

## **SMM639: People Analytics Key Findings**

Below are the emerging key insights of the possible sources of bias that led to misclassified adults in the incorrect economic threshold (income > or <= USD 50K). Out of a total 15060 individual entries, 16.95% was inaccurate.

### **Key Findings- Categorical**

**Race:** The largest count of incorrect predictions arising from race has been with the 'White' race whereby a total of 2290 individuals were predicted in the wrong category. This is 17.66% of the total white population from the dataset. Using the Bokeh zoom-in function, upon looking closer the Asian-pac-Islander race, despite not being a large number of our sample population, suffers from 19.12% of misclassified adults.

**Marital Status:** 28.40% of married individuals have been incorrectly predicted in the threshold, this a noteworthy aspect and the highest category amongst all the categorical variables of a misclassification instance.

**Work class category:** The ML model has inaccurately placed 15.14% of private sector employees under the wrong economic status. Local government employees and self-employed individuals also are affected by 21.49% and 21.20% respectively of errors from misclassification.

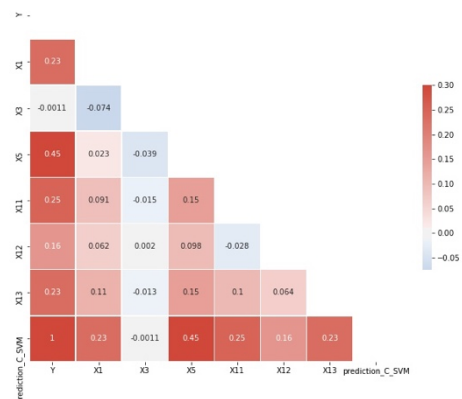
### **Key Findings- Continuous**

**Age:** Ages 30 to 45 are the ones with the highest misclassified count of individuals. Moreover, upon reaching age 60 there is a high percent of individuals that the ML model was unable to classify threshold accurately.

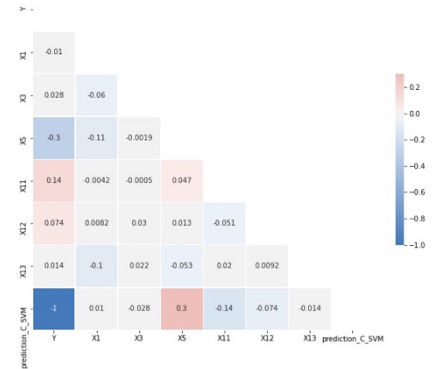
**Hours-per-week:** Individuals working between 20 - 40 hours have the most possible bias in the model and this pattern continues amongst those working more than 40 hours however with a reduced intensity.

### **Key Findings- Correlation and Confusion:**

**Classified:** Although the correlation between educ with capgain and hpw is weak, these are one of the highest correlations identified in the matrix (0.15). Individuals with higher levels of education work more hours in a week and generate higher capital gains. It is worth considering when managing employees of different education backgrounds to meet company diversity policies.



**Misclassified:** As the age of an employee increases, their education level “decreases”. This is incorrect, as education can only increase with one’s age. Although the correlation is not strong (-0.11), this is important for firms to consider maintaining a strong reputation for employment.



**Confusion metric:** True vs Predicted Above and Below threshold helps us to draw that 40.40% (1031) of the misclassified individuals were inaccurately placed above 50K income threshold, and the remaining 59.60% (1521) misclassifications were of those cases incorrectly placed below the threshold.

