Project Design

I used the First Impressions V2 (CSPR'17) data set to build an algorithm that takes a 15 second video clip and scores the subject on how well they made a first impression. I created a model based on the images of the video, the audio, and the words transcribed to score the first impression. I then tied together the three predictions with a final OLS model to come up with one final prediction.

Data

Researchers at the University of Barcelona put together this data set of 10,000 videos. They were transcribed and annotated through Amazon Mechanical Turk. Each video is an HD video of someone speaking into the camera for 15 seconds. All genders and ethnicities are represented and there doesn't appear to be any bias of the first impression scores solely based on those factors.

Tools

Keras/tensorflow - Used for preprocessing of image data and deep learning models
Pandas - Used to store tabular data
Numpy - Used to manipulate 3D and 4D arrays for modelling purposes
FFMPEG - Used to strip audio from mp4 files
Librosa - Used to extract audio features from mp3 files
Gensim - Used to handle word embeddings and preprocess text
CV2 - Used to convert videos to jpegs and jpegs to numpy arrays

Algorithms

Image model
I used the vgg16 model in conjunction with an LSTM to handle the images. The vgg16 model was used to extract features about the images. Then those features were fed through an LSTM to understand the temporal patterns between each still frame in the video (20 frames were extracted for each video)

Audio model
I used a random forest to predict the audio score. I used this algorithm because I didnt have a lot of time to search hyper parameters and I understand this algorithm the best. This cut down on the search space and uncertainty involved with modelling audio data

Text model
I used pre-trained word embeddings from the google news 20 word2vec model. These embeddings were fed into a bidirectional LSTM to understand the sequential nature of these transcripts.