**Classifier Variance Analysis by LinkedIn Data**

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**

This paper comparatively examines accuracy of automated ethnicity classifiers. A survey was conducted in which respondents self-reported ethnicity and, optionally, a URL to their LinkedIn profile. Name and image data was scraped from URLs provided. NamePrism, NamSor, and Kairos ethnicity estimates were produced based on LinkedIn input data. These estimates were compared against the self-reported ethnicities. The relation of interest is any correlation between machine estimate error and other individual factors identified on the survey. Particularly, ethnicity and gender.

This approach yielded a low response rate. 35 respondents yielded 19 LinkedIn values, 12 of which were accessible. 9 of these 12 included a processable profile image. Of the 35 responses, 18 were derived from a Facebook Ad campaign, 6 were obtained through a Fiverr survey service, 10 were obtained through SurveyCircle, and 1 was obtained through a personal Facebook post. All responses were collected between May 5 and May 19, 2018.

**II. Why It Was Done**

This study was immediately motivated by a need to provide standard sociological controls in a study of online learning[[1]](#footnote-1). The online learning study involved scraping user profile data which included names and profile images, but no direct reporting of ethnicity. Using classifiers, name and image information can be input into a process which produces an estimate of ethnicity. However, prior research failed to indicate an error rate on such estimates.

Prior studies compare classifiers using a metric called agreement. Agreement occurs when two classifiers produce the same estimate. When such agreement occurs, prior researchers took this to indicate mutual validation. However, this could equally indicate mutual invalidation. Instead of looking at agreement, this study looks at applied accuracy by comparing estimates to self-reported ethnicity.

This study is practically valuable for academic and business reasons. The academic value of choosing a better estimate of ethnicity is that more accurate models and factor estimates will become available and relationships which were previously insignificant or unclear will clarify to some degree. One business application comes in the form of candidate screening. Ethnicity is a factor in a model of employee productivity, and better models of employee productivity lead to improved hiring and firing decisions which reduce turnover and other costs to business.

Some classifiers are built on samples from outside of the United States[[2]](#footnote-2). This is another reason to doubt the accuracy of such classifiers, and it becomes a particular concern in a business application in the United States. The Unites States is a leader in online education, which compounds the original motivation.

**III. What I Found**

1. Image-based classifier Kairos had an error rate of 1/9. NamePrism had an error rate of 0/12 and Namsor had an error rate of 3/12.
   1. Notice that the average image-based classifier error is not significantly different from the average name-based classifer. Classifier error was about 12%.
   2. Image-based error appears orthogonal to name-based error. They did not error on the same records. So, these two approach can be complimentary when the required input data is available.
      1. However, any time multiple factors are used to measure the same target parameter, explanatory power will be split between those factors.
      2. As a result, including multiple proxies of ethnicity may result in improper mutual non-identification of factor significance, particularly when orthogonal classifier error is small or when ethnicity has weak significance in the regression of interest to begin with.
   3. Namsor, NamePrism, and Kairos offer some similar data, but in some cases a service offers something unique. This means that NamePrism may be preferred on ethnicity estimations, but it is not generally preferred for all use cases.
2. Among 12 valid LinkedIn responses, only 1 person self-reported as non-White, and only 1 was female. These were not the same individual.
3. Classifiers may do a better job at identifying biological sex and ethnicity compared to socially identified or self-reported gender or ethnicity.
   1. Three cases support this. In one case, an individual self-reported a gender of other, but this is not a classifier-provided gender. In a second case, a man named Salah who clearly appears Middle Eastern identified as White, ostensibly because Middle Eastern was not a selectable option. One classifier identified him as African, and it is plausible he is genetically closer to the average African rather than the average White. In a final case, a Hispanic-looking individual identified as White. One classifier identified him by name as Portuguese. Brazilians are a group of Portuguese-descendent Latin Americans. It’s plausible this person, Tiago, may be Brazilian and therefore genetically closer to the average Hispanic rather than the average White person.
4. All errors occurred among samples who did not report where they live. However, they all provided linkedIn information indicating United States residence.
5. All findings in the small sample exploration are with low confidence attributable to sample size.
   1. Namsor may overestimate the likelihood of black ethnicity when applied to United States residents
   2. Kairos may overestimate the likelihood of being Hispanic or non-white among people who describe themselves as white

1. See 2-pager-udacity-github.docx [↑](#footnote-ref-1)
2. TODO: Find reference. It was NamePrism or Namsor. Were the agreement studies also on non-US samples? [↑](#footnote-ref-2)