**Udacity Classifier Variation Analysis**

This 2-pager summarily describes what was done, why it was done, and what I found. What was found takes the form of 5+ talking points.

**I. What Was Done**

A number of data augmentation services exist. This paper investigates a specific set of classifiers used for data augmentation. I use each classifier to obtain the same parameter estimate, and I compare their levels of confidence and cross-correlation. I try to identify a preferred classifier. I also check about the value-added by using a first name vs a full name, a single image vs two images, and a single sample-wide image source vs two or three. I also check whether people with unusable or missing images vary in systematic or important ways, where unusable images include images with multiple people or irrelevant images. I obtained images from Udacity, GitHub, and LinkedIn. Classifiers utilized include Kairos, NamePrism, NamSor, Kairos, Genderize, Survey, and Manual Review. In addition to reviewing for comparative efficacy and cost on commonly estimated variables, I also discuss non-common data associated with some classifiers.

**II. Why it Was Done**

Certain standard sociological controls exist. These include age, sex, and ethnicity. A number of different data augmentation services exist, which can derive estimates for those data based on other data. One common pattern is to base such estimates on the name of an individual. Another approach is to use images. When a person’s profile on a public site exists, it is sometimes associated with a useable image. It’s also often possible to find additional images about this person by searching online or by following integrated sites and taking those profile images. Obtaining additional images, however, may have a large cost, so it may not be net beneficial to obtain such data if the gains in model usefulness are small. Also, non-common data inputs or outputs, or cost considerations, might contextually optimize a different solution than the simply most accurate solution.

**III. What I Found**

**Appendix**