST DATABASE of handwritten digits. http://yann.lecun.com/exdb/mnist/